HEALTHY COMMUNITIES DATA AND INDICATORS PROJECT

TECHNICAL MANUAL

By

Dulce M. Bustamante-Zamora, PhD
Neil Maizlish, PhD

Office of Health Equity
California Department of Public Health
Richmond, California

and

Diem Tran, MPH, MS

Institute for Health & Aging
University of California San Francisco School of Nursing
San Francisco, California

October 12th, 2014

ACKNOWLEDGEMENTS

The Office of Health Equity of the California Department of Public Health and the Institute for Health & Aging, University of California, San Francisco, gratefully acknowledge the Strategic Growth Council for their support of the *Healthy Communities Data and Indicators Project* (Interagency Agreement 3010-355). The following organizations and individuals generously contributed their time and expertise:

Patrick Fox and Claudia Chaufan, The Institute for Health & Aging, University of California, San Francisco
Linda Rudolph, Public Health Institute
Connie Mitchell, Steve Wirtz, Lynn Walton-Haynes, Kathy Dervin, Anne Millington, California Department of Public Health
Glen Campora, Anda Draghici, Linda Wheaton, and staff at the California Department of Housing and Community Development
Allison Joe, Deputy Director, Strategic Growth Council
Louise Bedsworth and Elizabeth Baca, Governor's Office of Planning and Research
Scott Gregory, California Technology Agency
Trish Kelley, Applied Development Economics, Inc.
Hans Harinder, Office of Highway System Information and Performance, Caltrans
Douglas Kuczynski, Demographic Research Unit, California Department of Finance
Bay Area Regional Health Inequities Initiative, Sandi Galvez, Executive Director; Amy Smith, Program Manager; Matt Beyers and Bradley Jacobson, Data Committee Co-Chairs; Jennifer Henn, Napa County Public Health Division; Pam Stoddard and Anandi Sujeer, Santa Clara County Public Health Department; Rochelle Ereman, Marin County Public Health Department; Abigail Kroch, Contra Costa Health Services; and other Bay Area health departments
Tracy Delaney and Carla Blackmar, Public Health Alliance of Southern California; Data Committee, Danyte Mockus and Travers Ichinose, Chairs; Deirdre Browner, San Diego County Public Health Services; and other southern California local health departments
Kathleen Grassi, Cindy Valencia, and Kelley Rose, Merced County Department of Public Health
Andrew Deckert, MD, Health Officer; Stephanie Taylor, Epidemiologist; and Benjamin O'Neil, Epidemiologist; Shasta County Department of Public Health
Jake McKenzie, Councilperson, City of Rohnert Park
Barbara Payne, Councilperson, City of Galt
Brendan Reed, City of Chula Vista
Jonathan London, Cassie Hartzog, and Teri Greenfield, Center for Regional Change, University of California, Davis
Elisa Arias, Ping Chang, and Dan Gallagher, Southern California Association of Governments
Rye Baerg, Safe Routes to School National Partnership
Chanell Fletcher, ClimatePlan
Susan Stuart, Alyssa Roy, and Rosni Shah, Santa Clara County Public Health Department
Ashley Kranz and Dennis Parker, San Diego County Health and Human Services Agency

Suggested Citation

Bustamante-Zamora DM, Maizlish N, Tran D. *Healthy Communities Data and Indicators Project: Technical Manual*. Richmond, CA and San Francisco, CA: California Department of Public Health Office of Health Equity and University of California, San Francisco, Institute for Health & Aging; 2014. www.cdph.ca.gov/programs/Pages/HealthyCommunityIndicators.aspx

Version No.	Date	Changes/Comments	Author
1	9/30/14	First version of Technical Manual	Dulce Bustamante-Zamora
2	10/12/14	Second version	Neil Maizlish

TABLE OF CONTENTS

Acronyms

ABC	California Department of Alcohol and Beverage Control
ACS	American Community Survey
ARB	Air Resources Board
BLS	U.S. Bureau of Labor Statistics
CalEPA	California Environmental Protection Agency
CalTrans	California Department of Transportation
CDC	Centers for Disease Control and Prevention
CDE	California Department of Education
CDPH	California Department of Public Health
CHAS	Consolidated Planning Comprehensive Housing Affordability Strategy
CHIS	California Health Interview Survey
CHP	California Highway Patrol
DOF	California Department of Finance
EPA	U.S. Environmental Protection Agency
FBI	Federal Bureau of Investigations
HCDIP	Healthy Communities Data and Indicators Project
HUD	U.S. Department of Housing and Urban Development
LAUS	Local Area Unemployment Statistics
MPO	Metropolitan Planning Organization
MTC	Metropolitan Transit Commission
NHTS	National Household Travel Survey
SACOG	Sacramento Area Council of Governments
SANDAG	San Diego Association of Governments
SCAG	Southern California Association of Governments
SGC	Strategic Growth Council
SWITRS	Statewide Integrated Traffic Records System
TIMS	Transportation Injury Mapping System
USDA	U.S. Department of Agriculture

HEALTHY COMMUNITIES DATA AND INDICATORS PROJECT (HCDIP)

Background

Chronic illness and injury are the leading causes of death and disability in California, and chronic illnesses account for an ever-growing share of escalating health care costs. Major risk factors for obesity, chronic illness, and injury include poor nutrition, sedentary lifestyles, smoking, and alcohol use. Each of these risk behaviors is profoundly influenced by people's social, physical, and economic environments. Chronic illness and injury are also the key drivers of very significant and persistent inequities in health outcomes, including substantial differences in life expectancy and years of potential life lost based on income level and race/ethnicity.

The creation of healthy social, economic, and physical environments that promote healthy behaviors and positive health outcomes requires coordination and collaboration across multiple sectors, including transportation, housing, education, agriculture and others. Statistical metrics, or indicators, are needed to help local, regional, and state public health and partner agencies assess community environments and plan for healthy communities that optimize public health. Currently, there is neither a clearly identified set of indicators nor ready access to data that can be used to plan for healthy communities and evaluate the impact of policy and environmental changes on health.

Goals and Objectives

The goal of *Healthy Communities Data and Indicator Project* (HCDIP) is to provide a standardized set of statistical measures (indicators), data files, and tools for planning healthy communities and evaluating the impact of plans, projects, policy, and environmental changes on community health. A list of 60 indicators has been proposed, and, for 21, data files are available at the CDPH website (http://www.cdph.ca.gov/programs/Pages/Healthy CommunityIndicators.aspx). The data files can be filtered for multiple California geographies (census tract, city, county, or region), time periods between 2000 and 2010, and race/ethnicity stratifications. Accompanying the data files are Illustrative .pdf reports using templates that integrate explanatory text, maps, graphs, and tables for selected geographies. The procedures for producing the indicator data files and templates for explanatory text, graphs, maps, and tables are documented in this technical manual. The underlying data model that supports the HCDIP is provided in Figure 1. The basic strategy of the data model is to manage all project content and administrative data in a coordinated relational database, or meta-database. Such a relational database can facilitate the reporting of indicators in a web interface.

With funding by the Strategic Growth Council (SGC), the HCDIP is a 2-year collaboration of the California Department of Public Health (CDPH) and the University of California, San Francisco (UCSF) to create and disseminate indicators linked to the Healthy Communities Framework (Figure 2). The main tasks of the project are:

1. Identify a standardized, core set of valid indicators that define a healthy community
2. Identify methods to construct indicators at different geographic scales (e.g. census tract, zip code, city, county, etc.)
3. Disseminate technical documentation that allows local, county, regional, and state stakeholders to produce indicators
4. Develop a multi-agency plan for centralized data collection, analysis, and reporting of indicators, and
5. Create a demonstration website that stakeholders and CDPH can use to pilot test selected healthy community indicators.

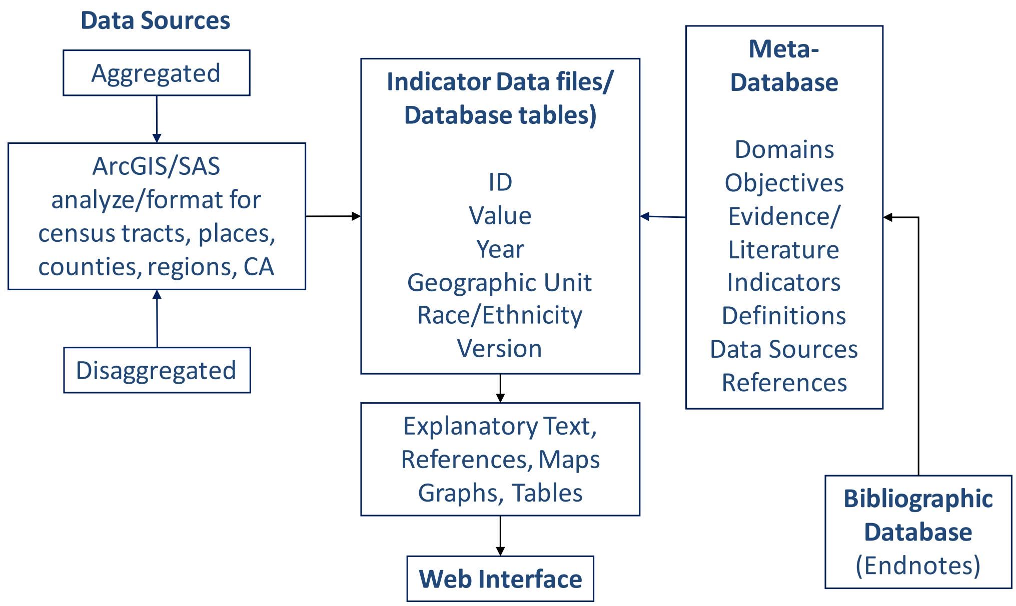


Figure 1. Conceptual Data Model for Healthy Communities Data and Indicators Project

The project builds partnerships with California organizations whose work encompasses public health, health equity, and municipal, regional, and statewide planning, and coordinates with other SGC-funded data and indicator projects.

Healthy Communities Framework and Aspirational Goals

The Healthy Communities Framework was developed by the SGC Health in All Policies Task Force with extensive public discussion and input from community stakeholders and public health organizations. The framework identifies 20 key attributes or aspirational goals of a healthy community through all stages of life, clustered in five broad categories (Figure 2): meets basic needs of all (housing, transportation, nutrition, health care, livable communities, physical activity), environmental quality and sustainability, adequate levels of economic and social development, health and social equity, and social relationships that are supportive and respectful.

Description of Indicators and Process of Selection

Indicators are usually a population-based measurement with a numerator, denominator, and time period, ideally capable of being stratified by population characteristics (race/ethnicity for assessment of inequity) or geography.

$$Indicator = \frac{What\ Happened?}{To\ Whom?}\ and\ When\ and\ Where?$$

The HCDIP identified approximately 60 indicators that address the Healthy Community Framework (Table 1) for the state of California. When possible, the indicators in the project included data for at least three points in time between 2000 and 2010, five geographic levels (8,057 Census tracts, 1,523 cities/towns and Census designated places, 58 counties, 14

What is a Healthy Community?

A Healthy Community provides for the following through all stages of life:

MEETS BASIC NEEDS OF ALL

- Safe, sustainable, accessible, and affordable transportation options
- Affordable, accessible and nutritious foods, and safe drinkable water
- Affordable, high quality, socially integrated, and location-efficient housing
- Affordable, accessible and high quality health care
- Complete and livable communities including quality schools, parks and recreational facilities, child care, libraries, financial services and other daily needs
- Access to affordable and safe opportunities for physical activity
- Able to adapt to changing environments, resilient, and prepared for emergencies
- Opportunities for engagement with arts, music and culture

QUALITY AND SUSTAINABILITY OF ENVIRONMENT

- Clean air, soil and water, and environments free of excessive noise
- Tobacco- and smoke-free
- Green and open spaces, including healthy tree canopy and agricultural lands
- Minimized toxics, green house gas emissions, and waste
- Affordable and sustainable energy use
- Aesthetically pleasing

ADEQUATE LEVELS OF ECONOMIC AND SOCIAL DEVELOPMENT

- Living wage, safe and healthy job opportunities for all, and a thriving economy
- Support for healthy development of children and adolescents
- Opportunities for high quality and accessible education

HEALTH AND SOCIAL EQUITY

SOCIAL RELATIONSHIPS THAT ARE SUPPORTIVE AND RESPECTFUL

- Robust social and civic engagement
- Socially cohesive and supportive relationships, families, homes and neighborhoods
- Safe communities, free of crime and violence

Figure 2. Healthy Communities Framework, California Health in All Policies Task Force, 2010 (http://sgc.ca.gov/docs/HiAP_Task_Force_Report-_Dec_2010.pdf)

regions, state), and race/ethnicity stratification (nine groups, including total). For some indicators other geographical levels like Census county divisions and Consolidated Metropolitan Statistical Areas where included depending on data availability and stakeholder input. Throughout this document the geographical levels will be referred to as Census tract (CT), place (PL, includes cities and towns), county division (CD), counties (CO), Consolidated Metropolitan Statistical Areas (R4), regions (RE), and the state (CA).

The following criteria were used to select the indicators:

- Validity
    - Measures what it purports to measure
    - Evidence linking indicator to health outcomes

- Technical and Data Properties
    - Data source(s) owned and collected by a recognized organization
    - Timeliness (time lag and frequency of updates)
    - Data quality (completeness, missing data, accuracy)
    - Variety of geographic levels available, including Census tract
    - Administrative accessibility (public domain, proprietary, confidentiality, costs)
    - Current use and acceptability to stakeholders
    - Straightforward mechanics of data collection, aggregation, and reporting

- Usable and Understandable to Users

Table 1. What is a Healthy Community and its indicators? (v3.10-12-14)

Aspirational Goal/Domain	Indicators
Meets basic needs of all	
Safe, sustainable, accessible and affordable transportation options*	✓ Annual number of fatal and severe road traffic injuries per population and per miles traveled by transport mode
	✓ Annual miles traveled by occurrence and by mode
	✓ Percent of residents aged 16 years and older mode of transportation to work
	✓ Percent of population residing within <½ a mile of a major transit stop
	• Percent of household income spent on transportation
	✓ Percent of population aged 16 years or older whose commute to work is 10 minutes /day or more by walking or biking
Affordable, accessible and nutritious foods	✓ Average cost of a market basket of nutritious food items relative to income
	• Percent of population within ½ mile of a full-service grocery store, fresh produce market, or store with fresh produce
	• Percent of adults who consume ≥5 servings of fruits & vegetables a day
	✓ Modified retail food environment index (mRFEI)
Affordable, high quality, socially integrated and location-efficient housing	✓ Percent of households paying more than 30% (or 50%) of monthly household income towards housing costs
	✓ Percent of household overcrowding (> 1.0 persons per room) and severe overcrowding (> 1.5 persons per room)
	• Household by type of family and head of household
	• Neighborhood Completeness Index (<½ mile radius for 7 out of 9 common public services and 9 of 12 common retail services)
	• Housing to jobs ratio
	• Jobs:housing match (e.g., Percent of adult working population who could find a job that matches their general occupational qualifications within a specified travel radius of their residence)
	• Neighborhood change: 5-year change in number of households by income and race/ethnicity
	• Residential racial segregation: isolation index
Affordable, accessible and high quality health care	• Percent of adults aged 18 - 64 years without health insurance
	• Patient satisfaction rating by medical group
Complete and livable communities	• Neighborhood Completeness Index (<½ mile radius for 7 out of 9 common public services and 9 of 12 common retail services)
Access to affordable and safe opportunities for physical activity	• Percent of adults getting moderate/vigorous daily exercise
	• Percent of children scoring 6 of 6 on Fitnessgram

Table 1. What is a Healthy Community and its indicators? (v3.10-12-14)

Aspirational Goal/Domain	Indicators
Able to adapt to changing environments, resilient, and prepared for emergencies	• Cities and counties with adopted climate action plans and FEMA-approved local hazard mitigation plans • Environmental resilience index (index composed of places with climate action and hazard mitigation plans and other Healthy Community Indicators -unemployment, lacking health insurance, educational attainment, income inequality, and registered voters-)
Opportunities for engagement with arts, music and culture	• Per capita revenue in nonprofit arts organizations • Percent of workers employed in artistic occupations
Quality and sustainability of environment	
Clean air, soil and water, and environments free of excessive noise	✓ Annual average number of unhealthy days of ozone air pollution ✓ Annual mean ambient concentration of fine particulate matter (PM2.5)
	• Percent of households/population near busy roadways
	• Average daily water use per capita
	✓ Percent of the population served by community water systems not meeting regulations of the Safe Drinking Water Act
	• Average daytime and night time outdoor noise levels*
Tobacco and smoke free	• Prevalence of smoking in adults and youth
Green and open spaces, including agricultural lands	✓ Percent of population within ½ mile of park, beach, open space, or coastline
	• Acres of parkland per 1,000 residents
	• Acres of cropland converted to developed land
	• Tree canopy coverage (urban areas)
Minimized toxics, GHG emissions and waste	• Toxicity-weighted concentrations of modeled chemical releases to air from facility emissions and off-site incineration
	• Total pounds of selected active pesticide ingredients (filtered for hazard and volatility) used in production-agriculture per square mile
	• Annual per capita greenhouse gas (GHG) emissions
	• Total waste diversion (per capita disposal rate)
Affordable and sustainable energy use	• Energy costs relative to household income*
	• Percent of electricity from renewable sources
Adequate levels of economic, social development	
Living wage, safe and healthy job opportunities for all	✓ Unemployment rate: percent of the population in the labor force who are unemployed
	✓ Overall, concentrated, and child (0 to 18 years of age) poverty rate
	• Number and rate of fatal and nonfatal occupational injuries by industry
	✓ Living wage and percent of families with incomes below the living wage
Support for healthy development of children and adolescents	• Percent of children who are kindergarten ready* • Percent of resilient high school students
	• Number of licensed daycare center slots per 1,000 children aged 0-5 years

Table 1. What is a Healthy Community and its indicators? (v3.10-12-14)

Aspirational Goal/Domain	Indicators
Opportunities for high quality and accessible education	• Mean score of Academic Performance Index (API) ✓ High school or greater educational attainment in the population aged 25 years and older
Health and social equity	
	• Race/ethnicity equity score as a composite of multiple core indicators, including median income ✓ Distribution of household income relative to the number of households, expressed on a 0 to 1 scale (Gini Index) • Place-based equity score as a composite of multiple core indicators calculated for census tracts
Social relationships that are supportive and respectful	
Robust social and civic engagement	✓ Percent of adults (18 years or older) who are registered voters; percent of adults who voted in general elections
Socially cohesive and supportive relationships, families, homes and neighborhoods	• Number of forcible rapes per 100,000 population • Percent of children (<18 years) reported with neglect or physical or sexual abuse
Safe communities, free of crime and violence	✓ Number of violent crimes per 1,000 population ✓ Percent of the population within ¼ mile of alcohol outlets by type of establishment sales

• Indicator under construction
✓Indicator available at http://www.cdph.ca.gov/programs/Pages/HealthyCommunityIndicators.aspx
* Indicator might not be feasible to construct

Indicator Vetting Process with Potential Users

The HCDIP engaged potential users in the development of the indicators. A learning community of potential users in local, regional, and state government from the Bay Area, Southern California, and rural counties was consulted to jointly pilot a set of indicators, taking into account the priorities of the participating organizations, data availability, timeliness, geographic scale, users' resources and capacity, and end uses. Three pilot projects were conducted between September 2012 and March 2014:

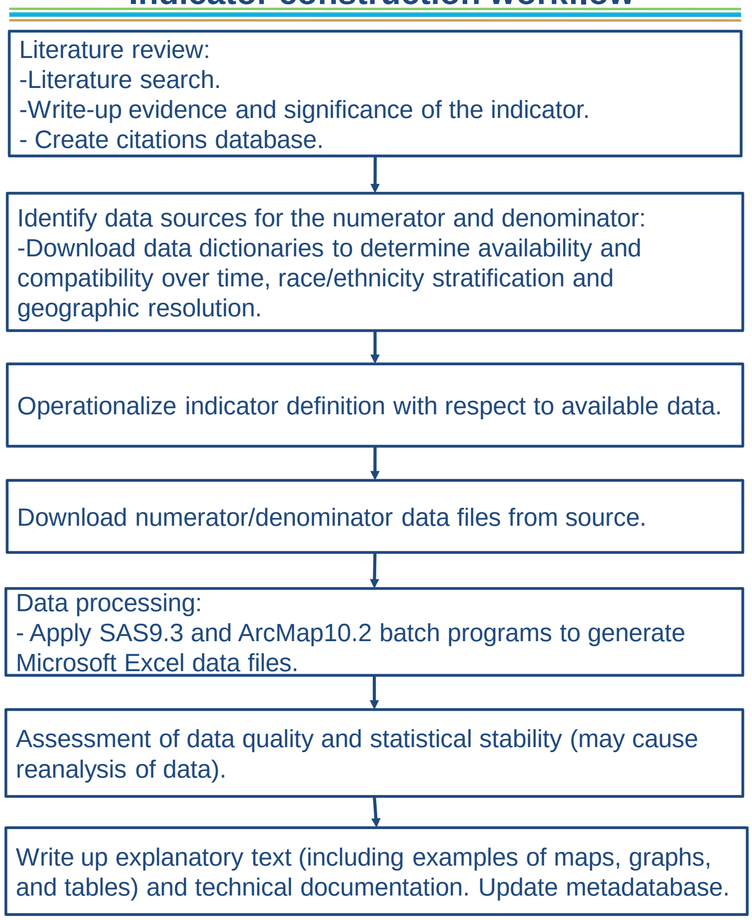
- Bay Area Pilot co-sponsored by the Bay Area Regional Inequities Initiative (BARHII)
  - Dates: September 2012 to January 2013
  - Participants:
    - Mayor of Rohnert Park
    - Planner from Association of Bay Area Governments
    - Epidemiologists from Santa Clara, Alameda, San Mateo, and Napa County Health Departments
- Southern California Pilot co-sponsored with Public Health Alliance of Southern California (PHASoCal)
  - Dates: March to October 2013
  - Participants:
    - Epidemiologists from Long Beach, Pasadena, Riverside, San Diego, Los Angeles, Orange, Santa Barbara, and Ventura Health Departments
    - Public Health Institute
    - Coalition for Clean Air
    - Safe Routes to School National Partnership
    - San Diego Association of Governments (SANDAG)
    - Southern California Association of Governments (SCAG)
- Rural Pilot
  - Dates: February to March 2014
  - Participants:
    - Shasta County Health Department
    - Merced County Health Department
    - UC Davis Center for Regional Change
    - Councilperson from City of Galt, Sacramento County

Methodology Overview

Indicator Construction Workflow

The HCDIP followed a similar workflow for the construction of all indicators. Variations to this workflow occurred when preprocessing of the source data was necessary.

Indicator construction workflow



<u>Data Sources</u>

Most indicators required the utilization of a single source of data but others required two or more sources. In most cases, organizations provided data in a machine readable format (*.csv, *.txt, *.xls, *.xlsx) that were downloaded from a website. Other organizations provided data in such formats only upon request. Reformatting of these files into an orthogonal layout was necessary in some cases. For other sources data were only be available in non-readable formats (*.doc, *.pdf) and preprocessing was necessary to convert the data into a readable format. In some cases data were only available at the organization's website and manual extraction of the tables (using copy and paste) or web-scrapping SAS programs were necessary.

The following is a list of all the sources utilized up to date for the completed indicators:

1. California Air Resources Board, Air Monitoring Network (http://www.arb.ca.gov/aaqm/mldaqsb/amn.htm)
2. California Department of Alcohol Beverage Control (http://www.abc.ca.gov/)
3. California Department of Finance, Demographic Research Unit (http://www.dof.ca.gov/research/demographic/)

4. California Department of Public Health, Division of Drinking Water and Environmental Management, Drinking Water Program (http://www.cdph.ca.gov/programs/Pages/DWP.aspx)
5. California Department of Transportation, Office of Highway System Information and Performance, Highway Performance Monitoring System (http://www.dot.ca.gov/hq/tsip/hpms/index.php)
6. California Protected Areas Database (http://projects.atlas.ca.gov/projects/cpad)
7. California Secretary of State, Elections, Voter Registration Statistics (http://www.sos.ca.gov/elections/elections_u.htm)
8. Centers for Disease Control and Prevention, Division of Nutrition, Physical Activity, and Obesity (http://www.cdc.gov/obesity/resources/reports.html)
9. Federal Bureau of Investigations, Uniform Crime Reports (http://www.fbi.gov/about-us/cjis/ucr/ucr)
10. Living Wage Calculator website (http://livingwage.mit.edu/)
11. Metropolitan Planning Organizations
    a. Metropolitan Transportation Commission (http://www.mtc.ca.gov/)
    b. San Diego Association of Governments (http://www.sandag.org/)
    c. Sacramento Area Council of Governments (http://www.sacog.org/)
    d. Southern California Association of Governments (http://www.scag.ca.gov/Pages/default.aspx)
12. Statewide Database, University of California Berkeley Law, Center for Research (http://statewidedatabase.org/index.html)
13. Transportation Injury Mapping System (data from the Statewide Integrated Traffic Records System, California Highway Patrol) (http://tims.berkeley.edu)
14. U.S. Bureau of Labor Statistics, Local Area Unemployment Statistics (http://www.bls.gov/lau/)
15. U.S. Census, Census 2000 and American Community Survey (http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml (see Appendix A for Census products commonly used by HCDIP)
16. U.S. Department of Agriculture, Center for Nutrition Policy and Promotion, Cost of Food at Home (http://www.cnpp.usda.gov/USDAFoodCost-Home.htm)
17. U.S. Department of Housing and Urban Development, Consolidated Planning Comprehensive Housing Affordability Strategy (http://www.huduser.org/portal/datasets/cp.html)
18. U.S. Department of Transportation, National Household Travel Survey (http://nhts.ornl.gov/)

The specific indicators for which each data source was used are presented in the individual indicator chapters.

Data Processing: Geocoding and Aggregation

Some indicators required geographic analysis using the geospatial processing program ArcMap10.2 (2013, ESRI, Redlands, CA) before a population-based metric could be obtained. For instance, indicators measuring the percent of the population living in proximity to transit stops or alcohol outlets first required the geocoding of these stops or outlets, respectively. The second step was to create “crow’s fly” buffers around the geocoded features and later estimate the population within those buffers by allocating U.S. Census blocks centroids to the buffers. In other cases it was necessary to geocode the place of occurrence of an event such as the location of traffic collisions, in order to determine the number and rate of injuries in a geographic area. Other data preprocessing was conducted in Excel or in the analytical software SAS 9.3 (2012, SAS Institute Inc., Cary, NC).

After preprocessing, data was analyzed in SAS. The complexity of the analysis varied with indicator and data source. An example of SAS coding for the "Poverty Rate" indicator can be found in Appendix B. For some indicators estimates and their margins of error were available from the U.S. Census and no further analysis was necessary to build the indicator, except for the calculation of the standard error and the population weighted average at the regional level. In other cases, Census estimates were used as the numerator and denominator of the indicator which could be a ratio, proportion or a percent; standard errors then had to be calculated using the approximate method http://www.census.gov/acs/www/Downloads/data_documentation-/Accuracy/ACS_Accuracy_of_Data_2013.pdf). Data aggregation was necessary to construct indicators that were linear combinations of estimates over multiple Census categories; standard errors of the linear combinations were also calculated using the approximate method.

The numerator and denominator data for some indicators originated from different sources. For example, the rate of road traffic injuries used data on the number of injuries from the California Highway Patrol in the numerator and the total population in a geographic area in the denominator from the U.S. Census.

Data Quality, Statistical Reliability, and Data Suppression

For indicators requiring aggregation or calculated weighted averages, manual checks were performed to confirm that the calculations were error-free.

When standard errors were published by the data source, the statistical reliability of the estimates was assessed using the relative standard error (RSE), which is calculated by dividing the standard error of the estimate by the estimate itself (or 1– estimate for proportions > 0.5), then multiplying that result by 100. Statistical reliability of an estimate is an expression of the relative amount of error that is deemed acceptable for the practical use of the estimate. A low amount of error with respect to the estimate is desirable. In the HCDIP we consider a threshold of RSE < 30% to indicate statistically reliable estimates.

When the majority (≥ 50%) of the population lived in Census tracts with unreliable estimates, the indicator data for that geographical level was not included in the data files. The same applied for cities and towns. More information can be found in the chapter Reliability Analysis of the Toolbox section.

Meta-Database and Data Dictionary

The HCDIP maintained a Microsoft Access database to manage the metadata giving information about each indicator. Additionally, each indicator data file included a data dictionary describing the variables in the indicator file. A description of the metadata and the data dictionary are presented in the *Metadata and Project Standards* sections of this manual.

Purpose and Structure of this Manual

The intended audience of this technical manual is epidemiologists or data analysts interested in replicating the construction of the indicators. It presents definitions, data sources, data preprocessing, and methodological details on the construction of the 21 indicators completed up to date. It also presents data sources and methodological notes about the indicators under construction. Each indicator is presented as a separate chapter in the manual in two separate sections “Completed Indicators” and “Indicators under Construction.”

Each indicator chapter has the following structure:

1. Short indicator title
2. Definition
    a. Full indicator title
    b. Indicator id number (internal use number for unambiguous identification of indicators)
    c. Healthy community framework domain and aspirational goal (Figure 2)
    d. Description of significance and health connection
    e. Summary of evidence
    f. Key references
    g. Detailed definition: numerator, denominator, strata, interpretation
    h. Data description: sources, years available, updating frequency, geographies available
    i. Methodology
    j. Limitations or comments
3. Data processing
4. Data acquisition and preprocessing

Items 3 and 4 are not included for indicators under construction, and items 2.d - 2.f are replaced by a short rationale for the selection of the indicator.

The indicator chapters will often make reference to acronyms, Census table types, methodological tools, formulas, or standards that are described in other sections of the manual.

The “Toolbox” section presents some of the methodological tools used to obtain raw data or to preformat the data, and the formulas for the calculation of standard errors, confidence intervals, relative standard error, regional level estimates, deciles, relative risk, and reliability analysis.

The “Healthy Communities Data and Indicators Project Standards” section presents the standards of all the components of the project including the basic structure of the SAS programs, output files, geographical lookup files, meta-database and data dictionary, maps, graphs, and tables.

The “Appendices” section presents lists of acronyms, Census table types, and other descriptive lists that are relevant for the HCDIP.

COMPLETED INDICATORS

Annual Miles Traveled by Occurrence

Definition

Full Title: Annual miles traveled by occurrence and by mode

Indicator ID: 39

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Safe, sustainable, accessible and affordable transportation options

Description of Significance and Health Connection:

Miles traveled by individuals and their choice of mode – car, truck, public transit, walking or bicycling – have a major impact on mobility and population health. Miles traveled by automobile offers extraordinary personal mobility and independence, but it is also associated with air pollution, greenhouse gas emissions linked to global warming, road traffic injuries, and sedentary lifestyles. Active modes of transport – bicycling and walking alone and in combination with public transit – offer opportunities for physical activity, which has many documented health benefits. Risks of injury in traffic collisions are greatest for motorcyclists, pedestrians, and bicyclists and lowest for bus and rail passengers. Minority communities bear a disproportionate share of pedestrian-car fatalities; Native American male pedestrians experience 4 times the death rate Whites or Asian pedestrians, and African-Americans and Latinos experience twice the rate as Whites or Asians. Miles traveled is influenced by affordability and quality of mode choices. Car ownership is lower in low-income and minority populations, who use a greater share of public transit and may spend a proportionally larger amount of their income and time budget on transportation than higher income groups. Increased time burden may increase stress of daily living. However, use of public transport is associated with increased walking.

Summary of Evidence:

Emissions from motor vehicles powered by fossil fuels are proportional to vehicle miles traveled and account for approximately one-third of California's annual emissions of air pollutants such as fine particulates and precursors of ozone. These air pollutants have established links to increased mortality, hospital admissions, and other adverse health effects. Numerous epidemiological studies have documented that physical activity, including that related to walking and bicycling, decrease risks of cardiovascular disease and stroke, colon and breast cancer, and dementia and depression. Miles traveled is also associated with road traffic injuries, although injury rates of bicyclists and pedestrians tend to level off as their miles traveled and mode share increases.

Key References:

- California Air Resources Board. *Estimated Annual Average Emissions*, California. Sacramento, CA: California Air Resources Board. 2008. Accessed July 19th 2013.
- McKenzie B, Rapino, M. *Commuting in the United States: 2009*. U.S. Census Bureau. Washington, DC. 2011. Accessed July 19th 2013.

- Tran HT, Alvarado A, Garcia C, Motallebi N, Miyasato L, Vance W. *Methodology for Estimating Premature Deaths Associated with Long-term Exposures to Fine Airborne Particulate Matter in California (Draft: Staff Report).* Sacramento, CA: California Air Resources Board. 2009. Accessed August 16th, 2012.
- Woodcock J, Edwards P, Tonne C, Armstrong BG, Ashiru O, Banister D, et al. Public health benefits of strategies to reduce greenhouse-gas emissions: urban land transport. *Lancet* 2009; 374(9705):1930-1943.
- Jacobsen PL. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Injury Prevention* 2003; 9(3): 205–209.

Detailed Definition:

- Annual miles traveled per capita
  - Numerator: annual number of miles traveled by mode of transportation
  - Denominator 1: total population
  - $Indicator\ 1\ (ratio) = \frac{Numerator}{Denominator\ 1}$
- Annual miles traveled per square mile
  - Numerator: annual number of miles traveled by mode of transportation
  - Denominator 2: area in square miles
  - $Indicator\ 2\ (ratio) = \frac{Numerator}{Denominator\ 2}$
- Annual miles traveled (cities/towns)
  - Indicator: annual number of miles traveled by mode of transportation
- Stratification: three modes of transport (pedestrian, bicycle, motor vehicles)
- Interpretation: healthier communities will show lower rates of annual vehicle miles traveled (AVMT), and higher rates of pedestrian/bicycle miles traveled.

Data Description:

- Data sources:
  - Numerator:
    - Pedestrian/Bicycle: U.S. Department of Transportation, Federal Highway Administration, 2009 National Household Travel Survey (NHTS) (http://nhts.ornl.gov).
    - Motor vehicles: California Public Road data (CPR); Division of Research, Innovation and System Information; Office of Highway System Information & Performance; Highway Performance Monitoring System; California Department of Transportation (http://www.dot.ca.gov/hq/tsip/hpms/datalibrary.php).
  - Denominator 1: Historical Population and Housing Estimates for Cities, Counties, and the State, 2000-2010, Demographic Research Unit, Department of Finance (DOF) (http://www.dof.ca.gov/research/demographic/reports/estimates/e-8/2000-10/view.php). American Community Survey (ACS) 5-year estimates 2006-2010 Census tract population data, table B01003 (http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml).
  - Denominator 2: 2010 Census Gazetteer files, U.S. Census Bureau, http://www.census.gov/geo/maps-data/data/gazetteer2010.html.
- Years available: 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2002-2004, 2005-2007, 2006-2010, 2008-2010
- Updated: annually

- Geographies available: cities/towns, consolidated metropolitan statistical area (CMSA), county, region (derived), and state.

Methodology:

Numerator: Data on the daily vehicle miles traveled for cities, counties, and the state were obtained from CPR. Vehicles included cars, trucks, buses, and motorcycles. Daily miles traveled were multiplied by 365 to estimate annual miles. Annual miles traveled by bicyclists and pedestrians for CMSAs (county clusters for major metropolitan areas) and California was estimated between 2002 and 2010 from the NHTS by applying the annual rate of change between 2001 and 2009. Denominator 1: Annual population counts from 2002 to 2010 for California's 481 incorporated cities and 58 counties were abstracted from DOF reports. July-centered population counts were obtained by averaging January counts for consecutive years, except for 2010, which used an April enumeration. Three-year population averages were also calculated for 2002-2004, 2005-2007, and 2008-2010. A five year population average was also calculated for 2006-2010. Denominator 2: Data on the land area in square miles for places and counties was obtained from the 2010 Census Gazetteer files. Decile rankings of counties were calculated. Regions were based on counties of metropolitan transportation organizations (MPO) as reported in the 2010 California Regional Progress Report (Appendix C).

Limitations:

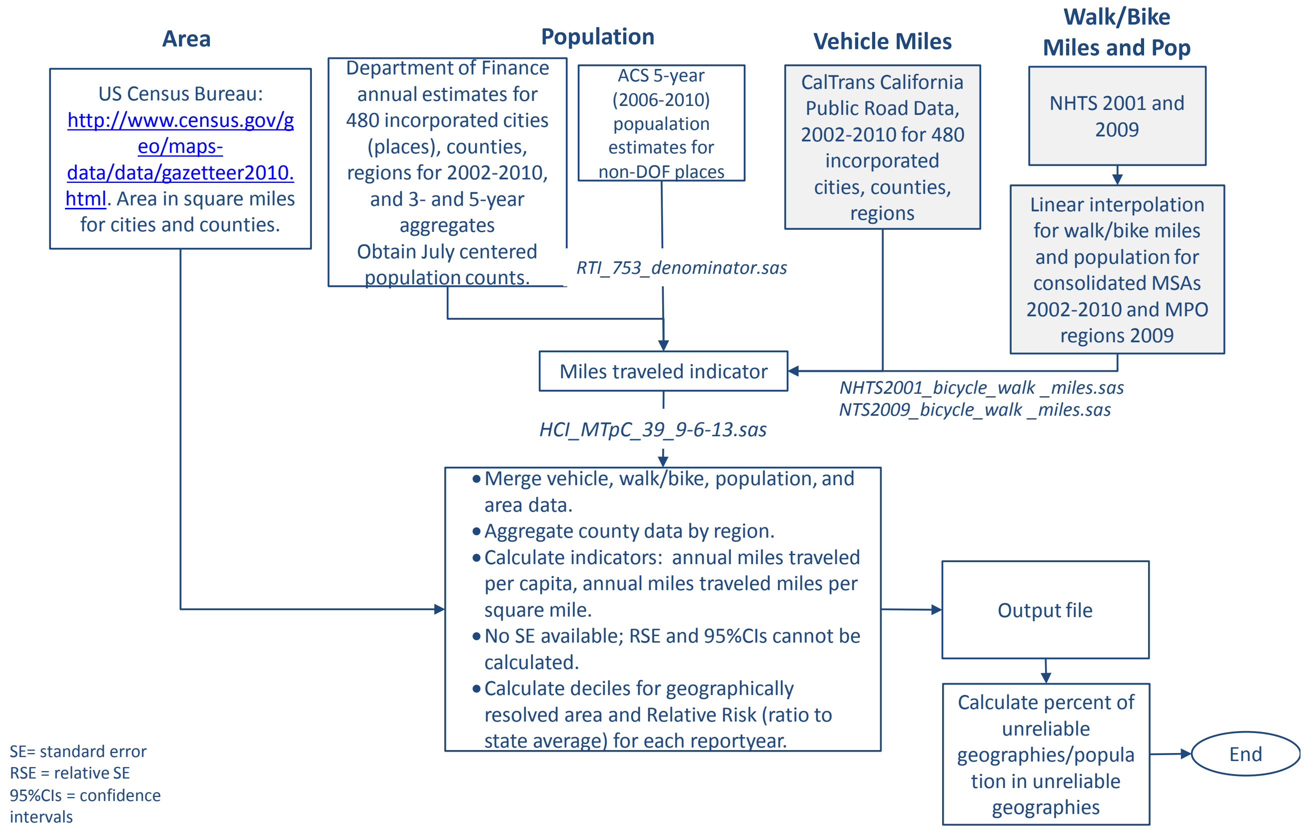
Data were not available to disaggregate different motorized modes (car, bus, truck, motorcycle), stratify by race/ethnicity, or to calculate standard errors. Data were not statistically stable for pedestrians and bicyclist at city geographies. City-specific estimates do not include miles traveled in unincorporated areas of counties. The annual miles traveled data is by occurrence and the population data is by residence; because of this discrepancy, it is not recommended to construct this indicator for cities since it could result in misleading outcomes (for example, cities with small populations but high through mileage could overestimate annual miles per capita).

Data Processing

Four programs were used to preformat the source data and to create the indicator file.

1. Numerator and Denominator preformatting: SAS file: RTI_753_denominator_7-15-13.sas (Note: this file was created for the Road Traffic Injuries indicator, for which population and annual vehicle miles travelled are both denominators)
2. Bicycle/Pedestrian Annual Data: SAS files: NHTS2001_bicycle_walk_miles.sas, NHTS2009_bicycle_walk_miles.sa
3. Indicator: SAS file: HCI_MTpC_39_9-6-13.sas. Daily vehicle miles traveled are imported into this file and converted to annual vehicle miles travelled (daily * 365).

Data Processing: Miles traveled by occurrence



<u>Data Acquisition and Preprocessing</u>

Numerator:

1. Data on the annual miles traveled by bicycle or walking for the years 2001 and 2009 was obtained from the National Household Travel Survey for CMSAs and the state. Linear interpolation was used to obtain estimates of annual miles traveled and population for the years 2002 to 2008.
2. Data on daily vehicle miles traveled (DVMT) for cities and counties in California were available from the California Public Road Data (PRD) library in .pdf format (Table 6 of the reports). PRD is published by the Division of Research, Innovation and System Information (DRISI); Office of Highway System Information & Performance; Highway Performance Monitoring System (HPMS); California Department of Transportation: http://www.dot.ca.gov/hq/tsip/hpms/datalibrary.php.
3. Hans Harinder (harinder.hans@dot.ca.gov) from DRISI was contacted to obtain the data in Excel format. Table 6 of the PRD reports for years 2000-2007 and 2010-2011 was obtained in Excel format. Data for years 2008 and 2009 could not be obtained in Excel format, thus data was extracted from the .pdf files by converting the .pdf file into a Microsoft Word file, using Adobe Acrobat Professional. Each table in the Word document was then copied and pasted into an Excel file.

4. Once all tables 6 for all years were in Excel format they were manually formatted into orthogonal tables (Figure 3).

Original table

	A	B	C	D	E	F	G	H	I	J	K
1					TABLE 6						
2											
3			2001 MAINTAINED MILEAGE & DAILY VEHICLE MILES OF TRAVEL ESTIMATES BY JURISDICTION								
4											
5								DAILY VEHICLE MILES			
6					MAINTAINED MILES			OF TRAVEL (DVMT) [1,000]			
7		COUNTY	JURISDICTION	RURAL	URBAN	TOTAL		RURAL	URBAN	TOTAL	
8											
9		ALAMEDA									
10		cities:	ALAMEDA	0.00	120.86	120.86		0.00	496.03	496.03	
11			ALBANY	0.00	27.15	27.15		0.00	112.31	112.31	
12			BERKELEY	0.00	222.32	222.32		0.00	938.03	938.03	
13			DUBLIN	3.84	61.94	65.78		34.08	165.82	199.91	
14			EMERYVILLE	0.00	19.75	19.75		0.00	93.62	93.62	
15			FREMONT	0.00	420.57	420.57		0.00	1,430.18	1,430.18	
16			HAYWARD	0.00	236.64	236.64		0.00	1,225.06	1,225.06	
17	14		LIVERMORE	0.84	251.58	252.41		4.74	710.99	715.74	
18			NEWARK	0.00	97.63	97.63		0.00	392.08	392.08	
19			OAKLAND	0.00	809.89	809.89		0.00	3,485.82	3,485.82	
20			PIEDMONT	0.00	43.60	43.60		0.00	91.93	91.93	
21			PLEASANTON	3.50	165.41	168.91		19.62	791.41	811.03	
22			SAN LEANDRO	0.00	173.28	173.28		0.00	700.72	700.72	
23			UNION CITY	0.00	148.32	148.32		0.00	439.74	439.74	
24											
25		other:	COUNTY (UNINCORPORATED)	223.76	251.16	474.92		465.35	946.22	1,411.57	
26			PORT OF OAKLAND	0.00	0.40	0.40		0.00	0.50	0.50	
27			STATE HIGHWAY	32.19	175.21	207.41		2,976.87	19,592.68	22,569.55	
28			STATE PARK SERVICE	1.00	0.00	1.00		0.10	0.00	0.10	
29											
30		ALAMEDA Total		265.13	3,225.72	3,490.85		3,500.77	31,613.13	35,113.90	
31		ALPINE									
32			BUREAU OF INDIAN AFFAIRS	1.10	0.00	1.10		0.07	0.00	0.07	
33			COUNTY (UNINCORPORATED)	133.49	0.00	133.49		34.43	0.00	34.43	
34			STATE HIGHWAY	82.71	0.00	82.71		129.52	0.00	129.52	
35			STATE PARK SERVICE	3.95	0.00	3.95		0.38	0.00	0.38	
36			US FOREST SERVICE	64.61	0.00	64.61		2.05	0.00	2.05	
37											
38		ALPINE Total		285.86	0.00	285.86		166.45	0.00	166.45	

Orthogonal table

	A	B	C	D	E	F	G	H	I	J	K
1	reporty	geotype	county_name	geoname	type	Maintained	Maintained	Maintained	DVMTRural	DVMTUrban	DVMTTotal
1184	2001	PL	Alameda	Alameda	city	0.00	120.86	120.86	0.00	496027.79	496027.79
1185	2001	PL	Alameda	Albany	city	0.00	27.15	27.15	0.00	112314.50	112314.50
1186	2001	PL	Alameda	Berkeley	city	0.00	222.32	222.32	0.00	938028.20	938028.20
1187	2001	PL	Alameda	Dublin	city	3.84	61.94	65.78	34083.66	165824.84	199908.50
1188	2001	PL	Alameda	Emeryville	city	0.00	19.75	19.75	0.00	93618.55	93618.55
1189	2001	PL	Alameda	Fremont	city	0.00	420.57	420.57	0.00	1430183.78	1430183.78
1190	2001	PL	Alameda	Hayward	city	0.00	236.64	236.64	0.00	1225060.89	1225060.89
1191	2001	PL	Alameda	Livermore	city	0.84	251.58	252.41	4743.28	710992.49	715735.77
1192	2001	PL	Alameda	Newark	city	0.00	97.63	97.63	0.00	392075.30	392075.30
1193	2001	PL	Alameda	Oakland	city	0.00	809.89	809.89	0.00	3485815.20	3485815.20
1194	2001	PL	Alameda	Piedmont	city	0.00	43.60	43.60	0.00	91932.57	91932.57
1195	2001	PL	Alameda	Pleasanton	city	3.50	165.41	168.91	19624.50	791406.35	811030.85
1196	2001	PL	Alameda	San Leandro	city	0.00	173.28	173.28	0.00	700716.63	700716.63
1197	2001	PL	Alameda	Union City	city	0.00	148.32	148.32	0.00	439735.42	439735.42
1198	2001	PL	Amador	Amador	city	3.30	0.00	3.30	1171.19	0.00	1171.19
1199	2001	PL	Amador	Ione	city	0.73	10.77	11.50	410.17	8281.66	8691.83
1200	2001	PL	Amador	Jackson	city	28.20	0.00	28.20	12474.31	0.00	12474.31
1201	2001	PL	Amador	Plymouth	city	7.00	0.00	7.00	3759.40	0.00	3759.40
1202	2001	PL	Amador	Sutter Creek	city	18.76	0.00	18.76	9961.36	0.00	9961.36
1203	2001	PL	Butte	Biggs	city	10.90	0.00	10.90	7167.40	0.00	7167.40
1204	2001	PL	Butte	Chico	city	12.30	169.55	181.86	4477.02	726483.49	730960.51
1205	2001	PL	Butte	Gridley	city	22.82	0.00	22.82	21624.89	0.00	21624.89
1206	2001	PL	Butte	Oroville	city	0.00	73.97	73.97	0.00	196669.25	196669.25
1207	2001	PL	Butte	Paradise	city	0.25	97.68	97.93	670.75	366508.94	367179.69
1208	2001	PL	Calaveras	City of Angels	city	29.08	0.00	29.08	12590.40	0.00	12590.40
1209	2001	PL	Colusa	Colusa	city	27.60	0.00	27.60	11223.64	0.00	11223.64
1210	2001	PL	Colusa	Williams	city	15.10	0.00	15.10	18178.28	0.00	18178.28
1211	2001	PL	Contra Costa	Antioch	city	0.00	227.70	227.70	0.00	622401.42	622401.42
1212	2001	PL	Contra Costa	Brentwood	city	0.00	65.01	65.01	0.00	80196.06	80196.06
1213	2001	PL	Contra Costa	Clayton	city	0.00	37.97	37.97	0.00	72348.83	72348.83
1214	2001	PL	Contra Costa	Concord	city	0.00	337.44	337.44	0.00	1417883.72	1417883.72
1215	2001	PL	Contra Costa	Danville	city	0.00	140.80	140.80	0.00	441513.37	441513.37
1216	2001	PL	Contra Costa	El Cerrito	city	0.00	72.95	72.95	0.00	156016.91	156016.91
1217	2001	PL	Contra Costa	Hercules	city	0.00	52.30	52.30	0.00	85547.19	85547.19
1218	2001	PL	Contra Costa	Lafayette	city	0.00	93.73	93.73	0.00	348387.41	348387.41
1219	2001	PL	Contra Costa	Martinez	city	0.00	110.75	110.75	0.00	352795.88	352795.88
1220	2001	PL	Contra Costa	Moraga	city	0.00	51.55	51.55	0.00	153922.60	153922.60
1221	2001	PL	Contra Costa	Oakley	city	3.00	115.87	118.87	10402.55	139917.56	150320.11
1222	2001	PL	Contra Costa	Orinda	city	0.00	93.06	93.06	0.00	286996.87	286996.87
1223	2001	PL	Contra Costa	Pinole	city	0.00	52.96	52.96	0.00	163865.46	163865.46
1224	2001	PL	Contra Costa	Pittsburg	city	1.70	137.82	139.52	13291.69	550082.41	563374.10
1225	2001	PL	Contra Costa	Pleasant Hill	city	0.00	116.73	116.73	0.00	462160.01	462160.01
1226	2001	PL	Contra Costa	Richmond	city	0.00	260.23	260.23	0.00	911489.48	911489.48
1227	2001	PL	Contra Costa	San Pablo	city	0.00	48.89	48.89	0.00	242697.95	242697.95

Figure 3. Comparison of original table layout and orthogonal table layout for Public Roads Data table 6.

5. The orthogonal tables for years 2002 to 2010 were combined into a single Excel file and the following edits were conducted:
   a. Only the “cities” jurisdictions were retained in the reduced orthogonal table as places (geotype=PL). The “other” jurisdictions (i.e., Bureau of Indian Affairs, State Highway, State Park Service, etc.) were removed, although they are included in the county total.
   b. County totals were retained in the orthogonal table as county (geotype=CO).
   c. State totals were retained in the orthogonal table as California (geotype=CA).

d. All names were transformed from all uppercase letters to first uppercase letter follow by lower case letters using the “proper” command in Excel.
e. Name adjustments:
    i. Angels Camp was changed to City of Angels.
    ii. El Paso de Robles was changed to Paso Robles.
    iii. Santa Marino was changed to San Marino.
    iv. Bell Flower was changed to Bellflower.
    v. Carmel by the sea was changed to Carmel.
    vi. Saint Helena was changed to St. Helena.
    vii. La Canada-Flintridge was changed to La Canada.
    viii. West Morland was changed to Westmorland.
f. The final orthogonal table had 6,430 records.
g. Other observations:
    i. The places Bear Valley (Alpine), Broadmoor (San Mateo), Stallion Springs (Kern), Rolling Hills (Los Angeles), and Kensington (Contra Costa), did not appear in the PRD.
    ii. There was no DVMT data for the following places-years: Arcata (Humboldt County, 2008), Rancho Cordova (Sacramento, 2002-2003), and Menifee and Wildomar (Riverside, 2002-2008).

6. The variable of interest among those included in the table was Daily Vehicle Miles Traveled (Total). The DVMT were manually multiplied by 1,000 in the Excel file as part of the formatting, given that data in the table was given in units that had been divided by 1,000. (Note: the DVMT was later converted to annual vehicle miles traveled (AVMT) in the SAS program).
7. More information about the PRD and the survey methods can be found in Chapter 6 of the March 2013 Highway Performance Monitoring System Field Manual http://www.fhwa.dot.gov/policyinformation/hpms/fieldmanual/ explains the sampling method.
    a. All roadway functional classes, including local public roads, are included in the DVMT.
    b. There are 13 motor vehicle classifications included in the PRD (Figure 3): Classes 1 to 3 - Motorcycles, Passenger Cars & Pickup trucks; classes 4 to 7 - Single Unit Trucks; and classes 8 to 13 – Combination Trucks.

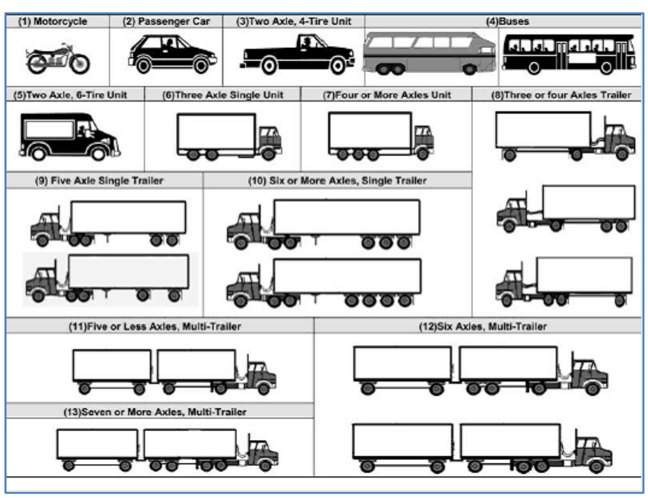


Figure 4. Motor Vehicle Classifications included in the Public Roads Data.

8. Data on the annual miles traveled by pedestrians and bicyclist for the years 2001 and 2009 was obtained from the National Household Travel Survey (NHTS, http://nhts.ornl.gov/download.shtml). The NHTS person and trip files were downloaded. Annual miles traveled by bicyclists and pedestrians for CMSAs (county clusters for major metropolitan areas) and California was estimated between 2002 and 2010 from the NHTS by applying the annual rate of change between 2001 and 2009.

Denominator 1: Total population (and group quarters population) from the Department of Finance

Population estimates (including population in group quarters) for California, its counties and cities for the years 2002-2010 was obtained from Douglas Kuczynski (Douglas.Kuczynski@dof.ca.gov) from the Department of Finance (DOF). The data obtained are equivalent to the file E8_2000-2010_Report_ByYear_Final_EOC.xls available at http://www.dof.ca.gov/research/demographic/reports/estimates/e-8/2000-10/view.php. City population estimates for 2010 for April and January, and January estimates for all other years (2000-2009) were available in the file. July-centered population counts were obtained by averaging January estimates for consecutive years. For example, the population for July 2002 was obtained as follows:

$$Population_{Jul02} = \frac{Population_{Jan02} + Population_{Jan03}}{2}$$

Denominator 2: Area in square miles

Data on the area in square miles for places and counties were obtained from the U.S. Census Bureau gazetteer files website: http://www.census.gov/geo/maps-data/data/gazetteer2010.html. Data for cities (http://www.census.gov/geo/maps-data/data/docs/gazetteer/2010_place_list_06.txt) and counties (http://www.census.gov/geo/maps-data/data/docs/gazetteer/counties_list_06.txt) were copied from the website and pasted into .txt files. The .txt files were imported into Excel. The variable of interest for the denominator was ALAND_SQMI which contained area information in square miles.

Summary of data sources and indicator availability by geography and by mode

Table 2 presents the variables used to construct the indicator by data source, and Table 3 presents the indicators reported by geography and by mode.

Table 2. Data Sources and Variables used for the Annual Miles Traveled by Occurrence Indicator

Data Source	Variable denominator	Variable numerator
California Public Road Data (Vehicles)		Daily vehicle miles traveled (DVMT) converted to annual vehicle miles traveled (AVMT)
Department of Finance	July centered total population	
Census Gazetteer	Area in square miles (ALAND_SQMI)	
National Household Travel Survey (Bicycle, Walking)		Annual miles traveled
National Household Travel Survey (Bicycle, Walking)	Total population	

Table 3. Annual Miles Traveled by Occurrence Indicator Reported by Geography and by Mode

Geography	Pedestrian/Bicycle		Motor vehicles	
	Annual miles traveled per capita	**Annual miles traveled per square mile**	**Annual miles traveled per capita**	**Annual miles traveled per square mile**
City			Y (numerator only)	
CMSA	Y			
County			Y	Y
Region			Y	Y
State	Y		Y	Y

Mode of Transport to Work

Definition

Full Title: Percent of residents aged 16 years and older mode of transportation to work

Indicator ID: 42

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Safe, sustainable, accessible and affordable transportation options

Description of Significance and Health Connection:

Commute trips to work represent 19% of travel miles in the United States. The predominant mode – the automobile - offers extraordinary personal mobility and independence, but it is also associated with health hazards, such as air pollution, motor vehicle crashes, pedestrian injuries and fatalities, and sedentary lifestyles. Automobile commuting has been linked to stress-related health problems. Active modes of transport – bicycling and walking alone and in combination with public transit – offer opportunities for physical activity, which is associated with lower rates of heart disease and stroke, diabetes, colon and breast cancer, dementia and depression. Risks of injury in traffic collisions are greatest for motorcyclists, pedestrians, and bicyclists and lowest for bus and rail passengers. Minority communities bear a disproportionate share of pedestrian-car fatalities; Native American male pedestrians experience 4 times the death rate of white or Asian pedestrians, and African-Americans and Latinos experience twice the rate of whites or Asians.

Summary of Evidence:

Emissions from motor vehicles account for approximately 1/3 of California's annual emissions of air pollutants. Among them, fine particulates and precursors of ozone have established links to increased mortality, hospital admissions, exacerbation of asthma symptoms, and other adverse health effects in numerous epidemiologic studies. The risk of road traffic injuries is strongly related with the mode of transportation: motorcyclists, pedestrians and bicyclists experience 2-20 times the fatal injury rate as car occupants. Numerous epidemiological studies have documented that physical activity decreases risks of cardiovascular disease and stroke, colon and breast cancer, and dementia and depression. Active transport increases opportunities for physical activity. Several recent studies associated long commutes in automobiles (>60 minutes) with decreases in aggregate health-related activities and reduced time for sleep, physical activity, and food preparation.

Key References:

- California Air Resources Board. Estimated Annual Average Emissions. Sacramento, CA: California Air Resources Board; 2008. Accessed July 19th, 2013.
- Tran HT, Alvarado A, Garcia C, Motallebi N, Miyasato L, Vance W. Methodology for Estimating Premature Deaths Associated with Long-term Exposures to Fine Airborne Particulate Matter in California (Draft: Staff Report). Sacramento, CA: California Air Resources Board; 2009. Accessed August 16th, 2012.

- Drechsler D, Garcia C, Tran H, Mehadi A, Nystrom M, Propper R, et al. Review of the California Ambient Air Quality Standard For Ozone. Vol 4. Sacramento, CA: California Environmental Protection Agency, Air Resources Board; 2005. Accessed January 4th, 2013.
- Centers for Disease Control and Prevention. CDC Recommendations for Improving Health through Transportation Policy. National Center for Environmental Health; 2008. Accessed November 2nd, 2013.
- Beck LF, Dellinger AM, O'Neil ME. Motor vehicle crash injury rates by mode of travel, United States: using exposure-based methods to quantify differences. Am J Epidemiol, 2007; 166(2): 212-218.
- Santos A, McGuckin N, Nakamoto HY, Gray D, Liss S. Summary of Travel Trends: 2009 National Household Travel Survey. Washington, DC: Federal Highway Administration; 2011. Accessed November 2nd, 2013.

Detailed Definition:

- Numerator: number of workers 16 years and older by mode of transportation
- Denominator: number of workers 16 years and older
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: eight race/ethnicity groups (African American, American Indian Alaska Native –AIAN-, Asian, Latino, multiple, Native Hawaiian Other Pacific Islander –NHOPI-, other and white); five modes of transportation (bicycle, car-truck-van, public transportation, walked, and worked at home).
- Interpretation: healthier communities will show (1) a lower percentage of workers using motorized vehicle to commute; (2) a higher percentage of workers walking, biking or taking public transportation to work which indicates higher opportunities for physical activity.

Data Description:

- Data source: U.S. Census Bureau, Census 2000 and American Community Survey (ACS), http://factfinder2.census.gov/.
- Years available: 2000, 2005-2007, 2008-2010, 2006-2010.
- Updated: 3 and 5 year intervals
- Geographies available: Census tracts, cities/towns, counties, regions (derived), and state

Methodology:

Data from the Census 2000 (table PCT055) and the ACS (tables DP03 and B08301) were downloaded from http://factfinder2.census.gov. The modes of transportation included: 1) bicycle, 2) car-truck-van, 3) public transportation, 4) walked, and 5) worked at home. The car-truck-van category was subdivided into carpooled or drove alone. The denominator was the total population aged 16 years and older that had a paid job in the week previous to the survey, and the numerator was the number of people within that population using each mode. For 2005-2007, 2008-2010, and 2006-2010, Table B08301 was used for the car-truck-van and bicycle modes, and Table DP03 was used for other modes. The percent of residents mode of transportation and its standard error were calculated from population counts of the numerator and denominator ($\times$ 100) using binomial approximation or abstracted directly from Table DP03.

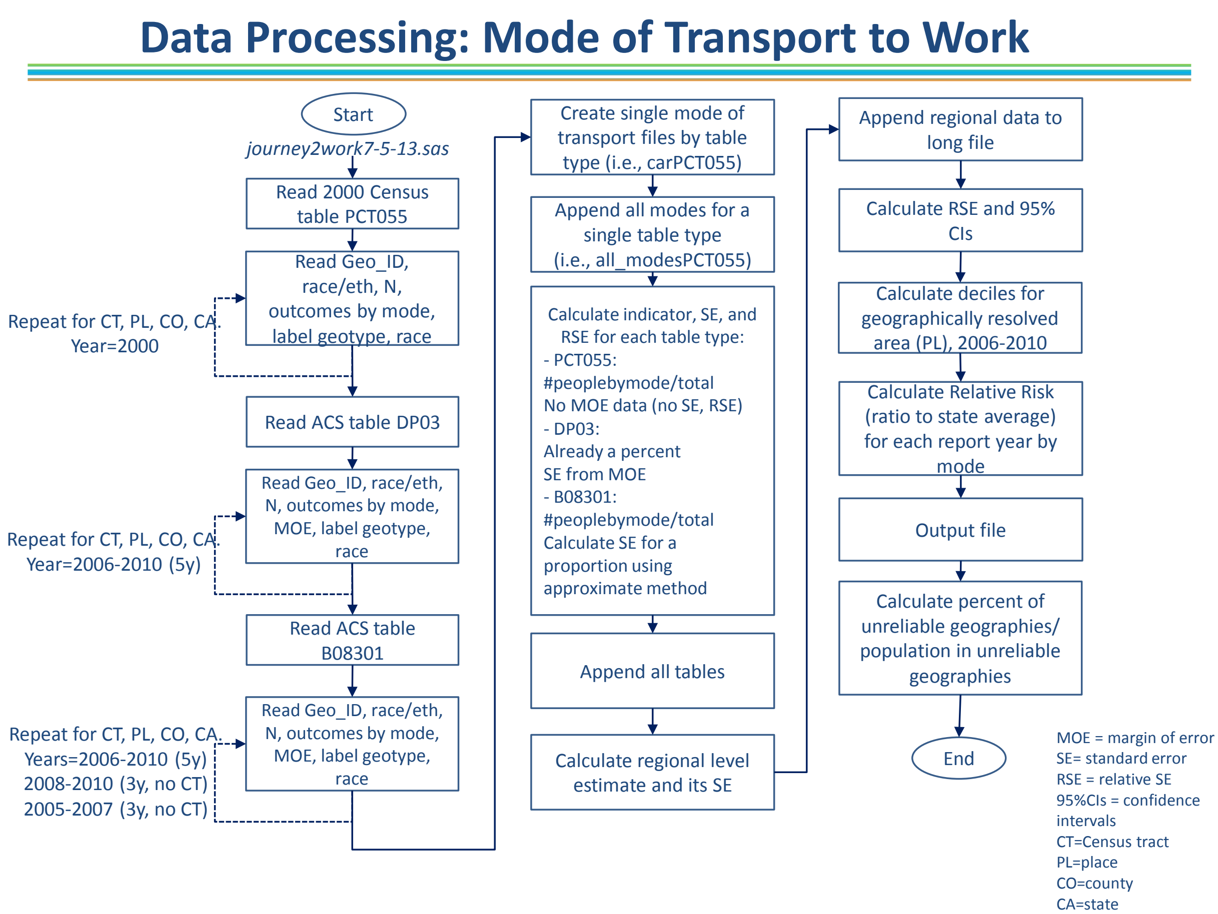
Relative standard errors, 95% confidence intervals, and decile ranking of places were also calculated. Regions were based on counties of metropolitan transportation organizations (MPO) as reported in the 2010 California Regional Progress Report (Appendix C).

Limitations/Comments:

The denominator of the indicator is limited to individuals with paid work. Commute to school or other types of trips were not included. Only the principal mode based on daily frequency or longest distance was used in the case of multi-modal trips on the same day or during the sample week. Commute trips to work tend to be longer in distance and are more likely to be made by automotive means, thus this indicator might depict a higher automotive mode share than if other type of trips were included. Race/ ethnicity data were not available for census tracts. Margin of error was not available for the year 2000. Taxicab was included in public transportation in 2000, but not for other years.

Data Processing

SAS program: journey2work7-5-13.sas



Data Acquisition and Preprocessing

The detailed table PCT055 from the 2000 Census, profile table DP03, and detailed table B08301 from the American Community Survey (2005-2007, 2008-2010, and 2006-2010) were downloaded using the American FactFinder (See Toolbox section, Downloading Census Data from American Fact Finder chapter). These tables correspond to the Census topic "Commuting." A total of 18 Census tables representing four datasets (time periods), four geographies, and eight race/ethnicities and the total population, were downloaded (See Table 4). The variables in the tables that contained the information of interest were identified prior to data processing in SAS (See Table 5) using the metadata tables downloaded from the Census.

Table 4. Census Tables Downloaded from American Fact Finder for the Mode of Transport to Work Indicator by Data Set, Geographies and Race and Ethnic Groups

Dataset (Time period)	Geographies	Race/Ethnic Groups	Table
2010 ACS 5-year Selected Population Tables (2006-2010)	Census Tract -140	Total population only	DP03 B08301
	Place – 160	Total and eight race/ethnicity groups	DP03 B08301
	County – 050	Total and eight race/ethnicity groups	DP03 B08301
	State – 040	Total and eight race/ethnicity groups	DP03 B08301
2010 ACS 3-year estimates (2008-2010)	Census Tract -140		No data
	Place - 160	Total and eight race/ethnicity groups	B08301
	County - 050	Total and eight race/ethnicity groups	B08301
	State - 040	Total and eight race/ethnicity groups	B08301
2007 ACS 3-year estimates (2005-2007)	Census Tract -140		No data
	Place - 160	Total and eight race/ethnicity groups	B08301
	County - 050	Total and eight race/ethnicity groups	B08301
	State - 040	Total and eight race/ethnicity groups	B08301
2000 SF4 Sample Data (2000)	Census Tract -140	Total population only	PCT055
	Place - 160	Total and eight race/ethnicity groups	PCT055
	County - 050	Total and eight race/ethnicity groups	PCT055
	State - 040	Total and eight race/ethnicity groups	PCT055

Table 5. Variables Used to Construct the Mode of Transport to Work Indicator by Census Table and Mode

Mode	Variable denominator	Variable numerator	Percent (indicator)	Margin of Error	Standard Error of the Percent
PCT055, Year: 2000					
Car, truck, or van	VD01	VD02	$\frac{VD02}{VD01} \times 100$	Not available	Not available
Car, truck, or van - Drove alone	VD01	VD03	〃	〃	〃
Car, truck, or van – Carpooled	VD01	VD04	〃	〃	〃
Public transportation (including taxicab)	VD01	VD05	〃	〃	〃
Bicycle	VD01	VD13	〃	〃	〃
Walked	VD01	VD14	〃	〃	〃
Worked at home	VD01	VD16	〃	〃	〃
DP03, Years: 2006-2010					
Car, truck, or van - Drove alone	Not available	Not available	HC03_VC29	HC04_VC29	$\frac{HC04\_VC29}{1.645}$
Car, truck, or van – Carpooled	〃	〃	HC03_VC30	HC04_VC30	〃
Public transportation (excluding taxicab)	〃	〃	HC03_VC31	HC04_VC31	〃
Walked	〃	〃	HC03_VC32	HC04_VC32	〃
Worked at home	〃	〃	HC03_VC34	HC04_VC34	〃
B08301, Years 2005-2007, 2008-2010, 2006-2010					
Car, truck, or van	HD01_VD01	HD01_VD02	$\frac{HD01\_VD02}{HD01\_VD01} \times 100$	HD02_VD01 (den) HD02_VD02 (num)	$\frac{1}{HD01\_VD01}\sqrt{[\frac{HD02\_VD01}{1.645}]^2 - \frac{HD01\_VD02^2}{HD01\_VD01^2}[\frac{HD02\_VD02}{1.645}]^2}$
Bicycle	HD01_VD01	HD01_VD18	〃	HD02_VD01 (den) HD02_VD18 (num)	〃
B08301, Years: 2005-2007, 2008-2010					
Car, truck, or van - Drove alone	HD01_VD01	HD01_VD03	$\frac{HD01\_VD18}{HD01\_VD03} \times 100$	HD02_VD01 (den) HD02_VD03 (num)	$\frac{1}{HD01\_VD01}\sqrt{[\frac{HD02\_VD01}{1.645}]^2 - \frac{HD01\_VD03^2}{HD01\_VD01^2}[\frac{HD02\_VD03}{1.645}]^2}$
Car, truck, or van – Carpooled	HD01_VD01	HD01_VD04	〃	HD02_VD01 (den) HD02_VD04 (num)	〃
Public transportation (excluding taxicab)	HD01_VD01	HD01_VD10	〃	HD02_VD01 (den) HD02_VD10 (num)	〃
Walked	HD01_VD01	HD01_VD19	〃	HD02_VD01 (den) HD02_VD19 (num)	〃
Worked at home	HD01_VD01	HD01_VD21	〃	HD02_VD01 (den) HD02_VD21 (num)	〃

The symbol 〃 indicates “same as above”

Walkable Distance to High Quality Public Transit

Definition

Full Title: Percent of population residing within ½ mile of a major transit stop

Indicator ID: 51

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Safe, sustainable, accessible and affordable transportation options

Significance and Health Connection:

A strong and sustainable transportation system supports safe, reliable, and affordable transportation opportunities for walking, bicycling, and public transit, and helps reduce health inequities by providing more opportunities for access to healthy food, jobs, health care, education, and other essential services. Active and public transportation promote health by enabling individuals to increase their level of physical activity, potentially reducing the risk of heart disease and obesity, improving mental health, and lowering blood pressure. Further, the transition from automobile-focused transport to public and active transport offers environmental health benefits, including reductions in air pollution, greenhouse gases and noise pollution, and leads to greater overall safety in transportation. Compared to public transit, a higher portion of trips by automobiles are associated with traffic accidents and increased air pollution, which are linked to increased rates of respiratory illness and heart disease.

Summary of evidence:

Individuals who live close to transit are more likely to be transit users and drive their cars less than people residing far from transit. Increased access to active and public transit is associated with increases in physical activity, which reduces risks of chronic disease and obesity.

Key References:

- Ewing R, Cervero R. Travel and the built environment: A meta-analysis. *JAPA* 2010; 76(3): 265-294.
- Frank LD, Andresen M, Schmid T. Obesity relationships with community design, physical activity, and time spent in cars. *Am J Prev Med* 2004; 27(2): 87-96.
- Besser LM, Dannenberg AL. Walking to public transit: Steps to help meet physical activity recommendations. *Am J Prev Med* 2005; 29(4): 273-280.
- Centers for Disease Control and Prevention. *CDC Recommendations for Improving Health through Transportation Policy*. National Center for Environmental Health. 2008. http://www.cdc.gov/transportation/docs/final-cdc-transportation-recommendations-4-28-2010.pdf

Detailed Definition:

- Numerator: number of people residing within ½ mile of bus/rail/ferry stop whose headway (waiting time) is less than 15 minutes during peak commute hours.
- Denominator: total population.

- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$

- Stratification: eight race/ethnicity groups (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white).
- Interpretation: healthier communities will have a higher percentage of the population residing within ½ mile of bus/rail/ferry stop whose headway (waiting time) is less than 15 minutes during peak commute hours.

Data Description:

- Data Sources: 2012 Transit Stops from the San Diego Association of Governments (www.sandag.org/), the Southern California Association of Governments (www.scag.ca.gov), and the Metropolitan Transportation Commission (www.mtc.ca.gov); 2008 Transit Stops from the Sacramento Council of Governments (www.sacog.org); 2010 block-level population data by race and ethnicity from the U.S. Census Bureau (provided by California State Data Center at the California Department of Finance)
- Years available: 2012 for SCAG, MTC, and San Diego Regions; 2008 for SACOG region.
- Updated: decennially
- Geographies available: Census tracts, cities/towns, counties, and region
    - SCAG region (Imperial, Los Angeles, Orange, Riverside, San Bernardino, and Ventura)
    - San Diego County
    - SACOG region (Placer, Sacramento, and Yolo)
    - Bay Area MTC region (Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, Sonoma)

Methodology:

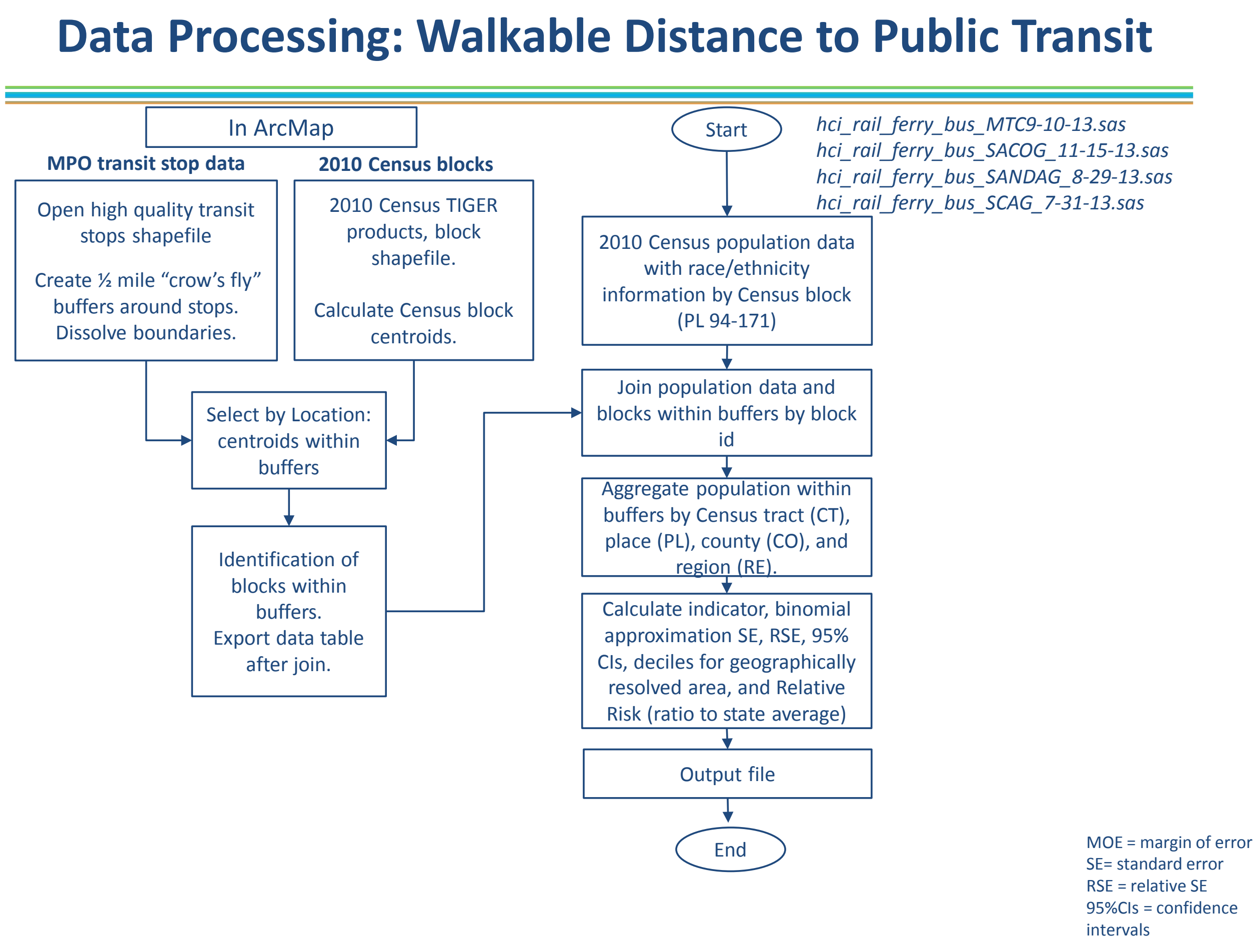
Transit stops included those served by one or more fixed route transit service with a frequency of 15 minutes or less during peak hours (6-9AM, 3-6PM). For the SCAG and MTC regions, stops with multiple routes whose average frequency was 15 minutes or less were included (e.g. 2 different bus routes with 30 minute frequencies each). Geospatial software (ArcMAP 10.1) was used to identify census blocks with centroids inside ½ mile buffers of the transit stops. Block-level 2010 Census redistricting data (100% counts by race/ethnicity) was merged with blocks inside the transit access area, and population counts were aggregated by census tract, cities/towns, county, and region. For each geography level and race/ethnicity strata, rates of transit access were calculated. Standard errors, relative standard errors, and 95% upper and lower confidence intervals, and decile rankings for cities and census tracts were calculated.

Strengths and limitations:

Transit stops and service are subject to change and this analysis may not reflect recent changes. Census blocks are designated as inside or outside of transit buffers based on block centroids, which may result in small under- or overestimates of the population within buffer areas. The population data are from a slighter earlier time period (2010) than the transit data (2012), which may introduce a small error if the population numbers or demographics have changed. This indicator measures geographic access; however, other characteristics of public transit, such as affordability and personal safety (e.g. crime), also impact transit use.

Data Processing

SAS programs:
hci_rail_ferry_bus_MTC9-10-13.sas
hci_rail_ferry_bus_SACOG_11-15-13.sas
hci_rail_ferry_bus_SANDAG_8-29-13.sas
hci_rail_ferry_bus_SCAG_7-31-13.sas



Data Acquisition and Preprocessing

Population Data at the Census Block Level, 2010

The 2010 Census block shapefile was downloaded from the TIGER website on 11/20/12 (http://www.census.gov/geo/maps-data/data/tiger-data.html). See Toolbox section, Downloading Census TIGER Shapefiles and Reprojecting in California Teale Albers Coordinate System chapter. The block centroids were calculated in ArcMap and a new shapefile of block centroids was created. See Toolbox section, Calculating Centroids in ArcMap and Creating a Centroid Shapefile chapter.

Population at the Census Block level with race/ethnicity information was obtained from the Department of Finance, Demographic Research Unit, 2010 redistricting data from the U.S. Census Bureau (Public Law 94-171) for California. (http://www.dof.ca.gov/research/demographic/state_census_data_center/census_2010/#PL94, http://www2.census.gov/census_2010/01-Redistricting_File--PL_94-171/).

Transit Stops

Data for local and regional transit stops and routes were obtained from the following contacts in Metropolitan Planning Organizations:

- Metropolitan Transit Commission (MTC): Stella Wotherspoon
    - o 2012 transit stops shapefiles from MTC was reprojected to NAD 1983 Teale Albers California (Meters). Both regional and local transit stops were used to construct the indicator.
- San Diego Association of Governments (SANDAG): Tom King
    - o 2012 data for transit stops with peak headways of 15 minutes or less (in accordance with SB375) was obtained from SANDAG (July, 2013). See forwarded e-mail from Tom King. Note: SANDAG's method for classifying stops may be different and/or less complex than SCAG's method.
- Southern California Association of Governments (SCAG): Jung Seo
    - o 2012 transit stops with headways of 15 minutes or less (in accordance with SB 375) shapefile was acquired from SCAG (July, 2013). SCAG's methodology for choosing transit stops is described in methodology document (Data Source and Procedure - TPP (Amended).docx). Note: SCAG's method for analysis of stops may be more detailed and/or different than SANDAG's method. Also see forwarded e-mail from Jung Seo.
- Sacramento Area Council of Governments (SACOG): Kazey Lizon
    - o 2008 transit stops (buffers) with headways of 15 minutes or less (in accordance with SB 375) shapefile was acquired from SACOG (November, 2013)

The following preprocessing was conducted for each transit stops shapefile except SACOG for which buffers were already provided:

1. The buffer tool in ArcMap was used to create a ½ mile buffer around each transit stop and the dissolve tool was used to dissolve all stops buffered areas into 1 new buffer shapefile. The select by location tool was used to select the Census blocks that have their centroid within regional/local transit stops buffer. See Toolbox "Creating Buffers around Features and Select by Location in ArcMap".
2. The resulting table was exported as an Excel file that was later imported into SAS for analysis.

Retail Food Environment

Definition

Full Title: Modified Retail Food Environment Index (mRFEI)

Indicator ID: 75

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Affordable, accessible and nutritious foods

Description of significance and health connection:

An adequate, nutritious diet is a necessity at all stages of life. Pregnant women and their developing babies, children, adolescents, adults, and older adults depend on adequate nutrition for optimum development and maintenance of health and functioning. Inadequate diets in children can impair intellectual performance and have been linked to frequent school absence and poorer educational achievement. Nutrition also plays a significant role in causing or preventing a number of illnesses, such as cardiovascular disease, some cancers, obesity, type-2 diabetes, and anemia. These weight-associated illnesses are no longer restricted to adults as the prevalence of obesity has more than doubled in children in the last 40 years. Obese children have an increased risk of heart disease and of becoming obese adults.

Lower income families are less likely to have a nutritious diet than those with higher incomes. Peoples' food choices and their likelihood of being overweight or obese are also influenced by their food environment: the foods available in their neighborhoods including stores, restaurants, schools, and worksites. There is a strong association between consumption of calorie-dense foods with low nutritional value and being overweight or obese when one or more calorie-dense meals are consumed per week. High-fat and high-sugar foods are available at most elementary and middle schools. Since the 1970s, the number of fast food restaurants has more than doubled and the proportion of daily calorie intake from foods eaten away from home has increased.

Summary of Evidence:

Measures of food availability in the environment include distance to food retailers, cost of foods, or density of food outlets. Due to the lack of standardization of food environment metrics and differences among populations studied, it is difficult to generalize the evidence on the relationship between the food environment and health. Nevertheless, various cross-sectional and longitudinal studies show a positive association between the density of fast-food restaurants and/or convenience stores with Body Mass Index (BMI), obesity and overweight rates; and a negative association with fruit and vegetable intake. The size of this relationship can vary with race/ethnicity. In California, adults living in cities or counties with 16.7% healthy food retailers or less had a 20 percent higher prevalence of obesity and a 23 percent higher prevalence of diabetes than adults living in areas with 25.0% healthy food retailers or more; this relationship held true regardless of household income, race/ethnicity, age, gender, or the physical activity levels of respondents.

Key References:

- California Center for Public Health Advocacy. *Searching for Healthy Food: The Food Landscape in California Cities and Counties*. 2007. Accessed November 8th, 2013.
- Robert Wood Johnson Foundation Commission to Build a Healthier America. *Improving the Health of All Americans through Better Nutrition*. 2009. Accessed November 8th, 2013.
- California Center for Public Health Advocacy, PolicyLink, UCLA Center for Health Policy Research. *Designed for Disease: The Link Between Local Food Environments and Obesity and Diabetes*. 2008. Accessed November 8th, 2013.
- U.S. Department of Agriculture and U.S. Department of Health and Human Services. *Dietary Guidelines for Americans.* 7th ed, Washington, DC: U.S. Government Printing Office; 2010. Accessed November 8th, 2013.
- Papas MA, Alberg, AJ, Ewing R, Helzlsouer KJ, Gary TL, Kalssen AC. The built environment and obesity. *Epidemiol Rev*. 2007; 29(1):129-143.
- Gibson DM. The neighborhood food environment and adult weight status: estimates from longitudinal data. *Am J Public Health.* 2011; 101(1): 71-78.
- Zenk SN, Lachance LL, Schulz AJ, Mentz G, Srimathi K, Ridella W. Neighborhood retail food environment and fruit and vegetable intake in a multiethnic urban population. *Am J Health Promot*. 2009; 23(4): 255-264.

Detailed Definition:

- Numerator: number of healthy food retailers
- Denominator: total food retailers (number of healthy plus number of less healthy food retailers)
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Healthy food retailers include supermarkets, larger grocery stores, supercenters, and produce stores. Less healthy food retailers include fast food restaurants, small grocery stores, and convenience stores.
- Stratification: eight race/ethnicity groups (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white)
- Interpretation: healthier communities will have a higher percentage of healthy food retailers.

Data Description:

- Data sources: Division of Nutrition, Physical Activity, and Obesity from the Centers for Disease Control and Prevention (CDC): www.cdc.gov/obesity/downloads/2_16_mrfei_data_table.xls. U.S. Census Bureau, 2000 Census DP-1 population table (http://factfinder2.census.gov).
- Years available: 2009
- Updated: unknown
- Geographies available: Census tracts, and population weighted averages for cities/towns, counties, regions, and state

Methodology:

Data on the mRFEI or percentage of healthy food retailers for California Census tracts (2000) was downloaded from the CDC. The CDC methodology considered that healthy food retailers include supermarkets, larger grocery stores, supercenters, and produce stores within census tracts or ½ mile from the tract boundary. Less healthy food retailers include fast food restaurants, small grocery stores, and convenience stores within census tracts or ½ mile from the tract boundary. Population weighted mRFEI averages by race/ethnicity were calculated for cities/towns, counties, regions, and the state, from census tract data. Regions were based on counties of metropolitan transportation organizations (MPO) regions as reported in the 2010 California Regional Progress Report (Appendix C).

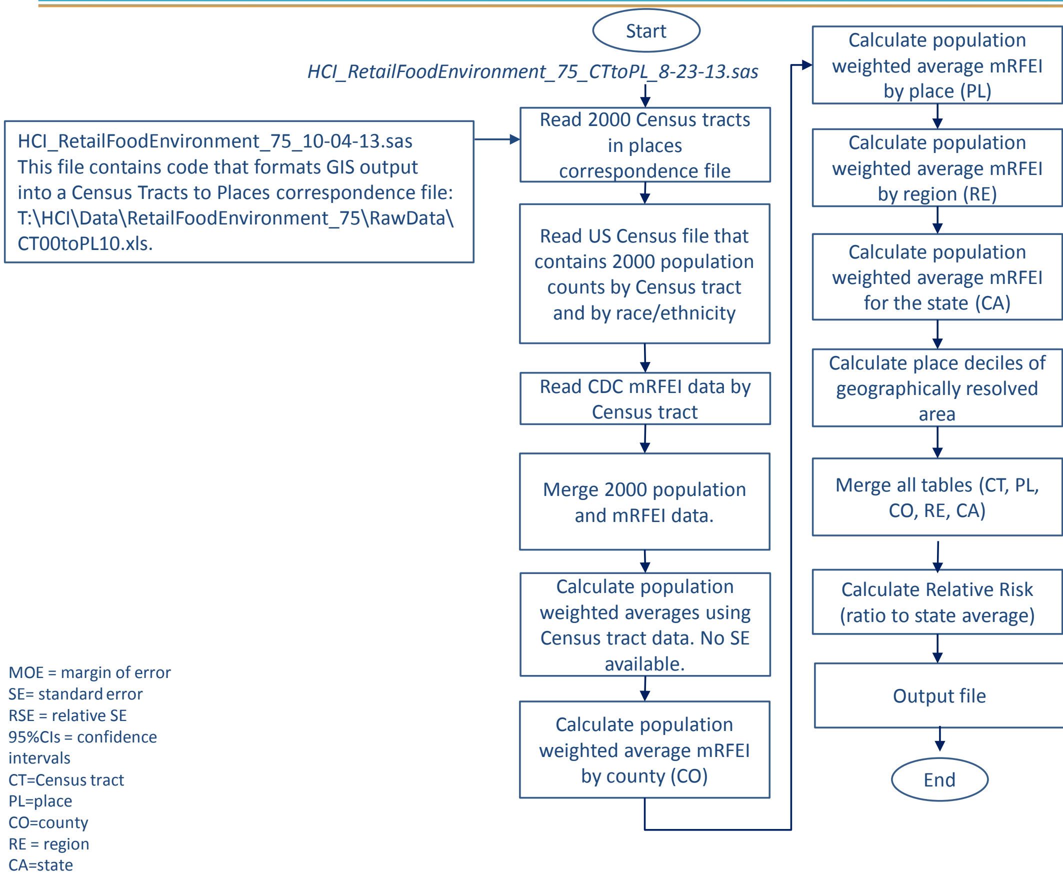
Limitations:

Travel distances to food retailers are not considered in this indicator. The number and type of food retailers by Census tract is not available in the CDC dataset due to contractual agreements with private companies. Ground verification has shown that business lists from private companies can have poor performance when used to measure the food environment. There are no immediate plans to update this indicator. The mRFEIs for places were obtained after intersecting the 2000 census tract centroids with places; information loss might have occurred if a census tract was partially contained within a place, but its centroid was outside the place limits. An mRFEI of 11 was reported for California by the CDC, which is lower than the population weighted average of 12.2 obtained here for this indicator maybe suggesting slightly higher indexes in populous areas.

<u>Data Processing</u>

SAS program: HCI_RetailFoodEnvironment_75_8-23-13.sas

Data Processing: Retail Food Environment



<u>Data Acquisition and Preprocessing</u>

1. Data on the modified retail food environment index (mRFEI) for California Census tracts (2000) was available from the Division of Nutrition, Physical Activity, and Obesity from the Centers for Disease Control and Prevention (http://www.cdc.gov/obesity/resources/reports.html).
2. The CDC used retailer data from the years 2008 and 2009 to calculate the mRFEI. According to the CDC methodology (ftp://ftp.cdc.gov/pub/Publications/dnpao/census-tract-level-state-maps-mrfei_TAG508.pdf), healthy food retailers include supermarkets, larger grocery stores, supercenters, and produce stores within census tracts or ½ mile from the tract boundary. The following stores as defined by North American Industry Classification Codes (NAICS) were included: supermarkets and larger grocery stores (NAICS 445110; supermarkets further defined as stores with >= 50 annual payroll employees and larger grocery stores defined as stores with 10–49 employees); fruit and vegetable markets (NAICS 445230); warehouse clubs (NAICS 452910). Fruit and vegetable markets include establishments that sell produce and include markets and permanent stands. Less healthy food retailers include fast food restaurants, small grocery stores, and convenience stores within census tracts or ½ mile from the tract boundary. Fast food stores were defined according to NAICS code 722211(fast food restaurants). Convenience stores were defined according to NAICS code 445120

(convenience stores) or NAICS code 445110 (small groceries) where the number of employees was three or fewer.

3. Before calculating the weighted average for places, the tracts within places were determined by intersecting tract centroids (see Toolbox section, Calculating Centroids in ArcMap and Creating a Centroid Shapefile chapter) and places (see Toolbox section, Intersecting Two Shapefiles in ArcMap). After the intersection, the resulting data table was exported as an Excel file containing Census tracts within places. A SAS program (HCI_RetailFoodEnvironment_75_CTtoPL_8-23-13.sas) was used to format the results of the intersection into a lookup table of the correspondence between 2000 census tracts (CT) and 2000 places (PL): CT00toPL00.xls.

Population weighted averages to obtain place and county estimates

The original variable in the CDC file was the mRFEI by Census tract and the original variable from the Census 2000 population was total population by Census tract by race/ethnicity. The weighted average $mRFEI_{i,j}$ for place (or county) i and race/ethnicity j was calculated with the formula:

$$mRFEI_{i,j} = \frac{\sum_{k=1}^{n} mRFEI_{i,k} \times N_{i,j,k}}{\sum_{k=1}^{n} N_{i,j,k}}$$

Were n is the number of Census tracts in place (or county) i. $N_{i,j,k}$ is the population in Census tract k within place (or county) i for race/ethnicity j. $mRFEI_{i,k}$ is the mRFEI for Census tract k within place (or county) i.

Percent of Households Incurring Cost Burdens

<u>Definition</u>

Full Title: Percent of households paying more than 30% (or 50%) of monthly household income towards housing costs

Indicator ID: 106

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Affordable, high quality, socially integrated and location-efficient housing

Description of significance and health connection:

Affordable, quality housing is central to health, conferring protection from the environment and supporting family life. Substandard housing is associated with increased risks of injury and respiratory ailments. Homes can be a source of exposure to radon, lead, asbestos or other hazardous agents. In children, lead exposure increases the risk of neurological impairment and developmental delays. Chronic homelessness is associated with higher rates of injuries, cancer, cardiovascular disease, substance addictions, mental disorders and death. Children and adolescents with transient housing have impaired academic performance. Housing costs—typically the largest, single expense in a family's budget—also impact decisions that affect health. As housing consumes larger proportions of household income, families have less income for nutrition, health care, transportation, education, etc. Severe cost burdens may induce poverty—which is associated with developmental and behavioral problems in children and accelerated cognitive and physical decline in adults. Low-income families and minority communities are disproportionately affected by the lack of affordable, quality housing.

Summary of evidence:

Controlled studies of the impact of housing characteristics or cost burdens on specific health outcomes are limited. However, cohort studies have documented adverse effects to health. Moisture linked to household mold was associated with respiratory illness, nausea, and fatigue. Lead abatement in residential housing was associated with abnormally, elevated blood lead levels in children. Overcrowding was associated with higher incidence of tuberculosis. Housing insecurity, especially triggered by poverty, was associated with behavioral problems in children and excessive school absences.

Key References:

- Baggett TP, Hwang SW, O'Connell JJ et al. Mortality among homeless adults in Boston: shifts in causes of death over a 15-year period. *JAMA Intern Med* 2013; 173:189-195.
- Robert Wood Johnson Foundation. Commission to Build a Healthier America. Housing and Health, Issue Brief 2; 2008.
- Stone ME. What is housing affordability? The case for the residual income approach. Housing Policy Debate 2006; 17:151–184.
- Thomson H, Thomas S, Sellstrom E, Petticrew M. Housing improvements for health and associated socio-economic outcomes. Cochrane Database Sys Rev 2013; 2:335.

- Office of the Deputy Prime Minister. The Impact of Overcrowding on Health & Education: A Review of Evidence and Literature. Wetherby, UK: Office of the Deputy Prime Minister Publications; 2004.
- California Department of Housing & Community Development. Housing and Health. 2013.

Detailed Definition:

- Numerator: number of households paying > 30% (and > 50%) of monthly household income toward monthly housing costs—rent and utilities or mortgage, utilities, property tax, insurance, home association fees, etc.
  - Note: at the Census tract level the definition of the numerator was the number households paying ≥ 30% (and ≥ 50%) of monthly household income toward monthly housing costs.
- Denominator: total number of households.
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: three types of housing tenure (owner-occupied, renter-occupied and total households); two income levels of renter/householder (households with a monthly household income at ≤ 30% and all levels of HUD-adjusted median family income); and eight race/ethnicity groups (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white).
- Interpretation: healthier communities will have the lowest (or zero) percentage of households paying more than 30% of monthly household income toward monthly housing costs.

Data Description:

- Data source: U.S. Department of Housing and Urban Development (HUD), Consolidated Planning Comprehensive Housing Affordability Strategy (CHAS) data (Tables 8 and 9, http://www.huduser.org/portal/datasets/cp.html); U.S. Census Bureau, American Community Survey (ACS) (Tables DP04, B25070 and B25091, http://factfinder2.census.gov/)
- Years available: 2006-2010
- Updated: 3 and 5 year intervals
- Geographies available: Census tracts, places, counties, regions (derived), and state

Methodology:

In CHAS data, housing cost burden estimates were pre-calculated for renter- and owner-occupied households. To derive the percent of households with a >30% cost burden (numerator), estimates from two cost strata (30%-50% and >50%) were summed and divided by the denominator. For the percent of households incurring a >50% cost burden, the estimate for the > 50% severely cost-burdened was divided by the same denominator. Both derived percentages were calculated for each type of housing tenure, where the number of total households was obtained by aggregating the estimates of renter- and owner-occupied households. The indicators and standard errors were calculated using the approximate method for the geographies of place, county, region, and state. Relative standard errors (RSE), 95% confidence intervals, and decile ranking of places were also calculated. Regions were based on counties of metropolitan transportation organizations (MPO) as reported in the 2010 California Regional Progress Report (Appendix C). Census tract estimates using CHAS data were statistically unstable (RSEs ≥ 30% for the majority of census tracts) and therefore not presented.

The > 30% cost burden indicator was calculated as follows:

Percent of households with a $>$ 30% cost burden, households at all levels of HAMFI

$$= \frac{\text{Number of households paying monthly housing costs } > 30\% \text{ of monthly household income}}{\text{Total number of households}} \times 100$$

Percent of households with a $>$ 30% cost burden, households with a monthly household income at $\leq$ 30% of HAMFI

$$= \frac{\begin{array}{c}\text{Number of households paying monthly housing costs } > 30\% \text{ of monthly household income} \\ \text{and having a monthly household income at } \leq \text{ 30\% of HAMFI}\end{array}}{\text{Total number of households with a monthly household income at } \leq \text{ 30\% of HAMFI}} \times 100$$

The > 50% cost burden indicator was calculated as follows:

Percent of households with a $>$50% cost burden, households at all levels of HAMFI

$$= \frac{\text{Number of households paying monthly housing costs } > 50\% \text{ of monthly household income}}{\text{Total number of households}} \times 100$$

Percent of households with a $>$ 50% cost burden, households with a monthly household income at $\leq$ 30% of HAMFI

$$= \frac{\begin{array}{c}\text{Number of households paying monthly housing costs } > 50\% \text{ of monthly household income} \\ \text{and having a monthly household income at } \leq \text{ 30\% of HAMFI}\end{array}}{\text{Total number of households with a monthly household income at } \leq \text{ 30\% of HAMFI}} \times 100$$

Stable estimates at the census tract level were provided by the ACS data. Here, the definition of the indicator was the percent of households incurring housing costs ≥ 30% (and ≥ 50%) of monthly household income. The denominator was renter and owner-occupied households paying housing costs; rent-free and mortgage-free households were excluded. The percent of cost-burdened households for renters and homeowners were pre-calculated in the ACS data profile DP04 tables. To derive the count of cost-burdened households (numerator) that were specific to each type of household, two cost strata percentages (30-34% and ≥35%) were summed and then multiplied by the denominator. The percent of severely cost-burdened households (≥ 50% cost burden) was calculated from the ACS B25070 and B25091 tables using household counts specific to each type of household, dividing by the total number of households specific to the type of household and multiplying by 100. The overall cost-burdened and severely cost-burdened were derived by aggregating household counts of renters and homeowners with a mortgage (combined households). The indicators and their standard errors were calculated using the approximate method (see Toolbox section, Standard Error Calculations chapter). Relative standard errors, 95% confidence intervals, and decile ranking of places and census tracts were also calculated. Regions were based on counties of metropolitan transportation organizations (MPO) as reported in the 2010 California Regional Progress Report (Appendix C).

The indicator was calculated as follows:

Percent of households with a ≥ 30% cost burden (overall cost-burdened)

$$= \frac{\text{Number of combined households paying monthly housing costs} \geq 30\% \text{ of monthly household income}}{\text{Total number of combined households paying monthly housing costs}} \times 100$$

Percent of households with a ≥ 50% cost burden (severely cost-burdened)

$$= \frac{\text{Number of combined households paying monthly housing costs} \geq 50\% \text{ of monthly household income}}{\text{Total number of combined households paying monthly housing costs}} \times 100$$

Limitations:

The housing cost burden estimates do not adjust for differences in household size. Estimates for the survey period 2006-2010 are bisected by the Great Recession (2008), marked by a large increase in home foreclosures, and house/rental price instability. Due to changes in definitions and sampling, HUD does not recommend making comparisons to prior years' estimates (e.g., 2005-2007 and 2000). ACS data are available at census tract geographies, albeit with a definition of cost burden that is different than that of CHAS.

Data Processing

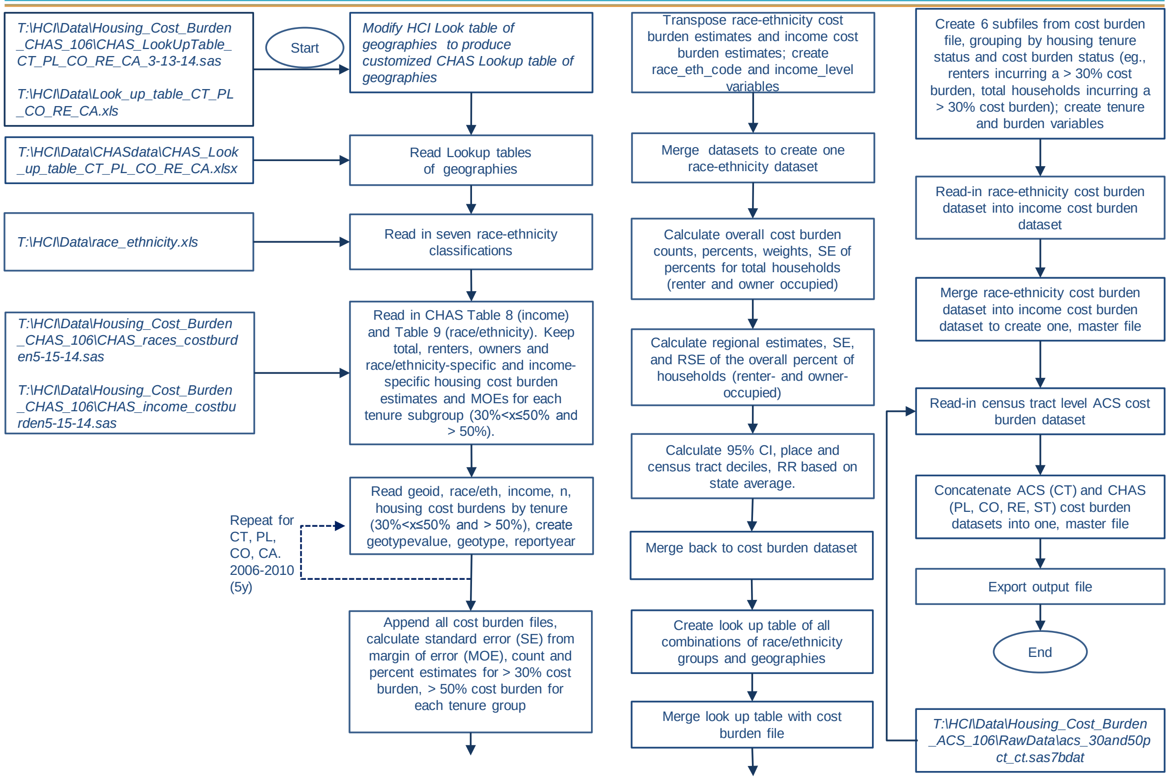
To produce the ACS housing cost burden indicators, run this sequence of files:

- Creates ≥ 50% cost burden analytic dataset:
  - rent_mortgage_cost_burden extraction2006-2010_<date>.sas
  - rent_mortgage_cost_burden<date>.sas
- Creates ≥ 30% cost burden analytic dataset and the final ACS analytic dataset
  - DP04extraction2006-2010_<date>.sas
  - DP04housgcostburden<date>.sas (this file appends the two cost burden data sets: ≥ 30%, ≥ %50).

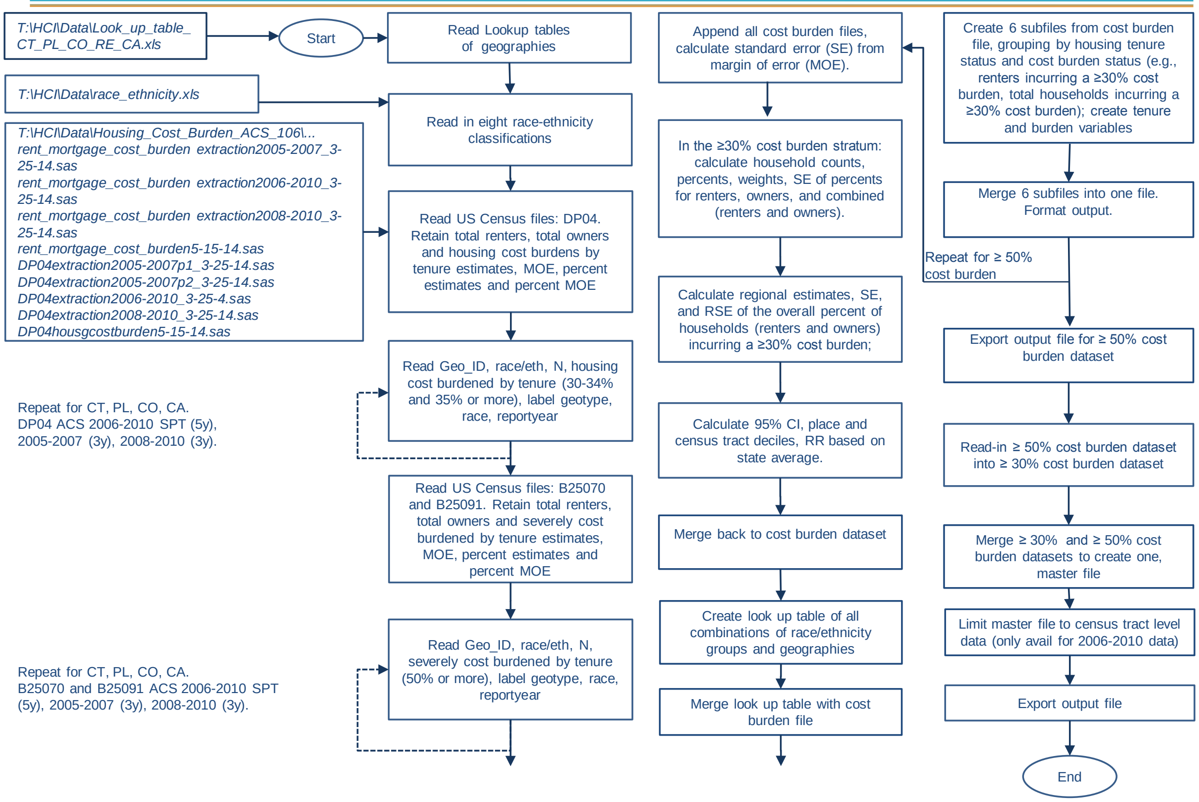
To generate the CHAS housing cost burden indicators, run this sequence of files:

- CHAS_LookUpTable_CT_PL_CO_RE_CA_<date>.sas
- CHAS_races_costburden<date>.sas
- CHAS_income_costburden<date>.sas (this file appends the race/ethnicity-specific cost burden data to that of the income-level specific cost burden data).

Indicator Analysis Overview: Percent of Owner-Occupied, Renter-Occupied and Total Households Incurring a > 30% Cost Burden and > 50% Cost Burden using CHAS data 2006-2010



Indicator Analysis Overview: Renting households, Mortgage-paying households and Combined Households that are Cost Burdened (≥30% of income) and Severely Cost Burdened (≥ 50% of income) using ACS data



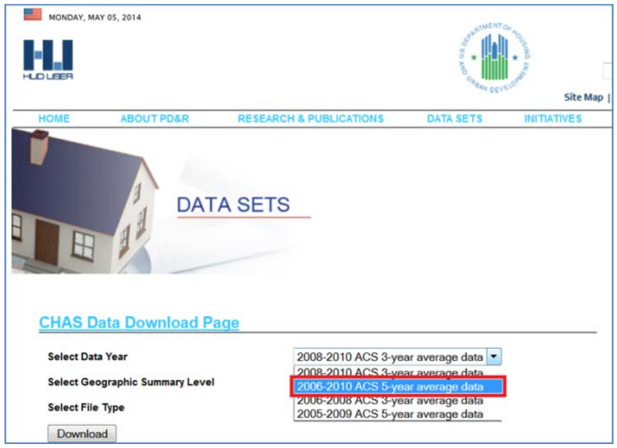
Data Acquisition and Preprocessing

U.S. Department of Housing and Urban Development (HUD), Consolidated Planning Comprehensive Housing Affordability Strategy (CHAS) data, Tables 8 and 9

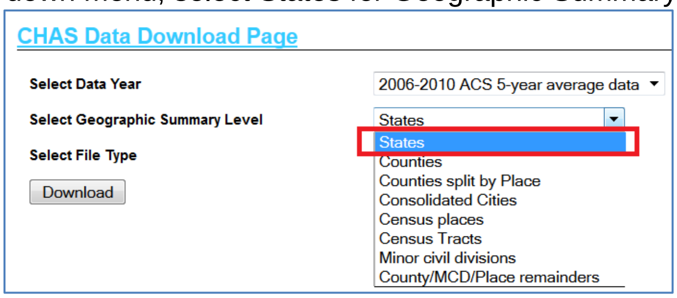
The following steps show how to download state-level CHAS data as an example. This process was repeated for other geographic levels (places and counties). Prior to downloading, file folders were created —named by survey source, years and geography—to prevent the CHAS files from being overwritten and assist in file identification. As an example, a folder named *CHAS_2006-2010States* was created for state data.

1. Go to: http://www.huduser.org/portal/datasets/cp/CHAS/data_download_chas.html

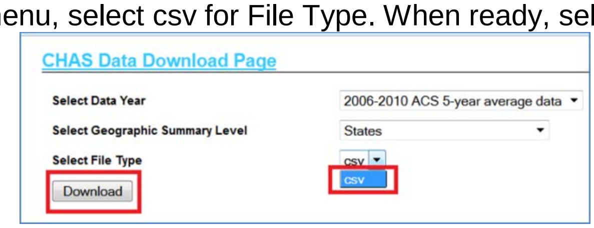
2. From the drop-down menu, select *2006-2010* for *Data Year.*



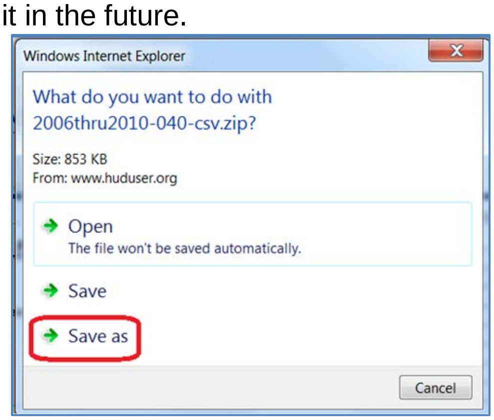
3. From the drop-down menu, select States for Geographic Summary Level



4. From the drop-down menu, select csv for File Type. When ready, select Download.



5. A new message window will appear. To download the zip file, select Save As, choose the path location where you want to save the file, assign the file a name and then select *Save*. It is recommended that each zip file be renamed to reflect the survey source, period and geography. This action will prevent overwriting and assist in file identification. An example of a filename might be *CHAS_2006-2010States.zip*. Remember to note where you saved the file so you can retrieve it in the future.



6. For the remaining geographic-level data (places and counties), repeat steps 1-5, making sure to choose the corresponding geographic summary level of interest. When finished, there should be a total of four downloaded zip files (one for each geography).
7. Unpack the zip files of all the datasets to reveal the csv files. Tables 8 and 9 for each geography will be contained in the unpacked files.
8. Table 8 contains data on owner occupied and renter occupied cost burden by household income. Table 9 contains data on owner occupied and renter occupied cost burden by race/ethnicity. CHAS data provides data dictionaries with more information about the variables contained in the files.

Table 6. CHAS 2006-2010 (Table 9) Variables used to construct the Housing Cost Burden by Race/ethnicity Indicator for Places, Counties, Regions and the State

Raw Variables and Renamed Variables
Total Occupied housing units: T9_est1=total T9_est2=owners T9_est38=renters
Owner occupied housing units and race/ethnicity: T9_est3=white_owners (white alone, non-Hispanic) T9_est8=afram_owners (African American alone, non-Hispanic) T9_est13=asian_owners (Asian alone, non-Hispanic) T9_est18=amian_owners (American Indian or Alaska Native alone, non-Hispanic) T9_est23=pacis_owners (Pacific Islander alone, non-Hispanic) T9_est28=latin_owners (Hispanic, any race) T9_est33=multi_owners (multiple races, non Hispanic)
Renter occupied housing units and race/ethnicity: T9_est39=white_renters T9_est44=afram_renters T9_est49=asian_renters T9_est54=amian_renters T9_est59=pacis_renters T9_est64=latin_renters T9_est69=multi_renters
Owner occupied housing units and race/ethnicity and housing cost burden is greater than 30% but less than or equal to 50%: T9_est5=white_owners30to50 T9_est10=afram_owners30to50 T9_est15=asian_owners30to50 T9_est20=amian_owners30to50 T9_est25=pacis_owners30to50 T9_est30=latin_owners30to50 T9_est35=multi_owners30to50
Renter occupied housing units and race/ethnicity and housing cost burden is greater than 30% but less than or equal to 50%: T9_est41=white_renters30to50 T9_est46=afram_renters30to50 T9_est51=asian_renters30to50 T9_est56=amian_renters30to50 T9_est61=pacis_renters30to50 T9_est66=latin_renters30to50 T9_est71=multi_renters30to50
Owner occupied housing units and race/ethnicity and housing cost burden is greater than 50%: T9_est6=white_owners50 T9_est11=afram_owners50 T9_est16=asian_owners50 T9_est21=amian_owners50 T9_est26=pacis_owners50 T9_est31=latin_owners50 T9_est36=multi_owners50

Renter occupied housing units and race/ethnicity and housing cost burden is greater than 50%: T9_est42=white_renters50 T9_est47=afram_renters50 T9_est52=asian_renters50 T9_est57=amian_renters50 T9_est62=pacis_renters50 T9_est67=latin_renters50 T9_est72=multi_renters50
The margin of error variables of the estimates: T9_moe1: Margin of error for estimate T9_est1 T9_moe2: Margin of error for estimate T9_est2 T9_moe38: Margin of error for estimate T9_est38 T9_moe3: Margin of error for estimate T9_est3 Etc.

Table 7. CHAS 2006-2010 (Table 8) Variables used to construct the Housing Cost Burden by Income Indicator for Places, Counties, Regions and the State

Raw Variables and Description
T8_est1: Total occupied housing units T8_est2: Owner occupied housing units T8_est68: Renter occupied housing units
Owner occupied housing units T8_est3: household income is less than or equal to 30% of HAMFI T8_est16: household income is greater than 30% but less than or equal to 50% of HAMFI T8_est29: household income is greater than 50% but less than or equal to 80% of HAMFI T8_est42: household income is greater than 80% but less than or equal to 100% of HAMFI T8_est55: household income is greater than 100% of HAMFI
Renter occupied housing units T8_est69: household income is less than or equal to 30% of HAMFI T8_est82: household income is greater than 30% but less than or equal to 50% of HAMFI T8_est95: household income is greater than 50% but less than or equal to 80% of HAMFI T8_est108: household income is greater than 80% but less than or equal to 100% of HAMFI T8_est121: household income is greater than 100% of HAMFI
Owner occupied housing units T8_est7: household income is less than or equal to 30% of HAMFI and housing cost burden is greater than 30% but less than or equal to 50% T8_est10: household income is less than or equal to 30% of HAMFI and housing cost burden is greater than 50% T8_est20: household income is greater than 30% but less than or equal to 50% of HAMFI and housing cost burden is greater than 30% but less than or equal to 50% T8_est23: household income is greater than 30% but less than or equal to 50% of HAMFI and housing cost burden is greater than 50% T8_est33: household income is greater than 50% but less than or equal to 80% of HAMFI and housing cost burden is greater than 30% but less than or equal to 50% T8_est36: household income is greater than 50% but less than or equal to 80% of HAMFI and housing cost burden is greater than 50% T8_est46: household income is greater than 80% but less than or equal to 100% of HAMFI and housing cost burden is greater than 30% but less than or equal to 50% T8_est49: household income is greater than 80% but less than or equal to 100% of HAMFI and housing cost burden is greater than 50% T8_est59: household income is greater than 100% of HAMFI and housing cost burden is greater than 30% but less than or equal to 50% T8_est62: household income is greater than 100% of HAMFI and housing cost burden is greater than 50%
Renter occupied housing units T8_est73: household income is less than or equal to 30% of HAMFI and housing cost burden is greater than 30% but less than or equal to 50% T8_est76: household income is less than or equal to 30% of HAMFI and housing cost burden is greater than 50% T8_est86: household income is greater than 30% but less than or equal to 50% of HAMFI and housing cost burden is greater than 30% but less than or equal to 50%

T8_est89: household income is greater than 30% but less than or equal to 50% of HAMFI and housing cost burden is greater than 50% T8_est99: household income is greater than 50% but less than or equal to 80% of HAMFI and housing cost burden is greater than 30% but less than or equal to 50% T8_est102: household income is greater than 50% but less than or equal to 80% of HAMFI and housing cost burden is greater than 50% T8_est112: household income is greater than 80% but less than or equal to 100% of HAMFI and housing cost burden is greater than 30% but less than or equal to 50% T8_est115: household income is greater than 80% but less than or equal to 100% of HAMFI and housing cost burden is greater than 50% T8_est125: household income is greater than 100% of HAMFI and housing cost burden is greater than 30% but less than or equal to 50% T8_est128: household income is greater than 100% of HAMFI and housing cost burden is greater than 50%
The margin of error variables of the estimates: T8_moe1: Margin of error for estimate T8_est1 T8_moe2: Margin of error for estimate T8_est2 T8_moe68: Margin of error for estimate T8_est68 T8_moe3: Margin of error for estimate T8_est3 Etc.

U.S. Census Bureau, American Community Survey (ACS), Table DP04

Table DP04 at the Census tract level was downloaded from the American FactFinder (http://factfinder2.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t) for the dataset 2010 ACS 5-year Selected Population Tables and the race/ethnic groups 01-Total, 400-Hispanic/Latino, 451- White, 453- Black, 455- American Indian/Alaska Native, 457- Asian, 459-NHOPI, 461- Other, and 463- Multiple. Table DP04 is a Census data profile-quick table that includes multiple topics and “Owner Costs & Fees” and “Renter Housing Costs & Fees” are among them. Table 8 contains the names of the variables from DP04 that were used to construct the indicator.

U.S. Census Bureau, American Community Survey (ACS), Tables B25070, B25091

Tables B25070 (Gross rent as a percentage of household income in the past 12 months) and B25091 (Mortgage status by selected monthly owner costs as a percentage of household income in the past 12 months) are detailed tables from the ACS. Detailed tables are included in the summary tables which are a series of comma-delimited text files that can be found on the Census Bureau’s FTP site www2.census.gov. Instructions on how to download tables from the FTP site and how to extract the variables of interest can be found in the Toolbox section, Obtaining American Community Survey Data from Summary Files chapter.

Alternatively, all detailed tables can be obtained from the FactFinder by directly typing the table name in the “topic or table name” of the “Advanced Search” section and following the steps in the Toolbox section, Downloading Census Data from American FactFinder chapter to select the dataset, geographies, and race/ethnic groups of interest.

Table 8 contains the names of the variables from B25070 and B25091 (as presented in the summary files) that were used to construct the indicator.

Table 8. ACS 2010 5-year Selected Population Tables Variables used to construct the Housing Cost Burden Indicator at the Census Tract Level

Table	Raw Variables and Renamed Variables
DP04	Total Owners and Renters Occupied Housing Units: HC01_VC155=owners HC02_VC155=owners_moe* HC01_VC191=renters HC02_VC191=renters_moe
	Owner occupied housing units and housing cost burden is greater or equal than 30% but less than or equal to 34%: HC01_VC159=owners30to34 (housing units) HC02_VC159=owners30to34_moe HC03_VC159=owners30to34_pct (percent) HC04_VC159=owners30to34_pct_moe Owner occupied housing units and housing cost burden is greater or equal than 35%: HC01_VC160=owners35 HC02_VC160=owners35_moe HC03_VC160=owners35_pct HC04_VC160=owners35_pct_moe
	Renter occupied housing units and housing cost burden is greater or equal than 30% but less than or equal to 34%: HC01_VC196=renters30to34 HC02_VC196=renters30to34_moe HC03_VC196=renters30to34_pct HC04_VC196=renters30to34_pct_moe Renter occupied housing units and housing cost burden is greater or equal than 35%: HC01_VC197=renters35 HC02_VC197=renters35_moe HC03_VC197=renters35_pct HC04_VC197=renters35_pct_moe
B25070	B25070e1: total, renter occupied housing units B25070m1: moe for e1 B25070e10: renter occupied housing units for which gross rent is 50% or more of gross income B25070m10: moe for e10
B25091	B25091e2: owners occupied housing units with a mortgage B25091m2: moe for e2 B25091e11: owners occupied housing units with a mortgage, mortgage cost is 50% or more of gross income B25091m11: moe for e11

*moe stands for margin of error

Percent of Household Crowding

<u>Definition</u>

Full Title: Percent of household overcrowding (> 1.0 persons per room) and severe overcrowding (> 1.5 persons per room)

Indicator ID: 137

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Affordable, high quality, socially integrated and location-efficient housing

Description of significance and health connection:

Residential crowding has been linked to an increased risk of infection from communicable diseases, a higher prevalence of respiratory ailments, and greater vulnerability to homelessness among the poor. Residential crowding reflects demographic and socioeconomic conditions. Older-adult immigrant and recent immigrant communities, families with low income and renter-occupied households are more likely to experience household crowding. A form of residential overcrowding known as "doubling up"—co-residence with family members or friends for economic reasons—is the most commonly reported prior living situation for families and individuals before the onset of homelessness.

Summary of evidence:

Population-based and cohort studies have found adverse associations between residential crowding and health outcomes and family and social relationships. Children in crowded households experienced more conflicts/problems within the family and at school, excessive school absences and lower scores in reading and math exercises. Household crowding is associated with a higher incidence of tuberculosis and prevalence of respiratory conditions. Several studies have identified household crowding to be a major risk factor in measles outbreaks and related-deaths, mumps infection, and meningitis and meningococcal disease.

Key References:

- Burr JA, Mutchler JE. Housing characteristics of older Asian Americans. *J Cross Cult Gerontol* 2012; 27(3):217-37.
- Burr JA, Mutchler JE, Gerst K. Patterns of residential crowding among Hispanics in later life: immigration, assimilation, and housing market factors. *J Gerontol B Psychol Sci Soc Sci* 2010; 65(6):772-82.
- Burström B, Diderichsen F, Smedman L. Child mortality in Stockholm during 1885-1910: the impact of household size and number of children in the family on the risk of death from measles. *Am J Epidemiol* 1999; 149(12):1134-41.
- Clark WAV, Deurloo MC, Dieleman FM. Housing consumption and residential crowding in US housing markets. *J Urban Aff* 2000; 22(1):49-63.
- Marin M, Nguyen HQ, Langidrik JR, Edwards R, Briand K, Papania MJ, et al. Measles transmission and vaccine effectiveness during a large outbreak on a densely populated island: implications for vaccination policy. *Clin Infect Dis* 2006; 42(3):315-9.

- Muhsen K, Aboudy Y, Mendelson E, Green MS, Cohen D. Prevalence of mumps antibodies in the Israeli population in relation to mumps vaccination policy and incidence of disease. *Epidemiol Infect* 2008; 136(5):688-93.
- National Alliance to End Homelessness, Homelessness Research Institute. The State of Homelessness in America; 2014. http://www.endhomelessness.org/page/-/files/2014_State_Of_Homelessness_final.pdf
- Nelson GE, Aguon A, Valencia E, Oliva R, Guerrero ML, Reyes R, et al. Epidemiology of a mumps outbreak in a highly vaccinated island population and use of a third dose of measles-mumps-rubella vaccine for outbreak control--Guam 2009 to 2010. *Pediatr Infect Dis J* 2013; 32(4):374-80.
- Office of the Deputy Prime Minister. The Impact of Overcrowding on Health & Education: A Review of Evidence and Literature. Wetherby, UK; 2004. http://dera.ioe.ac.uk/5073/1/138631.pdf
- Weitzman M, Baten A, Rosenthal DG, Hoshino R, Tohn E and Jacobs DE. Housing and child health. *Curr Probl Pediatr Adolesc HealthCare* 2013; 43:187-224.

Detailed Definition:

- Numerator: number of households with overcrowding (> 1.0 person per room, PPR) or severe overcrowding (> 1.5 PPR).
- Denominator: total number of households.
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: housing tenure (owner-occupied, renter-occupied, and total households); income level of renter/householder (households with a monthly household income ≤ 30% and all levels of HUD-adjusted median family income); race/ethnicity stratification is not available.
- Interpretation: healthier communities will tend to have a low percentage (or zero) of households with overcrowding.

Data Description:

- Data source: U.S. Department of Housing and Urban Development (HUD), Consolidated Planning Comprehensive Housing Affordability Strategy (CHAS) data, Table 10 (http://www.huduser.org/portal/datasets/cp/CHAS/data_download_chas.html).
- Years available: 2006-2010
- Updated: 3 and 5 year intervals
- Geographies available: cities/towns, counties, regions (derived), and state

Methodology:

In CHAS data, household overcrowding and severe household overcrowding estimates were pre-calculated for renter- and owner-occupied households. To derive the percent of household overcrowding (> 1.0 PPR), household estimates from two crowding strata (1.0 – 1.5 PPR and > 1.5 PPR) were summed (the numerator), divided by the denominator and multiplied by 100. For the percent of severe household overcrowding, household estimates for severe overcrowding (the numerator) were divided by the denominator and multiplied by 100. Both derived percentages (the indicator) were calculated for renter-occupied, owner-occupied and total households (includes renter- and owner-occupied households).

The indicators and standard errors were calculated using the approximate method for the geographies of census tract, place, county, region (derived), and state. Relative standard errors (RSE), 95% confidence intervals, and decile ranking of census tracts and places were also calculated. Regions were based on counties of metropolitan transportation organizations (MPO) as reported in the 2010 California Regional Progress Report (Appendix C). Census tract estimates using CHAS and ACS data were statistically unstable (RSE ≥ 30% for the majority of census tracts). As a result, census-tract level data are not presented.

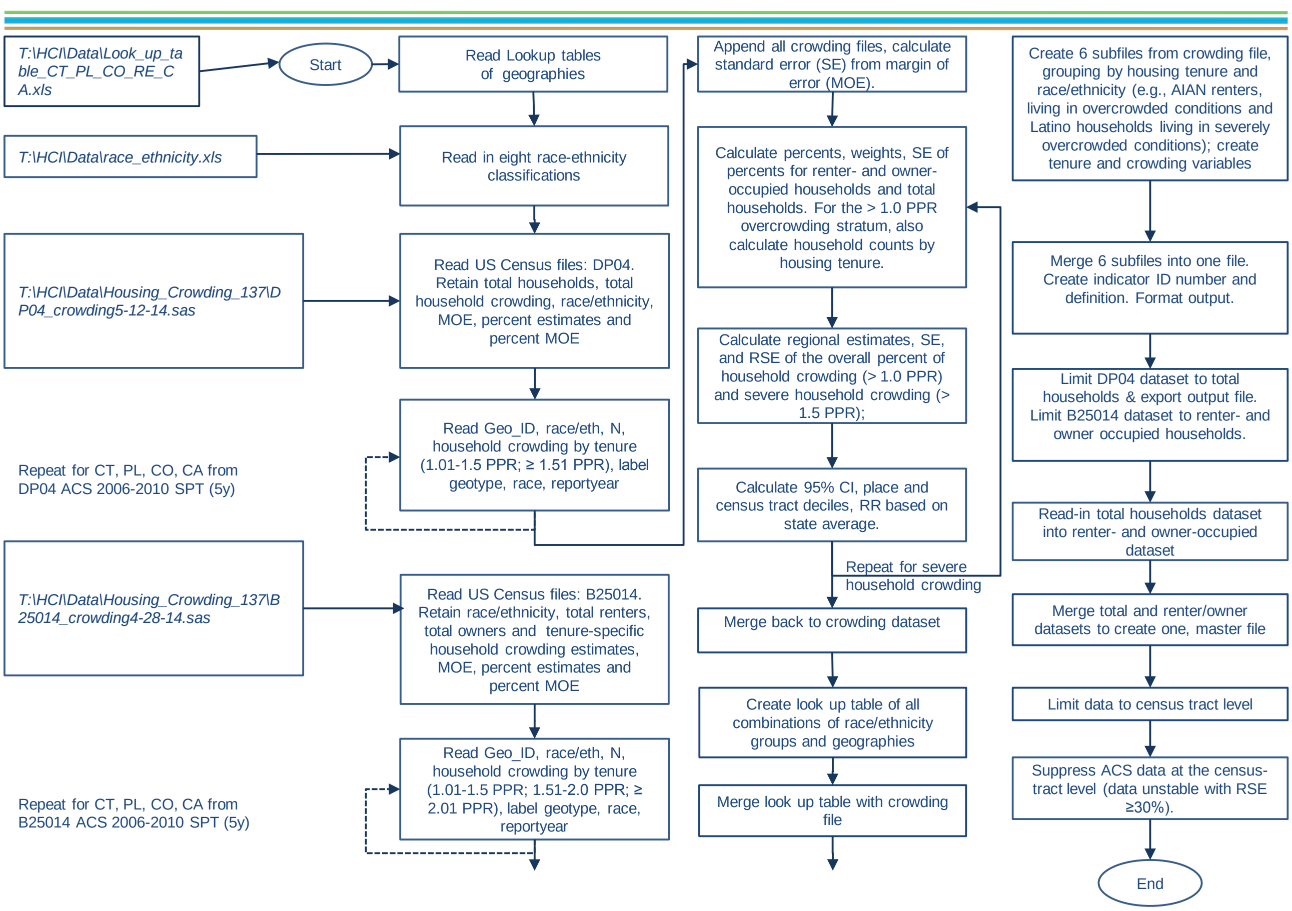
Limitations:

Race/ethnicity data was not available at the time of this analysis. Estimates for the survey period 2006-2010 are bisected by the Great Recession (2008), marked by a large increase in home foreclosures, and house/rental price instability. Due to changes in definitions and sampling, HUD does not recommend making comparisons to prior years' estimates.

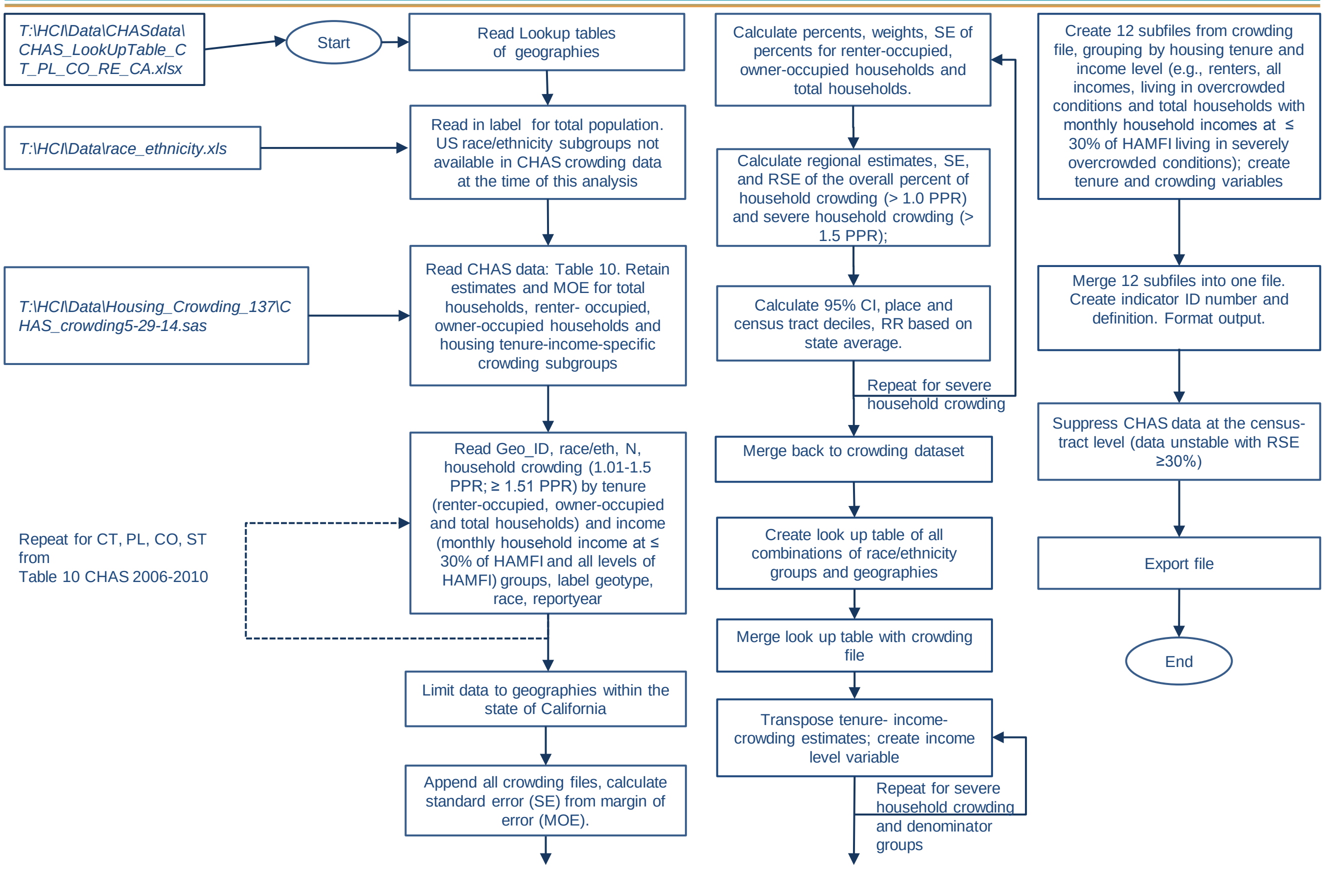
Data Processing

SAS program: CHAS_crowding7-22-14.sas

Data Processing: Percent of Household Crowding, ACS 2006-2010 data



Data Processing: Percent of Household Crowding, CHAS 2006-2010 data



Data Acquisition and Preprocessing

Data for this indicator was acquired following the same methods used for the "Percent of Households Incurring Cost Burdens" indicator. Refer to the instructions for data acquisition for that indicator to download the following tables to construct the "Percent of Household Crowding" indicator:

- U.S. Census Bureau, American Community Survey (ACS), Tables DP04 and B25014
- U.S. Department of Housing and Urban Development (HUD), Consolidated Planning Comprehensive Housing Affordability Strategy (CHAS) data, Table 10.

Table 9. CHAS 2006-2010 (Table 10) Variables used to construct the Household Crowding Indicator by Income for Places, Counties, Regions and the State

Raw Variables and Variable Description
T10_est1: Total occupied housing units T10_est2: Owner occupied housing units T10_est66: Renter occupied housing units
Owner occupied housing units T10_est3: persons per room is less than or equal to 1 T10_est4: persons per room is less than or equal to 1 and household income is less than or equal to 30% of HAMFI T10_est24: persons per room is greater than 1 but less than or equal to 1.5 T10_est25: persons per room is greater than 1 but less than or equal to 1.5 and household income is less than or equal to 30% of HAMFI T10_est45: persons per room is greater than 1.5 T10_est46: persons per room is greater than 1.5 and household income is less than or equal to 30% of HAMFI
Renter occupied housing units T10_est67: persons per room is less than or equal to 1 T10_est68: persons per room is less than or equal to 1 and household income is less than or equal to 30% of HAMFI T10_est88: persons per room is greater than 1 but less than or equal to 1.5 T10_est89: persons per room is greater than 1 but less than or equal to 1.5 less than or equal to 30% of HAMFI T10_est109: persons per room is greater than 1.5 T10_est110: persons per room is greater than 1.5 less than or equal to 30% of HAMFI
T10_moe1: Margin of error for estimate T10_est1 T10_moe2: Margin of error for estimate T10_est2 T10_moe3: Margin of error for estimate T10_est3 T10_moe4: Margin of error for estimate T10_est4 T10_moe24: Margin of error for estimate T10_est24 T10_moe25: Margin of error for estimate T10_est25 T10_moe45: Margin of error for estimate T10_est45 T10_moe46: Margin of error for estimate T10_est46 T10_moe66: Margin of error for estimate T10_est66 T10_moe67: Margin of error for estimate T10_est67 T10_moe68: Margin of error for estimate T10_est68 T10_moe88: Margin of error for estimate T10_est88 T10_moe89: Margin of error for estimate T10_est89 T10_moe109: Margin of error for estimate T10_est109 T10_moe110: Margin of error for estimate T10_est110

Unemployment Rate

Definition

Full Title: Unemployment rate: percent of the population in the labor force who are unemployed

Indicator ID: 290

Healthy Community Framework: Economic and social development

Aspirational Goal: Living wage, safe and healthy job opportunities for all

Description of Significance and Health Connection:

Unemployment is associated with higher rates of self-reported poor health, long-term illnesses, higher incidence of risky health behaviors (alcoholism, smoking), and increased mortality. These negative health outcomes affect not only the unemployed persons but can extend to their families. Longer unemployment can be associated with higher odds of negative health effects. Various explanations have been proposed for the link between poor health and unemployment; for example, economic deprivation that results in reduced access to essential goods and services. Another explanation is that unemployment causes the loss of latent functions (social contact, social status, time structure and personal identity) which can result in stigma, isolation and loss of self-worth. In present times, the safety net available to the unemployed is weaker than in the past due to the deterioration of employment rights, decrease in social support and welfare systems.

The population in the labor force is the civilian non-institutionalized population 16 years and over who have jobs or are actively looking for jobs. Persons in the labor force are classified as unemployed if they do not have a job, are currently available for work and have actively looked for work in the previous month (for instance, attending interviews, sending out resumes, or filling out applications). People that do not have a job and are not looking for one are considered not to be in the labor force. Women, youth (16-24 years of age), the least educated, and ethnic minorities are more likely to be unemployed.

Summary of Evidence:

Aggregate level studies (countries) found a positive association between higher unemployment and overall mortality and death due to cardiovascular disease and suicide; however a negative relationship was detected with deaths due to motor-vehicle accidents. Individual level longitudinal studies showed that the unemployed had higher rates of poor physical health, suicides, mental health problems (depression, stress, anxiety), and greater use of health care services. Other studies found reduced access to health care services and higher likelihood to delay looking for care among the unemployed.

Key References:

- Bambra C, Eikemo TA. Welfare state regimes, unemployment and health: a comparative study of the relationship between unemployment and self-reported health in 23 European countries. *J Epidemiol Community Health* 2009; 63: 92–98.

- Janlert U, Hammarström A. Which theory is best? Explanatory models of the relationship between unemployment and health. *BMC Public Health* 2009; 9:235.
- Bambra C. Yesterday once more? Unemployment and health in the 21st century. *J Epidemiol Community Health* 2010; 64:213e215.
- Bureau of Labor Statistics, Labor Force Statistics from the Current Population Survey, http://www.bls.gov/cps/cps_htgm.htm#concepts, Accessed June 2nd, 2014
- Employment Conditions Knowledge Network (EMCONET). Employment Conditions and Health Inequalities. 2007. Accessed June 2nd, 2014.
- Jin R, Chandrakant PS, Tomislav JS. The impact of unemployment on health: a review of the evidence. *Can Med Assoc J* 1995; 153(5): 529-540.
- Mossakowski K. The influence of past unemployment duration on symptoms of depression among young women and men in the United States. *A J Public Health* 2009; 99(10): 1826-1832.
- Pharr JR, Moonie S, Bungum T. The impact of unemployment on mental and physical health, access to health care and health risk behaviors. *ISRN Public Health 2012*; doi:10.5402/2012/483432.

Detailed Definition:

- Numerator: unemployed population
- Denominator: population in the labor force
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: eight race/ethnicity groups (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white)
- Interpretation: healthier communities will have the lowest percentage of unemployed population.

Data Description:

- Data sources: Local Area Unemployment Statistics (LAUS), Bureau of Labor Statistics (BLS), (http://www.bls.gov/lau/home.htm). U.S. Census Bureau, American Community Survey (ACS), 2006-2010 (http://factfinder2.census.gov).
- Years available: 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2006-2010 (5 year estimate)
- Updated: annually
- Geographies available: Census tracts, cities/towns, counties, county divisions, regions and state

Methodology:

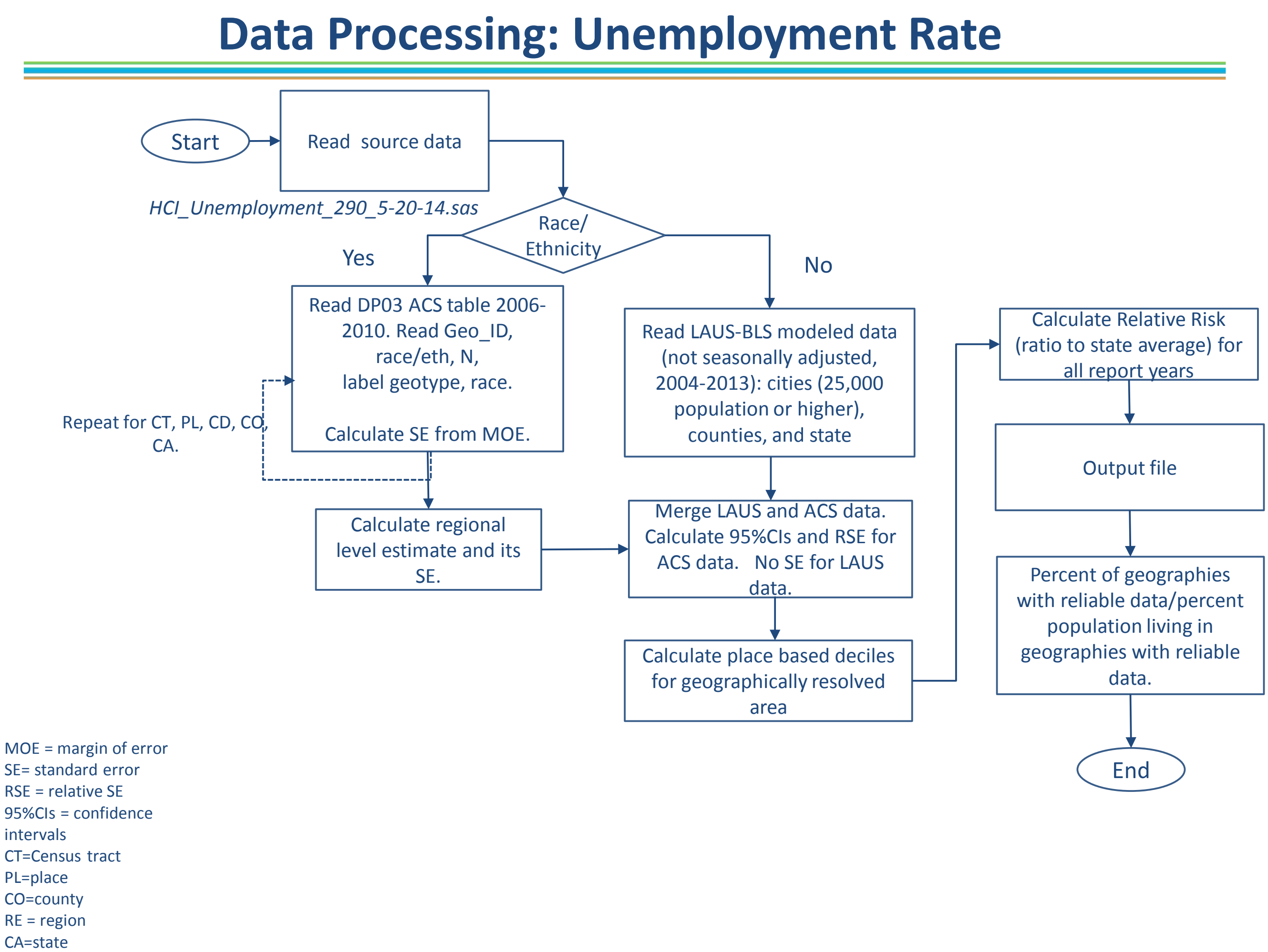
Annual estimates of the population in the labor force (denominator), unemployed population (numerator), and unemployment rate were downloaded for cities and towns (population 25,000 or above), counties, and the state from the LAUS database. Five year estimates of unemployment data (2006-2010) with race ethnicity stratification were obtained from the ACS (DP03 table). Standard errors were obtained for the ACS data, and relative standard errors and 95% upper and lower confidence intervals were calculated. Decile rankings of places and relative risk in relation to state average were calculated. Regions were based on counties of metropolitan transportation organizations (MPO) regions as reported in the 2010 California Regional Progress Report (Appendix C).

Limitations:

The definition of unemployment excludes people that do not have a job but are not actively looking for one and people living in group quarters. The annual unemployment estimates provided by BLS-LAUS for sub-national areas are modeled from country level data from the Current Population Survey and no standard error estimates are available. More details can be found here: http://www.bls.gov/lau/laumthd.htm and http://www.bls.gov/opub/hom/homch4.htm. The modeled unemployment data is not seasonally adjusted. Information about differences in the methods to estimate unemployment between BLS-LAUS and the ACS can be found here: http://www.bls.gov/lau/acsqa.htm.

Data Processing

SAS file: HCI_Unemployment_290_5-20-14.sas



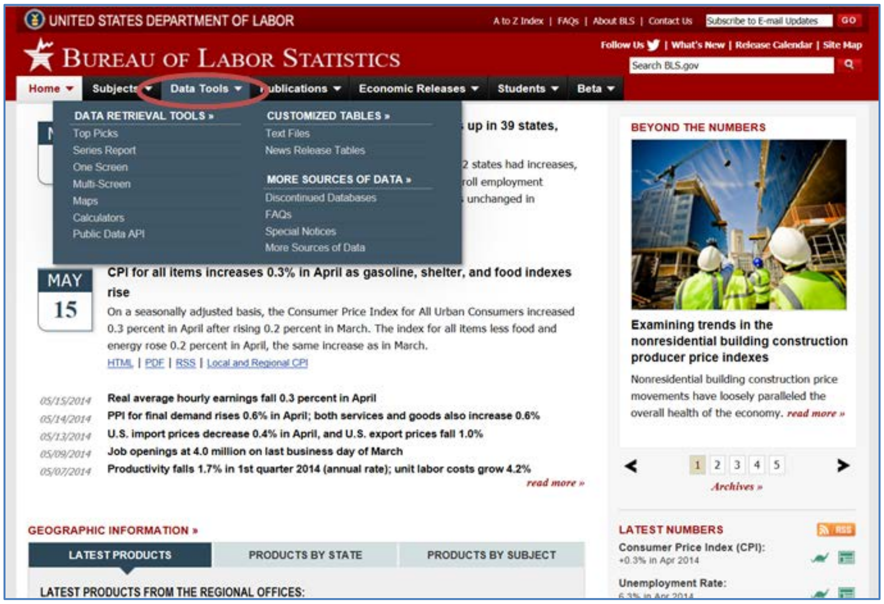
Data Acquisition and Preprocessing

Bureau of Labor Statistics – Local Area Unemployment Statistics

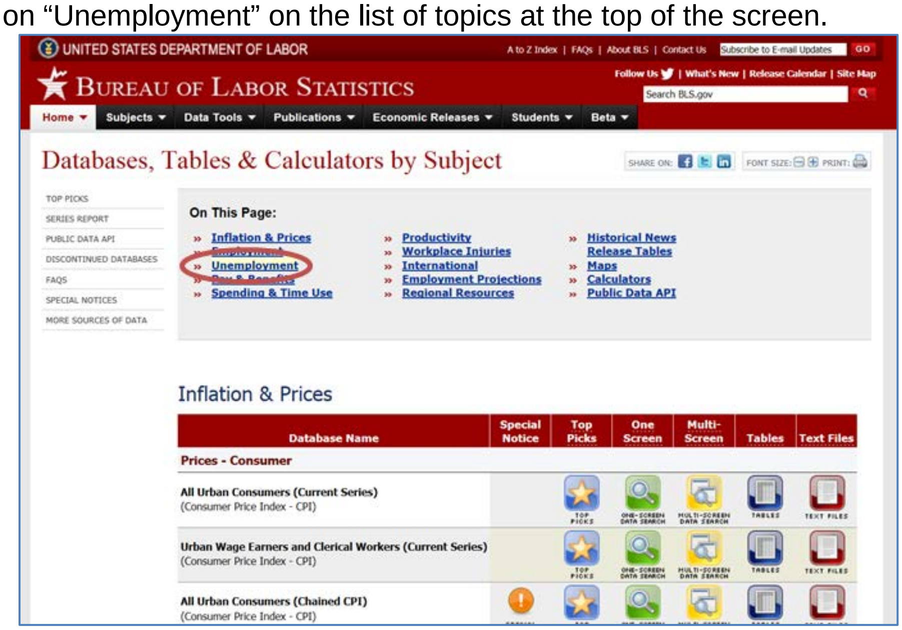
The following steps exemplify how to extract BLS-LAUS data for cities (population > 25,000) using the multi-screen data search feature in the BLS website. The same steps can be followed to extract data for counties and the state.

1. Go to the BLS home page http://www.bls.gov/home.htm.

2. Click on “Data Tools”.

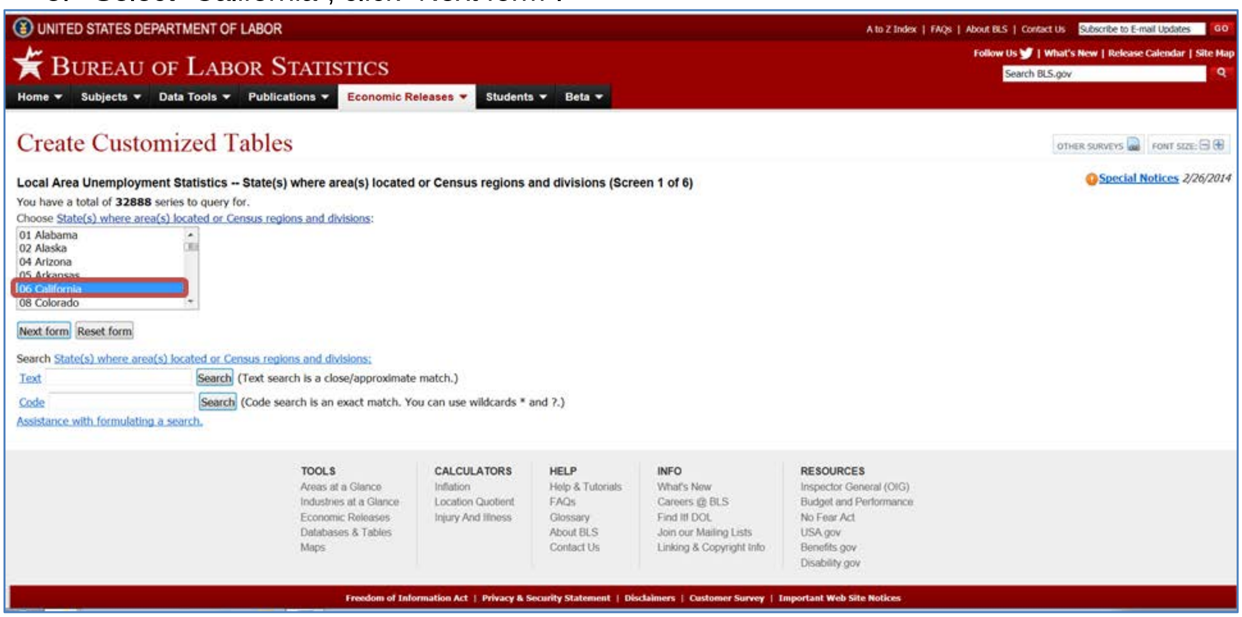


3. Click on “Unemployment” on the list of topics at the top of the screen.

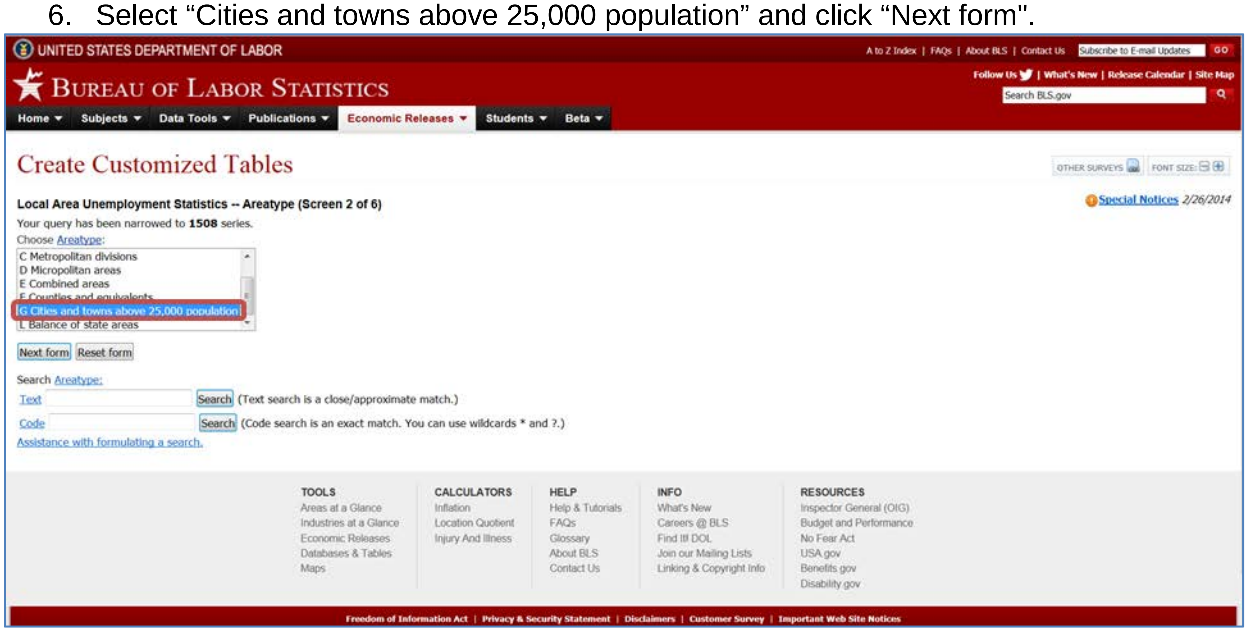


4. Click on Local Area Unemployment Statistics (LAUS) “Multi screen data search”.

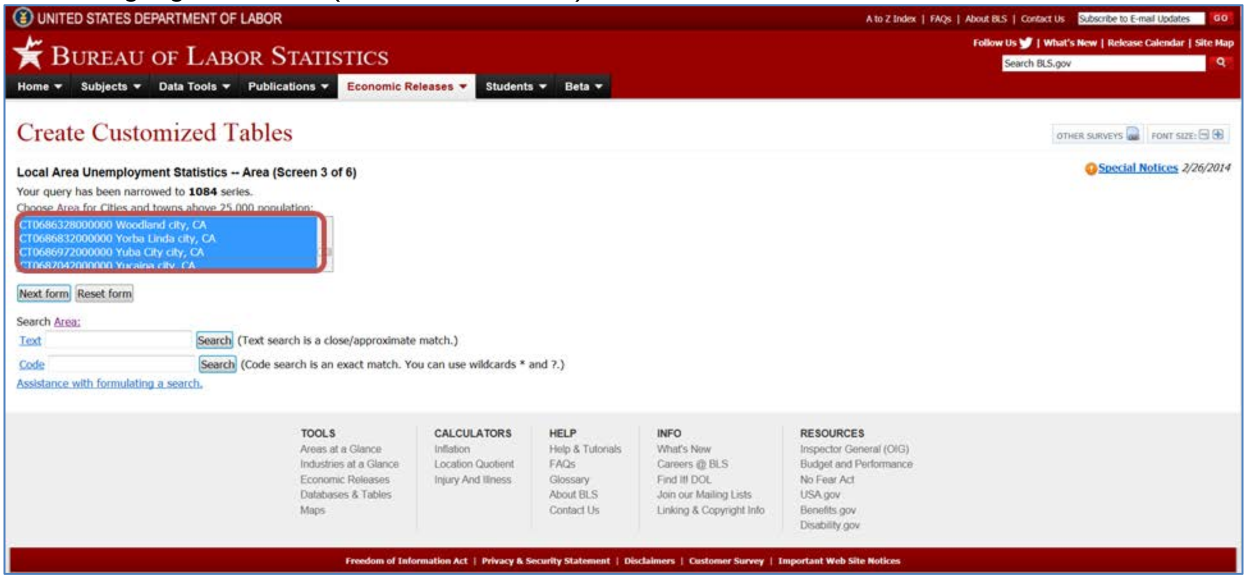
5. Select “California”, click “Next form”.



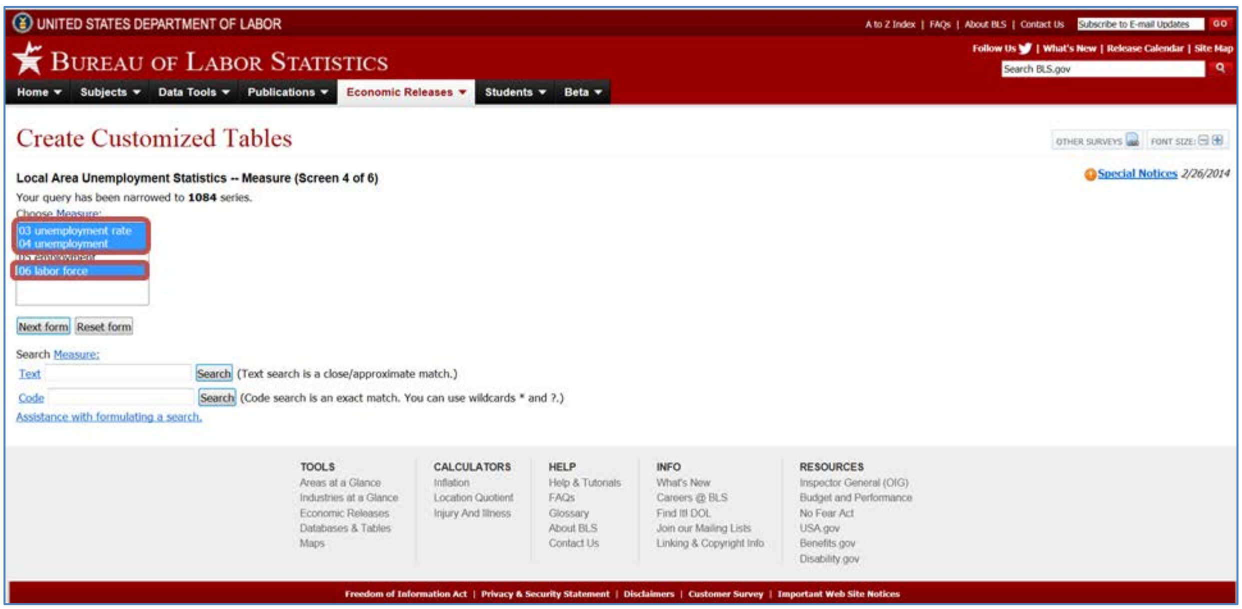
6. Select “Cities and towns above 25,000 population” and click “Next form".



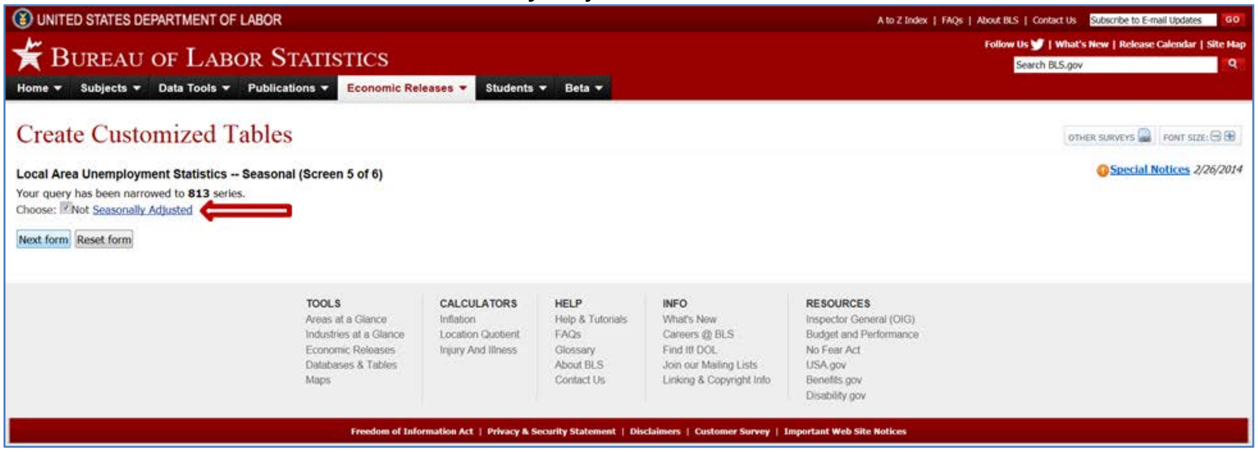
7. Highlight all cities (Shift+Down Arrow), click “Next form”.



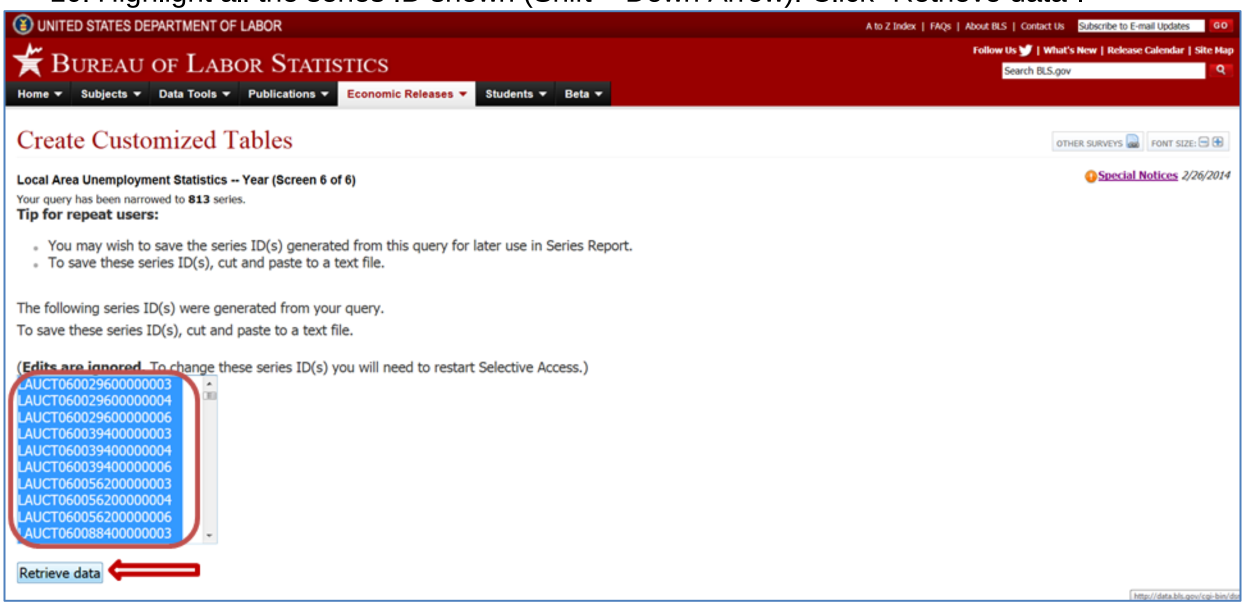
8. Highlight “unemployment rate”, “unemployment”, “labor force”, (Ctrl + Click on term) and click “Next form”.



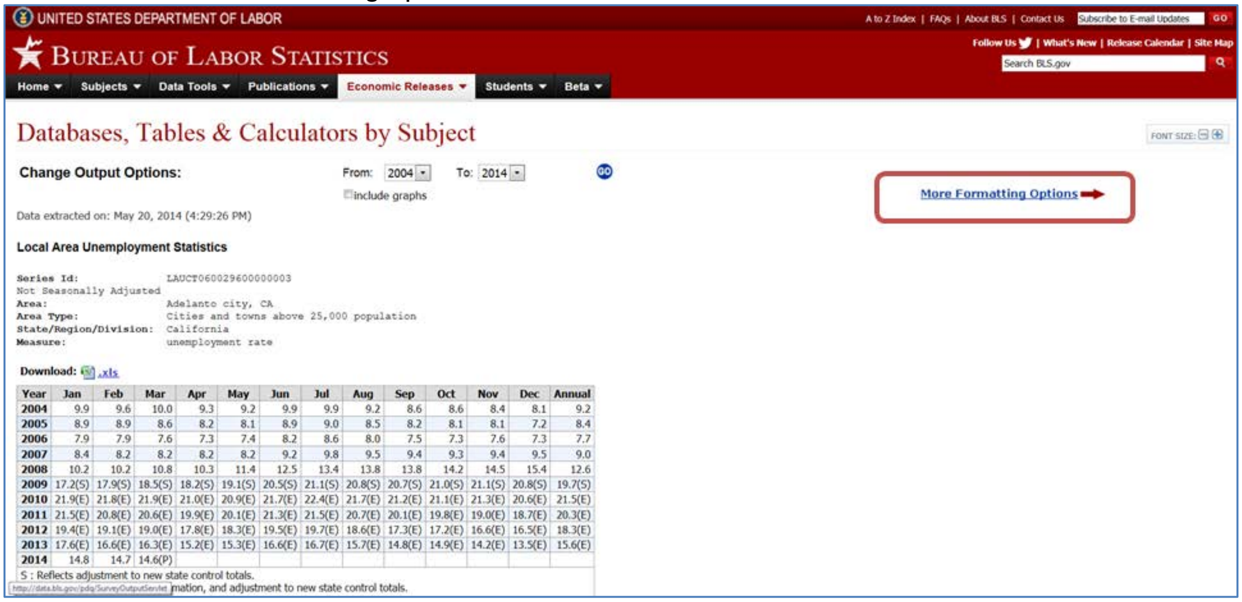
9. Check the box for “Not Seasonally Adjusted”, click “Next form”.



10. Highlight all the series ID shown (Shift + Down Arrow). Click “Retrieve data”.



11. Click “More formatting options”.

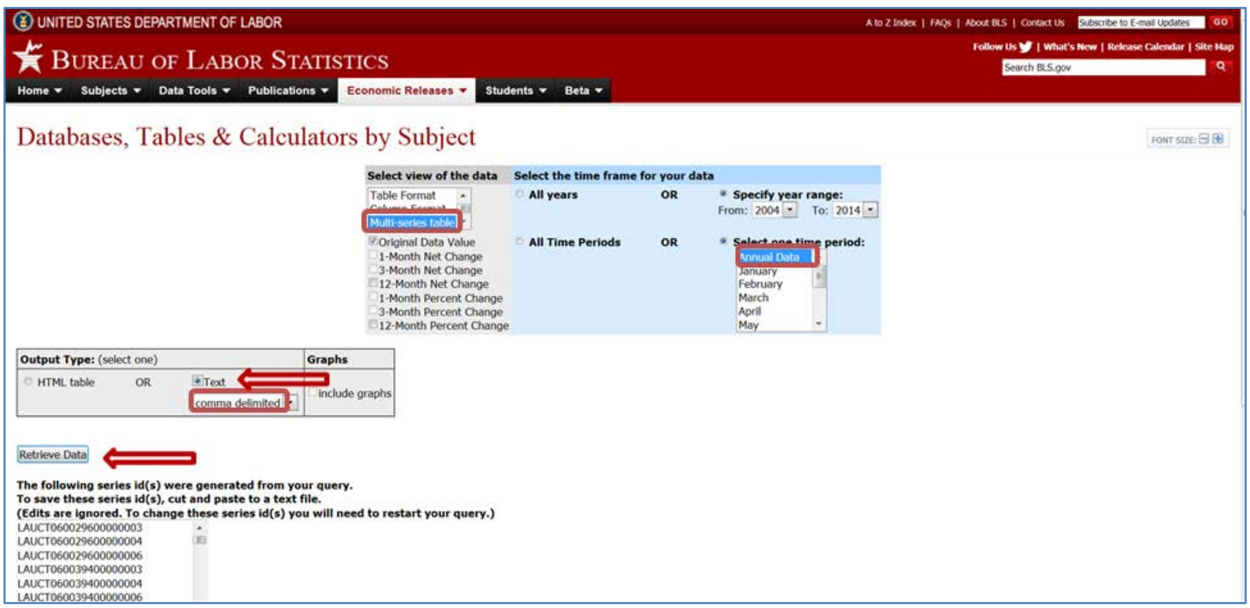


Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
2004	9.9	9.6	10.0	9.3	9.2	9.9	9.9	9.2	8.6	8.6	8.4	8.1	9.2
2005	8.9	8.9	8.6	8.2	8.1	8.9	9.0	8.5	8.2	8.1	8.1	7.2	8.4
2006	7.9	7.9	7.6	7.3	7.4	8.2	8.6	8.0	7.5	7.3	7.6	7.3	7.7
2007	8.4	8.2	8.2	8.2	8.2	9.2	9.8	9.5	9.4	9.3	9.4	9.5	9.0
2008	10.2	10.2	10.8	10.3	11.4	12.5	13.4	13.8	13.8	14.2	14.5	15.4	12.6
2009	17.2(S)	17.9(S)	18.5(S)	18.2(S)	19.1(S)	20.5(S)	21.1(S)	20.8(S)	20.7(S)	21.0(S)	21.1(S)	20.8(S)	19.7(S)
2010	21.9(E)	21.8(E)	21.9(E)	21.0(E)	20.9(E)	21.7(E)	22.4(E)	21.7(E)	21.2(E)	21.1(E)	21.3(E)	20.6(E)	21.5(E)
2011	21.5(E)	20.8(E)	20.6(E)	19.9(E)	20.1(E)	21.3(E)	21.5(E)	20.7(E)	20.1(E)	19.8(E)	19.0(E)	18.7(E)	20.3(E)
2012	19.4(E)	19.1(E)	19.0(E)	17.8(E)	18.3(E)	19.5(E)	19.7(E)	18.6(E)	17.3(E)	17.2(E)	16.6(E)	16.5(E)	18.3(E)
2013	17.6(E)	16.6(E)	16.3(E)	15.2(E)	15.3(E)	16.6(E)	16.7(E)	15.7(E)	14.8(E)	14.9(E)	14.2(E)	13.5(E)	15.6(E)
2014	14.8	14.7	14.6(P)										

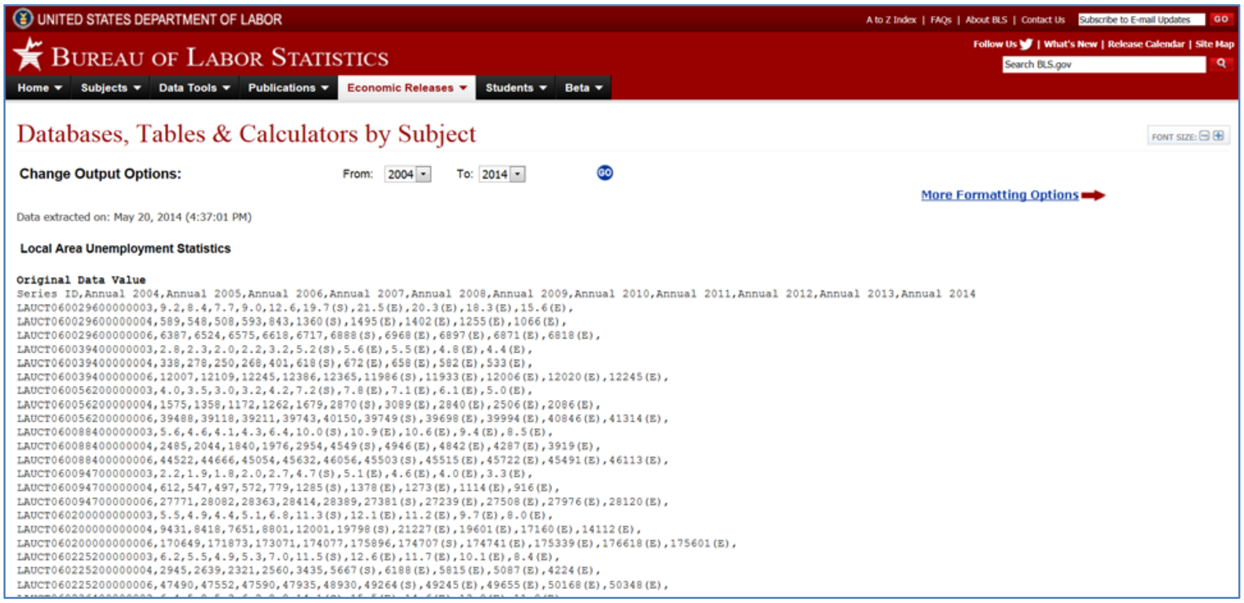
S : Reflects adjustment to new state control totals.

http://data.bls.gov/pdq/SurveyOutputServlet mation, and adjustment to new state control totals.

12. Select “Multi-series table”, “Annual Data”, “Output Type=text”, “comma delimited”, as shown below and then click “Retrieve data”.



13. Results will look like this:


UNITED STATES DEPARTMENT OF LABOR

A to Z Index | FAQs | About BLS | Contact Us | Subscribe to E-mail Updates | GO

BUREAU OF LABOR STATISTICS

Follow Us | What's New | Release Calendar | Site Map

Search BLS.gov

Home | Subjects | Data Tools | Publications | Economic Releases | Students | Beta

Databases, Tables & Calculators by Subject

FONT SIZE:

Change Output Options: From: 2004 To: 2014 GO

More Formatting Options

Data extracted on: May 20, 2014 (4:37:01 PM)

Local Area Unemployment Statistics

```
Original Data Value
Series ID,Annual 2004,Annual 2005,Annual 2006,Annual 2007,Annual 2008,Annual 2009,Annual 2010,Annual 2011,Annual 2012,Annual 2013,Annual 2014
LAUCT060029600000003,9.2,8.4,7.7,9.0,12.6,19.7(S),21.5(E),20.3(E),18.3(E),15.6(E),
LAUCT060029600000004,589,548,508,593,843,1360(S),1495(E),1402(E),1255(E),1066(E),
LAUCT060029600000006,6387,6524,6575,6618,6717,6888(S),6968(E),6897(E),6871(E),6818(E),
LAUCT060039400000003,2.8,2.3,2.0,2.2,3.2,5.2(S),5.6(E),5.5(E),4.8(E),4.4(E),
LAUCT060039400000004,338,278,250,268,401,618(S),672(E),658(E),582(E),533(E),
LAUCT060039400000006,12007,12109,12245,12386,12365,11986(S),11933(E),12006(E),12020(E),12245(E),
LAUCT060056200000003,4.0,3.5,3.0,3.2,4.2,7.2(S),7.8(E),7.1(E),6.1(E),5.0(E),
LAUCT060056200000004,1575,1358,1172,1262,1679,2870(S),3089(E),2840(E),2506(E),2086(E),
LAUCT060056200000006,39488,39118,39211,39743,40150,39749(S),39698(E),39994(E),40846(E),41314(E),
LAUCT060088400000003,5.6,4.6,4.1,4.3,6.4,10.0(S),10.9(E),10.6(E),9.4(E),8.5(E),
LAUCT060088400000004,2485,2044,1840,1976,2954,4549(S),4946(E),4842(E),4287(E),3919(E),
LAUCT060088400000006,44522,44666,45054,45632,46056,45503(S),45515(E),45722(E),45491(E),46113(E),
LAUCT060094700000003,2.2,1.9,1.8,2.0,2.7,4.7(S),5.1(E),4.6(E),4.0(E),3.3(E),
LAUCT060094700000004,612,547,497,572,779,1285(S),1378(E),1273(E),1114(E),916(E),
LAUCT060094700000006,27771,28082,28363,28414,28389,27381(S),27239(E),27508(E),27976(E),28120(E),
LAUCT060200000000003,5.5,4.9,4.4,5.1,6.8,11.3(S),12.1(E),11.2(E),9.7(E),8.0(E),
LAUCT060200000000004,9431,8418,7651,8801,12001,19798(S),21227(E),19601(E),17160(E),14112(E),
LAUCT060200000000006,170649,171873,173071,174077,175896,174707(S),174741(E),175339(E),176618(E),175601(E),
LAUCT060225200000003,6.2,5.5,4.9,5.3,7.0,11.5(S),12.6(E),11.7(E),10.1(E),8.4(E),
LAUCT060225200000004,2945,2639,2321,2560,3435,5667(S),6188(E),5815(E),5087(E),4224(E),
LAUCT060225200000006,47490,47552,47590,47935,48930,49264(S),49245(E),49655(E),50168(E),50348(E),
```

14. Copy and paste the displayed table into a text file (Notepad). Later open in Excel in a comma delimited file and save as .xlsx.

15. Each row of the table has a Series ID. An explanation (data dictionary) of LAUS series id formats can be found here: http://www.bls.gov/help/hlpforma.htm#LA. The characters at specific positions of the Series ID indicate the information and geographical location of the data.

Local Area Unemployment Statistics

Survey Overview The following is a sample format description of the Local Area Unemployment Statistics' series identifier:

```
                         1         2
              12345678901234567890
Series ID     LAUCN281070000000003
Positions    Value            Field Name
1-2          LA               Prefix
3            U                Seasonal Adjustment Code
4-5          CN               Area Type Code
6-18         2810700000000    Area Code
19-20        03               Measure Code
```

To assist you in formatting series IDs, access any of the following for a list of codes and their corresponding titles:

- Area Type Code
- Area Codes
- Measure Codes

For example, Pasadena City in Los Angeles County will have three rows in the recently created table. Each row will be identified by a Series ID. Here is how to interpret the IDs:

LAUCT065600000000003	Position 3, U=not seasonally adjusted Positions 4-5, CT= Cities and towns above 25,000 population Positions 6-7, 06=state of California Positions 8-12, 56000=Census place code Positions 19-20, 03=unemployment rate
LAUCT065600000000004	Positions 3 to 12, same as above Positions 19-20, 04=unemployment
LAUCT065600000000006	Positions 3 to 12, same as above Positions 19-20, 06=labor force

The crosswalk of series id and geographical names can be found here: http://download.bls.gov/pub/time.series/la/la.area.

American Community Survey, DP03 table

Follow instructions on how to download data from the ACS in the Toolbox (Downloading Census Data from American Fact Finder). Download table DP03, for the dataset 2010 ACS 5-year Selected Population Tables (2006-2010), for all geographical levels of interest (Census tract, place, counties, and the state), including all race ethnicities.

Variables of interest from the DP03 table: HC01_VC08, unemployed civilian labor force; HC01_VC12, civilian labor force; HC03_VC13, percent unemployment; and HC04_VC13 - percent margin of error. The standard error of the percent unemployment was calculated by dividing the margin of error by a factor of 1.645.

Educational Attainment

Definition

Full Title: High School or Greater Educational Attainment in the Population Aged 25 Years and Older

Indicator ID: 369

Healthy Community Framework: Adequate levels of economic and social development

Aspirational Goal: Opportunities for high quality and accessible education

Description of significance and health connection:

Education has broad impacts on standards of living and social interactions, with consequences for the health of individuals and communities. Through three inter-related pathways, education influences health: health knowledge and behaviors; employment and income; and social and psychological factors. Completion of formal education (e.g., high school) is a key pathway to employment and access to healthier and higher paying jobs that can provide food, housing, transportation, health insurance, and other basic necessities for a healthy life. Education is linked with social and psychological factors, including sense of control, social standing and social support. These factors can improve health through reducing stress, influencing health-related behaviors and providing practical and emotional support.

It is estimated that approximately 245,000 (10%) of the 2.4 million U.S. deaths in 2000 were attributable to low education. Low education is also associated with poorer self-rated health status, higher infant mortality rates, lower cancer screening rates, and many other health outcomes and health behaviors. It is estimated that raising the health of all Americans to that of college educated Americans would result in annual gains of over 1 trillion dollars of increased health value. Health burdens due to low educational attainment disproportionately impact African Americans and other race/ethnicities whose share of high school graduates is lower than Whites.

Summary of evidence:

There is an extensive body of research based on epidemiologic studies using national vital statistics (births and deaths) and health surveys, including the California Health Interview Survey, that establish the relationship between education and health. A recent study synthesized 24 population-based studies of mortality and education in the United States. Several of these studies were longitudinal studies following healthy populations over time. Greater educational attainment has been associated with health-promoting behaviors including consumption of fruits and vegetables and other aspects of healthy eating, engaging in regular physical activity, and refraining from excessive consumption of alcohol and from smoking.

Key References:

- Egerter S, Braveman P, Sadegh-Nobari T, Grossman-Kahn R, Dekker M. Issue Brief 6: Education and Health. Robert Wood Johnson Foundation Commission to Build a Healthier America. September 2009.

- Galea S, Tracy M, Hoggatt KJ, DiMaggio C, Karpati A. Estimated deaths attributable to social factors in the United States. Am J Public Health 2011;101(8):1456-1465.
- Dow W, Schoeni R. *Economic Value of Improving the Health of Disadvantaged Americans. Technical Report for Overcoming Obstacles to Health: Report from the Robert Wood Johnson Foundation to the Commission to Build a Healthier America*: Robert Wood Johnson Foundation 2008.
- California Health Interview Survey. Public Use Datafiles. Los Angeles, CA: UCLA Center for Health Policy Research (http://healthpolicy.ucla.edu/contact/Pages/default.aspx, Accessed on February 4, 2013).

Detailed definition:

- Numerator: population 25 years and older completing high school or associate, bachelors, or a more advanced degree
- Denominator: population 25 years and older
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: eight race/ethnicity groups (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white).
- Interpretation: healthier communities will have a higher percentage of adults that have completed high school or more advanced degrees

Data Description:

- Data Source: Educational Attainment Data from U.S. Census 2000 (Summary File - SF3): Table QT-P20; Educational Attainment Data from 2005-2007 and 2008-2010 American Community Survey: Table S1501 and DP02 for 2006-2010. All data downloaded from American Factfinder at http://factfinder2.census.gov
- Years available: 2000, 2005-2007, 2008-2010, 2006-2010
- Updated: 3 and 5 year intervals
- Geographies available: Census tracts, cities/towns, counties, regions, and state

Methodology:

Five-year data (2006-2010) at census tract and cities/towns geographies for this indicator and its margin of error were downloaded from the American Community Survey (ACS) website. To create a 3-point time series of California places and counties, data were downloaded from the U.S. Census (2000) and from the ACS (2005-2007 and 2008-2010). Standard errors, relative standard errors, and 95% upper and lower confidence intervals were calculated where margins of error were reported in ASC files. Regional estimates were based on county groupings associated with California metropolitan planning organizations as reported in the 2010 California Regional Progress Report (Appendix C).

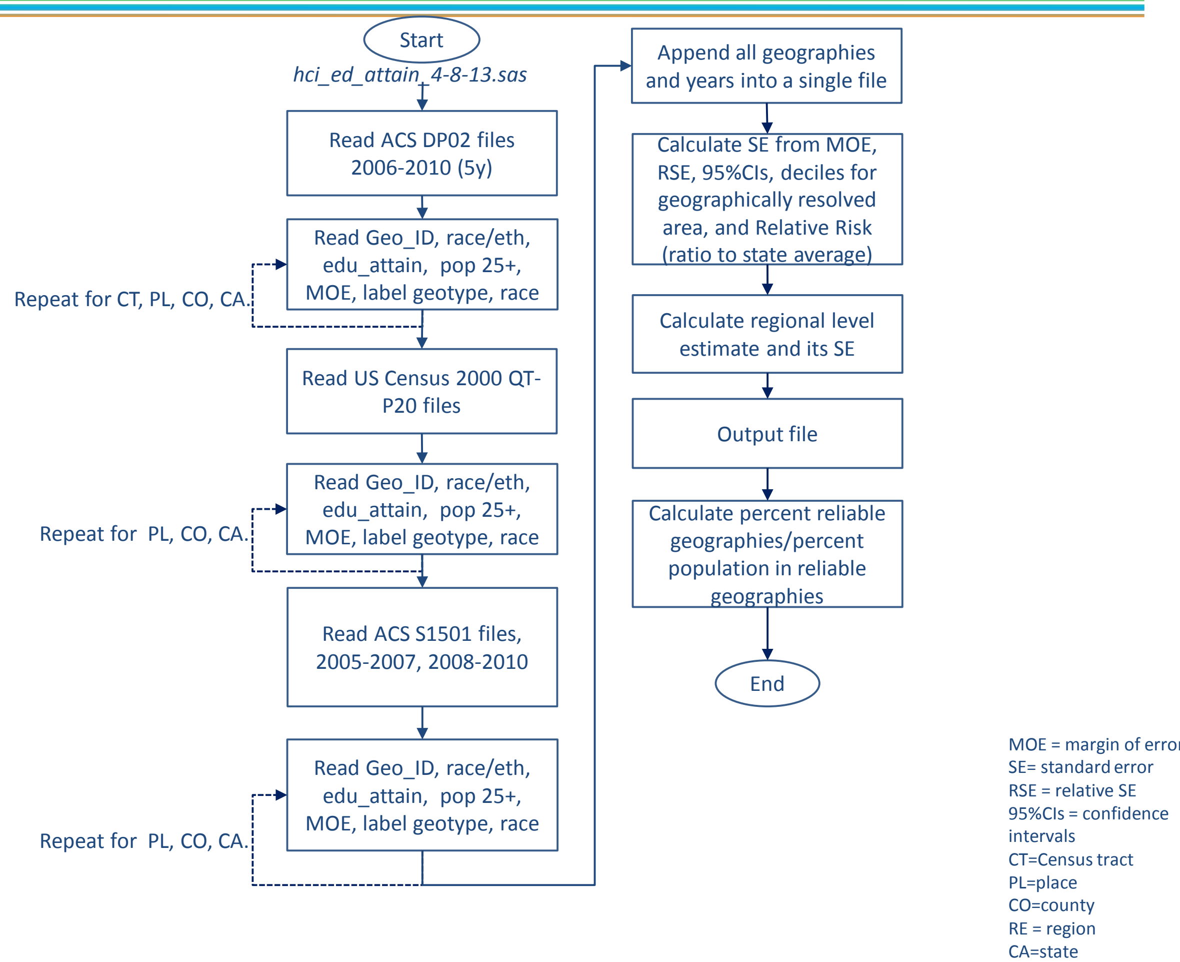
Limitations:

The American Community Survey does not publish data when there are fewer than 100 individuals in a geographic reporting area and when there are fewer than 50 respondents to survey questions. Data are not available for many small California cities and census tracts for race/ethnicity groups with relatively small numbers, generally American Indian/Alaska Native, Hawaiian Native and Other Pacific Islander, Other Race, and Multiple Races.

Data Processing

SAS file: hci_ed_attain_4-8-13.sas

Data Processing: Educational Attainment



Data Acquisition and Preprocessing

The quick table QT-P20 from the 2000 Census, profile table DP02, and detailed table S1501 from the American Community Survey (2005-2007, 2008-2010, and 2006-2010) were downloaded using the American FactFinder (See Toolbox section, Downloading Census Data from American Fact Finder chapter). These tables correspond to the Census topic “Educational Attainment.” A total of 17 Census tables representing four datasets (time periods), four geographies, and eight race/ethnicities and the total population, were downloaded (See Table 10). The variables in the tables that contained the information of interest were identified prior to data processing (See Table 11) using the metadata tables downloaded from the Census.

Table10. Census Tables Downloaded from American Fact Finder for the Educational Attainment Indicator by Data Set, Geographies and Race and Ethnic Groups

Dataset (Time Period)	Geographies	Race and Ethnic Groups	Census Table
Census 2000 Summary File 3 (SF 3) - Sample Data	Place – 160	Total population	QT-P20
	County – 050	Total population	QT-P20
	State – 040	Total population	QT-P20
2007 ACS 3-year estimates (2005-2007)	Place – 160	Total population Total population	DP02 S1501
	County – 050	Total population Total population	DP02 S1501
	State – 040	Total population	DP02
2010 ACS 3-year estimates (2008-2010)	Place – 160	Total population Total population	DP02 S1501
	County – 050	Total population Total population	DP02 S1501
	State – 040	Total population	DP02
2010 ACS 5-year estimates Selected Population Tables (2006-2010)	Census tract – 140	Total and eight race/ethnicity groups	DP02
	Place – 160	Total and eight race/ethnicity groups	DP02
	County – 050	Total and eight race/ethnicity groups	DP02
	State – 040	Total and eight race/ethnicity groups	DP02

Table 11. Variables Used to Construct the Educational Attainment Indicator by Census Table

Census Table	Variable Denominator	Variable Numerator	Variable Percent (Indicator)	Margin of Error of the Percent	Standard Error of the Percent
DP02	HC01_VC84 Estimate; EDUCATIONAL ATTAINMENT - Population 25 years and over	Calculated: Variable Denominator * Variable Percent /100	HC03_VC93 Estimate; EDUCATIONAL ATTAINMENT - Percent high school graduate or higher	HC04_VC93	$\frac{HC04\_VC93}{1.645}$
QT-P20	HC01_VC07 Both sexes; EDUCATIONAL ATTAINMENT (highest level) - Population 25 years and over	Calculated: Variable Denominator * Variable Percent /100	HC01_VC31 Both sexes; EDUCATIONAL ATTAINMENT (highest level) - Percent of population 25 years and over - Percent high school graduate or higher	Not available	Not available
S1501	HC01_EST_VC06 Total; Estimate; Population 25 years and over	Calculated: Variable Denominator * Variable Percent /100	HC01_EST_VC14 Total; Estimate; Population 25 years and over - Percent high school graduate or higher	HC01_MOE_VC14	$\frac{HC01\_MOE\_VC14}{1.645}$

Unsafe Drinking Water

Definition

Full Title: Percent of the population served by community water systems not meeting regulations of the Safe Drinking Water Act

Indicator ID: 426

Healthy Community Framework: Quality and sustainability of environment

Aspirational Goal: Clean air, soil and water, and environments free of excessive noise

Description of significance and health connection:

Drinking water is water that is used for domestic purposes, drinking, cooking, and personal hygiene. Drinking water is safe when it meets standards that limit the presence of harmful microbes and chemicals. Public Water Systems (PWS) are subdivided in community and non-community systems. A community water system (CWS) supplies water to the same population year-round. It serves at least 25 people at their primary residences or at least 15 residences that are primary residences.

All PWS are required to follow the standards and regulations set by the U.S. Environmental Protection Agency (USEPA) under the Safe Drinking Water Act (SDWA). Water contaminants are classified as microorganisms, water disinfectants, disinfection byproducts, inorganic chemicals, organic chemicals, and radionuclides. The USEPA sets limits on water maximum contaminant levels (MCLs), maximum residual disinfectant levels (MRDLs), regulates treatment techniques (TT), and regulates the compliance of PWS to monitoring and reporting (M/R) rules.

Over 20% of PWS in the U.S. failed provisions of the SDWA from 2004 to 2009, with the majority of the violations occurring within small and rural water systems. Unincorporated, low-income, rural communities of color spend over 10% of their annual incomes purchasing water. In California for instance, the Central and Salinas valleys use groundwater exposed to contamination by agricultural fertilizers and livestock manure. This contamination financially burdens the poorest mostly Latino communities.

Summary of evidence:

There were 33 drinking water-associated pathogen outbreaks (mainly *Giardia*, *Legionella*, Norovirus, *Shigella*, and *Campylobacter*) in the U.S. in 2009-2010 causing 1,040 cases of illness, 85 hospitalizations, and 9 deaths. California reported one Norovirus outbreak in a non-community PWS in 2010. The most common violations reported for California PWSs in 2008-2011 were 34% total coliform rule violations, 34% arsenic, and 10% nitrates. Prolonged exposures to arsenic (above the MCL) can cause skin damage, problems with the circulatory system, or an increased risk of certain cancers. Excessive levels of nitrate can cause serious illness, and in rare instances death in infants less than six-months of age. The presence of fecal coliforms is usually associated with direct contamination by sewage or animal wastes and indicates possible contamination with organisms that can cause disease.

Key References:

- World Health Organization. Health through Safe Drinking Water and Basic Sanitation. Accessed April 21st, 2014.
- Centers for Disease Control and Prevention. Public Water Systems. Accessed April 21st, 2014.
- United States Environmental Protection Agency. Drinking Water Contaminants. Accessed April 21st, 2014.
- Pannu C. Drinking Water and Exclusion: A Case Study from California's Central Valley. *Cal L Rev* 2012; 100(223): 227-268.
- California State Water Resources Control Board. Recommendations Addressing Nitrate in Groundwater (Report to the Legislature), 2013. Accessed April 21st, 2014.
- Centers for Disease Control and Prevention. Surveillance for Waterborne Disease Outbreaks Associated with Drinking Water and Other Nonrecreational Water — United States, 2009–2010. *MMWR* 2013; 62:714-720.
- California Department of Public Health. Division of Drinking Water and Environmental Management. Annual Compliance Reports. 2008, 2009, 2010, and 2011. Accessed June 19th, 2014.

Detailed Definition:

- Numerator: population living in a geographic area served by a community water system with 1 or more violations of the MCL/TT in a time period
- Denominator: population living in a geographic area served by a community water system
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: eight race/ethnicity groups (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white). Five violation categories (Treatment Technique – tt; MCL Total Coliform Rule-Bacteriologic – tcrb; arsenic, nitrates and all other MCL- other).
- Interpretation: healthier communities will have the lowest (or zero) percentage of the population served by community water systems with at least one violation.

Data Description:

- Data sources: California Department of Public Health (CDPH), Division of Drinking Water and Environmental Management (DDWEM), Drinking Water Program (DWP, http://www.cdph.ca.gov/programs/Pages/DWP.aspx), Annual Compliance Reports (http://www.cdph.ca.gov/certlic/drinkingwater/Pages/Publications.aspx). CDPH, California Environmental Health Tracking Program (CEHTP), Water System Boundary Tool (http://www.ehib.org/page.jsp?page_key=762), Water System Boundary shapefile (v. 11/11/2013). Department of Finance (DOF), Demographic Research Unit, Redistricting files (http://www.dof.ca.gov/research/demographic/state_census_data_center/census_2010/#PL94).
- Years available: 5 year aggregate 2008-2012
- Updated: every 5 years
- Geographies available: Census tracts, cities/towns, counties, county divisions, regions and state.

Methodology:

Part 1: Data on PWS with MCL/TT violations for any contaminant between 2008 and 2012 was obtained from the DWP-Annual Compliance Reports. Violations were classified in five categories: Treatment Technique (tt), MCL Total Coliform Rule-Bacteriologic (tcrb), arsenic, nitrates and all other MCL (other). A subset of data including only CWS and violations information was joined to the Water System Boundary shapefile using ArcMap. New shapefiles containing only CWS (CWSs) and CWS with violations by category (CWSVs) were created. Census block centroids (2010) were spatially joined to the polygons in CWSs and CWSVs to obtain the blocks in geographic areas served by CWS and the blocks in areas served by a CWS with violations. Part 2: These Census blocks were imported into SAS and merged with population (2010) data by block. The total population in geographic areas served by CWS (denominator) and the total population in geographic areas served by CWS with violations (numerator) were calculated for Census tracts, places, counties, county divisions, regions and state. Regions were based on counties of metropolitan transportation organizations (MPO) regions as reported in the 2010 California Regional Progress Report (Appendix C). Standard errors, relative standard errors, and 95% upper and lower confidence intervals, decile rankings of places and relative risk in relation to state average were calculated. The percentage of the population covered by CWS for which data was available was calculated.

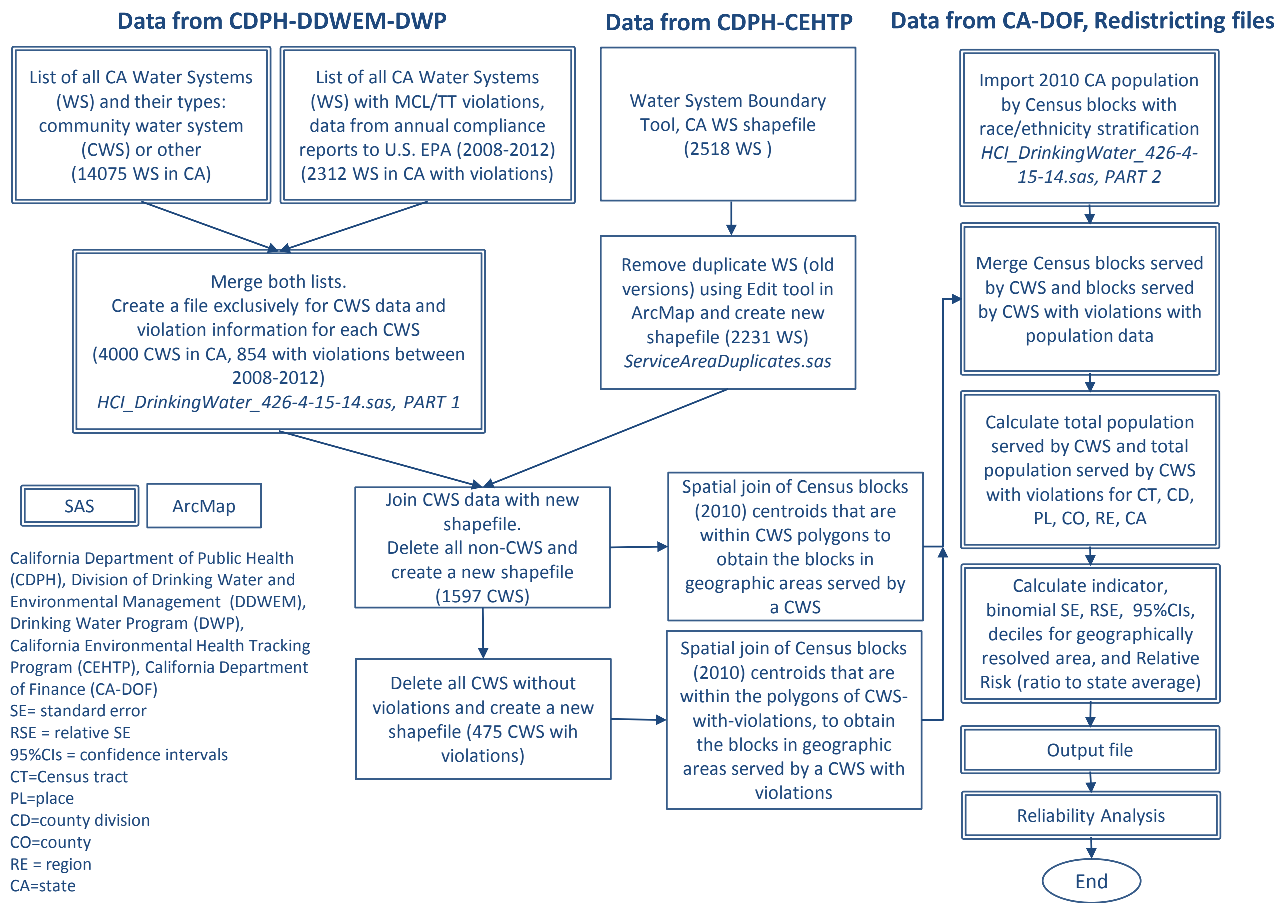
Limitations:

This indicator measures water availability but not consumption. The violations in the Annual Compliance Reports could have been corrected before serving the water to the public. MRDLs or M/R violations were not included. This indicator only refers to Community Water Systems (CWS) and not to private domestic wells or small CWS. A discussion of the many limitations regarding the collection of water quality data and its use for Public Health purposes can be found here http://www.ehib.org/page.jsp?page_key=153. Only 1597 of 4000 CWS in California and only 475 of the 854 CWS with 1 or more MCL/TT violations (2008-2012) are currently included in the Water Systems Boundary shapefile. Some water system boundaries might change in the future as the shapefile is updated.

Data Processing

SAS program: HCI_DrinkingWater_426-4-15-14.sas, PART 2

Data Processing: Unsafe Drinking Water



<u>Data Acquisition and Preprocessing</u>

Number and Type of Violations in Public Water System from Annual Compliance Reports

The annual reports (2008-2011) on the type and number of public water system violations from CDPH to the U.S. Environmental Protection Agency were downloaded in pdf format from: http://www.cdph.ca.gov/certlic/drinkingwater/Pages/Publications.aspx. Data for 2012 was obtained from Mark Bartson, Supervising Sanitary Engineer (mark.bartson@cdph.ca.gov). The "Appendix D - Summary of Violations by Violation Category in Each County" of these reports was converted from pdf to an Excel spreadsheet using Adobe Acrobat. Using Microsoft Excel, the data was formatted into an orthogonal table (file name: MCLTTviolations20082012.xlsx) with the following columns:

county_name
reportyear
contaminant_name
contaminant_name_orig
MCL_level
system_id

system_name
Population_affected
MCL_violations
TT_violations
Total_violations
Category

The column “contaminant_name_orig” retained the information on the full name of the contaminant as it appeared in the compliance report. The “contaminant_name” contained standardized names. The following categories were created:

contaminant_name	category
Aluminum	MCL-OTHER
Antimony	MCL-OTHER
Arsenic	MCL-OTHER
Asbestos	MCL-OTHER
Bromate	MCL-OTHER
Cadmium	MCL-OTHER
Carbon Tetrachloride	MCL-OTHER
Chlorine	MCL-OTHER
Chlorite	MCL-OTHER
Dibromochloropropane	MCL-OTHER
EWSTR	TT
Fluoride Natural Source	MCL-OTHER
Gross alpha	MCL-OTHER
Haa5 Haloacetic Acids (Five)	MCL-OTHER
LCR	TT
Nitrate (as NO3)	MCL-OTHER
Nitrate + Nitrite (As N)	MCL-OTHER
Nitrate nitrogen (NO3-N)	MCL-OTHER
Nitrite (As N)	MCL-OTHER
Nitrite Nitrogen (No2-N)	MCL-OTHER
Perchlorate	MCL-OTHER
SWTR	TT
Tetrachloroethylene	MCL-OTHER
Thallium	MCL-OTHER
Total Coliform Rule Bacteriologic	MCL-TCRB
Total Dissolved Solids	MCL-OTHER
Total Organic Carbon	MCL-OTHER
Total Trihalomethanes (TTHM)	MCL-OTHER
Uranium	MCL-OTHER

Information on the Type of Public Water System

The Annual Compliance Reports did not contain information on the type of water system (Community or non-Community). This information was obtained in an Excel file from Eric Swing, Associate Sanitary Engineer, Northern California Drinking Water Field Operations Branch (file name: All CA PWS original E Swing_4-10-2014.xlsx). The file contained the following columns:

Water System No.
Water System Name
State Type
Fee Code
Service Connections
Population

This file was used as lookup file to assign PWS to the CWS category.

Water System Boundaries

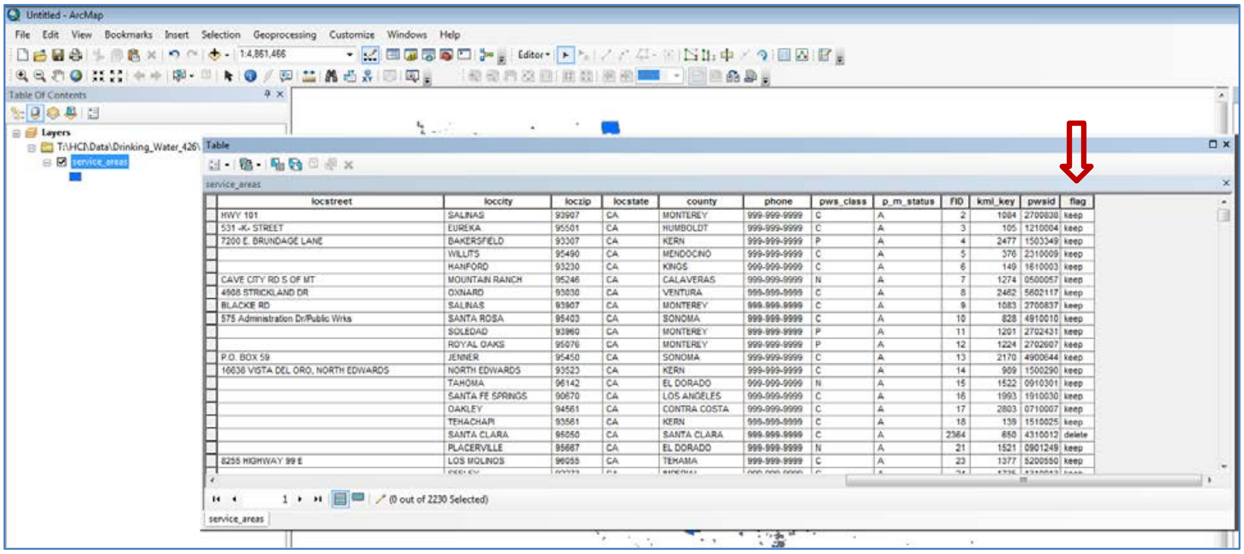
The public use dataset of the Water Systems Boundary Tool of the California Environmental Health Tracking Program was downloaded in shapefile (.shp) format from the CEHTP website: http://www.ehib.org/tools/water/archive/service_areas.zip (version 11-Nov-2013 23:04 5.8M downloaded in March 2014). The geographic coordinate system of the original shapefile (service_areas_ORIGINAL.shp) was GCS_WGS_1984. The geographic coordinate system was reprojected to GCS_North_American_1983 (service_areas.shp).

After exploration and consultation with Dan Meltzer (intern at EHIB) it was determined that the Water System Boundary shapefile contains both old and recent versions for some water systems (polygons). Before using the shapefile it was necessary to remove the duplicates as follows:

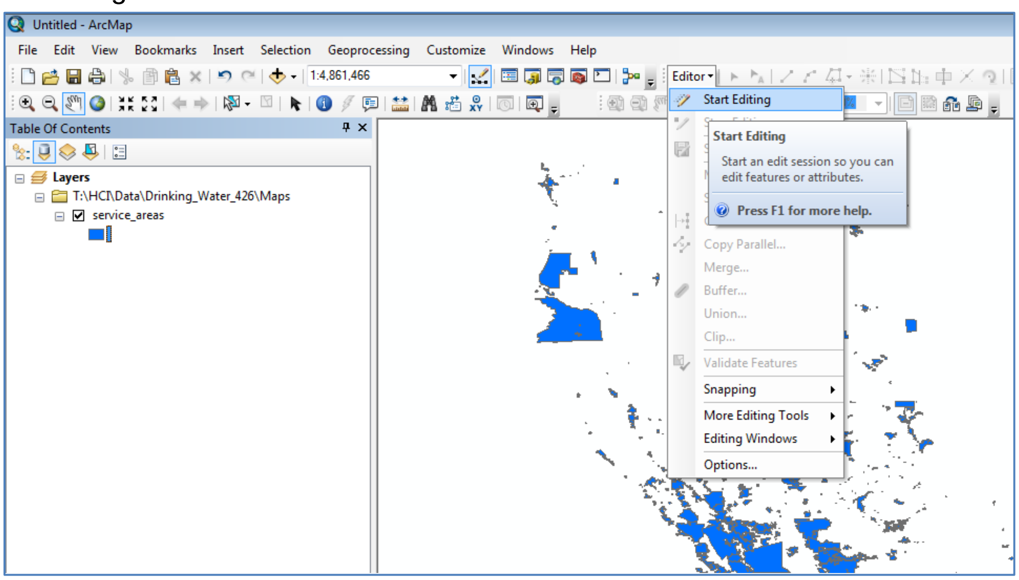
1. Export the Water System Boundary shapefile table as an Excel file (Service_Area_original.xlsx).

2. Import the table into SAS (ServiceAreaDuplicates.sas). Identify which polygons have duplicates and determine which polygons are the recent or older versions using the kml_key field of the table. Higher values of kml_key represent recent versions while the smaller values represent old versions and are flagged to be deleted later in ArcMap.

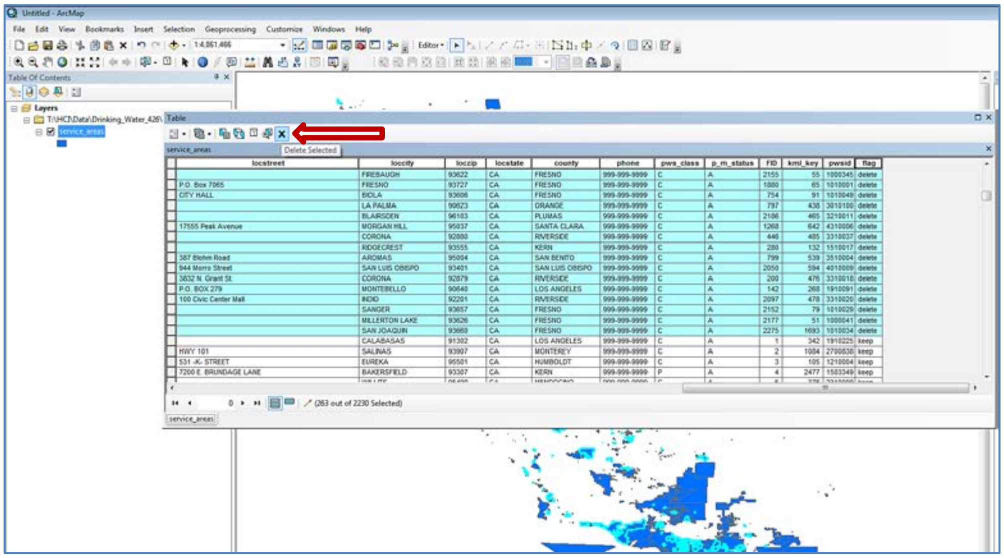
3. The file identifying the polygons to delete (duplicate_systems_flags.xlsx, flag field=keep or delete) was exported and joined with the shapefile.



4. The duplicates were deleted using the “editor” in Arc Map: click on “Editor” and then “Start Editing.”



5. Once in editor mode, the polygons were selected by attribute (flag=delete) and manually deleted. After deleting, the edits were saved under a new shapefile name (service_areas2.shp).



6. There were 2,518 water system polygons in the original shapefile (service_areas.shp). After removing duplicates there were 2,231 water systems (service_areas2.shp).

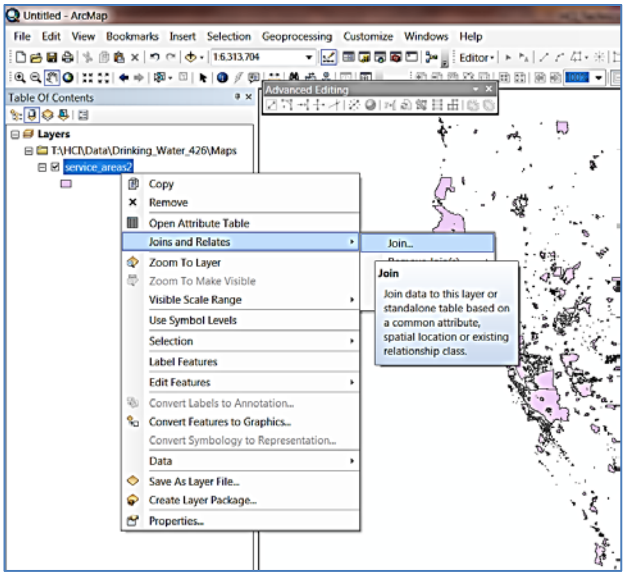
Creating Community Water Systems (CWS) boundary shapefiles: merging violations, type of water systems, and boundaries

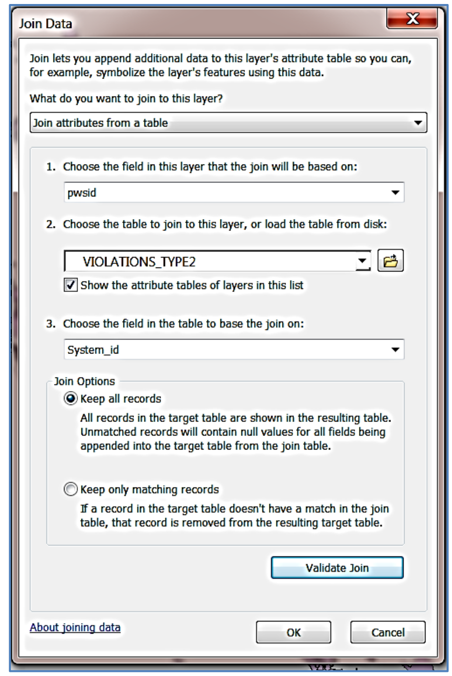
Data on violations (MCLTTviolations20082012.xlsx) and data on the total water systems (watersystemsCDPH.xlsx) and type of water system (All CA PWS.xlsx) was imported into SAS (HCI_DrinkingWater_426-6-16-14.sas, PART 1). The three files were merged to create a file with all the community water systems in California (4000) and an identification field for those with at least one violation in the years 2008-2012. Five categories of violations were created:

- Treatment Technique (tt)
- MCL Total Coliform Rule-Bacteriologic (tcrb)
- arsenic
- nitrates
- all other MCL (other)

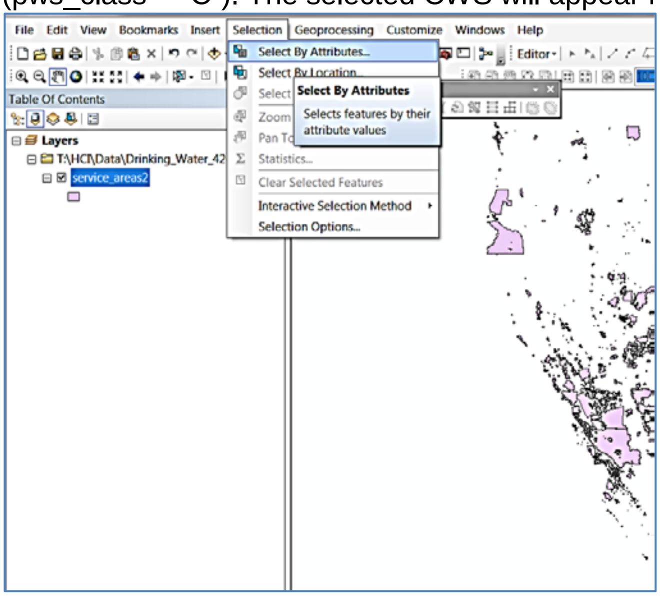
ArcMap was used to create individual shapefiles for community water systems, community water systems with violations, and community water systems with violations by type. The creation of these shapefiles was necessary to correctly allocate Census blocks (see below) into individual CWS.

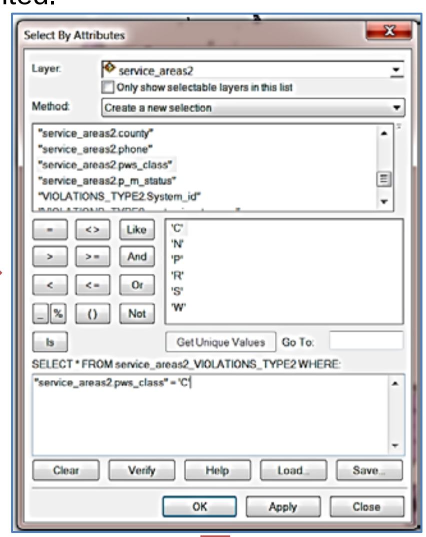
1. The Excel file containing information on CWS with violations (GisFileCWSViolations2.xls) was joined with the shapefile of public water systems with no duplicates (service_areas2.shp) using the water system id as the field to join both files.

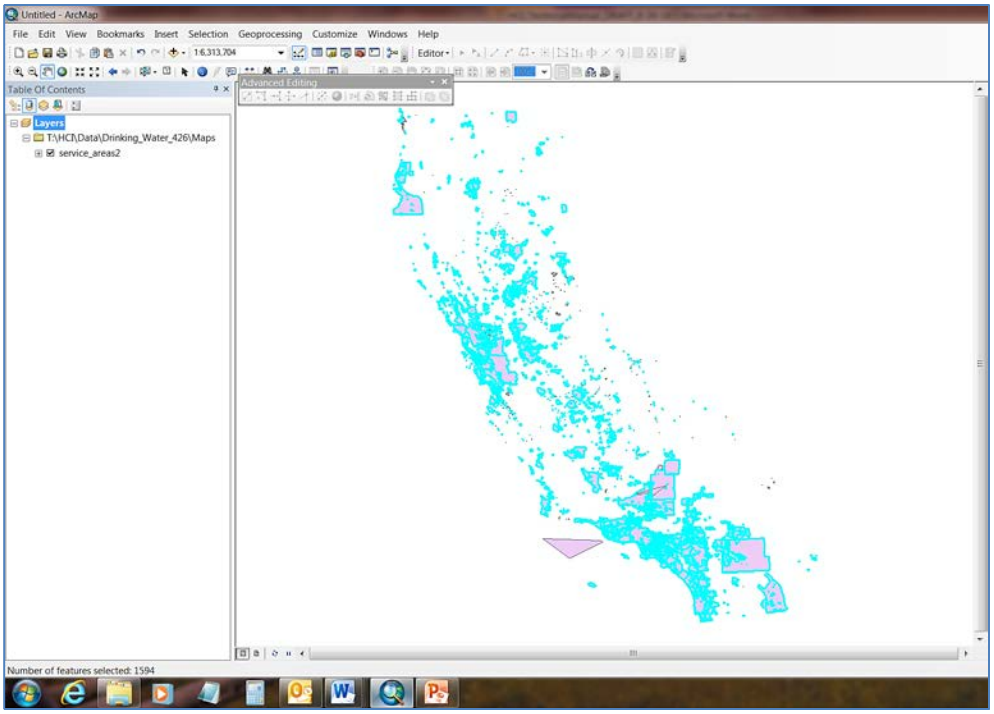




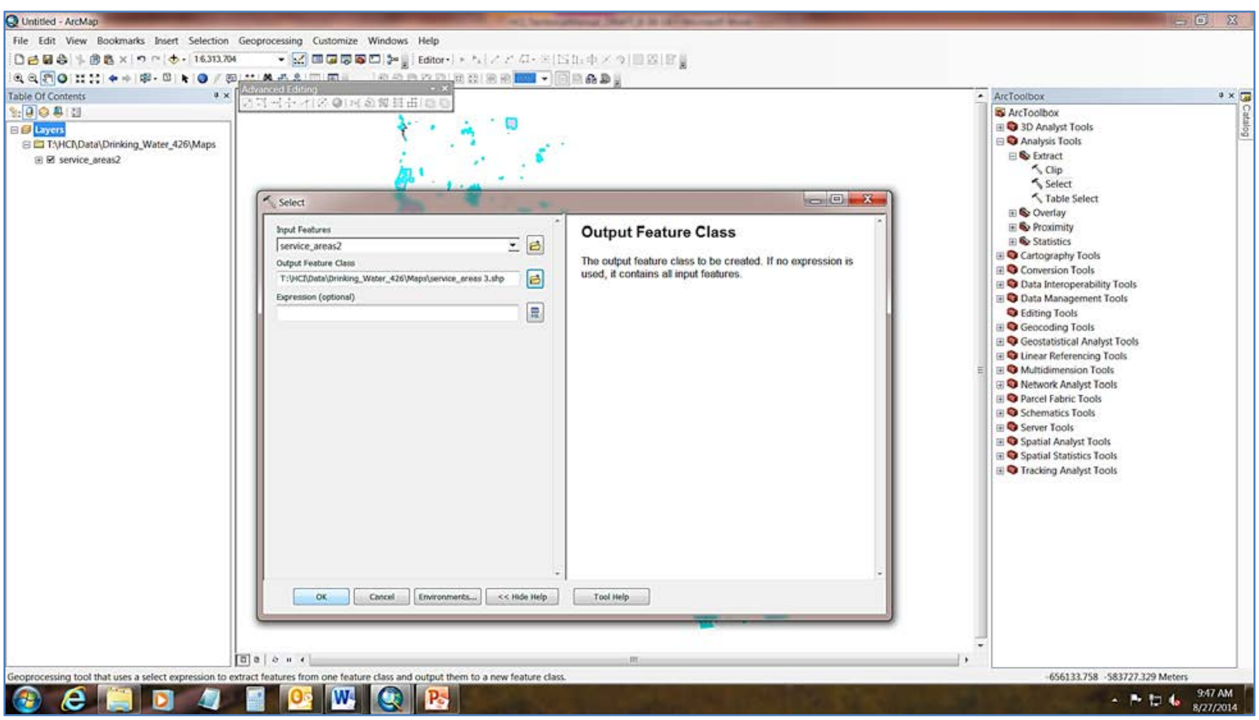
2. The "select by attributes" tool was used to select all water systems type "C" or Community (pws_class = "C"). The selected CWS will appear highlighted.



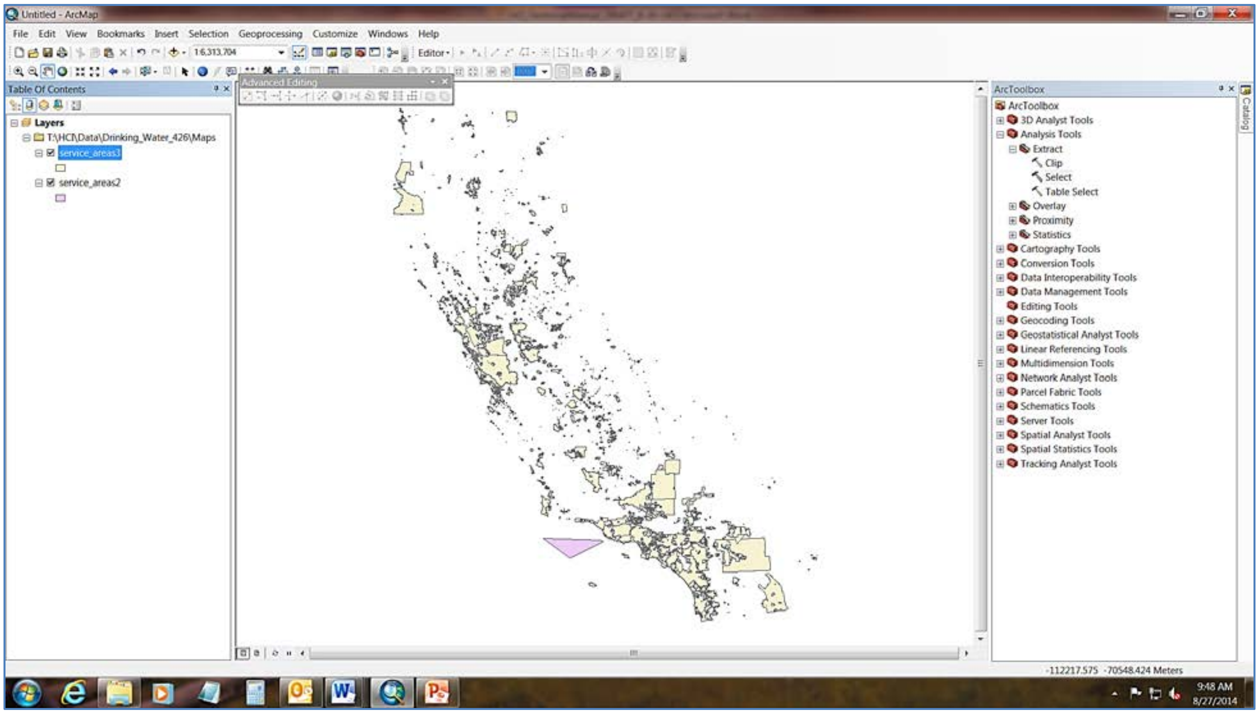




3. The “Select” tool (ArcToolbox > Analysis Tools > Extract > Select) was used to create a new shapefile containing only the CWS that were selected or highlighted in the previous step (service_areas3.shp).



4. The new shapefile is added to the project.

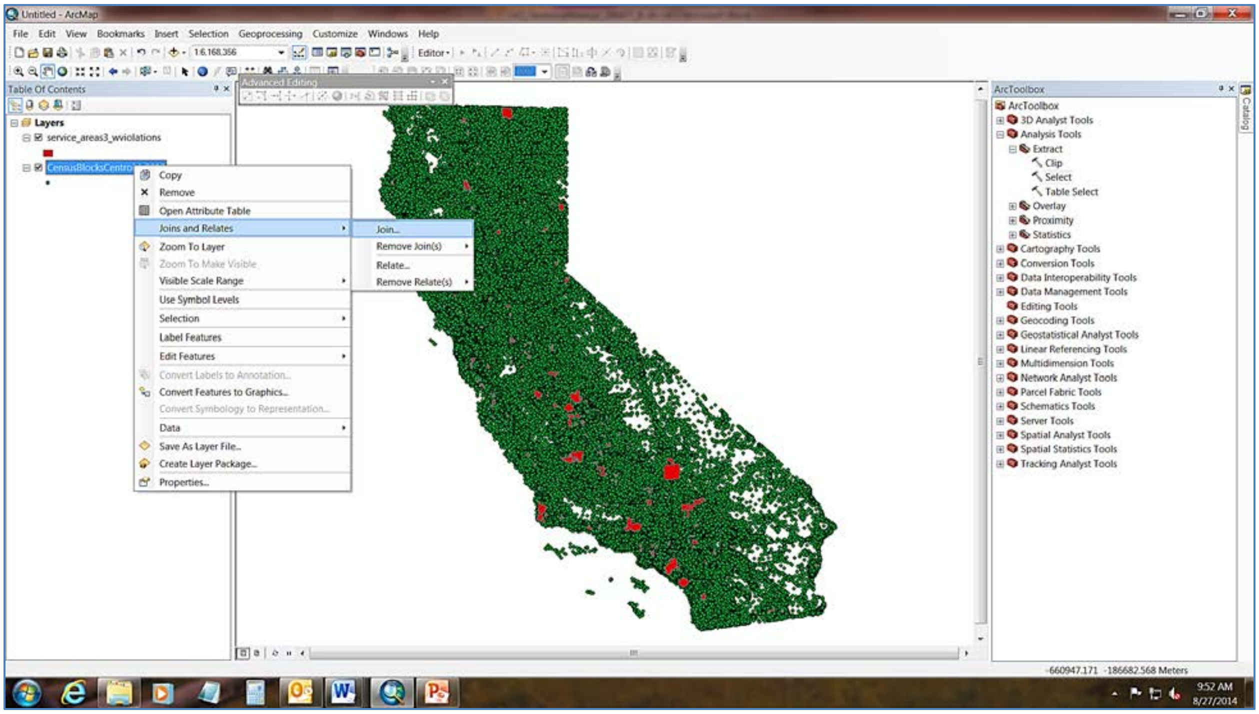


5. The same steps 1 to 4 were followed to create six individual shapefiles with boundaries for CWS by violation type:
   a. All violations (service_areas3_wviolations.shp)
   b. Treatment Technique (service_areas3_wviolationsTT.shp)
   c. MCL Total Coliform Rule-Bacteriologic (service_areas3_wviolationsMCLTCRB.shp)
   d. Arsenic (service_areas3_wviolationsMCLARSENIC.shp)
   e. Nitrates (service_areas3_wviolationsMCLNITRATES.shp)
   f. all other MCL (service_areas3_wviolationsMCLOTHER2.shp)

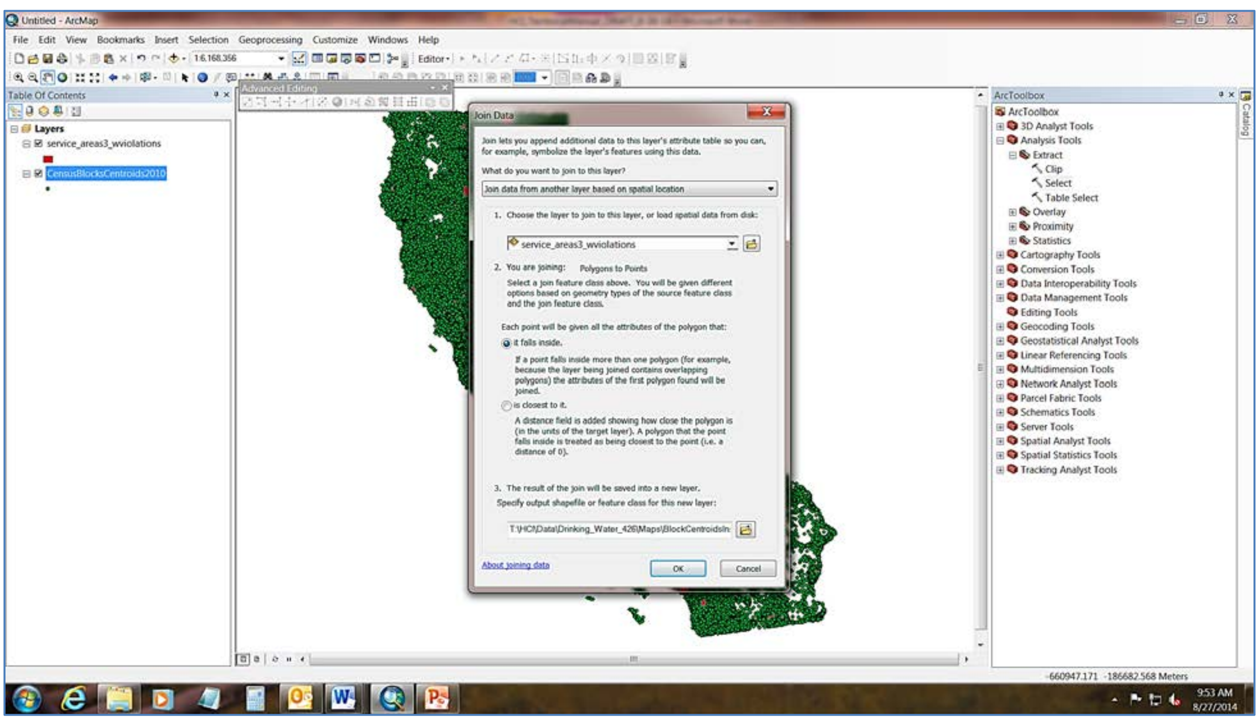
Identifying Census blocks centroids within (CWS) boundaries using spatial join

The “Spatial Join” tool in ArcMap was used to allocate Census blocks centroids within the polygons of CWS with violations. The assumption was that the population living in those blocks was served by those CWS.

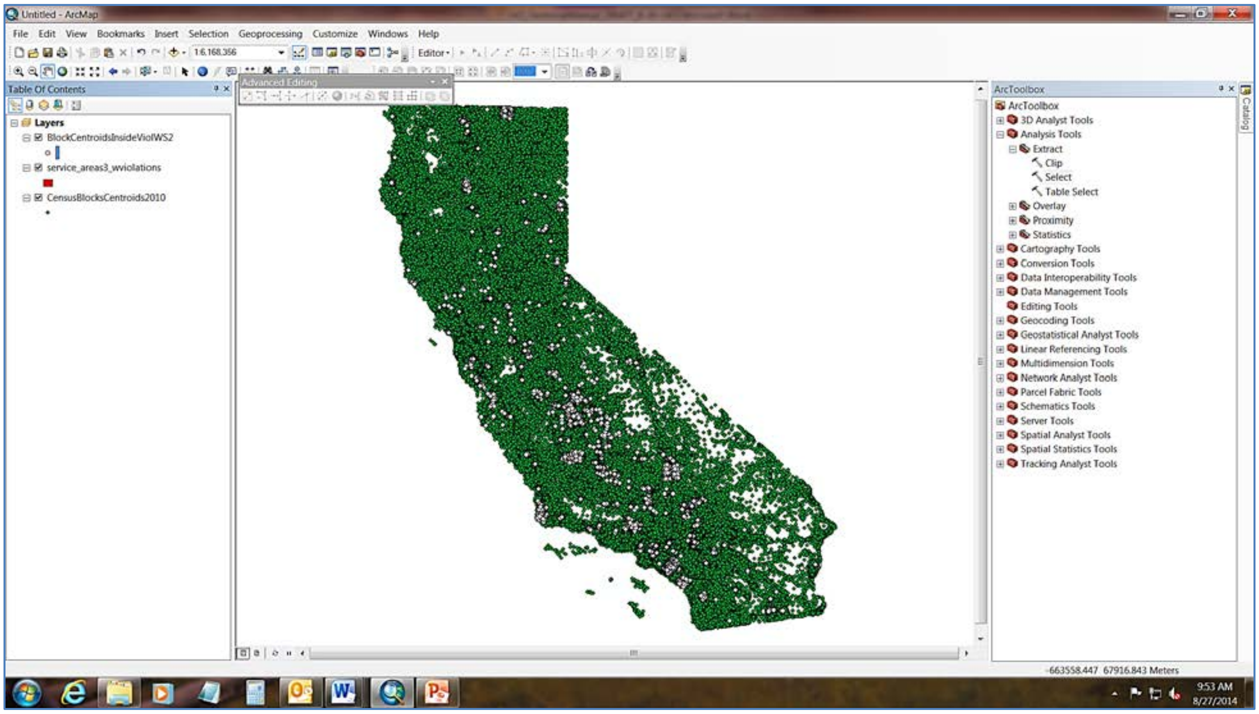
1. A shapefile of Census blocks centroids (2010 vintage) was created (see Toolbox section, Calculating Centroids in ArcMap and Creating a Centroid Shapefile chapter).
2. The centroids were spatially joined to the CWS with violations polygons (service_area3_wviolations.shp):
   a. Right click on the centroids shapefile > Joins and Relates > Join.



b. Select "Join data from another layer based on spatial location" in the drop down menu. Choose the CWS with violations layer as the "layer to join to this layer", and select the option of giving each point the attributes of the polygon in which it falls inside. Provide a name for the resulting layer and click OK.



c. The new layer of block centroids within CWS with violations will be added.



d. The attribute table of the new shapefile was exported in .txt format, and later saved in Excel format to be imported into SAS.

3. Five other spatial joins were conducted for CWS with violations by violation type:
   a. service_areas3_wviolationsTT.shp
   b. service_areas3_wviolationsMCLTCRB.shp
   c. service_areas3_wviolationsMCLARSENIC.shp
   d. service_areas3_wviolationsMCLNITRATES.shp
   e. service_areas3_wviolationsMCLOTHER2.shp

Population Data at the Census Block Level, 2010

Population at the Census Block level with race/Hispanic origin information was obtained from the Department of Finance, Demographic Research Unit, 2010 redistricting data from the U.S. Census Bureau (Public Law 94-171) for California. (http://www.dof.ca.gov/research/demographic/state_census_data_center/census_2010/#PL94, http://www2.census.gov/census_2010/01-Redistricting_File--PL_94-171/).

Access to Parks

Definition

Full Title: Percent of residents within ½ mile of a park, beach, open space, or coastline

Indicator ID: 469

Healthy Community Framework: Environmental Quality

Aspirational Goal: Green and open spaces, including agricultural lands

Description of significance and health connection:

As communities become increasingly more urban, parks and the protection of green and open spaces within cities increase in importance. Parks and natural areas buffer pollutants and contribute to the quality of life by providing communities with social and psychological benefits such as leisure, play, sports, and contact with nature. Parks are critical to human health by providing spaces for health and wellness activities. Of these, physical activity helps prevent and manage many chronic conditions, such as diabetes, obesity, arthritis, hypertension, cardiovascular disease, some cancers, cognitive function, and depressive disorders. Access to physical activity resources is particularly important for rural residents, who are at highest risk of poor health compared with their suburban and urban counterparts.

Summary of evidence:

An extensive body of research indicates that built environment factors correlate with better health. A recent systematic review of 204 articles showed that built environment factors, including levels of open space, were associated with increased levels of physical activity and walking. Further, an extensive body of research indicates that the presence of parks is correlated with physical activity. A recent systematic review of 20 studies examining the influence of the built environment and physical activity showed positive associations between health and environments with pleasant aesthetics, trails, safety/crime, parks, and walkable destinations. Another recent review of 50 studies reported that in general the presence of parks and recreation settings correlates with physical activity, specifically in the form of exercise or utilitarian functions, such as walking.

Key References:

- Bedimo-Rung AL, Mowen AJ, Cohen DA. The Significance of Parks to Physical Activity and Public Health: A Conceptual Model. *Am J Prev Med* 2005; 28 (2 Suppl 2):159-168.
- Chiesura A. The role of urban parks for the sustainable city. *Landscape Urban Plan* 2004; 68(1):129-138.
- Durand CP, Andalib M, Dunton GF, Wolch J, Pentz MA. A Systematic Review of Built Environment Factors Related to Physical Activity and Obesity Risk: Implications for Smart Growth Urban Planning. *Obes Rev* 2011; 12(501):e173-e182.
- Godbey GC, Caldwell LL, Floyd M, Payne LL. Contributions of leisure studies and recreation and park management research to the active living agenda. *Am J Prev Med* 2005; 28 (2 Suppl 2):150-158.

- Kaczynski AT, Henderson KA. Parks and recreation settings and active living: a review of associations with physical activity function and intensity. *J Phys Act Health* 2008; 5:619-632.

Detailed Definition:

- Numerator: number or residents within ½ mile of a park greater than 1 acre, or a beach, open space, or coastline
- Denominator: total number of residents
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: eight race/ethnicity groups (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white).
- Interpretation: healthier communities will have a higher percentage of residents within ½ of a park greater than 1 acre, or a beach, open space, or coastline

Data Description:

- Data source: California Protected Areas Database (CPAD version 1.8, 2012), maintained by GreenInfo Network, accessed September, 2012 from CALANDS website at http://www.calands.org/ . 2010 block-level population data by race and ethnicity from the U.S. Census Bureau (provided by California State Data Center at the California Department of Finance)
- Years available: 2010
- Updated: 10-year intervals
- Geographies available: Census tracts, cities/towns, counties, regions, and state

Methodology

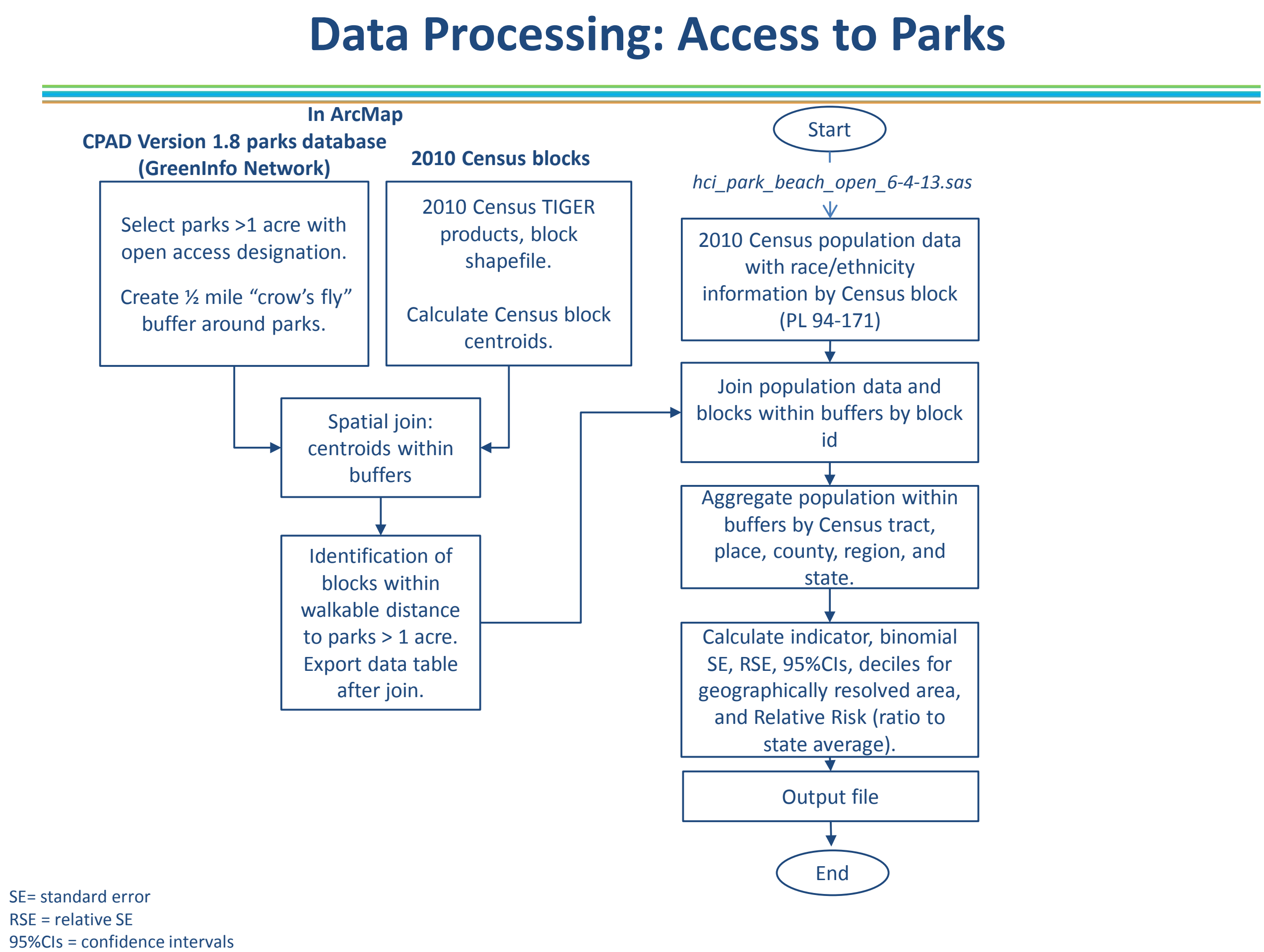
The California Protected Areas Database (updated 2012) was obtained as a shape file from the CALANDS website. The database includes open space lands including parks, as well as open space lands with other uses, including: recreation, forestry, historical/cultural, habitat conservation, water supply, scenic areas, flood control, agricultural/ranching, and general open space. Parks greater than 1 acre with 'Open Access' designation were selected for analysis. Half mile buffers were created around parks. Census blocks with centroids inside the parks buffer area were selected. 2010 block-level Census redistricting data (100% counts by race/ethnicity) were merged with blocks inside the parks buffer area. Block data were aggregated by census tract, cities/towns, county, region, and state. The percent of residents' access to parks were calculated for each geographic level and for race/ethnicity strata. Regions were based on counties of metropolitan transportation organizations (MPO) regions as reported in the 2010 California Regional Progress Report (Appendix C). Standard errors, relative standard errors, and 95% upper and lower confidence intervals were calculated.

Limitations:

The California Protected Areas Database does not include tribal lands, lands used for active military purposes, and properties protected through easements. The indicator takes into account the travel distance to park borders, but does not take into account points of entry. The indicator does not take into account the quality of park facilities, level of maintenance, specific amenities and services offered, or safety issues. While the indicator only measures "walkable" distance, transportation to parks through private or public transit was not considered. Census blocks are designated as inside or outside of park buffers based on block centroids, which can result in some misclassification of population within buffer areas. The indicator does not include “mini parks” or “pocket parks”, sometimes defined as less than 1 acre. The indicator only includes beach and coastline areas that are part CPAD, and known to be accessible to the public.

Data Processing

SAS program: hci_park_beach_open_6-4-13.sas



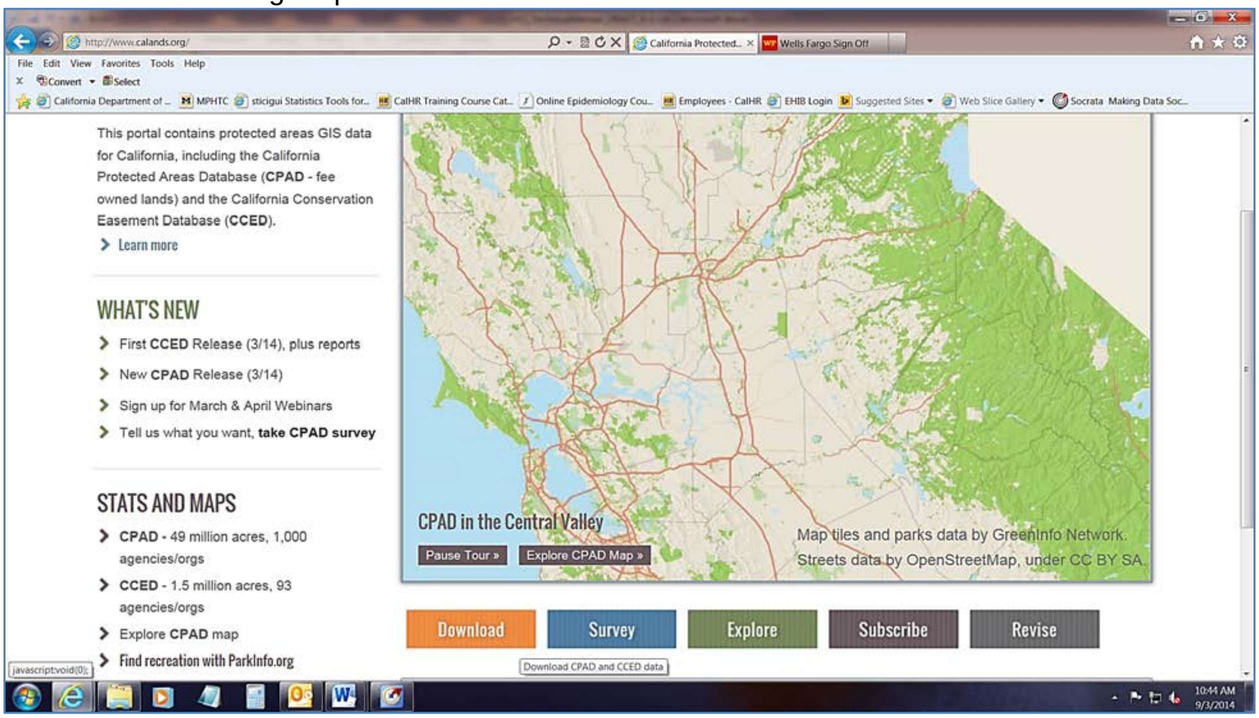
Data Acquisition and Preprocessing

Parks Data

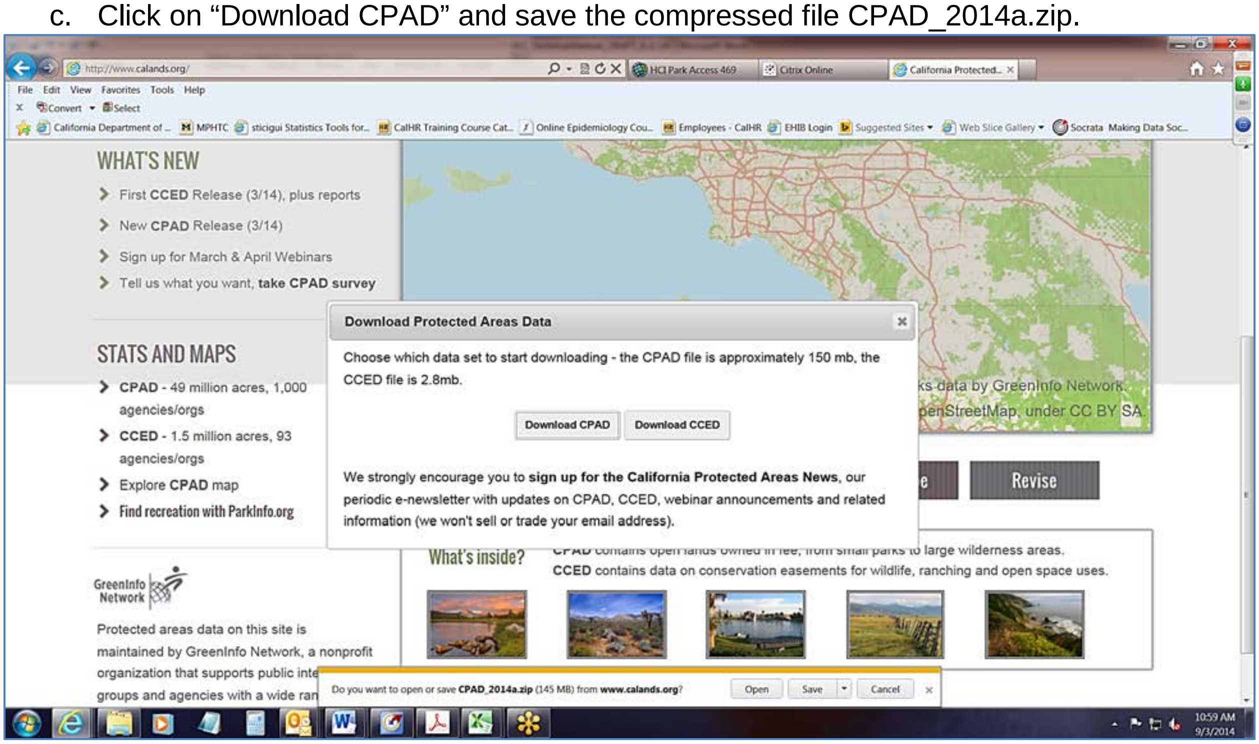
1. Data for parks and open spaces were acquired from the California Protected Areas Database (CPAD, Version 1.8, updated July 2nd, 2012), downloaded as an ESRI shapefile from www.CALANDS.org (accessed September, 2012). CPAD is prepared, maintained, and updated by GreenInfo Network (www.greeninfonetwork.org). CPAD includes open space lands that have been protected primarily for open space uses

through fee ownerships (it is not a database of all public lands). More detailed data description of CPAD is available in the Data Manual that accompanies the CPAD geographical data. The following downloading steps are for the most recent CPAD release of March 2014.

2. Data download steps for CPAD:
    a. Go to http://www.calands.org
    b. Click on orange square on left: "Download"



    c. Click on "Download CPAD" and save the compressed file CPAD_2014a.zip.



3. Parks were classified based on acreage:
    a. Open CPAD_2014a_Holdings.shp in ArcMap.
    b. Open attributes table and sort ascending by acreage (field = 'GIN_ACRES').
    c. Select all records with acreage >1.0 and export data into new shapefile.
    d. Use new shapefile and 'Select by Attributes' to select records with Access = 'Open access'.
    e. Export selection into new shapefile.

4. Buffers were created around parks with acreage > 1.0:
    a. Using new shapefile (step 3.e.), use 'Buffer' tool in ArcMap to create ½ mile buffer around California parks/open space.
    b. Use 'Dissolve' tool to dissolve parks buffers into new shapefile.

5. Create shapefile of 2010 TIGER blocks within buffer area
    a. Open TIGER 2010 block shapefile in ArcMap.
    b. Use 'Select by Location', and select blocks that have their centroid within dissolved parks buffer shapefile.

6. Export file from ArcMap in Excel format and import into SAS.

Population Data at the Census Block Level, 2010

Population at the Census Block level with race/Hispanic origin information was obtained from the Department of Finance, Demographic Research Unit, 2010 redistricting data from the U.S. Census Bureau (Public Law 94-171) for California. (http://www.dof.ca.gov/research/demographic/state_census_data_center/census_2010/#PL94, http://www2.census.gov/census_2010/01-Redistricting_File--PL_94-171/).

Income Inequality

Definition

Full title: Distribution of household income relative to the number of households, expressed on a 0 to 1 scale (Gini Index)

Indicator ID: 556

Healthy Community Framework: Communities have health and social equity

Aspirational Goal: Social equity

Description of Significance and Health Connection:

Income is linked to acquiring resources for healthy living. Both household income and the distribution of income across a society independently contribute to the overall health status of a community. On average Western industrialized nations with large disparities in income distribution tend to have poorer health status than similarly advanced nations with a more equitable distribution of income. Approximately 119,200 (5%) of the 2.4 million U.S. deaths in 2000 are attributable to income inequality. The pathways by which income inequality act to increase adverse health outcomes are not known with certainty, but policies that provide for a strong safety net of health and social services have been identified as potential buffers.

Summary of evidence:

Many cross sectional, ecological studies have compared western industrialized countries, including the United States, along a gradient of a health outcome and the corresponding gradient of income inequality using the Gini index, a measure of inequality of income and wealth. A Gini index of 1 represents maximum inequality or unequal distribution of income; a Gini index of 0 represents maximum equality when each household has the same income. Studies using this index often show a linear relationship between increasing income inequality and poorer health outcomes such as life expectancy, infant mortality, obesity, mental illness, homicide, etc. Several, large longitudinal studies that followed healthy participants at baseline were combined to estimate the number of U.S. deaths in 2000 attributable to income inequality.

Key References:

- Commission on the Social Determinants of Health. Closing the Gap in a Generation: Health Equity through Action on the Social Determinants of Health. Final Report of the Commission on Social Determinants of Health. Geneva: World Health Organization; 2008. http://www.who.int/social_determinants/thecommission/finalreport/en/index.html. Accessed on February 4, 2013.
- Wilkinson R, Pickett K. The Spirit Level: Why Equal Societies Almost Always Do Better. London, Pilgrim Press; 2009.
- Galea S, Tracy M, Hoggatt KJ, DiMaggio C, Karpati A. Estimated deaths attributable to social factors in the United States. *Am J Public Health* 2011;101(8):1456-1465.

Detailed definition:

- Cumulative percentage of household income relative to the cumulative percentage of the number of households expressed on a 0 to 1 scale called the Gini Index (0 = equal incomes for all households, 1 = all income is concentrated in 1 household)
- Stratification: race/ethnicity not available
- Interpretation: healthier communities will have a tendency to income equality. It is recommended to compare communities within the same income decile; this will help interpret if households are for instance, “equally rich” or “equally poor.”

Data Description:

- Data source: Gini Index data from the American Community Survey: Table B19083. http://factfinder2.census.gov
- Years available: 2005-2007 (3-year), 2008-2010 (3-year), and 2006-2010 (5-year)
- Updated: 3 and 5-year intervals
- Geographies available: Cities/towns, counties, regions (derived), state

Methodology:

Data for the Gini index, including margin of error, were downloaded from the American Community Survey (ACS). Three-year data were collected for cities, towns, and counties over 60,000 in population for the intervals 2005-2007 and 2008-2010. Five-year ACS data 2006-2010 were also collected for cities and towns with populations generally greater than 20,000. Standard errors, relative standard errors, and 95% upper and lower confidence intervals were calculated from margins of error. Regional estimates were based on county groupings associated with California metropolitan planning organizations as reported in the *2010 California Regional Progress Report* (http://www.dot.ca.gov/hq/tpp/offices/orip/Collaborative%20Planning/Files/CARegionalProgress_2-1-2011.pdf). Information on the median household income for places and counties (obtained from table DP03 of the ACS) and their grouping in deciles are included for those users interested in comparing the Gini index within the same income decile.

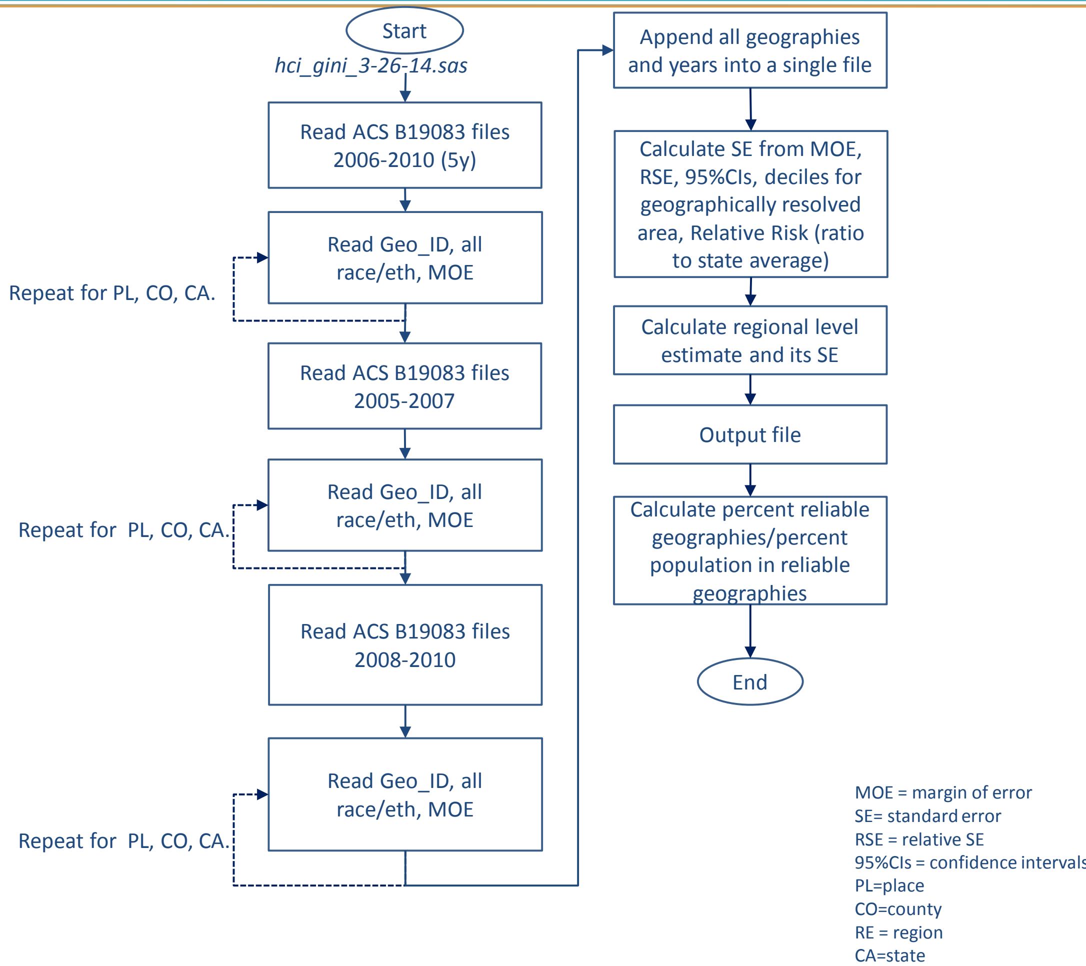
Limitations:

Calculations of the Gini coefficient from the American Community Survey at geographies below census place were not statistically reliable. There was no race/ethnicity information available for the Gini coefficient.

Data Processing

SAS file: hci_gini_3-26-14.sas

Data Processing: Income Inequality



Data Acquisition and Preprocessing

The quick table QT-P11, profile table DP03, and detailed table B19083 from the American Community Survey (2005-2007, 2008-2010, and 2006-2010) were downloaded using the American FactFinder (See Toolbox section, Downloading Census Data from American Fact Finder chapter). A total of 24 Census tables representing four datasets (time periods) and three geographies were downloaded (See Table 12). The variables in the tables that contained the information of interest were identified prior to data processing (See Table 13) using the metadata tables downloaded from the Census.

Table 12. Census Tables Downloaded from American Fact Finder for the Income Inequality Indicator by Data Set, Geographies and Race and Ethnic Groups

Dataset (Time period)	Geographies	Race and Ethnic groups	Census tables
2010 ACS 5-year Selected Population Tables (2006-2010)	Place - 160	Total population	B19083, QT-P11, DP03
	County - 050	Total population	B19083, QT-P11, DP03
	State - 040	Total population	B19083, QT-P11
2010 ACS 3-year Selected Population Tables (2008-2010)	Place - 160	Total population	B19083, QT-P11, DP03
	County - 050	Total population	B19083, QT-P11, DP03
	State - 040	Total population	B19083, QT-P11
2007 ACS 3-year Selected Population Tables (2005-2007)	Place - 160	Total population	B19083, QT-P11, DP03
	County - 050	Total population	B19083, QT-P11, DP03
	State - 040	Total population	B19083, QT-P11

Table 13. Variables Used to Construct the Income Inequality Indicator by Census Table

Census Table	Indicator	Margin of Error	Standard Error
B19083	HD01_VD01, Gini Index	HD02_VD01	$\frac{HD02\_VD01}{1.645}$
Supporting variables			
QT-P11	HD01_S01, Total households		
DP03	HC01_VC85, Median household income		

Voter Registration and Participation

Definition

Full Title: Percent of adults (18 years or older) who are registered voters; percent of adults who voted in general elections

Healthy Community Framework: Social relationships that are supportive and respectful

Aspirational Goal: Robust social and civic engagement

Description of significance and health connection:

Political participation can be associated with the health of a community through two possible mechanisms: through the implementation of social policies or as an indirect measure of social capital. Disparities in political participation across socioeconomic groups can influence political outcomes and the resulting policies could have an impact on the opportunities available to the poor to live a healthy life. Lower representation of poorer voters could result in reductions of social programs aimed toward supporting disadvantaged groups. Although there is no direct evidentiary connection between voter registration or participation and health, there is evidence that populations with higher levels of political participation also have greater social capital. Social capital refers to the existence of trust and mutual aid among the members of a society and participation of its members in civic associations. There is evidence of a positive association between social capital and lower mortality rates and higher self-assessed health ratings.

Political participation is directly related to the socioeconomic and demographic characteristics of individuals, with lower levels of participation observed in people with low income and low education levels. The most common reasons for non-participation in the 2008 elections included disapproval of candidate choices, busyness, illness, transportation, and registration/administrative problems. African Americans and Latinos were more likely than whites to cite transportation and registration/ administrative problems (lack of I.D., difficulty finding the correct polling place, not receiving absentee ballots, waiting in long lines) while white voters were most likely to cite disapproval of candidate choices. The population eligible to register to vote in California includes residents who are U.S. citizens and 18 years of age or older on Election Day. People in prison, on parole, serving a state prison sentence in county jail or a sentence for a felony, on post release community supervision, or that have been found mentally incompetent by a court are not eligible to register to vote.

Summary of evidence:

The evidence supporting a relationship between voter participation and health is largely indirect. The relationship appears to be mediated by social capital, for which there are multiple longitudinal community studies, including several from California, that show a positive association with health outcomes.

Key References:

- Baum FE, Bush RA, Modra CC, Murray CJ, Cox EM, Alexander KM, Potter RC. Epidemiology of participation: an Australian community study. *J Epidemiol Community Health.* 2000; 54: 414-423.
- Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda county residents. *Am J Epidemiol*. 1979; 109(2): 186-204.
- Braveman PA, Egerter S, Woolf SH, Marks JS. When do we know enough to recommend action on the social determinants of health? *Am J Prev Med.* 2011 (40)1S1: S58-S66 .
- Gilbert KL, Quinn SC, Goodman RM, Butler J, Wallace J. A meta-analysis of social capital and health: a case for needed research. *J Health Psychol*. 2013; 18(11): 1385-1399.
- Hero RE, Tolbert CJ. A racial/ethnic diversity interpretation of politics and policy in the states of the U.S. *Am J Political Science.* 1996; 40(3): 851-871.
- Kawachi I. Social capital and community effects on population and individual health. *Ann N Y Acad Sci.* 1999; 896: 120-130.
- Alvarez RM, Ansolabehere S, Berinsky A, Lenz G, Stewart III C, Hall T. 2008 Survey of the Performance of American Elections. http://vote.caltech.edu/sites/default/files/Final%20report20090218.pdf, Accessed June 2nd 2014
- California Secretary of State, Voter Registration, http://www.sos.ca.gov/elections/elections_vr.htm, Accessed May 5th 2014

Detailed Definitions:

- Percent of adults (18 years or older) who are registered voters among those eligible.
  - Numerator: Adults eligible to vote
  - Denominator: Eligible adults registered to vote
  - $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Percent of adults who voted in general elections among those who registered.
  - Numerator: Eligible adults registered to vote
  - Denominator: Eligible adults registered to vote that voted in the election
  - $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: two types of participation (registered of those eligible, registered/eligible; voted of those registered, voted/registered).
- Interpretation: healthier communities will have higher percentages of eligible people registered to vote and higher percentages of registered people voting in elections.

Data Description:

- Data sources: Statewide Database, University of California Berkeley Law, Center for Research, http://statewidedatabase.org. California Secretary of State, Elections Division, Reports of Registration, http://www.sos.ca.gov/elections. California Department of Finance, Demographic Unit (http://www.dof.ca.gov/research/demographic/data/race-ethnic/2000-2010/index.php). U.S. Census Bureau, Decennial Census 2010 (http://factfinder2.census.gov).
- Years available: 2002, 2004, 2006, 2008, 2010
- Updated: every two years

- Geographies available:
    - Percent of adults (18 years or older) who are registered voters among those eligible: counties, regions (derived), and state
    - Percent of adults who voted in general elections among those who registered: Census tracts, cities/towns, counties, regions (derived), and state.

Methodology:

Estimates of the number of people who are eligible to vote were obtained from the Secretary of State's Reports of Registration (15 days prior to a general election) for counties and the state. Eligible population is obtained by subtracting from the total population counts published by the Department of Finance, the population that is 17 years or below, non-citizens, felons in prison, and supervised felon parolees. Complete enumeration data at the Census block level on the number of people 18 years and over who registered to vote and who voted in the general elections was obtained from the Statewide Database. Data was aggregated into Census tracts, cities/towns, counties, regions, and the state. Regional estimates of population eligible to vote were also obtained. Decile rankings of places and relative risk in relation to state average were calculated. Regions were based on counties of metropolitan transportation organizations (MPO) regions as reported in the 2010 California Regional Progress Report (Appendix C). Standard errors, relative standard errors, and 95% upper and lower confidence intervals were calculated. Information on the population 18 years and over or voting age population (VAP) for the state and counties was obtained from the Department of Finance for all years available and is included for those interested. Estimates of the VAP for cities/towns and Census tracts were obtained from the Census 2010.

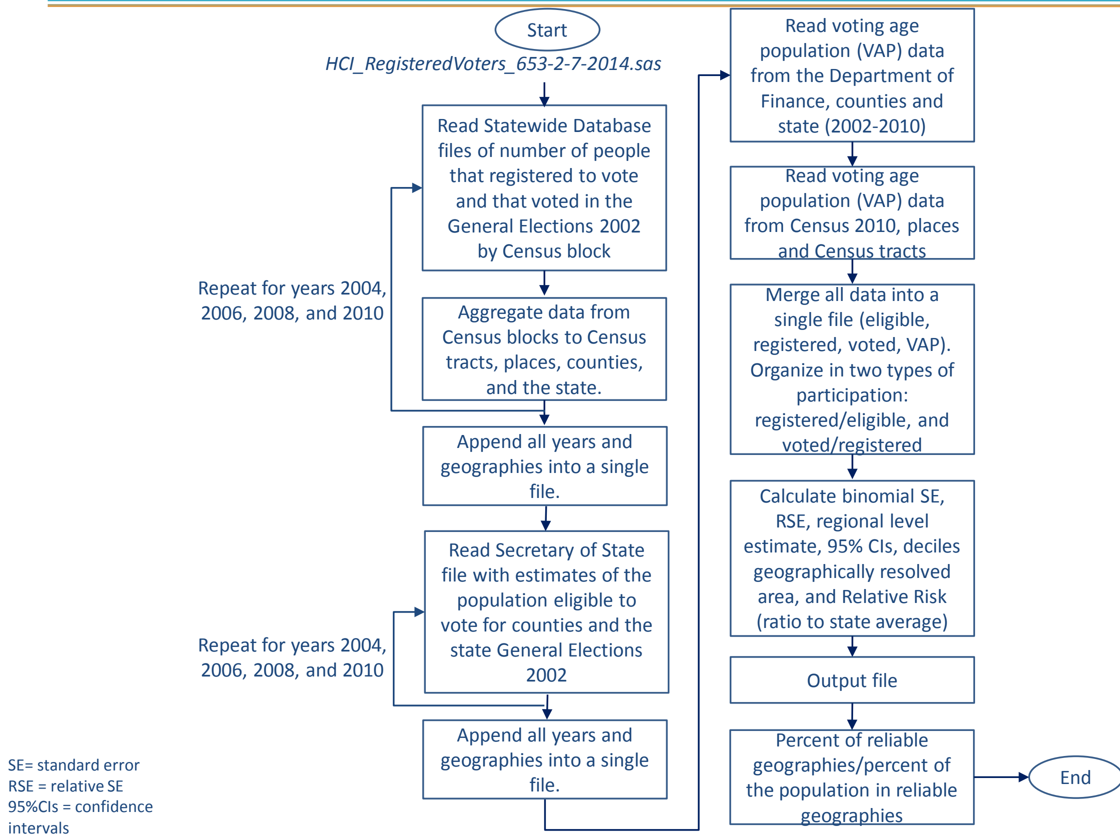
Limitations:

No race/ethnicity information is available for this indicator. Eligible population data is only available at the county and state levels. Registration and voting are only indirect measures of social capital, which has been found more directly associated with health status.

<u>Data Processing</u>

SAS program: HCI_RegisteredVoters_653-2-7-2014.sas

Data processing: Voter Registration and Participation



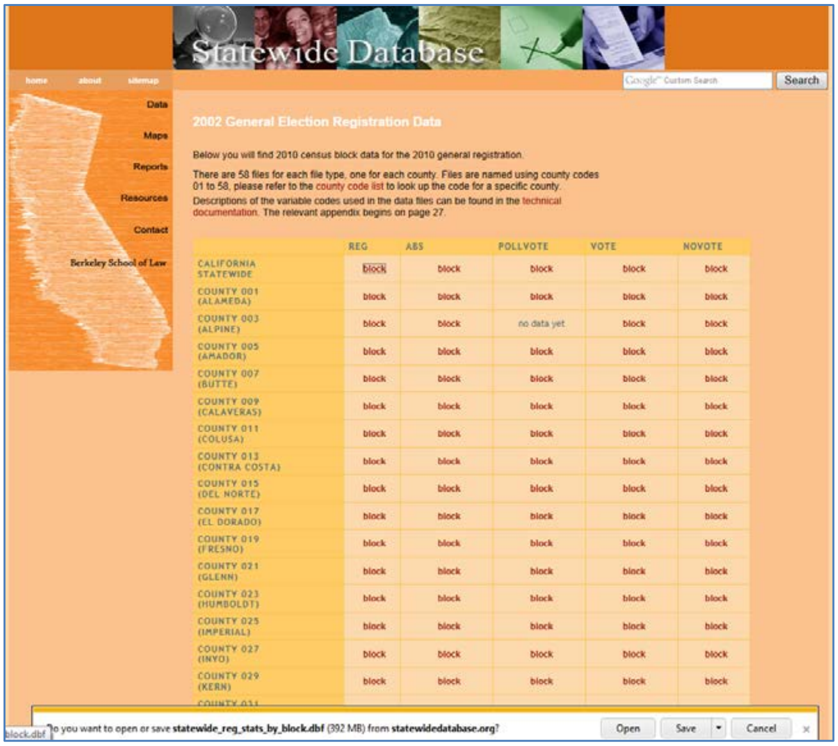
Data Acquisition and Preprocessing

Number of people who registered to vote and number of people that voted by Census block, California

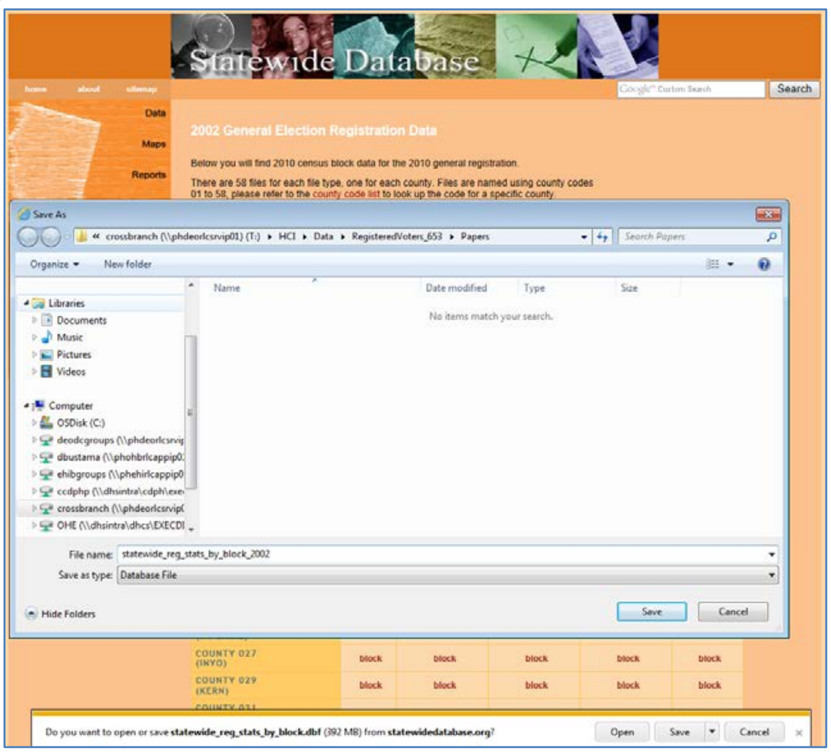
Data is located in the Statewide Database from U.C. Berkeley Law, Center for Research. http://statewidedatabase.org/d10/index_registration.html. Download all General Election Data 2002, 2004, 2006, 2008, 2010, REG and VOTE files at the census block level. Data are in .dbf format. The REG file contains "registration data for all registered voters", and the VOTE file contains "registration data for all voters that voted" (see http://statewidedatabase.org/info/metadata/precinct_data.html). The variables of interest in the REG and VOTE files are "totreg" (total registration) and "geoid" (Census block identification number) (see http://statewidedatabase.org/info/metadata/SOR_codebook.html)

The following example shows how to download the files for the 2002 General Election.

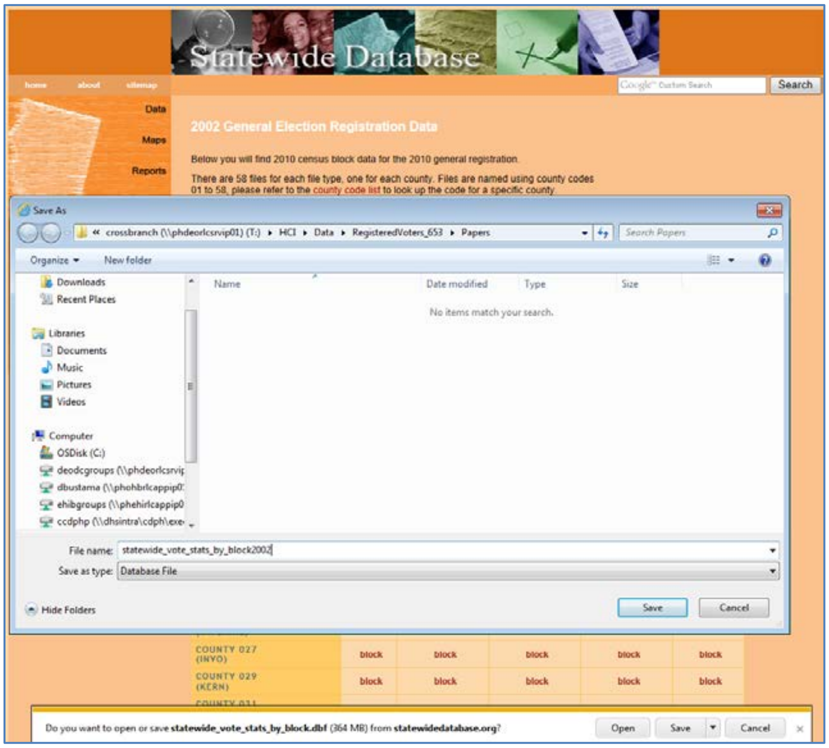
1. Go to the 2002 General Election Data page (http://statewidedatabase.org/d10/g02_registration.html)
2. Click on the word "block" under the REG column, CALIFORNIA STATEWIDE row. A message to either open or download the .dbf file will appear.



3. Save the data file adding the year at the end of the name provided (2002), to avoid overwriting when the next file is downloaded.



4. Repeat the previous two steps with the voting data by clicking on the word “block” under the VOTE column, CALIFORNIA STATEWIDE row.



5. Repeat for General Election Data for years 2004, 2006, 2008, and 2010.

The five registration .dbf files and the five voting .dbf files were imported into SAS and converted into sas data sets. Only the “totreg” (total registration) and “geoid” (Census block identification number) were retained for each year, and a reportyear variable was added.

Number of people who are eligible to vote in California

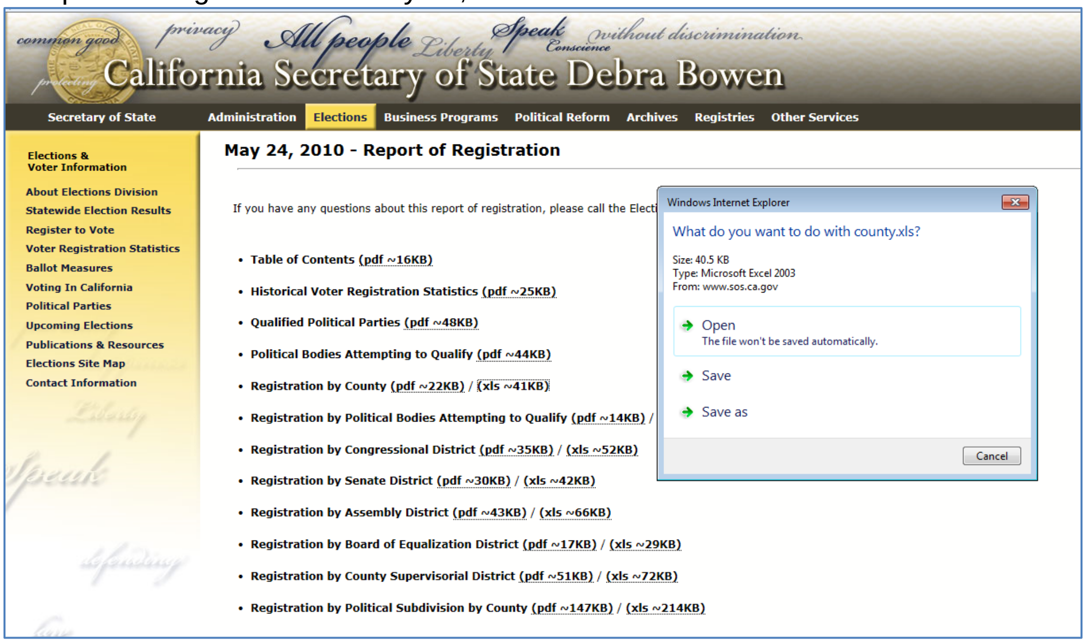
The California Secretary of State is required by law to produce a Report of Registration detailing voter registration levels throughout California at specific times prior to each statewide election and once during odd numbered years. The Secretary of State Elections Division website has historic voter registration information by county and for the state (http://www.sos.ca.gov/elections/elections_u.htm). (Contact person at the Secretary of State, Elections Division, Rhonda Pascual Rhonda.Pascual@sos.ca.gov)

To obtain an estimate of the number of people who are eligible to vote in California and its counties the 15 Day Report of Registration, Registration by County, Excel files, were downloaded for the following dates:

- November 5, 2002, General Election: http://www.sos.ca.gov/elections/ror/ror-pages/15day-gen-02/county.xls
- November 2, 2004, General Election: http://www.sos.ca.gov/elections/ror/ror-pages/15day-presgen-04/county.xls
- November 7, 2006, General Election: http://www.sos.ca.gov/elections/ror/ror-pages/15day-gen-06/county.xls

- November 4, 2008, General Election: http://www.sos.ca.gov/elections/ror/ror-pages/15day-presgen-08/county.xls
- November 2, 2010, General Election: http://www.sos.ca.gov/elections/ror/ror-pages/15day-prim-10/county.xls

The following screenshot shows the California Secretary of State website and the download page for reports of registration for May 24, 2010:



The Excel files were reformatted to make them suitable to be read by SAS (Figure 5). Empty rows were deleted and only population counts were retained (percentages were eliminated).

Original layout:

County	Eligible	Registered	Democratic	Republican	American Independent	Green	Libertarian	Natural Law	Reform	Miscellaneous	Declined to State
Alameda	955,139	662,339	369,434	130,083	10,533	15,461	3,381	2,102	1,395	8,514	121,436
Percent		69.34%	55.78%	19.64%	1.59%	2.33%	0.51%	0.32%	0.21%	1.29%	18.33%
Alpine	915	828	284	311	33	7	5	1	5	4	178
Percent		90.49%	34.30%	37.56%	3.99%	0.85%	0.60%	0.12%	0.60%	0.48%	21.50%
Amador	24,968	19,129	7,368	8,668	496	153	127	14	88	236	1,979
Percent		76.61%	38.52%	45.31%	2.59%	0.80%	0.66%	0.07%	0.46%	1.23%	10.35%
Butte	149,948	113,988	40,394	48,144	3,128	2,697	831	99	253	1,403	17,039
Percent		76.02%	35.44%	42.24%	2.74%	2.37%	0.73%	0.09%	0.22%	1.23%	14.95%
Calaveras	32,167	24,971	9,052	11,253	638	301	367	29	103	118	3,110
Percent		77.63%	36.25%	45.06%	2.55%	1.21%	1.47%	0.12%	0.41%	0.47%	12.45%
Colusa	10,578	7,512	2,959	3,435	191	22	35	2	16	24	828
Percent		71.02%	39.39%	45.73%	2.54%	0.29%	0.47%	0.03%	0.21%	0.32%	11.02%
Contra Costa	654,002	484,640	237,561	155,308	8,530	4,718	2,378	1,001	1,569	4,610	68,965
Percent		74.10%	49.02%	32.05%	1.76%	0.97%	0.49%	0.21%	0.32%	0.95%	14.23%

Reformatted layout:

Geoname	Eligible	Registered	Democratic	Republican	American	Green	Libertarian	NaturalLaw	Reform	Other	Declined
Alameda	955139	662339	369434	130083	10533	15461	3381	2102	1395	8514	121436
Alpine	915	828	284	311	33	7	5	1	5	4	178
Amador	24968	19129	7368	8668	496	153	127	14	88	236	1979
Butte	149948	113988	40394	48144	3128	2697	831	99	253	1403	17039
Calaveras	32167	24971	9052	11253	638	301	367	29	103	118	3110
Colusa	10578	7512	2959	3435	191	22	35	2	16	24	828
Contra Costa	654002	484640	237561	155308	8530	4718	2378	1001	1569	4610	68965
Del Norte	16740	12067	4687	4699	384	122	82	4	61	64	1964
El Dorado	116640	89640	29551	42684	2190	966	642	76	318	617	12596
Fresno	475513	344359	148208	151151	6010	1660	1336	307	727	819	34141
Glenn	15959	11346	4083	5378	295	39	38	11	17	68	1417
Humboldt	95608	76514	33246	22677	1796	5396	736	73	224	448	11918
Imperial	75335	51271	27936	14259	967	125	182	60	76	461	7205
Inyo	13078	10387	3465	5000	301	102	61	5	13	87	1353
Kern	399124	259786	98561	124323	6440	681	1329	221	749	612	26870
Kings	65706	45872	19016	20851	942	87	130	46	84	494	4222
Lake	44343	31593	14487	10611	937	500	236	39	152	128	4503
Lassen	16112	14584	4858	6363	614	67	102	13	62	65	2440

Figure 5. Comparison of original table layout and reformatted table layout for report of registration data.

Voting age population

Data on the voting age population is included in the indicator file for reference but it is not used in the calculation of an indicator. Data on the population 18 years and older for the counties and the state (July, 2000-2010) was obtained from the Department of Finance, Demographic unit (http://www.dof.ca.gov/research/demographic/data/race-ethnic/2000-2010/index.php) in Excel file format (http://www.dof.ca.gov/research/demographic/data/race-ethnic/2000-2010/documents/Intercensal_2000-2010_Total_Age-Race.xls). Data is presented at 1 year age increments from age 0 to age 100.

Data on population by age and by place (or Census tract) was downloaded from the American Fact Finder website (see Toolbox section, Downloading Census Data from American FactFinder Chapter). Table DP1 (2010 Summary File SF1 100% Data) was downloaded for both place and Census tracts.

Violent Crime Rate

Definition

Full Title: Number of Violent Crimes per 1,000 Population

Indicator ID: 752

Healthy Community Framework: Social relationships that are supportive and respectful

Aspirational Goal: Safe communities, free of crime and violence

Description of significance and health connection:

Safe neighborhoods that are free of crime and violence are an integral component of healthy neighborhoods. Injuries – both physical and mental – from violence and crime contribute greatly to the overall burden of disease and death. Violent crimes disproportionately affect communities of color and young adults. Post-traumatic stress in victims, families, and community members add significantly to mental health problems and are risk factors for chronic illnesses like cardiovascular disease. In addition to direct physical and mental impacts, fear of crime and violence inhibit the use of community assets and social interaction that promote health. These include stifling opportunities for physical exercise at nearby parks and playgrounds, or walking or bicycling as a way to commute to local destinations for basic needs. Furthermore, violence contributes to negative perceptions of neighborhoods and impacts real estate, housing, and economic development.

Summary of evidence:

Ten percent of all deaths in young California adults aged 15-44 years are related to assault and homicide. In 2010, California law enforcement agencies reported 1,809 murders, 8,331 rapes, and over 95,000 aggravated assaults. African Americans in California are 11 times more likely to die of assault and homicide than Whites.

Key References:

- Fowler PJ et al. Community violence: A meta-analysis on the effect of exposure and mental health outcomes of children and adolescents. *Dev Psychopathol* 2009; 21(1):227-259.
- Takagi D, Ken'ichi I, Kawachi I. Neighborhood social capital and crime victimization: Comparison of spatial regression analysis. *Soc Sci Med* 2012.
- EpiCenter: California Injury Data Online. Assault/Homicide. Sacramento, CA: California Department of Public Health. http://epicenter.cdph.ca.gov/ReportMenus/InjuryDataByTopic.aspx. Accessed on February 24, 2013.

Detailed Definition:

- Numerator: number of violent crimes
- Denominator: total population (from FBI data)

- $Indicator\ (rate) = \frac{Numerator}{Denominator} \times 1{,}000$
- Stratification: race/ethnicity stratification not available
- Interpretation: healthier communities will have lower rates of violent crimes

Data Description:

- Data source: Uniform Crime Reports from the Federal Bureau of Investigation http://www.fbi.gov/about-us/cjis/ucr/ucr
- Years available: 2006, 2007, 2008, 2009, 2010
- Updated: annually
- Geographies available: cities/towns, counties, regions (derived), state

Methodology:

The Uniform Crime Reports is a nationwide, cooperative statistical effort of nearly 18,000 city, university and college, county, state, tribal, and federal law enforcement agencies that voluntarily report data on crimes discovered by police and those reported to the police by the general public. The Federal Bureau of Investigation compiles these reports in a standard format annually. Four types of major crimes fall into the category of violent crimes: 1) murder and non-negligent manslaughter, 2) forcible rape, 3) robbery, 4) aggravated assault. These tend to be more reliably reported than other less serious crimes, but underreporting has been well documented.

UCR crime reports were downloaded as Excel files from the FBI website (http://www.fbi.gov/about-us/cjis/ucr/ucr, Table 8). The violent crime rate per 1,000 populations was calculated for each town/city by dividing the number of violent crimes by the total population and multiplying by 1,000. Standard errors were calculated using the Poisson distribution and were used to compute 95% confidence intervals. Relative standard errors (to assess statistical reliability) were computed by dividing the standard error by the violent crime rate. Cities and towns were assigned a 5-digit U.S. Census place code based on a database join matching city names in the UCR file to a reference file of California place names, census city and county FIPS codes, and population counts from the 2010 Census. City name matches were manually inspected for concordance with population counts and errors were corrected for a few false positive matches. To obtain county rates, the number of violent crimes and population of cities were aggregated by county FIPS codes. To obtain regional rates, the number of violent crimes and population of counties were aggregated by county FIPS conforming to the metropolitan transportation organizations (MPO) regions as reported in the 2010 California Regional Progress Report (Appendix C).

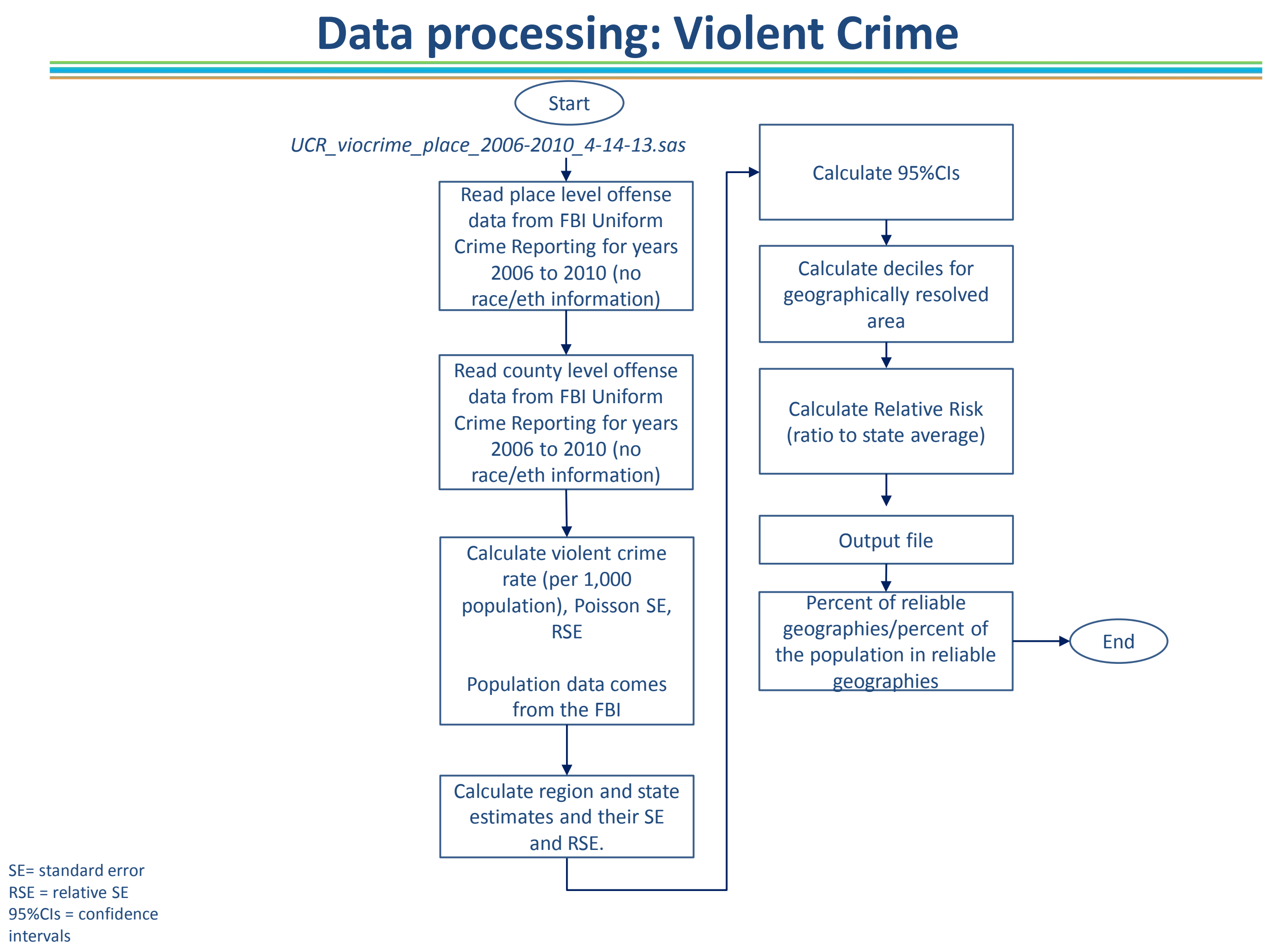
Limitations:

Crime data are based on incidents that are reported to law enforcement agencies. Limitations in the use of these data are detailed at http://www.fbi.gov/about-us/cjis/ucr/ucr-statistics-their-proper-use. Because some crimes such as sexual assaults are underreported, crime data are not always accurate. Even with underreporting, it still may be possible to observe trends over time if reporting biases hold steady within a jurisdiction. The Healthy Communities Indicators Project is also exploring an indicator based on injury rates from assaults treated in emergency departments. This measure has the advantage of being less susceptible to reporting biases, and can provide information at zip code geographies and for different race/ethnicity groups.

However, this alternative provides information on where the victim resides rather than where the assault occurred.

Data Processing

SAS program: UCR_viocrime_place_2006-2010_4-14-13.sas



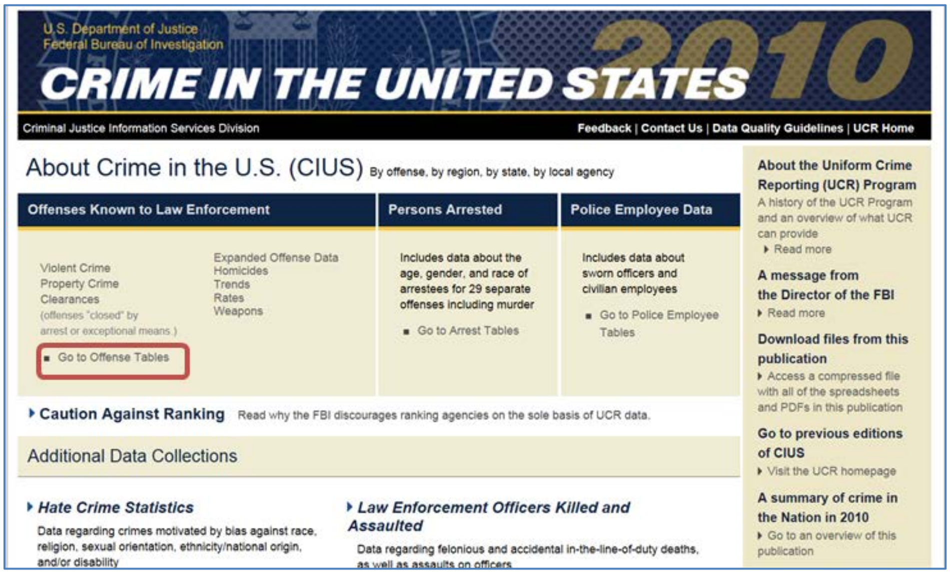
Data Acquisition and Preprocessing

Data was downloaded from the Federal Bureau of Investigation website on September 5th, 2012.

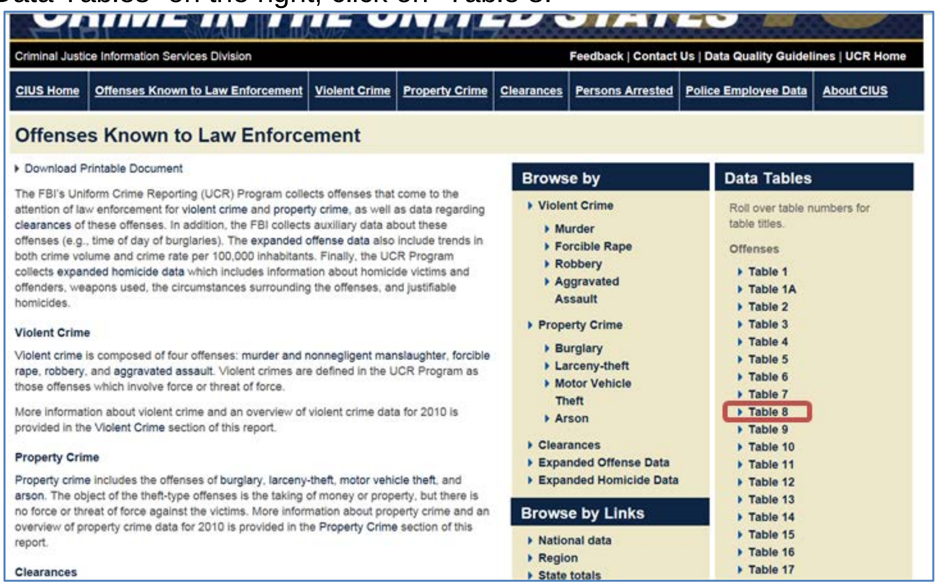
Data download steps:
1. Go to https://www.fbi.gov/about-us/cjis/ucr/ucr.
2. Under “Stats and Services” click on “Crime Statistics/UCR.”

3. Click on “2010” under “Uniform Crime Reports.”

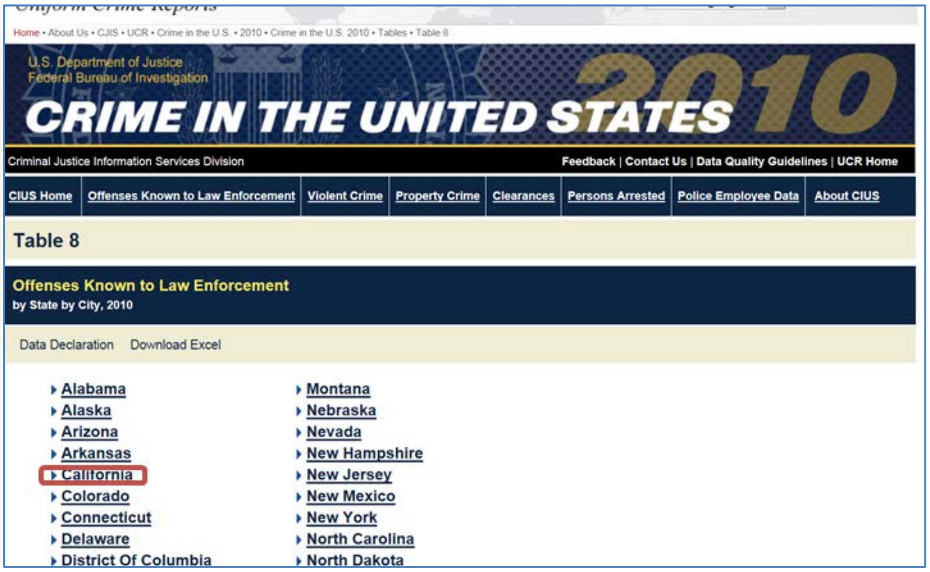
4. Click on “Go to Offense Tables.”



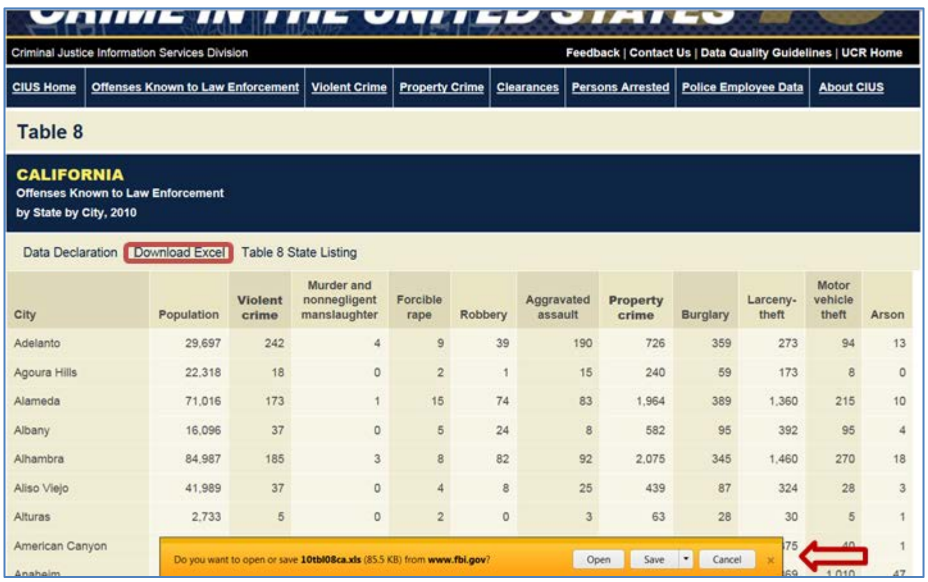
5. Under “Data Tables” on the right, click on “Table 8.”



6. Click on “California.”



7. Table of ‘Offenses Known to Law Enforcement’ for California by City should appear. Click on ‘Download Excel’



City	Population	Violent crime	Murder and nonnegligent manslaughter	Forcible rape	Robbery	Aggravated assault	Property crime	Burglary	Larceny-theft	Motor vehicle theft	Arson
Adelanto	29,697	242	4	9	39	190	726	359	273	94	13
Agoura Hills	22,318	18	0	2	1	15	240	59	173	8	0
Alameda	71,016	173	1	15	74	83	1,964	389	1,360	215	10
Albany	16,096	37	0	5	24	8	582	95	392	95	4
Alhambra	84,987	185	3	8	82	92	2,075	345	1,460	270	18
Aliso Viejo	41,989	37	0	4	8	25	439	87	324	28	3
Alturas	2,733	5	0	2	0	3	63	28	30	5	1

The downloaded Excel file was converted to a Comma Separated Value file (.csv). The footnotes and the first four rows of the original table were edited out. The place name "La Canada Flintridge" was changed to "La Canada." Note: Lake Shastina is not a Census Designated Place (unincorporated city near Weed, Siskiyou County that has its own police force). The variables of interest were the first three columns of the file: city, population, and violent crime.

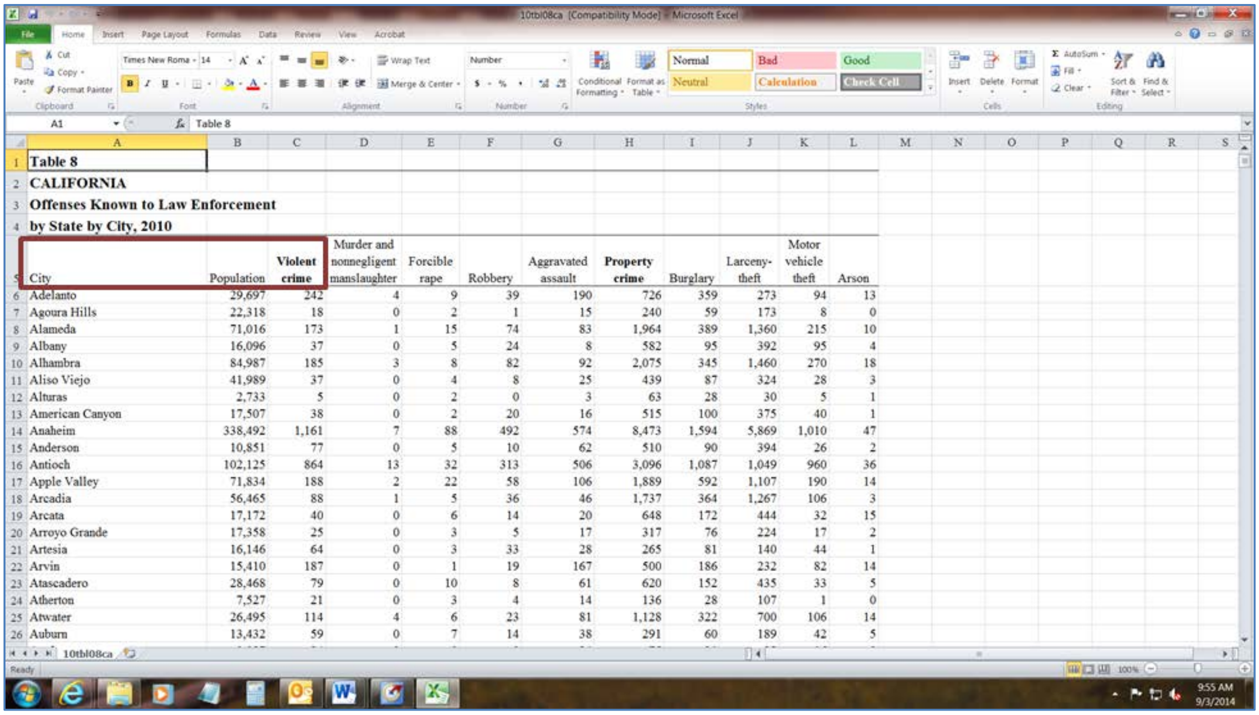


Table 8

CALIFORNIA

Offenses Known to Law Enforcement

by State by City, 2010

City	Population	Violent crime	Murder and nonnegligent manslaughter	Forcible rape	Robbery	Aggravated assault	Property crime	Burglary	Larceny-theft	Motor vehicle theft	Arson
Adelanto	29,697	242	4	9	39	190	726	359	273	94	13
Agoura Hills	22,318	18	0	2	1	15	240	59	173	8	0
Alameda	71,016	173	1	15	74	83	1,964	389	1,360	215	10
Albany	16,096	37	0	5	24	8	582	95	392	95	4
Alhambra	84,987	185	3	8	82	92	2,075	345	1,460	270	18
Aliso Viejo	41,989	37	0	4	8	25	439	87	324	28	3
Alturas	2,733	5	0	2	0	3	63	28	30	5	1
American Canyon	17,507	38	0	2	20	16	515	100	375	40	1
Anaheim	338,492	1,161	7	88	492	574	8,473	1,594	5,869	1,010	47
Anderson	10,851	77	0	5	10	62	510	90	394	26	2
Antioch	102,125	864	13	32	313	506	3,096	1,087	1,049	960	36
Apple Valley	71,834	188	2	22	58	106	1,889	592	1,107	190	14
Arcadia	56,465	88	1	5	36	46	1,737	364	1,267	106	3
Arcata	17,172	40	0	6	14	20	648	172	444	32	15
Arroyo Grande	17,358	25	0	3	5	17	317	76	224	17	2
Artesia	16,146	64	0	3	33	28	265	81	140	44	1
Arvin	15,410	187	0	1	19	167	500	186	232	82	14
Atascadero	28,468	79	0	10	8	61	620	152	435	33	5
Atherton	7,527	21	0	3	4	14	136	28	107	1	0
Atwater	26,495	114	4	6	23	81	1,128	322	700	106	14
Auburn	13,432	59	0	7	14	38	291	60	189	42	5

Road Traffic Injuries

Definition

Full Title: Annual number of fatal and severe road traffic injuries per population and per miles traveled by transport mode

Indicator ID: 753

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Safe, sustainable, accessible and affordable transportation options

Description of significance and health connection:

Transportation accidents are the second leading cause of death in California for people under the age of 45 and account for an average of 4,018 deaths per year (2006-2010). Risks of injury in traffic collisions are greatest for motorcyclists, pedestrians, and bicyclists and lowest for bus and rail passengers. Minority communities bear a disproportionate share of pedestrian-car fatalities; Native American male pedestrians experience 4 times the death rate as Whites or Asians, and African-Americans and Latinos experience twice the rate as Whites or Asians.

Key References:

- Chang D. *National pedestrian crash report*. Washington, DC: National Center for Statistics and Analysis, National Highway Traffic Safety Administration, U.S. Department of Transportation; 2008. Report No.: DOT HS 810 968. Accessed October 27, 2013.
- Beck LF, Dellinger AM, O'Neil ME. Motor vehicle crash injury rates by mode of travel, United States: Using exposure-based methods to quantify differences. *Am J Epidemiol* 2007; 166(2):212-218.
- Leaf WA, Preusser DF*. Literature review on vehicle travel speeds and pedestrian injuries*. National Highway Traffic Safety Administration, U.S. Department of Transportation; 1999. Report No.: DOT HS 809 021. Accessed October 27, 2013.
- California Department of Public Health. *Vital Statistics Query System*, 2006-2010. Accessed October 27, 2013.
- Naumann RB, Beck LF. Motor vehicle traffic-related pedestrian deaths – United States, 2001-2010. *MMWR* 2013; 62(15): 277-282.

Detailed Definition:

- Annual number of fatal and severe road traffic injuries per population by transport mode
    - Numerator: Annual number of injuries
    - Denominator 1: Total population
    - $Indicator\ 1\ (rate) = \frac{Numerator}{Denominator\ 1} \times 100{,}000$
- Annual number of fatal and severe road traffic injuries per miles traveled by transport mode
    - Numerator: Annual number of injuries
    - Denominator 1: Annual number of miles traveled
    - $Indicator\ 2\ (rate) = \frac{Numerator}{Denominator\ 2} \times 1{,}000{,}000{,}000$

- Stratification: victim mode of transport (bicyclist, bus, car/pickup, motorcycle, pedestrian, truck, vehicles), severity of injury (fatal, severe).
- Interpretation: healthier communities will show lower rates (or zero) of severe and fatal injuries for any mode of transportation.

Data Description:

- Data sources:
    - Numerator: Statewide Integrated Traffic Records System (SWITRS), California Highway Patrol (CHP), 2002-2010 data from the Transportation Injury Mapping System (TIMS) (http://www.tims.berkeley.edu).
    - Denominator 1: Historical Population and Housing Estimates for Cities, Counties, and the State, 2000-2010, Demographic Research Unit, Department of Finance (DOF) (http://www.dof.ca.gov/research/demographic/reports/estimates/e-8/2000-10/view.php). U.S. Census Bureau, 2006-2010 American Community Survey (http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml).
    - Denominator 2: California Public Road data (CPR); Division of Research, Innovation and System Information; Office of Highway System Information & Performance; Highway Performance Monitoring System; California Department of Transportation http://www.dot.ca.gov/hq/tsip/hpms/datalibrary.php). U.S. Department of Transportation, Federal Highway Administration, 2009 National Household Travel Survey (NHTS) (http://nhts.ornl.gov). All data accessed 7/2013.
- Years available: 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2002-2004, 2005-2007, 2008-2010, 2006-2010.
- Updated: 1, 3, and 5 year intervals
- Geographies available: Census tracts, cities/towns, counties, regions (derived), consolidated metropolitan statistical areas (CMSA), and state

Methodology:

Numerator: Collision data for severe and fatal injuries occurring between 2002 and 2010 were downloaded from TIMS and geocoded to the 2010 U.S. Census tracts and places. Specific coordinates within counties were not reported for 7.7% of injuries, which could not be geocoded to a specific city or Census tract. CHP defines severe injuries as those other than fatal injuries that include the following: broken or fractured bones; dislocated or distorted limbs; severe lacerations; skull, spinal, chest or abdominal injuries that go beyond other visible injuries; unconsciousness at or when taken from collision scene; and severe burns. Fatal injuries are deaths from collisions occurring within 30 days after the collision date. Victim mode of transport was classified into 6 groups: car/pickup, truck, bus, motorcycle, pedestrian, and bicyclist. Denominator 1: annual total population for 480 incorporated cities, counties, and California from 2002 to 2010 was acquired from DOF. Three-year population averages were calculated for 2002-2004, 2005-2007, and 2008-2009. Total population counts for census tracts and 1043 census designated places for the period 2006-2010 were obtained from the American Community Survey. Denominator 2: Daily vehicle miles traveled for cities, counties, and California from 2002 to 2010 was abstracted from CPR and multiplied by 365 to estimate annual vehicle miles traveled. Annual miles traveled by bicyclists and pedestrians for CMSAs (county clusters for major metropolitan areas) and California was estimated between 2002 and 2010 from the NHTS by applying the annual rate of change between 2001 and 2009. Three and five-year averages were calculated. Indicator: the rate of collisions was calculated as injuries (severe or fatal) per 100,000 people (denominator 1) and injuries (severe or fatal) per 10^9 miles traveled (denominator 2), for each of the 6 victim modes of transport. The standard error was calculated using a Poisson approximation as: $rate/\sqrt{injuries}$. Relative standard errors, 95% confidence intervals, and decile ranking of places were also calculated. Regions were based on counties of metropolitan transportation organizations as reported in the 2010 California Regional Progress Report (Appendix C).

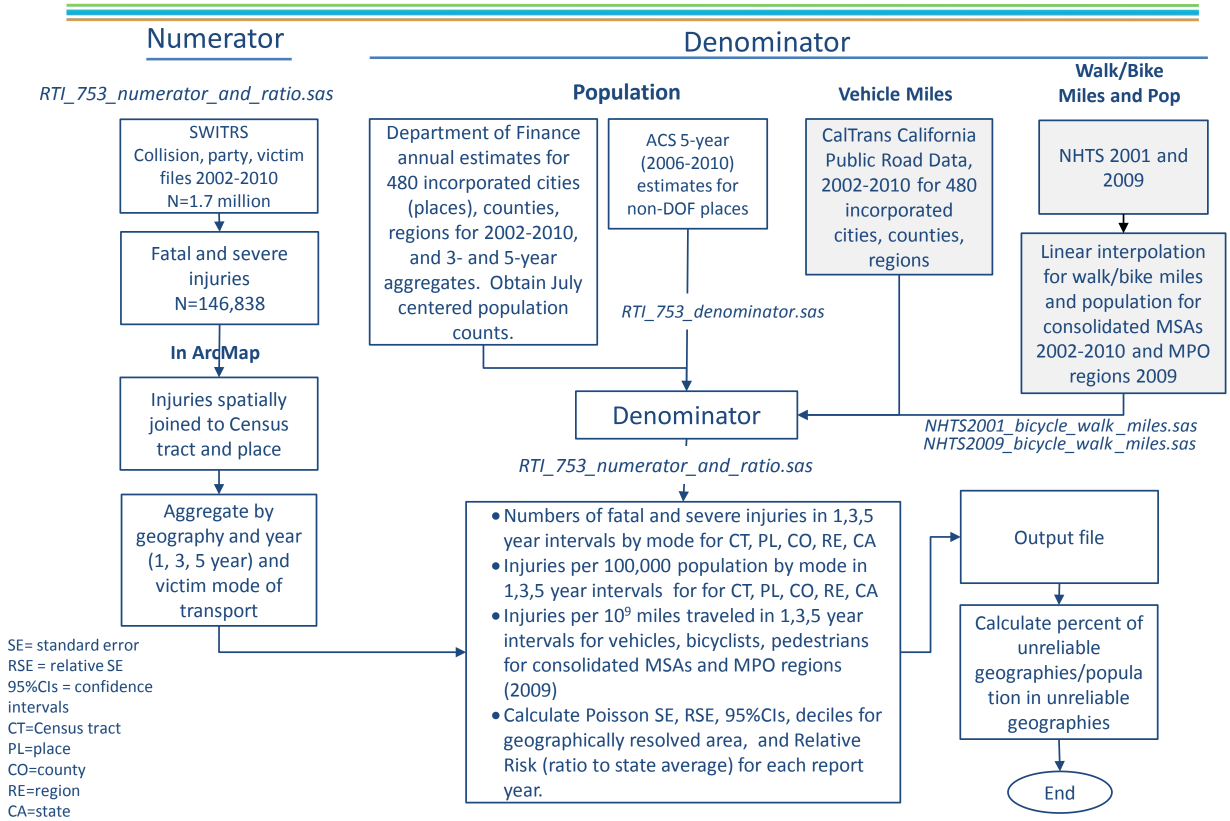
Strengths and limitations:

SWITRS provides information on injury occurrence for census tracts and cities, geographies not currently available from death certificates or data on hospitalizations or emergency room visits. SWITRS undercounts "non-traffic" injuries that occur off of public roads. The occurrence of the injury may not match the geography where the victim resides, although this discrepancy decreases at larger geographic units (counties, states). Compared to death certificates and hospitalizations, SWITRS is known to undercount both fatal and severe injuries. This may be especially true for victims that are low income, do not have health insurance, or are undocumented. Furthermore, collision data do not address disability and mental health impacts or economic losses.

Data Processing

SAS programs:
RTI_753_denominator_7-19-13.sas
RTI_753_numerator_and ratio_6-16-14.sas

Data Processing: Road Traffic Injuries



Data Acquisition and Preprocessing

Table 14 shows the years, modes and geographies combinations for which data was available to construct indicators 1 and 2.

Numerator

Numerator data is from the Statewide Integrated Traffic Records System (SWITRS), California Highway Patrol (CHP). It was downloaded from the Transportation Injury Mapping System (TIMS) (http://www.tims.berkeley.edu) of the Safe Transportation Research and Education Center (SafeTREC) at the University of California, Berkeley, for the years 2002 to 2010.

The codebooks showing the structure of SWITRS files and the variables available are in Appendix C. The downloaded file contained information for 1.7 million geocoded collisions of all severity levels. The file was imported into SAS 9.3 to create a new file with a subset of only serious or fatal injuries (N=146,838).

The geocoded collisions were displayed in ArcMap (see Toolbox Displaying XY Data in ArcMap). Spatial join methods were used to assign the location of occurrence of collisions to places and Census tracts (see Identifying Census blocks centroids within CWS boundaries using spatial join, Unsafe Drinking Water indicator). The attribute tables after the join were exported as .dbf files and later converted to an Excel format.

Denominators 1 and 2

This indicator uses two of the same data sources as the "Annual miles traveled by occurrence" indicator. Please refer to the Data Acquisition and Preprocessing section of that indicator chapter to read about how to download data for denominator 1, total population, and denominator 2, annual miles traveled by mode.

Table 14. Numerators and Denominators for Road Traffic Injuries Indicator, 2002-2010, Healthy Community Indicators Project

		Numerator (SWITRS)* Geography						Population#- Denominator 1 Geography						Miles Traveled¥ - Denominator 2 Geography					
Time Period	Mode(s)	Census Tract (CT)	Place (PL)	County (CO)	Region (RE)	CMSA (R4)†	State (CA)	Census Tract (CT)	Place (PL)	County (CO)	Region (RE)	CMSA (R4)	State (CA)	Census Tract (CT)	Place (PL)	County (CO)	Region (RE)	CMSA (R4)	State (CA)
Annual, 2002-2010	Vehicles §	Y	Y	Y	Y	N	Y	N	Y	Y	Y	N	Y	N	Y	Y	Y	N	Y
	Car/Pick up, Bus, Truck, Motorcycle	Y	Y	Y	Y	N	Y	N	Y	Y	Y	N	Y						
	Walk, Bicycle+	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	N	N	N	Y	Y
	All modes	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	N	N	N	Y	Y
3-Year (2002-2004, 2005-2007, 2008-2010)	Vehicles	Y	Y	Y	Y	N	Y	N	Y	Y	Y	N	Y	N	Y	Y	Y	N	Y
	Car/Pick up, Bus, Truck, Motorcycle	Y	Y	Y	Y	N	Y	N	Y	Y	Y	N	Y						
	Walk, Bicycle+	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	N	N	N	Y	Y
	All modes	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	N	N	N	Y	Y
5-Year (2006-2010)	Vehicles	Y	Y	Y	Y	N	Y	Y‡	Y	Y	Y	N	Y	N	Y	Y	Y	N	Y
	Car/Pick up, Bus, Truck, Motorcycle	Y	Y	Y	Y	N	Y	Y‡	Y	Y	Y	N	Y						
	Walk, Bicycle+	Y	Y	Y	Y	Y	Y	Y‡	Y	Y	Y	Y	Y	N	N	N	N	Y	Y
	All modes	Y	Y	Y	Y	Y	Y	Y‡	Y	Y	Y	Y	Y	N	N	N	N	Y	Y

* Serious and fatal injuries are separate numerators
# Department of Finance population estimates by year, 2002-2010
¥ California Public Roads Data, Department of Transportation. All motorized vehicles combined.
† CMSA=Consolidated Metropolitan Statistical Area
§ All motorized vehicles combined for risk-based estimates based on vehicle miles traveled (Highway Performance Monitoring System, CalTRANS)
+ Miles traveled data from the National Household Travel Survey
‡ ACS 5-year population estimates

Poverty Rate

<u>Definition</u>

Full title: Overall, Concentrated, and Child (under 18 years of age) Poverty Rate

Indicator ID: 754

Healthy Community Framework: Economic and Social Development

Aspirational goal: Living wage, safe and healthy job opportunities for all

Description of significance and health connection:

Poverty limits the acquisition of basic material necessities and it can impact the ability to live a healthy life. It restricts people's access to housing, food, education, jobs, and transportation. Poverty is associated with societal exclusion and higher incidence and prevalence of mental illness. Poor people are more likely to live in dangerous or under-resourced environments and to work in hazardous conditions, with greater risk of injury, and greater exposure to pesticides, lead, and outdoor air pollution. Low income people are more likely to be uninsured and to have limited access to quality health care; are more likely to suffer from chronic diseases like diabetes and heart disease, acute and chronic stress, and to die prematurely. It's been estimated that 133,250 (6%) of the 2.4 million U.S. deaths in 2000 could be attributed to poverty.

The children of the poor grow up in environments that are harmful to their development and health and the schools they attend typically provide substandard education. Low educational attainment among poor children will affect their job prospects and income level in adulthood, which could perpetuate the poverty cycle across generations. Poverty tends to cluster geographically and the concentrated poverty rate, defined as the share of the poor living in neighborhoods where at least 40% of individuals live below the poverty level, has risen from 9.1% (2000) to 10.5% (2005-2009). Poverty is more common among ethnic minorities, the unemployed, the disabled, the homeless, the uneducated, and migrants.

Summary of evidence:

There is ample evidence strongly linking poverty and health, with a progressive improvement in health measures with increasing income. The prevalence of psychiatric disorders, including neurotic disorders, functional psychoses and alcohol and drug dependence is consistently more common among lower income people. Self-reported poor or fair health status, as opposed to good or better, is 5 times more common among the poor (<100% Federal Poverty Level [FPL]) than high income people (≥400% FPL). Diabetes incidence rates (age-standardized) decrease from 11.2% among the poor to 5.4% among high income people. Uninsured rates for the poor and near-poor (<300% FPL) range between 30-39% and are significantly higher than those among the non-poor (≥300%FPL). Poor to middle income people (<400%FPL) have worse access to care and receive worse care than high income people. Householders earning an annual salary ≤$24,999 are five times more likely to live in inadequate housing than those earning ≥$75,000. In a longitudinal study it was observed that people living in concentrated poverty areas had higher rates of obesity, diabetes, and depression, than those that moved out of those same areas.

Key References:

- Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. *National Healthcare Disparities Report 2012*. AHRQ Publication No. 13-0003, 2013.
- Braveman P, Egerter S. *Overcoming Obstacles to Health: Report from the Robert Wood Johnson Foundation to the Commission to Build a Healthier America*. Robert Wood Johnson Foundation, 2008. Accessed July 16th, 2012.
- Braveman PA, Cubbin C, Egerter S, Williams DR, Pamuk E. Socioeconomic disparities in health in the United States: what the patterns tell us. *Am J Public Health*. 2010; 100(S1): S186–S196.
- Centers for Disease Control and Prevention. *Health Disparities and Inequalities Report - United States, 2011*. Morbidity and Mortality Weekly Report, 2011; 60: 1-113.
- Galea S, Tracy M, Hoggatt KJ, DiMaggio C, Karpati A. Estimated deaths attributable to social factors in the United States. *Am J Public Health* 2011; 101(8):1456-1465.
- Gennetian L, Ludwig J, McDade T, Sanbonmatsu L. Why Concentrated Poverty Matters. *Pathways Spring* 2013; 10-13. http://ssrn.com/abstract=2326579 Accessed November 4th 2013.
- Kaplan G. *The Poor Pay More – Poverty's High Cost to Health*. The Robert Wood Johnson Foundation; 2009. Accessed Nov. 5th, 2013.
- Kneebone E, Nadeau C, Berube A. *The Re-Emergence of Concentrated Poverty: Metropolitan Trends in the 2000s*. Metropolitan Policy Program, Brookings Institution, 2011. Accessed November 5th, 2012.
- Murali V, Oyebode F. Poverty, social inequality and mental health. *Adv Psychiatr Treat* 2004; 10: 216-224.
- Wilkinson R, Marmot M, eds. *Social Determinants of Health: The Solid Facts*. 2nd ed. Copenhagen, Denmark: WHO; 2003. Accessed July 12th, 2012.

Detailed definition:

- Poverty:
  - Numerator: Number of individuals or children (under 18 years of age) whose income in the past 12 months is below the poverty level.
  - Denominator: Total number of individuals or children
  - $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Concentrated poverty:
  - Numerator: Total number of poor people in all the Census tracts with ≥40% overall poverty within the cities/towns, county, region, or the state.
  - Denominator: Total number of poor people in the cities/towns, county, region, or the state.
  - $\frac{\sum_{i=1}^{n} Number\ of\ poor\ people_i}{Total\ number\ of\ poor\ people} \times 100$, where n is the total number of Census tracts with ≥40% overall poverty in the cities/towns, county, region, or the state.
- Stratification: eight race/ethnicity groups (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white). 3 types of poverty (overall, child, and concentrated).
- Interpretation: healthier communities will have the lowest (or zero) percentage of the population living in poverty.

Data Description:

- Data source: U.S. Census Bureau, Decennial Census 2000, and American Community Survey (ACS): 2005-2007, 2008-2010, and 2006-2010 (http://factfinder2.census.gov).
- Years available:
  - Poverty: 2000, 2005-2007 (3-year), 2008-2010 (3-year), and 2006-2010 (5-year)
  - Concentrated poverty: 2000 and 2006-2010 (5-year)
- Updated: each year with rolling 3- and 5-years estimates.
- Geographies available:
  - Poverty: Census tracts, cities/towns, counties, regions (derived), state
  - Concentrated poverty: cities/towns, counties, regions (derived), state

Methodology:

The percentage of all individuals and children living below the poverty level where obtained from the Census 2000 (DP3 table) and the ACS (DP03 table). Standard errors of the estimates were obtained from the margin of error provided by the Census (not available for 2000). Relative standard errors and 95% confidence intervals were calculated. Concentrated poverty was calculated using Census tract (CT) data aggregated to city/town, county, region, or the state level. ArcGIS was used to assign CT to the city/town in which the tract centroid was positioned. The number of poor people was calculated using the percentage of people in poverty and the total population by CT, obtained from the Census 2000 (DP1) and the ACS (B01001). Census tracts were excluded from the concentrated poverty calculation if (a) 50% or more of the population lived in group quarters; (b) 50% or more of the population was enrolled in college or graduate school; (c) a federal or state prison or youth correctional facility (YCF) was located in the CT; or (d) total population was 500 people or less. Group quarters and school enrolled population were obtained from Census 2000 (DP1 and QTP19) and ACS (B26001 and S1401). Lists of state and federal prisons and YCF were obtained from the Department of Corrections and the Federal Bureau of Prisons. Prisons were manually geocoded using published geographical coordinates (https://wiki.toolserver.org/view/GeoHack) and validated by observation using Google Maps and published addresses. Decile rankings of places and relative risk were calculated for all three poverty measures. Regional estimates were based on county groupings associated with California metropolitan planning organizations as reported in the 2010 California Regional Progress Report (Appendix C).

Limitations:

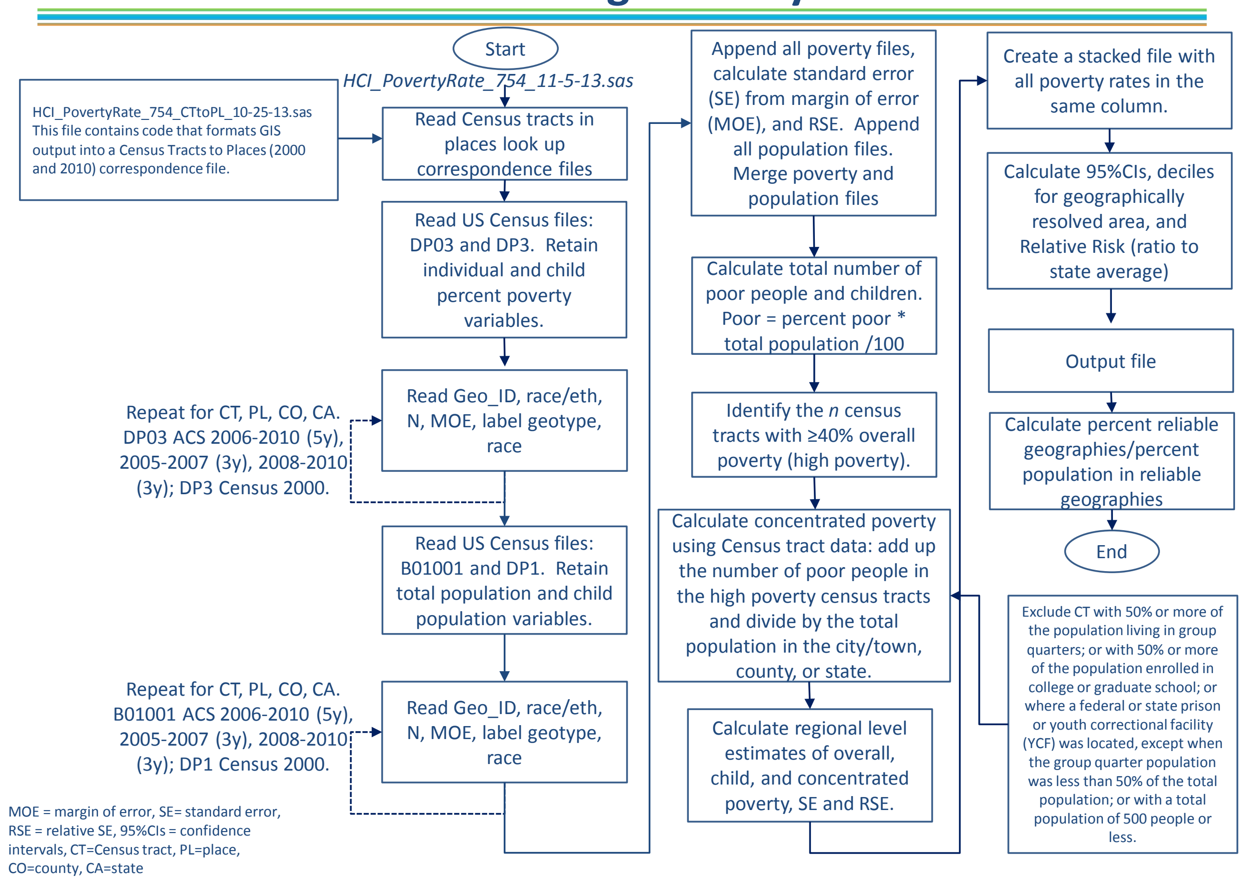
Poverty levels are not adjusted for differences in the cost of living. Caution should be exercised when comparing Census 2000 and ACS. The ACS collects income data on an ongoing monthly basis while the Census 2000 collected income data for a fixed period of time (1999). The ACS reports poverty for “people under 18 years” and the Census 2000 for “related children less than 18 years.” Concentrated poverty for cities/towns was obtained after intersecting the Census tract (CT) centroids with the city/town. Underestimation could have occurred if a CT was partially contained within a city, but its centroid was outside the city limits. Overestimation could have occurred when the population living in the CT but outside the city limits was included in the calculation.

Beginning in 2006, the population in group quarters (GQ) is included in the ACS. The part of the group quarters population in the poverty universe (for example, people living in group homes or those living in agriculture workers’ dormitories) is many times more likely to be in poverty than people living in households. Direct comparisons of the data would likely result in erroneous conclusions about changes in the poverty status of all people in the poverty universe.

Data Processing

SAS programs:
HCI_PovertyRate_754_CTtoPL_10-25-13.sas
HCI_PovertyRate_754_10-25-13.sas (See Appendix B)

Data Processing: Poverty Rate



Data Acquisition and Preprocessing

The percentage of all individuals and children living below the poverty level was precalculated and was obtained from tables DP3 (Census 2000) and DP03 (American Community Survey - ACS – years 2005-2007, 2008-2010, and 2006-2010). In order to present numerator and denominator estimates, the total number of people were obtained from tables DP1 (Census 2000) and B01001 (ACS). Group quarter population and student population data was used to determine if certain tracts should be excluded from the concentrated poverty calculation. Group quarter population was obtained from Census 2000 table DP1 and ACS table B26001; school enrollment was obtained from Census 2000 table QTP19 and ACS table S1401.

For instructions on how to obtain the data see the Toolbox section, Downloading Census Data from American FactFinder chapter. A total of 31 Census data tables were downloaded (see Table 15). Table 16 shows the variables used to construct the indicator file.

Table 15. Census Tables Downloaded from American FactFinder for the Poverty Rate Indicator by Data Set, Geographies and Race and Ethnic Groups

Dataset (Time period)	Geographies	Race and Ethnic groups	Census tables
2010 ACS 5-year Selected Population Tables (2006-2010)	Census Tract -140	Total and eight race/ethnicity groups Total population only	DP03, B01001 S1401 (school enrollment), B26001 (group quarters)
	Place - 160	Total and eight race/ethnicity groups	DP03, B01001
	County - 050	Total and eight race/ethnicity groups	DP03, B01001
	State - 040	Total and eight race/ethnicity groups	DP03, B01001
2010 ACS 3-year Selected Population Tables (2008-2010)	Census Tract -140		
	Place - 160	Total population	DP03, B01001
	County - 050	Total population	DP03, B01001
	State - 040	Total population	DP03, B01001
2007 ACS 3-year Selected Population Tables (2005-2007)	Census Tract -140		
	Place - 160	Total population	DP03, B01001
	County - 050	Total population	DP03, B01001
	State - 040	Total population	DP03, B01001
2000 SF4 Sample Data (2000)	Census Tract -140	Total and eight race/ethnicity groups Total population only	DP3, DP1 QTP19 (school enrollment)
	Place - 160	Total and eight race/ethnicity groups	DP3, DP1
	County - 050	Total and eight race/ethnicity groups	DP3, DP1
	State - 040	Total and eight race/ethnicity groups	DP3, DP1

Table 16. Variables Used to Construct the Poverty Rate Indicator by Census Table Number

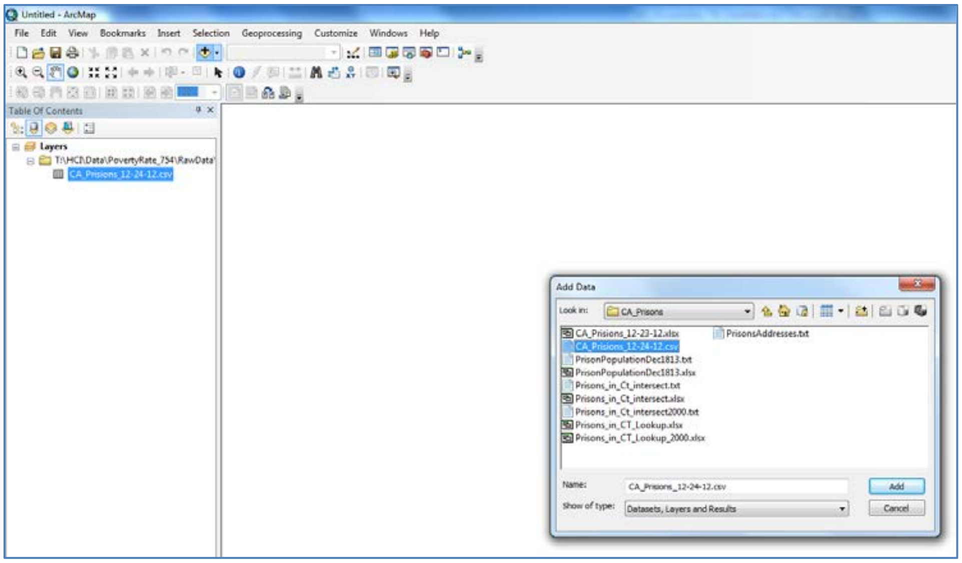
Table num. den.	Variable denominator	Variable numerator	Table percent	Percent	Standard Error Percent
Overall Poverty					
B01001	HD01_VD01 Total population	Calculated Variable denominator * variable percent/100	DP03 (2005-2007)	HC01_EST_VC112 Percent; PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12 MONTHS IS BELOW THE POVERTY LEVEL - All people	$\frac{HC01\_MOE\_VC112}{1.645}$
B01001	HD01_VD01 Total population	Calculated Variable denominator * variable percent/100	DP03 (2008-2010, 2006-2010)	HC03_VC166 Percent; PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12 MONTHS IS BELOW THE POVERTY LEVEL - All people	$\frac{HC04\_VC166}{1.645}$
DP1 (2000)	HC01_VC01 Total population	Calculated Variable denominator * variable percent/100	DP3 (2000)	Percent; POVERTY STATUS IN 1999 (below poverty level) - Individuals - Percent below poverty level	Not available
Child Poverty					
B01001	Aggregate across ages and sex: Male HD01_VD03 < 5 years, HD01_VD04 5-9 years, HD01_VD05 10-14 years, HD01_VD06 15-17 years, Female HD01_VD27 < 5 years, HD01_VD28 5-9 years, HD01_VD29 10-14 years, HD01_VD30 15-17 years	Calculated Variable denominator * variable percent/100	DP03 (2005-2007)	HC01_EST_VC114 Percent; PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12 MONTHS IS BELOW THE POVERTY LEVEL - Under 18 years	$\frac{HC01\_MOE\_VC114}{1.645}$

B01001	Aggregate across ages and sex: Male HD01_VD03 < 5 years, HD01_VD04 5-9 years, HD01_VD05 10-14 years, HD01_VD06 15-17 years, Female HD01_VD27 < 5 years, HD01_VD28 5-9 years, HD01_VD29 10-14 years, HD01_VD30 15-17 years	Calculated Variable denominator * variable percent/100	DP03 (2008-2010, 2006-2010)	HC03_VC167 Percent; PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12 MONTHS IS BELOW THE POVERTY LEVEL - Under 18 years	$\frac{HC04\_VC167}{1.645}$
DP1 (2000)	HC01_VC19 Population over 18 years of age	Calculated Variable denominator * variable percent/100	DP3 (2000)	HC02_VC111 Percent; POVERTY STATUS IN 1999 (below poverty level) - Individuals - Related children under 18 years - Percent below poverty level	Not available

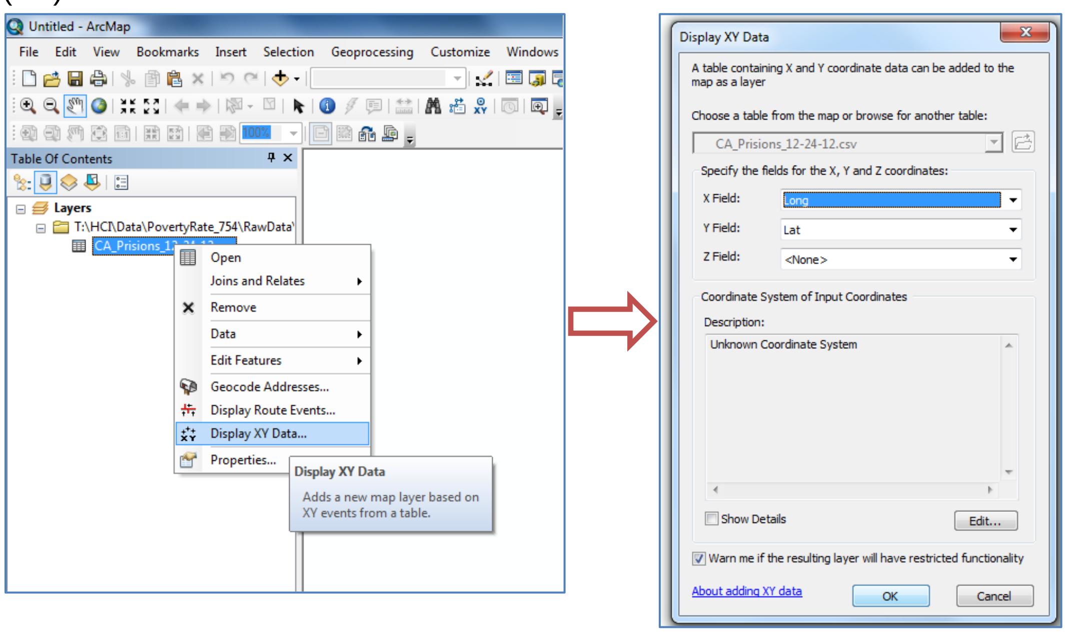
Geocoding of CA State and Federal Prisons and Youth Camps

It was necessary to identify the Census tracts where prisons are located in California, in order to exclude them from the concentrated poverty calculation. A list of state prisons and youth correctional facilities with addresses was obtained from a law firm website (http://www.wekeepyououtofjail.com/resources/california_prisons/). The list was updated and verified using information from the California Department of Corrections (http://www.cdcr.ca.gov/). A list of federal prisons was obtained from the Federal Bureau of Prisons (http://www.bop.gov/). Each institution was manually geocoded (Lat Long columns added) using the geographical coordinates published in Wikipedia through GeoHack (https://wiki.toolserver.org/view/GeoHack); geocoding was validated by observation using Google Maps (https://maps.google.com/) and the published addresses for the institutions. Following are the steps to assign the prisons to Census tracts for 2010 and 2000 using ArcMap.

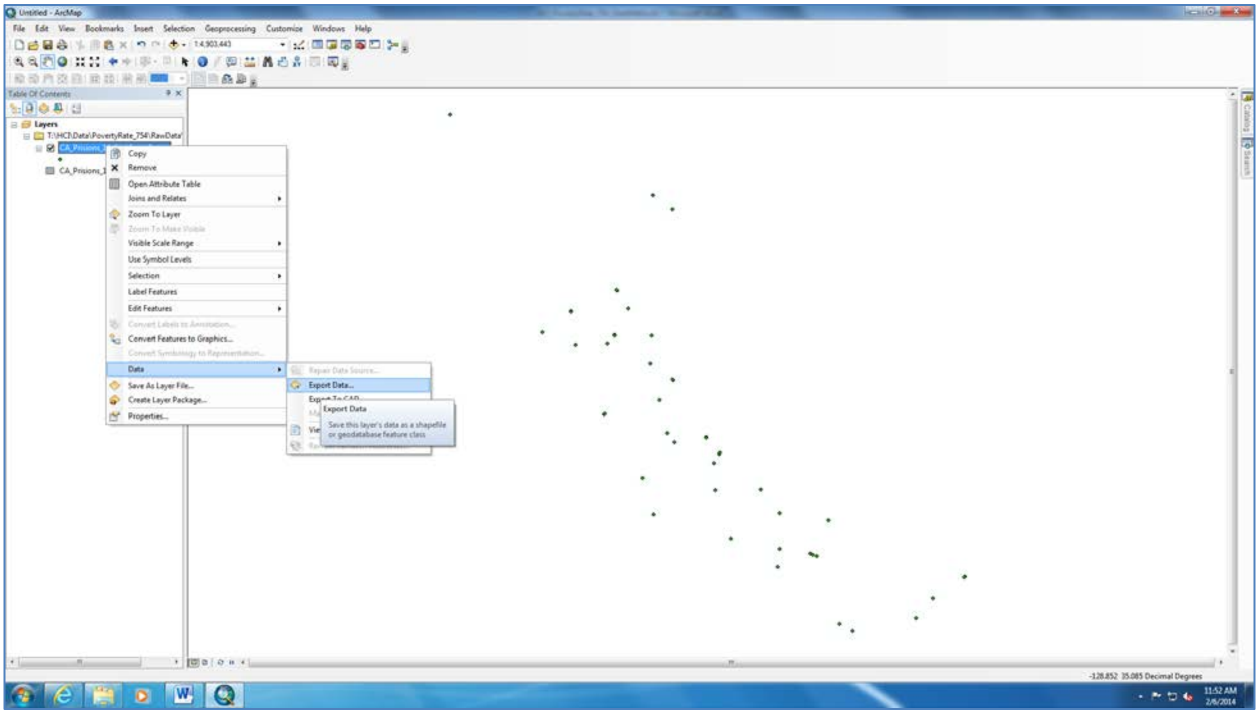
1. Add the data file with the geolocation of prisons into ArcMap:



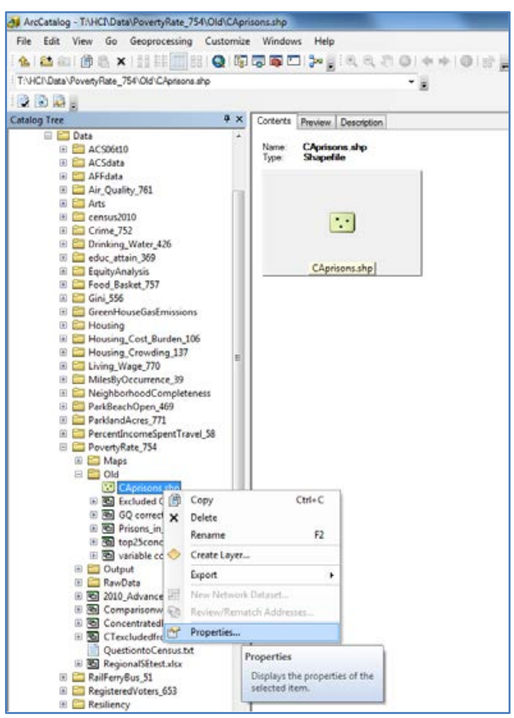
2. After the table is added in the Table of Contents, right click the table name, go to Display XY Data. A new window will open. Verify the X Field (Long) and Y Field (Lat). Click ok.



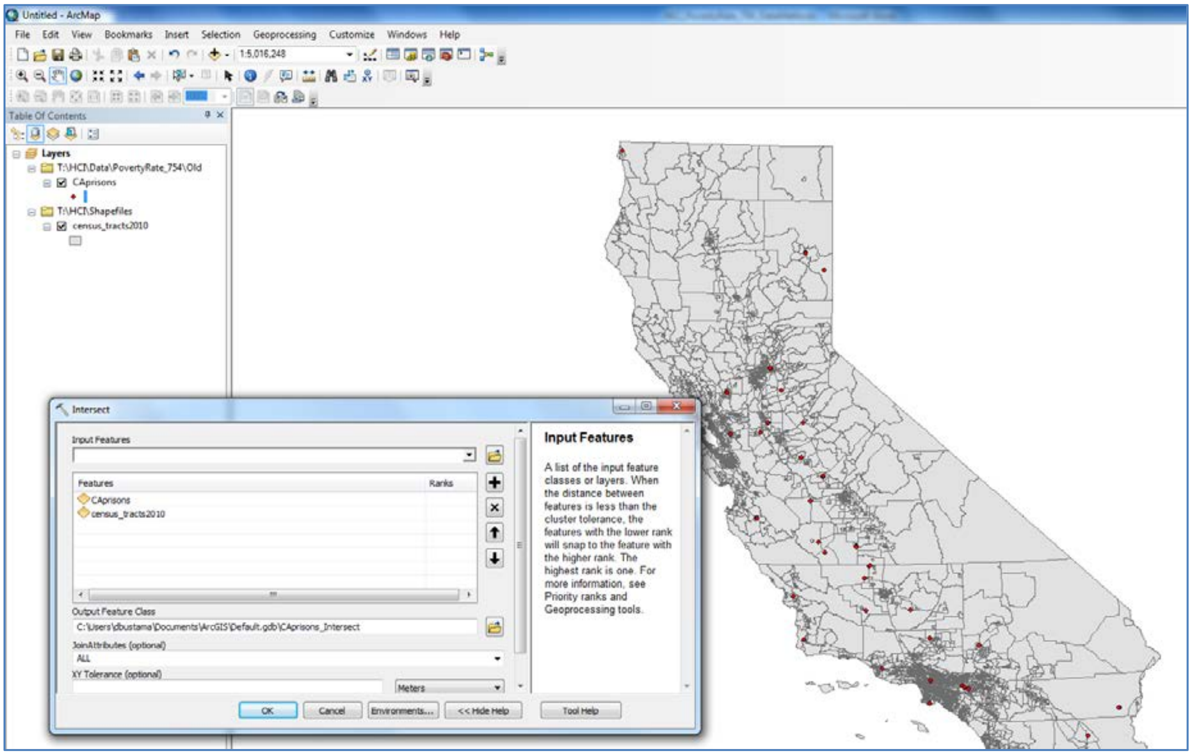
3. A message will appear indicating that the data does not have object id field, click ok.
4. To add an object id field it is necessary to create a shapefile: right click on the x,y layer, go to Data, click on Export Data, give the layer a name (CAprisons for instance) and save. It is not necessary to add the layer to the current map.



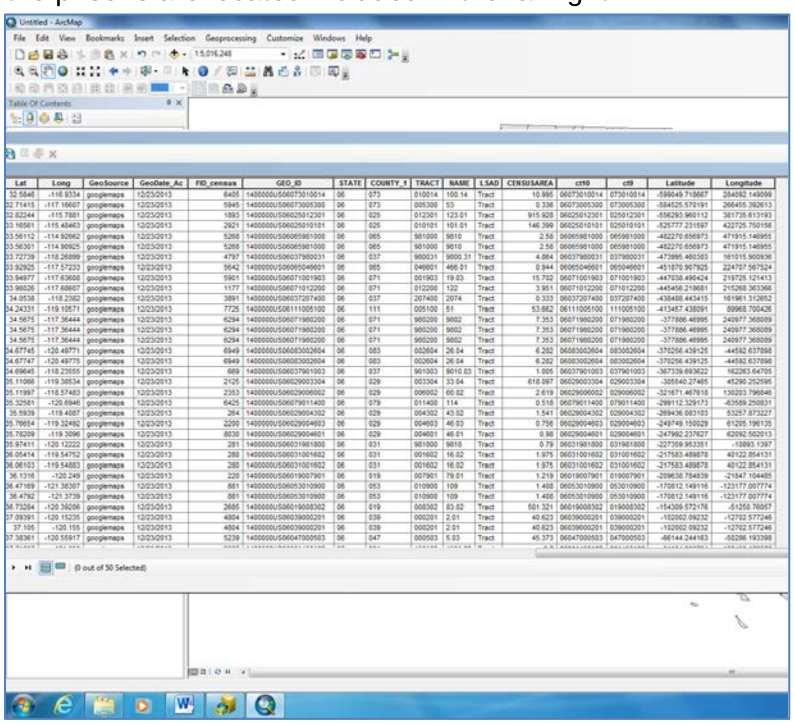
5. It is necessary to add a geographic system to the newly created shapefile CAprisons. Close the map project. Open ArcCatalog and find CAprisons. Right click on the file name click on Properties. In the Properties window select XY coordinate system, then the Geographic Coordinate System folder, and then the North America folder, and select NAD 1983. Once the geographic coordinate system appear under “Current coordinate system” click Apply.



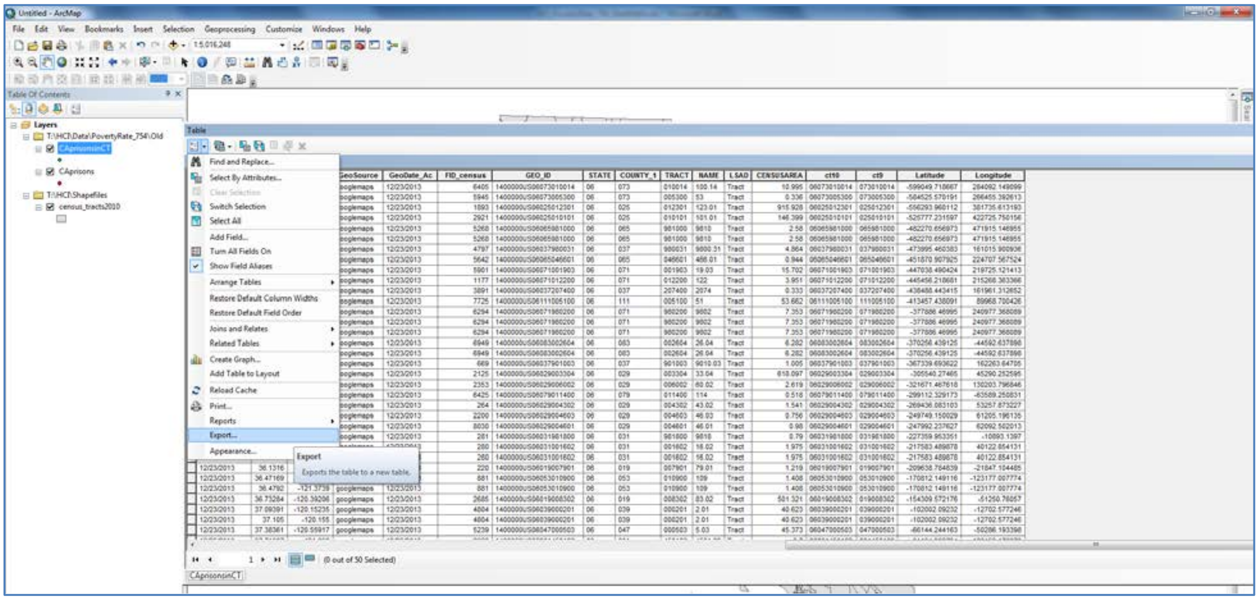
6. Open a new ArcMap project. Add the Census tracts 2010 shapefile layer, and then add the CAprisons layer. Go to Geoprocessing and select Intersect. Input the two layers in the map. Provide an output file name (CAprisonsinCT).

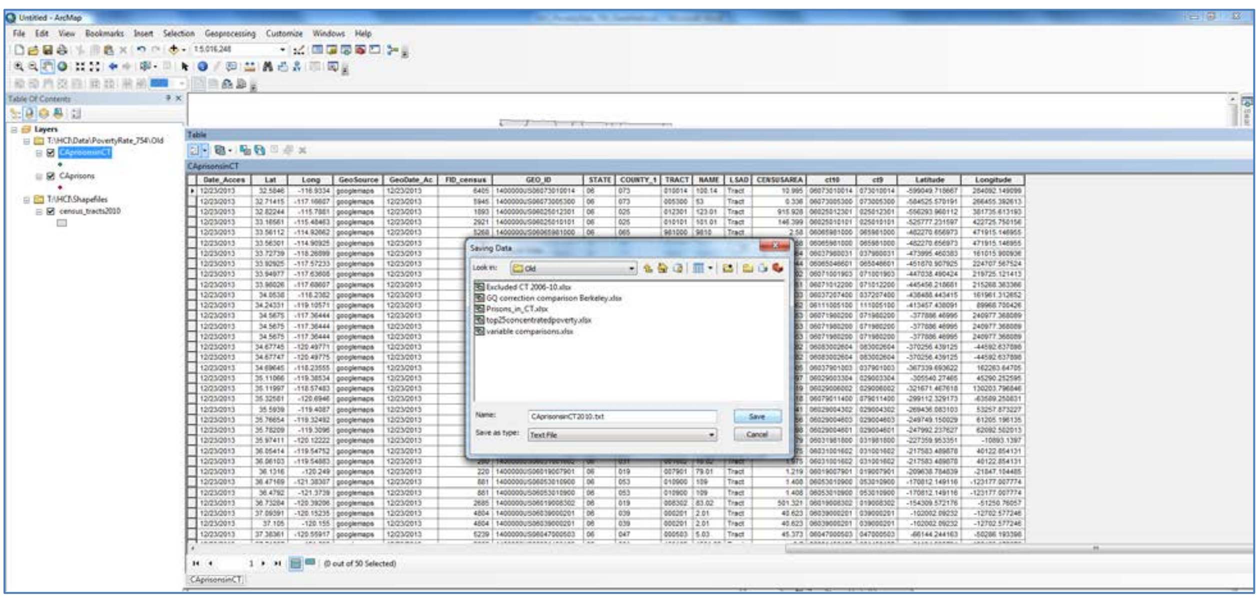


7. Open the attribute table of the resulting file. It should have the information on the CT where the prisons are located included in the far right.



8. Export the table as a txt file. The table can later be saved as an Excel file and formatted to use it in SAS as a lookup table of the tracts that have state or federal prisons for the year 2010.





9. The same process was conducted using the 2000 census tracts shapefile. Prisons that opened after 2000 were removed from the resulting Excel file.

Food Affordability

<u>Definition</u>

Full Title: Average cost of a market basket of nutritious food items relative to income

Indicator ID: 757

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Affordable, accessible and nutritious foods

Description of significance and health connection:

An adequate, nutritious diet is a necessity at all stages of life. Pregnant women and their developing babies, children, adolescents, adults, and older adults depend on adequate nutrition for optimum development and maintenance of health and functioning. Inadequate diets can impair intellectual performance and have been linked to more frequent school absence and poorer educational achievement in children. Nutrition also plays a significant role in causing or preventing a number of illnesses, such as cardiovascular disease, some cancers, obesity, type 2 diabetes, and anemia.

At least two factors influence the affordability of food and the dietary choices of families – the cost of food and family income. The inability to afford food is a major factor in food insecurity, which has a spectrum of effects including anxiety over food sufficiency or food shortages; reduced quality or desirability of diet; and disrupted eating patterns and reduced food intake.

Inadequate diet and physical inactivity are responsible for approximately 17% of deaths in the United States. In 2011, approximately 16% of Californians experienced food insecurity at some time during the course of the year. Approximately one-third of California's low income households comprised of single adults with children experienced food insecurity in 2005. Low-income African Americans (37.5%) and Latinos (32.6%) report a higher frequency of food insecurity than either Whites (28.6%) or Asians (24.1%). Low-income residents living in the Central Valley are more likely to be food insecure than other Californians.

Summary of evidence:

There is a large body of scientific research that shows nutrition disorders can be caused by an insufficient intake of food or certain nutrients, by an inability of the body to absorb and use nutrients, or by overconsumption of certain foods. Examples include obesity caused by excess energy intake, anemia caused by insufficient intake of iron, and impaired sight because of inadequate intake of vitamin A.

Key References:

- Mokdad et al. Actual Causes of Death in the United States, 2000. JAMA. 2004;291:1238-1245
- Harrison GG, Sharp M, Manalo-LeClair G, Ramiriez A, McGarvey N. Food security among California's Low-Income Adults Improves. But Most Severely Affected Do Not Share Improvement. Los Angeles: UCLA Center for Health Policy Research; 2007.

- Agricultural Research Service. Report of the Dietary Guidelines Advisory Committee on the Dietary Guidelines for Americans, 2010. Washington, DC: Department of Agriculture and United States Department of Health and Human Services; May 2010. http://www.cnpp.usda.gov/Publications/DietaryGuidelines/2010/DGAC/Report/2010DGACReport-camera-ready-Jan11-11.pdf, Accessed January 20, 2013.

Detailed Definition:

- Numerator: dollar amount to purchase the annual market basket of foods in USDA's low plan for meal eaten at home for a female-headed household with children under 18 years of age
    - $Annual\ cost\ per\ person\ \times number\ of\ persons\ in\ the\ household$
- Denominator: annual inflation-adjusted median household income
- $Indicator\ (ratio)\ = \frac{Numerator}{Denominator}$
- Stratification: eight race/ethnicity groups (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white).
- Interpretation: healthier communities will have lower food costs relative to income

Data Description:

- Data Sources: Center for Nutrition Policy and Promotion. Official USDA Food Plans: Average U.S. Cost of Food at Home at Four Levels http://www.cnpp.usda.gov/USDAFoodCost-Home.htm. American Community Survey. Households and families. Table S1101. Washington, DC: U.S. Census Bureau. http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml. American Community Survey. Median Income in the Last 12 Months by Family Type. Table B19126. Washington, DC: U.S. Census Bureau. http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml
- Years available: 2006 – 2010
- Updated: annually
- Geographies available: cities/towns, county, region, state

Methodology:

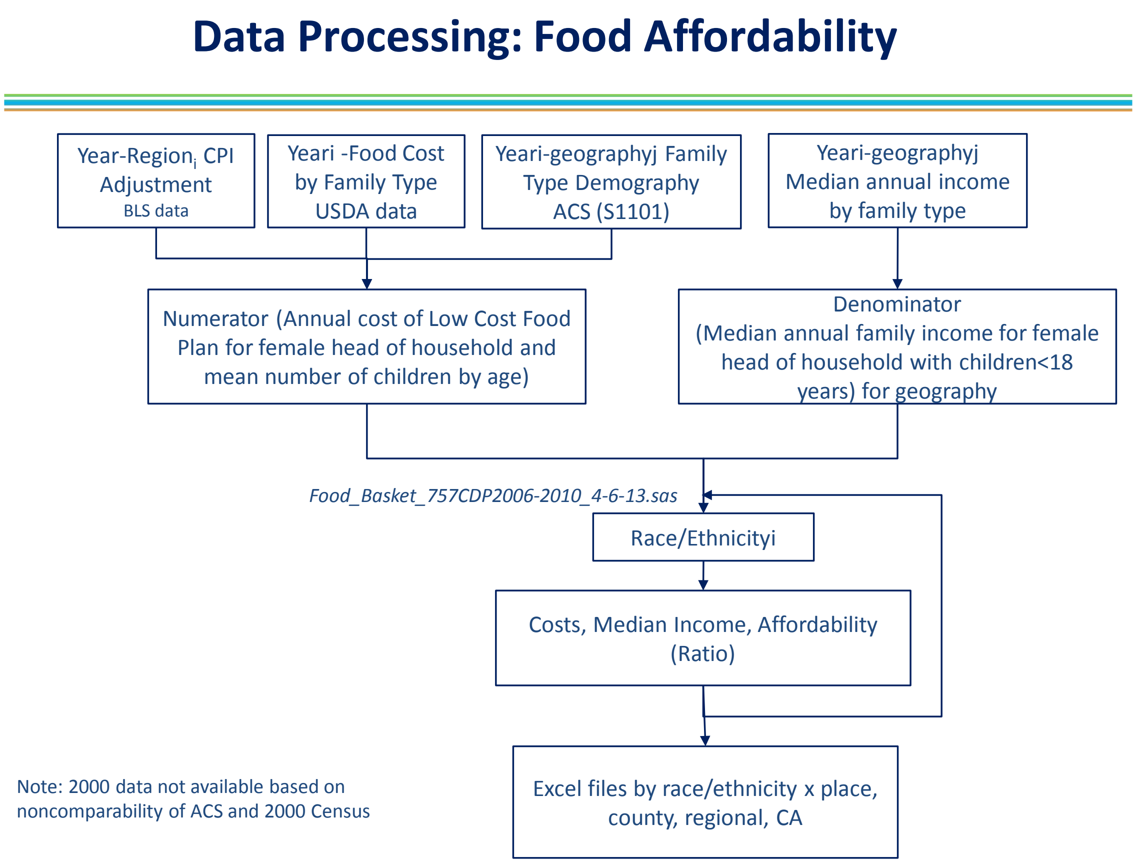
The annual cost of food is based on the USDA's low-cost food plan, which includes a market basket of items that families would have to purchase to provide a nutritious diet for each family member. To determine the costs, the USDA conducts a monthly national market basket survey of food items. The USDA tabulates per person costs by age for children <11 years, and age and gender for those aged 12-71+ years. For the HCDIP, family costs were the sum of costs for the female head of household and the per child-cost multiplied by the area average number of children under 18 years of age, taking into account their age distribution. The USDA annual costs were expressed in constant 2010 dollars and adjusted for regional differences (Los Angeles, Bay Area, San Diego, California average) based on the Consumer Price Index (CPI) food at home. Standard errors of the ratio were calculated using published margins of error for the average number of children and median income in the ACS. Standard errors were used to calculate 95% confidence intervals and relative standard errors. Regional estimates were based on county groupings associated with California metropolitan planning organizations as reported in the 2010 California Regional Progress Report (Appendix C).

Limitations:

Geographic variation in food costs is only considered at the regional level. For some families a substantial number of meals may be eaten outside of the home, and these costs are not included. Standard errors of the ratio (cost/income) are underestimated because standard errors for food costs were not available for USDA survey data.

<u>Data Processing</u>

SAS program: Food_Basket_757CDP2006-2010_4-14-13.sas



<u>Data Acquisition and Preprocessing</u>

Numerator: Cost of the Annual Market Basket of Foods

The numerator was calculated using data from multiple sources. To obtain the dollar amount to purchase the annual market basket of foods in USDA's low plan for meal eaten at home for a female-headed household with children under 18 years of age, the first step was to obtain data on the cost of foods comprising the USDA Low Cost Meals at Home plan.

The data was obtained from the Official USDA Food Plans: Monthly Cost of Food at Home at Four Levels, U.S. Average, June 2000-2010, downloaded from http://www.cnpp.usda.gov/USDAFoodCost-Home.htm (accessed 1/7/13). The USDA tabulates per person costs by age for children <11 years, and age and gender for those aged 12-71+ years.

Foods are selected to meet nutritional requirements for families of a given household size and age distribution. Annual costs are based on the June survey of a national market basket. Costs, driven by the amount and type of food, vary by age and gender of family members and household size. Costs can be regionally adjusted by the percent difference in the national CPI Food at Home to 3 California regional CPIs for Food at Home and a synthetic estimate for other areas of California.

The Bureau of Labor Statistics Website provides a Food at Home CPI for urban consumers (Los Angeles, San Diego, San Francisco-Oakland) and CPI Total for the U.S. http://data.bls.gov/pdq/querytool.jsp?survey=cu. The California Department of Finance provides Total CPI data http://www.dof.ca.gov/html/fs_data/latestecondata/FS_Price.htm. For other regions in California, The California Food at Home CPI was estimated from the ratio of US total and food and the California Total, for which published values at the California Department of Finance were available. The adjustment factor was calculated as follows:

$$California\ CPI_{food\ at\ home} = \frac{California\ CPI_{total} \times U.S.CPI_{food\ at\ home}}{U.S.CPI_{total}}$$

The next step was to obtain data on the average number of persons in female-head households with children under 18 years of age for California, its cities, towns, and counties. Data was available from Table S1101 of the American Community Survey (ACS), Households and Families, for the dataset 2006-2010 5-Year Estimates was downloaded using the FactFinder (accessed 1/7/13, see Toolbox Downloading Census Data from American FactFinder).

ACS data do not provide median income for family types that specify the number or ages of children. A definition of a family type that minimizes differential underestimation of annual food costs (numerator) appears to be a female head of household with children under 18 years of age, adjusting for number of children and their age distribution. This poses the question "What are the food costs relative to income for female headed families with children <18 years, given their specific demographic characteristics in a geographic area?" This makes comparisons between areas somewhat dependent on the number of children in families.

The numerator was calculated as follows:

$$Numerator = Cost_{19-50y} \times \sum_{i} Average\ number\ of\ children \times p_i \times Cost_i$$

Where

i are the children age categories <7 years and 7-18 years

p is the proportion of children in female headed households in the *i*th age category

$Cost_{<7y}$ is arithmetic average of costs from 1yr to 4-5year categories in USDA tables

$Cost_{7\text{-}18yrs}$ is arithmetic average of the average of $Costs_{Males7\text{-}18yrs}$ and $Costs_{Females7\text{-}18yrs}$ in USDA tables

$Cost_{<7\ and\ 7\text{-}18yrs}$ is arithmetic average of costs of $Cost_{<7y}$ and $Cost_{7\text{-}18yrs}$

The average number of children and p_i are from ACS Table S1101.

Because the average number of children was close to 2, per person costs used the 5% upward adjustment factor that USDA used in its food at home cost tables.

To create 5-year average U.S. costs for census tract and place analysis, for each year, 2006-2010, the CPI food at home for 2010 was subtracted from the CPI for food at home for each year. The difference was expressed as a percent different from 2010 and each year costs were converted to constant 2010 dollars. The 5-year average was based on five annual costs in constant dollars.

To create regional adjusted figures, the U.S. 2010 CPI for food at home was used as the reference value for CPI differences. Annual costs were converted to constant 2010 dollars and 5 annual constant dollar values were averaged for the regional cost.
The regions were defined from BLS documents:

LA Region	Los Angeles, Orange, Riverside, San Bernardino, and Ventura Counties
SF-Oakland	Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, San Benito, Santa Clara, Santa Cruz, Sonoma, and Solano Counties
SD	San Diego County

Denominator: Median Income by Family Type

Median Family Income in the Past 12 Months (In 2010 Inflation-Adjusted Dollars) By Family Type By Presence Of Own Children Under 18 Years, Table B19126, 2006-2010 American Community Survey 5-Year Estimates, was downloaded using the FactFinder (accessed 1/7/13, see Toolbox section, Downloading Census Data from American FactFinder chapter).

Ozone Air Pollution

Definition

Full Title: Annual Average Number of Unhealthy Days of Ozone Air Pollution

Indicator ID: 761

Healthy Community Framework: Quality and sustainability of the environment

Aspirational Goal: Clean air, soil and water, and environments free of excessive noise

Description of significance and health connection:

Clean air is a fundamental building block of human health. Air pollution from fixed and mobile sources (e.g. factories and cars, respectively) is a complex mixture of gases, fumes, and particles released into the atmosphere from the combustion of fossil fuels and evaporation of solvents. Ozone that forms at the ground-level and fine particulate matter (described separately) are two indicators of air pollution that are linked to short- and long-term adverse health effects. In California, the Air Resources Board estimated that 630 deaths, 4,200 hospital admissions, and 4.7 million lost school days could be prevented each year if California met its current statewide standard of 0.070 ppm for ozone (8-hour average). Most California residents are currently exposed to levels at or above the current State ozone standard during some parts of the year. Besides harming human health, ground level ozone can harm crops and potentially alter food quality and costs. Ozone can cause significant damage to a wide range of materials found in the built environment, which can shorten material life span and increase maintenance costs.

Summary of evidence:

Laboratory studies in which human subjects were exposed to measured concentrations of ozone for brief periods demonstrate that ozone can reduce lung function, increase respiratory symptoms, increase airway hyper-reactivity, and increase airway inflammation. Numerous community-based epidemiologic studies have shown that exposure to ozone is also associated with decreased lung function, respiratory symptoms, hospitalizations for cardiopulmonary causes, emergency room visits for asthma, and premature death. At higher daily concentrations, ozone increases asthma attacks, hospital admissions, daily mortality, and days of restricted activity and school absences. Children may be more affected by ozone than the general population due to effects on the developing lung and to relatively higher exposure than adults.

Key References:

- Drechsler D, Garcia C, Tran H, Mehadi A, Nystrom M, Propper R, et al. Review of the California Ambient Air Quality Standard For Ozone. Vol 4. Table B-5: California Annual Health Impacts of Current Ozone Concentrations Compared to the State 8-hour Ozone Standard of 0.070 ppm. Sacramento, CA: California Environmental Protection Agency, Air Resources Board; 2005. http://www.arb.ca.gov/carbis/research/aaqs/ozone-rs/rev-staff/vol4.pdf. Accessed on January 4, 2013.

- Committee On Estimating Mortality Risk Reduction Benefits From Decreasing Tropospheric Ozone Exposure. Estimating Mortality Risk Reduction and Economic Benefits from Controlling Ozone Air Pollution. Washington, DC: National Academy of Sciences; 2008. http://www.nap.edu/catalog/12198.html. Accessed on January 4, 2013.

Detailed Definition:

- Three-year annual average number of days from May to October that exceed the 8-hour federal ozone standard of 0.075 ppm.
- Stratification: eight race/ethnicity groups (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white).
- Interpretation: healthier communities will have lower averages of ozone air pollution

Data Description:

- Data source: Air Monitoring Network, Air Resources Board (CARB) http://www.arb.ca.gov/aqmis2/aqmis2.php
- Years available: three-year average of 2007-2009
- Updated: not stated
- Geographies available: zip codes, and population weighted averages for place, county, region, California

Methodology:

A network of approximately 300 air monitoring stations is maintained in California by the California Air Resource Board, local air pollution control districts, tribes and federal land managers. Ozone formation is strongly influenced by sunlight and temperature, so the preponderance of unhealthy days occurs from May through October. An unhealthy day is defined as a day (from May to October) in which the daily maximum value exceeded the federal standard. A statistical average of ozone measurements was calculated for each zip code based on monitoring stations within a 50 km radius. To calculate averages for cities, places, and regions, the number of unhealthy days was weighted by the population in the zip code based on U.S. 2010 Census files that parsed zip codes by place and county. To calculate race/ethnicity specific weighted averages, the proportion of race/ethnicities by zip code was calculated from U.S. 2010 Census files. Regional estimates were based on county groupings associated with California metropolitan planning organizations as reported in the 2010 California Regional Progress Report (Appendix C).

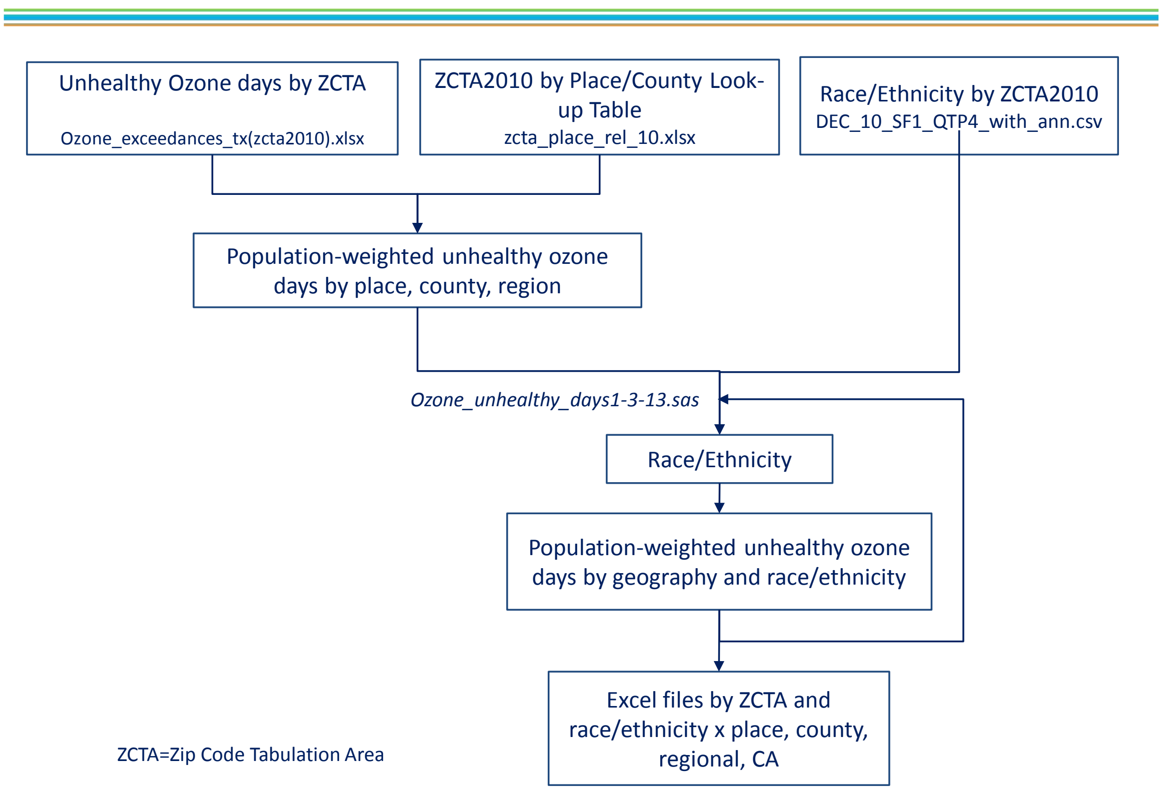
Limitations:

Geographic coverage was not complete because of the limited number and geographic extent of monitoring stations. The uncertainty of the interpolated values increases with distance from the nearest monitor. According to the Air Resources Board, values for areas greater than 50 km from the nearest monitor are very imprecise, and should be regarded as speculative. They are included for the sake of completeness, but should not be relied upon. Even within populated areas, monitoring stations are often located in areas that cannot detect highly localized areas of pollution that significant numbers or sensitive subgroups (daycare centers, schools, hospitals, etc.) in the population may encounter. Data were not available to present standard errors.

Data Processing

SAS program: Ozone_unhealthy_days4-14-13.sas

Data Processing: Ozone Air Pollution



Data Acquisition

Zip code specific ozone unhealthy healthy days data were acquired from Laura August (Laura.August@oehha.ca.gov), Office of Environmental Health Hazard Assessment (OEHHA) on 11/26/2012.

Living Wage

Definition

Full Title: Living wage and percent of families with incomes below the living wage

Indicator ID: 770

Healthy Community Framework: Economic and Social Development

Aspirational Goal: Living wage, safe and healthy job opportunities for all

Description of significance and health connection:

The living wage is the wage or annual income that covers the cost of the bare necessities of life for a worker and his/her family. These necessities include housing, transportation, food, childcare, health care, and payment of taxes. Health insurance has a direct impact on health through the provision of clinical preventive services (e.g., immunization, cancer screening) and treatment for illness and injury. Transportation is means to access jobs, education, and health-promoting goods and services, and may improve health through physical activity associated with walking and bicycling. Housing protects us from the elements, and functions as sanitary space for food preparation and other activities of daily living. Substandard housing can expose occupants to molds, air contaminants, and overcrowded conditions that impact respiratory, mental health, and other health conditions. Child care is a setting for safety and possible transmission of communicable diseases, but it aids in early child development. Quality child care programs have documented health impacts into adulthood. As discussed in other indicators in the Healthy Community Data and Indicators Project, low income populations and non-white race/ethnic have disproportionately lower wages, poorer housing, and higher levels of food insecurity. Several California studies have estimated the impacts of raising the wage rate in low-wage earners to the level of a living wage. These impacts include reduced mortality, improved mental health and self-rated health status, reduced disability at home and work, and greater educational attainment in their children.[1-2]

Summary of evidence:

There are scores of systemic reviews and hundreds of individual research publications documenting that absolute and relative level of income and/or socioeconomic status are associated with a wide range of health outcomes. Adverse health outcomes of low income occur over the entire lifecycle (e.g., infant mortality, all causes and specific diseases, self-reported health status, mental health). The most compelling evidence is from studies that follow different socioeconomic groups within entire communities or large populations (>400,000) over time. The mortality/income gradient appears to be steepest in the lowest income groups (<$20,000/yr), but persists at higher income levels, although it is less pronounced.[5]

Key References:

- Bhatia R, Katz M. Estimation of health benefits from a local living wage ordinance. *Am J Public Health.* 2001; 91(9):1398-1402.

- Cole B, Shimkhada R, Morgenstern M, Kominski G, Fielding J, Wu S. Projected health impact of the Los Angeles City living wage ordinance. *J Epidemiol Community Health.* 2005; 59:645-650.
- Commission on the Social Determinants of Health. *Closing the Gap in a Generation: Health Equity Through Action on the Social Determinants of Health. Final Report of the Commission on Social Determinants of Health.* SI. Geneva: World Health Organization; 2008. Accessed on February 4th 2013.
- Galea S, Tracy M, Hoggatt KJ, DiMaggio C, Karpati A. Estimated deaths attributable to social factors in the United States. *Am J Public Health.* 2011; 101(8):1456-1465.
- Backlund E, Sorlie PD, Johnson NJ. A comparison of the relationships of education and income with mortality: the National Longitudinal Mortality Study. *Soc Sci Med.* 1999; 49:1373-1384.

Detailed Definition:

- The living wage is defined as the hourly wage rate or annual income that a sole provider working full time (2080 hours/yr) must earn to provide his/her family a minimum standard of living, covering costs of food, child care, health insurance, housing, transportation, and other necessities. An annualized budget is calculated for 8 different types of families (1 adult; 1 adult-1 child; 1 adult-2 children; 1 adult-3 children; 2 adults; 2 adults-1 child; 2 adults- 2 children; 2 adults-3 children). The living wage for two family types (an adult with two children and two adults with two children) was applied against the income distribution of single women with children and married couples with children, respectively. The number of families that fell below the living wage, expressed as a percentage of all families of that type, is the indicator.
- Stratification: family type and race/ethnicity
- Interpretation: healthier communities will have lower percentage (or zero) families living below the living wage

Data Description:

- Data sources: Living Wage Calculator (http://livingwage.mit.edu/), American Community Survey (Table B19131)
- Years available: 2010
- Updated: periodically
- Geographies available: cities/towns, county, region (derived), state

Methodology:

The data on living were extracted from the *Living Wage Calculator* website (http://livingwage.mit.edu/) and re-formatted without changes. The data represents a synthesis of multiple data sources including U.S. Department of Agriculture's 2010 low-cost food plan (food costs); Parents and the High Cost of Child Care – 2011 Update, National Association of Child Care Resource and Referral Agencies (childcare); 2010 Consumer Expenditure Survey and the 2010 wave of the Medical Expenditure Panel Survey (healthcare); 2010 Fair Market Rents produced by U.S Department of Housing and Urban Development (housing); 2010 Consumer Expenditure Survey (transportation); and federal payroll taxes as well as federal and state income taxes for the 2011 tax year (taxes).

Income data were tabulated from sequence tables (B19139) of the American Community Survey (ACS), 2006-2010, and stratified by race/ethnicity (county, region, state). Prevailing (median) wages for selected occupations in 2010 were downloaded from the Employment Development Department, labor market information website (www.labormarketinfo.edd.ca.gov) based on the 1st quarter of the Occupational and Employment Statistics (OES) Survey, 2009.

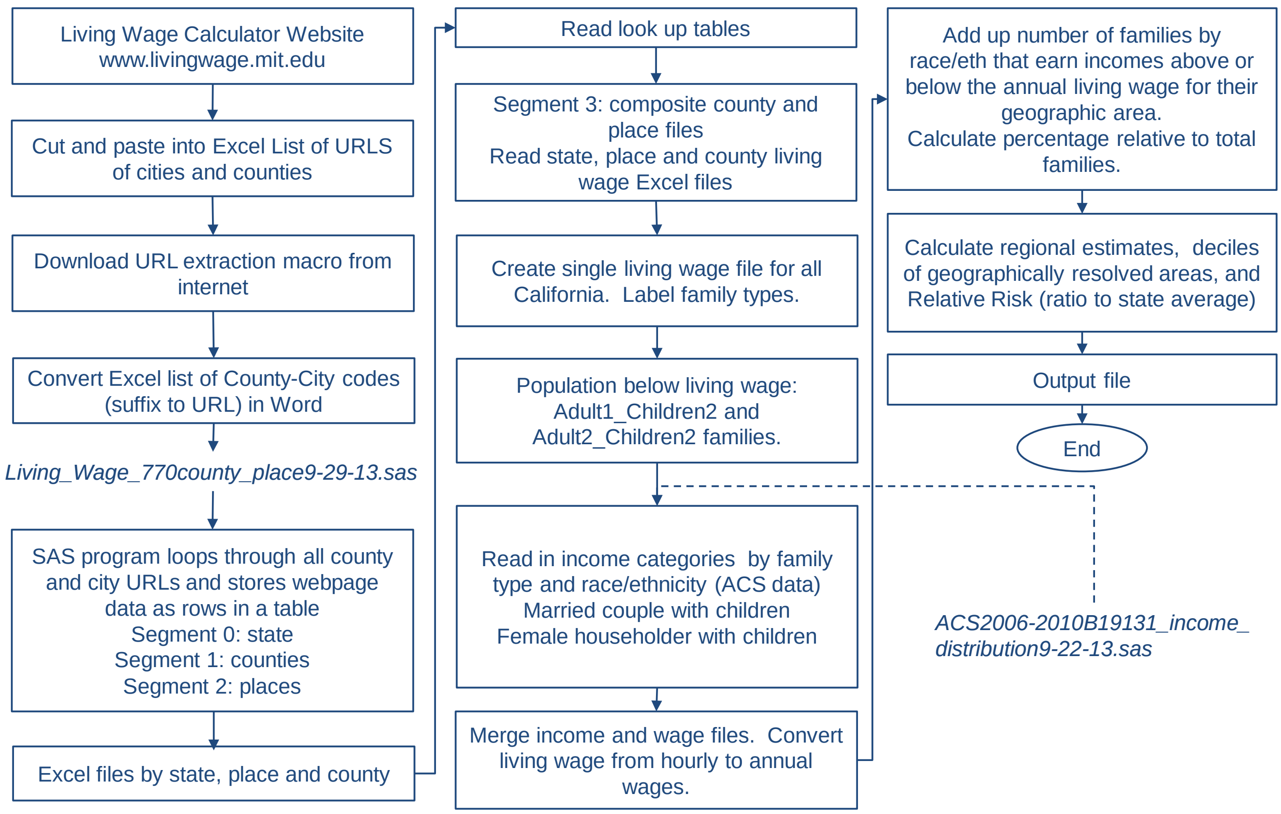
Limitations:

The living wage is a no-frills, minimum standard of living that should be considered a step above the poverty rate and not a lifestyle most middle class Americans would desire. It does not include income set aside for children's post-secondary education, pension, retirement, or savings for wealth accumulation (investments, home ownership). The budget also does not include money for restaurant meals or entertainment, leisure activities, or vacations. Regional cost adjustments were available for some of the cost categories. However, local variation in costs within regions was not accounted for. Family income for married couples can reflect two earners. Standard errors for percent of families below the living wage were not calculable from ACS data.

Data Processing

SAS programs:
ACS2006-2010B19131_income_distribution9-22-13.sas
Living_Wage_770county_place9-29-13.sas

Data Processing: Living Wage



Data Acquisition and Preprocessing

Living Wage

A SAS program iteratively loops through the living wage website (http://livingwage.mit.edu/) and extracts the living wage data specific to a California county or place from individual web pages. The data extracted was living wage by 8 family types (1 Adult; 1 Adult-1 Child; 1 Adult-2 Children; 1 Adult-3 Children; 2 Adults; 2 Adults-1 Child; 2 Adults- 2 Children; 2 Adults-3 Children.) The arguments for the webpages are the URLs (http://livingwage.mit.edu/counties/xxxxx) of the 5-digit FIPS code of a county (e.g. 06001 for Alameda Co.) and 10-digit county-census place codes (http://livingwage.mit.edu/placeses/yyyyyyyyyy) [e.g. 0600153000 for Oakland]. The Living Wage calculator reports hourly living wages; these were converted to annual wages (*2080)

Table 17. Example of Types of Wages by Family Type that can be obtained from the Living Wage website

	1 Adult	1 Adult, 1 Child	1 Adult, 2 Children	1 Adult, 3 Children	2 Adults	2 Adults, 1 Child	2 Adults, 2 Children	2 Adults, 3 Children
Living Wage	$11.20	$22.70	$26.33	$33.28	$16.73	$20.80	$22.15	$27.29
Poverty Wage	$5.21	$7.00	$8.80	$10.60	$7.00	$8.80	$10.60	$12.40
Minimum Wage	$8.00	$8.00	$8.00	$8.00	$8.00	$8.00	$8.00	$8.00

The xxxxx and yyyyyyyyyy suffixes were obtained by cut-and-pasting the web page of the cities into an Excel spreadsheet, and then using a Visual Basic macro (downloaded from the internet) to extract the URL from the hypertexted list of 58 counties and 1082 cities. The Excel column of codes was pasted into a Word file and the paragraph marks were stripped using a global search-and-replace command using ^p as the character for find and space " " as its replacement. The long continuous list of 5 or 10 digit numbers were pasted into the SAS program as the macro loop list argument.

There were 37 CDPs from 2000 that were discontinued in 2010, and these are excluded from the output. 2010 CDPs for which no data were available appear in the output file as missing data.

- 2000 Census Place code — Place Name
- 02132 — Angels City city
- 04503 — Bayview-Montalvin CDP
- 04541 — Baywood-Los Osos CDP
- 06933 — Blackhawk-Camino Tassajara CDP
- 08848 — Buena Vista CDP
- 10032 — Calwa CDP
- 17030 — Crescent City North CDP
- 20536 — East Blythe CDP
- 20550 — East Compton CDP
- 24730 — Foothill Ranch CDP
- 29399 — Gerber-Las Flores CDP
- 31375 — Groveland-Big Oak Flat CDP
- 34407 — Homewood Canyon-Valley Wells CDP
- 39173 — Laguna CDP
- 39257 — Laguna West-Lakeside CDP
- 39521 — Lakehead-Lakeshore CDP
- 49194 — Moraga town
- 50090 — Murrieta Hot Springs CDP
- 50706 — Nebo Center CDP
- 51186 — Newport Coast CDP
- 52452 — North Woodbridge CDP
- 54358 — Orosi CDP
- 55837 — Parkway-South Sacramento CDP
- 56871 — Phoenix Lake-Cedar Ridge CDP
- 58356 — Portola Hills CDP
- 58982 — Quail Valley CDP
- 67143 — San Joaquin Hills CDP
- 70784 — Sedco Hills CDP
- 71074 — Seven Trees CDP
- 73461 — South Woodbridge CDP
- 73472 — South Yuba City CDP
- 75826 — Sun City CDP
- 77049 — Sunol-Midtown CDP
- 78680 — Tierra Buena CDP
- 80868 — Tustin Foothills CDP

- 81008 Twentynine Palms Base CDP
- 84186 West Compton CDP

Income distribution by family type

Information on income distribution by family type and by race/ethnicity for place, counties, and the state of California were obtained from Table B19131 (Family Type By Presence Of Own Children Under 18 Years By Family Income In The Past 12 Months (In 2010 Inflation-Adjusted Dollars), Universe: Families) from the 2006-2010 American Community Survey (ACS) Selected Population Tables (SPT). A SAS data extraction program was used to obtain Table B19131 from the Summary Files (see Toolbox Obtaining American Community Survey Data from Summary Files). The summary files for the 2006-2010 Special Population Tables for California were downloaded from the U.S. Census FTP site (http://www2.census.gov/acs2010_SPT_AIAN/SelectedPopulationTables/California/ca.acs2010.zip).

The income categories by family type reported by the ACS are:

- $10,000 to $14,999
- $15,000 to $19,999
- $20,000 to $24,999
- $25,000 to $29,999
- $30,000 to $34,999
- $35,000 to $39,999
- $40,000 to $44,999
- $45,000 to $49,999
- $50,000 to $59,999
- $60,000 to $74,999
- $75,000 to $99,999
- $100,000 to $124,999
- $125,000 to $149,999
- $150,000 to $199,999
- $200,000 or more

Family types include:

- Married-couple family, with own children under 18 years
- Married-couple family, no own children
- Male householder, no wife present, with own children under 18 years
- Male householder, no wife present, no own children
- Female householder, no husband present, with own children under 18 years
- Female householder, no husband present, no own children

Each geography (PL, CO) was assigned to an ACS income category based on its annual living wage:
if annual living wage <= $10,000 AND annual living wage ≠ 0 then category = 2;
if annual living wage >= $10,000 AND annual living wage < $15,000 then category = 3;
if annual living wage >= $15,000 AND annual living wage < $20,000 then category = 4;
etc.

The cumulative number of families whose income category is lower than the income category for the geographic area was calculated. This cumulative number of families was the numerator of the indicator. The denominator was the total number of families of each type on each of the geographies.

A sensitivity analysis was conducted to examine how the indicator results would vary using the living wage for families of “2 Adults- 2 Children” and living wage of the average for families 2 Adults-1 Child, 2 Adults- 2 Children, and 2 Adults-3 Children. The analysis suggested that the different thresholds produce small differences in results and the slightly more conservative results (based on 2 children) were used as the threshold for data distribution.

Proximity to Alcohol Outlets

Definition

Full Title: Percent of the population within ¼ mile of alcohol outlets by type of establishment sales

Indicator ID: 774

Healthy Community Framework: Social relationships that are supportive and respectful

Aspirational Goal: Safe communities, free of crime and violence

Description of significance and health connection:

Excessive alcohol consumption caused approximately 88,000 deaths and 2.5 million years of potential life lost annually in the U.S. during 2006-2010, making it the fourth leading preventable cause of death. Evidence shows that high density and proximity to alcohol outlets in neighborhoods is associated with higher rates of binge drinking and associated harms, like drinking and driving, motor vehicle-related pedestrian injuries, child abuse and neglect, youth drinking, intimate partner violence, and violent crime. In California, the rate of alcohol-attributable deaths (ADD/year/100,000 population, 2006-2010) is higher for males (43.9) and African Americans (36.6) in comparison with the total population (29.4). Low-income and minority neighborhoods are more likely to have higher concentrations of stores selling alcohol.

Alcohol outlet density is controlled by the states and local regulations. In California the number of on-sale and off-sale alcohol licenses at the county level is restricted based upon the ratio of number of current licenses to the population within each Census tract. Additional licenses may be allowed based on a showing of public convenience or necessity. Limiting alcohol outlet density through the use of regulatory authority (e.g., licensing and zoning) is a public health strategy to prevent deaths and harms associated with excessive alcohol consumption.

Summary of evidence:

Multiple studies provide empirical evidence that higher alcohol outlet density and closer proximity to alcohol outlets is positively associated with outcomes like excessive alcohol consumption and other alcohol related harms like injuries and violence. However, some studies have found variations in the patterns; for example, four California cities showed higher rates of heavy drinking in high income neighborhoods with low alcohol outlet density than in lower income neighborhoods.

Key References:

- Gonzalez K, Roeber J, Kanny D, et al. Alcohol-attributable deaths and years of potential life lost - 11 States, 2006–2010. *MMWR* 2014; 63(10): 213-216.
- Michigan Department of Community Health. Bureau of Disease Control, Prevention & Epidemiology. The Association of Increased Alcohol Outlet Density & Related Harms: Summary of Key Literature. Accessed May 19th, 2014.
- California Department of Alcohol and Drug Programs. Racial/Ethnic Disparities – A data informed perspective, 2013. Accessed May 19th, 2014.

- ABC Act (Business and Professions Code Sections 23815-23827). Accessed May 19th, 2014.
- Community Anti-Drug Coalitions of America, Center on Alcohol Marketing and Youth, Johns Hopkins Bloomberg School of Public Health. Regulating Alcohol Outlet Density: an Action Guide. Accessed May 19th, 2014.
- Task Force on Community Preventive Services. Preventing Excessive Alcohol Consumption: Regulation of Alcohol Outlet Density. Accessed June 15th, 2014.
- Popova S, Giesbrecht N, Bekmuradov D, Patra J. Hours and days of sale and density of alcohol outlets: impacts on alcohol consumption and damage: a systematic review. *Alcohol Alcsm* 2009; 44(5): 500–516.
- Campbell CA, Hahn RH, Elder R, et al. The effectiveness of limiting alcohol outlet density as a means of reducing excessive alcohol consumption and alcohol-related harms. *Am J Prev Med* 2009; 37(6): 556-569.
- Truong KA, Sturm R. Alcohol environments and disparities in exposure associated with adolescent drinking in California. *Am J Pub Health* 2009; 99(2): 264-270.
- Pollack CE, Cubbin C, Ahn D, Winkleby M. Neighbourhood deprivation and alcohol consumption: does the availability of alcohol play a role? *Int J Epidemiol* 2005; 34:772–780

Detailed Definition:

- Numerator: population living within ¼ of a mile of an alcohol outlet
- Denominator: total population
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: eight race/ethnicity groups (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white). Three types of establishments or licenses (on-sale, off-sale, total).
- Interpretation: healthier communities will have a lower percentage of the population living within ¼ of a mile of an alcohol outlet.

Data Description:

- Data sources: California Department of Alcohol and Beverage Control (ABC), raw data April 2014 (http://www.abc.ca.gov/datport/ABC_Data_Export.zip). California Department of Public Health, California Environmental Health Tracking Program (CEHTP), Browser Based Geocoder (https://www.ehib.org/page.jsp?page_key=357). Department of Finance (DOF), Demographic Research Unit, Redistricting files, 2010 (http://www.dof.ca.gov/research/demographic/state_census_data_center/census_2010/#PL94).
- Years available: 2014
- Updated: ABC raw data is updated weekly; block population data is available every 10 years
- Geographies available: Census tracts, cities/towns, counties, county divisions, regions and state.

Methodology:

Part 1: The addresses of all establishments with active off-sale (ABC types 20 and 21) and on-sale licenses (ABC types 40, 41, 42, 47 and 48) (cross-sectional April 23rd 2014) were obtained from the raw data file. Using the Browser Based Geocoder (CEHTP), 97.7% of the 67,617 alcohol outlets with active licenses were successfully geocoded. Geocoded addresses were imported into ArcMap and displayed as X, Y points. Buffers of ¼ mile around each point were created for establishments with off-sale and on-sale alcohol licenses separately, and later for establishments with one or both types of licenses (total). Census blocks centroids (2010) were spatially joined to the buffers for each type of establishment. <u>Part 2</u>: The Census blocks within buffers were imported into SAS and merged with total population (2010) counts by block. The number of people that live within ¼ mile of alcohol outlets by race/ethnicity and by type of establishment sales (numerator) was calculated for each Census block and later aggregated to Census tracts, cities/towns, counties, county divisions, regions and state levels. The denominator was the total number of people on each geographic level. Standard errors, relative standard errors, and 95% upper and lower confidence intervals were calculated. Decile rankings of places and relative risk in relation to state average were calculated. Regions were based on counties of metropolitan transportation organizations (MPO) regions as reported in the 2010 California Regional Progress Report (Appendix C).

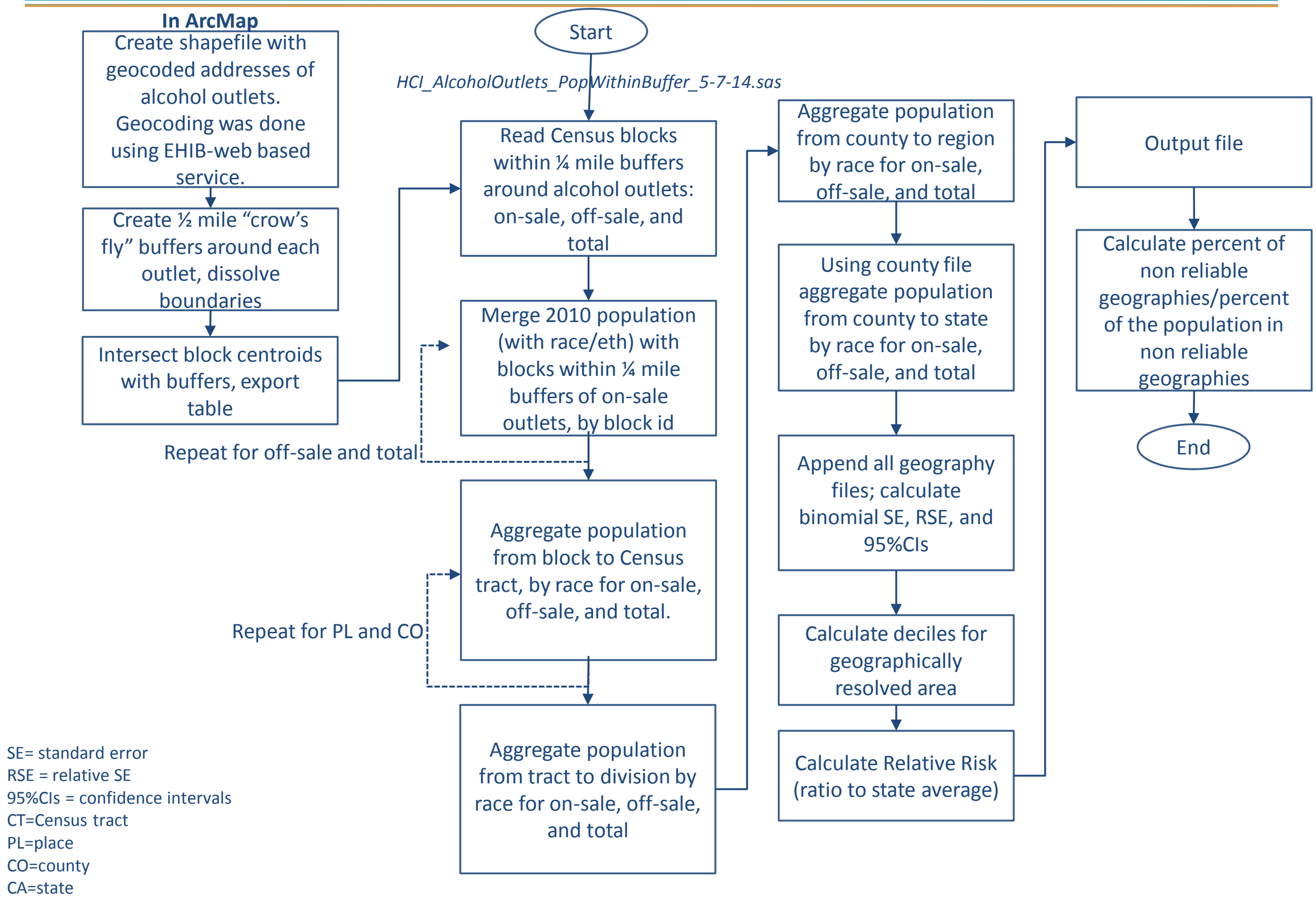
Limitations:

The indicator was constructed with ABC alcohol outlet data from April 23rd 2014 and population data from 2010. Updates of the indicator should use ABC data from the same date; alternatively a new version could be created using data from the end of the calendar or fiscal year. Because proximity to outlets is only an indicator of an increased likelihood of alcohol related problems, it is not a direct measure of alcohol related problems.

<u>Data Processing</u>

SAS program: HCI_AlcoholOutlets_PopWithinBuffer_5-7-14.sas

Data Processing: Proximity to Alcohol Outlets

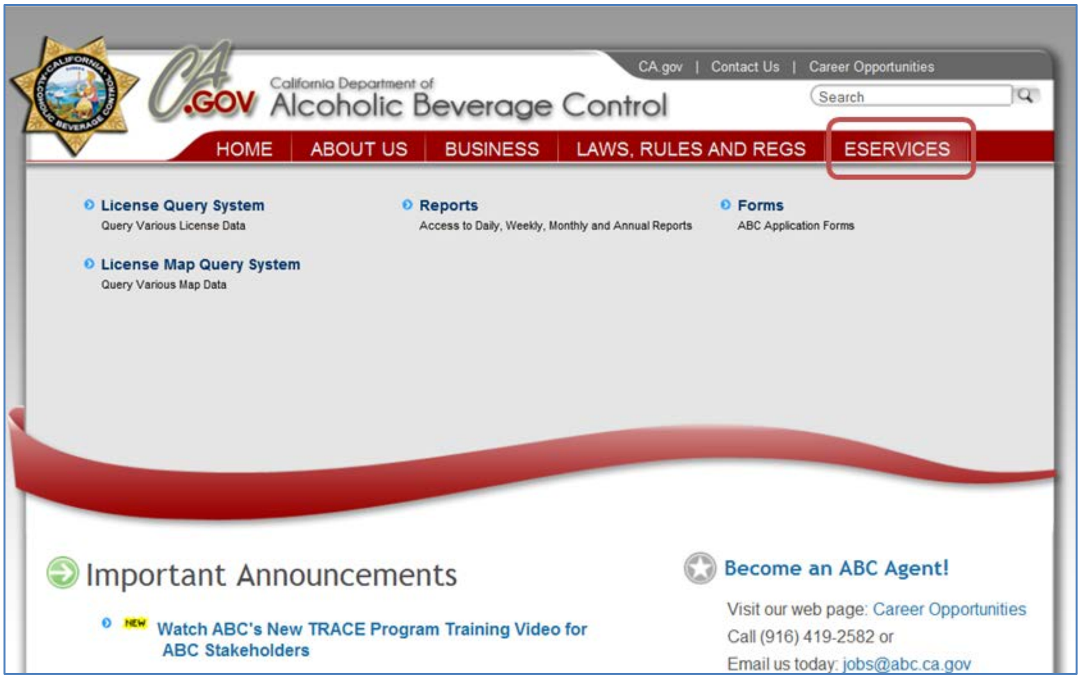


<u>Data Acquisition and Preprocessing</u>

Raw Data File from California Department of Alcohol and Beverage Control (ABC)

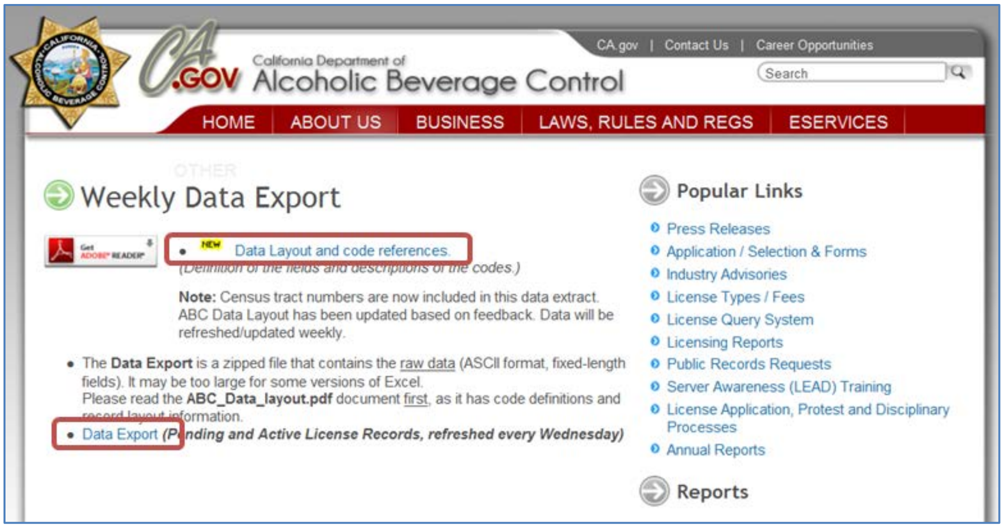
The raw data file contains addresses and license numbers of all outlets with active and pending licenses by license type. It is updated weekly. The raw data file is in SAS format and each variable has a fixed number of characters. The companion data layout and codes reference explains the variables in the raw data file and their number of characters.

1. Go to the ABC homepage (http://www.abc.ca.gov/) and click on "ESERVICES".



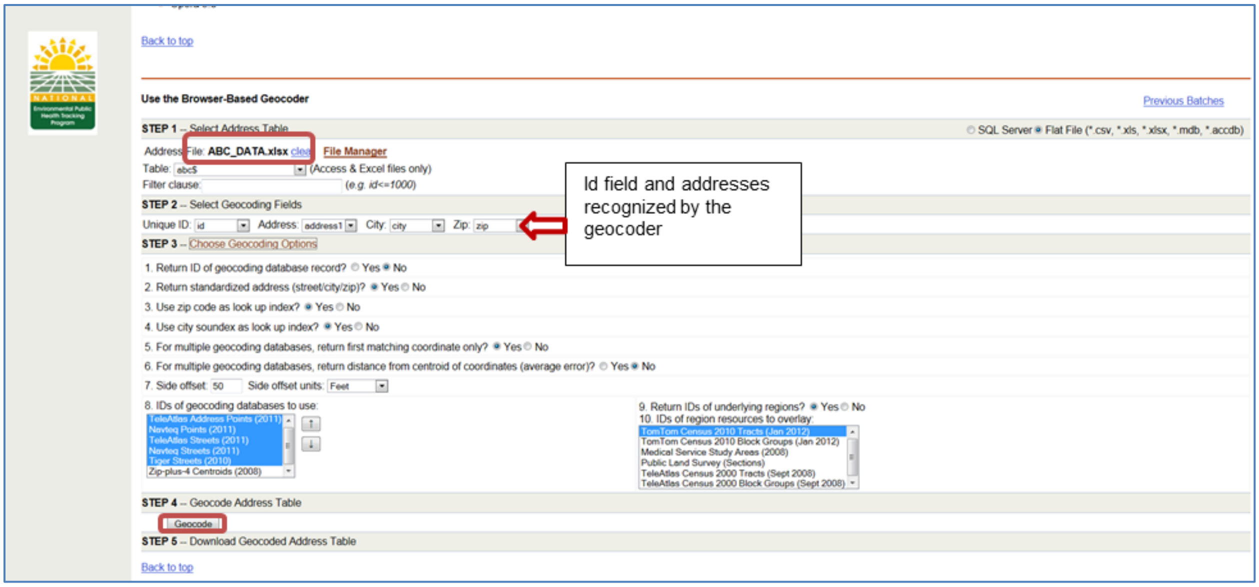
2. Under E-Services click on Raw Data Export (Weekly).

3. Download the raw data file (click on “Data Export”) and the companion “Data layout and code references.”



Selection of On-Sale and Off-Sale License Types and Geocoding of ABC Raw Data

1. Raw data was imported into SAS using the infile statement (HCI_AlcoholOutlets_774_GeocodingData-5-6-14.sas). A reduced file was created retaining only active licenses of the following types:
   a. TYPE 20 Off-Sale Beer and Wine
   b. TYPE 21 Off-Sale General
   c. TYPE 40 On-Sale Beer
   d. TYPE 41 On-Sale Beer and Wine - Eating Place
   e. TYPE 42 On-Sale Beer and Wine - Public Premises
   f. TYPE 47 On-Sale General Eating Place
   g. TYPE 48 On-Sale General Public Premises.
2. The reduced file contained the following variables that identify the addresses of alcohol outlets: filenum, address1, city, zip. The reduced file was exported as an Excel file.
3. A new “id” column was added to the Excel file numbering all addresses from 1 to n. Adding this “id” is important for the geocoding process to work.
4. An account was created at the CDPH-Environmental Health Tracking Program Geocoding Service: http://www.ehib.org/page.jsp?page_key=137. The Browser Based Geocoder was used to geocode the alcohol outlets addresses contained in the raw data file from ABC. The default geocoding options were used.



5. The geocoder produced a .csv file with the following variables:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	filenum	id	address1	city	zip	StdStreet	StdCity	StdZip	StdError	Status	Score	Side	X	Y	RegionIDs	MetadataID	
2	529465	1	3 JUSTIN C	NAPA	94558-27	3 Justin Ct	Napa	94558-2749		M	100		-122.322	38.3229	6.1E+09	1-teleatlas_points_zi	
3	453020	2	777 CASIN	BLUE LAKE	95525	777 CASIN	BLUE LAKE	95525	3.1	M	100	L	-123.999	40.8852	6E+09	3-navteq_streets_zip	
4	521761	3	31 DAVEN	DAVENPO	95017	31 DAVEN	DAVENPO	95017	3.1	M	100		-122.194	37.0111	6.1E+09	1-teleatlas_points_zi	

6. The variables include standardized address variables (Std prefix), geocoding error codes, status, score, X,Y coordinates for the address and other information. The results of the geocoder were summarized and 97.70% of all addressed were matched by the geocoder. Of those, 73.52% were matched with a high score and no errors. All of the matched addresses (regardless of score or errors) were used to construct the indicator, given that none of the errors were considered highly problematic.

Geocoding results, EHIB Browser Based Geocoder
ABC data/ 5-2-14/Dulce Busatamante

Status	Number	Percent
U=Unmatched	1511	2.23
T=Tied	44	0.07
M=Matched	66062	97.70
Total	67617	

M=Matched

Score	Error term	Number	Percent
>=80 (99.68)	yes	6636	10.05
>=80 (98.60)	no	48569	73.52
<80 (63.32)	yes	1461	2.21
<80 (63.38)	no	9396	14.22
Total		66062	

Score is an indicator of how well the address matched the database

The lower the score the higher the probaility that an address was matched to an incorrect location

M=Matched, score >=80 with errors

Error code	Error term	Number	Percent
14.3	+4 changed	2910	43.85
3.1	Street not found	810	12.21
4.1	Address not found	779	11.74
11.1	Street phonetic match used	397	5.98
11.9	Street Standardized	394	5.94
14.2	ZIP changed	287	4.32
10.2	City determined from ZIP	262	3.95
11.2	Predirection dropped	154	2.32
12.1	Unit Standardized	153	2.31
11.3	Predirection added	151	2.28
13.1	Leftovers found	90	1.36
6.1	Multiple streets match	52	0.78
11.5	Suffix added	51	0.77
11.4	Suffix dropped	37	0.56
5.2	+4 unavailable	33	0.50
10.3	Acceptable city name used	15	0.23
11.8	Predirection standardized	13	0.20
11.6	Postdirection dropped	12	0.18
11.11	Postdirection standardized	10	0.15
11.7	Postdirection added	10	0.15
2.1	City not found	8	0.12
12.3	Unit not found	5	0.08
6.2	Multiple addresses match	1	0.02
10.1	City phonetic match used	1	0.02
12.4	Unit unverified	1	0.02
		6636	

M=Matched, score <80 with errors

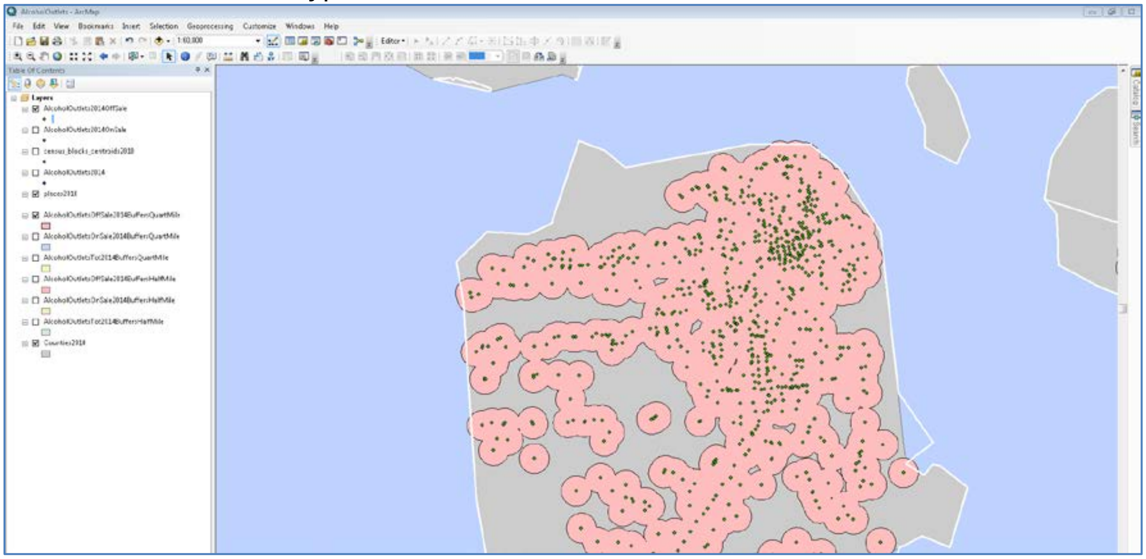
Error code	Error term	Number	Percent
14.3	+4 changed	596	40.71
4.1	Address not found	239	16.33
3.1	Street not found	211	14.41
11.9	Street Standardized	87	5.94
11.1	Street phonetic match used	65	4.44
14.2	ZIP changed	60	4.10
10.2	City determined from ZIP	45	3.07
11.2	Predirection dropped	39	2.66
11.3	Predirection added	28	1.91
12.1	Unit Standardized	26	1.78
13.1	Leftovers found	20	1.37
11.5	Suffix added	18	1.23
10.3	Acceptable city name used	7	0.48
6.1	Multiple streets match	6	0.41
5.2	+4 unavailable	5	0.34
11.4	Suffix dropped	4	0.27
2.1	City not found	3	0.20
11.8	Predirection standardized	3	0.20
6.2	Multiple addresses match	1	0.07
11.6	Postdirection dropped	1	0.07
		1464	

Census Blocks Centroids within ¼ Mile Buffers around Outlets

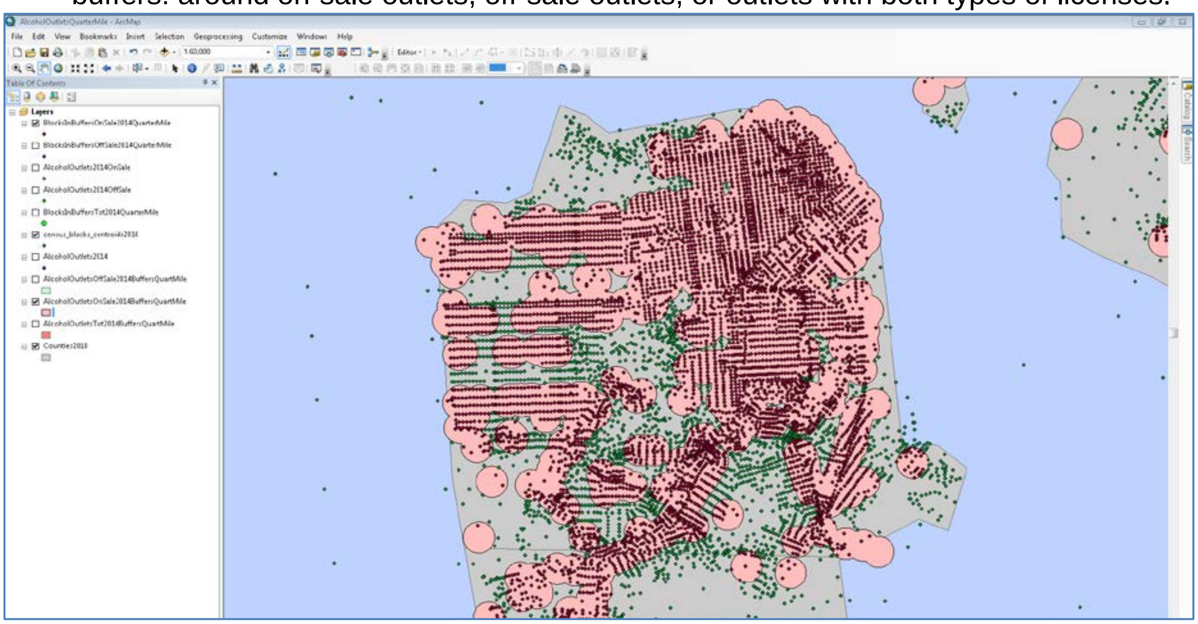
1. The .cvs output file from the geocoder was saved as an Excel file and imported into ArcMap. Only the matched (M) addresses were used. The X,Y points were displayed using the "Display X Y data option" (see Toolbox Displaying XY Data in ArcMap). The GCW_WGS_1984 geographic coordinate system was defined for the data points (this is the system used by the Browser Based geocoder).

2. The X,Y points were reprojected from GCW_WGS_1984 geographic coordinate system to NAD 1983 California Teal Albers (Meters) (see Toolbox Downloading census TIGER shapefiles and reprojecting in California Teale Alberts coordinate system).

3. One quarter mile buffers were created around each alcohol outlet X,Y point (dissolve type output, see Toolbox Creating Buffers around Features and Select by Location in ArcMap). Three types of buffers were created: around on-sale outlets, off-sale outlets, or outlets with both types of licenses.



4. Using a shapefile of the Census block centroids (2010) for California, the Census blocks centroids were spatially joined to the buffer: all centroids that fall within the buffers (see Identifying Census blocks centroids within (CWS) boundaries using spatial join under the Unsafe Drinking Water indicator). The spatial join was conducted for the three types of buffers: around on-sale outlets, off-sale outlets, or outlets with both types of licenses.



5. The Census blocks within buffers were exported as sas7bdat files to be used as raw data to construct the indicator: blocksinbuffersoffsale2014quart, blocksinbuffersonsale2014quart, and blocksinbufferstotal2014quart.

Population Data at the Census Block Level, 2010

Population at the Census Block level with race/Hispanic origin information was obtained from the Department of Finance, Demographic Research Unit, 2010 redistricting data from the U.S. Census Bureau (Public Law 94-171) for California. (http://www.dof.ca.gov/research/demographic/state_census_data_center/census_2010/#PL94, http://www2.census.gov/census_2010/01-Redistricting_File--PL_94-171/).

Fine Particulate Air Pollution

Definition

Full Title: Annual Mean Ambient Concentration of Fine Particulate Matter ($PM_{2.5}$)

Indicator ID: 776

Healthy Community Framework: Quality and sustainability of environment

Aspirational Goal: Minimize toxics, greenhouse gas emissions, and waste

Description of significance and health connection:

Clean air is a fundamental building block of human health. Air pollution from fixed and mobile sources (e.g. factories and cars, respectively) is a complex mixture of gases, fumes, and particles released into the atmosphere from the combustion of fossil fuels and evaporation of solvents. Ozone that forms at the ground-level (described elsewhere) and fine particulate matter (PM) are two indicators of air pollution that are linked to short- and long-term adverse health effects. Particulate matter that has an aerodynamic diameter of 2.5 microns or less is called $PM_{2.5}$ and is capable of reaching deep into the lungs causing a host of diseases including lung cancer, heart disease, respiratory disease, and acute respiratory infections, particularly in children. In California, the Air Resources Board estimated that, given the $PM_{2.5}$ levels between 2004 and 2006, over 9300 deaths could be prevented each year if California met its current statewide $PM_{2.5}$ standard of 12 $\mu g/m^3$.

Summary of evidence:

Based on numerous community-based epidemiologic studies, both short-term and long-term exposures to $PM_{2.5}$ increase the risk of cardiovascular disease and mortality, and are linked to adverse respiratory outcomes such as chronic obstructive lung disease, hospital and emergency department admissions for asthma, increased respiratory symptoms, altered pulmonary function, and pulmonary inflammation among asthmatic children. While not definitive, evidence is accumulating for $PM_{2.5}$ effects on low birth weight and infant mortality, especially due to respiratory causes during the post-neonatal period.

Key References:

- Tran HT, Alvarado A, Garcia C, Motallebi N, Miyasato L, Vance W. Methodology for Estimating Premature Deaths Associated with Long-term Exposures to Fine Airborne Particulate Matter in California (Draft: Staff Report). Sacramento, CA: California Air Resources Board; 2009. www.arb.ca.gov/research/health/pm-mort/pm-mort_final.pdf. Accessed on 8/16/12.
- Drechsler D, Garcia C, Tran H, Mehadi A, Nystrom M, Propper R, et al. Review of the California Ambient Air Quality Standard For Ozone. Vol 4. Table B-5: California Annual Health Impacts of Current Ozone Concentrations Compared to the State 8-hour Ozone Standard of 0.070 ppm.. Sacramento, CA: California Environmental Protection Agency, Air Resources Board; 2005. http://www.arb.ca.gov/carbis/research/aaqs/ozone-rs/rev-staff/vol4.pdf. Accessed on January 4, 2013.

- Particulate Matter Integrated Science Assessment Project Team. Integrated Science Assessment for Particulate Matter. Research Triangle Park, NC: U.S. Environmental Protection Agency; 2009. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=216546#Download. Accessed on January 4, 2013.

Detailed definition:

- Three-year annual mean concentration of $PM_{2.5}$ (average of quarterly means, 2007-2009)
- Stratification: eight race/ethnicity groups (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white).
- Interpretation: healthier communities will have lower averages of fine particulate air pollution.

Data Description:

- Data source: Air Monitoring Network, Air Resources Board (CARB) http://www.arb.ca.gov/aqmis2/aqmis2.php
- Years available: 3-year average of 2007-2009
- Updated: not stated
- Geographies available: zip codes, and population weighted average for place, county, region, California

Methodology:

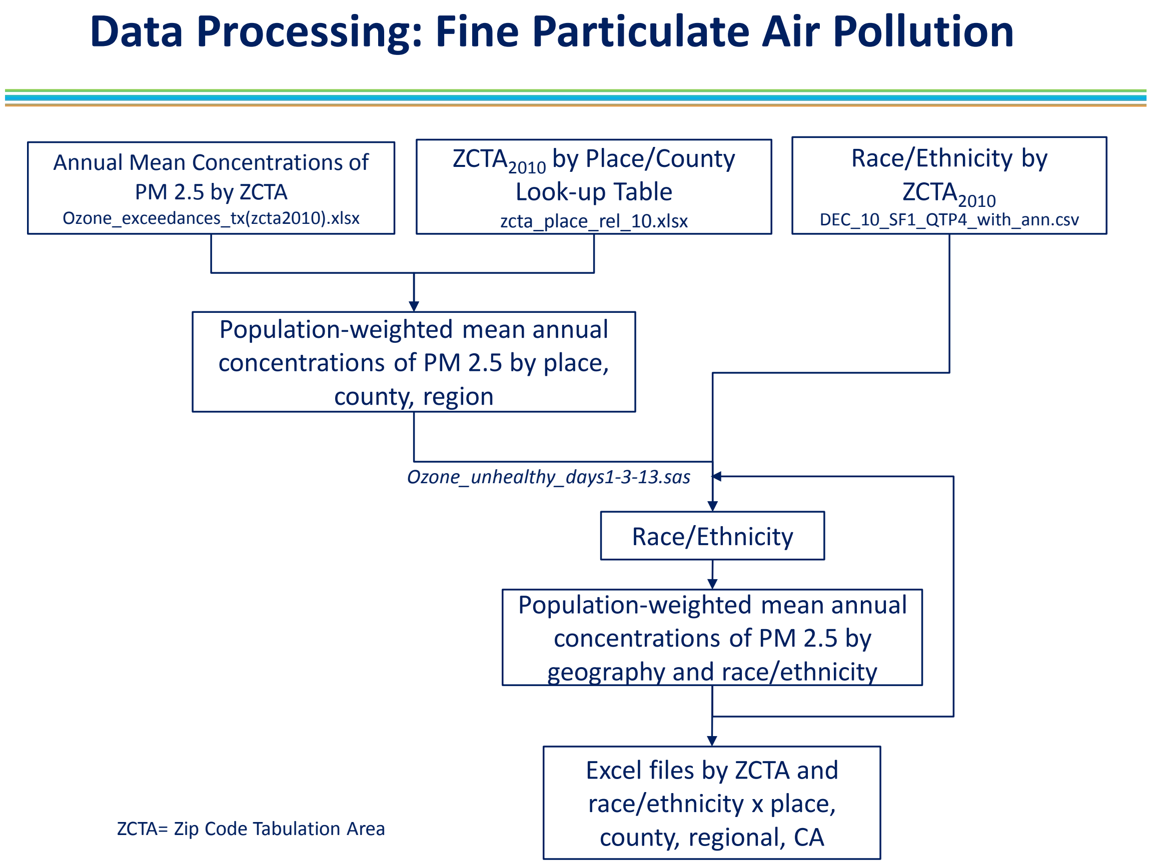
A network of approximately 300 air monitoring stations is maintained in California by the California Air Resources Board, local air pollution control districts, tribes and federal land managers. Of these, approximately 80 measure $PM_{2.5}$ concentrations. A statistical average of 12 quarterly means (2007 to 2009) was calculated for each zip code based on these monitoring stations. To calculate averages for cities, places, and regions, the mean $PM_{2.5}$ concentration by zip code was weighted by the population in the zip code based on U.S. 2010 Census files that parsed zip codes by place and county. To calculate race/ethnicity specific weighted averages, the proportion of race/ethnicities by zip were calculated from U.S. 2010 Census files. Regional estimates were based on county groupings associated with California metropolitan planning organizations as reported in the 2010 California Regional Progress Report (Appendix C).

Limitations:

Geographic coverage was not complete because of the limited number and geographic extent of monitoring stations. The uncertainty of the interpolated values increases with distance from the nearest monitor. According to the Air Resources Board, values for areas greater than 50 km from the nearest monitor are very imprecise, and should be regarded as speculative. They are included for the sake of completeness, but should not be relied upon. Even within populated areas, monitoring stations are often located in areas that cannot detect highly localized areas of pollution that significant numbers or sensitive subgroups (daycare centers, schools, hospitals, etc.) in the population may encounter. Data were not available to present standard errors.

Data Processing

SAS program: PM25levels4-14-13.sas



Data Acquisition

Zip code specific fine particulate air pollution unhealthy healthy days data were acquired from Laura August (Laura.August@oehha.ca.gov), Office of Environmental Health Hazard Assessment (OEHHA) on 11/26/2012.

Percent of Workers' Daily Commute 10 or More Minutes by Walking or Biking

Definition

Full Title: Percent of population aged 16 years or older whose commute to work is 10 or more minutes/day by walking or biking

Indicator ID: 778

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Safe, sustainable, accessible and affordable transportation options

Description of significance and health connection:

Active modes of transport, bicycling and walking alone and in combination with public transit, offer opportunities to incorporate physical activity into the daily routine. Physical activity is associated with lowering rates of heart disease and stroke, diabetes, colon and breast cancer, dementia and depression. Automobile commuting is associated with health hazards, such as air pollution, motor vehicle crashes, pedestrian injuries and fatalities, and sedentary lifestyles. Consequently the transition from automobile-focused transport to public and active transport offers environmental health benefits, including reductions in air pollution, greenhouse gases and noise pollution, and may lead to greater overall safety in transportation.

People who walk to public transit tend to be non-White, from lower income households, and to live in large urban areas with access to rail and bus systems. Car ownership, race, gender, and built environment features, including safety infrastructure, affect modal choices regarding active transport. In some areas, college education is associated with greater use of non-motorized modes of transportation. Risks of injury in traffic collisions are higher for pedestrians and bicyclists, but lowest for bus and rail passengers.

Summary of evidence:

There are dozens of longitudinal epidemiologic studies that have documented improved health outcomes with increasing physical activity, including bicycling and walking. Active transportation can contribute to the U.S. Surgeon General's recommended physical activity goals for adults of at least 120 to 150 minutes per week (17-22 minutes per day) of moderate-to-vigorous activity, which lowers the risk of early death, heart disease, high blood pressure, diabetes, stroke, colon cancer, breast cancer, depression, cognitive decline, and osteoporosis.

Key References:

- de Nazelle A, Nieuwenhuijsen MJ, Antó JM, et al. Improving health through policies that promote active travel: a review of evidence to support integrated health impact assessment. *Environ Int*. 2011; 37(4):766-77.
- Cavill N, Kahlmeier S, Rutter H, Racioppi F, Oja P. *Economic Assessment of Transport Infrastructure and Policies: Methodological Guidance on the Economic Appraisal of Health Effects Related to Walking and Cycling*. World Health Organization Regional Office for Europe. 2007. Accessed October 25th, 2013.

- Plaut PO. Non-motorized commuting in the US. *Transport Res D-TR E*. 2005; 10(5): 347-356.
- Physical Activity Guidelines Advisory Committee. *Physical Activity Guidelines Advisory Committee Report*. Washington, DC: U.S. Department of Health and Human Services. 2008. Accessed October 25th, 2013.
- Pucher J, Buehler R, Bassett DR, Dannenberg AL. Walking and cycling to health: a comparative analysis of city, state, and international data. *Am J Public Health*. 2010; 100(10):1986-1992.
- Freeland AL, Banerjee SN, Dannenberg AL, Wendel AM. Walking associated with public transit: moving toward increased physical activity in the United States. *Am J Public Health*. 2013; 103(3): 536-542.

Detailed Definition:

- Numerator: population aged 16 years or older whose commute to work is 10 or minutes /day by walking or biking
- Denominator: population aged 16 year or older who are employed and work away from home
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: eight race/ethnicity groups (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white). Two modes of transportation (walking, biking).
- Interpretation: healthier communities will have higher percentages of workers walking/biking.

Data Description:

- Data sources: U.S. Census Bureau, American Community Survey (ACS): 2005-2007, 2008-2010, and 2007-2011 (http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml). U.S. Department of Transportation, Federal Highway Administration, National Household Travel Survey (NHTS): 2001 and 2009 (http://nhts.ornl.gov). All data accessed 9/2013.
- Years available: 2001, 2005-2007, 2008-2010, 2009, 2007-2011.
- Updated: 1, 3, 5, 10 year intervals
- Geographies available: cities/towns, counties, regions (derived), consolidated metropolitan statistical area (CMSA) and state

Methodology:

Data on walking to work by travel time (one way) for the population aged 16 years or older was obtained from the ACS, table B08134. The table provides nine travel time groupings (less than 10 min, 10-14 min, 15-19 min, 20-24 min, 25-29 min, 30-34 min, 35-44 min, 45-59 min, 60 or more min). The estimated number of people who walk 10 min or more was calculated as the difference between the total numbers of people who walked to work minus those that walked less than 10 min. The standard error of the difference was estimated with the formula: $\sqrt{SE(no.of\ people\ who\ walk\ to\ work)^2 + SE(no.of\ people\ who\ walk\ less\ than\ 10\ min\ to\ work)^2}$. The percent was calculated as the number of workers who walk 10 min or more over the total number of workers ($\times$100). Its standard error was calculated using the approximate method. Data on biking to work by travel time (one way) was obtained from the NHTS person file. To make population and bicycling definitions comparable to the ACS, NHTS survey participants were subset for people 16 year or older who had been employed and traveled to work during the reference week of the survey. Excluded were people that had been employed but worked from home once a week or more (NHTS 2001), or people who worked from home ten days or more per month (NHTS 2009). The estimated number of people who bike 10 minutes or more, the estimated number of people who biked to work, and the percent of people who biked 10 min or more to work and its standard error were calculated using the survey sample weights. Relative standard errors, 95% confidence intervals, and decile ranking of places were calculated. Regions were based on counties of metropolitan transportation organizations (MPO) as reported in the 2010 California Regional Progress Report (Appendix C).

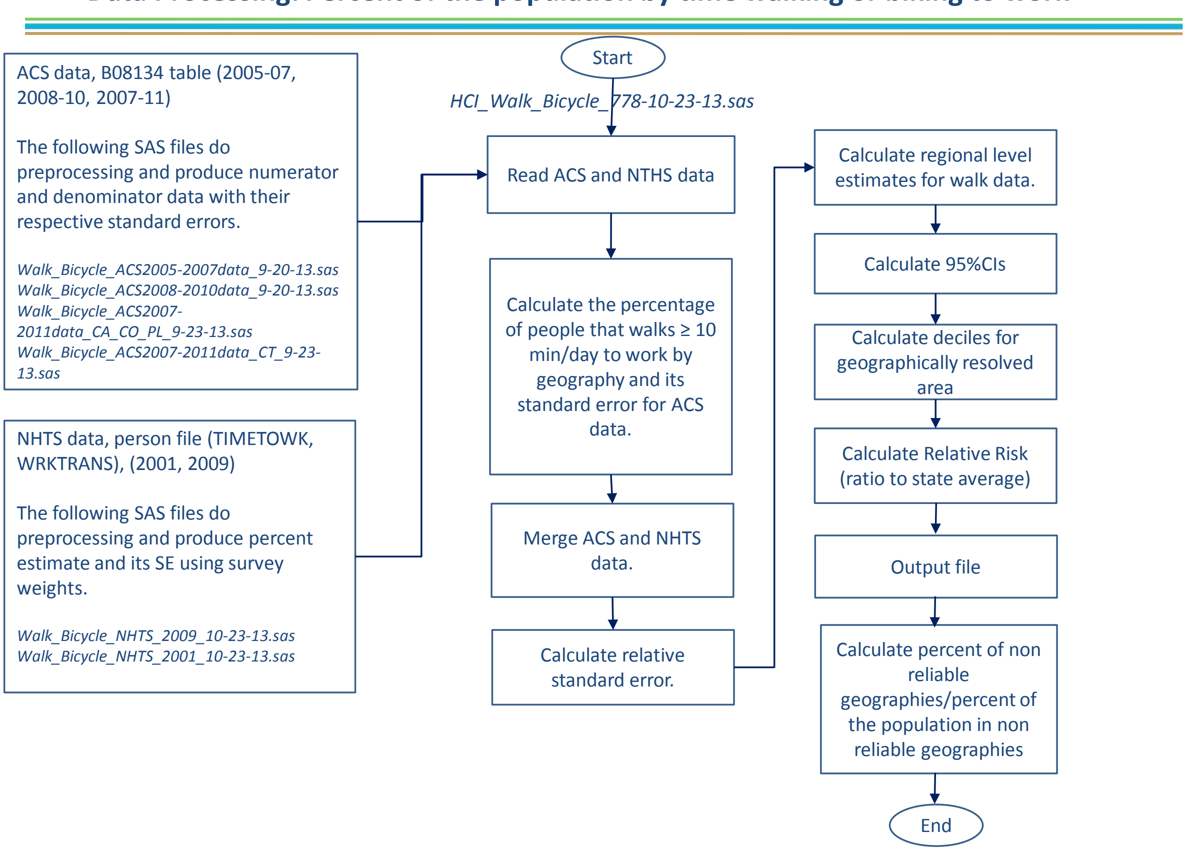
Limitations:

The indicator is limited to individuals with paid work. Other types of commute trips (i.e., travel to school) are not included. The indicator does not reflect walking or biking to public transportation. Race/ethnicity data was only available from the NHTS. Biking data is only available at the CMSA and state level, with limited county level data (NHTS 2009). Travel time data was missing for 16% of workers in the 2001 NHTS and for 18% of workers in the 2009 NHTS. Calculations of the indicator at Census tract level were not statistically reliable.

<u>Data Processing</u>

SAS Program: HCI_Walk_Bicycle_778-10-23-13.sas

Data Processing: Percent of the population by time walking or biking to work



Data Acquisition and Preprocessing

ACS data on the Means of Transportation to Work by Travel Time to Work is found in the detailed table B08134. SAS programs were used to extract the data from the 2007-2011 ACS summary files (see Toolbox section, Obtaining American Community Survey Data from Summary Files chapter), using a sequence number and SAS programs (Table 18).

Table 18. ACS Datasets, Tables, and SAS extraction programs used for the Percent of Workers' Daily Commute 10 or More Minutes by Walking or Biking Indicator

ACS Dataset	Table and sequence number	SAS data extraction program
3-year estimates 2005-2007	B08134, 0041	Walk_Bicycle_ACS2005-2007data_9-20-13.sas
3-year estimates 2008-2010	B08134, 0041	Walk_Bicycle_ACS2008-2010data_9-20-13.sas
5-year estimates 2007-2011	B08134, 0027	Walk_Bicycle_ACS2007-2011data_CA_CO_PL_9-23-13.sas Walk_Bicycle_ACS2007-2011data_CT_9-23-13.sas

The files produced by the SAS programs contained the following variables:

- n_time_mode: number of workers (16 years or older) that traveled a certain amount of time (10 minutes or less, 10 to 14 minutes, etc.) to work by mode (Car, Bus, Bicycle, etc), for a particular geography.
- n_total: number of workers (16 years or older) that traveled to work, for a particular geography.
- n_time_mode_se: standard error of n_time_mode.
- n_total_se: standard error of n_total.

The person file of the National Household Travel Surveys (NHTS) 2001 and 2009 was downloaded from http://nhts.ornl.gov/download.shtml. The variables of interest in the survey were TIMETOWK and WORKTRANS.

The SAS program used to extract the information from the Person File conducted the following preprocessing:

- Subset California data.
- Categorize individuals by age (0-15, 16+), race/ethnicity (White, African American, Asian, AIAN, NHOPI, Multiple, Latino, Other), and work status (employed, unemployed, works from home).
- Reclassify mode to work categories into bicycle, car, walk, bus, rail-ferry, streetcar-trolley, and others.
- Classify travel time to work responses into two categories: < 10 minutes, ≥ 10 minutes.
- Create a new file for employed individuals, age 16+. Create a variable identifying those that bike 10 min/day or more to work ("bike").
- Use the SURVEYMEANS procedure in SAS to estimate characteristics of a survey population by using statistics computed from a survey sample. The NHTS provides the person sampling weights. (Include geography and/or race/ethnicity in the domain statement; include "bike" in the class and var statements; use house in the cluster statement; and use the person sampling weights, wtperfin, in the weight statement).

- Percentage of people that bikes ≥ 10 min/day to work by geography and race/ethnicity and its standard error.
- Percentage of people that bikes ≥ 10 min/day to work by geography (all race/ethnicities combined) and its standard error.

NHTS-ACS compatibility

Travel Time to Work: ACS "data on travel time to work were derived from answers to Question 34. This question was asked of people who indicated in Question 29 that they worked at some time during the reference week, and who reported in Question 31 that they worked outside their home. Travel time to work refers to the total number of minutes that it usually took the worker to get from home to work during the reference week. The elapsed time includes time spent waiting for public transportation, picking up passengers in carpools, and time spent in other activities related to getting to work."

The NHTS interview included the following questions on travel time and travel mode to work, that are compatible with questions 31 and 34 of the ACS.

- ACS
    - 31. How did this person usually get to work LAST WEEK?
    - 34. How many minutes did it usually take this person to get from home to work LAST WEEK?
- NHTS 2001:
    - E15. (N_F6) How many minutes did it usually take {you/SUBJECT} to get from home to work last week? (TIMETOWK)
    - E16. (N_F7) How did {you/SUBJECT} usually get to work last week? (WRKTRANS)
- NHTS 2009:
    - E15. How many minutes did it usually take {you/SUBJECT} to get from home to work last week? (TIMETOWK)
    - E16. How did {you/SUBJECT} usually get to work last week? (WRKTRANS)

The responses to these questions can be found in the NHTS Person File, for both 2001 and 2009.

Employment: According to the U.S. Census the "employed" category includes "all civilians 16 years old and over who either (1) were "at work," that is, those who did any work at all during the reference week as paid employees, worked in their own business or profession, worked on their own farm, or worked 15 hours or more as unpaid workers on a family farm or in a family business; or (2) were "with a job but not at work," that is, those who did not work during the reference week but had jobs or businesses from which they were temporarily absent due to illness, bad weather, industrial dispute, vacation, or other personal reasons. Excluded from the employed are people whose only activity consisted of work around the house or unpaid volunteer work for religious, charitable, and similar organizations; also excluded are all institutionalized people and people on active duty in the United States Armed Forces."

In order to obtain NHTS data on workers 16 years or older that would be comparable with ACS data, information on age, employment, and frequency of working from home, was used to categorize individuals as employed, unemployed, or works from home, as shown in Figure 6 (Q. refers to the question number).

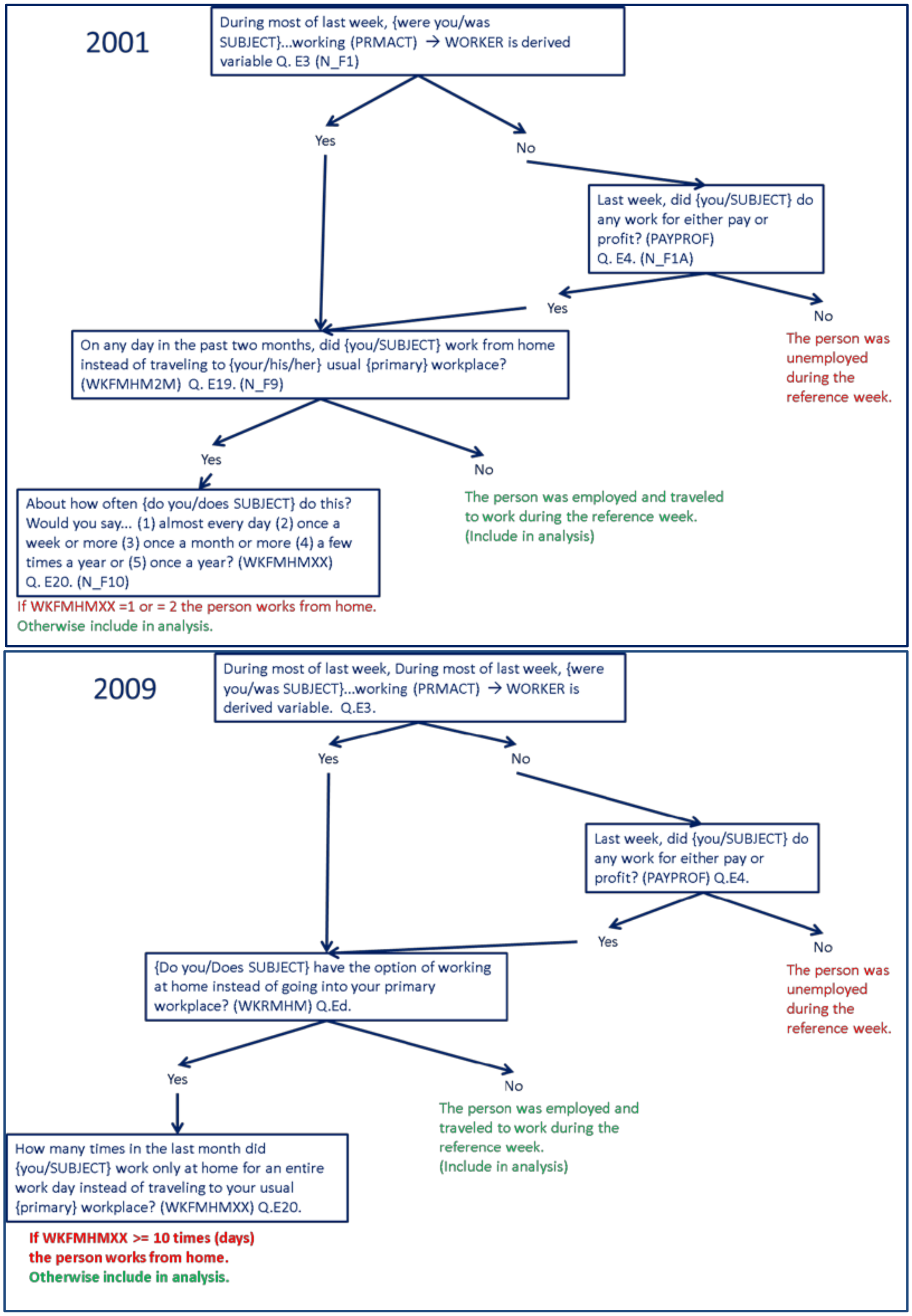


Figure 6. Decision Rules to Determine Persons Working from Home in the 2001 and 2009 National Household Travel Survey.

Indicators Under Construction

Percent of Household Income Spent on Transportation

Full Title: Percent of household income spent on transportation

Indicator ID: 58

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Safe, sustainable, accessible and affordable transportation options

Rationale: Evidence supports that health outcomes improve with available income; spending on transportation competes with other basic necessities; indicator understandable to wide range of users; California Regional Progress Report 2010 includes this indicator.

Detailed Definition:

- Numerator: not provided by the data source
- Denominator: not provided by the data source
- Indicator: Percent of household income spent on transportation (modeled)
- Stratification: eight household types (Dual-Income Family, Low Income, Regional Moderate, Regional Typical, Retirees, Single Person Very Low Income, Single Professional, Single Worker). No race/ethnicity data available from source.
- Interpretation: healthier communities will have lower transportation costs

Data Description:

- Data sources: Location Affordability Portal, U.S. Department of Housing and Urban Development, U.S. Department of Transportation, http://www.locationaffordability.info/lai.aspx?url=download.php
- Years available: 2006-2010
- Updated: with every ACS 5-year estimates release
- Geographies available: Core Based Statistical Areas (CBSA), counties, places, region (derived), state (derived)

Methodology:

Modeled data on location affordability for California (2006-2010) can be downloaded from the Location Affordability Portal for CBSAs, counties, and places in California. Data is available for 8 household types. Six different location affordability measures are included in the data files:

(1) housing and transportation costs as a percent of income,
(2) housing and transportation costs as a percent of income, owners,
(3) housing costs as a percent of income, owners,
(4) housing and transportation costs as a percent of income, renters,
(5) housing costs as a percent of income, renters, and
(6) transportation costs as a percent of income.

The indicator is the affordability measure number 6. Travel costs include auto ownership, driving, and transit costs (http://www.locationaffordability.info/About_Data.aspx#CalculatingTrans). No additional calculations are necessary for the indicator. No standard error information is available, confidence intervals and relative standard error cannot be calculated. Place deciles and relative risk can be calculated.

Limitations/Comments:

The data are derived from a model; no standard errors or race/ethnicity information is available. The Location Affordability Index is primarily intended for use by researchers, developers, planners, and policymakers to help enhance their understanding of combined housing and transportation cost burdens, analyze the relationships between the affordability landscape and other factors (e.g. transportation infrastructure, development pressure, etc.), and communicate this information to the public and stakeholders. The interpretation for housing cost estimates for a specific household type in a given location is the expected amount spent on housing in that location between 2006 and 2010 (http://www.locationaffordability.info/about_lai.aspx).

Methodological details on the calculation of transportation costs can be found here: http://www.locationaffordability.info/About_Data.aspx#CalculatingTrans. The following figure illustrates the model from which the data is derived.

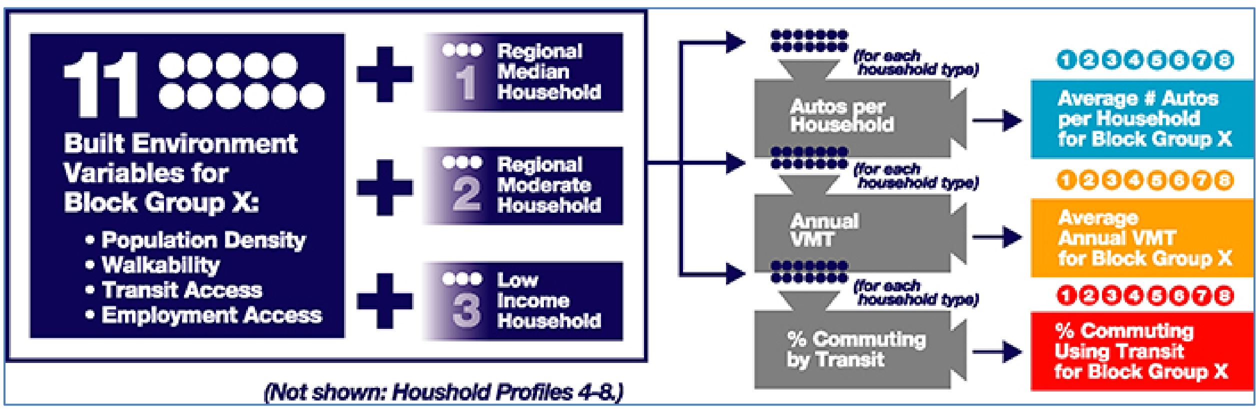


Figure 7. Location Affordability Index Model (taken from http://www.locationaffordability.info/About_Data.aspx#CalculatingTrans).

Access to High Quality Food Retailers

Full Title: Percent of the population within ½ mile full-service grocery store, fresh produce market, or store with fresh produce

Indicator ID: 76

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Affordable, accessible and nutritious foods

Rationale: Employs a commonly used definition of walking distance to access health promoting food resources and understandable to wide range of users. Evidence supports increased access to healthy foods promotes their consumption. Data are available from public and proprietary data sources and are feasibly geocoded for Census tracts or blocks. Indicator used by the California Department of Public Health and county health departments to measure healthy food access.

Detailed Definition:

- Numerator: number of people within ½ mile full-service grocery store, fresh produce market, or store with fresh produce
- Denominator: total population
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: eight race/ethnicity strata (African American, AIAN, Asian, Latino, Multiple, NHOPI, White, total)
- Interpretation: healthier communities will have higher percentages of people within ½ mile full-service grocery store, fresh produce market, or store with fresh produce

Data Description:

- Potential government data sources:
  - Store/market data: California Board of Equalization (BOE, http://www.boe.ca.gov). United States Department of Agriculture (USDA), Food and Nutrition Service (FNS), Supplemental Nutrition Assistance Program (SNAP), retail locator (http://www.fns.usda.gov/snap/retailerlocator). United States Department of Agriculture (USDA), Economic Research Service (ERS), Food Access Research Atlas (http://www.ers.usda.gov/data/fooddesert#). California Department of Public Health, Nutrition Education and Obesity Prevention Branch, Network for a Healthy California - GIS Map Viewer (http://gis.cdph.ca.gov/cnn/, data from multiple sources).
  - Population data: Population counts, Census block level, 2010, U.S. Census Bureau Redistricting File for California (http://www2.census.gov/census_2010/01-Redistricting_File--PL_94-171/California/).
- Potential proprietary data sources:
  - Store/market data: Dun & Bradstreet (http://www.dnb.com/); Trade Dimensions TDLinx (a Nielsen company, www.nielsen.com). InfoUSA (www.infousa.com).
- Years available: Unknown (BOE); 2014 (SNAP); 2010 (ERS); current year (proprietary data sources)
- Updated: Annually (BOE); unknown (SNAP); unknown (ERS); annually or semiannually (proprietary data sources)

- Geographies available: Census tract, cities/towns, county, region, state

Methodology:

Geocoded data of full-service grocery store, fresh produce market, or store with fresh produce needs to be obtained from government data sources; otherwise data needs to be purchased from proprietary sources. Farmers' Market data can be obtained from the Network for a Healthy California – GIS map viewer. The original source for the data is the USDA: http://search.ams.usda.gov/farmersmarkets/default.aspx#. Half mile "crow's fly" buffers will be calculated around the geocoded location of stores using ArcMap; the number of households within the buffers will be estimated by allocating Census block centroids into buffers. Data will be imported into SAS for analysis; the percent of households within ½ mile of stores will be estimated. The standard error can be calculated using the binomial approximation. Relative standard error, confidence intervals, deciles, and relative risk can be calculated.

Limitations/Comments:

The HCDIP has met with the State Geospatial Information Office at the California Technology Agency to request his assistance in piloting geocoded establishment data from the Board of Equalization business establishment registration database. This information is required to construct several indicators related to neighborhood completeness/walkability, and availability of grocery and other food establishments. Pilot data obtained from the Board of Equalization's business establishment registrations suggests that the 4-digit coding detail using the North American Industrial Classification System (NAICS) may not be sufficient to identify different types of retail establishments. While further evaluation is necessary, it appears that, unless more detailed coding is available via BOE, data from commercial vendors who do additional detailed coding will be only alternative source for this type of establishment level data.

As an alternative data source, a database containing the number and percent of people at least ½ mile from stores at the Census tract level (2010 population and store data) is available to download from the USDA's Food Access Research Atlas (http://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data.aspx). To calculate this percentage, he USDA created a directory of stores, supermarkets, supercenters, and large grocery stores within the United States, including Alaska and Hawaii using geocoded information on stores authorized to accept Supplemental Nutrition Assistance Program (SNAP, data available from the USDA retail locator) and the 2010 NACS/Nielsen TDLinx.

Small retail food stores with annual sales volume of less than $500,000 are generally not included in proprietary databases, but are included in government databases. (Wang et al. 2008, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1564032/).

The USDA includes the following types of retail food stores in their Food Access Research Atlas: "Retail food stores that offer a full range of food products - including fresh meat and poultry, produce, dairy, dry and packaged foods, and frozen foods - and that have at least $2 million or more in annual sales. Stores meeting these criteria include one of three store types: (1) supercenters: large stores usually 100,000 square feet or more of floor space, with a separate grocery area and general merchandise area under a single roof; (2) supermarkets: stores that are typically smaller than a supercenter and that primarily sell food and nonfood grocery products; and (3) large grocery stores: stores that sell a full range of foods and have at least $2 million in annual sales, but are not as large as supermarkets."

Daily Fruit and Vegetable Intake in Adults

Full Title: Percent of adults who consume ≥5 servings of fruits and vegetables a day

Indicator ID: 85

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Affordable, accessible and nutritious foods

Rationale: Widely used indicator in public health for healthy nutritional intake based on evidence of cancer prevention and other health outcomes; indicator is understandable to wide range of users. Measured biennially in California adults at the county level in the California Health Interview Survey. Mentioned in the CDPH's 2010 California Obesity Prevention Plan 2010. Indicator may be stratified by race/ethnicity, income, and educational level to assess potential inequities.

Detailed Definition:

- Numerator: number of adults who consume ≥5 servings of fruits and vegetables a day
- Denominator: number of adults
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: seven race/ethnicity strata (African American, AIAN, Asian, Latino, Multiple, NHOPI, White, total)
- Interpretation: healthier communities will have higher percentages of adults who consume ≥5 servings of fruits and vegetables a day

Data Description:

- Data sources: California Health Interview Survey (CHIS), UCLA Center for Health Policy Research, Confidential data, http://healthpolicy.ucla.edu/chis/data/Pages/confidential.aspx
- Years available: 2001, 2005
- Updated: unknown
- Geographies available: county, region (derived), state

Methodology:

SAS code provided by CHIS will be used as a template to obtain the estimate and its standard error using the replicate weights method (Proc Surveymeans). Based on information included in the data dictionary of the CHIS public use files, the variable of interest is “FV_5PLUS: 5+ FRUIT/VEGS. A DAY?” for 2001 and “FV: DAILY SERVINGS OF FRUITS & VEGETABLES” for 2005. Confidence intervals, relative standard error, deciles, and relative risk can be calculated.

Limitations/Comments:

The CHIS variable FV is a constructed variable. Constructed variables are put together by the CHIS Data Production team or by Westat, Inc., the data collection contractor. Constructed variables are usually based on multiple questions asked during the interview.

Recent data is not available for this indicator and it is unknown if the data will be updated in the future. Survey questions are added, removed, and modified in each cycle of CHIS. Questions are added to meet stakeholders' needs and to monitor emerging public health concerns. Questions are removed to reduce the length of the survey interview and save data collection costs when questions are no longer relevant for public health surveillance, or when they are not funded by a sponsor. Occasionally, changes are made to question wording based on methodological evaluations or user feedback that strongly suggests that changes will produce better data; otherwise, questions are kept consistent to aid in trending.

As a potential alternative to CHIS, the Behavioral Risk Factors Surveillance System (BRFSS) of the Centers for Disease Control and Prevention (http://www.cdc.gov/brfss/) has estimates of the number of adults who have consumed fruits and vegetables five or more times per day for California (state only) for the years 2002, 2003, 2005, 2007 and 2009. Data can be disaggregated in five race/ethnicity groups (white, black, Hispanic, other and multiracial). Data for selected Metropolitan Areas in California is available for 2009.

Adults without Health Insurance

Full Title: Percent of adults aged 18-64 years without health insurance

Indicator ID: 187

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Affordable, accessible and high quality health care

Rationale: Widely used indicator of lack of access to medical services by health care and state and county public health agencies; and understandable to wide range of users. Data collected biennially in California Health Interview Survey for counties and annually for most California cities by the American Community Survey. Indicator (CHIS) may be stratified by race/ethnicity, income, and educational level to assess potential inequities.

Detailed Definition:

- Numerator: number of adults aged 18-64 years without health insurance
- Denominator: number of adults
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: seven race/ethnicity strata (African American, AIAN, Asian, Latino, Multiple, NHOPI, White, total)
- Interpretation: healthier communities will have a low percentage (or zero) of adults aged 18-64 years without health insurance

Data Description:

- Data sources: California Health Interview Survey (CHIS), UCLA Center for Health Policy Research, Confidential data, http://healthpolicy.ucla.edu/chis/data/Pages/confidential.aspx
- Years available: 2001, 2003, 2005, 2007, 2009, 2011-12
- Updated: continuous survey, published every two years
- Geographies available: county, region (derived), state

Methodology:

SAS code provided by CHIS will be used as a template to obtain the estimate and its standard error using the replicate weights method (Proc Surveymeans). Based on information included in the data dictionary of the CHIS public use files, the variable of interest is “INSURED: Currently insured” for 2001 and “INS: Currently insured” for all other years. Confidence intervals, relative standard error, deciles, and relative risk can be calculated.

Limitations/Comments:

INS is a constructed variable. Constructed variables are put together by the CHIS Data Production team or by Westat, Inc., the data collection contractor. Constructed variables are usually based on multiple questions asked during the interview. In general, constructed variables can be identified by their variable names, which are acronyms and/or abbreviations.

The INS Currently Insured variable indicates the current insurance status of the respondent. Cases that are assigned a value of 1 (covered) for any of the following variables are considered to be currently insured (INS=1): INSMC, INSMD, INSHF, INSEM, INSPR, INSML, INSOG, INSOT. The cases assigned a value of 2 (not covered) for all of those variables are considered to not be currently insured (INS=2). (Source: http://healthpolicy.ucla.edu/chis/data/public-use-data-file/Documents/CV2011-12_Adult_PUF_6.19.2014.pdf)

Survey questions are added, removed, and modified in each cycle of CHIS. Questions are added to meet stakeholders' needs and to monitor emerging public health concerns. Questions are removed to reduce the length of the survey interview and save data collection costs when questions are no longer relevant for public health surveillance, or when they are not funded by a sponsor. Most CHIS questions are included in every CHIS cycle. Occasionally, changes are made to question wording based on methodological evaluations or user feedback that strongly suggests that changes will produce better data; otherwise, questions are kept consistent to aid in trending.

As a potential alternative to CHIS, health insurance data is available for Census tracts and places from the American Community Survey 2006-2010 period (Table DP03), but without race/ethnicity information.

The ACS began asking questions about health insurance coverage in 2008. Because 2008 was the first year of collection, the Census Bureau limited the number and type of data products to simple age breakdowns of overall, private, and public coverage status. The evaluation of the 2008 data suggested that the data were of good quality, so the Census Bureau expanded the data products to include estimates of the specific types of coverage along with estimates about social, economic, and demographic details for people with and without health insurance.

For the 2008 data released September 2009, there was no eligibility edit applied. The eligibility edit that was developed for the 2009 was applied to the 2008 data during spring 2010. New estimates of health insurance coverage with this data are available.

Occupational Injuries

Full Title: Number and rate of fatal and nonfatal occupational injuries by industry

Indicator ID: 297

Healthy Community Framework: Adequate levels of economic, social development

Aspirational Goal: Living wage, safe and healthy job opportunities for all

Rationale: Easily understood, widely used indicator of acute occupational health conditions produced annually by California Department of Industrial Relations Cal/OSHA.

Detailed Definition:

- Numerator: number of occupational injuries
- Denominator: total civilian employed population by industry
- $Indicator\ (rate) = \frac{Numerator}{Denominator} \times 1{,}000$
- Stratification: type of injury (fatal, nonfatal); type of industry (13 categories)
- Interpretation: healthier communities will have lowest (or zero) rates of occupational injuries

Data Description:

- Data sources:
  - o Numerator: California Department of Industrial Relations, Census of fatal occupational injuries (http://www.dir.ca.gov/dosh/cfoi/cfoi.htm) from the Office of Occupational Safety and Health; fatal occupational injuries and illnesses (https://www.dir.ca.gov/oprl/fatal.htm), non-fatal occupational injuries and illness (https://www.dir.ca.gov/OPRL/nonfatal.htm) from the Office of Policy, Research, and Legislation.
  - o Denominator: American Community Survey (DP03) 1-year state level estimates of the civilian employed population (16 years and older) by industry (http://factfinder2.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t).
- Years available: 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012
- Updated: annually
- Geographies available: state

Methodology:

The number of fatal injuries and non-fatal injuries registered in the state are available from the Department of Industrial Relations. Tables are in pdf format. Non-fatal injuries are available in Excel format but require re-formatting. Injuries are presented by industry type. Race/ethnicity information is available for fatal injuries, but there is no cross tabulation with industry type. State level estimates of the civilian employed population (16 years and older) by industry can be obtained from the American Community Survey 1-year estimates. The standard error can be estimated using the Poisson approximation. Confidence intervals, relative standard error, deciles, and relative risk can be calculated.

Limitations/Comments:

The Department of Industrial Relations publishes non-fatal injury incidence rates at the state level by industry using the following formula:

[1] Incidence rates represent the number of injuries and illnesses per 100 full-time workers and		
were calculated as: (N/EH) x 200,000 where		
	N	= number of injuries and illnesses
	EH	= total hours worked by all employees during
		the calendar year
	200,000	= base for 100 equivalent full-time workers
		(working 40 hours per week, 50 weeks per year).

Following is the list of 13 industry classes for which injury data is reported:

- Agriculture, Forestry, Fishing and Hunting
- Mining, Quarrying, and Oil and Gas Extraction
- Utilities
- Construction
- Manufacturing
- Wholesale Trade
- Retail Trade
- Transportation and Warehousing
- Information
- Finance and Insurance
- Real Estate and Rental and Leasing
- Professional, Scientific, and Technical Services
- Management of Companies and Enterprises
- Administrative and Support and Waste Management and Remediation Services
- Educational Services
- Health Care and Social Assistance
- Arts, Entertainment, and Recreation
- Accommodation and Food Services
- Other Services (except Public Administration)
- Public Administration

Academic Performance Index

Full Title: Mean score of Academic Performance Index (API)

Indicator ID: 341

Healthy Community Framework: Adequate levels of economic, social development

Aspirational Goal: Opportunities for high quality and accessible education

Rationale: Widely used indicator based on standardized tests of California public school students' proficiency in English language and math subjects. Evidence links educational success with positive health and social outcomes. Reported annually by the California Department of Education and may be stratified by race/ethnicity and economic disadvantage to assess equity.

Detailed Definition:

- The Academic Performance Index (API) is a single number, ranging from a low of 200 to a high of 1000, which reflects a school's, and local educational agencies (LEAs), or a student group's performance level, based on the results of statewide assessments. Its purpose is to measure the academic performance and improvement of schools. The state has set 800 as the API target for all schools to meet. Schools that fall short of 800 are required to meet annual growth targets until that goal is achieved. API targets vary for each school and student group. The API is calculated by converting a student's performance on statewide assessments across multiple content areas into points on the API scale. These points are then averaged across all students and all tests. The result is the API. An API is calculated for schools, LEAs, and for each student group with 11 or more valid scores at a school or an LEA. (Source: Analysis, Measurement, and Accountability Reporting Division California Department of Education, http://www.cde.ca.gov/ta/ac/ap/documents/apiexecsummary.pdf)
- $Indicator = \ Mean\ Score\ of\ Academic\ Performance\ Index$
- Stratification: seven race/ethnicity groups (African American, AIAN, Asian, Latino, NHOPI, white, Multiple)
- Interpretation: healthier communities will have higher API or student performance level

Data Description:

- Data sources: California Department of Education (CDE), Testing and Accountability, Academic Performance Index data files, http://www.cde.ca.gov/ta/ac/ap/apidatafiles.asp
- Years available: 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013
- Updated: annually
- Geographies available: school district, county, region (derived), state

Methodology:

The base and growth API for multiple cycles can be downloaded from the DOE Testing and Accountability website. "Each reporting cycle begins with a Base API. The Base API is calculated using the assessment results of the previous year and the Growth API is calculated using the assessment results of the current year. For example, the 2012 Base API is calculated using results of statewide testing from spring 2012 and the 2013 Growth API is calculated using results of statewide testing from spring 2013. Any changes in the API calculations, such as adding a new assessment, begin with the Base API. Therefore, the calculation methods for the Base API might not be the same across years. However, the Base API and Growth API within a reporting cycle must use the same calculation method. The 2012 Base API is subtracted from the 2013 Growth API to show how much a school's API changed from 2012 to 2013 (referred to as 2012–13 API growth). This determines whether a school meets its API growth target. The Base API Report includes the Base API, targets, and ranks. The Growth API Report includes the Growth API, growth achieved, and whether or not targets were met."

API data will be extracted from the CDE files and reformatted in SAS. The Base API, Growth API, and the change within a cycle will be reported. There is no standard error for the API therefore confidence intervals or relative standard error will not be calculated. Deciles and relative risk will be calculated. The "number of students included in the Academic Performance Index (API)" will be added as an extra column in the file.

Limitations/ Comments:

API cannot be directly compared from year to year because the grades, the tests and the test weights might change from cycle to cycle. "API comparisons must be based on the same tests with the same test weights. If the API indicators and methodology remained the same from year to year, there would be no need for two API reports. However, complication arises in year-to-year comparisons of the API when changes to the API must be made. From one year to the next, assessments may be added or taken away from the set of API indicators, or API rules may change. For example, in the 2012–13 API cycle, performance levels of grade eight and nine results on the CST in General Mathematics were not lowered. Also, the test weights (relative emphasis on each test) or rules for inclusions/exclusions in the API can also change." "Because new indicators are added to the API, test weights may change, and methodologies may change from one cycle to the next, it is inappropriate to compare APIs across reporting cycles. It is appropriate, however, to compare the Base and Growth APIs within a reporting cycle as well as to compare the amount of API growth (i.e., change in the API) of different reporting cycles." (Source: Analysis, Measurement, and Accountability Reporting Division California Department of Education, http://www.cde.ca.gov/ta/ac/ap/documents/apiexecsummary.pdf)

The API will not be produced for the period 2013-2015 due to California's transition to a new testing system. API will be released again in 2016 (personal communication Academic Accountability team: aau@cde.ca.gov).

Proximity to Busy Roadways

Full Title: Percent of population near busy roadways

*Indicator ID***:** 399

Healthy Community Framework: Quality and sustainability of the environment

Aspirational Goal: Clean air, soil and water, and environments free of excessive noise

Rationale: Easily understandable indicator associated with noise, pedestrian safety, and air pollution, which impact health outcomes. Roadway classification data are available for all roadways in California but actual vehicle volume/counts must be compiled from municipal and county sources not consistently available in electronic format.

Detailed Definition:

- Numerator: population living within 300 meters of busy roads (>10,000 vehicles per day)
- Denominator: total population
- $Indicator = \frac{Numerator}{Denominator} \times 100$
- Stratification: no race/ethnicity stratification available
- Interpretation: healthier communities will have lower percentages of the population living near busy roadways

Data Description:

- Data sources: California Department of Public Health (CDPH), Environmental Health Investigations Branch (EHIB), California Environmental Health Tracking Program (CEHTP), http://www.ehib.org/page.jsp?page_key=130&pmn=PERPOPNBR
- Years available: 2004
- Updated: unknown
- Geographies available: county, region (derived), state

Methodology:

The California Environmental Health Tracking Program has calculated the percent of the population by county, living within 300 meters of busy roads (>10,000 vehicles per day) using 2004 data from CalTrans Highway Performance Monitoring System (HPMS) and data from the 2000 U.S. Census. The data can be aggregated to the region level. Standard error could be estimated using the binomial approximation. The CEHTP offers a traffic volume linkage tool that allows the user to obtain information about how much traffic passes near a specific location. The CEHTP service could assist HCDIP to identify busy roadways at the place and Census tract level using GIS tools. If the data is available in GIS format, it can be spatially joined with Census block centroids (300 meters buffer) to estimate the percent of population near busy roadways at finer geographical levels.

Limitations/Comments:

Traffic data are sampled in various locations at intermittent times and do not distinguish between automotive and truck traffic. The traffic data are more representative of urban populations than rural populations; rural populations can be exposed to pollution from busy roads, but fewer individuals are exposed than in urban areas. (http://www.ehib.org/page.jsp?page_key=130&year=2006&pmn=PERPOPNBR)

The definition of this indicator (defining busy roadway and buffer size) needs further input from other state agencies (transit, housing, health) and stake holders to reconcile versions that have different distance buffers and volumes.

As a potential alternative to the CEHTP data, traffic volume data by segment and by type of vehicle (truck, others) are available for the State Highway System (244 highways: interstates, California state routes, or U.S. State routes; see Table 19 below) from the Traffic Census, Division of Traffic Operations, California Department of Transportation (http://www.traffic-counts.dot.ca.gov/, http://traffic-counts.dot.ca.gov/2013all/). Data is available in Excel and GIS formats. The data reports "back annual average daily traffic (AADT)," traffic South or West of the count location and is the total volume for the year divided by 365 days, and "Ahead annual average daily traffic (AADT)," traffic North or East of the count location and is the total volume for the year divided by 365 days. "The traffic count year is from October 1st through September 30th. Very few locations in California are actually counted continuously. Traffic Counting is generally performed by electronic counting instruments moved from location throughout the State in a program of continuous traffic count sampling. The resulting counts are adjusted to an estimate of annual average daily traffic by compensating for seasonal influence, weekly variation and other variables which may be present."

Table 19. List of Interstate Highways, California State Routes, and U.S. State Routes with Traffic Volume Data

1	35	68	103	136	172	210	259	805
2	36	70	104	137	173	211	260	880
3	37	71	105	138	174	213	261	905
4	38	72	107	139	175	215	262	980
5	39	73	108	140	177	216	263	
6	40	74	109	142	178	217	265	
7	41	75	110	144	180	218	266	
8	43	76	111	145	182	219	267	
9	44	77	112	146	183	220	269	
10	45	78	113	147	184	221	270	
12	46	79	114	149	185	222	271	
13	47	80	115	150	186	223	273	
14	49	82	116	151	187	225	275	
15	50	83	118	152	188	227	280	
16	51	84	119	153	189	229	281	
17	52	85	120	154	190	232	282	
18	53	86	121	155	191	233	283	
19	54	87	123	156	192	236	284	
20	55	88	124	158	193	237	299	
22	56	89	125	160	195	238	330	
23	57	90	126	161	197	241	371	
24	58	91	127	162	198	242	380	

25	59	92	128	163	199	243	395
26	60	94	129	164	200	244	405
27	61	95	130	165	201	245	505
28	62	96	131	166	202	246	580
29	63	97	132	167	203	247	605
32	65	98	133	168	204	253	680
33	66	99	134	169	205	254	710
34	67	101	135	170	207	255	780

Per Capita Water Use

Full Title: Average daily water use per capita

Indicator ID: 416

Healthy Community Framework: Quality and sustainability of the environment

Aspirational Goal: Clean air, soil and water, and environments free of excessive noise

Rationale: Average daily use (gallons) per capita stratified by user (residential, institutional, industrial, etc.) is an indicator used in the water conservation plan of multiple California state agencies (DWR, CDPH, SWRCB, ARB, PUC, Energy, Bay-Delta Authority); easily understandable indicator available at hydrologic region and updated through annual progress reports.

Detailed Definition:

- Numerator: total withdrawals of water in Mgal/d
- Denominator: total population
- $Indicator = \frac{Numerator}{Denominator}$
- Stratification: water user (industrial, irrigation, livestock, mining, public supply, domestic, thermoelectric, aquaculture)
- Interpretation: healthier communities will have lower per capita water usage

Data Description:

- Data sources: United States Geological Survey, Water Data for the Nation, Water Use Data for the Nation, http://waterdata.usgs.gov/ca/nwis/wu. United States Geological Survey, California Water Science Center, Water Use in California, http://ca.water.usgs.gov/water_use/2010-california-water-use.html.
- Years available: 2000, 2005, 2010
- Updated: every 5 years (years ending in 0 and 5)
- Geographies available: county, region (derived), state

Methodology:

Multiple use categories are available to construct the indicator. There are differences in the water user category available for 2000-2005 and 2010 (see Table 20 below). Water use per capita by county can be calculated for each of the water users available, but the USGS only reports per capita use for the Domestic Public Supply and Domestic Self-Supplied categories; the HCDIP should evaluate if reporting per capita figures for all the user categories. Additionally some of the user categories might be aggregated to an “Others” group to reduce file content. The USGS provides total population data per county per year that can be used as denominator. No standard errors are published with the daily water use estimates.

Table 20. Categories of Water Users Available in the USGS Database

2000, 2005	2010
• Public Supply deliveries to domestic • Public Supply deliveries to commercial • Public Supply deliveries to industrial • Public Supply deliveries to thermoelectric • Public Supply total deliveries • Commercial total consumptive use • Domestic total consumptive use • Industrial total consumptive use • Total Thermoelectric Power total consumptive use • Fossil-fuel Thermoelectric Power total consumptive use • Geothermal Thermoelectric Power total consumptive use • Nuclear Thermoelectric Power total consumptive use • Thermoelectric Power (Once-through cooling) total consumptive use • Thermoelectric Power (Closed-loop cooling) total consumptive use • Mining total consumptive use • Livestock (Stock) total consumptive use • Livestock (Animal Specialties) total consumptive use • Aquaculture total consumptive use • Irrigation, Total consumptive use, fresh • Hydroelectric Power total off stream surface-water withdrawals	• Industrial, self-supplied total withdrawals, total (fresh+saline) • Irrigation, total withdrawals, fresh • Livestock, total withdrawals, fresh • Mining, total withdrawals, total (fresh+saline) • Public Supply, total withdrawals, total (fresh+saline) • Domestic, total self-supplied withdrawals, fresh • Thermoelectric, total withdrawals, total (fresh+saline) • Aquaculture, total withdrawals, fresh • Total withdrawals, total (fresh+saline)

Limitations/Comments:

Data is gathered by the USGS from multiple sources across the state to report total water withdrawals (see http://ca.water.usgs.gov/water_use/california-water-use-resources.html for 2010 data sources). There is significant uncertainty associated with the estimates. Power plant and mining water use data are considered the most accurate. Water withdrawals for crop irrigation are the daily average daily quantities used over a year and do not represent actual daily rates. According to USGS, withdrawals for irrigation are among the least accurate estimates because they are not measured directly and must be calculated based on crop acreage, crop coefficients, stage ratios, irrigation-system efficiency, and precipitation. For all other categories, the daily quantity of water that is withdrawn may be based on daily usage or more often, monthly usage that is averaged to a daily quantity.

Tree Canopy Coverage

Full Title: Tree Canopy Coverage

Indicator ID: 458

Healthy Community Framework: Quality and sustainability of the environment

Aspirational Goal: Green and open spaces, including agricultural lands

Rationale: Evidence links tree canopy to positive health effects due to mitigation of air pollution, UV exposure, and urban heat islands and creation of environments that reduce stress and neighborhood violence; understandable to range of users; data are available but require geographic processing of output from mathematical models of satellite imagery for census tract detail, which is currently available only for a few California cities. Cal/Fire mentions urban planning and forestry in its annual reports/strategic plan.

Detailed Definition:

- Numerator: area with canopy cover
- Denominator: total area
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: eight race/ethnicity strata (African American, AIAN, Asian, Latino, Multiple, NHOPI, White, total).
- Interpretation: healthier communities will have higher values of tree canopy coverage

Data Description:

- Data sources: Multi-Resolution Land Characteristics Consortium, National Land Cover Database 2011 (NLCD 2011), U.S. Forest Service Tree Canopy analytical, http://www.mrlc.gov/nlcd11_data.php. Population data at the Census block level for the year 2010, U.S. Census Bureau Redistricting File for California (http://www2.census.gov/census_2010/01-Redistricting_File--PL_94-171/California/).
- Years available: 2011
- Updated: unknown
- Geographies available: Census tract, city, county, region (derived), state

Methodology:

One of the products included in the NLCD is the U.S. Forest Service Tree Canopy analytical. This product has two layers: percent tree canopy cover and standard error. The percent canopy cover value ranges from 0 to 100 percent for 30 x 30 m pixel cells. The standard error value ranges from 0 to 45 percent and represents the uncertainty of the model in estimating the canopy value for a cell. The 2011 tree canopy product is not comparable to the 2001 product.

The percent tree canopy by Census tract could be estimated by averaging all the pixel values (U.S. National Atlas Equal Area (Lambert azimuthal) projection in the 1983 North American Datum) across census block groups, then averaging up to census tracts using population weighting (by race/ethnicity) (this method was used by CDPH-EHIB-CEHTP). The standard error for all the pixel values could also be averaged and weighted to obtain an estimate of uncertainty. The relative standard error, confidence interval, deciles, and relative risk can be calculated.

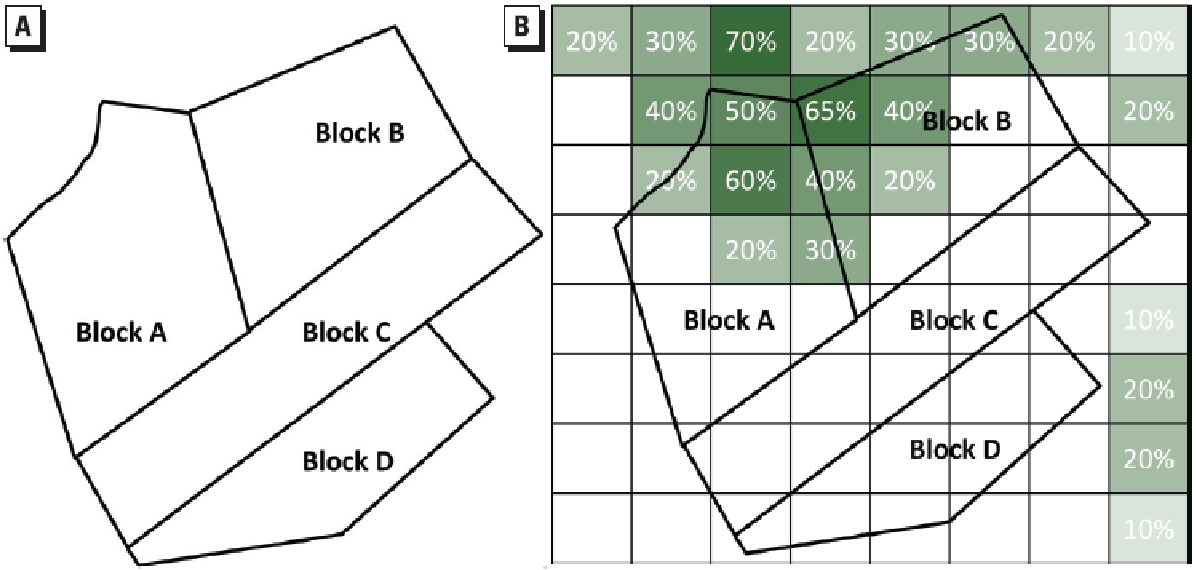


Figure 8. NLCD tree canopy overlay over a group of blocks (Taken from Jesdale et al. 2013, *Environ Health Perspect* 121(7): 811-817).

Limitations/Comments:

Studies have shown that the NLCD is valuable for analyzing land cover at the regional level but it can largely underestimate (10% on average) tree canopy at the local level (city) because it is derived from relatively coarse, 30-meter resolution satellite imagery. (http://planning.co.cuyahoga.oh.us/canopy/background.html, http://www.stateforesters.org/urban_forest_canopy_cover_primer)

Cropland Converted To Developed Land

Full Title: Acres of cropland converted to developed land

Indicator ID: 489

Healthy Community Framework: Quality and sustainability of the environment

Aspirational Goal: Green and open spaces, including agricultural lands

Rationale: Understandable indicator reported biennially for counties or sections of counties by the Department of Conservation to assess magnitude of urban encroachment and lost potential of agricultural land. An indirect measure of urban sprawl, which has been linked to built environments with additional health burdens such increased travel-related air pollution.

Detailed Definition:

- Numerator: agricultural land converted to urban and build up land in a 2 year cycle
- Denominator: total agricultural land converted in a 2 year cycle
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: none
- Interpretation: healthier communities will have lower losses of agricultural land

Data Description:

- Data sources: California Department of Conservation, Division of Land Resource Protection (DLRP), Farmland Mapping and Monitoring Program (FMMP), http://www.conservation.ca.gov/dlrp/fmmp/Pages/Index.aspx
- Years available: 2010-2012, 2008-2010, 2006-2008, 2004-2006, 2002-2004
- Updated: Biannually
- Geographies available: selected counties, state

Methodology:

The FMMP publishes biannual reports of changes in California's agricultural resources. The reports are published by county (49/58), except for Lassen, Sierra, and Plumas counties which are grouped in the Sierra Valley region. The Figure 9 shows the coverage area of the FMMP survey program.

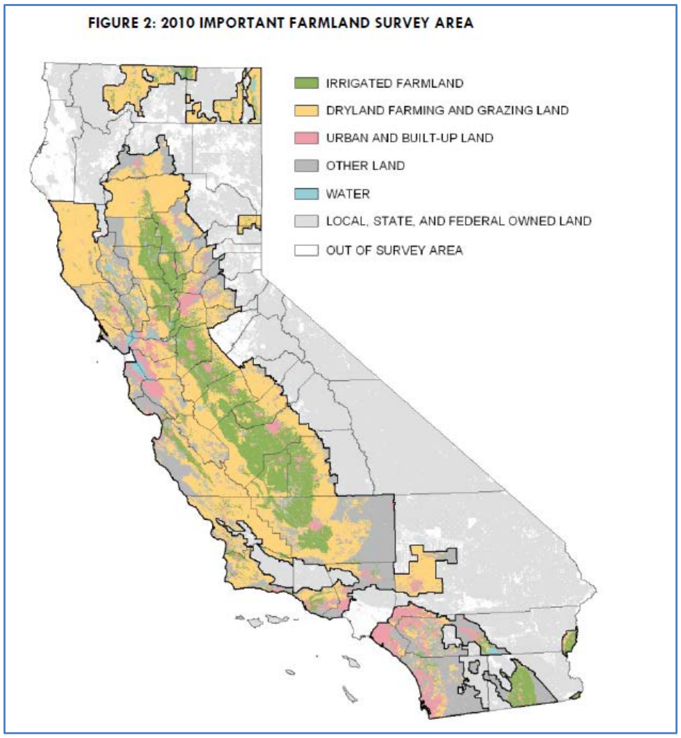


Figure 9. Farmland Mapping and Monitoring Program Survey Area.

The data by county can be downloaded in Excel format for all the years available, however the tables will require reformatting. The following is an example of the data layout that could be constructed:

Reportyear (a)	LandUse (b)	AcreageYear1 (c)	AcreageYear2 (d)	NetChangeAcres (e)	AgToUBU/ UBUToAg (f)	Percent (|e/f|*100)
2008-2010	Agricultural (Ag)	545,096	539,830	-5,266	3,937	74.76
2008-2010	UrbanBuiltUp (UBU)	315,679	321,553	5,874	41	0.69

The 74.76% represents the percent of agricultural land lost that changed to urban and built up land. The 0.69% represents the percent of urban built up land lost that changed to agriculture. The standard error for this indicator (percent) could be estimated using the binomial approximation. Relative standard error, confidence intervals, deciles, and relative risk can be calculated.

Limitations/Comments:

Shapefiles contain data on land uses for individual years (2002, 2004, 2006, 2008, and 2010) for the state are available from FMMP. These shapefiles could be used to conduct original analysis on the change from Agricultural land to urban build up land, but more research would be needed to obtain support from the FMMP on the methods they use for the calculation. Some of the categories that are included in the shapefiles are not reported at the county level by the FMMP and it is unclear if they are considered in the calculation (for example, categories CI, sAC, or Z, see below). The shapefiles contain the following categories of land uses (http://www.conservation.ca.gov/dlrp/fmmp/mccu/Pages/map_categories.aspx):

- Confined animal agriculture (CI)
- Urban and built up land (D)

- Grazing land (G)
- Farmland of local importance (L)
- LP (?)
- Prime farmland (P)
- Farmland of statewide importance (S)
- Unique farmland (U)
- Vacant or disturbed land (V)
- Water (W)
- Z (?)
- Non agricultural vegetation (nV)
- Semi agricultural and rural commercial land (sAC)

Pesticide Use

Full Title: Total pounds of selected active pesticide ingredients (filtered for hazard and volatility) used in production-agriculture per square mile

Indicator ID: 510

Healthy Community Framework: Quality and sustainability of environment

Aspirational Goal: Minimized toxics, GHG emissions and waste

Rationale: Widely used indicator to assess intensity of agricultural pesticide applications. Department of Pesticide Regulation publishes these data annually and data are mapped by CDPH Environmental Tracking Program and other groups. Acute health effects of toxic pesticides are well documented and there is evidence linking pesticide exposure to adverse birth outcomes, developmental disorders, cancers, and other chronic diseases in human populations. Data are compiled at sub-section land grids.

Detailed Definition:

- Numerator: pounds of pesticide
- Denominator: area in square miles
- $Indicator\ (ratio) = \frac{Numerator}{Denominator}$
- Stratification: eight race/ethnicities (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white)
- Interpretation: healthier communities will have lower ratios of pesticide use

Data Description:

- Data sources: Information Portal (CalPIP) (http://calpip.cdpr.ca.gov/main.cfm), Department of Pesticide Regulation. Office of Environmental Health Hazard Assessment, California Communities Environmental Health Screening Tool: CalEnviroScreen Version 2.0 (CalEnviroScreen 2.0), http://oehha.ca.gov/ej/ces2.html. Population data Census tract level for the year 2010, U.S. Census Bureau Redistricting File for California (http://www2.census.gov/census_2010/01-Redistricting_File--PL_94-171/California/).
- Years available: 3- year average, 2009-2011
- Updated: with each new version of CalEnviroScreen
- Geographies available: Census tract, cities/towns, county, region, state

Methodology:

Data from CalPIP previously analyzed by CalEnviroScreen can be obtained from their website. CalEnviroScreen only includes a subset of 69 chemicals filtered for hazard and volatility, given that higher volatility could increase the risk of exposure. Production pesticide use (total pounds of selected active ingredient) for Meridian-Township-Range-Section (MTRS) records were matched to Census tracts using a match file created in the GIS software ArcMap and later divided by each census tract's area. (http://oehha.ca.gov/ej/pdf/CES20Finalreport2014.pdf)

To calculate averages for cities, places, and regions, the mean pounds of pesticide by Census tract will be weighted by the population in the Census tract based on U.S. 2010 Census files that parse Census tracts by place and county. To calculate race/ethnicity specific weighted averages, the proportion of race/ethnicities by Census tract will be calculated from U.S. 2010 Census files. There is no standard error on the data provided by CalEnviroScreen; the relative standard error cannot be calculated. Deciles and relative risk can be calculated.

Limitations/Comments:

Annual raw data of pesticide use can be obtained from the Department of Pesticide Regulation for years prior to 2009-2011. More research would is needed to determine the pesticides to include in the calculation and on how to match the MTRS to Census tracts using ArcMap. An alternative indicator of pounds of pesticide use per capita could be constructed using the raw data.

Toxic Chemicals Released to the Environment

Full Title: Toxicity-weighted concentrations of modeled chemical releases to air from facility emissions and off-site incineration

Indicator ID: 511

Healthy Community Framework: Quality and sustainability of environment

Aspirational Goal: Minimized toxics, GHG emissions and waste

Rationale: Widely used indicator of presence of toxic chemical generation, storage, and potential release based on US EPA's/California Toxic Release Inventory (TRI). Indicator mentioned in Department of Toxic Substance Control strategic plan. Information is available for Zip codes. Information in this indicator has been combined with other characteristics of neighborhoods to assess cumulative health impacts in vulnerable populations and equity.

Detailed Definition:

- Toxicity-weighted concentrations of modeled chemical releases to air from facility emissions and off-site incineration
- Stratification: eight race/ethnicity strata (African American, AIAN, Asian, Latino, Multiple, NHOPI, White, total).
- Interpretation: healthier communities will have lower amounts (or zero) of toxic chemicals released to the environment.

Data Description:

- Data sources: U.S. Environmental Protection Agency, Risk-Screening Environmental Indicators (RSEI), http://www.epa.gov/opptintr/rsei/index.html. Office of Environmental Health Hazard Assessment, California Communities Environmental Health Screening Tool: CalEnviroScreen Version 2.0 (CalEnviroScreen 2.0), http://oehha.ca.gov/ej/ces2.html. Population data Census tract level for the year 2010, U.S. Census Bureau Redistricting File for California (http://www2.census.gov/census_2010/01-Redistricting_File--PL_94-171/California/).
- Years available: 2010
- Updated: annually
- Geographies available: Census tract, cities/towns, county, region (derived), state

Methodology:

Data from RSEI previously analyzed by CalEnviroScreen can be obtained from their website. RSEI is a model that uses the reported quantities of Toxic Release Inventory (TRI) (http://www2.epa.gov/toxics-release-inventory-tri-program) on releases and transfers of chemicals to estimate the impacts associated with each type of air and water release or transfer by every TRI facility. RSEI produces a hazard-based result, calculated by multiplying the pounds released by the chemical-specific toxicity weight for the exposure route (oral or inhalation) associated with the release. If there is no toxicity weight available for the chemical, then the hazard score is zero. CalEnviroScreen has extracted RSEI toxicity-weighted concentrations of modeled chemical releases to air in grid format and converted it to Census tracts.

To calculate averages for cities, places, and regions, the toxicity weighted concentrations by Census tract will be weighted by the population in the Census tract based on U.S. 2010 Census files that parse Census tracts by place and county. To calculate race/ethnicity specific weighted averages, the proportion of race/ethnicities by Census tract will be calculated from U.S. 2010 Census files. There is no standard error on the data provided by CalEnviroScreen; the relative standard error cannot be calculated. Deciles and relative risk can be calculated.

Limitations/Comments:

Facilities are required to report to TRI if they have 10 or more full-time employees, operate within a set of industrial sectors outlined by TRI, and manufacture more than 25,000 pounds or otherwise use more than 10,000 pounds of any listed chemical during the calendar year. Lower reporting thresholds apply for PBT chemicals (10 or 100 pounds) and dioxin-like chemicals (0.1 gram). (http://oehha.ca.gov/ej/pdf/CES20Finalreport2014.pdf)

"RSEI is subject to the limitations of the underlying data sources and models that it incorporates, in addition its own limitations:

- RSEI relies exclusively on TRI-reported data for release estimates.
- TRI data do not cover all sources of TRI chemicals; they only include industrial releases from certain industries that meet specific reporting thresholds.
- TRI does not include all toxic chemicals.
- RSEI does not provide scores for all TRI chemicals because information required for modeling, such as toxicity data, is not available for every chemical.
- RSEI does not cover all exposure routes or all chronic health effects.
- Toxicity weights are based only on chronic human toxicity.
- Toxicity weights do not address acute human toxicity or environmental toxicity.
- Dermal and food ingestion pathways (other than fish consumption) and other indirect exposure pathways are not evaluated.
- RSEI uses a number of simplifying assumptions."

(http://www.epa.gov/opptintr/rsei/pubs/using_rsei.html)

Per Capita Greenhouse Gas Emissions

Full Title: Annual Per Capita Green House Gas (GHG) Emissions

Indicator ID: 514

Healthy Community Framework: Quality and sustainability of environment

Aspirational Goal: Minimized toxics, GHG emissions and waste

Rationale: Widely used indicator mentioned in CARB Climate Change Scoping Plan (AB32) and Regional Targets Advisory Committee (SB375).

Detailed Definition:

- Numerator: million metric tons of carbon dioxide equivalents per unit of time
- Denominator: population of geographic unit per unit of time
- $Indicator\ (rate) = \frac{Numerator}{Denominator}$
- Stratification: nine sectors of GHG emissions (Agriculture & Forestry, Commercial, Electricity Generation (Imports), Electricity Generation (In State), Industrial, Military, Not Specified, Residential, Transportation). Four greenhouse gas categories (CO_2, N_2O, CH_4, other).
- Interpretation: healthier communities will have lower per capita GHG emissions

Data Description:

- Data Source:
  - Numerator: California Environmental Protection Agency, Air Resources Board (ARB), Greenhouse Gas Inventory Data - 2000 to 2012, http://www.arb.ca.gov/cc/inventory/data/data.htm.
  - Denominator: Historical Population and Housing Estimates for Cities, Counties, and the State, 2000-2010, Demographic Research Unit, Department of Finance (DOF) (http://www.dof.ca.gov/research/demographic/reports/estimates/e-8/2000-10/view.php). Years 2010-2014 can be found at http://www.dof.ca.gov/research/demographic/reports/estimates/e-5/2011-20/documents/E-5_2014_Internet_Version.xls
- Years available: 2000-2012
- Updated: annually
- Geographies available: state

Methodology:

Numerator data can be downloaded in Excel format from the California Air Resources Board Greenhouse Gas Inventory Data. Data is disaggregated in many subsector levels, activities, and type of gas. Data will be imported into SAS and will be aggregated only by sector and type of gas. Data will be merged with total population counts for the state by year to calculate the ratio. Standard error for the ratio and relative standard error can not be calculated. Deciles and relative risk can be calculated.

Limitations/Comments:

The greenhouse gas (GHG) emissions inventory provides estimates of the amount of GHGs emitted to the atmosphere by human activities within California. The inventory includes estimates for carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), sulfur hexafluoride (SF_6), hydrofluorocarbons (HFCs), and perfluorocarbons (PFCs) and nitrogen trifluoride (NF_3). The current GHG inventory covers years 2000 to 2012. The emission estimates are statewide estimates that rely both on state, regional or national data sources, and on aggregated facility-specific emissions reports. In particular, 2009 through 2012 facility level data from the GHG Mandatory Reporting Program were used for compiling statewide emissions from electricity generation, refineries, cement plants, lime and nitric acid production. The amount of GHG emissions are expressed in millions of metric tonnes of CO_2 equivalent (mmtCO_2e). The CO_2 equivalence calculations are based upon the IPCC Fourth Assessment Report's (IPCC, 2007) global warming potentials. In preparation for each new edition of the GHG inventory, recalculations are made to correct errors, incorporate new methodologies or, most commonly, to reflect changes in statistical data supplied by other agencies. Emission and sink estimates are recalculated for all years in order to maintain a consistent time-series of estimates within the inventory. Only facilities emitting more than the regulation's thresholds are required to report their emissions. As a consequence, reported emissions represent the totality of emissions in sectors where all facilities are over the threshold (i.e. cement manufacturing and petroleum refining) but not in the other sectors, such as electricity generation.

Per Capita Disposal Rate

Full Title: Total waste diversion (per capita disposal rate)

Indicator ID: 524

Healthy Community Framework: Quality and sustainability of environment

Aspirational Goal: Minimized toxics, GHG emissions and waste

Rationale: Used by Cal/Recycle to facilitate state reporting requirements (revised 2007) for municipal waste diversion. Annual results are reported by municipality.

Detailed Definition:

- Numerator: $\text{Annual disposal tons} * \frac{2{,}000 \text{ lbs}}{\text{ton}}$
- Denominator: total population
- $Indicator\ (ratio) = \frac{Numerator}{Denominator}$
- Stratification: none
- Interpretation: healthier communities will have lower per capita disposal rates

Data Description:

- Data source:
    - Numerator: California Department of Resources Recycling and Recovery (CalRecycle), Disposal Reporting System (DRS). For statewide, http://www.calrecycle.ca.gov/lgcentral/goalmeasure/disposalrate/default.htm. For jurisdiction by facility, http://www.calrecycle.ca.gov/LGCentral/Reports/DRS/Destination/JurDspFa.aspx
    - Denominator: Total population estimates are available from the California Department of Finance, http://www.dof.ca.gov/research/demographic/reports/estimates/e-5/2011-20/documents/E-5_2014_Internet_Version.xls
- Years available: 2007, 2008, 2009, 2010, 2011, 2012, 2013
- Updated: annually
- Geographies available: cities/towns, county, region, state

Methodology:

Quarterly disposal tonnage data can be downloaded in Excel or csv format by jurisdiction (city/town, unincorporated county) and by disposal facility from the Disposal Reporting System for 20 jurisdictions at a time. Multiple files would need to be downloaded and concatenated for the numerator. The annual disposal tonnage can be calculated by aggregating quarterly data and convert. After pre-processing data will be imported into SAS and merged to population data to calculate the ratio. Standard error for the ratio and relative standard error can not be calculated. Deciles and relative risk can be calculated.

Limitations/comments:

According to CalRecycle, this indicator assesses how jurisdictions perform, compared to their own 50 percent per capita disposal target (i.e., jurisdictions are not compared to other jurisdictions or the statewide average). A comparison of the reported annual per capita disposal rate to the 50 percent per capita disposal target should be the yardstick for progress.

For comparisons over time, consider using 2007 and later years due to the definitional changes/reporting requirements implemented in 2007 (due to SB 1016).

The "per capita disposal" measured in 2007 and later years is the disposal (tons) divided by jurisdiction population (residents) or in some cases jurisdiction industry employment (employees) to obtain disposal by individual. Per capita disposal is adjusted to account for the maximum “transformation” credit (10%). "Transformation" means incineration, pyrolysis, distillation, or biological conversion other than composting. "Transformation" does not include composting, gasification, or biomass conversion. There is no specified biomass credit under SB 1016. However, materials diverted from a disposal facility to a biomass facility will result in a disposal reduction. **“**Disposal**"** means all waste created by all sources within each jurisdiction (including businesses, government agencies and residents) which is disposed at CalRecycle-permitted landfills or CalRecycle-permitted transformation facilities, or is exported from the state. CalRecycle tracks tons of waste disposed by each jurisdiction using its Disposal Reporting System (DRS). In some cases, the numerator uses “historical disposal” instead of “disposal.” DRS defines “historical disposal”, measured in tons, as the total tons landfilled in-state and exported that are reported at the time. “Historical disposal” does not include transformation.

Electricity from Renewable Sources

Full Title: Percent of Electricity from Renewable Sources

Indicator ID: 534

Healthy Community Framework: Quality and sustainability of environment

Aspirational Goal: Affordable and sustainable energy use

Rationale: Widely used, easily understandable indicator reported annually by the Independent System Operator for California. Indicator used to monitor California progress towards 33% renewable share of electricity generation by 2020.

Detailed Definition:

- Numerator: aggregate of renewable sources (biomass, geothermal, small hydroelectric, solar and wind generation) from total electricity system power, expressed in gigawatt-hours (GWh)
- Denominator: Total electricity system power (GWh) generated in-state
- $Indicator = \frac{Numerator}{Denominator} \times 100$
- Stratification: fuel type (natural gas, nuclear, coal, large hydroelectric, other, biomass, geothermal, small hydroelectric, solar and wind generation)
- Interpretation: healthier communities will have a higher percentage of energy from renewable resources

Data Description:

- Data Source: California Energy Comission, Energy Almanac (http://energyalmanac.ca.gov/), California Electricity Statistics & Data, Total Electricity System Power (http://energyalmanac.ca.gov/electricity/total_system_power.html)
- Years available: 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013
- Updated: Annually
- Geographies available: state

Methodology:

Data on system power will be web-scrapped from the Energy Almanac website using a SAS program. Data will be reported by type of fuel. The percent will be calculated using the in-state total electricity system power as denominator given that the fuel type of imports is not detailed in the table. The standard error of the percent can be estimated using the binomial approximation and the relative standard error, deciles, and relative risk can also be calculated.

Limitations/comments:

Due to changes in legislation, definitions and data collection methods for “unspecified sources of power”, comparisons across time can be made between 2009 and later years. Prior to 2009, data on unspecified sources of power were not collected. Instead, when the fuel type was unknown, the imported amount was multiplied by the percentages of a generation mix to estimate the amounts contributed by each fuel type. This generation mix was calculated using an averaging methodology based on the generation mixes of the Northwest and Southwest regions of the US. The estimated amounts by fuel type were then added to the corresponding fuel type categories, even though the fuel types were in fact unknown. The CEC described these procedures and estimates as, "not a good methodology to follow but at the time it was the only one available." Consequently, making comparisons from earlier years to 2009 and later years are not recommended (personal communication with Michael Nyberg, CEC, 9-18-14).

Because most solar photovoltaic systems on residential households and businesses are less than 1 megawatt, data on these installations is not collected. Data from the Total Electricity System Power "cannot be used to track the state's progress for the Renewable Portfolio Standard (RPS) program due to the intricacies, nuances, and special requirements of the RPS legislation."

The California Energy Commission (CEC) considers renewable energy sources to be "a power source other than a conventional power source within the meaning of Section 2805 of the Public Utilities Code. Section 2805 states: " 'Conventional power source' means power derived from nuclear energy or the operation of a hydropower facility greater than 30 megawatts or the combustion of fossil fuels, unless cogeneration technology, as defined in Section 25134 of the Public Resources Code, is employed in the production of such power (Final Committee Draft, Renewable Energy Program: Overall Program Guidebook, Third Edition. California Energy Commission, Efficiency and Renewable Energy Division. Publication number: CEC-300-2010-008-CMF)." At this time, electricity generation produced from renewable sources includes: biomass, geothermal, small hydroelectric, solar and wind generation. The category, small hydroelectric (hydroelectric plants producing at less than 30 MW capacity), did not exist prior to 2001 so the data will show as "N/A" for this fuel type (http://www.energy.ca.gov/hydroelectric/).

Total electricity system power is defined as the annual total energy requirement for all load serving entities with end-use loads in California, including self-generation supply for combined heat and power, and other non-utility served loads from power plants that are 1 megawatt and larger in nameplate capacity.

Data on energy generation by county can be accessed at the QFER CEC-1304 Power Plant Owner Reporting Database (http://energyalmanac.ca.gov/electricity/web_qfer/).

Exercise in Adults

Full Title: Percent of adults getting moderate/vigorous daily exercise

Indicator ID: 624

Healthy Community Framework: Meets basic needs of all

Aspirational goal: Access to affordable and safe opportunities for physical activity

Rationale: Strong scientific evidence links physical activity to improved health outcomes for cardiovascular disease, several types of cancers, diabetes, and mental health conditions. Measured biennially in adults for California counties in the California Health Interview Survey. Mentioned in the CDPH's 2010 California Obesity Prevention Plan 2010. Indicator may be stratified by race/ethnicity, income, and educational level to assess potential inequities.

Detailed Definition:

- Numerator: Number of adults getting moderate/vigorous daily exercise
- Denominator: Number of adults
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: seven race/ethnicity strata (African American, AIAN, Asian, Latino, Multiple, NHOPI, White, total)
- Interpretation: Healthier communities will have a higher percentage of adults getting moderate/vigorous daily exercise

Data Description:

- Data sources: California Health Interview Survey (CHIS), UCLA Center for Health Policy Research, Confidential data, http://healthpolicy.ucla.edu/chis/data/Pages/confidential.aspx
- Years available: 2005, 2007, 2009
- Updated: Unknown
- Geographies available: county, region (derived), state

Methodology:

SAS code provided by CHIS will be used as a template to obtain the estimate and its standard error using the replicate weights method (Proc Surveymeans). Based on information included in the data dictionary of the CHIS public use files, the variables of interest are "AE24: HARD/VIGOROUS EXERCISE IN PAST 7 DAYS" and A26: MODERATE PHYSICAL ACTIVITIES IN PAST 7 DAYS . Confidence intervals, relative standard error, deciles, and relative risk will be calculated.

Limitations/Comments:

2009 is the most recent year of data available from CHIS and it is unknown if the data will be updated in the future. Survey questions are added, removed, and modified in each cycle of CHIS. Questions are added to meet stakeholders' needs and to monitor emerging public health concerns. Questions are removed to reduce the length of the survey interview and save data collection costs when questions are no longer relevant for public health surveillance, or when they are not funded by a sponsor. Occasionally, changes are made to question wording based on methodological evaluations or user feedback that strongly suggests that changes will produce better data; otherwise, questions are kept consistent to aid in trending.

As a potential alternative to CHIS, the Behavioral Risk Factors Surveillance System (BRFSS) of the Centers for Disease Control and Prevention (http://www.cdc.gov/brfss/) has data on physical activity in adults for California (state only). Data can be disaggregated in five race/ethnicity groups (white, black, Hispanic, other and multiracial). For 2003, 2005, 2007, and 2009 the estimate is the percentage of "adults with 30+ minutes of moderate physical activity five or more days per week, or vigorous physical activity for 20+ minutes three or more days per week." For 2011 the estimate is the percentage of adults who "participated in 150 minutes or more of Aerobic Physical Activity per week."

Another potential data sources is the data in the County Health Rankings and Roadmap (http://www.countyhealthrankings.org/). The data available is for the "Percent of adults aged 20 and over reporting no leisure-time physical activity" at the county level. The data are based on "indirect model-dependent estimates. Bayesian multilevel modeling techniques are used to obtain estimates." The data source is the National Diabetes Surveillance System; the model uses three years of data from CDC's Behavioral Risk Factor Surveillance System (BRFSS) and data from the U.S. Census Bureau's Population Estimates Program.

Percent of Children Scoring 6/6 on Fitnessgram

Full Title: Percent of children scoring 6/6 on Fitnessgram

Indicator ID: 627

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Access to affordable and safe opportunities for physical activity

Rationale: Evidence links physical activity to improved educational outcomes, which are associated with improved health outcomes. Fitnessgram is the standardized measure of physical fitness for K-12 public school students published annually for school districts by the California Department of Education and may be stratified by race/ethnicity to assess potential health inequities.

Detailed Definition:

- Numerator: number of children scoring 6/6 on the Fitnessgram test
- Denominator: total number of children who took the test
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: seven race/ethnicity groups (African American, AIAN, Asian, Latino, multiple, NHOPI, and white); three grades (5th, 7th, and 9th).
- Interpretation: Healthier communities will have a larger percentage of children scoring 6/6/ on the Fitnessgram.

Data Description:

- Data sources: California Department of Education (CDE), Physical Fitness Testing (PFT) Program, http://www.cde.ca.gov/ta/tg/pf/pftresearch.asp
- Years available: 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013
- Updated: Annually
- Geographies available: School district, county, region (derived), state

Methodology:

Data on the number of children that obtained 6/6 in the Fitnessgram test, the total number of children who took the test, and the percent, can be downloaded from the PFT program research files, at the CDE website. The research files contain the results for the whole state by school, district, county, and the state. They also contain multiple race/ethnicity, gender, and income strata. The results are presented for three grades: 5th, 7th, and 9th. Five tables are included in the research files: 1) Fitness areas, 2) Meeting Health Fitness Zone (HFZ), 3) Gender summary, 4) Ethnicity summary, and 5) Economic summary. The table of interest for this indicator is Table 2, Meeting HFZ. No additional calculations are necessary for the indicator. Standard errors can be estimated using the binomial approximation; confidence intervals and relative standard error can be calculated using the binomial standard error. Place deciles and relative risk can be calculated.

Limitations/Comments:

The test is only administered to children in public schools to all students in grades five, seven, and nine, whether or not they are enrolled in a physical education class or participate in a block schedule. Students include those enrolled in elementary, high, and unified school districts, county offices of education, and charter schools. School districts should also test all students in alternate programs, including, but not limited to, continuation schools, independent study, community day schools, county community schools, and nonpublic schools. Students who are physically unable to take the entire test battery are to given as much of the test as conditions permit. (Education Code (EC) Section 60800 and the California Code of Regulations, Title 5, Section 1041). School districts have flexibility in the implementation of the methodologies for certain fitness areas (i.e., different measuring tools) which can produce unknown variation in the data. Details about the test can be found here: http://www.cde.ca.gov/ta/tg/pf/pftprogram.asp

The Fitnessgram is a fitness test developed by the Cooper Institute; it evaluates six fitness areas: 1) aerobic capacity, 2) body composition, 3) muscle strength, endurance and flexibility, 4) trunk extensor strength and flexibility, 5) upper body strength and endurance, and 6) flexibility. A passing result in all six areas (6/6) of the test represents a level of fitness that offers some protection against the diseases associated with physical inactivity. The 6/6 score does not allow for the study of the heterogeneities of how the different areas contributed to the overall score.

Forcible Rape Rate

Full Title: Number of Forcible Rapes per 100,000 Population

Indicator ID: 690

Healthy Community Framework: Social relationships that are supportive and respectful

Aspirational Goal: Socially cohesive and supportive relationships, families, homes and neighborhoods

Rationale: A widely used indicator of incidents reported to police (with concerns of significant underreporting). Compiled for cities and reported annually by U.S. Dept. of Justice (FBI). Monitoring statistics of this type is mentioned in the CDPH's California Statewide Policy Recommendations for the Prevention of Violence Against Women.

Detailed Definition:

- Numerator: number of forcible rape offenses known to law enforcement
- Denominator: total population
- $Indicator\ (rate) = \frac{Numerator}{Denominator} \times 100{,}000$
- Stratification: race/ethnicity stratification not available
- Interpretation: healthier communities will have lower (or zero) rates of forcible rape

Data Description:

- Data sources: Uniform Crime Reports (http://www.ucrdatatool.gov/Search/Crime/Crime.cfm), U.S. Department of Justice, Federal Bureau of Investigation.
- Years available: 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012
- Updated: Annually
- Geographies available: City, county, region (derived), state

Methodology:

Counts (and rates per 100,000 population) of forcible rape offenses know to law enforcement for California can be downloaded from the UCR website. Standard errors can be estimated using the Poisson approximation; relative standard errors, 95% confidence intervals, deciles and relative risk can be calculated. Data is reported by "Local Agencies" not cities or counties, therefore city and county names and geographical codes will need to be added manually following the same methodology as the Violent Crime Rate indicator.

Limitations/Comments:

Sexual assault is usually underreported. The FBI definition of forcible rape changed in December 2011, the effects of the definition will be seen in reported crime data until after January 2013. The FBI will not revise data from previous years.

UCR definitions of forcible rape:

Previous definition: The carnal knowledge of a female forcibly and against her will. Rapes by force and attempts or assaults to rape, regardless of the age of the victim, are included. Statutory offenses (no force used—victim under age of consent) are excluded.

Revised definition: Penetration, no matter how slight, of the vagina or anus with any body part or object, or oral penetration by a sex organ of another person, without the consent of the victim.

The revised definition:

- Includes either Male or Female Victims or Offenders.
- Includes instances in which the victim is incapable of giving consent because of temporary or permanent mental or physical incapacity, (e.g., due to the influence of drugs or alcohol or because of age).
- Reflects the various forms of sexual penetration understood to be Rape.

Sources:
http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2010/crime-in-the-u.s.-2010/violent-crime/rapemain
http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2012/crime-in-the-u.s.-2012/methodology

Children Reported with Neglect, Physical or Sexual Abuse

Full Title: Percent of children (under 18 years) reported with neglect or physical or sexual abuse

Indicator ID: 741

Healthy Community Framework: Social relationships that are supportive and respectful

Aspirational Goal: Socially cohesive and supportive relationships, families, homes and neighborhoods

Rationale: A standard indicator produced annually for Census tracts and counties by California Department of Social Services and UC Berkeley. Child abuse is linked to adverse health and social outcomes that endure into adulthood. May be stratified by race/ethnicity to assess equity.

Detailed Definition:

- Numerator: number of children suffering sexual abuse, physical abuse, severe neglect, general neglect, exploitation, or emotional abuse
- Denominator: population 0-17 years of age
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: six race/ethnicity strata (African American, Asian-NHOPI, AIAN, Latino, multiple, white)
- Interpretation: healthier communities will have lower (or zero) percent of children reported with physical or sexual abuse

Data Description:

- Data sources:
    - Numerator: California Child Welfare Indicators Project, UC Berkeley, http://cssr.berkeley.edu/ucb_childwelfare.
    - Denominator: 2000-2010 - Estimates of Race/Hispanics Population with Age and Gender Detail. http://www.dof.ca.gov/research/demographic/data/. Report P-3: Population Projections by Race/Ethnicity, Detailed Age, and Gender, 2010-2060, http://www.dof.ca.gov/research/demographic/reports/projections/P-3/.
- Years available: 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013
- Updated: Annually
- Geographies available: County, region (derived), state

Methodology:

Data is available from the California Child Welfare Indicators Project, Child Maltreatment Allegations/Child Count. Data tables for single years, single counties, with Census ethnic group, and allegation type (include all types of allegations) detail can be generated and downloaded one at the time (Excel download option available). For this indicator, 580 tables will be generated and downloaded: 58 counties × 10 years. Population data (0-17 years) is availble from the DOF population projections. Child maltreatment data and population data will be imported into SAS and joined to calculate the indicator. Standard error can be estimated using the Poisson approximation. Confidence intervals, relative standard error, relative risk, and deciles can be calculated.

Limitations/Comments:

The number of children with neglect or abuse could be underreported. The Child Welfare Indicators Project counts each child with a maltreatment allegation only once for each analysis year. If a child has more than one allegation in a specific year, they are counted one time in the category of the most severe occurrence. More methodological details can be found here http://cssr.berkeley.edu/ucb_childwelfare/Allegations.aspx, http://cssr.berkeley.edu/cwscmsreports/methodologies/default.aspx?report=Referrals.

Household Types

Full Title: Household by type of family and head of household

Indicator ID: 746

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Affordable, high quality, socially integrated and location-efficient housing

Rationale: Widely used description of community demographic structure. Indicator published by U.S. Census at 1-3 year intervals for most California cities and at census tract in 5 year intervals. Indicator recommended by the California Department of Housing and Community Development in county/local general plans. Indicator can be stratified by race/ethnicity to assess potential differences.

Detailed Definition:

- Numerator: number of households by type of family and head of household
- Denominator: total number of households
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: five types of family (married couple; married couple with own children under 18 years; female householder, no husband present; female householder, no husband present, with own children under 18 years of age; and nonfamily households). Eight race/ethnicities of the head of household (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white)
- Interpretation: households with two parents will generally have more economic resources that could conduce to better health outcomes for the family

Data Description:

- Data sources: American Community Survey (ACS), http://factfinder2.census.gov
- Years available: 2005-2007, 2008-2010, and 2006-2010
- Updated: 1, 3, and 5 year intervals
- Geographies available: Census tract, cities/towns, county, region (derived), state

Methodology:

The ACS table DP02 can be downloaded from http://factfinder2.census.gov. The types of family include: 1) married couple, 2) married couple with own children under 18 years, 3) female householder, no husband present, 4) female householder, no husband present, with own children under 18 years of age and 5) nonfamily households. The ACS DP02 table also includes the family types 6) male householder, no wife present, and 7) male householder, no wife present, with own children under 18 years. These last two types of family could be included in the file if considered necessary.

The denominator is the total number of households, and the numerator is the number of households by type of family. The percent can be calculated from these two estimates. The standard error of the percent can be calculated using the formula for the standard error for a proportion or percent (approximate method). Confidence intervals, relative standard error, deciles, and relative risk can also be calculated.

Limitations/Comments:

Family households and married-couple families do not include same-sex married couples even if the marriage was performed in a state issuing marriage certificates for same-sex couples. Same sex couple households are included in the family household category if there is at least one additional person related to the householder by birth or adoption. A nonfamily household is a householder living alone or with nonrelatives only. Same-sex couple households with no relatives of the householder present are tabulated in nonfamily households. The relationship categories for the most part can be compared to previous ACS years and to similar data collected in the decennial census, CPS, and SIPP. (Source: ACS Subject Definitions 2010)

According to the U.S. Census guidance for data users, the household/family types are not comparable between the 2000 and the 2010 5- year and 3- year ACS (http://www.census.gov/acs/www/guidance_for_data_users/comparing_2010/).

Neighborhood Completeness Index

Full Title: Neighborhood Completeness Index

Indicator ID: 747

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Complete and livable communities

Rationale: Geographic proximity to retail and public services are increasingly being used by researchers and commercial vendors to rate neighborhood environments. The NCI is based on the experience of the San Francisco Department of Public Health. Data are available for purchase from proprietary sources or require administrative permission from public agencies. Data require geocoding and frequent updating to avoid inaccuracy.

Detailed Definition:

- Numerator: population living in complete neighborhood
- Denominator: total population
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: eight race/ethnicity strata (African American, AIAN, Asian, Latino, Multiple, NHOPI, White, total). Two types of establishments (public services, retail or private establishments)
- Interpretation: healthier communities will have higher percentages of the population living in complete neighborhoods

Data Description:

- Potential government data sources: California Board of Equalization (BOE, http://www.boe.ca.gov). Population data at the Census block level for the year 2010, U.S. Census Bureau Redistricting File for California (http://www2.census.gov/census_2010/01-Redistricting_File--PL_94-171/California/).
- Potential proprietary data sources: InfoUSA (http://www.infousa.com/); Dun & Bradstreet (http://www.dnb.com/); Trade Dimensions TDLinx (a Nielsen company, www.nielsen.com);
- Years available: unknown (BOE); current year (proprietary data sources)
- Updated: annually (BOE); annually or semiannually (proprietary data sources)
- Geographies available: Census tract, city, county, region, state

Methodology:

The X-Y coordinates and NAIC code, or a geocoded GIS layer, for 12 types of retail/private establishments and 9 types of public services (see Tables 21 and 22) for all of California in a recent year will be requested to the BOE or obtained from a proprietary data source. For some of the public establishments data will be obtained from publicly available databases. Half mile “crow’s fly” buffers will be drawn around each establishment using ArcMap. Census block centroids will be spatially joined to each of the 21 types of establishments (centroid falling within the buffer). The results of the joins by type of establishment will be exported to SAS. The total number of retail establishments and public services by block will be calculated and the "complete" Census blocks will be identified: blocks with 9/12 retail establishments or higher and 7/9 public services or higher will be identified. The scored blocks will be joined to the population data. The population in complete blocks will be aggregated up to the Census tract level and the percent of tract population in complete blocks will be calculated. The standard error of the percent can be estimated using the binomial approximation. Relative standard error, confidence intervals, deciles, and relative risk can be calculated.

Table 21. Neighborhood Completeness Index Retail Establishments/Services Categories (N=12)

Retail Establishment Type	NAIC Codes, 2007
Auto repair shops	811111 General Automotive Repair 811112 Automotive Exhaust System Repair 811113 Automotive Transmission Repair 811118 Other Automotive Mechanical and Electrical Repair and Maintenance
Banks and credit unions*	521110 Federal Reserve Banks 522110 Commercial Banks 522120 Savings Institutions 522130 Credit Unions, Federally Chartered
Beauty and barber shops	812111 Barber Shops 812112 Beauty Salons/Shops- Unisex Hair Care
Bike repair shops	451110 New bicycles
Dry cleaners	812320 Dry Cleaning (on premises)
Eating establishments	722110 Restaurants (Full Service-With Wait Staff) 722211 Restaurants (Fast Food) 722212 Restaurants (Cafeterias) 722213 Restaurants (Non-Alcoholic Snack and Beverage)
Gyms	713940 Physical Fitness Facilities
Hardware stores	444130 Hardware Stores 444110 Home Centers
Laundromats	812310 Laundromats
Pharmacies	446110 Pharmacies and Drug Stores
Retail food markets	445110 Grocery Stores
Video rental stores/movie theaters	532230 Video Tape Rental 512131 Motion Picture Theaters

* (Check Cashing Agencies 522320, excluded) http://www.census.gov/econ/industry/#

Table 22. Neighborhood Completeness Index Public Services Categories (N=9). Note: some of the public services selected might be modified depending on data availability

Service Type	NAIC or Potential source	NAIC Codes, 2007
Post offices (from proprietary or publicly available database)	Postal delivery services, local, operated by U.S. Postal Service Postal delivery services, local, operated on a contract basis Postal services operated by U.S. Postal Service	491110
Public School (from proprietary or California Department of Education Database)	Elementary schools High schools Junior high schools Kindergartens Middle schools	611110
Medical Facilities (from proprietary or publically available database)	Walk-in physicians' offices MDs' (medical doctors), mental health, offices Medical doctors' (MDs), mental health, offices Mental health physicians' offices Pediatricians', mental health, offices Physicians', mental health, offices	621111
Medical Facilities (from proprietary or publically available database)	Hospitals, general medical and surgical	622110
Community Park or Playground	Access to Parks indicator, HCDIP	
Community garden	Data source not determined	
Library	California Public Library Inventory	
Recreation Center	Data source not determined	
Civic Spaces	Data source not determined	
Transit Stops	Transit Access indicator, HCDIP	

Limitations/Comments:

The HCDIP has met with the State Geospatial Information Office at the California Technology Agency to request his assistance in piloting geocoded establishment data from the Board of Equalization business establishment registration database. This information is required to construct several indicators related to neighborhood completeness/walkability, and availability of grocery and other food establishments. Pilot data obtained from the Board of Equalization's business establishment registrations suggests that the 4-digit coding detail using the North American Industrial Classification System (NAICS) may not be sufficient to identify different types of retail establishments. While further evaluation is necessary, it appears that, unless more detailed coding is available via BOE, data from commercial vendors who do additional detailed coding will be only alternative source for this type of establishment level data. The estimated cost for an enterprise license for all programs of the Department of Public Health is $100,000 (Personal Communication, Elizabeth Harris, InfoUSA, 2014).

Smoking in Adults and Youth

Full Title: Prevalence of smoking in adults and youth

Indicator ID: 755

Healthy Community Framework: Quality and sustainability of the environment

Aspirational Goal: Tobacco and smoke free

Rationale: Widely used indicator used by health care and state and county public health agencies, including CDPH Tobacco Control Program; understandable to wide range of users. Data collected biennially in California Health Interview Survey for counties and annually in the California Healthy Kids Survey at the level for school districts. Indicator may be stratified by race/ethnicity, income, and educational level to assess potential inequities.

Detailed Definition:

- Numerator: number of persons who smoke
- Denominator: total number of persons
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: eight race/ethnicity strata (African American, AIAN, Asian, Latino, Multiple, NHOPI, White, total), two age groups (adults -18 to 64 years of age; youth – 12 to 17 years of age).
- Interpretation: healthier communities will have lower (or zero) percent of persons who smoke.

Data Description:

- Data sources: California Health Interview Survey, UCLA Center for Health Policy Research, Confidential data, http://healthpolicy.ucla.edu/chis/data/Pages/confidential.aspx
- Years available: 2001, 2003, 2005, 2007, 2009, 2011-12
- Updated: CHIS is a continuous survey published every two years
- Geographies available: county, region (derived), state

Methodology:

SAS code provided by CHIS will be used as a template to obtain the estimate and its standard error using the replicate weights method (Proc Surveymeans). Based on information included in the data dictionary of the CHIS public use files, the variable of interest is “SMOKER: CURRENT SMOKER” for 2001 and “SMKCUR: CURRENT SMOKER” for all other years. Confidence intervals, relative standard error, deciles, and relative risk can be calculated.

Limitations/Comments:

SMKCUR is a constructed variable. Constructed variables are put together by the CHIS Data Production team or by Westat, Inc., the data collection contractor. Constructed variables are usually based on multiple questions asked during the interview. In general, constructed variables can be identified by their variable names, which are acronyms and/or abbreviations.

The SMKCUR variable is derived from questionnaire items AE15 and AE15A. If the adult indicated smoking every day (AE15A=1) or some of the days (AE15A=2) then the respondent was considered to be a current smoker (SMKCUR=1). If the respondent indicated never smoking more than 100 cigarettes in one's lifetime (AE15=2) or not smoking cigarettes daily (AE15A=3) then he or she was considered to be a non-smoker (SMKCUR=2). (Source: http://healthpolicy.ucla.edu/chis/data/public-use-data-file/Documents/CV2011-12_Adult_PUF_6.19.2014.pdf)

Survey questions are added, removed, and modified in each cycle of CHIS. Questions are added to meet stakeholders' needs and to monitor emerging public health concerns. Questions are removed to reduce the length of the survey interview and save data collection costs when questions are no longer relevant for public health surveillance, or when they are not funded by a sponsor. Occasionally, changes are made to question wording based on methodological evaluations or user feedback that strongly suggests that changes will produce better data; otherwise, questions are kept consistent to aid in trending.

As a potential alternative to CHIS, the California Healthy Kids Survey (CHKS, http://chks.wested.org/), WestEd and the California Department of Education (CDE), produce teen smoking data by county. However, "with the ending of Title IV funding, districts are no longer required to administer the CHKS every two years to be in compliance with No Child Left Behind (unless they have carryover funding). However, CDE is strongly encouraging districts to continue to administer the surveys at least once every two years."

Additionally, the California Department of Public Health, Tobacco Control Program conducts the California Student Tobacco Survey (CSTS) (http://www.cdph.ca.gov/programs/tobacco/Pages/CTCPEvaluationResources.aspx). The CSTS is a large-scale, in-school student survey of tobacco use. Surveys are collected for both middle (grades 6-8) and high school (grades 9-12) students. The sampling strategy includes stratification by 12 regions in California, with sample sizes that allow for the calculation of regional and some individual county estimates.

Patient Satisfaction

Full Title: Patient Satisfaction Rating by Medical Group

Indicator ID: 758

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Affordable, accessible and high quality health care

Rationale: Patient experience with the health care system is a widely accepted component of health care quality. Based on a validated, nationally standardized questionnaire of commercially insured managed care patients and reported annually at the medical group level within counties by the California Office of the Patient Advocate.

Detailed Definition:

- Numerator: number of adult members of commercial HMO health plans who responded using the most positive responses for each of six quality measures.
- Denominator: sample of adult (18-64 years of age) members of commercial HMO health plans who had at least one doctor visit during the year.
- The percentage of members in the sample who responded using the most positive responses for each quality measure is translated into a star rating as shown in Table 22. The indicator is the star rating by medical group.
- Stratification: medical group (varies by county)
- Interpretation: highly rated medical groups are considered to provide a high quality interaction between the patient and the health care system.

Data Description:

- Data sources: California Office of the Patient Advocate (OPA), Health Care Quality Medical Group Report Card http://reportcard.opa.ca.gov/rc/medicalgroupcounty.aspx
- Years available: 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013
- Updated: Annually
- Geographies available: county, state

Methodology:

The OPA produces report cards on medical group's quality of care with data from the California Healthcare Performance Information System, Inc. (CHPI), Patient Assessment Survey's (PAS) of patient experience. The PAS contains a set of six patient experience quality measures:

- Patients Rate Their Medical Group summary rating
- Communicating with Patients
- Timely Care and Service
- Helpful Office Staff
- Coordinating Patient Care
- Health Promotion

The survey is based on the industry-standard Consumer Assessment of Healthcare Providers and Systems (CAHPS), Clinician and Group Survey (CG-CAHPS©) with some customization for topics of local interest (http://reportcard.opa.ca.gov/rc/pdfs/2014%20Edition%20PAS_Final.pdf). Star rating data can be found in the OPA website and can be requested in machine readable format (http://reportcard.opa.ca.gov/rc/medicalgroupabout.aspx) however it might not contain the information necessary to calculate a standard error.

Limitations/Comments:

The OPA data is reported by medical group and by county, but no data is available at the respondent level thus there is no race/ethnicity detail for equity analysis. The data available is for commercially insured patients only. The PAS scoring method was revised in 2013 therefore medical group scores for that year are not comparable to prior years (http://www.chpis.org/programs/results.aspx). The 2013 scores are not available in the OPA website but are published in the website: http://www.calqualitycare.org/. The 2013 survey can be found here http://www.chpis.org/attachments/PAS_2013_English_PCP_1E.pdf. The cut points for the interpretation of the star rating of care as shown in Table 23.

Table 23. Cut points and Interpretation of the Star Rating of Care.

Star Rating	Percent Positive Responses	Grade	Performance
4 stars	70%-100%	Excellent	Comparable to the top 90% of scores for all the medical groups across the specific quality measure.
3 stars	63%-69%	Good	Comparable to scores between 50% and 89% percent of ratings across all the medical groups for the specific quality measure.
2 stars	54%-62%	Fair	Comparable to scores between 10% and 49% percent of ratings across all the medical groups for the specific quality measure.
1 star	0%-53%	Poor	Comparable to scores below the bottom 9% percent of ratings across all the medical groups for the specific quality measure.

A population based measure could be estimated if data on the coverage area by medical groups becomes available, although coverage areas could overlap and people at a particular location could seek for medical care outside the coverage area.

An alternative source of data for the indicator that would allow for population based measures is the CAHPS survey database (https://cahps.ahrq.gov/cahps-database/index.html) from the U.S. Agency for Healthcare Research and Quality (AHRQ). Data for the two components of the survey (Health Plan or Clinician and Group) can be requested for research purposes. Data is at the individual respondent level and includes "all information from the survey, except for respondent names, addresses, telephone numbers, and member ID numbers. These data files also include select information on survey respondents (gender, age, race, and self-reported health status). The CAHPS Database does not include any data that could be used to identify individual respondents. Certain survey administration data may also be included (e.g., mode of administration, survey language) and descriptive information (e.g., State)." Data requests are reviewed by a Management Committee with a review period of 1-2 weeks.

Resilient High School Students

Full Title: Percent of Resilient High School Students

*Indicator ID***:** 759

Healthy Community Framework: Adequate levels of economic, social development

Aspirational Goal: Support for healthy development of children and adolescents

Rationale: Indicator based on questionnaire responses of California public secondary school students surveyed annually by the California Department of Education in the California Healthy Kids Survey. Evidence links youth resilience measures to positive youth development and health outcomes. Because this survey module is optional, data may not be consistently available for all school districts.

Detailed Definition:

- Numerator: public school students in grades 7, 9, and 11 with high levels of total community assets (key factors promoting resiliency)
- Denominator: public school students in grades 7, 9, and 11
- $Indicator = \frac{Numerator}{Denominator} \times 100$
- Stratification: eight race/ethnicities (African American, American Indian/Alaska Native, Asian, Hispanic, Native Hawaiian/Pacific Islander, white, multiracial, other, total)
- Interpretation: healthier communities will have a higher percentage of school students with high levels of total community assets

Data Description:

- Data sources: California Healthy Kids Survey (CHKS, http://chks.wested.org/), WestEd and the California Department of Education (CDE); data available at http://www.kidsdata.org, Lucile Packard Foundation for Children's Health.
- Years available: 2008-2010
- Updated: unknown
- Geographies available: county, region (derived), state

Methodology:

Resilience is the ability for youth to strive in the face of adversity, and it is associated with healthy development and the avoidance of risky behavior, such as abuse of alcohol and other drugs, early sexual activity, and suicidal thoughts/actions. The California Healthy Kids Survey evaluates resiliency based on three key factors or assets that have been found to promote resilience in young people, (1) caring relationships with adults (supportive connections to others who model and support healthy development and wellbeing); (2) high expectations by adults (consistent communication of direct and indirect messages that youth can and will succeed); and (3) opportunities for participation and contribution (involvement of youth in relevant, engaging, and interesting activities with opportunities to contribute). The data available at www.kidsdata.org is the percent of public school students in grades 7, 9, and 11 with high levels for all three assets by county and by race/ethnicity. The percent of students with medium and low levels for all three assets is also available. Data disaggregated by asset is also available. No standard error information is available. Deciles and relative risk can be calculated.

Limitations/Comments:

With the ending of Title IV funding, school districts are no longer required to administer the CHKS every two years to be in compliance with No Child Left Behind (unless they have carryover funding). However, CDE is strongly encouraging districts to continue to administer the surveys at least once every two years. The Resilience and Youth Development module of CHKS is optional and is administered to a subsample of students taking the core modules.

Given the high likelihood that the data will not be updated for a large sample of the student population, the HCDIP could use “High school drop-out rates” or the “Percent of the population aged 18 - 24 years, with less than a high school education” as indicators of economic and social development. Data for the former is available from the Department of Education (http://dq.cde.ca.gov/dataquest/, Student Demographics, Dropouts) for school districts and counties, with race/ethnicity and grade detail, years 2002-2012. Data for the latter is available from the American Community Survey, table B15001, with Census tract and race/ethnicity data for the period 2006-2010.

Licensed Daycare Centers

Full Title: Number of licensed daycare center slots per 1,000 children aged 0-5 years

Indicator ID: 760

Healthy Community Framework: Adequate levels of economic, social development

Aspirational Goal: Support for healthy development of children and adolescents

Rationale: Indicator is associated with availability of childcare settings that may promote early childhood development, which in turn is associated with positive health and social outcomes. Increasing the availability of early child care settings is mentioned in the strategic plan of California Children and Families Commission (First Five). Licensing data from the Dept. of Social Services may be feasibly geocoded and combined with census tract demographic data.

Detailed Definition:

- Numerator: number of licensed daycare center slots
- Denominator: population 0-5 years of age
- $Indicator\ (rate) = \frac{Numerator}{Denominator} \times 1{,}000$
- Stratification: eight race/ethnicities (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white)
- Interpretation: healthier communities will have higher availability of early child care facilities

Data Description:

- Data sources:
  - Numerator: California Department of Social Services, Community Care Licensing Facility Search, https://secure.dss.ca.gov/CareFacilitySearch/home/selecttype/
  - Denominator: California Department of Finance, California and its Counties Population by Age, Race/Hispanics, and Gender: 2000–2010, http://www.dof.ca.gov/research/demographic/data/race-ethnic/2000-2010/index.php; U.S. Census Bureau, 2010 Census Summary File 1, Profile of General Population and Housing Characteristics, DP-1, http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml.
- Years available: 2014
- Updated: CDSS updates the public database weekly, population data with age and race breakdown for Census tracts and cities/towns is available every 10 years
- Geographies available: Census tract, cities/towns, county, region (derived), state

Methodology:

The addresses of currently licensed infant centers (serves ages 0-2) and child care centers (serves ages 2-5) are available online from CDSS. At present, data would need to be web scrapped using a SAS program. A downloadable spreadsheet might be available in the near future. The addresses will be geocoded using the California Environmental Health Tracking Service Web Based Geocoder. The geocoded locations will be imported into ArcMap to determine the Census tract were the centers are located. Location data by Census tract, city, and county will be imported into SAS and merged with population data of children 0-5 years to calculate the indicator. The standard error could be estimated using the Poisson approximation. Relative standard error, confidence intervals, deciles and relative risk can be calculated.

Limitations/Comments:

Daycare center data is from 2014 and population data is from 2010. Historical data of daycare centers might not be available. The center's address data should be downloaded at the same date every year (i.e. June 30th) to construct a time series.

Workers Employed in Artistic Occupations

Full Title: Percent of workers employed in artistic occupations

Indicator ID: 766

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Opportunities for engagement with arts, music and culture

Rationale: This indicator is used in national projects to assess the magnitude of investment in organized arts activities.

Detailed Definition:

- Numerator: number of employed civilians 16 years and over whose occupation is in the arts, entertainment, and recreation industries
- Denominator: total number of employed civilians 16 years and over
- $Indicator\ (percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: eight race/ethnicities (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white)
- Interpretation: healthier communities will have a higher percentage of workers in artistic occupations

Data Description:

- Data sources: U.S. Census 2000 SF4 (table DP-3); American Community Survey (table DP03), http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml.
- Years available: 2000, 2005-2007, 2008-2010, 2006-2010
- Updated: 3- and 5- year intervals
- Geographies available: Census tract, cities/towns, county, region (derived), state

Methodology:

The percent of the civilian employed population 16 years and over employed in the “Arts, entertainment, and recreation, and accommodation and food services industries” and its margin of error can be obtained from the U.S. Census Bureau (except no margin of error is available for Census 2000). Standard errors can be obtained from the margin of error; relative standard errors, confidence intervals, place deciles and relative risk can be calculated.

Limitations/Comments:

The NAICS code, Census code, and Category Titles for the :Arts, entertainment, and recreation, and accommodation and food services industries” have not changed since 1997 to 2012. Therefore data for the 2000 Census and ACS seems comparable for this category (other industry categories have changed due to the appearance and disappearance of some industries). (Source http://www.census.gov/people/io/files/2007-2012-Final-Census-Industry-Code-changes.xls).

There are other ACS and Census tables with related data. Table QT-P28 from the 2000 U.S. Census contains the percent of the civilian employed population 16 years and over employed in "Arts, design, entertainment, sports, and media occupations." Table B240550 of the ACS has the same percent for the "Arts, entertainment, and recreation," although there is no 5 year estimate.

Table 24. Industry Categories used by the U.S. Census Bureau that are Relevant for the Workers Employed in Artistic Occupations Indicator

NAICS code	Census code	Category title
711	8560	Performing arts, spectator sports, and related industries
712	8570	Museums, art galleries, historical sites, and similar institutions
71395	8580	Bowling centers
713 exc. 71395	8590	Other amusement, gambling, and recreation industries
7211	8660	Traveler accommodation
7212, 7213	8670	Recreational vehicle parks and camps, and rooming and boarding houses
722 exc. 7224	8680	Restaurants and other food services

Revenue in Nonprofit Arts Organizations

Full Title: Per capita revenue in nonprofit arts organizations

Indicator ID: 767

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Opportunities for engagement with arts, music and culture

Rationale: This indicator is used in national projects to assess the magnitude of investment in organized arts activities.

Detailed Definition:

- Numerator: total revenue in a fiscal year
- Denominator: total population
- $Indicator\ (ratio) = \frac{Numerator}{Denominator}$
- Stratification: none
- Interpretation: healthier communities will have higher investment in organized art activities

Data Description:

- Data sources:
    - Numerator: National Center for Charitable Statistics (http://nccs.urban.org/index.cfm), Search by Location tool.
    - Denominator: Total population data from the California Department of Finance, Demographic Research Unit.
- Years available: 2012. Revenue data for 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, and 2011 is available for purchase.
- Updated: annually
- Geographies available: Cities/towns, county, region (derived), state

Methodology:

Data on the total revenue for the most recent year available (2012) can be obtained for free for Census tracts, counties, primary metropolitan statistical areas, selected cities, and the state. Historical revenue data and custom tables can be purchased from NCCS for an estimated cost of $1,000.00 (personal communication with Jon Durnford, contractor for NCCS, jon@datalake.net). Population data for each geographical level could be obtained from the DOF, by year, for selected cities, counties and the state. Standard error cannot be calculated for this indicator. Deciles and relative risk can be calculated.

Limitations/Comments:

The NCCS already calculates "Nonprofit Public Charity Activities per Capita - Arts (2012)" for large statistical areas and counties; data for Census tracts is outdated. The NCCS Core Files include approximately 60 financial variables from the Form 990-series (Form 990, Form 990-EZ, Form 990-PF) for those organizations required to file. The Core File data is updated annually and NCCS has Core File data going back to 1989. The total revenue reported in Form 990 for all organizations is calculated by NCCS.

Housing to Jobs Ratio

Full Title: Housing to jobs ratio

Indicator ID: 768

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Affordable, high quality, socially integrated and location-efficient housing

Rationale: Widely used indicator of location-efficient housing. Housing shortages near job centers (and vice versa) promotes the use of personal passenger automobile in the absence of transit. Increased automobile use is associated with air-pollution, road traffic injuries, and greenhouse gas emissions.

Detailed Definition:

- Numerator: total job count
- Denominator: total occupied household count
- $Indicator\ (ratio) = \frac{Numerator}{Denominator}$
- Stratification: no race ethnicity information; three types of industry (farm, goods producing, service providing)
- Interpretation: healthier communities will have sufficient housing for workers

Data Description:

- Data sources: Current Employment Statistics Survey (CESS), Employment by Industry Data, California Employment and Development Department (CEDD), http://www.labormarketinfo.edd.ca.gov/LMID/Employment_by_Industry_Data.html, U.S. Census Bureau, American Community Survey, 1-year estimates, table DP04 (http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml).
- Years available: 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012
- Updated: annually
- Geographies available: county, region (derived), state

Methodology:

The county level data of CESS by industry can be downloaded (2000 to present by month) from the CEDD website. July centered employment estimates by industry for California counties can be calculated. Employment data can be joined to housing data and the ratio can be calculated. Standard error cannot be calculated for this indicator. Deciles and relative risk can be calculated.

Limitations/Comments:

There are multiple sources of employment data that differ in the survey tools used, excluded and included population, and geographical details. The Current Employment Statistics Survey from the Bureau of Labor Statistics is the official source of employment data. Table 25 compares sources of employment data.

Table 25. Comparison of Sources of Employment Data

Data source	Time detail	Geographical detail	Type of data	Not included
Census of Governments http://www.census.gov/govs/	Years 2002, 2007, 2012 (March)	County	State and Local government employees: Full-Time Employment; Part-Time Employment; Part-Time Hours; Full-Time Equivalent Employment; March Full-Time and Part-Time Employment;	Federal employees, private industry
Current Employment Statistics Survey, Employment by Industry Data http://www.labormarketinfo.edd.ca.gov/LMID/Employment_by_Industry_Data.html **Official employment statistics**	Years 2002-2014 by month	County, selected metropolitan areas	Total Industry Employment counts the number of jobs by the place of work. Reports number of employees by industry (SS-NAICS). Not seasonally adjusted. Survey methods: http://www.labormarketinfo.edd.ca.gov/LMID/Methodology_for_Industry_Employment.html Includes federal, state, and local employees.	This does not include business owners, the self-employed, unpaid family workers, or private household workers. If someone holds more than one job, they may be counted more than once.
Quarterly Census of Employment and Wages (QCEW), California Regional Economies Employment (CREE) http://www.labormarketinfo.edd.ca.gov/Content.asp?pageid=1016	2002-2012	County	Employment by NAICS code. Federal, state, and local employees' data from CES. Private industry data from QCEW; only employees covered under Unemployment Insurance programs.	Excludes employees not covered by unemployment insurance. Not official data.
Longitudinal Employer-Household Dynamics (LEHD), Quarterly	2002-2012 (4 Quarters)	County	Race/ethnicity, education, sex of workers by industry. Estimate of the	Excludes Federal government (includes state and local) workers and workers not

Workforce Indicators (WQI), http://lehd.ces.census.gov/data/#qwi Data from QCEW			total number of jobs on the first day of the reference quarter. Beginning-of-quarter employment counts are similar to point-in-time employment measures, such as the QCEW (see: www.bls.gov/cew/). Estimate of the number of jobs on the last day of the quarter.	covered by unemployment insurance. Examples of jobs that are not covered by UI include some agricultural jobs, railroad employment, self-employment, and other exceptions that vary from state to state.
Longitudinal Employer-Household Dynamics (LEHD), Origin Destination Employment Statistics (LODES), http://lehd.ces.census.gov/data/#qwi	2002-2011	Census block (very large files)	Three types of files: (1) number of workers by origin destination (workplace and residence census block), (2) number of workers by residence block with demographic characteristics and industry type; (3) number of workers by workplace with demographic characteristics by industry type.	Race ethnicity data only for 2009-2011; excludes federal workers except 2010-11.

Jobs: Housing Match

Full Title: Housing to jobs match

Indicator ID: 769

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Affordable, high quality, socially integrated and location-efficient housing

Rationale: In conjunction with housing to jobs ratio, provides additional information on whether the geographic distribution of job skills the work force in job centers is balanced with the skills of the residential population. This has implications on travel and whether the resident population has job skills commensurate with income to afford housing near job centers.

Detailed Definition:

- Numerator: total job count by industry
- Denominator: total employed residents by industry
- $Indicator\ (ratio) = \frac{Numerator}{Denominator}$
- Stratification: six race groups (white alone, African American alone, American Indian or Alaska Native alone, Asian alone, Native Hawaiian or other Pacific islander alone, two or more race groups) and two ethnicity groups (not Latino, Latino). Industry types (NAICS sectors 11, 21, 22, 23, 31-33, 42, 44-45, 48-49, 51, 52, 53, 54, 55, 56, 61, 62, 71, 72, 81 and 92).

Data Description:

- Data sources: Longitudinal Employer-Household Dynamics (LEHD), Origin Destination Employment Statistics (LODES), http://lehd.ces.census.gov/data/#qwi, Residence Area (RAC) and Workplace Area (WAC) Characteristic data.
- Years available: 2010, 2011
- Updated: annually
- Geographies available: Census tract, place, county, region (derived), state

Methodology:

This indicator will require intensive data aggregation given that all the demographic characteristics and industry information are in separate large files. The naming convention of the RAC (similar for WAC) files describes that there are 10 files for the different “segments of the workforce”, and 5 files for different job types:

RAC
Filenames of the RAC datasets are described by the following template:
[ST]_rac_[SEG]_[TYPE]_[YEAR]_1.csv.gz where
[ST] = lowercase, 2-letter postal code for a chosen state
[SEG] = Segment of the workforce, can have the values of “S000”, “SA01”, “SA02”, “SA03”, “SE01”, “SE02”, “SE03”, “SI01”, “SI02”, or “SI03”.
[TYPE] = Job Type, can have a value of “JT00” for All Jobs, “JT01” for Primary Jobs, “JT02” for All Private Jobs, “JT03” for Private Primary Jobs, “JT04” for All Federal Jobs, or “JT05” for Federal Primary Jobs.
[YEAR] = Year of job data. Can have the value of 2002-2011 for most states.

Data on workers place of residence (RAC) and workplace (WAC) characteristics will be downloaded and imported into SAS. Data will be aggregated from the block level to the tract level by race, ethnicity, and industry types. After aggregation, residence and workplace data will be merged and the ratio will be calculated. Standard error cannot be calculated for this indicator. Deciles and relative risk can be calculated.

Limitations/Comments:

There are multiple sources of employment data that differ in the survey tools used, excluded and included population, and geographical details. The Current Employment Statistics Survey from the Bureau of Labor Statistics is the official source of employment data. See Table 25 comparing employment data sources.

In the RAC and WAC data, a job is counted if a worker is employed with positive earnings during the reference quarter (Quarter 2, April – June) as well as in the quarter prior to the reference quarter. Worker data is limited to those who are covered by unemployment insurance. Federal workers are only included in LODES 6 and 7, in the 2010 and 2011 releases. LODES data is available from 2002-2011, but only LODES 7 uses 2010 Census blocks, all other versions use 2000 Census blocks. Sex, race/ethnicity and education details for workers are available in LODES 5, 6, and 7 (RAC and WAC files).

List of NAICS sectors in LODES (RAC and WAC files):

- 11 (Agriculture, Forestry, Fishing and Hunting)
- 21 (Mining, Quarrying, and Oil and Gas Extraction)
- 22 (Utilities)
- 23 (Construction)
- 31-33 (Manufacturing)
- 42 (Wholesale Trade)
- 44-45 (Retail Trade)
- 48-49 (Transportation and Warehousing)
- 51 (Information)
- 52 (Finance and Insurance)
- 53 (Real Estate and Rental and Leasing)
- 54 (Professional, Scientific, and Technical Services)
- 55 (Management of Companies and Enterprises)
- 56 (Administrative and Support and Waste Management and Remediation Services)
- 61 (Educational Services)
- 62 (Health Care and Social Assistance)
- 71 (Arts, Entertainment, and Recreation)
- 72 (Accommodation and Food Services)
- 81 (Other Services [except Public Administration])
- 92 (Public Administration)

Segments of the workforce:

- S000 Total number of jobs
- SA01 Number of jobs of workers age 29 or younger
- SA02 Number of jobs for workers age 30 to 54
- SA03 Number of jobs for workers age 55 or older
- SE01 Number of jobs with earnings $1250/month or less
- SE02 Number of jobs with earnings $1251/month to $3333/month
- SE03 Number of jobs with earnings greater than $3333/month
- SI01 Number of jobs in Goods Producing industry sectors
- SI02 Number of jobs in Trade, Transportation, and Utilities industry sectors
- SI03 Number of jobs in All Other Services industry sectors

Parkland Area

Full Title: Acres of parkland per 1,000 population

Indicator ID: 771

Healthy Community Framework: Quality and sustainability of environment

Aspirational Goal: Green and open spaces, including agricultural lands

Rationale: Widely used indicator for the amount of park land and open space available to residents of a geographic area.

Detailed Definition:

- Numerator: acres of parkland
- Denominator: total population
- $Indicator\ (rate) = \frac{Numerator}{Denominator} \times 1{,}000$
- Stratification: eight race/ethnicities (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white)
- Interpretation: healthier communities will have higher rates of parkland area per 1,000 population

Data Description:

- Data source:
    - Numerator: California Protected Areas Database (CPAD version 2014a), maintained by GreenInfo Network (http://www.calands.org/).
    - Denominator: Population data, year 2010, U.S. Census Bureau Redistricting File for California (http://www2.census.gov/census_2010/01-Redistricting_File--PL_94-171/California/).
- Years available: 2010
- Updated: 10-year intervals (Census data)
- Geographies available: Census tracts, cities/towns, counties, regions, and state

Methodology:

The CPAD includes open space lands including parks, as well as open space lands with other uses, including: recreation, forestry, historical/cultural, habitat conservation, water supply, scenic areas, flood control, agricultural/ranching, and general open space. The park area in acres (variable GIS_ACRES in CPAD) for a particular jurisdiction will be calculated using "Clip" ArcMap tool: the polygons with the geographical boundaries for Census tracts, places, and counties will be clipped with the park boundaries file and a new shapefile containing only the park polygons within jurisdictions will be created. The "Calculate Geometry" tool will be used to obtain the total number of park acres within a jurisdiction. These data will be joined with population data for each jurisdictions by race/ethnicity in SAS to calculate the indicator. Standard error cannot be calculated for this indicator. Deciles and relative risk can be calculated.

Limitations/Comments:

The California Protected Areas Database does not include tribal lands, lands used for active military purposes, and properties protected through easements. The indicator would not take into account the quality of park facilities, level of maintenance, specific amenities and services offered, or safety issues. The indicator would only include beach and coastline areas that are part CPAD, and known to be accessible to the public.

Neighborhood Change

Full Title: 5-year change in number of households by income and race/ethnicity (neighborhood change or gentrification).

Indicator ID: 772

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Affordable, high quality, socially integrated and location-efficient housing

Rationale: This is a measure of community succession and/or gentrification. Displacement of low-income and minority populations is negatively associated with adverse mental health outcomes and the disruption of social networks.

Detailed Definition:

- Numerator: 2000 to 2006-2010 difference in the number of households by income and race/ethnicity
- Denominator: 2000 number of households by income and race/ethnicity
- $Indicator(percent) = \frac{Numerator}{Denominator} \times 100$
- Stratification: eight race/ethnicities (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white), income categories (<$10,000, $10,000-$14,999, $15,000-$24,999, $25,000-$34,999, $35,000-$49,999, $50,000-$74,999, $75,000-$99,000, $100,000-$149,000, $150,000-$199,999, $200,000 or more)
- Interpretation: it is desirable for communities to have less income inequality and/or more ethnic diversity

Data Description:

- Data sources: U.S. Census Bureau, Census 2000, DP-3 table. Census 2000 5% Public Use Microdata Files (PUMS, https://www.census.gov/census2000/PUMS5.html). American Community Survey (ACS) 2006-2010 5-year file, table DP03 (http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml).
- Years available: 2000 to 2006-2010
- Updated: every 5 years
- Geographies available: Census tract, cities/towns, county, region (derived), state

Methodology:

The income distributions between Census 2000 and the 2006-2010 ACS will be compared by first converting the income distribution cutpoints to 2010 dollars using and adjusting factor (https://usa.ipums.org/usa/cpi99.shtml). Linear interpolation assuming continuous distribution of household along each of the income distribution categories will be used to obtain the number and percent of households by income and by race/ethnicity. The results obtained from the linear interpolation at the state level will be compared with income distributions obtained for 2000 using PUMS data to calibrate the interpolation. "PUMS files have state-level Census 2000 data containing individual records of the characteristics for a 5 percent sample of people and housing units." (https://www.census.gov/census2000/PUMS5.html) PUMS data and the geographic equivalency files that show the relationship between the PUMS and standard Census 2000 geographic concepts will be imported into SAS. The household income for each individual record will be converted to 2010 dollars using and adjustment factor (https://usa.ipums.org/usa/cpi99.shtml). The number and percent of households by race/ethnicity in 10 different income categories (<$10,000, $10,000-$14,999, $15,000-$24,999, $25,000-$34,999, $35,000-$49,999, $50,000-$74,999, $75,000-$99,000, $100,000-$149,000, $150,000-$199,999, $200,000 or more) and it's standard error by Census tract will be estimated (this analysis could be conducted in SAS or using the Data Ferret tool for PUMS data, http://www.census.gov/acs/www/data_documentation/data_ferrett_for_pums/). The 2000 estimates obtained from PUMS will be used as input to create estimates within 2010 tract boundaries using a publicly available tool that allows the bridging of data between the 2000 and 2010 Census tracts (http://www.s4.brown.edu/us2010/Researcher/LTDB1.htm). The results of the bridging will be compared with 2010 household estimates by race/ethnicity for the 10 income categories (table DP03) to calculate a percent change. Standard errors can be calculated using 2000 standard errors derived from the PUMS data; 2010 standard errors are available. The approximate method would be used to calculate the standard error of the percent change.

Limitations/Comments:

According to the U.S. Census "Direct comparisons of income and earnings distributions between Census 2000 and the 2010 ACS are not possible due to inflation. Users interested in making distribution comparisons need to inflation adjust individual income records using the Public Use Microdata Samples (PUMS) files from Census 2000." (http://www.census.gov/acs/www/guidance_for_data_users/comparing_2010/)

Updating of this indicator is only possible every 5 years after the release of the American Community Survey Special Population Tables that include mutually exclusive race/ethnicity categories. The next release (2011-2015) will be in 2016.

Residential Segregation

Full Title: Residential racial segregation: isolation index

Indicator ID: 773

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Affordable, high quality, socially integrated and location-efficient housing

Rationale: In meta-analyses of longitudinal epidemiologic studies, residential segregation has been associated with adverse mortality outcomes even after statistically adjusting for individual's race/ethnicity.

Detailed Definition:

- The isolation index is the minority-weighted average of the minority proportion in a geographical area (cities/towns, county, region, state).

$$\sum_{i=n}^{n}\left[\left(\frac{x_i}{X}\right)\times\left(\frac{x_i}{t_i}\right)\right]$$

n the number of Census tracts in the geographical area
x_i the minority population of Census tract i
t_i the total population of Census tract i
X the sum of all x_i (the total minority population)

- Stratification: eight race/ethnicities (African American, AIAN, Asian, Latino, multiple, NHOPI, other and white)
- Interpretation: The isolation index is the average proportion of people of a particular race/ethnicity in neighborhoods where people of that race/ethnicity live. Healthier communities will have lower racial segregation or index values closer to 1/number of race-ethnicities indicating the same proportion of each minority in a geographical area.

Data Description:

- Data sources: U.S. Census 2000, American Community Survey (ACS), http://factfinder2.census.gov
- Years available: 2000, 2006-2010
- Updated: 5 year intervals
- Geographies available: Cities/towns, county, region (derived), state

Methodology:

Census tract level data with race/ethnicity stratification can be downloaded from the American Fact Finder. The isolation index measures “the extent to which minority members are exposed only to one another” and is computed as the minority-weighted average of the minority proportion in each area. Isolation is a segregation measure referring to the degree of potential contact, or the possibility of interaction, between people of the same racial group. For instance, the black isolation index provides the average proportion of neighbors that are black, for the average neighborhood where blacks live (diversitydata.org). Data on the number of people by race/ethnicity will be added as columns in the data file for reference. Standard errors cannot be calculated for this index.

Limitations/Comments:

There are a large number of racial segregation indices; the U.S. Census provides a review document with formulas to calculate all indices. The U.S. Census does not include the race/ethnicity groups “Multiple” (two or more races) in the calculation of indices. Additionally, indices are only calculated for metropolitan areas with at least 10 tracts and 3 percent or 20,000 or more minority population. An adjustment formula needs to be applied to the index when more than two race ethnicities are present in one area. (Source U.S. Census)

Another index that could be calculated for this indicator is the exposure index. “Exposure is a segregation measure referring to the degree of potential contact, or the possibility of interaction, between group members of two racial groups within the average neighborhood of an area. For instance, the white-black exposure index provides the average proportion of neighbors that are black, for the average neighborhood where whites live. When the racial groups are the same (black-black exposure), this is same-group exposure, or isolation. In this example, it means that the average black person lives in a neighborhood where X proportion of his/her neighbors are black.” (Source diversitydata.org)

Three groups are using the isolation index: Harvard, http://diversitydata.sph.harvard.edu/, Michigan, http://enceladus.isr.umich.edu/race/racestart.asp, and U.S. Census https://www.census.gov/hhes/www/housing/housing_patterns/pdf/app_b.pdf.

Cities and Counties with Climate and Hazard Mitigation Plans

Full Title: Cities and counties with adopted climate action plans and FEMA-approved local hazard mitigation plans

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Able to adapt to changing environments, resilient, and prepared for emergencies

Rationale: Cities and counties are responding to the threat of climate change by developing strategies and written plans to reduce their carbon emissions and adapt and prepare for unavoidable consequences of climate change. Climate action and hazard mitigation plans are indicators of responding to this and traditional environmental threats.

Methodology:

Cities and counties with adopted climate actions plans are compiled by the Governor's Office of Planning and Research based on annual surveys. The listing of California counties with approved hazard mitigation plans is compiled by the California Office of Emergency Services (http://www.calema.ca.gov/HazardMitigation/Pages/FEMA-Approved-Local-Hazard-Mitigation-Plans.aspx) and the Federal Emergency Management Agency (http://www.fema.gov/multi-hazard-mitigation-plan-status).

Environmental Resilience Index

Full Title: Environmental resiliency, climate change adaptation, and emergency preparedness index

Indicator ID: 756

Healthy Community Framework: Meets basic needs of all

Aspirational Goal: Able to adapt to changing environments, resilient, and prepared for emergencies

Rationale: In addition to the planning processes in the preceding indicator, community resiliency has been measured by combining individual indicators across many domains of social vulnerabilities, proximity to hazards, adaptive capacity, and social capital. Because the Healthy Community Indicators already have such measures, a resiliency index will be composed from indicators form several of the domains of the HCI (unemployment, lacking health insurance, educational attainment, income inequality, and registered voters).

Detailed Definition: A resilience index composed of places with climate action and hazard mitigation plans and other HCDIP indicators

Data Description: Data will be obtained from previously constructed indicators. The indicator will be constructed at county and city/town levels if data is available. The most recent data available will be used for this index.

Methodology:

The methodology developed by Cutter et al. (2010, *JHSEM*, 7(1)) of a composite index will be applied. According to this methodology there are five components of disaster resilience and a composite index can be created with indicators that fall within each of these components. The following is a list of the resilience components and examples of HCDIP indicators within each component:

- Social Resilience: Educational attainment, adults without health insurance
- Economic Resilience: Unemployment rate, income inequality, percent of households incurring cost burdens
- Institutional Resilience: Number of cities and counties with climate and hazard mitigation plan
- Infrastructure Resilience
- Community Capital: Registered voters

To construct the index the indicators will be normalized, rescaled, and aggregated (equally weighted summation) within each composite. A second aggregation occurs over the composites to obtain a value from 0-5 that can be used to rank or compare cities, towns or counties in terms of their environmental resilience.

Limitations/Comments:
There is currently no infrastructure resilience indicator included in the HCDIP.

Race/Ethnicity and Place Based Equity Scores

Definition:

Inequities are differences in indicator values (outcomes) that are avoidable, unfair, preventable, and rooted in social position such as race/ethnicity, social/economic class, educational attainment, occupation, place (urban/rural), tribal status, gender, sexual orientation, or other social disadvantage.

Measurement of Equity:

We propose 3 scores for the equity dimensions of race/ethnicity, income, and place. Each score can be applied at the level of a city, county, or region by drawing on information from census tracts and individuals, depending on data availability.

The *race/ethnicity* composite is made up from scores of a selection of 15-20 individual core indicators (Table 26) that have data available for each race/ethnicity:

Table 26. Indicators Having Race/Ethnicity Detail

• Mode of transport to work
• Food affordability
• Daily fruit and vegetable intake in adults
• Percent of household crowding
• Household types
• Adults without health insurance
• Exercise in adults (county)
• Percent of children scoring 6 of 6 on fitnessgram (school district, county)
• Resilient high school students
• Academic Performance Index (API)
• Percent of household crowding*
• Educational attainment
• Smoking in youth (school district, county) and adults (county)
• Unemployment rate
• Poverty rate
• Children reported with neglect or physical or sexual abuse (county)

*No race/ethnicity information available as of 10-10-14.

For each indicator, a pair-wise comparison of different groups (e.g. African American vs. non-Hispanic White) will identify the race/ethnicity groups with the largest difference based on the highest and lowest indicator values. If the ratio of highest to lowest indicator value is statistically different than 1 (parity), then 1 point will be added to the composite score. This scoring of ratios is repeated for each indicator. It is still necessary to clarify how to create the composite for places that do not have data for all indicators or when race/ethnicities are defined differently between data sources. After the scores are summed and divided by the total number of indicators in the composite, the results can be expressed on scale from 0% to 100%, where higher scores indicate less race/ethnicity equity:

$$\text{Race/Ethnicity Equity Composite} = 100 \times \frac{\sum \begin{array}{c}\text{Count of indicators with significant}\\ \text{difference between highest and lowest groups}\end{array}}{\text{Number of Indicators}}$$

Income inequality is measured by the Gini coefficient, which is in wide use by economists, and is routinely reported at the city level in the American Community Survey. The Gini coefficient indicates the evenness of income distribution in a community on a 0 to 1 scale. A value of 0 indicates that each household generates an equal share of annual income and a value of 1 indicates that a small number (percent) of households generate nearly all of annual income of that community.

Using the same income data stratified by race/ethnicity, a complementary equity measure associated with *displacement/gentrification* can be created for a city, describing the change over time in the number (percent) of households by race/ethnicity and annual income in a specific geographic area.

The *place* composite is also made up of a score from a selection of the core indicators above (Table 27) and the following with data available for census tracts:

Table 27. Indicators Having Census Tract Detail

• Road traffic injuries
• Annual miles traveled by occurrence*
• Walkable distance to high quality public transit*
• Percent of household income spent on transportation
• Access to high quality food retailers
• Retail food environment index*
• Percent of households incurring cost burdens
• Neighborhood Completeness Index
• Proximity to busy roadways
• Access to parks
• Tree canopy coverage (urban areas)
• Toxic chemicals released to the environment
• Pesticide use
• Licensed daycare centers
• Voter registration and participation
• Children reported with neglect or physical or sexual abuse
• Poverty rate
• Educational attainment

*No Census tract data (estimate and/or standard error) available as of 10-10-14.

The place equity score for a city can be calculated from the value of indicators in each of the city's census tracts. The ratio of the highest to lowest indicator values will be taken from the top and bottom 20% (quintile) of census tracts. If the ratio of highest to lowest indicator value is statistically different than 1 (parity), then 1 point will be added to the composite score. After the scores are summed for each indicator and divided by the total number of indicators in the composite, the results can be expressed on scale from 0% to 100%, where higher scores indicate less place-based or neighborhood level equity within a jurisdiction.

$$\text{Place-Based Equity Composite} = 100 \times \frac{\sum \begin{array}{c}\text{Count of indicators with significant difference between}\\ \text{the highest and lowest quintiles of census tracts}\end{array}}{\text{Number of Indicators}}$$

<u>Pair-wise Comparisons for Race/Ethnicity Equity Composite:</u>

Pilot Study with Two Indicators

A pilot study of pair-wise race/ethnicity comparisons was conducted using the educational attainment and poverty rate (overall poverty) indicators. The poverty rate indicator used the data and the definition of the U.S. Census and is the percent of all individuals whose income in the past 12 months is below the poverty level. The educational attainment measures the percent of individuals 25 years or above, that have a high school degree or higher.

Pair-wise comparisons using the z-test were conducted to identify the race/ethnicity groups with the largest difference based on the highest and lowest indicator values. Data from the ACS 2006-2010 5-year estimate file was used for the comparisons. Tests were carried out using the method recommended by the U.S. Census using unpooled variances:

$$Z_{Place\ i} = \frac{Indicator_{RaceA,Place\ i} - Indicator_{RaceB,Place\ i}}{\sqrt{SE(Indicator_{RaceA,Place\ i})^2 + SE(Indicator_{RaceB,Place\ i})^2}}$$

.

If $Z_{Place\ i} < -1.645$ or $Z_{Place\ i} > 1.645$, then the difference in the indicator estimate between Race A and Race B is significant at the 90 percent confidence level.

All possible pair-wise comparison tests for all the race/ethnic groups available in a place were conducted, regardless of the reliability status of the indicator estimates (Note: a preliminary analysis showed that significant differences could be detected even when standard error was large in relation to the estimate.)

After all possible comparisons were conducted the two races with the largest significant difference were identified.

Preliminary Results of Pilot Study with Two Indicators

The pilot study found that 651 out of 1,523 places had data for two or more race/ethnicities and comparisons were possible for that subset of places (see Flowchart below). More places showed at least one significant difference for the poverty rate than for educational attainment. The number of significant differences was dependent on the number of race ethnicities available (see Table 28). One of the most common results for places was for Latinos to have the worst outcome and for Whites to have the best outcome, for both indicators (see Table 29). It is still necessary to contextualize the tests results with respect to the state average, given that significant difference between two race/ethnicities that have an educational attainment higher than 90% might not have the same relevance as higher magnitude differences (one race above and another race below the state average, for instance).

Results Overview Race/Ethnicity Pair-wise Comparisons

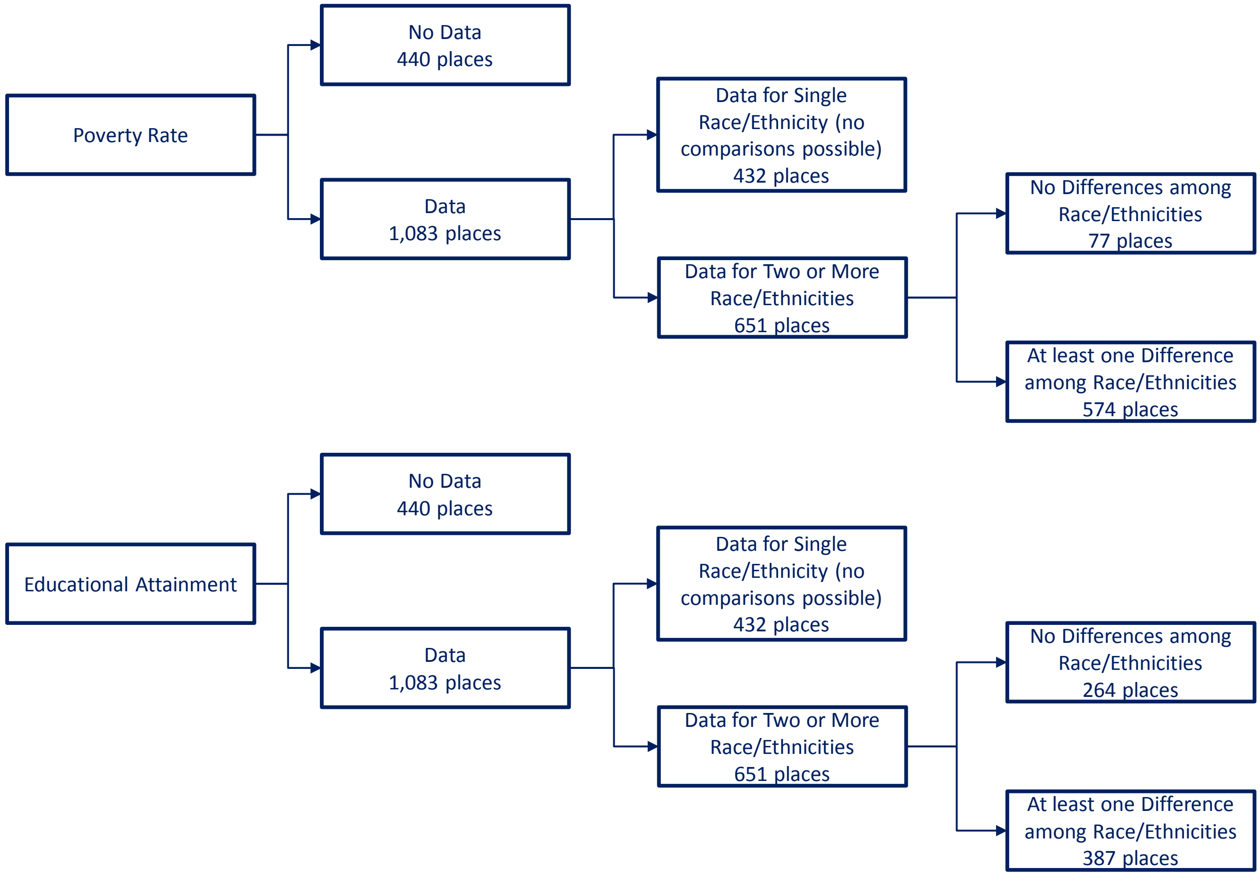


Table 28. Frequency of Significantly Different Pair-Wise Comparisons by Place and by Number of Race/ethnicities Available

	Number of race/ethnicities with data available in a place (number of possible comparisons)													
	Educational Attainment							Poverty Rate						
Number of significantly different pairs	**2 (1)**	**3 (3)**	**4 (6)**	**5 (10)**	**6 (15)**	**7 (21)**	**8 (28)**	**2 (1)**	**3 (3)**	**4 (6)**	**5 (10)**	**6 (15)**	**7 (21)**	**8 (28)**
0	51	19	6	1				174	47	27	16			
1	215	8	8					93	20	20	7			
2		58	9	1					38	14	18			
3		24	42	6					3	15	23	2		
4			19	37						15	32			
5			11	24	2					4	14	2		
6			1	36	2					1	27	1	1	
7				23	5						8	3	1	
8				18	1						3	6		
9				3	3						1	3		
10					4								2	
11					1	1						2		1
12					1	1							1	
13						2								
14							1							
15						1								2
16														1
17														1
18														1
19							1							
20							2							1
21							1							
22							1							
23														
24							1							
Total	266	109	96	149	19	5	7	267	108	96	149	19	5	7

Table 29. Frequency of Best and Worst Outcomes Based on the Largest Disparity between Two Race/Ethnicities

Largest disparity between:		Educational Attainment			Poverty Rate		
Best	Worst	Numerator	Denominator	Percent	Numerator	Denominator	Percent
African American	AIAN				1	18	5.6
	Asian	2	221	0.9	3	221	1.4
	Latino	91	252	36.1	15	251	6
	Multiple	1	180	0.6	4	180	2.2
	Other	1	13	7.7			
	White				1	244	0.4
AIAN	Latino	1	20	5.0	1	20	5
	White	0	20	0.0			
Asian	African American	1	221	0.5	21	221	9.5
	AIAN				1	18	5.6
	Latino	47	364	12.9	51	364	14
	Multiple				3	230	1.3
	Other				1	13	7.7
	White	1	380	0.3	12	380	3.2
Latino	African American				1	251	0.4
	Asian	0	364	0.0	5	364	1.4
	Multiple				3	235	1.3
	White	0	611	0.0	10	611	1.6
Multiple	African American				9	180	5
	Asian	2	230	0.9	10	230	4.3
	Latino	73	235	31.1	39	235	16.6
	White				2	234	0.9
NHOPI	African American				4	22	18.2
	Latino				5	22	22.7
Other	African American				2	13	15.4
	Latino				1	13	7.7
White	African American	3	245	1.2	32	244	13.1
	AIAN	0	20	0.0	4	20	20
	Asian	13	380	3.4	18	380	4.7
	Latino	336	611	55.0	119	611	19.5
	Multiple	2	234	0.9	7	234	3
	NHOPI				1	22	4.5
	Other				1	13	7.7

Numerator is the number of places where both race/ethnicities are present and significant differences were found. Denominator is the number of places where both race/ethnicities are present and a comparison is possible.

Pair-wise Comparisons for Place Equity Composite:

Pilot Study with Two Indicators

Similar to the race/ethnicity equity composite, a pilot study was conducted using the educational attainment indicator and the poverty indicator (overall poverty). The first step was to assign all Census tracts to places. Different methods were used to assign Census tracts to places given thtat the pilots for each indicator were conducted in different time periods. ArcMap was used to allocate 2010 Census tract centroids to place and then create a crosswalk that was used for the educational attainment indicator. This crosswalk used ct9 Census tract id's and could not be used for the poverty indicator. Data on Census blocks by place from the Redistricting files was used to obtain a crosswalk of Census tracts by place and used for the poverty rate indicator.

Only Census tracts with 500 or more people were included in the analysis and only data for all race/ethnicities combined was used. Data from the ACS 2006-2010 5-year estimate file was used for the comparisons.

Census tracts within a place with the highest and lowest percent of people (25 years or older) who have a high school degree or higher were identified and retained for further testing. Similarly for the poverty rate, the Census tracts with the highest and lowest percent of people in poverty were identified and retained.

Pair-wise comparisons using the z-test were conducted to identify if there was a significant difference between the two Census tracts identified. Tests were carried out using the method recommended by the U.S. Census using unpooled variances:

$$Z_{Place\ i} = \frac{Indicator_{CThigh,Place\ i} - Indicator_{CTlow,Place\ i}}{\sqrt{SE(Indicator_{CThigh,Place\ i})^2 + SE(Indicator_{CTlow,Place\ i})^2}}$$

.

If $Z_{Place\ i} < -1.645$ or $Z_{Place\ i} > 1.645$, then the difference in the indicator estimate between the two Census tracts is significant at the 90 percent confidence level.

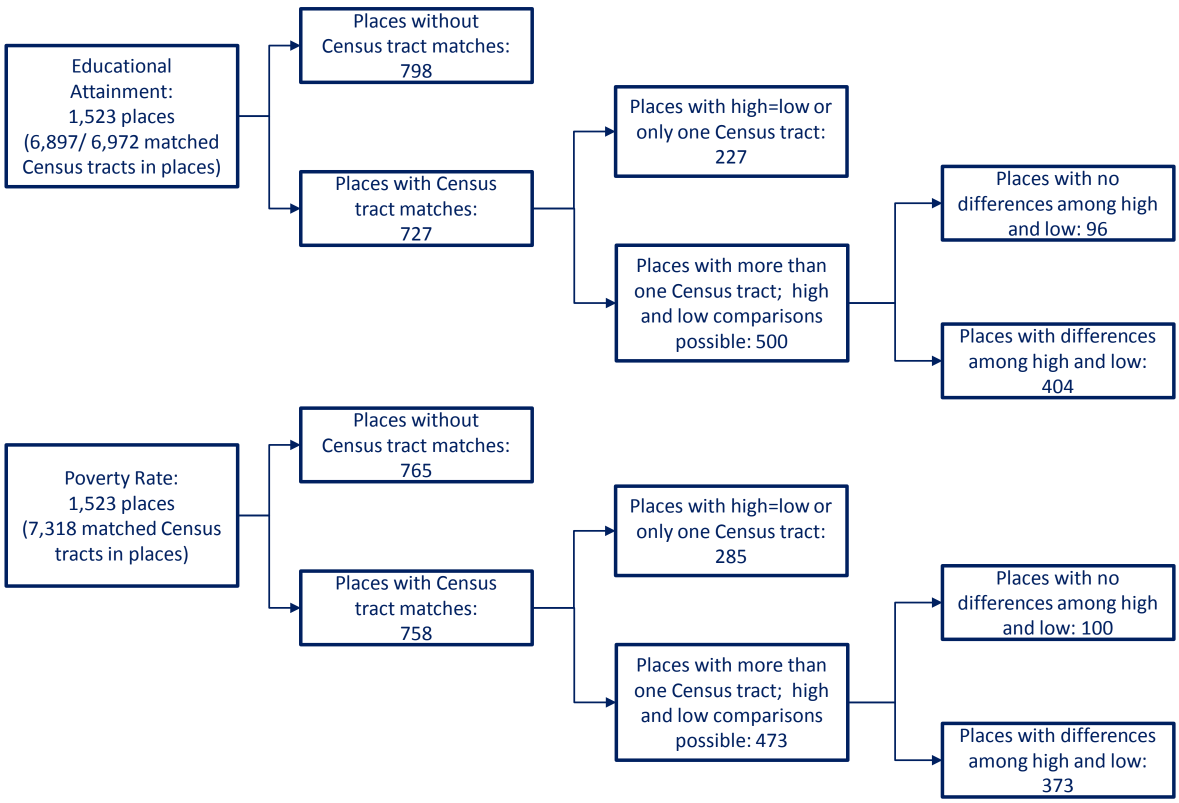
Preliminary Results of Pilot Study with Two Indicators

For the educational attainment indicator, we found that 500 out of 1,523 places had data for two or more Census tracts that could be compared with a test; 404 of those places had significant differences based on the Z test. For the poverty indicator, comparisons between Census tracts were possible for 473 places and 373 of those places showed significant differences using the Z test.

For educational attainment, the average percent significant difference between two Census tracts -when data was available- was 24.2% (min=3.6% for Cupertino, max=79.9% for Los Angeles). For poverty rate, the average percent significant difference between two Census tracts was 20.0% (min=3.9% for El Dorado Hills, max=69.0% for Atascadero).

Similar to the race/ethnicity comparisons presented above, it is still necessary to contextualize the tests results with respect to the state average, given that significant difference between two tracts that have an educational attainment higher than 90% might not have the same relevance as higher magnitude differences.

Results Overview Census tracts in Places Pair-wise Comparisons



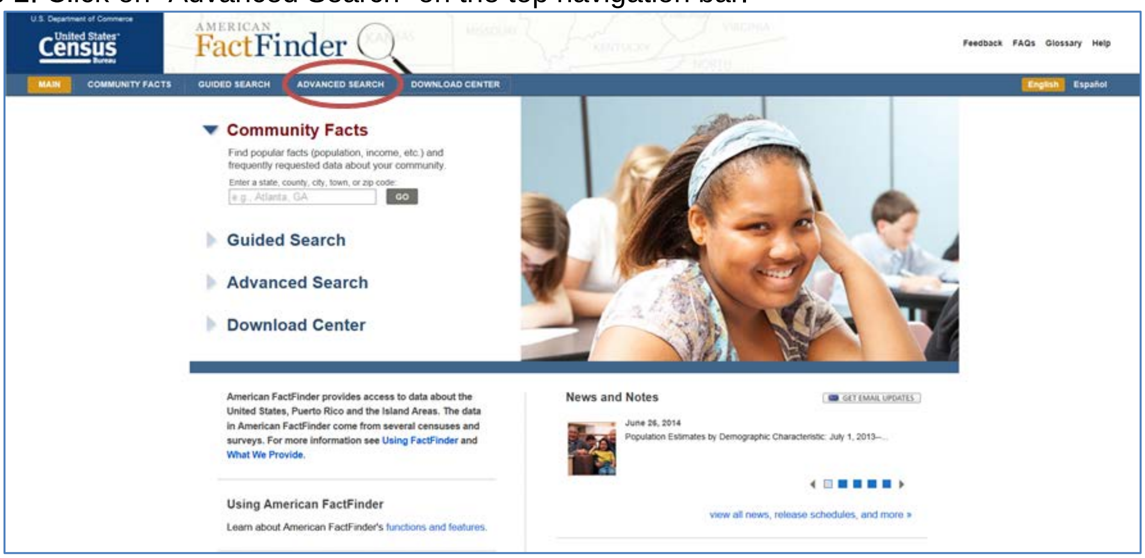
TOOLBOX

Downloading Census Data from American FactFinder

The following steps to download Census data are for a single table (DP03) from a single dataset (2010 ACS 5-year Selected Population Tables), at the county level geography and for 8 race/ethnicity groups and total population. The race/ethnicity groups selected are the standard for the HCDI project. Some tables will not have race/ethnicity information. The DP03 table used in this example contains information on multiple topics. In this example it is downloaded for the "Mode of Transportation to Work" indicator, and therefore the topic selected is "Commuting". These steps can be followed for all other Census tables under other topics. Please be aware that websites are prone to modification; the appearance shown here might not exactly match the appearance of the website in the near future. Use the steps here when helpful.

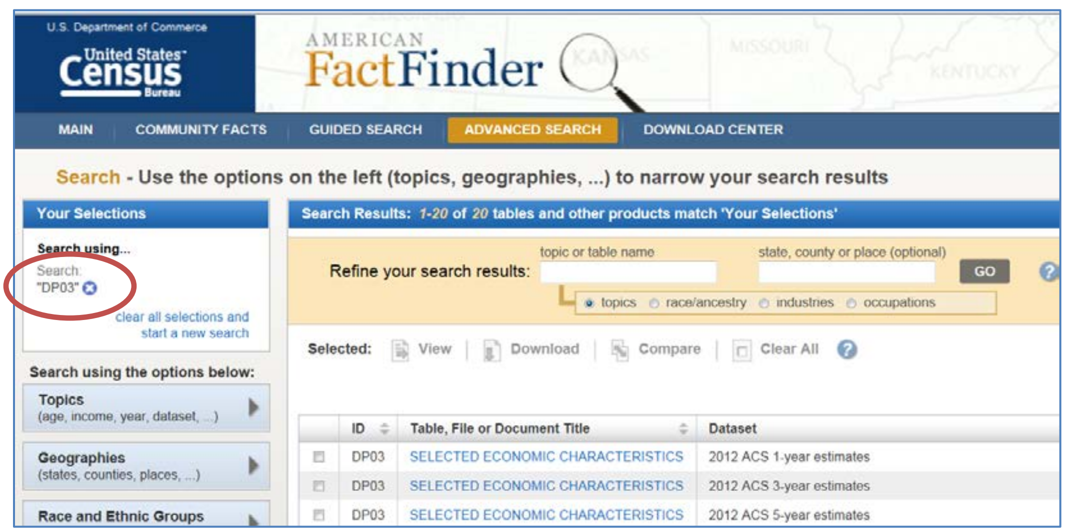
Step 1. Go to http://factfinder2.census.gov/ (accessed 8-5-14).

Step 2. Click on "Advanced Search" on the top navigation bar.

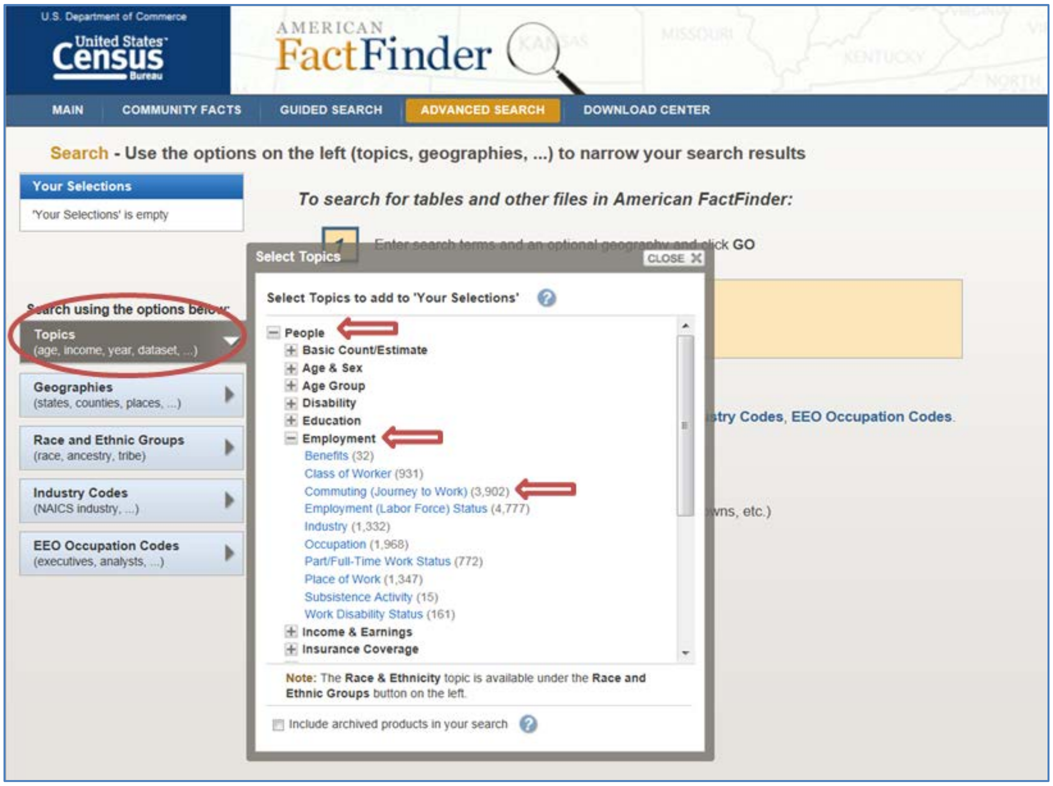


Step 3. If the number of the table of interest is known (DP03), type it in the "topic or table name" search box and click "GO".

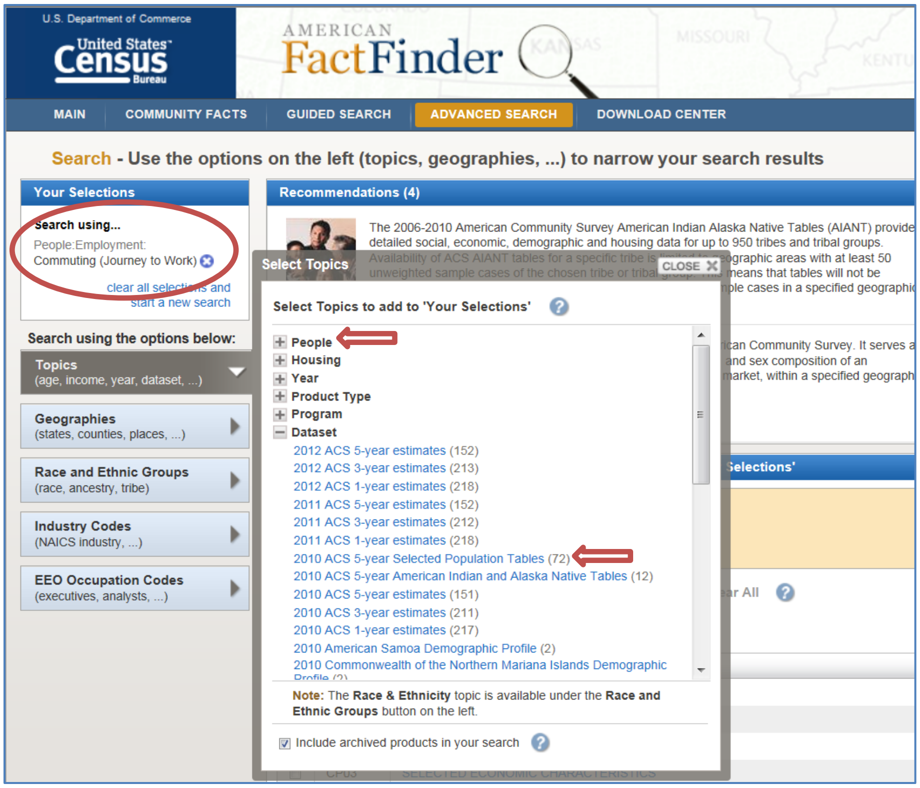
"DP03" will appear in the box "your selections" in the upper left corner of the screen.



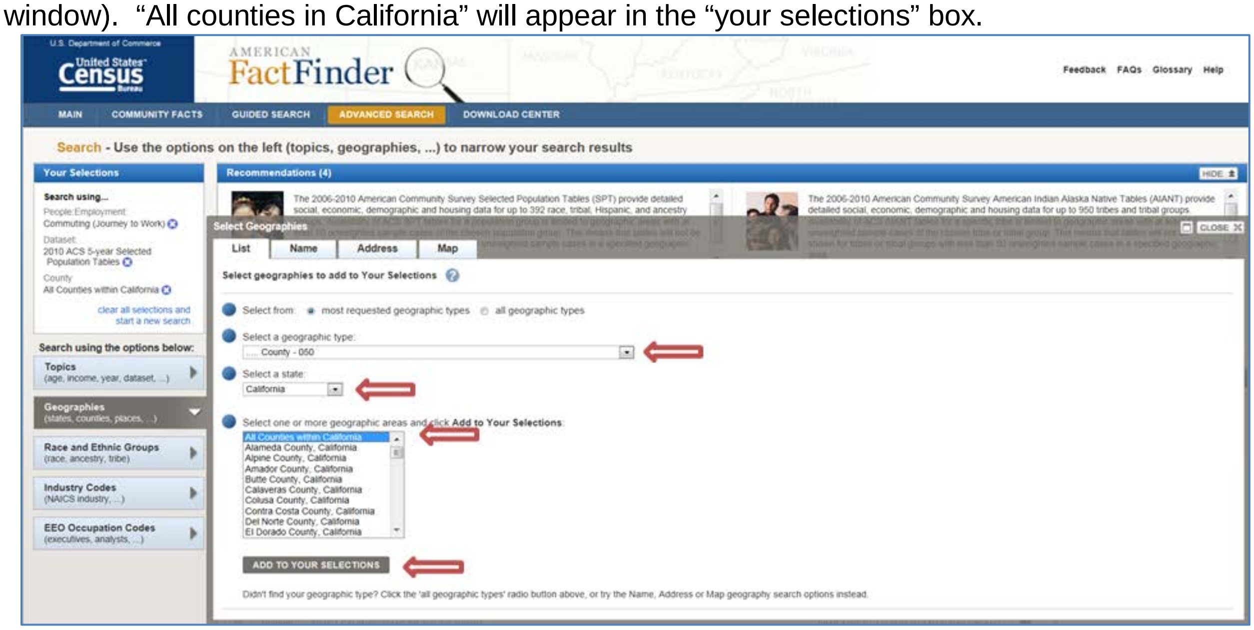
Step 4. If the table number is not known, it is possible to search by topic by typing a search term ("Commuting", for instance) in the "topic or table name" box (same as above) and click "GO". Alternatively, to search the topics that the Census has available, select the "Topics" box on the tool bar on the left hand side; a new window will appear. Then, select "People," then "Employment" and finally "Commuting (Journey to Work)." Click "Close" on the upper right corner of the new window. "Commuting (Journey to Work)" will later appear in the box "your selections" in the upper left corner of the screen. Note: When using the topic search options, multiple tables will appear in the results; it will be necessary to study the tables and select the table that contains the information needed.



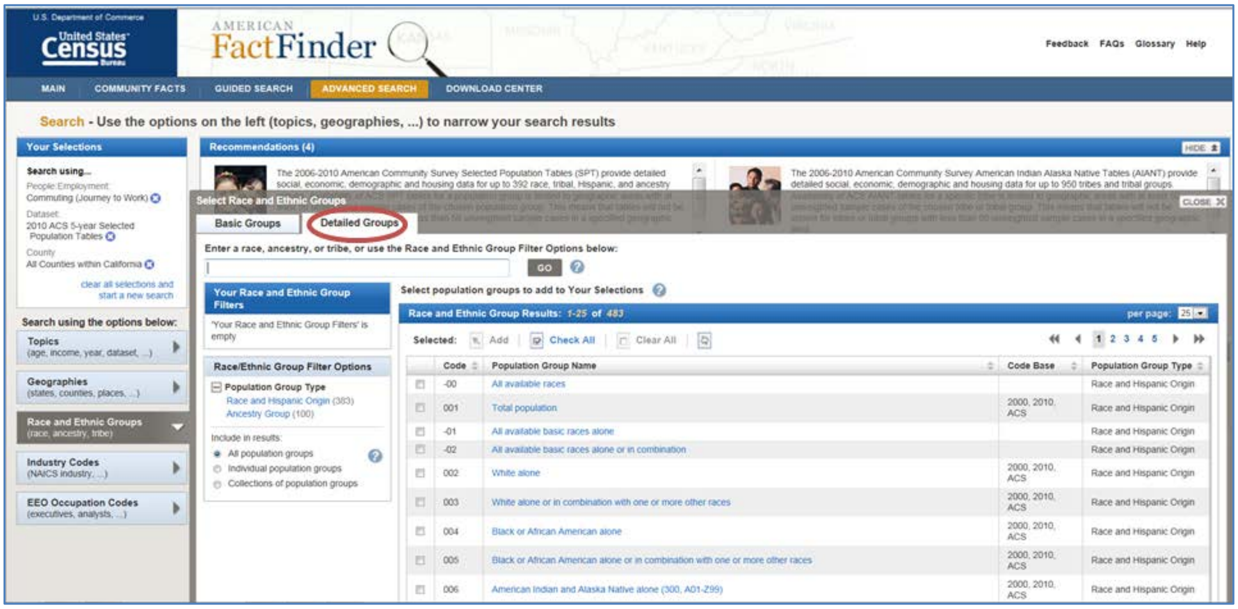
Step 5. The next step is to select the Census dataset (time period) for the selected table. Click on the "Topics" box to the left, scroll down the window and click on "Dataset," and then on "2010 ACS 5-year Selected Population Tables," which will later appear in the "your selections" box. Click "Close" on the upper right corner of the new window. Note: if older datasets do not appear, check "Include archived products in your search" at the bottom of the new window.



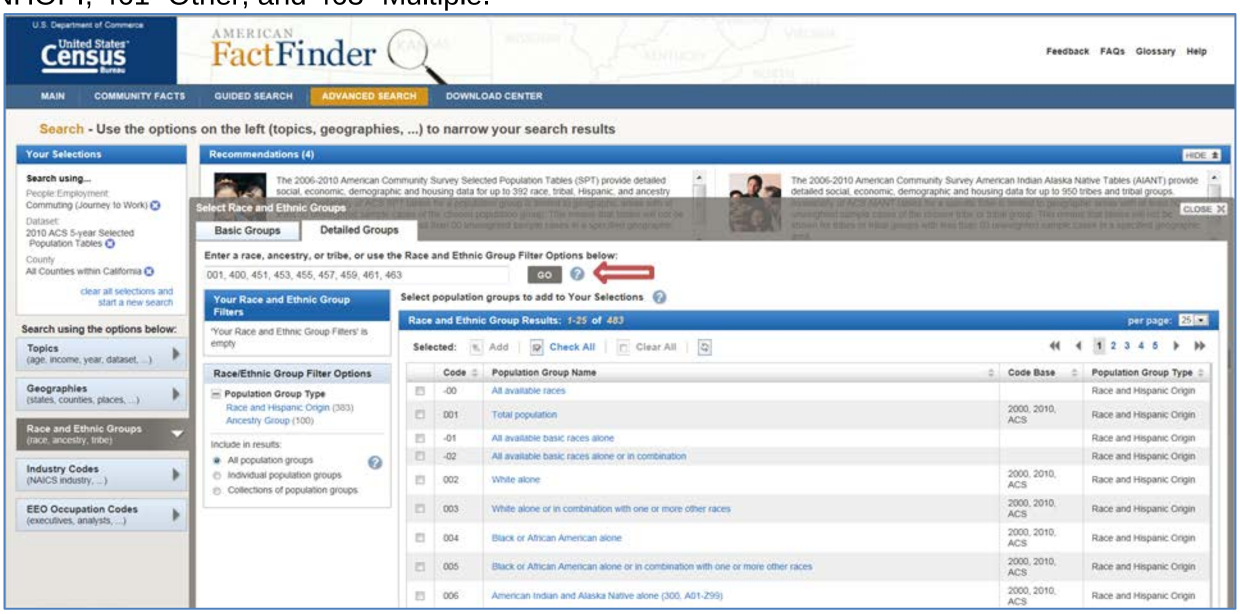
Step 6. On the tool bar on the left hand side select the "Geographies" box, a new window will appear. From the dropdown list "Select a geographic type", pick County – 050. From the dropdown list "Select a state", pick California. Finally, click on the phrase "All counties within California" that is inside the "Select one or more geographic areas and click add to your selections" box. Click "Add to your selections" and then click "Close" (upper right corner of the new window). "All counties in California" will appear in the "your selections" box.



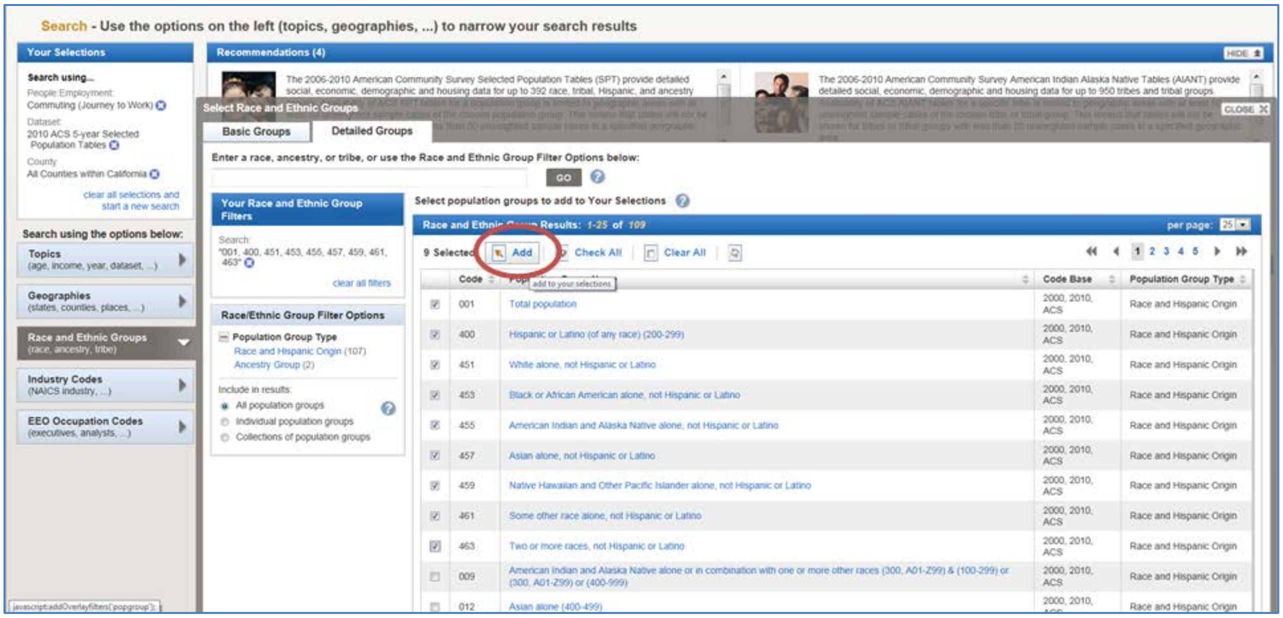
Step 7. On the tool bar on the left hand side select the “Race and Ethnic Groups” box, a new window with a list of race/ethnicity groups and their codes will appear. Go to the “Detailed Groups” tab.



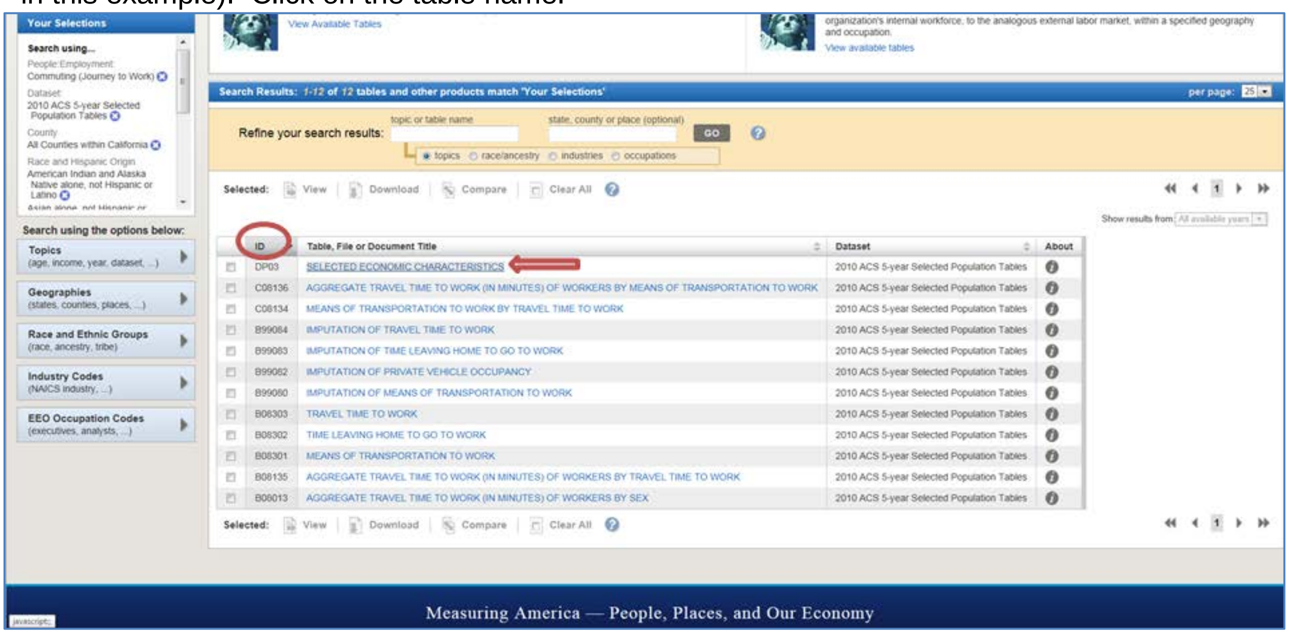
Step 8. Under the “Enter a race, ancestry, or tribe, or use the Race and Ethnic Group Filter Options below:” box, enter the following list of codes: “001, 400, 451, 453, 455, 457, 459, 461, 463”, and click “GO”. These codes correspond to the following race/ethnic groups: 01-Total, 400- Hispanic/Latino, 451- White, 453- Black, 455- American Indian/Alaska Native, 457- Asian, 459- NHOPI, 461- Other, and 463- Multiple.



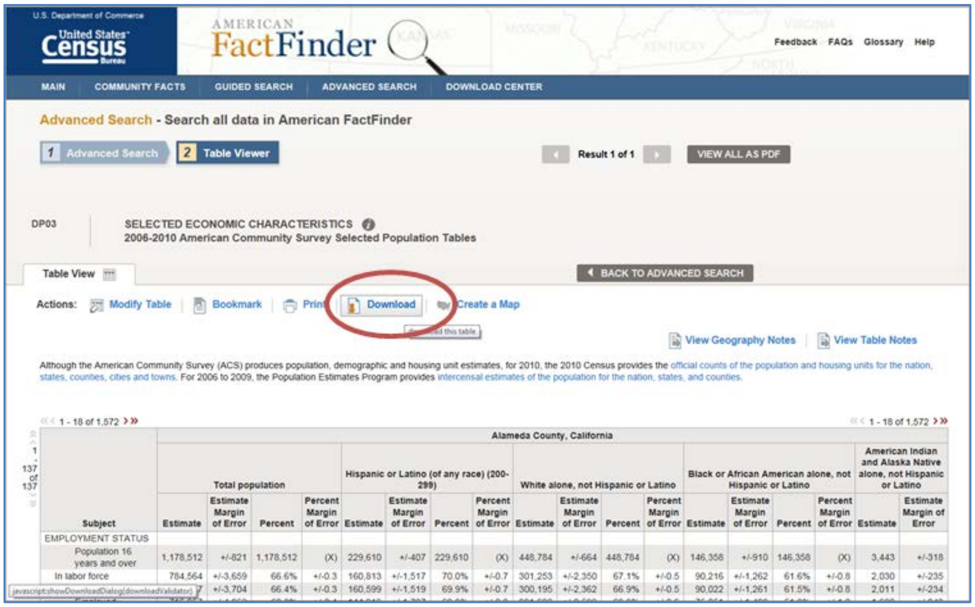
Step 9. A reduced list of race and ethnic groups will appear. Click on the empty boxes next to the codes to select the Population Group Name with codes 001, 400, 451, 453, 455, 457, 459, 461, and 463. Click "Add" at the top of the list. Close the "Select Race and Ethnic Groups" window (upper left corner). The list of the 9 selected race/ethnicities should appear in the “your selections” box.



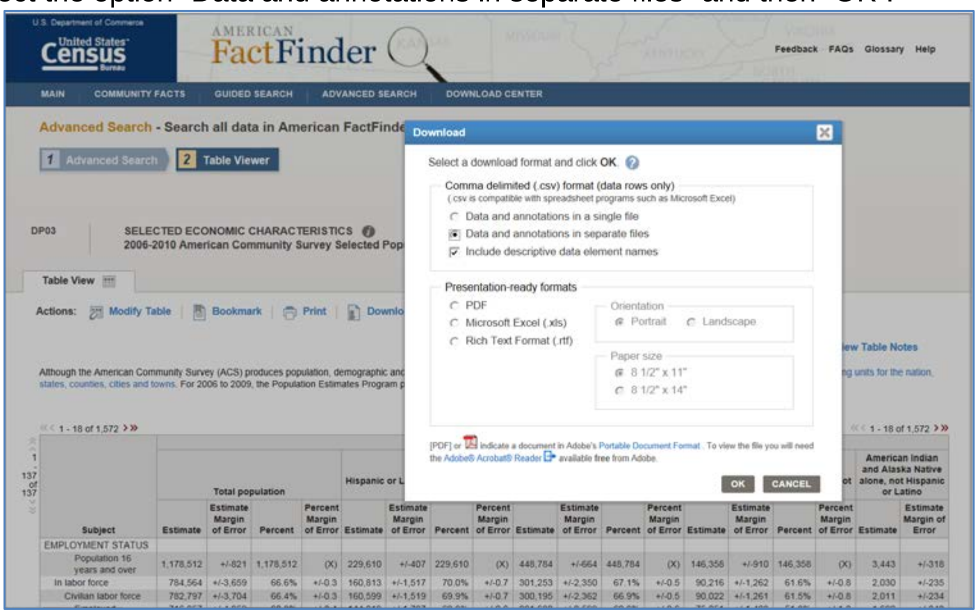
Step 10. Sort the resulting group of tables using the “ID” field and pick table of interest (DP03 in this example). Click on the table name.



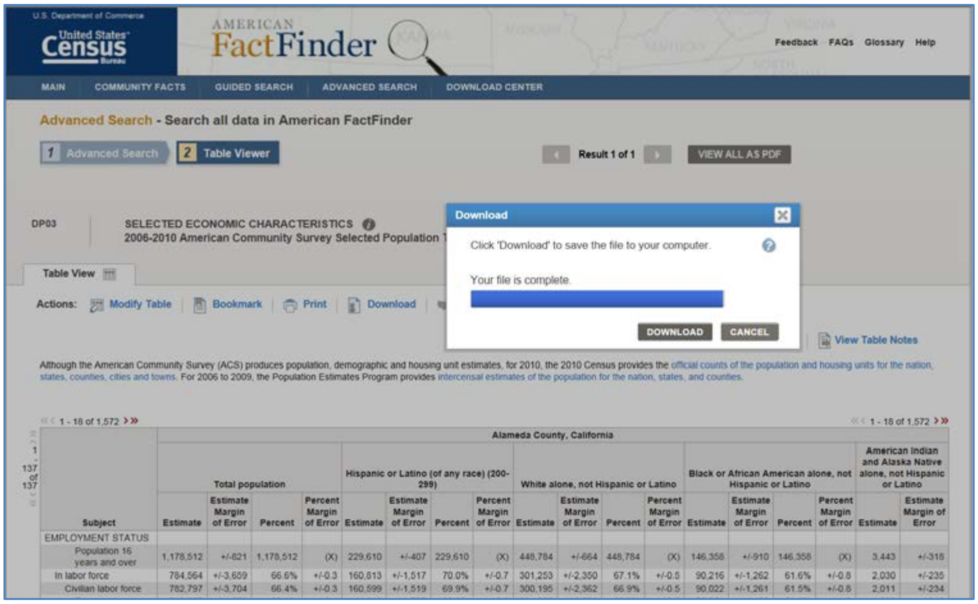
Step 11. The data will be displayed. Click on “Download” at the top of the table. A new window will appear.



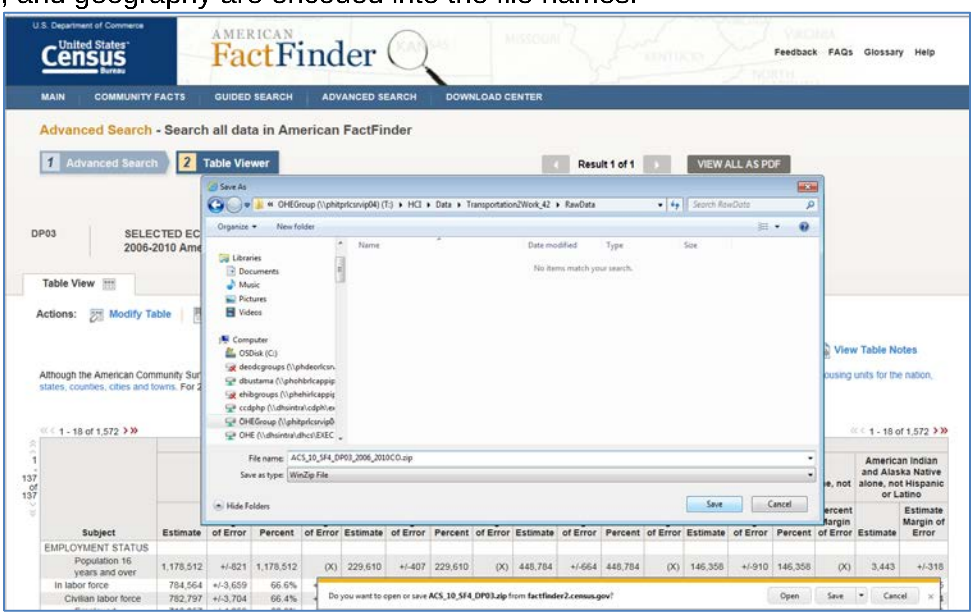
Step 12. Select the option “Data and annotations in separate files” and then “OK”.



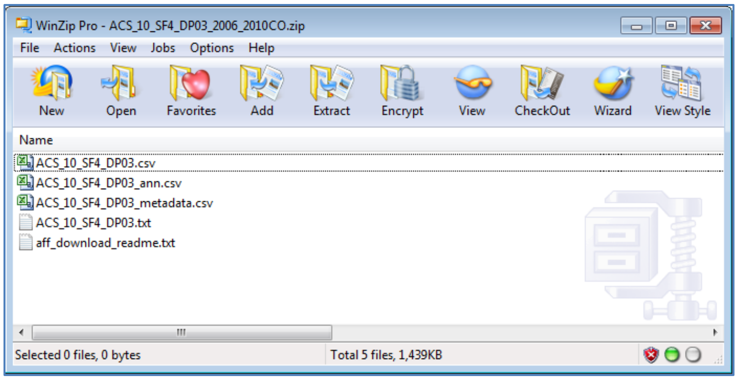
Step 13. A downloadable table will be created. When it is complete click “Download”.



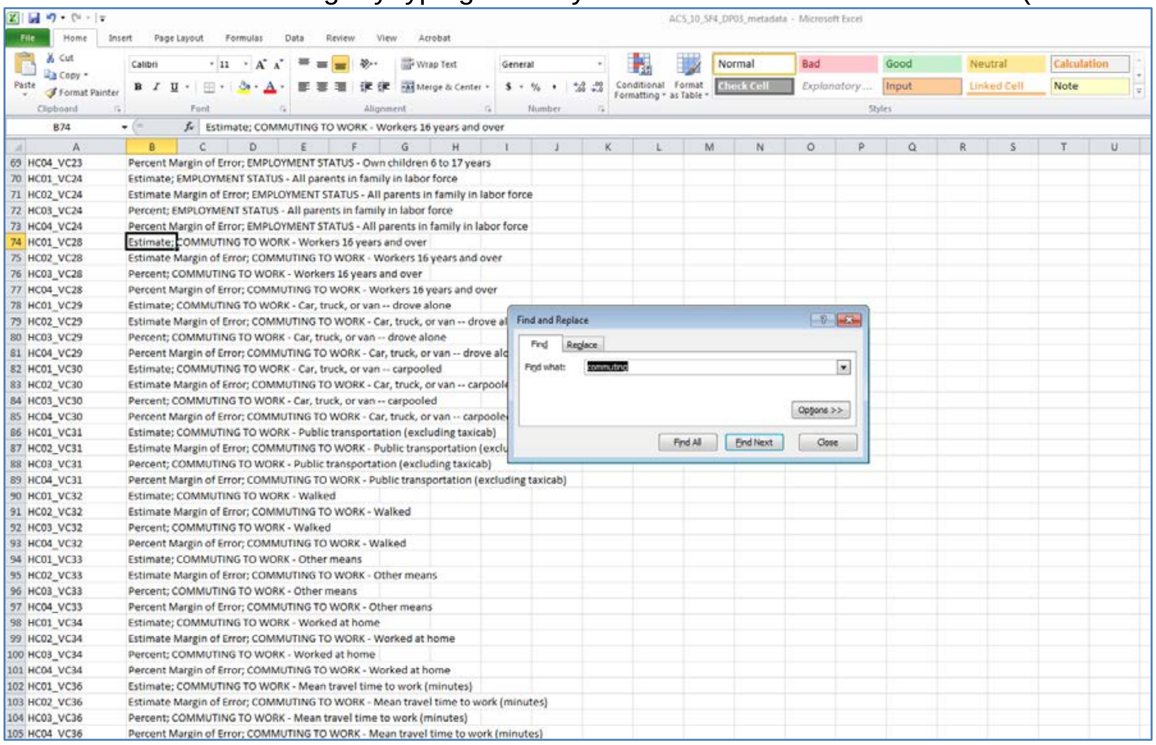
Step 14. Save the table by selecting the “save as” option and rename it to indicate the date (2006-2010) and geography (CO): ACS_10_SF4_DP03_2006_2010CO.zip, for instance. This step is especially necessary if more than one table will be downloaded because factfinder2 gives the same name to all tables of the same format. For the HCDIP, Census table name, report year, and geography are encoded into the file names.



Step 15. Open the zip file and extract the files ACS_10_SF4_DP03.csv and ACS_10_SF4_DP03_metadada.csv. The metadata file contains the information explaining the names of the variables in the data file.



For instance, for the "Transportation to work" indicator, one can search all the variables related to "Commuting" by typing that key word into the "Find" tool in Excel (see below).

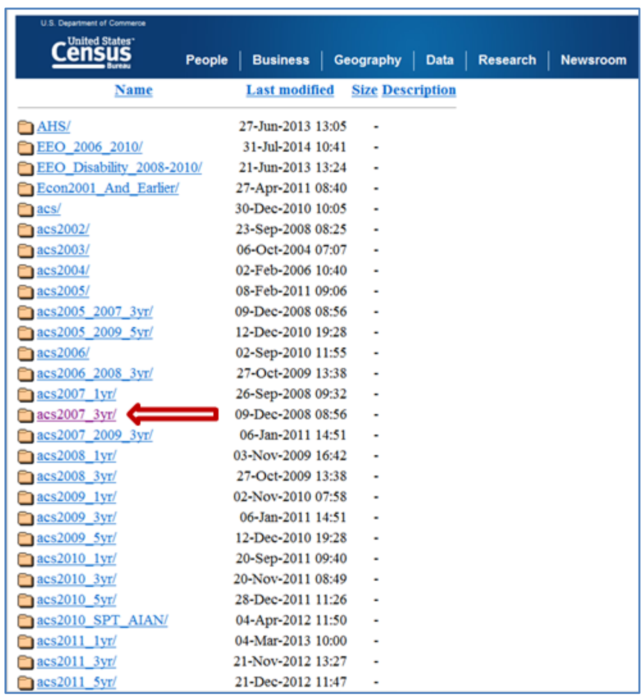


Obtaining American Community Survey Data from Summary Files

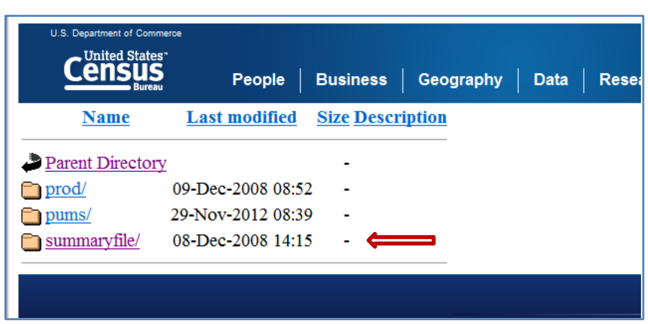
An ACS Summary File is a set of comma-delimited text files that contain all of the Detailed Tables for the ACS 1-year, 3-year or 5-year estimates (Detailed Tables include those in the Bxxxxx and Cxxxxx series). For example, the 2008 ACS 1-year Summary File includes hundreds of comma-delimited text files that are best viewed in spreadsheet or using statistical software (http://www.census.gov/acs/www/data_documentation/summary_file/). Summary Files and supporting files are available at the Census FTP site: http://www2.census.gov/. Using summary files requires SAS programming knowledge in order to understand, use and/or modify the SAS code provided by the Census to extract the tables.

Example: Downloading All Summary Files for ACS 3-year 2005-2007

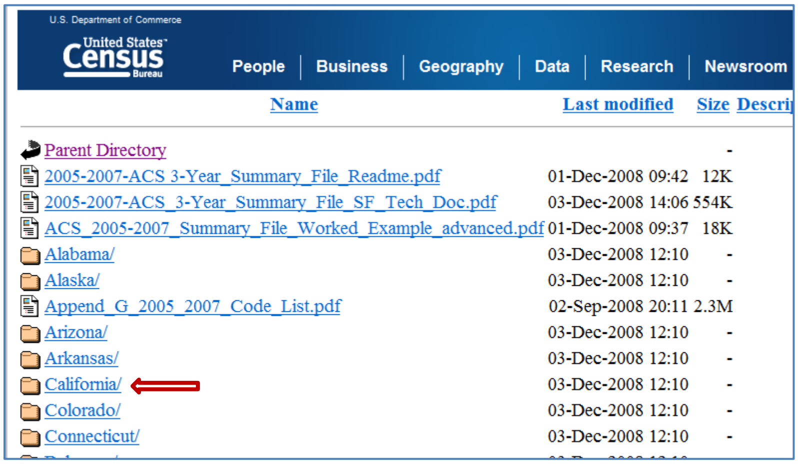
Step 1. Go to http://www2.census.gov/.
Step 2. All ACS summary files folders are labeled as “acs” followed by the most recent year included in the folder, and the number of years aggregated. For instance, the folder acs2007_3yr includes the aggregate for the years 2005, 2006, and 2007. Click on the “acs2007_3yr” folder:



Step 3. Click on the “summaryfile” folder:

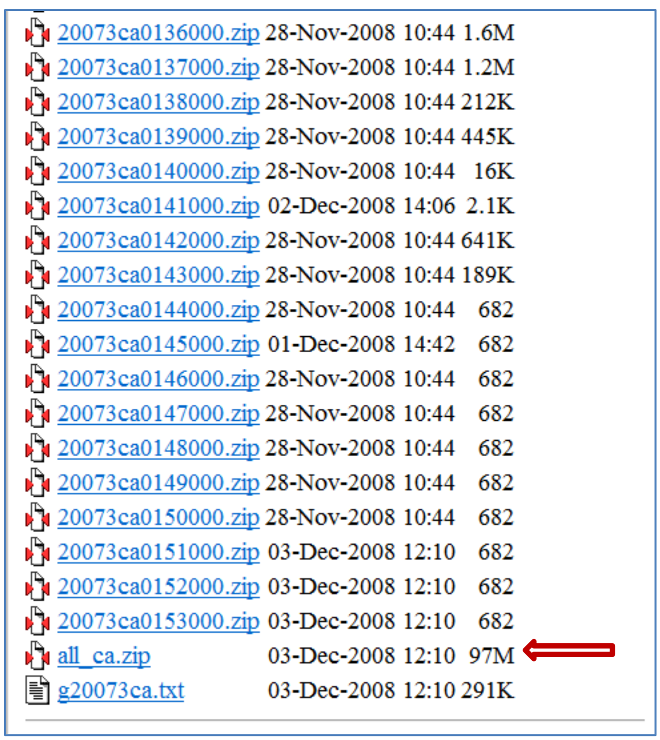


Step 4. Click on the “California” folder:



Name	Last modified	Size	Descri
Parent Directory		-	
2005-2007-ACS 3-Year_Summary_File_Readme.pdf	01-Dec-2008 09:42	12K	
2005-2007-ACS_3-Year_Summary_File_SF_Tech_Doc.pdf	03-Dec-2008 14:06	554K	
ACS_2005-2007_Summary_File_Worked_Example_advanced.pdf	01-Dec-2008 09:37	18K	
Alabama/	03-Dec-2008 12:10	-	
Alaska/	03-Dec-2008 12:10	-	
Append_G_2005_2007_Code_List.pdf	02-Sep-2008 20:11	2.3M	
Arizona/	03-Dec-2008 12:10	-	
Arkansas/	03-Dec-2008 12:10	-	
California/	03-Dec-2008 12:10	-	
Colorado/	03-Dec-2008 12:10	-	
Connecticut/	03-Dec-2008 12:10	-	

Step 5. To download all summary files for California, scroll until finding the “all_ca.zip” file, click on it and save to the computer. Note that when Census tract level data is available, there will be a separate “all_ca_CT.zip” file. Note: It is recommended that file folders, named by survey source and years, be created prior to downloading the summary files to prevent possible overwriting and assist in file identification. An example of a folder name would be ACS_SummaryFiles_2005-2007.



20073ca0136000.zip	28-Nov-2008 10:44	1.6M
20073ca0137000.zip	28-Nov-2008 10:44	1.2M
20073ca0138000.zip	28-Nov-2008 10:44	212K
20073ca0139000.zip	28-Nov-2008 10:44	445K
20073ca0140000.zip	28-Nov-2008 10:44	16K
20073ca0141000.zip	02-Dec-2008 14:06	2.1K
20073ca0142000.zip	28-Nov-2008 10:44	641K
20073ca0143000.zip	28-Nov-2008 10:44	189K
20073ca0144000.zip	28-Nov-2008 10:44	682
20073ca0145000.zip	01-Dec-2008 14:42	682
20073ca0146000.zip	28-Nov-2008 10:44	682
20073ca0147000.zip	28-Nov-2008 10:44	682
20073ca0148000.zip	28-Nov-2008 10:44	682
20073ca0149000.zip	28-Nov-2008 10:44	682
20073ca0150000.zip	28-Nov-2008 10:44	682
20073ca0151000.zip	03-Dec-2008 12:10	682
20073ca0152000.zip	03-Dec-2008 12:10	682
20073ca0153000.zip	03-Dec-2008 12:10	682
all_ca.zip	03-Dec-2008 12:10	97M
g20073ca.txt	03-Dec-2008 12:10	291K

Step 6. The zipfile is a large file and downloading can take a few minutes. After downloading contents of the zipfile should be extracted. The contents include four types of files (Table 30).

Table 30. Contents of the ACS Summary Files.

File	Contents
g20073ca.txt	Geography file: contains all the geographic identifiers (state, county, place, tract, etc.)
SequenceNumberTableNumberLookup.sas7bdat (this file can also be in .txt format)	Lookup table file: contains the description, number, and sequence number of the tables generated by the Census for a particular time frame.
e20073caYYYY000.txt	Estimates file: contains all the estimates for a particular sequence (YYYY).
m20073caYYYY000.txt	Margin of error file: contains all the margins of error for the estimates in a particular sequence (YYYY).

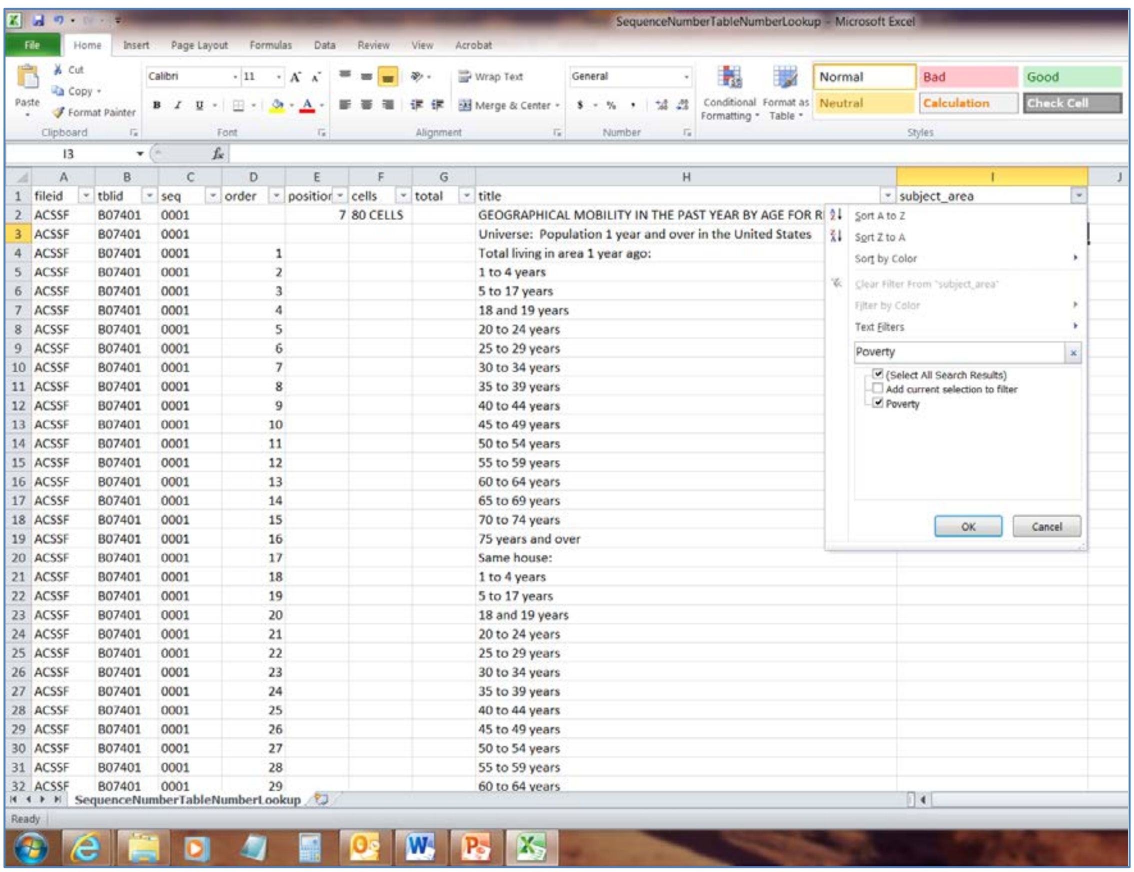
The geography, estimates, and margin of error files, are all linked by the LOGRECNO.

<u>Example: Extracting a Detailed Table from the Summary Files for ACS 3-year 2005-2007</u>

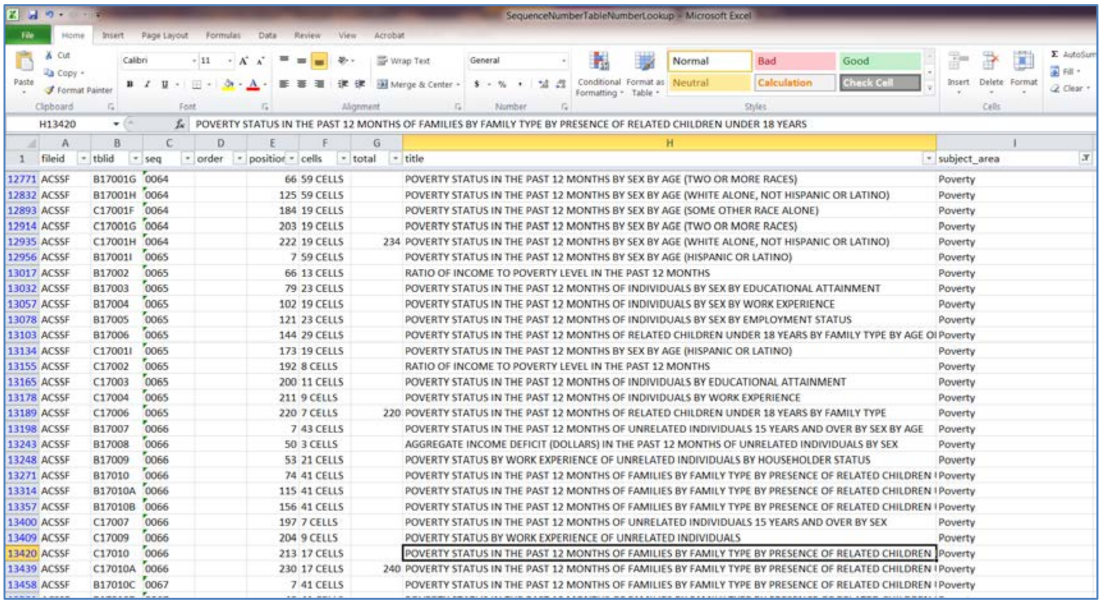
The first step to extract a detailed table from a summary file is to determine the number of the file (seq) where the table is contained and the position of the cell where the table starts. This information is contained in the SequenceNumberTableNumberLookup file. As an example, we will extract a table related to subject: poverty.

Step 1. Open the SequenceNumberTableNumberLookup file in Excel.

Step 2. Filter the field "subject_area" using the term "poverty."

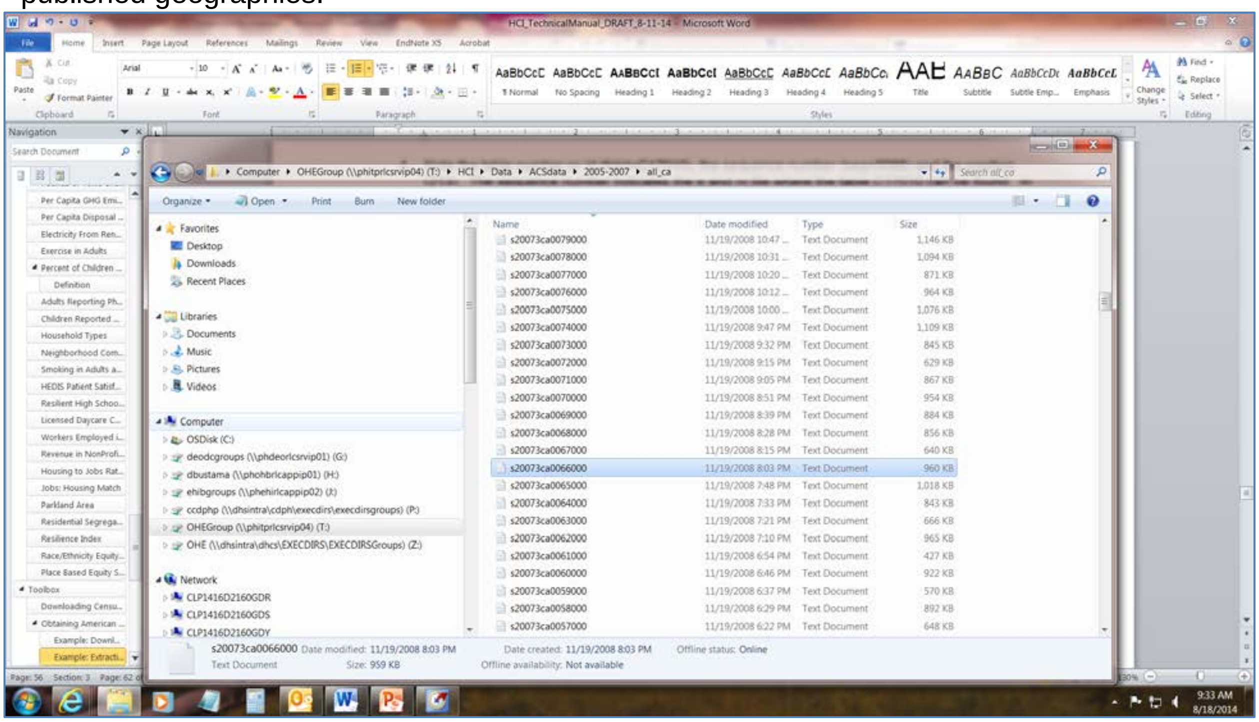


Step 3. Use the titles to find the table you are interested in. For this example we want the table entitled "POVERTY STATUS IN THE PAST 12 MONTHS OF FAMILIES BY FAMILY TYPE BY PRESENCE OF RELATED CHILDREN UNDER 18 YEARS."



1	fileid	tblid	seq	order	position	cells	total	title	subject_area
12771	ACSSF	B17001G	0064		66	59 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS BY SEX BY AGE (TWO OR MORE RACES)	Poverty
12832	ACSSF	B17001H	0064		125	59 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS BY SEX BY AGE (WHITE ALONE, NOT HISPANIC OR LATINO)	Poverty
12893	ACSSF	C17001F	0064		184	19 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS BY SEX BY AGE (SOME OTHER RACE ALONE)	Poverty
12914	ACSSF	C17001G	0064		203	19 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS BY SEX BY AGE (TWO OR MORE RACES)	Poverty
12935	ACSSF	C17001H	0064		222	19 CELLS	234	POVERTY STATUS IN THE PAST 12 MONTHS BY SEX BY AGE (WHITE ALONE, NOT HISPANIC OR LATINO)	Poverty
12956	ACSSF	B17001I	0065		7	59 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS BY SEX BY AGE (HISPANIC OR LATINO)	Poverty
13017	ACSSF	B17002	0065		66	13 CELLS		RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS	Poverty
13032	ACSSF	B17003	0065		79	23 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS OF INDIVIDUALS BY SEX BY EDUCATIONAL ATTAINMENT	Poverty
13057	ACSSF	B17004	0065		102	19 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS OF INDIVIDUALS BY SEX BY WORK EXPERIENCE	Poverty
13078	ACSSF	B17005	0065		121	23 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS OF INDIVIDUALS BY SEX BY EMPLOYMENT STATUS	Poverty
13103	ACSSF	B17006	0065		144	29 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS OF RELATED CHILDREN UNDER 18 YEARS BY FAMILY TYPE BY AGE OI	Poverty
13134	ACSSF	C17001I	0065		173	19 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS BY SEX BY AGE (HISPANIC OR LATINO)	Poverty
13155	ACSSF	C17002	0065		192	8 CELLS		RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS	Poverty
13165	ACSSF	C17003	0065		200	11 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS OF INDIVIDUALS BY EDUCATIONAL ATTAINMENT	Poverty
13178	ACSSF	C17004	0065		211	9 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS OF INDIVIDUALS BY WORK EXPERIENCE	Poverty
13189	ACSSF	C17006	0065		220	7 CELLS	220	POVERTY STATUS IN THE PAST 12 MONTHS OF RELATED CHILDREN UNDER 18 YEARS BY FAMILY TYPE	Poverty
13198	ACSSF	B17007	0066		7	43 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS OF UNRELATED INDIVIDUALS 15 YEARS AND OVER BY SEX BY AGE	Poverty
13243	ACSSF	B17008	0066		50	3 CELLS		AGGREGATE INCOME DEFICIT (DOLLARS) IN THE PAST 12 MONTHS OF UNRELATED INDIVIDUALS BY SEX	Poverty
13248	ACSSF	B17009	0066		53	21 CELLS		POVERTY STATUS BY WORK EXPERIENCE OF UNRELATED INDIVIDUALS BY HOUSEHOLDER STATUS	Poverty
13271	ACSSF	B17010	0066		74	41 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS OF FAMILIES BY FAMILY TYPE BY PRESENCE OF RELATED CHILDREN I	Poverty
13314	ACSSF	B17010A	0066		115	41 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS OF FAMILIES BY FAMILY TYPE BY PRESENCE OF RELATED CHILDREN I	Poverty
13357	ACSSF	B17010B	0066		156	41 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS OF FAMILIES BY FAMILY TYPE BY PRESENCE OF RELATED CHILDREN I	Poverty
13400	ACSSF	C17007	0066		197	7 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS OF UNRELATED INDIVIDUALS 15 YEARS AND OVER BY SEX	Poverty
13409	ACSSF	C17009	0066		204	9 CELLS		POVERTY STATUS BY WORK EXPERIENCE OF UNRELATED INDIVIDUALS	Poverty
13420	ACSSF	C17010	0066		213	17 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS OF FAMILIES BY FAMILY TYPE BY PRESENCE OF RELATED CHILDREN	Poverty
13439	ACSSF	C17010A	0066		230	17 CELLS	240	POVERTY STATUS IN THE PAST 12 MONTHS OF FAMILIES BY FAMILY TYPE BY PRESENCE OF RELATED CHILDREN I	Poverty
13458	ACSSF	B17010C	0067		7	41 CELLS		POVERTY STATUS IN THE PAST 12 MONTHS OF FAMILIES BY FAMILY TYPE BY PRESENCE OF RELATED CHILDREN I	Poverty

Step 4. Note the table number or id (tblid=C17010) and the sequence number (seq=0066). The sequence number indicates number of the file where the table C17010 can be found. In this case the estimates for the table are in the file e20073ca0066000.txt, and the margins of error of the estimates can be found in the file m20073ca0066000.txt. This applies to all published geographies.

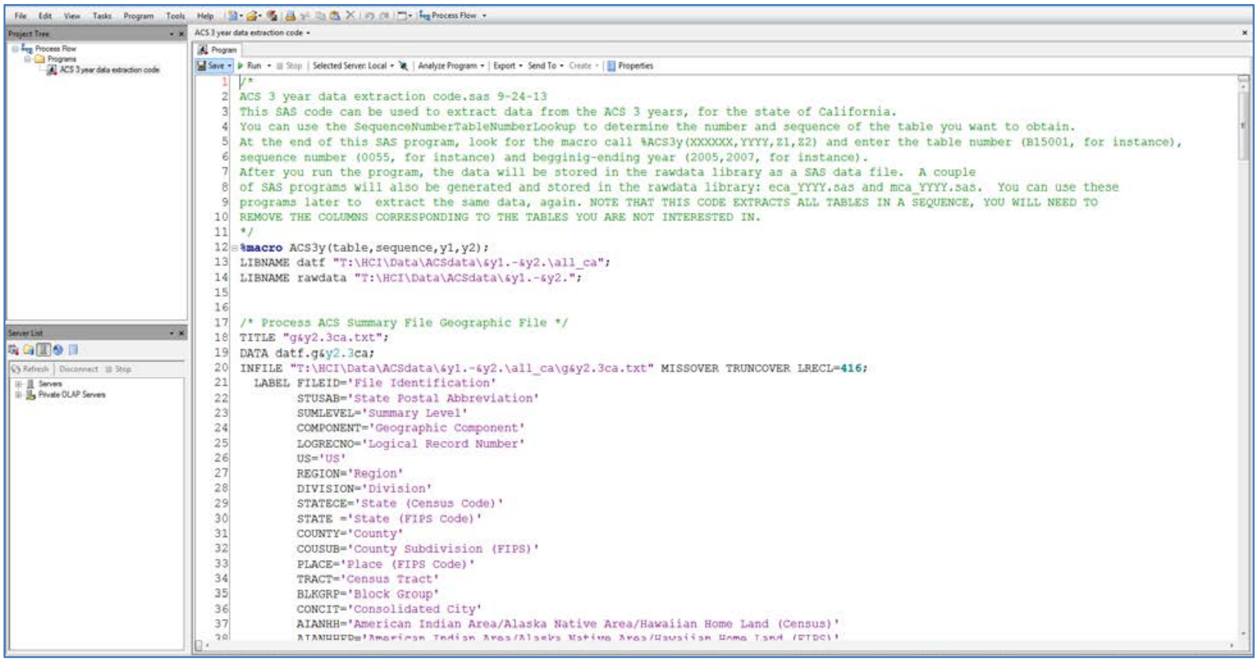


With each summary files release the U.S. Census provides SAS code to extract a particular detailed table. Advanced users can find all the documentation to implement the SAS code from the U.S. Census in the FTP site. The code for the 2005-2007 release is found here: http://www2.census.gov/acs2007_3yr/summaryfile/ and http://www2.census.gov/acs2007_3yr/summaryfile/Sample%20SAS%20Programs/. The file "summary_file_example_macros.sas" contains SAS code that can be used to extract data by seq, tblid, and state abbreviation ("al", "ca", etc.). Notice that the code will extract ALL the tables (tblid) within a sequence. For example, Seq=0066 (e20073ca0066000.txt) from the previous example includes 10 tables (Table 31).

Table 31. Detailed Tables included in a Single Sequence (0066) of the Summary Files.

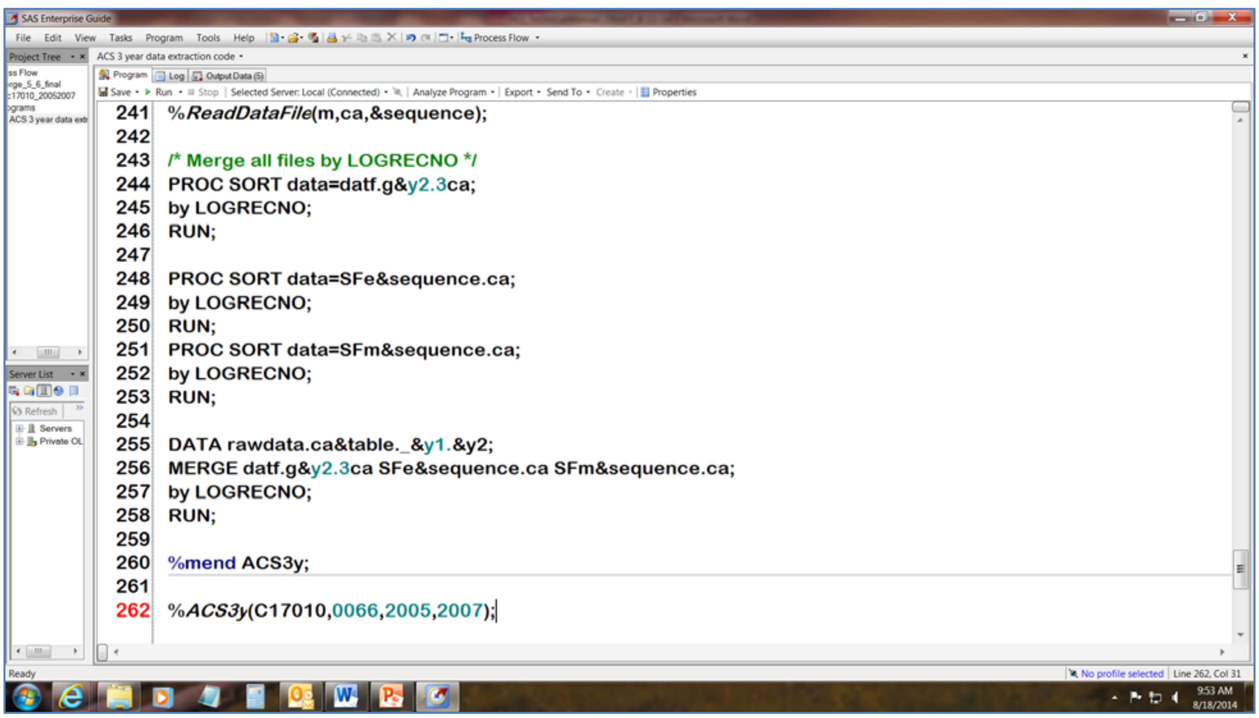
Tblid	seq	Title
B17007	0066	POVERTY STATUS IN THE PAST 12 MONTHS OF UNRELATED INDIVIDUALS 15 YEARS AND OVER BY SEX BY AGE
B17008	0066	AGGREGATE INCOME DEFICIT (DOLLARS) IN THE PAST 12 MONTHS OF UNRELATED INDIVIDUALS BY SEX
B17009	0066	POVERTY STATUS BY WORK EXPERIENCE OF UNRELATED INDIVIDUALS BY HOUSEHOLDER STATUS
B17010	0066	POVERTY STATUS IN THE PAST 12 MONTHS OF FAMILIES BY FAMILY TYPE BY PRESENCE OF RELATED CHILDREN UNDER 18 YEARS BY AGE OF RELATED CHILDREN
B17010 A	0066	POVERTY STATUS IN THE PAST 12 MONTHS OF FAMILIES BY FAMILY TYPE BY PRESENCE OF RELATED CHILDREN UNDER 18 YEARS BY AGE OF RELATED CHILDREN (WHITE ALONE HOUSEHOLDER)
B17010 B	0066	POVERTY STATUS IN THE PAST 12 MONTHS OF FAMILIES BY FAMILY TYPE BY PRESENCE OF RELATED CHILDREN UNDER 18 YEARS BY AGE OF RELATED CHILDREN (BLACK OR AFRICAN AMERICAN ALONE HOUSEHOLDER)
C17007	0066	POVERTY STATUS IN THE PAST 12 MONTHS OF UNRELATED INDIVIDUALS 15 YEARS AND OVER BY SEX
C17009	0066	POVERTY STATUS BY WORK EXPERIENCE OF UNRELATED INDIVIDUALS
C17010	0066	POVERTY STATUS IN THE PAST 12 MONTHS OF FAMILIES BY FAMILY TYPE BY PRESENCE OF RELATED CHILDREN UNDER 18 YEARS
C17010 A	0066	POVERTY STATUS IN THE PAST 12 MONTHS OF FAMILIES BY FAMILY TYPE BY PRESENCE OF RELATED CHILDREN UNDER 18 YEARS (WHITE ALONE HOUSEHOLDER)

For the HCDI project, we have downloaded and modified the U.S. Census code so it can be used for California data requiring seq, tblid, and year information for the extraction (example, ACS_3_year_data_extraction_code.sas).

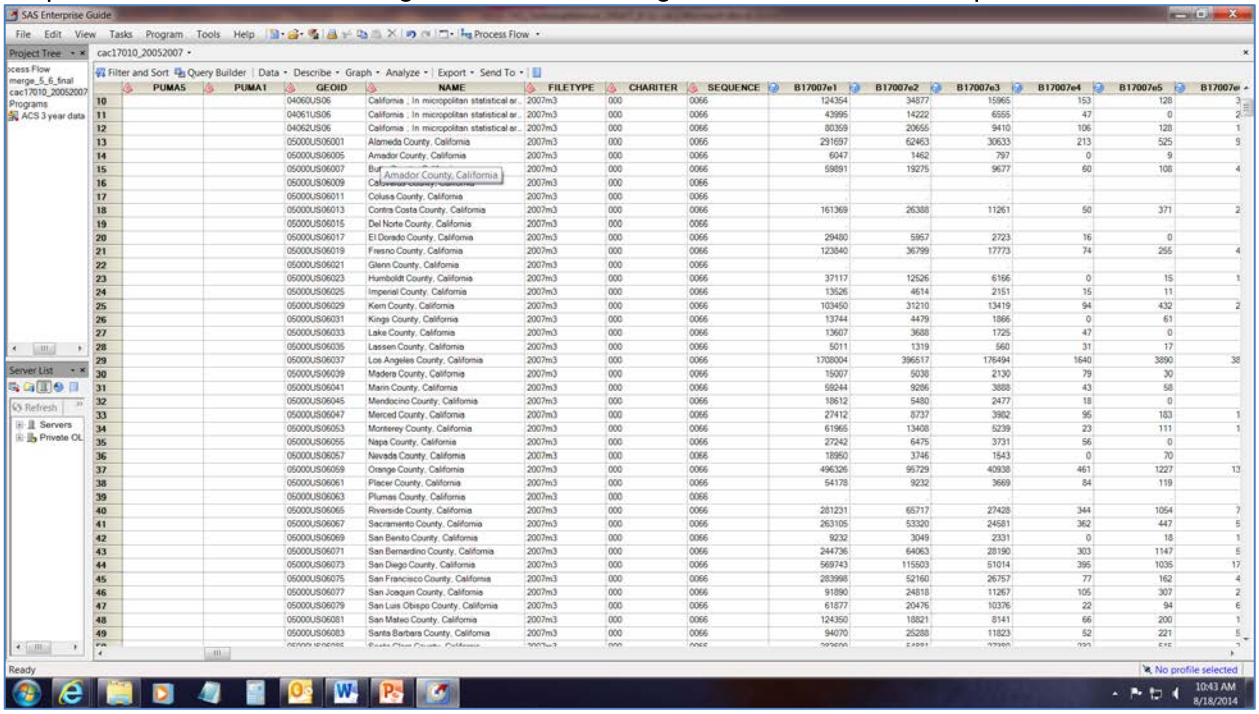


To use this extraction code, one first needs to obtain the tblid and seq numbers from the lookup table and then follow these steps:

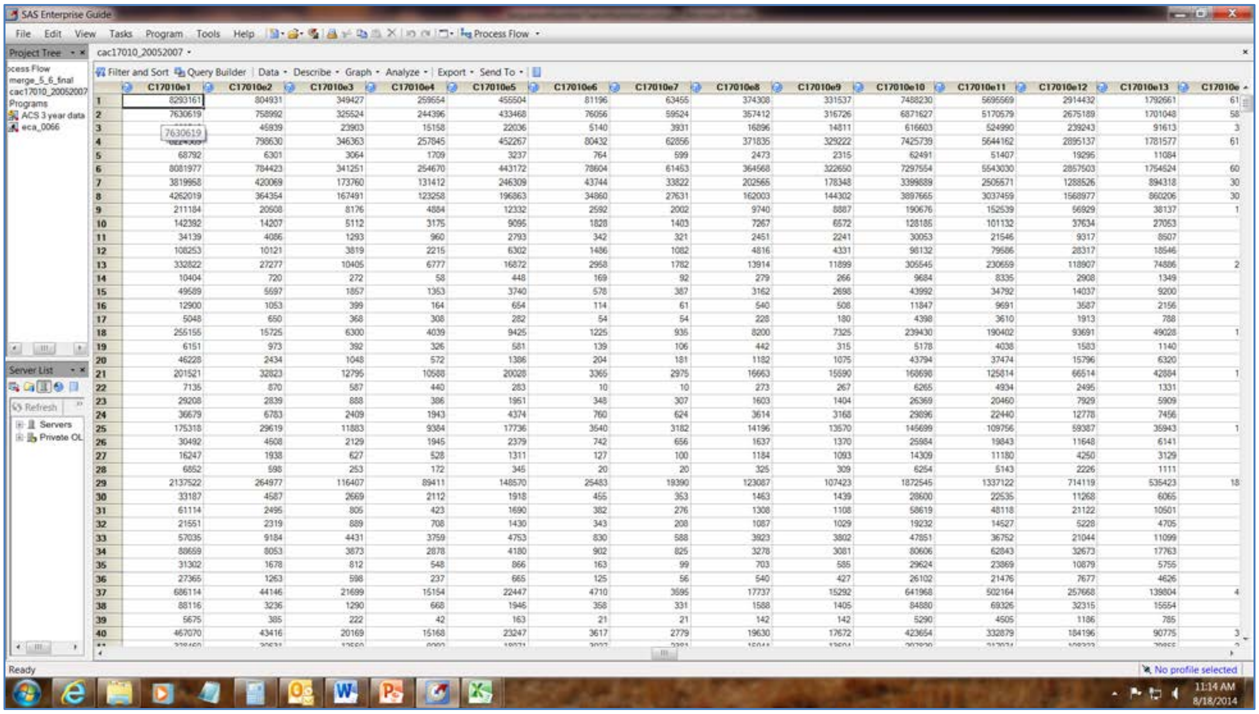
Step 1. Open the code file and go to the end of the SAS code, look for the macro call %ACS3y(XXXXX,YYYY,ZZZZ,ZZZZ). Replace the information in the macro call with the table number (C17010), sequence number (0066), and beginning (2005) and ending years (2007) of the dataset of interest. Click Run.



Step 2. A SAS data file will be generated containing all the tables in the sequence.



Step 3. The SAS macros contained in the code have merged the geographical identifiers from the geography file (g20073ca.txt), the estimates from the e file (e20073ca0066000.txt), and the corresponding margins of error from the m file (m20073ca0066000.txt); the merge is conducted using the using the LOGRECNO field. The data file that is generated after this merge has 533 columns. The first 51 columns in correspond to geographic and table identifiers coming from the geography file. The rest of the columns correspond to the data: 241 columns of estimates and 241 of the margins of error for the estimates. To find the columns for table C17010, one can scroll to the right of the data file and look for the variables labeled “C17010e1,” where e indicates “estimate” and the number (#1 in this example) indicates it is the first column or variable for this table.

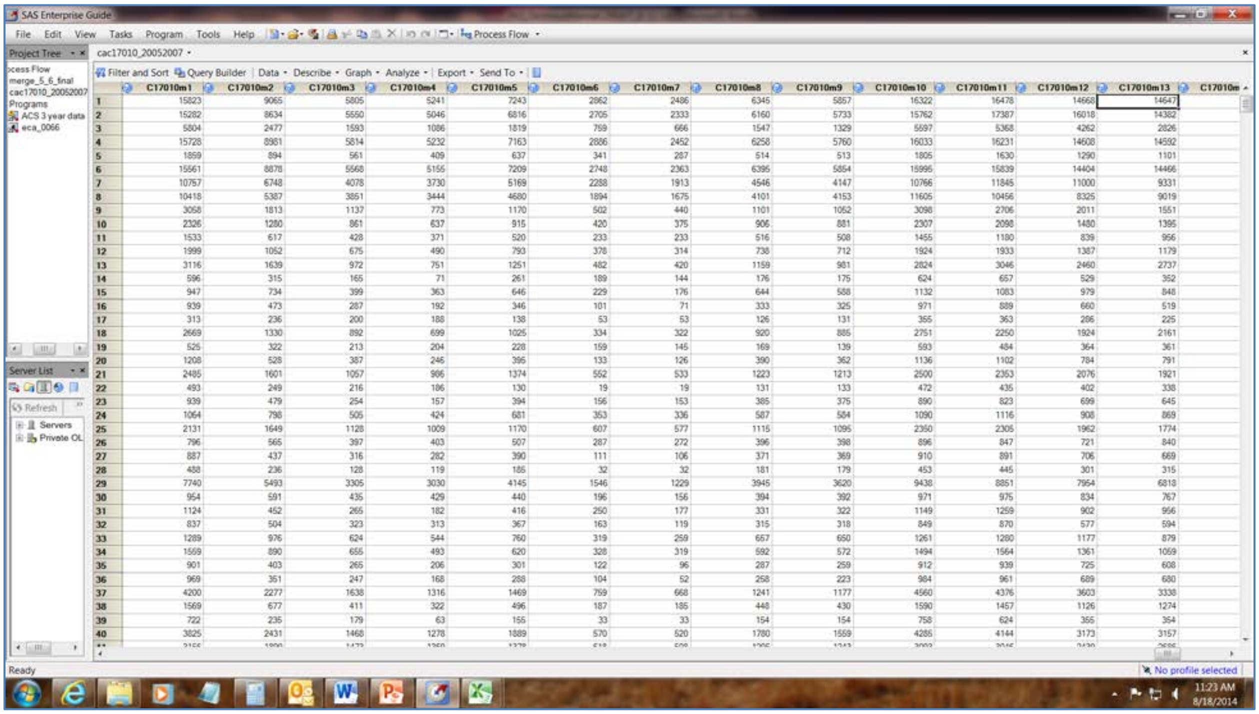


Step 4. The sequence lookup file has more information about the variables in the data file generated by the code. For this example, the sequence lookup (see table 32) indicates that table C17010 starts at column 213, that there are 17 cells for this table, that variable C17010e1 corresponds to the "Total:", in this case, total number of families. Variable C17010e2 corresponds to the number of families whose "income in the past 12 months is below the poverty level". Variable C17010e3 corresponds to the number of "married-couple families" whose income in the past 12 months is below the poverty level, and so forth.

Table 32. Variables Included in the Detailed Table C17010

C17010		213	17 CELLS	POVERTY STATUS IN THE PAST 12 MONTHS OF FAMILIES BY FAMILY TYPE BY PRESENCE OF RELATED CHILDREN UNDER 18 YEARS
n				Universe: Families
C17010	1			Total:
C17010	2			Income in the past 12 months below poverty level:
C17010	3			Married-couple family:
C17010	4			With related children under 18 years
C17010	5			Other family:
C17010	6			Male householder, no wife present
C17010	7			With related children under 18 years
C17010	8			Female householder, no husband present
C17010	9			With related children under 18 years
C17010	10			Income in the past 12 months at or above poverty level:
C17010	11			Married-couple family:
C17010	12			With related children under 18 years
C17010	13			Other family:
C17010	14			Male householder, no wife present
C17010	15			With related children under 18 years
C17010	16			Female householder, no husband present
C17010	17			With related children under 18 years

Step 5. The corresponding margins of error for each of the estimates are found in the same data file, only the variables will be labeled with an “m”. For instance “C17010m1” is the margin of error for the estimate “C17010e1.”



Step 6. As one continues working with the data file, one will need to retain only the columns corresponding to the tables of interest.

Note: two SAS programs will be generated by the Census code. For the previous example those files will be eca_0066.sas and mca_0066.sas. These programs can be saved and used to extract the sequence 066 data in the future.

Downloading Census TIGER Shapefiles and Reprojecting in California Teale Albers Coordinate System

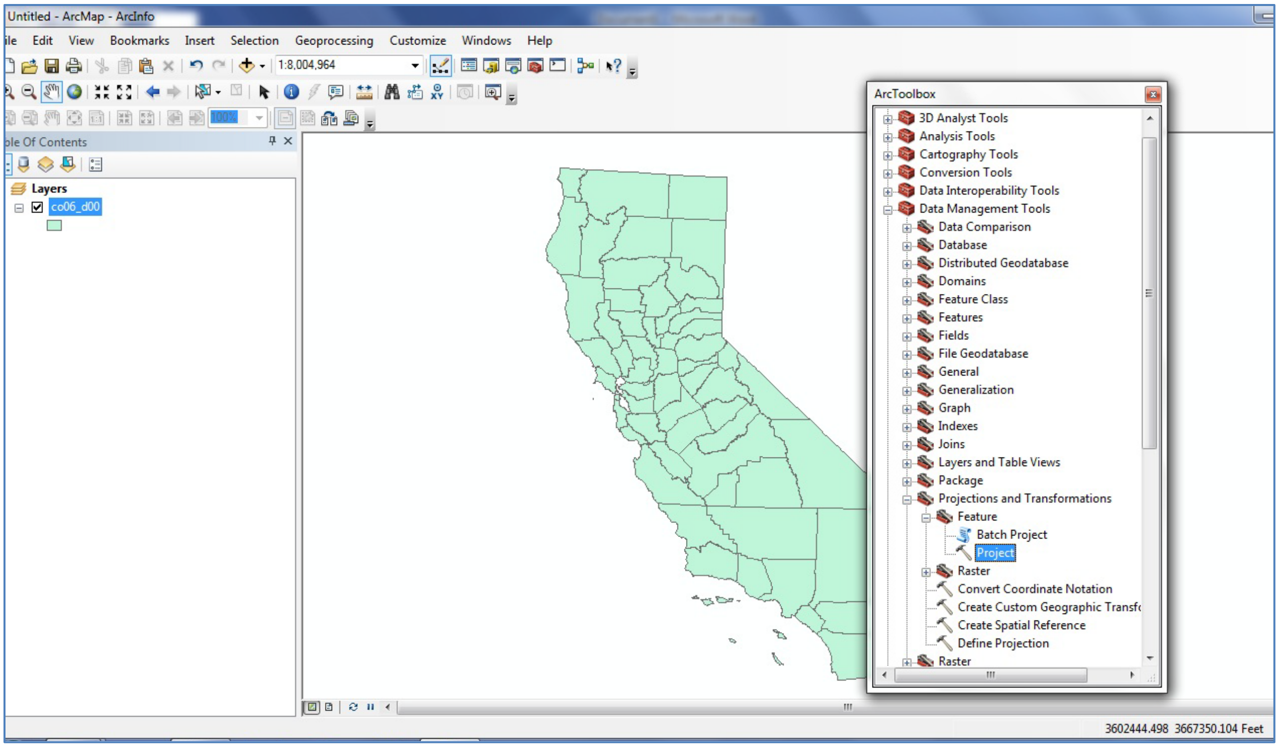
Download census cartographic boundary shapefiles (TIGER files, http://www.census.gov/geo/maps-data/data/tiger-cart-boundary.html) for California from the following locations:

Table 33. URL of TIGER Shapefiles for Multiple Geographies and Datasets.

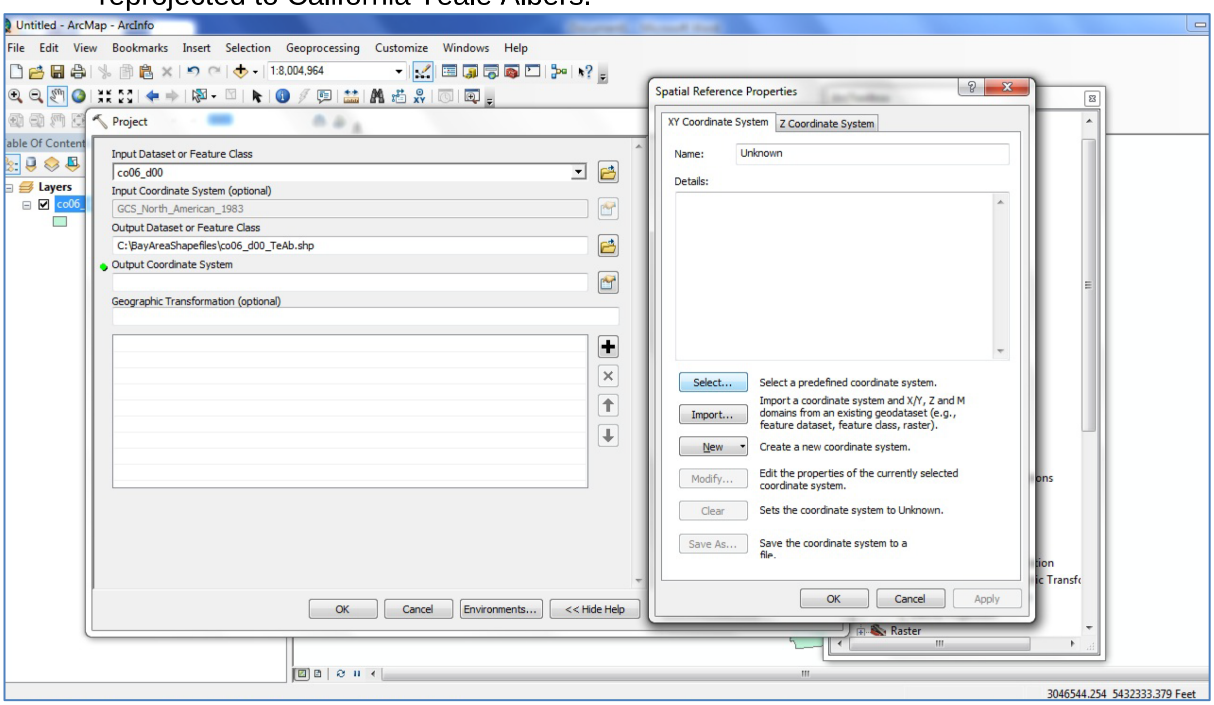
Dataset	Geography	URL
2000	Counties	http://www2.census.gov/geo/tiger/TIGER2010/COUNTY/2000/tl_2010_06_county00.zip
	Incorporated Places/ Census Designated Places	http://www2.census.gov/geo/tiger/TIGER2010/PLACE/2000/tl_2010_06_place00.zip
	Census Tracts	http://www2.census.gov/geo/tiger/TIGER2010/TRACT/2000/tl_2010_06_tract00.zip
	Census Blocks	http://www2.census.gov/geo/tiger/TIGER2010/BG/2000/tl_2010_06_bg00.zip
	5-digit Zip Code Tabulation Areas	http://www2.census.gov/geo/tiger/TIGER2010/ZCTA5/2000/tl_2010_06_zcta500.zip
2010	Counties	http://www2.census.gov/geo/tiger/TIGER2010/COUNTY/2010/tl_2010_06_county10.zip
	Incorporated Places/ Census Designated Places	http://www2.census.gov/geo/tiger/TIGER2010/PLACE/2010/tl_2010_06_place10.zip
	Census Tracts	http://www2.census.gov/geo/tiger/TIGER2010/TRACT/2010/tl_2010_06_tract10.zip
	Census Blocks	http://www2.census.gov/geo/tiger/TIGER2010/BG/2010/tl_2010_06_bg10.zip
	5-digit Zip Code Tabulation Areas	http://www2.census.gov/geo/tiger/TIGER2010/ZCTA5/2010/tl_2010_06_zcta510.zip

To reproject in California Teale Albers, unzip and open each file separately in ArcMap, following the procedure below:

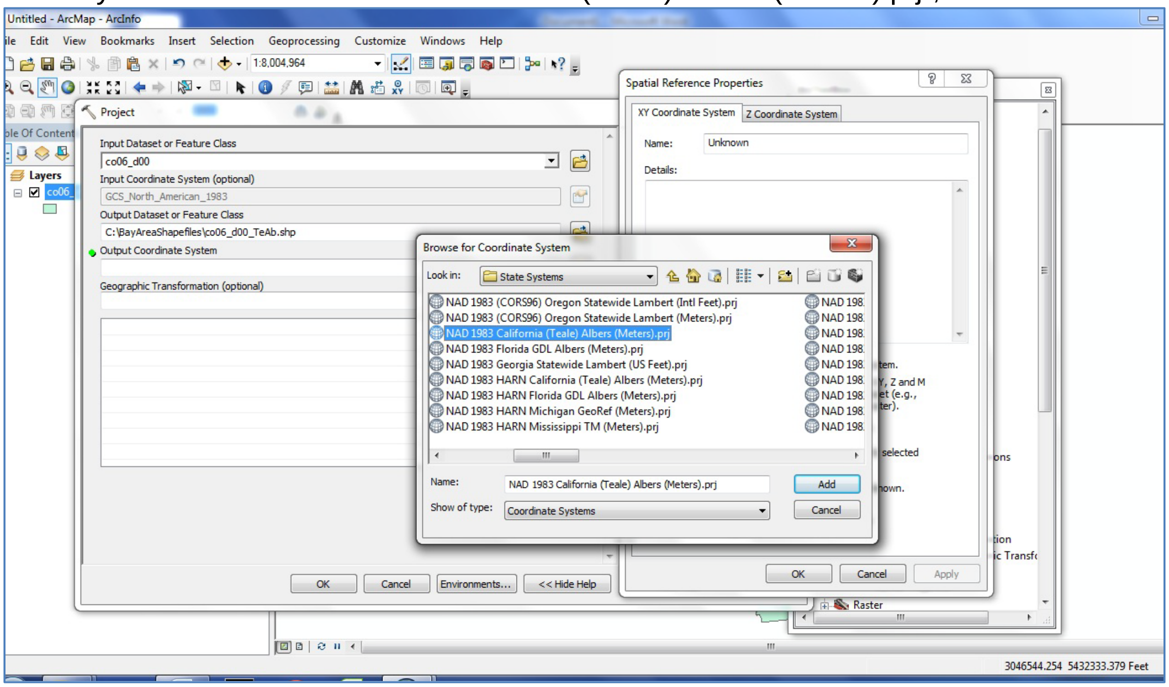
Step 1. Open shapefile, and in ArcToolbox go to Data Management Tools. Choose Projections and Transformations, select Feature, and then select Project.



Step 2. In Project window, choose Input Dataset or Feature Class (co06_d00.shp in example). Input Coordinate System may fill in automatically. Choose Output Dataset or Feature Class, adding "_TeAb" to filename, indicating that shapefile data is being reprojected to California Teale Albers.

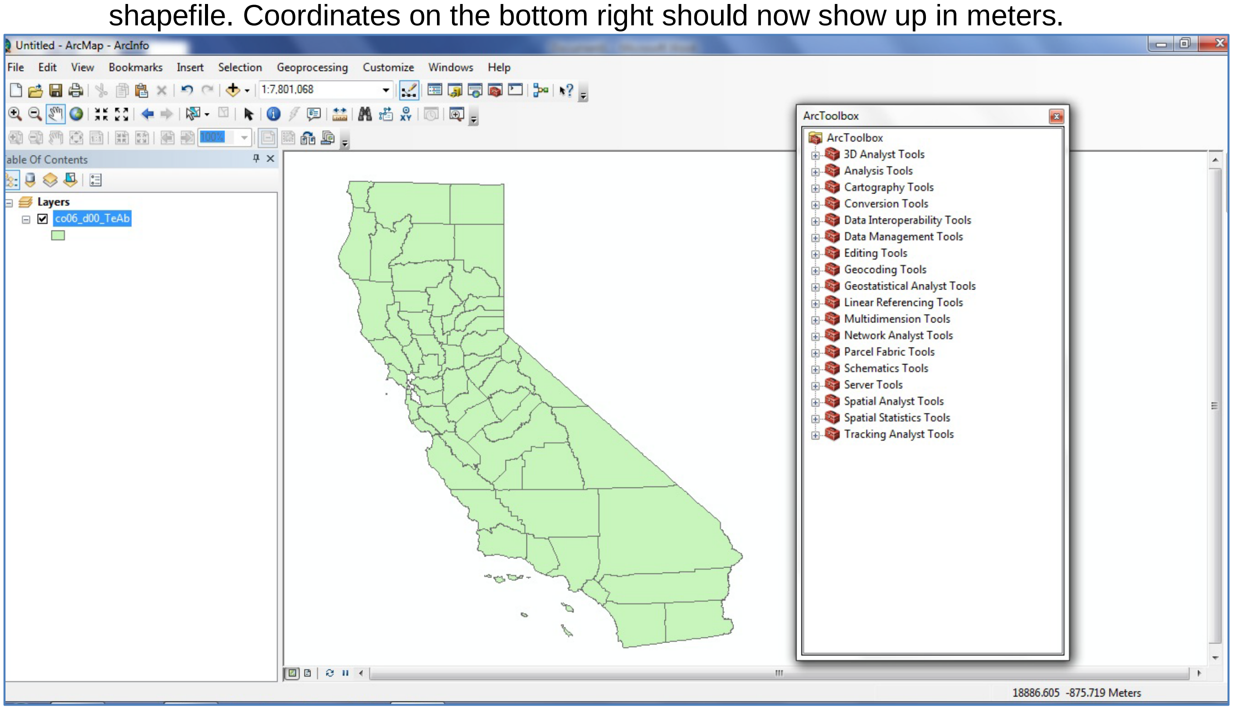


Step 3. For Output Coordinate System, click on box and 'Select a predefined coordinate system'. Select Projected Coordinate Systems, and then select State Systems. Select "NAD 1983 California (Teale) Albers (Meters).prj", and Add



Step 4. Click OK and the OK again, and the newly reprojected shapefile will be added to the ArcMap project.

Step 5. Close and reopen ArcMap. From blank map project, add newly reprojected shapefile. Coordinates on the bottom right should now show up in meters.



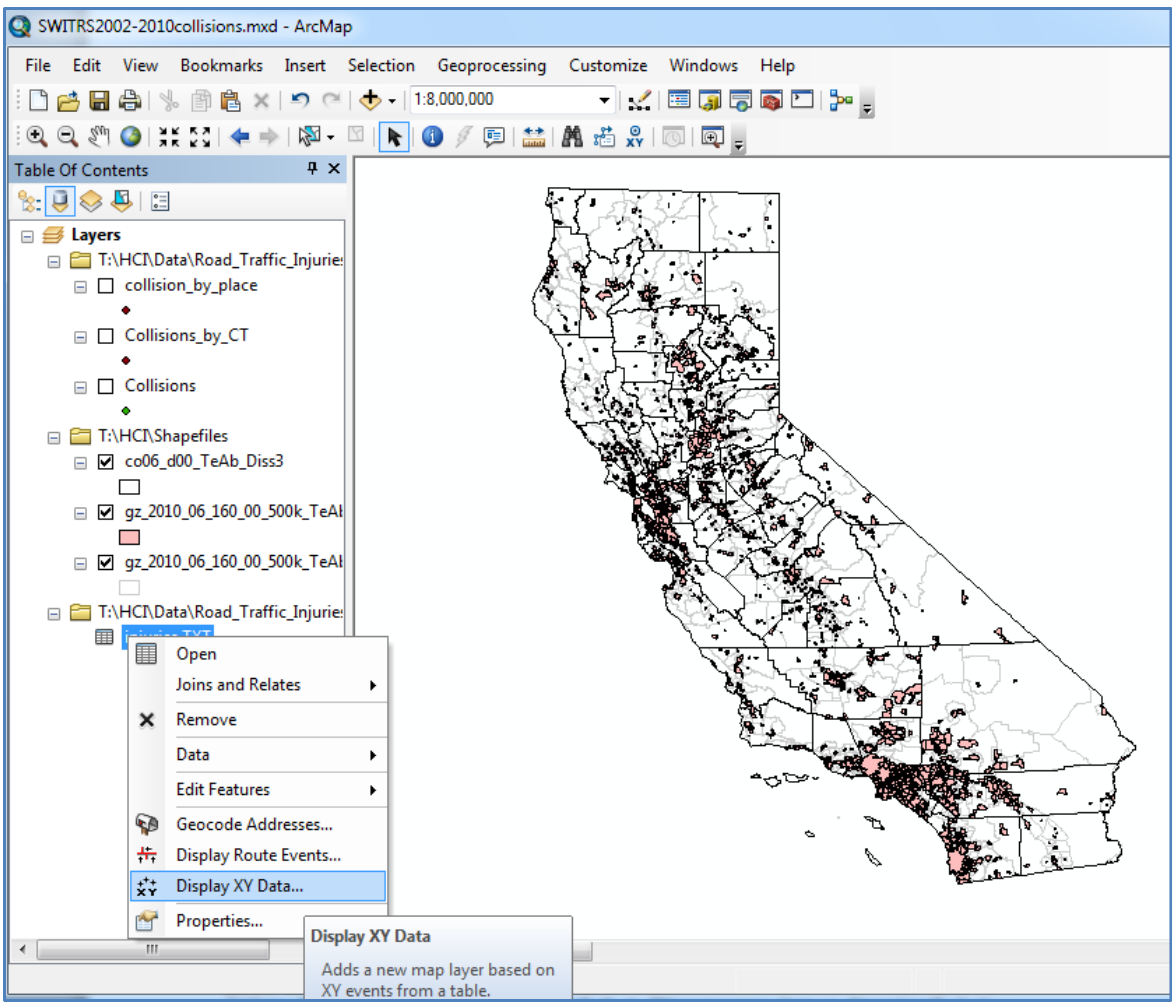
Displaying XY Data in ArcMap

The following steps illustrate how to display geocoded data (with XY coordinates) in ArcMap. Geocoded data could be provided by an organization (source) or it could be obtaining from a list of addresses whose XY coordinates are obtained from a geocoding service. The following example uses a table of the geocoded locations of traffic collisions. It is necessary to know in advance the geographic system (not projected) of the coordinates, in order to provide it to ArcMap.

Step 1. Add table of injuries (injuries.txt) to ArcMap.

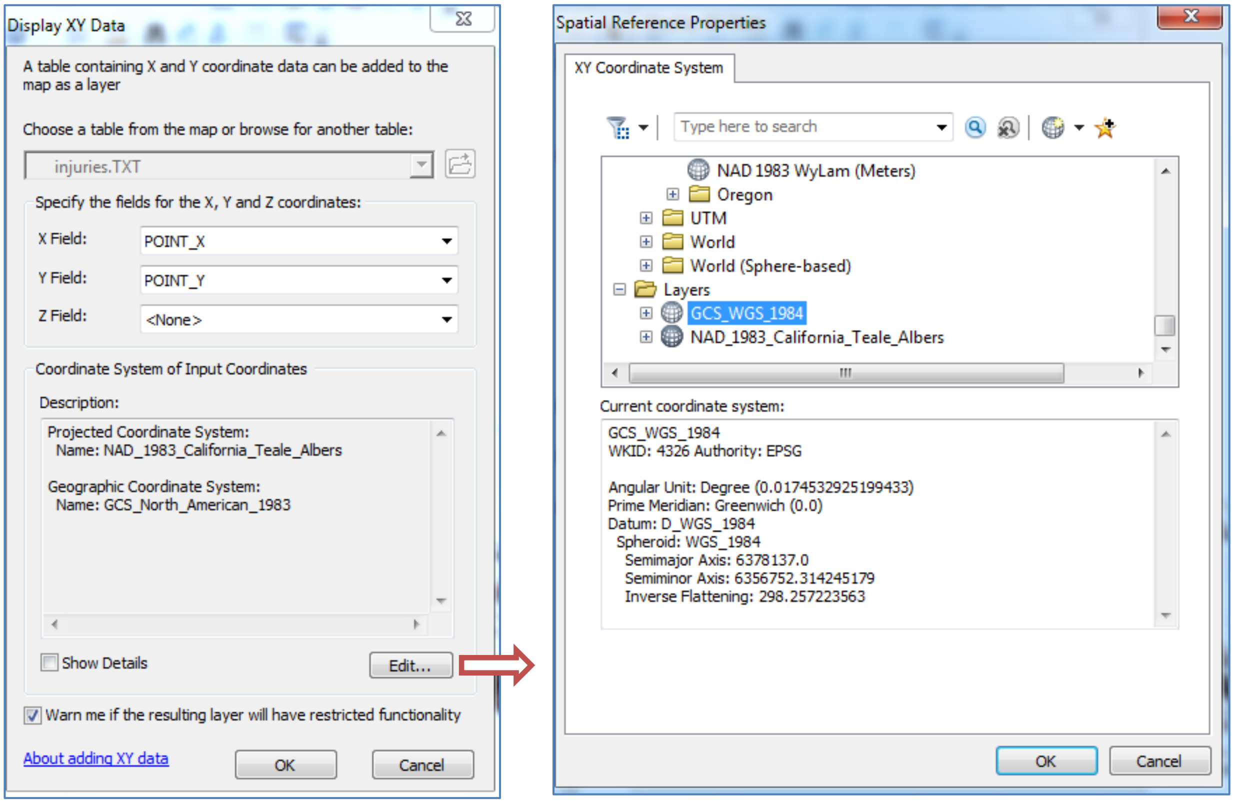
Step 2. Add XY injury coordinates using WGS84 layer (add geographic -not projected- coordinate system from list)

Step 3. Select Display XY Data

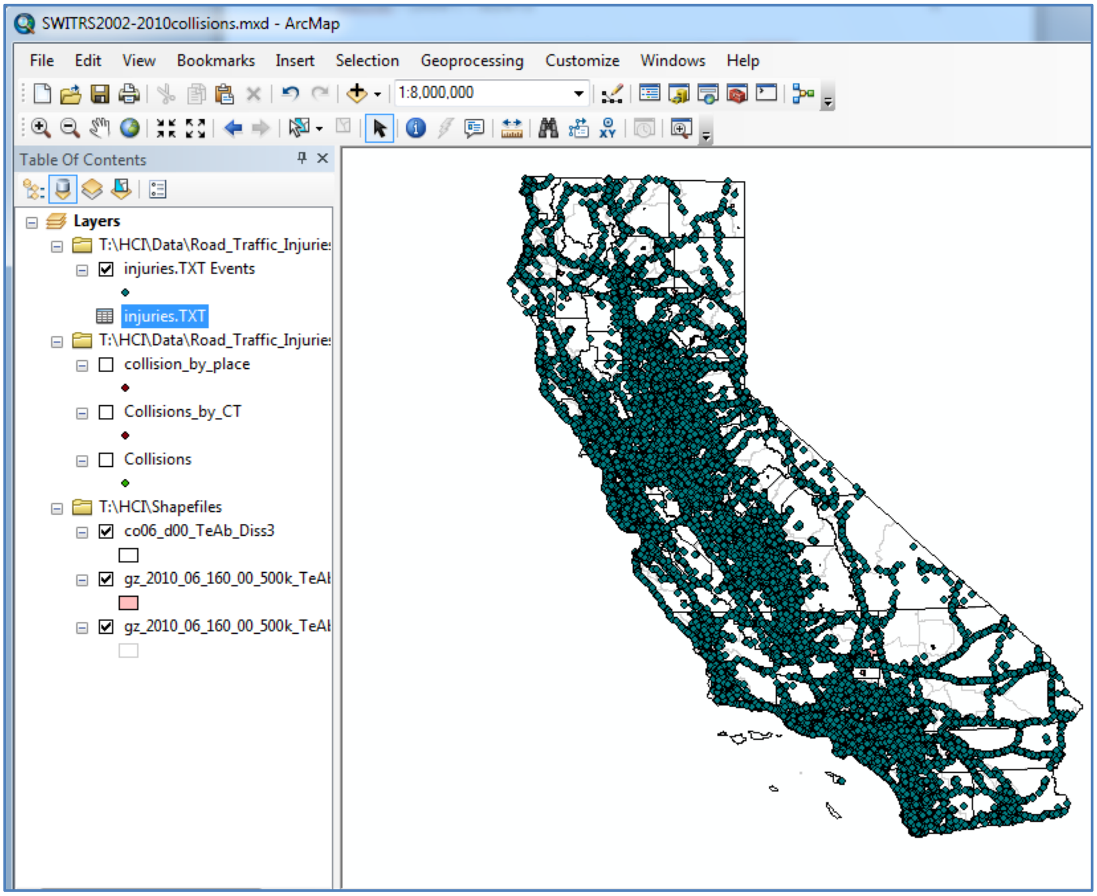


Step 4. In dialog box, use data source (SWITRS) defined fields for X (Point _X) and Y (Point_Y) coordinates.

Step 5. Select Edit to change to coordinate system of SWITRS data (WGS84) rather than coordinate system of map projection (NAD1983 Teal Albers).



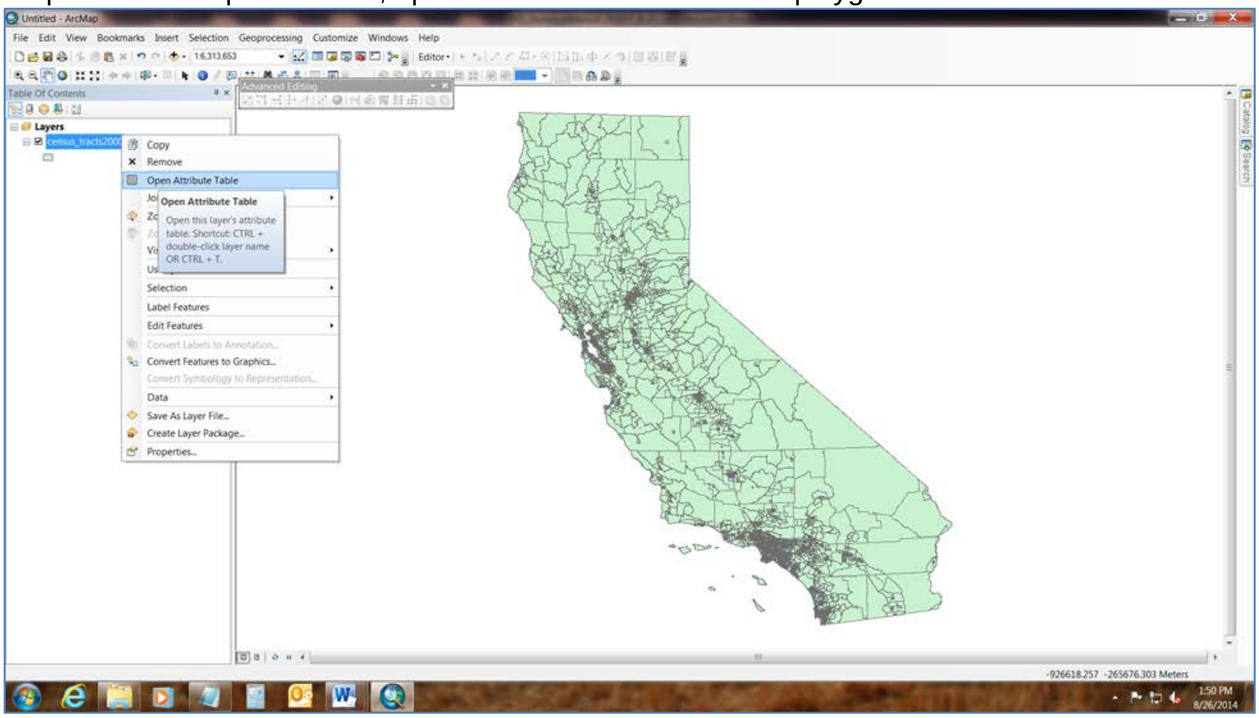
Step 6. Select OK and OK in previous dialog box. All the XY points where road traffic collisions occurred should map.


SWITRS2002-2010collisions.mxd - ArcMap
File Edit View Bookmarks Insert Selection Geoprocessing Customize Windows Help
1:8,000,000
Table Of Contents
Layers
T:\HCI\Data\Road_Traffic_Injurie
injuries.TXT Events
injuries.TXT
T:\HCI\Data\Road_Traffic_Injurie
collision_by_place
Collisions_by_CT
Collisions
T:\HCI\Shapefiles
co06_d00_TeAb_Diss3
gz_2010_06_160_00_500k_TeAl
gz_2010_06_160_00_500k_TeAl

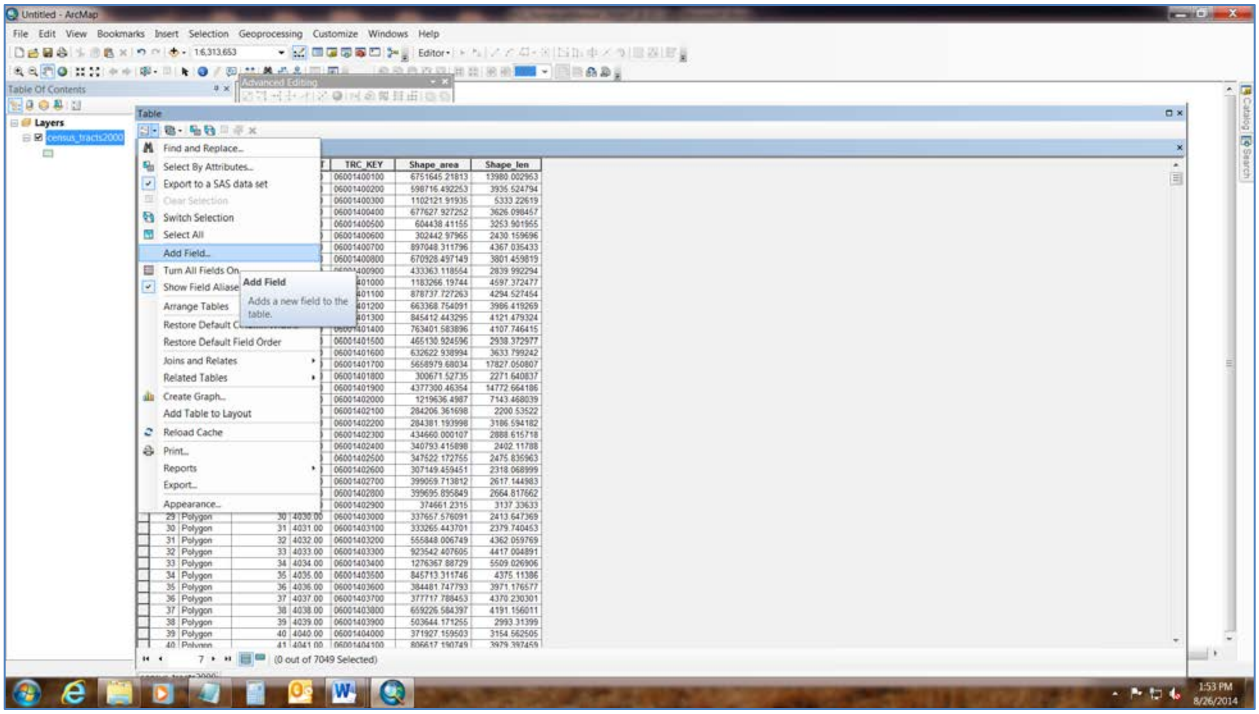
Calculating Centroids in ArcMap and Creating a Centroid Shapefile

Shapefiles of 2000 Census Tracts and Places in California were obtained from the U.S. Census TIGER files (see above). The centroid of each Census tract was calculated using the instructions at “Calculate the x,y coordinates of the feature using Calculate Geometry”, http://support.esri.com/en/knowledgebase/techarticles/detail/41027).

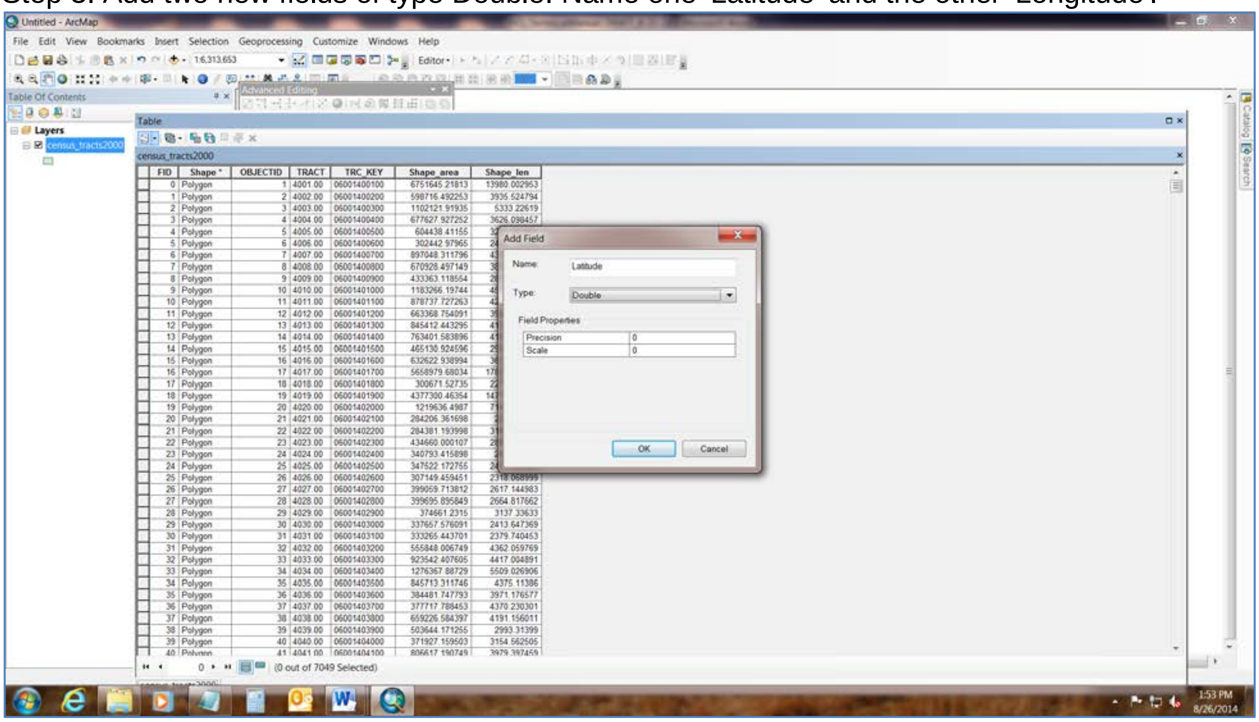
Step 1. In the map document, open the attribute table for the polygon feature class.



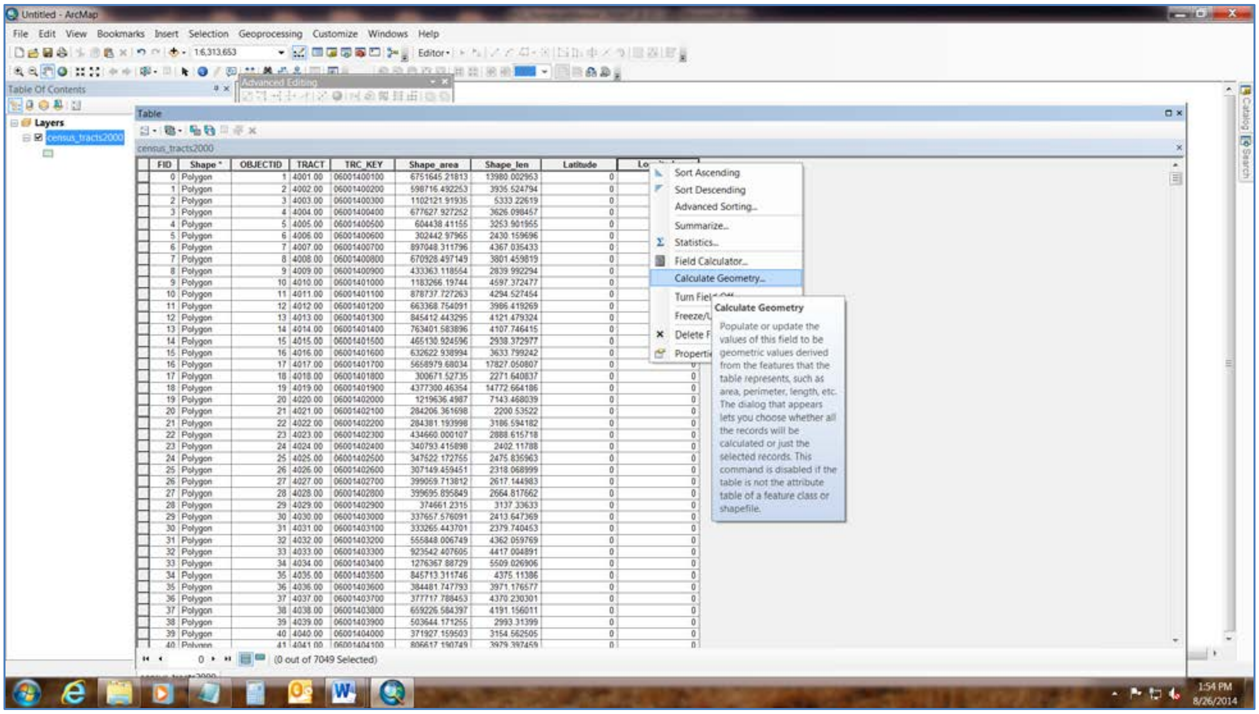
Step 2. In the attribute table, navigate to Table Options > Add Field.



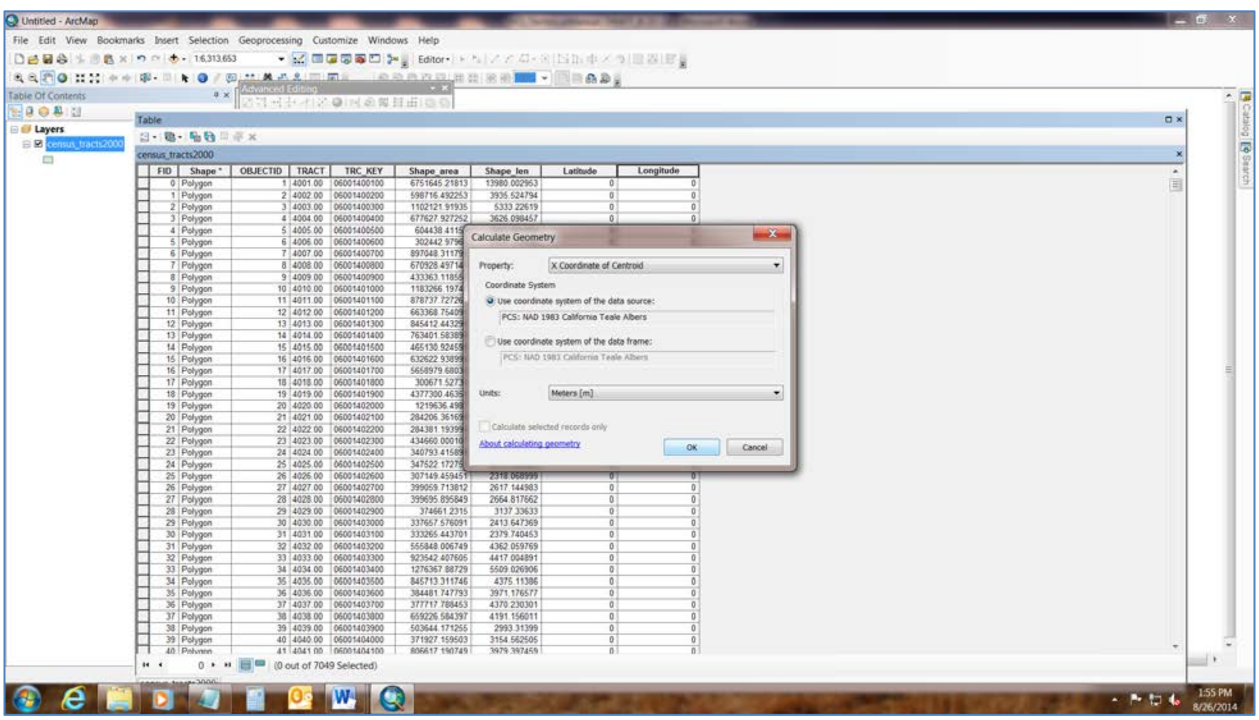
Step 3. Add two new fields of type Double. Name one 'Latitude' and the other 'Longitude'.



Step 4. Right-click the Longitude field and select Calculate Geometry.



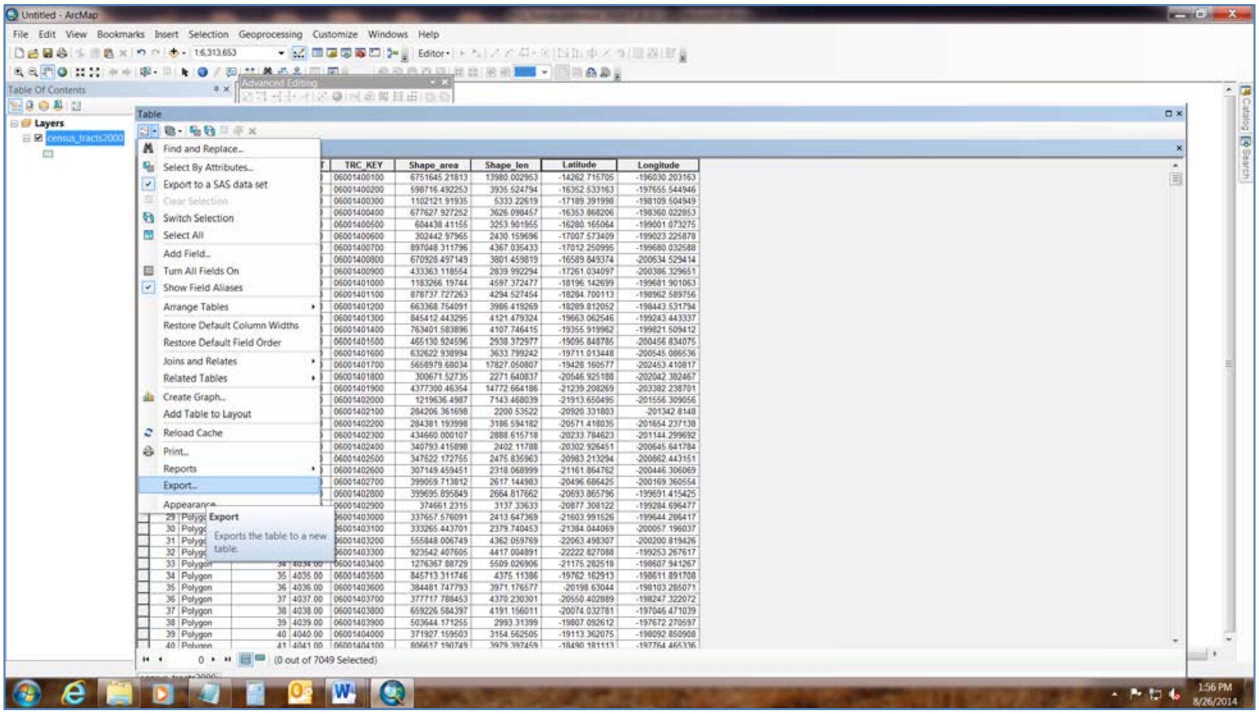
Step 5. In the Calculate Geometry dialog box, select 'X Coordinate of Centroid' from the Property drop-down menu. Click OK.

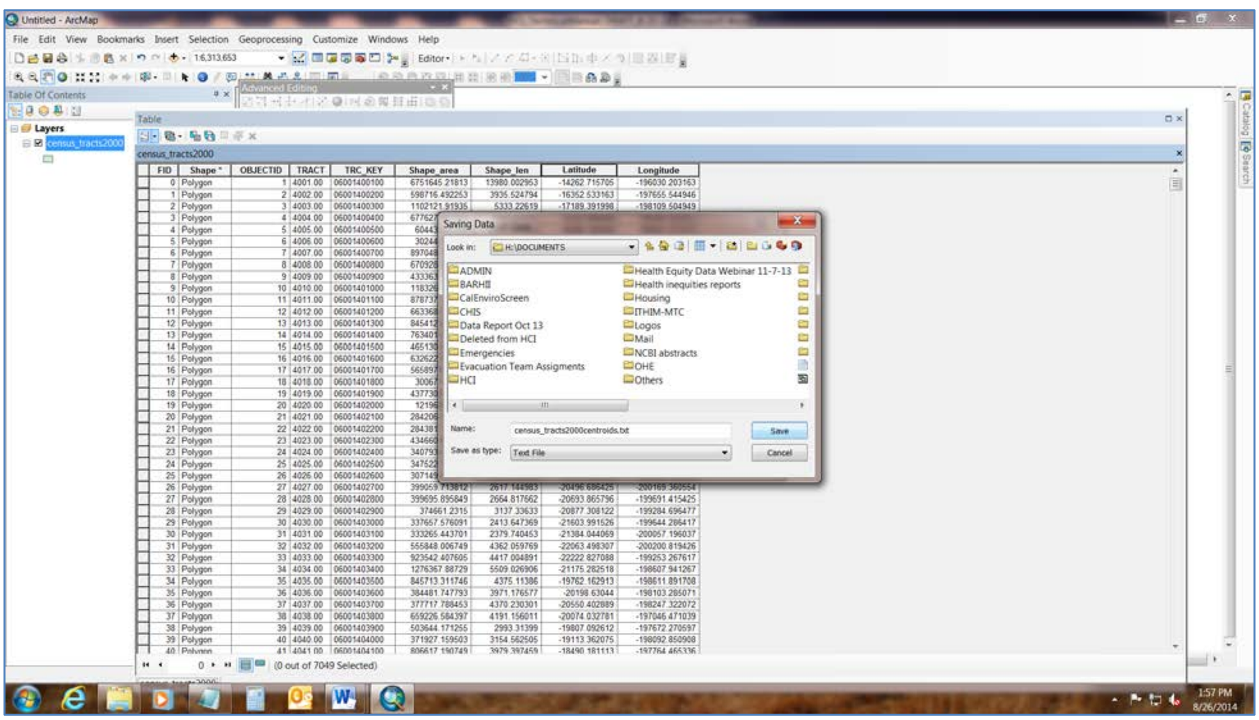


Step 6. Right-click the Latitude field and select Calculate Geometry.
Step 7. In the Calculate Geometry dialog box, select 'Y Coordinate of Centroid' from the Property drop-down menu. Click OK.

Step 8. Export to a table.

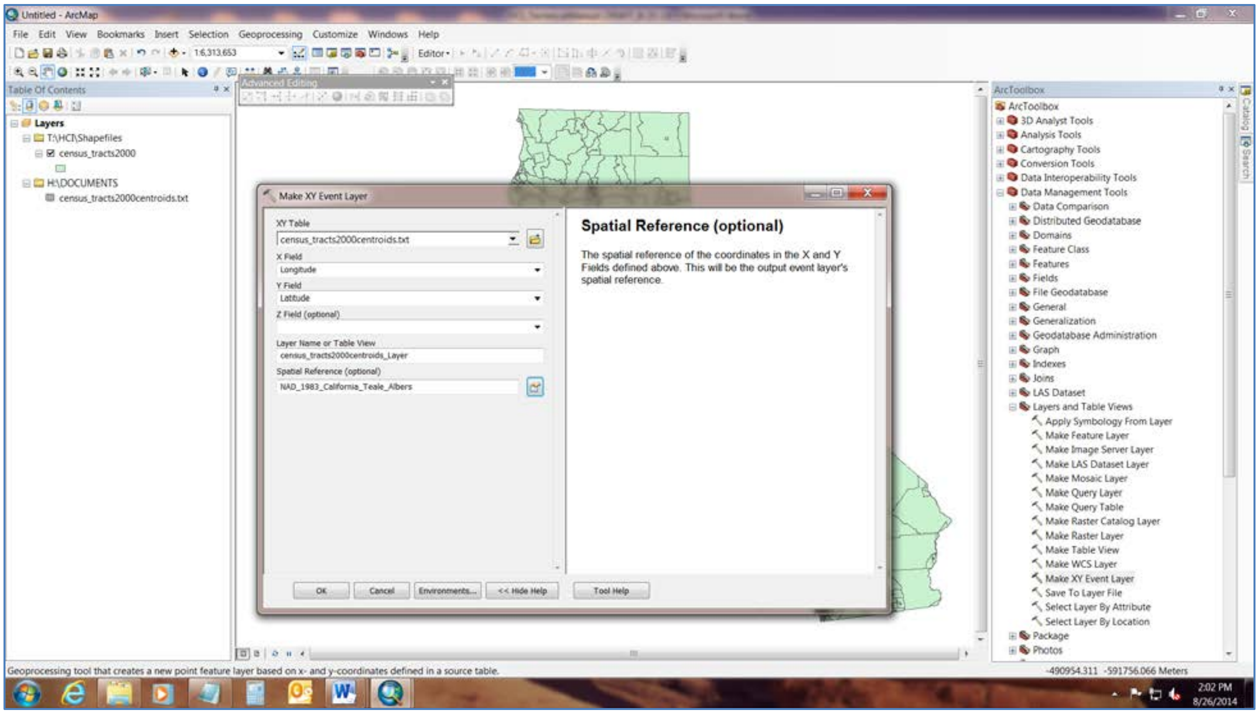
a. In the attribute table, select Table Options > Export.
b. Specify a name and location for the new table.
c. Save as .txt and add the new table to the contents.





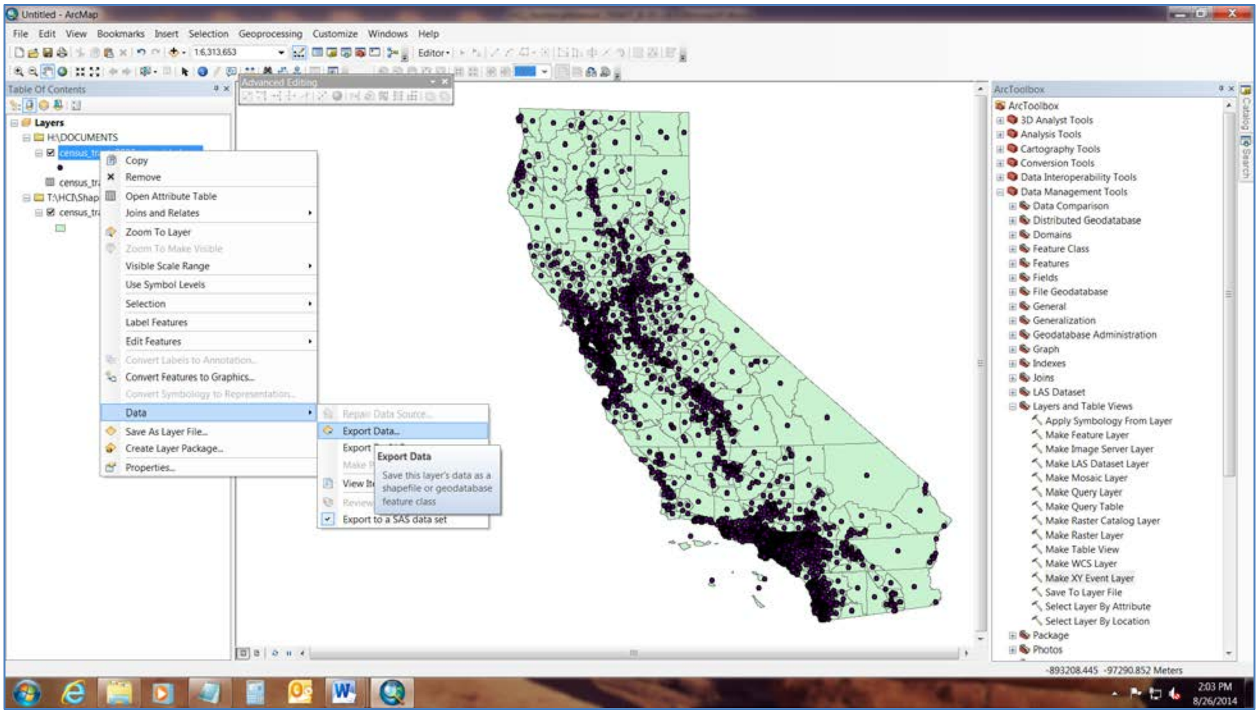
Step 9. Make an XY Event layer.

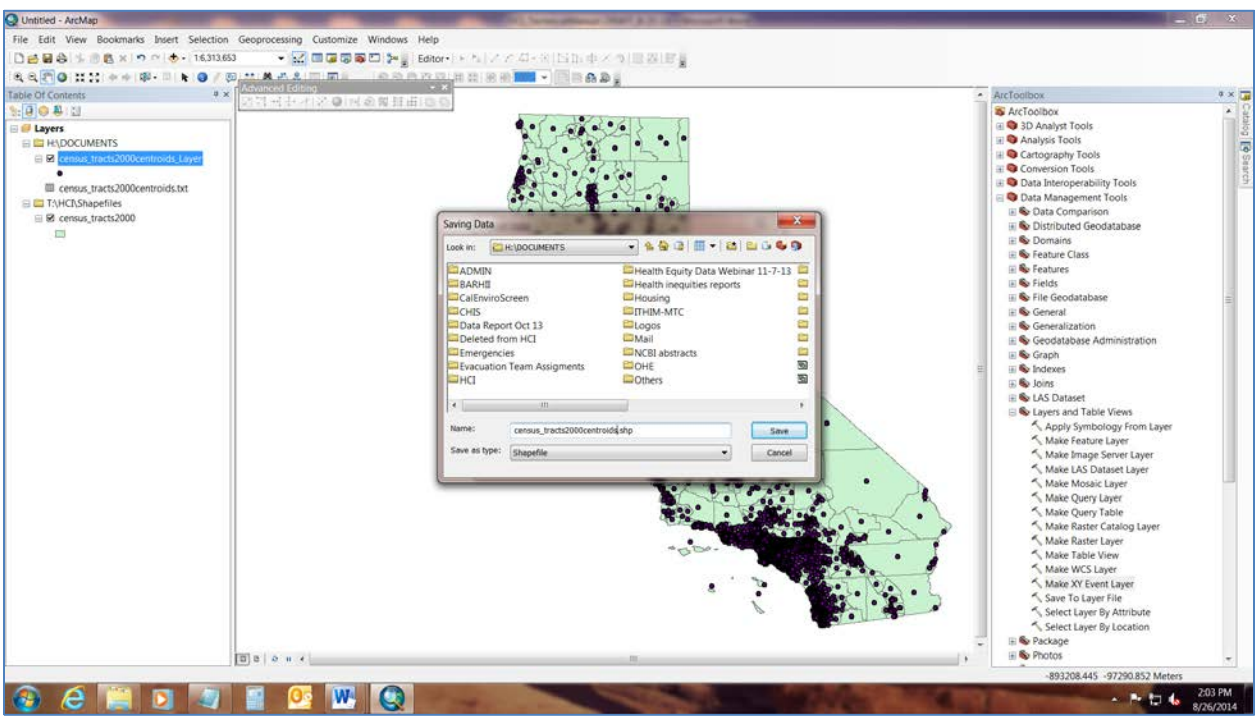
a. Navigate to ArcToolbox > Data Management Tools > Layers and Table Views > Make XY Event Layer.
b. For the X Field, select the Longitude field.
c. For the Y Field, select the Latitude field.
d. Name the new event layer.
e. Select the spatial reference or coordinate system.
f. Click OK.



Step 10. Export the Event Layer to a shapefile.

a. Right click on the XY Event Layer and go to “Data”
b. Select “Export Data”, indicate the coordinate system for the new shapefile (Select “this layer’s source data” which should be NAD_1983_California_Teale_Albers).
c. Provide a name for the new shapefile and click “OK.”

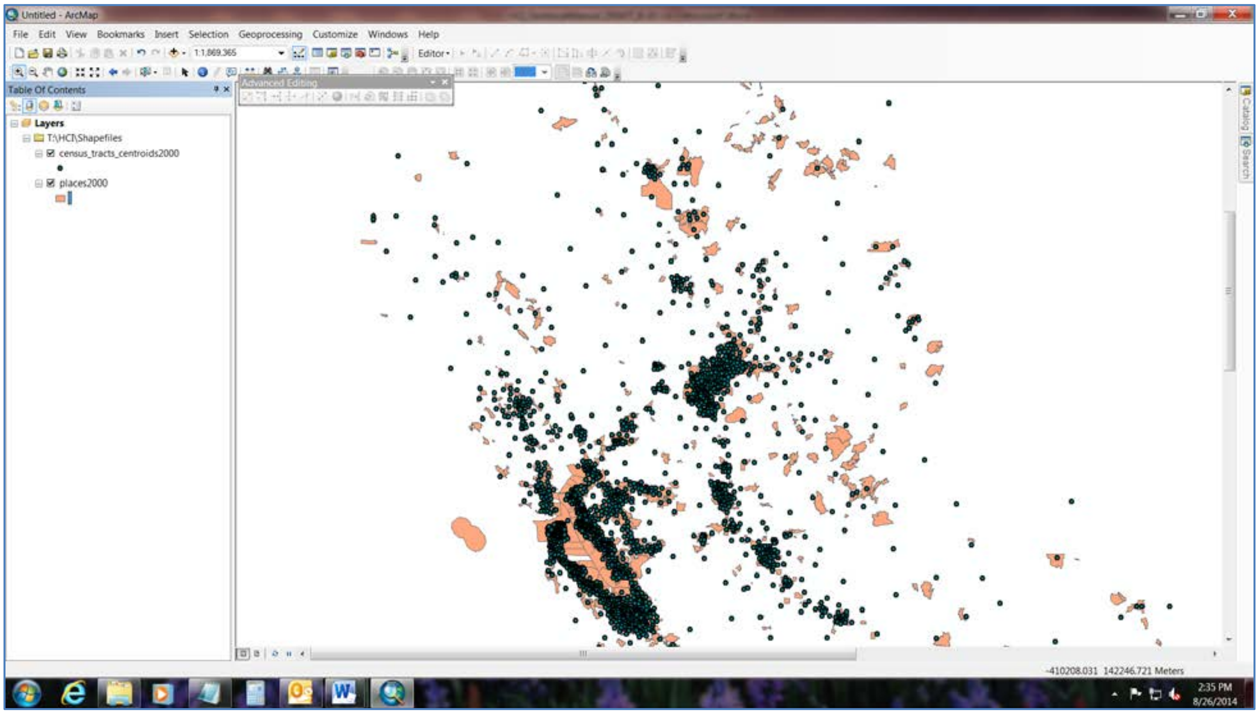

Untitled - ArcMap
File Edit View Bookmarks Insert Selection Geoprocessing Customize Windows Help
Table Of Contents
Layers
H:\DOCUMENTS
Copy
Remove
Open Attribute Table
Joins and Relates
Zoom To Layer
Zoom To Make Visible
Visible Scale Range
Use Symbol Levels
Selection
Label Features
Edit Features
Convert Labels to Annotation...
Convert Features to Graphics...
Convert Symbology to Representation...
Data
Save As Layer File...
Create Layer Package...
Properties...
Repair Data Source...
Export Data...
Export Data
Save this layer's data as a shapefile or geodatabase feature class
Export to a SAS data set
ArcToolbox
3D Analyst Tools
Analysis Tools
Cartography Tools
Conversion Tools
Data Interoperability Tools
Data Management Tools
Layers and Table Views
Make Feature Layer
Make XY Event Layer
Save To Layer File
Select Layer By Attribute
Select Layer By Location


Untitled - ArcMap
Table Of Contents
Layers
H:\DOCUMENTS
census_tracts2000centroids_Layer
census_tracts2000centroids.txt
T:\HCI\Shapefiles
census_tracts2000
Saving Data
Look in: H:\DOCUMENTS
ADMIN
BARHII
CalEnviroScreen
CHIS
Data Report Oct 13
Deleted from HCI
Emergencies
Evacuation Team Assigments
HCI
Health Equity Data Webinar 11-7-13
Health inequities reports
Housing
ITHIM-MTC
Logos
Mail
NCBI abstracts
OHE
Others
Name: census_tracts2000centroids.shp
Save as type: Shapefile
Save
Cancel

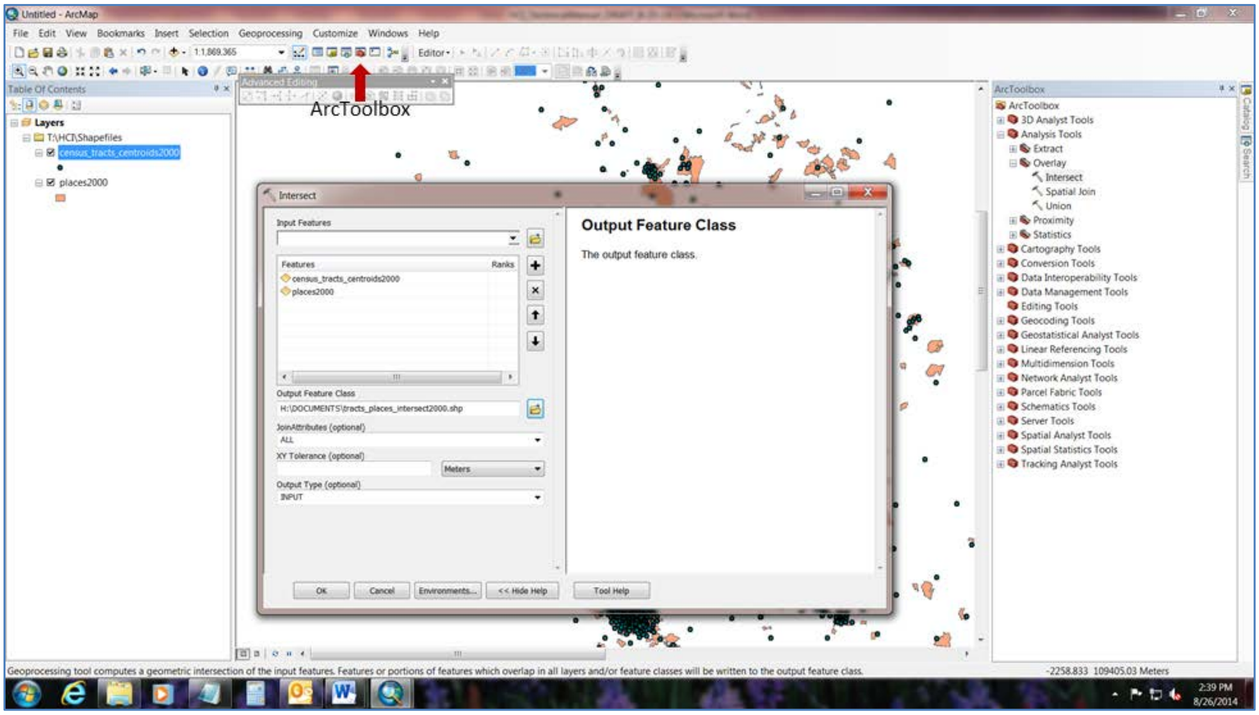
Intersecting Two Shapefiles in ArcMap

The intersection tool in ArcMap was used to allocate tracts or blocks centroids (points) to other geographical features (polygon). The following steps illustrate how to intersect tract centroids and places. The Toolbox chapter "Calculating Centroids in ArcMap and Creating a Centroid Shapefile" presents details on how to create a centroid shapefile.

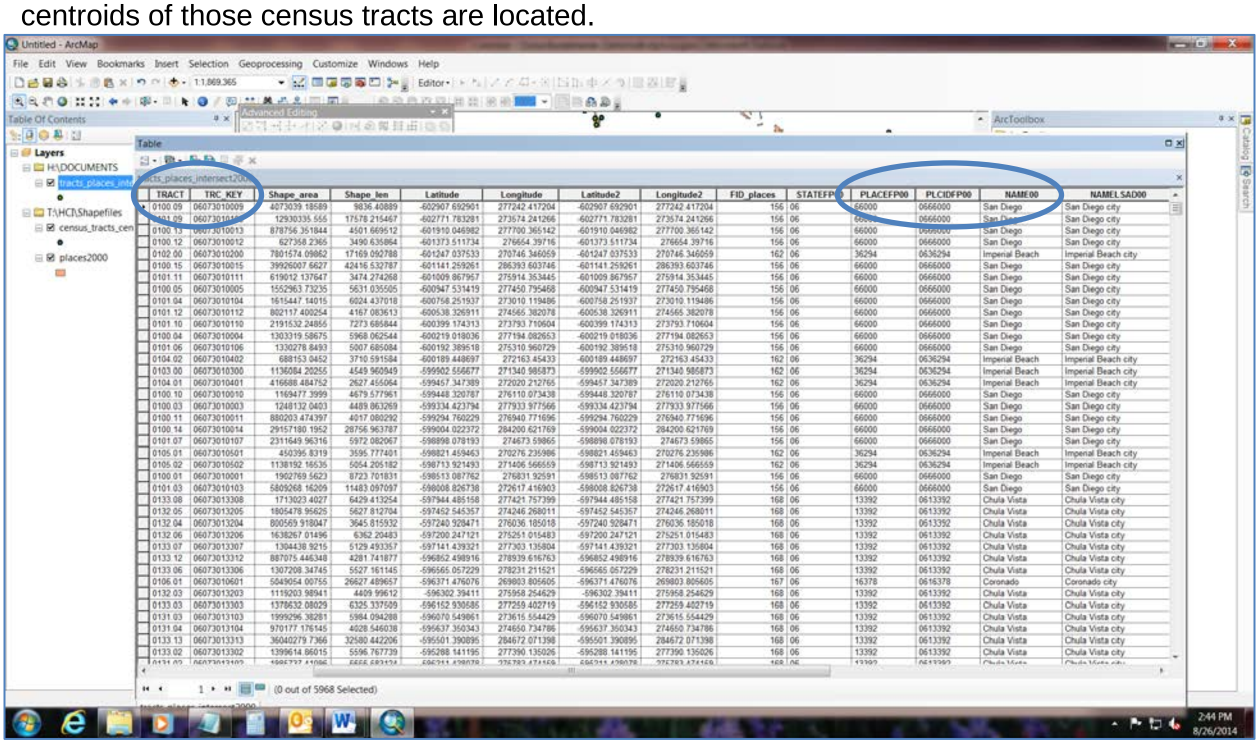
Step 1. Open the Census tract centroids (points) shapefile and places (polygon) shapefile.



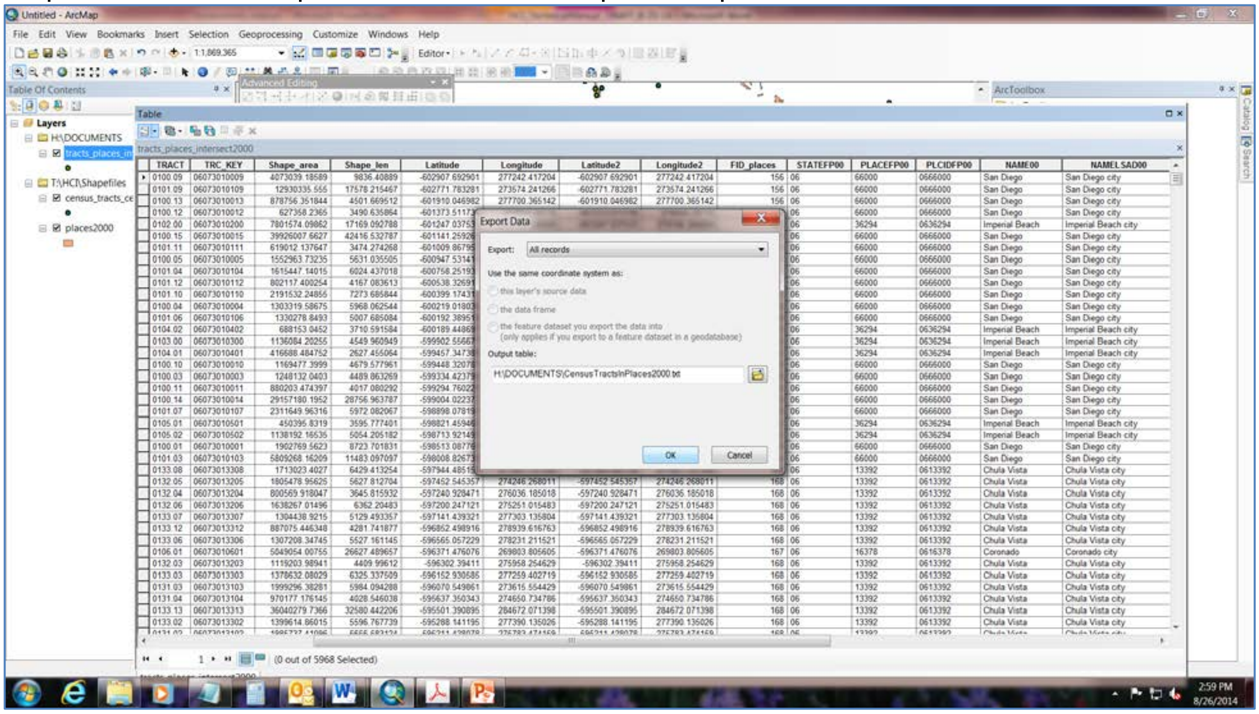
Step 2. Navigate to ArcToolbox > Analysis Tool > Overlay > Intersect. Select the 2000 Census tract centroids shapefile and the places shapefile from the "Input Features" dropdown list. Provide a new name for the "Output Feature Class" (.shp) and click "OK".



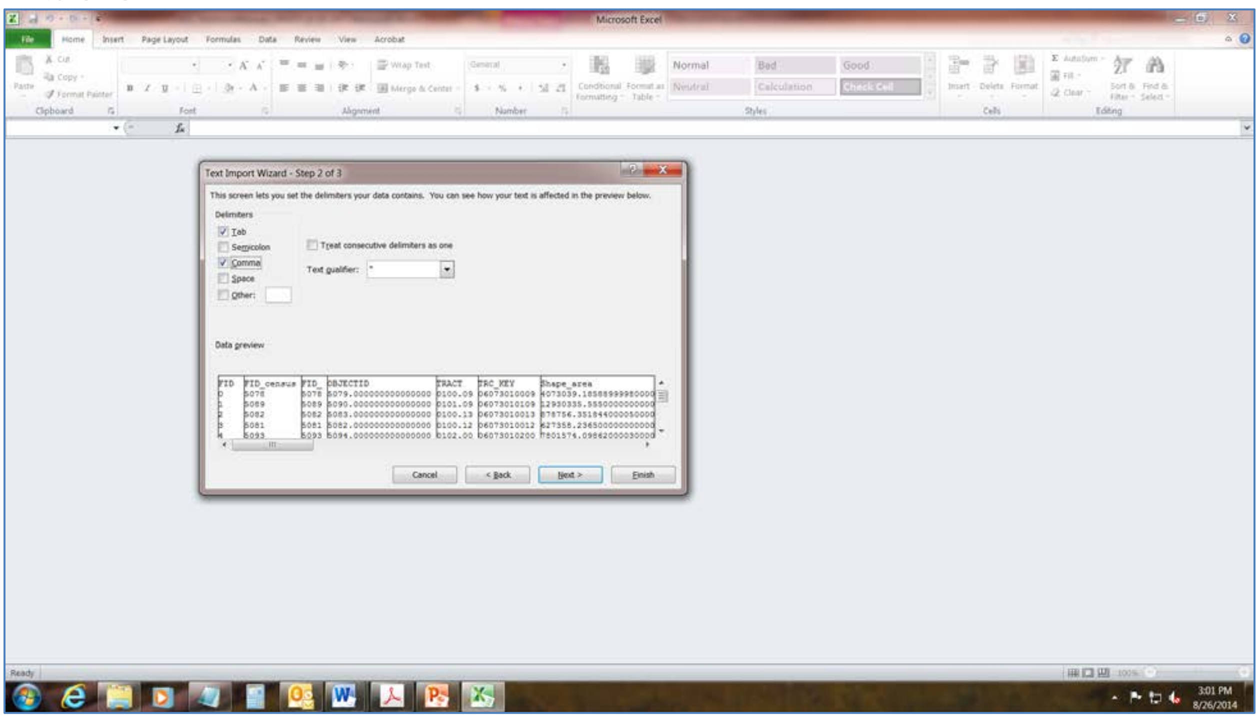
Step 3. The attribute table of the new shapefile (.shp) created includes the 2000 census tract codes (TRC_KEY) and the name (NAME) and code (PLACE) of places in California where the centroids of those census tracts are located.



Step 4. Go to “Table Options” and click on “Export.” Export the table as a .txt file.



Step 5. Open the .txt table in Excel and save it as an .xls file. This file is ready to be imported into SAS.



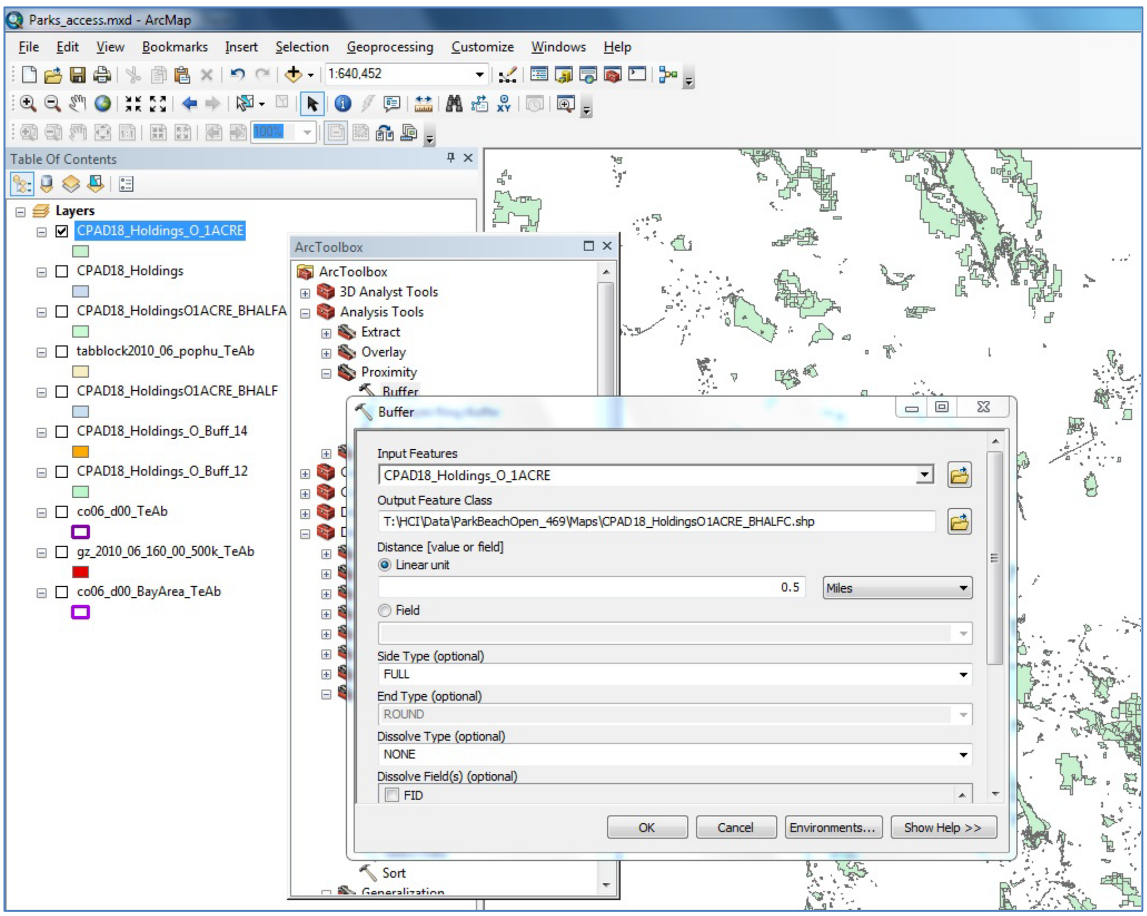
Creating Buffers around Features and Select by Location in ArcMap

The following example explains how to create half mile buffers (“crow’s fly”) around parks and open spaces (polygon features) and how to select the Census blocks (polygon features) that fall within the buffers. The same steps can be followed to create buffers around point features and to select the points that fall within buffers.

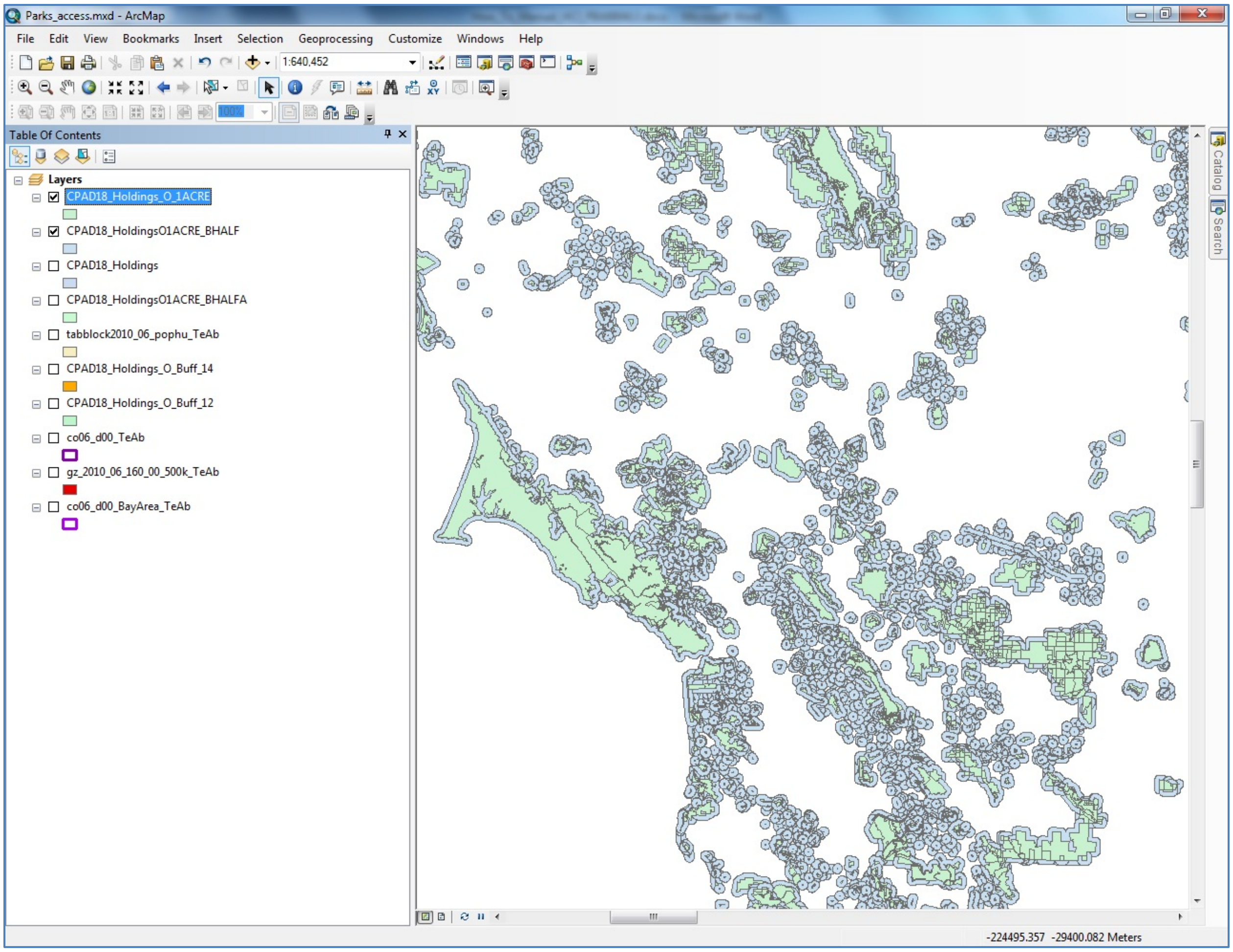
<u>Creating buffer around parks and open spaces</u>

Step 1. To create a half mile buffer around parks and open spaces (polygons), go to ArcToolbox > Analysis Tools > Proximity > Buffer.

a. For Input Features choose shapefile of parks over 1 acre designated as ‘Open Access’.
b. For Output Feature Class, choose the destination for the buffer file.
c. For Distance, select 0.5 Miles.
d. Leave Dissolve Type as None and click OK.

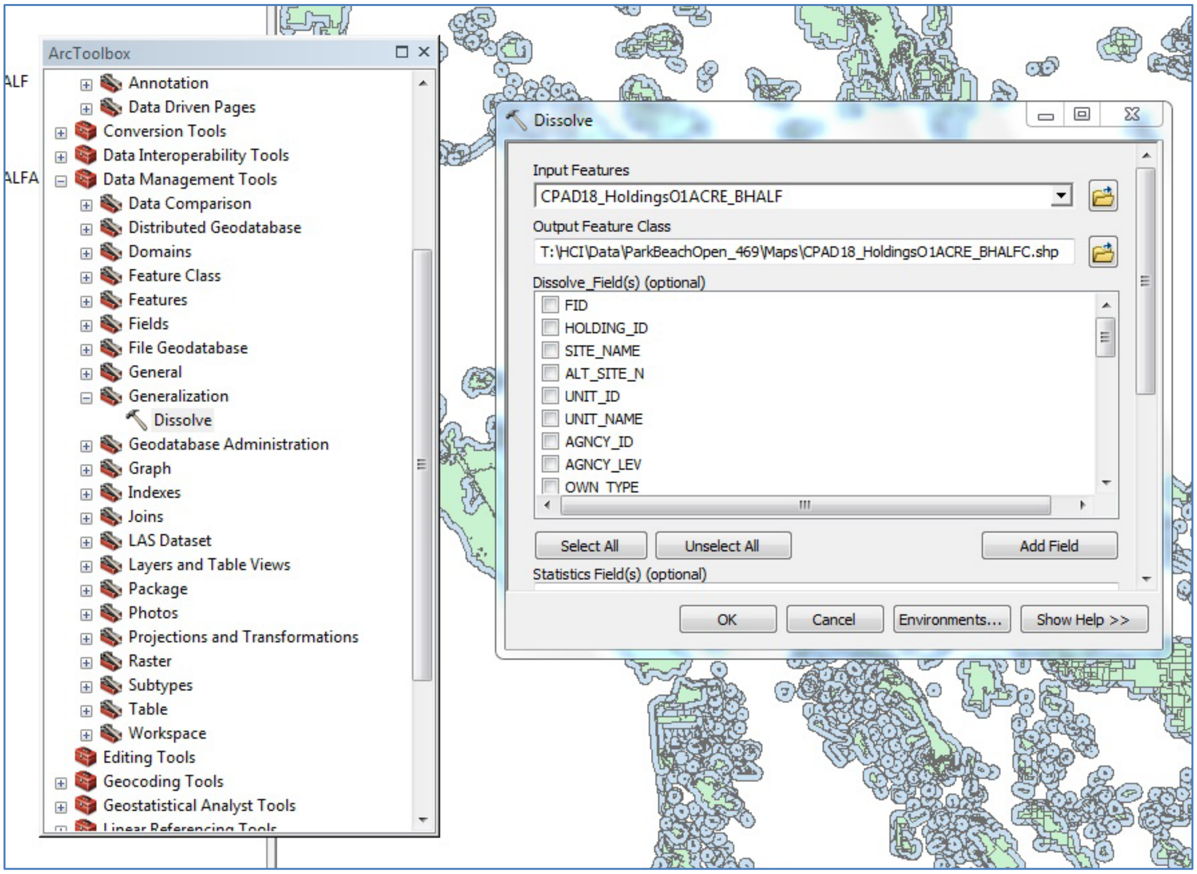


Step 2. After half mile buffer has been created, add it to map project to verify and check.



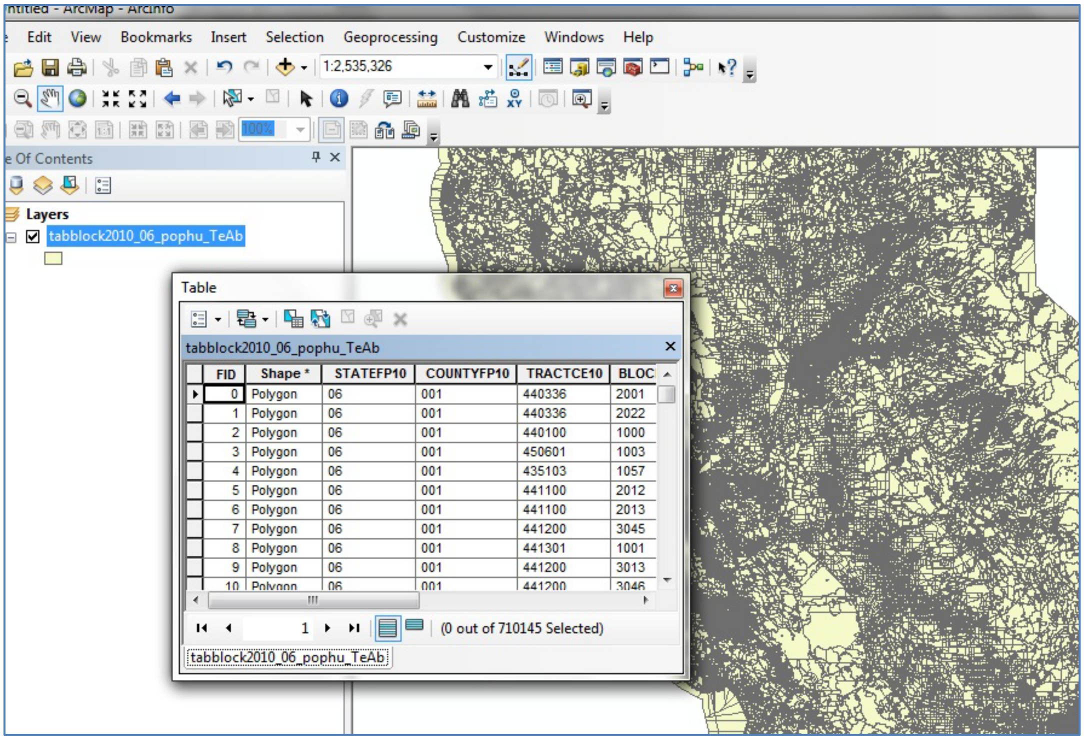
Step 3. Dissolve parks & open space buffer file into 1 shape (from several thousand shapes). In ArcToolbox select Data Management Tools > Generalization > Dissolve.

a. For Input Features select buffer file.
b. For Output Feature Class, select destination for output buffer file.

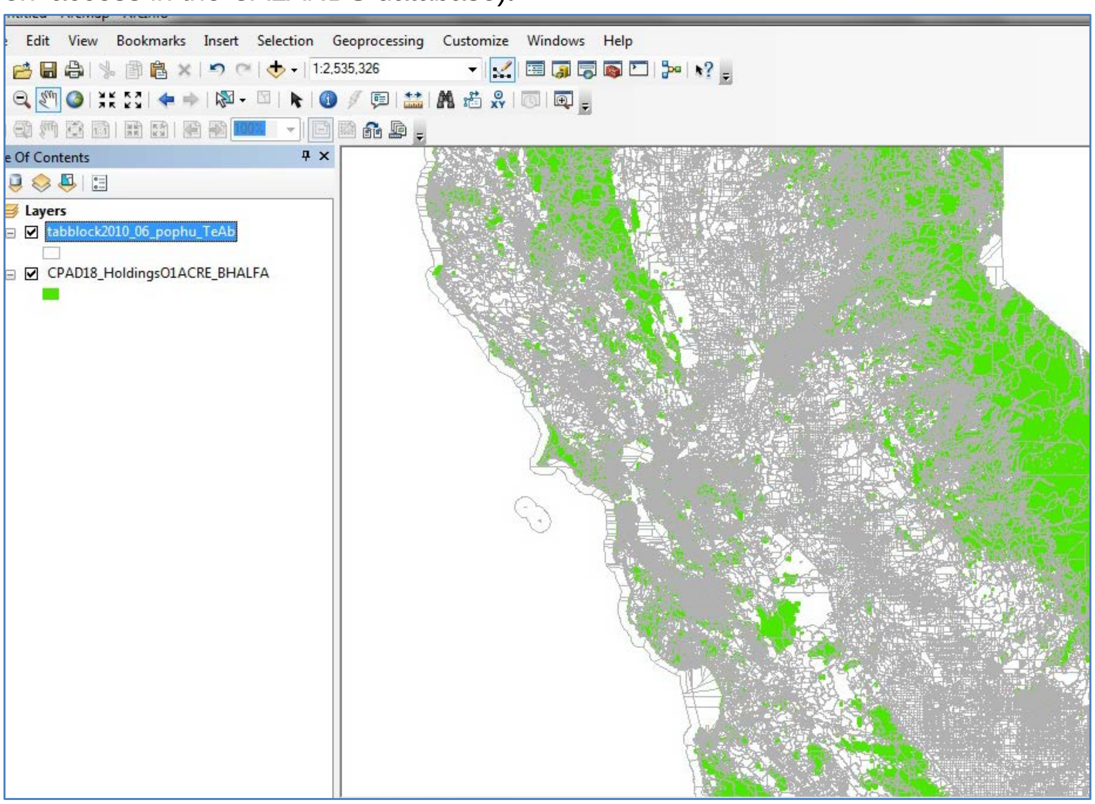


<u>Selecting Census blocks within parks and open space buffer</u>

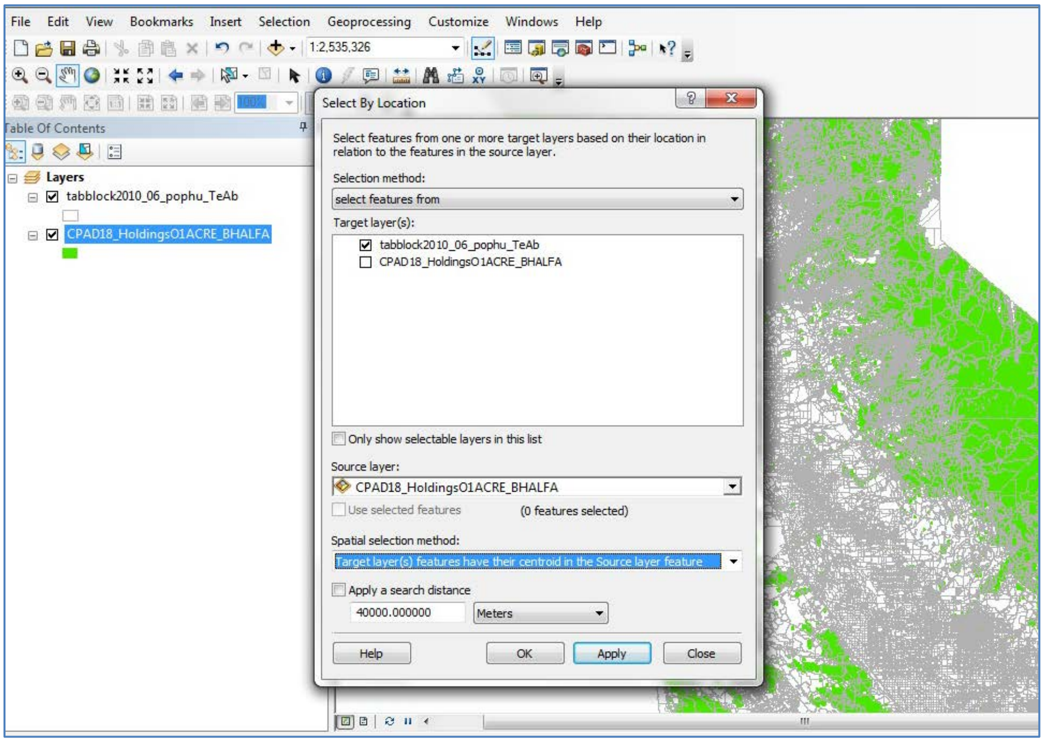
Step 1. In ArcMap, open the 2010 Census block file.



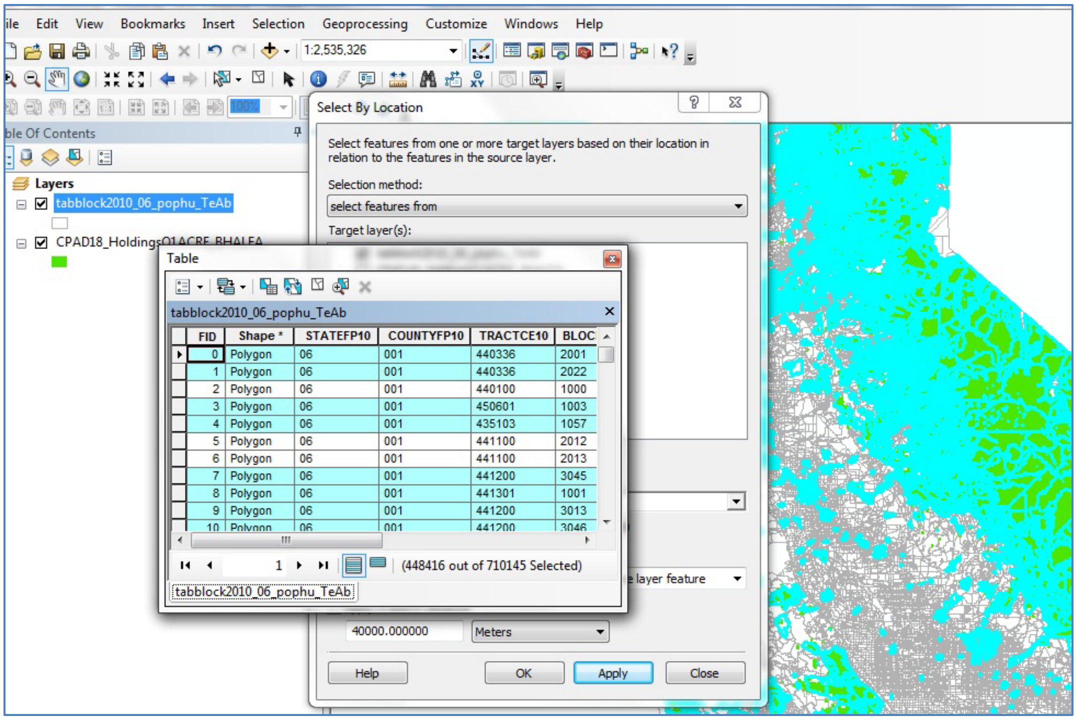
Step 2. Open the dissolved parks and open space buffer layer created above (see “Creating buffer around parks and open spaces”), a shapefile layer including all areas within ½ mile of a park/open space border (including only parks/open space above 1 acre in size and classified as ‘open’ access in the CALANDS database).



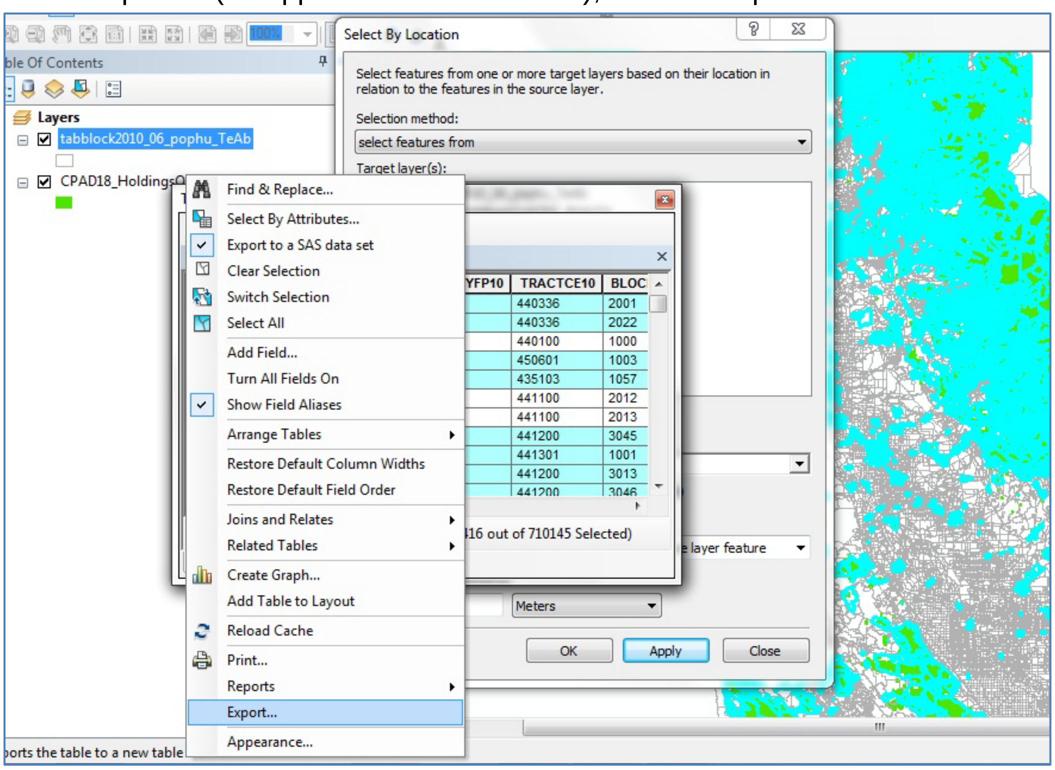
Step 3. From ‘Selection’ menu, choose ‘Select by Location.’ Select 2010 TIGER blocks (target layer) which have their centroid within the parks buffer area (source layer). Do not apply a search distance.



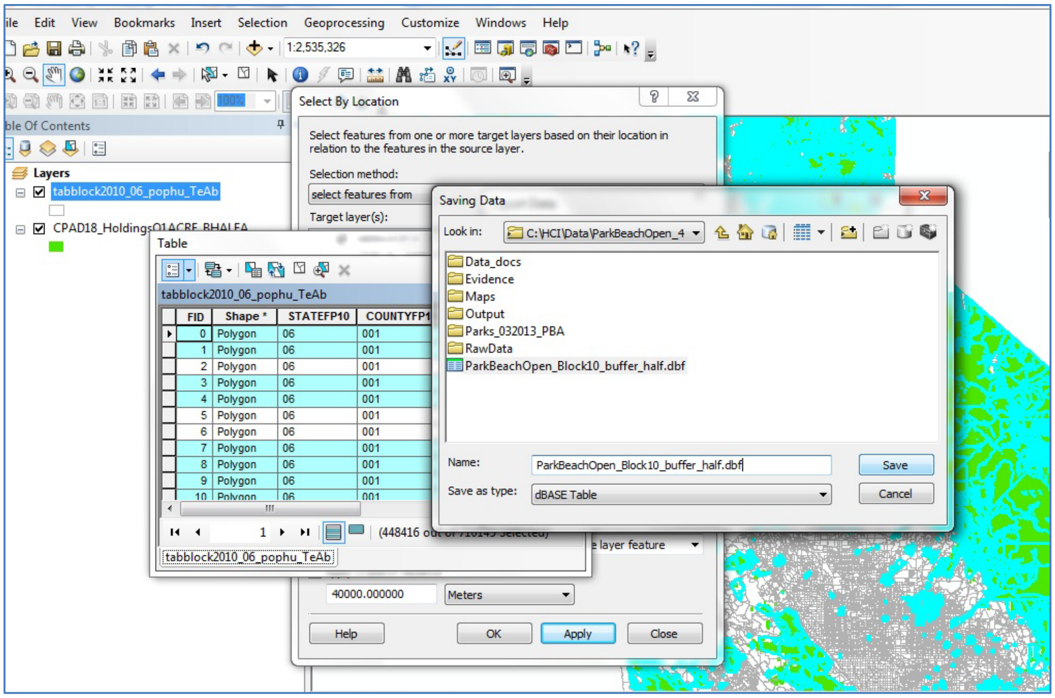
Step 4. When selection process is finished, right click and open 2010 TIGER block attribute table. The number of blocks selected will be highlighted.



Step 5. In ‘Table Options’ (on upper left corner of table), choose ‘Export’.



Step 6. For ‘Save as Type’ choose dBASE Table (.dbf).



Step 7. Open dbf block file in Excel and save as Excel file. File is now ready to be imported into SAS.

Standard Error Calculations

Standard Error from Margin of Error for U.S. Census Data

The standard error (SE) of a survey estimate ($\hat{X}$) indicates the extent to which this estimate is likely to deviate from the true population value. The U.S. Census publishes margins or error (MOE, 90% Confidence Level) for the estimates provided by the American Community Survey. The corresponding standard error can be computed using the following formula:

$$SE(\hat{X}) = \frac{MOE(\hat{X})}{1.645}$$

Standard Error of a Proportion or Percent for U.S. Census Data

A proportion (or percent) is a ratio where the numerator is a subset of the denominator.

$$\hat{P} = \frac{\hat{X}}{\hat{Y}}$$

Where $\hat{X}$ and $\hat{Y}$ are two survey estimates. The U.S. Census indicates that the standard error of a proportion can be approximated as:

$$SE(\hat{P}) = \frac{1}{\hat{Y}}\sqrt{[SE(\hat{X})]^2 - \frac{\hat{X}^2}{\hat{Y}^2}[SE(\hat{Y})]^2}$$

If $\hat{Q} = \hat{P} \times 100$ is the corresponding percent of the proportion, then $SE(\hat{Q}) = SE(\hat{P}) \times 100$.
If the value under the radical is negative, the minus sign should be replaced with a plus sign.

Standard Error of the Sum of Two or more Estimates for U.S. Census Data (Approximate Method)

The standard error for the sum or the difference between two or more sample estimates ($\hat{X}_1$ and $\hat{X}_2$) can be approximated using the following formula:

$$SE(\hat{X}_1 \pm \hat{X}_2) \approx \sqrt{SE(\hat{X}_1)^2 \pm SE(\hat{X}_2)^2}$$

This method will overestimate or underestimate the standard error if the estimates interact in a positive or negative way and if we do not have a covariance estimate. This method can be expanded to more than two estimates but the approximate estimate will become increasingly different from the standard error as the number of estimates increases. It is recommended to work with the fewest amounts of estimates as possible.

Binomial Approximation of the Standard Error of a Proportion or Percent

When working with data surveys different to the ACS standard errors might not be available. It is possible to approximate the standard error for a binomial variable (proportion or percent $\hat{P}$, a ratio where the numerator is a subset of the denominator) as follows:

$$\hat{P} = \frac{X}{N}$$

$$SE(\hat{P}) = \sqrt{\frac{\hat{P} \times (1 - \hat{P})}{N}}$$

where X is the number of people with an attribute and N is the total number of people.

Poisson Approximation of the Standard Error of a Rare Events

If R is a rate where the numerator X is not a subset of the denominator Y:

$$R = \frac{X}{Y} \times 1{,}000$$

And X is a count of a rare event (i.e., severe injuries in a traffic collision, number of crimes in a city), it is possible to approximate the standard error for R as follows:

$$SE(R) = \frac{R}{\sqrt{X}}$$

Confidence Intervals Calculation

Confidence intervals represent the range in which the population value is likely to lie. The upper and lower 95% confidence bounds for an estimate ($\hat{X}$) can be calculated using the estimate and its standard error as follows:

$$95CI_{UL} = \hat{X} + \mathrm{SE}(\hat{X}) \times 1.96$$
$$95CI_{LL} = \hat{X} - \mathrm{SE}(\hat{X}) \times 1.96$$

Relative Standard Error Calculation

The relative standard error (RSE) is the standard error $SE(\hat{X})$ expressed as a fraction of the estimate $\hat{X}$ and is usually displayed as a percentage. Estimates with a RSE of 30% or greater are subject to high sampling error and should be used with caution.

$$RSE = \frac{SE(\hat{X})}{\hat{X}} \times 100$$

If the estimate is a proportion or a percent, the relative standard error is calculated as follows:

$$RSE = \frac{SE(\hat{X})}{\hat{X}} \times 100, if\ \hat{X} \leq 50\%$$

$$RSE = \frac{SE(\hat{X})}{100 - \hat{X}} \times 100, otherwise$$

Regional Level Estimate and Standard Error Calculations

Regional estimates and their standard errors are calculated from county level estimates. The regional estimate can be calculated with the following population weighted formula:

$$\hat{X}_{Region} = \frac{\sum_{i=1}^{n} \hat{X}_i N_i}{\sum_{i=1}^{n} N_i}$$

where n is the number of counties in a region, $\hat{X}_i$ is the estimate for county i, and N_i is the total population for county i. The standard error of this estimate is:

$$SE(\hat{X}_{Region}) = \sqrt{\frac{\sum_{i=1}^{n} N_i^2 \times [SE(\hat{X}_i)]^2}{(\sum_{i=1}^{n} N_i)^2}}$$

The list of the regions from the 2010 California Regional Progress Report can be found in *Appendix X*.

Deciles Calculation

A decile is one of ten equal groups into which a population can be divided according to the distribution of values of a particular variable. For the indicators in this project, the deciles are calculated for a single geographical level (usually cities/towns), all race/ethnicities (total population), and the most recent time period. The RANK procedure in SAS (http://support.sas.com/documentation/cdl/en/proc/61895/HTML/default/viewer.htm#a000146840.htm) is used to compute ranks for the indicator variable using the groups=10 option to calculate deciles. The RANK procedure assigns each geography (i.e., each city) to a group value ranging from 0 to 9. The results are later rescaled to values ranging from 1 to 10. The deciles can be used to sort and compare all geographies from highest to lowest values.

Relative Risk Calculation

The risk of occurrence of the event described by the indicator (educational attainment in adults, access to park, etc.) with respect to the California average is calculated as follows:

$$CA_{RR} = \frac{\hat{X}_{ijk}}{\hat{X}_{ijCa}}$$

Where $\hat{X}_{ijk}$ is the indicator estimate for report year i, race/ethnicity j, and geography k. $\hat{X}_{ijCa}$ is the California level estimate for the same report year and race/ethnicity.

Reliability Analysis

After each indicator file is completed with all the data available a reliability analysis is conducted for each geography, year, race-ethnicity, and other strata combination in order to determine the percent of the population in that combination of strata with missing or unreliable data. Analysis by all strata combinations is not possible for some indicators.

Using SAS, each strata combination is classified into one of three data reliability categories: RSE <30%, RSE ≥30%, RSE not available or missing. Population data is later aggregated by strata and by reliability category. The example in Table 34 shows the results of the aggregation of population by different strata for the Poverty Rate Indicator. The results show that four race/ethnicity groups have 50% or more of the population living in places were the estimates for the child poverty indicator are not stable: AIAN, NHOPI, multiple and other. The NHOPI group has 56% of the population living in places were the overall poverty rate estimates are not stable.

After completing the analysis is necessary to evaluate if unreliable estimates should be published in the data files. Decisions were made on a case to case basis. For some indicators, data was not published when 50% or more of the population lived in places or Census tracts with unreliable estimates (Census tract geography was not included in the Percent of Workers' Daily Commute 10 or more Minutes by Walking or Biking indicator because they were largely unreliable or equal to zero). But for other indicators the data was included because it was important to publish the information for some strata given their importance to health (percent of workers who walk/bike in the Mode of Transport to Work indicator).

Table 34. Example of Results of Reliability Analysis for Poverty Rate Indicator for the Strata geotype=Place, reportyear=2006-2010, Child and Overall Poverty, and all Race/Ethnicity Groups

Race/ Ethnicity	Not available/ missing	RSE <30%	RSE ≥30% (A)	Total Population with data available (B)	Total Population (C)	Percent Unreliable of available (A/B)	Percent Unreliable (A/C)
Child Poverty							
AIAN	470	1,067	4,942	6,009	6,479	**82.2**	76.3
Asian	25,324	555,704	343,897	899,601	924,925	38.2	37.2
AfricanAm	9,445	394,519	109,087	503,606	513,051	21.7	21.3
Latino	16,257	3,954,718	427,637	4,382,355	4,398,612	9.8	9.7
NHOPI	1,890	1,490	10,193	11,683	13,573	**87.2**	75.1
White	71,699	1,342,674	1,023,734	2,366,408	2,438,107	43.3	42.0
Multiple	30,653	107,254	161,525	268,779	299,432	**60.1**	53.9
Other	421	5,325	5,537	10,862	11,283	**51.0**	49.1
Total	39,686	7,944,854	841,115	8,785,969	8,825,655	9.6	9.5
Overall Poverty							
AIAN	0	25,373	8,543	33,916	33,916	25.2	25.2
Asian	15,431	3,781,677	696,594	4,478,271	4,493,702	15.6	15.5
AfricanAm	12,891	1,656,673	328,438	1,985,111	1,998,002	16.5	16.4
Latino	6,790	11,944,634	757,778	12,702,412	12,709,202	6.0	6.0
NHOPI	1,161	24,777	31,774	56,551	57,712	**56.2**	55.1
White	10,090	12,375,404	1,369,580	13,744,984	13,755,074	10.0	10.0
Multiple	7,050	327,278	312,921	640,199	647,249	48.9	48.3
Other	0	19,784	17,141	36,925	36,925	46.4	46.4
Total	14,999	33,103,156	1,377,230	34,480,386	34,495,385	4.0	4.0

HEALTHY COMMUNITIES DATA AND INDICATORS PROJECT META-DATABASE

Structure of the Database

The HCDIP meta-database is implemented as an MS Access 2010 database file. The structure of the database follows the organization of the Healthy Community Framework (Figure 2), which links indicators to 20 aspirational goals that are organized in 5 domains related to meeting basic needs, a sustainable and quality environment, adequate social and economic development, health and social equity, and social relationships that are respectful. The entity-relationship diagram is shown in Figure 10. Database tables are described in Table 35.

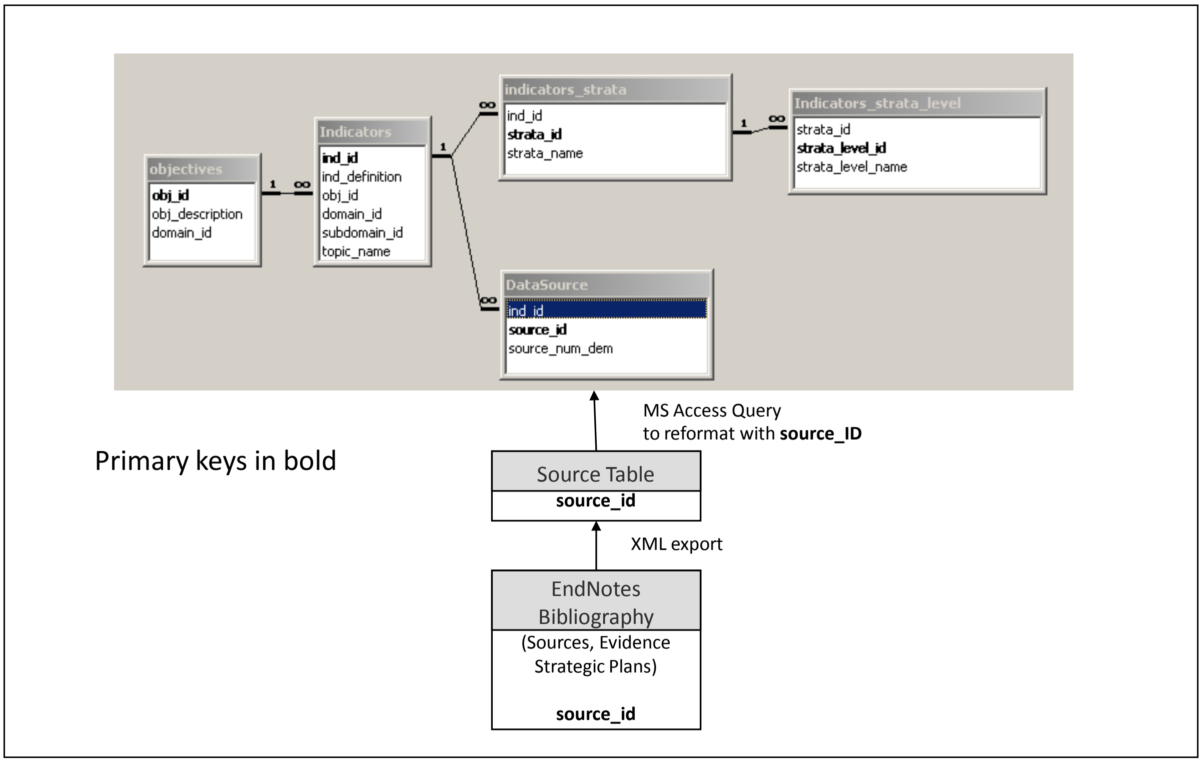


Figure 10. Entity-Relationship Diagram for Meta-Database of the Healthy Communities Data and Indicators Project

Table 35. Database Tables Comprising the HCDIP Meta-database

Table name	Description	No. of records
zDomains	Name and description of domains	5
Objectives	Listing of each aspiration	20
Indicators	Metadata and descriptive information on each indicator	60
Indicators_strata	Names of strata	15
Indicators_strata_levels	Levels of each strata	93
Followup	Administrative data to manage workload and progress of each indicator	20
Biblio_import_tabdl	Table of imported bibliographic references	1500+
Biblio_updated	Table with additional bibliographic information	1500+

The domains table (zDomains) functions as a descriptive look-up to associate domains with each objective. There is a one-to-many relationship between the *Healthy Community Framework* objectives and indicators. To create a manageable number of indicators, approximately 3 indicators were selected for each of the 20 objectives. Some indicators are stratified by race/ethnicity and other characteristics. In this data model the primary entity is the indicator. There is a one-to-many relationship between strata of an indicator and the indicator. The meta-database provides the standardized code lists (keys) for strata names and levels. Because some indicators can have a large number of strata, defining indicators as "indicators-strata name-strata level" creates a false impression of more content than actually exists. So, for the purposes of HCDIP, "indicators" are defined without strata.

A data dictionary for the meta-database was created as an Excel file (HCI_Metadata_DictionaryM-D-YY.xls) and essential details of the primary keys and names of fields in currant usage are presented in Table 36-39.

<u>Bibliographic References</u>

Bibliographic references were key-entered into an EndNotes 14 library file (HCI_BibliographyM-D-YY.enl) using only the templates for reports and journal articles. Several EndNotes bibliographic fields were "cannibalized" and given altered meanings so that information of a reference could be linked to an indicator. These fields included Accession Number which was used to designate the 3-digit indicator ID number (ind_id). There is a many-to-one relationship between references and indicators. When a reference applied to more than one indicator, the entry was copied in EndNote and the Accession Number field was changed to indicate a second (or higher) indicator to which the information applied. EndNote references were exported using an output style of tab delimited text format, which was specific to journal articles and reports. The tab delimited file uploaded into an MS access table ('biblio_import_tabdl'). Because the EndNotes library references are cumulative, older versions of 'biblio_import_tabdl' were deleted but the file structure was retained. The MS Access import tool for external data was used with the tab delimited and append options.

<u>Queries, Forms, and Reports</u>

Queries

Queries were written to recompose bibliographic references in MS access, create lists of indicators by domain and objective, create dynasets of strata and references for display in forms, filter indicators that were on the draft core list, and assign staff specific follow-up activities. Table 6 describes these queries and provides the SQL syntax.

Forms

The meta-database has two main forms: 'hci_summary' and 'hci_summary_followup'. The form 'hci_summary' (Figure 11) facilitates for data entry and display of meta-data and narrative content for a single indicator. There are two main subforms, each of which was created by dragging and dropping a table into the main form in which a link (parent) field was the indicator identification (ind_id). The subforms represent a many-to-one association between strata and references to each indicator. One subform (Indicators_strata subform1) displays the strata and the strata levels associated with a specific indicator. A second subform (Evidence_references subform) displays the bibliographic references. Applying a two character code to filter the ‘reftype’ (Type) field in the subform allows users to select references pertaining to data sources (DS), evidence (EV), or indicator projects (IR). The ‘hci_summary’ form was designed to fit on two letter-size printed pages (8.5 by 11 inches) with 0.5 inch margins and has a 12 point Arial font for text.

The second form ‘hci_summary_follow-up’ (Figure 12) facilitates key entry and display of administrative data on the status of follow-up included the indicator of interest, the issues that require follow up, the person assigned to conduct follow up, the actions taken, and a proposed deadline by which the follow-up active should take place, and the date follow-up was completed. Because there can be multiple issues with a single indicator, there is a many-to-one relationship between the indicators table and follow-up table reflected in the Follow-up subform.

Reports

The meta-database has one report to display the status of follow-up activities (Figure 13) based on the query ‘follow-up_status’.

Using the Meta-Database Model to Create Data Files

The meta-database serves as reference standard and repository for all content including metadata; explanatory text; and titles of graphs, tables, and maps. In the proof-of-concept demonstration web site (non-public) there is an explicit linkage between the meta-database and the data files (values table), which allows the seamless integration of content. However, The content of the report templates (pdf files) posted at www.cdph.ca.gov/programs/Pages/HealthyCommunityIndicators.aspx was not created with a direct link to the meta-database. Narrative content was cut and pasted from the MS Access data file.

The meta-database structure supports the creation of standardized data files in which the row entity is in a linear combination of primary keys of indicator, strata name, and strata level. Most downloadable Excel data files posted at the HCDIP website follow this format. This meta-database structure also supports the creation of data files with several related indicators appearing as separate columns for the same strata. This is done as a convenience to users to reduce the number of data files they handle for related indicators. This format was followed for the downloadable Excel poverty data file at the HCDIP website, which includes indicators for overall poverty (all ages), childhood poverty, and concentrated poverty in the same data file.

Table 36. Description of Primary Keys and Attribute Fields for Domains, Objectives, and Indicator Tables in the HCDIP Meta-database

Table	Name	Definition	Type	Width	Coding
zDomain	**Domain_id**	3-character code for domain	T	3	See code lists
	Domain_name	Name of Domain	T	20	See code lists
	Domain_desc	Description of this domain	M	64,000	
Objectives	**Domain_id**	3-character code for domain	T	3	Domain_id
	Obj_id	Number of objective within domain	N	3	Obj_id
	Obj_description	Healthy Community description of aspirational goal	T	255	Obj_description
Indicators	**ind_id**	ID/key of indicator	T	3	Autonumber from 1-999
	ind_definition	Definition of indicator in plain language	T	255	Free text
	obj_id	ID/key of objective	N	2	
	domain_id	3-character code for domain	T	3	
	PCDL	Preliminary Core Draft List	T	1	Y/N (blank=N)
	Comments	Comments about indicator	T	255	Free text
	ind_numerator	Technical definition of outcome (numerator)	T	255	Free text
	ind_denominator	Technical def'n of population to which outcome is applicable	N	255	Free text
	ind_evidence	Narrative of evidence linking to health outcome	T	64,000	Free text (memo field)
	ind_status	Administrative field used to indicate data has been checked	N	1	1, not started; 2, arranging data acquisition; 3, data acquired;4; processing; 5, ready to use
	years_av	Years Available	T	255	Year or range of years
	freq	Frequency of updates	T	255	
	ind_data_collection_method	Method of data collection	T	2	Survey, administrative, modeled
	ind_data_quality	Data quality	T	100	Free text
	ind_explanations_limitations	Indicator explanations and limitations about data source and calculation of this indicator	T	64,000	Free text (memo field)
	ind_significance	Narrative of the significance of indicator	T	64,000	Free text (memo field)
	ind_data_desc_summ	Indicator Data Description Summary	T	256	Free text

T, Text; N, Numeric, D, date
Note: primary keys in **bold**

Table 37. Description of Primary Keys and Attribute Fields for Indicator Values Table in the HCDIP Meta-database

Table	Name	Definition	Type	Width	Coding
Values	**ind_id**	HCI ID/key of indicator	N	3	
	ind_definition	Definition of indicator in plain language	T	255	Free text
	reportyear	Year(s) that the indicator was reported	T	9	CCYY-CCYY
	race_eth_code	numeric code for a race/ethnicity group	T	1	See strata table
	race_eth_name	Name of race/ethnic group	T	25	See strata table
	geotype	Type of geographic unit	T	2	CT,Census Tract; N,Neighborhood; PL,Place; ZC,Zipcode; CO, County; RE, region; CA, State
	geotypevalue	Value of geographic unit	T	See specifics under coding	FIPS ID; place code (5-characters in length), CT code (11-characters in length starting with 06), zip code, region id
	geoname	Name of geographic unit	T	25	place, county, region name se code list
	county_name	Name of county that geotype is in	T	25	county name
	county_fips	FIPS code of county that geotype is in	N	5	See Code list
	region_name	Name of region	T		See code list
	regioncode	Numeric code of region	T		
	numerator	Numeric value of numerator of indicator	N	double	
	denominator	Numeric value of denominator	N	double	
	percent	Numerator divided by denominator for basis	N	double	
	LL_95CI	Lower limit of 95% confidence interval	N	double	
	UL_95CI	Lower limit of 95% confidence interval	N	double	
	SE	Standard error of percent	N	double	
	RSE	Relative standard error (se/percent * 100) expressed as a percent	N	double	
	CA_decile	Statewide decile ranking	N	integer	
	CA_RR	Ratio of indicator in geographic area to state average	N	double	
	Version	Date/time stamp of version of data	T	1	mm/DD/CCYY hh:mm:ss

T, Text; N, Numeric, D, date

Table 38. Description of Primary Keys and Attribute Fields for Bibliography Table in the HCDIP Meta-database

Table	Name	Definition	Type	Width	Variable Name in EndNotes
Biblio_import_tabdl	rec-number	ID of the reference in EndNotes	N	Long Integer	rec-number
	Type	type of bibliographic document (report or journal article)	T	15	Reference Type
	Author	author(s) of report/journal article	T	150	Author
	Year	year of report/article	T	50	Year
	Title	title of report/article	T	150	Title
	journal_name	journal name	T	50	Secondary title
	City	city of institution (reports)	T	25	Place Published
	Institute	institution (reports)	T	50	Publisher
	Volume	journal volume	T	10	Volume
	Number	journal number	T	10	Number
	Pages	journal pages	T	15	Pages
	Reftype	Type of reference: data source (DS), evidence (EV), indicator project/report (IR), methods (ME), reference for evidence (RE)	T	2	Type of Article
	ind_id	ID/key of indicator	N	Integer	Accession Number
	N/D	Data source for numerator (N), denominator (D), or both (B), or quantitative indicator (Q)	T	1	Call Number
	Notes	notes	T	200	Comments on indicator
	URL	URL of report/article	T	250	URL
	endnote_import_date	Date imported from Endnotes to MS Access	D		
Additional fields in Biblio_updated					Coding
	data_own_type	public, private, or nonprofit ownership of data	T	25	Text (public, private, or nonprofit)
	year_1	1st year available	T	50	4-digit year
	freq	frequency of updates	T	50	Free text

T, Text; N, Numeric, D, date

Table 38. Description of Primary Keys and Attribute Fields for Bibliography Table in the HCDIP Meta-database

Table	Name	Definition	Type	Width	Coding
Followup	**ind_id**	ID/key of indicator	N	3	
	issue_id	ID/key of issue	N	3	
	person_id	Person assigned/responsible for follow-up	T	2	Initials
	date_assigned	date the issue was assigned for follow-up	D	10	
	issue_description	description of the issue to be followed-up	T	255	
	proposed_followup	Description of proposed follow-up	T	255	
	date_due	Date the proposed follow-up activity should be resolved/re-checked	D	10	
	date_done	Date the follow-up was actually completed	D	10	
	followup_done	Description of what follow-up was actually done	T	255	

T, Text; N, Numeric, D, date

Table 39. Strata and Strata Levels

Strata ID	Strata Name	Strata level ID	Strata Level Name
1	Race/ethnicity	1	American Indian/ Alaska Native
1	Race/ethnicity	2	Asian
1	Race/ethnicity	3	African American
1	Race/ethnicity	4	Latino
1	Race/ethnicity	5	Native Hawaiian Other Pacific Islander
1	Race/ethnicity	6	White
1	Race/ethnicity	7	Multiple
1	Race/ethnicity	8	Other
1	Race/ethnicity	9	Total
2	Geography (5 strata)	1	Census tract
2	Geography (5 strata)	2	Place
2	Geography (5 strata)	3	County
2	Geography (5 strata)	4	Region
2	Geography (5 strata)	5	State
3	Geography (4 strata)	1	Place
3	Geography (4 strata)	2	County
3	Geography (4 strata)	3	Region
3	Geography (4 strata)	4	State
4	Geography (4 strata with Zip code)	1	Zip code
4	Geography (4 strata with Zip code)	2	Place
4	Geography (4 strata with Zip code)	3	County
4	Geography (4 strata with Zip code)	4	Region
4	Geography (4 strata with Zip code)	5	State
5	Geography (6 strata with CMSA)	1	Census tract
5	Geography (6 strata with CMSA)	2	Place
5	Geography (6 strata with CMSA)	3	County
5	Geography (6 strata with CMSA)	4	Region
5	Geography (6 strata with CMSA)	5	Consolidated Metropolitan Statistical Area
5	Geography (6 strata with CMSA)	6	State
6	Geography (5 strata with CMSA)	1	Place
6	Geography (5 strata with CMSA)	2	County
6	Geography (5 strata with CMSA)	3	Region
6	Geography (5 strata with CMSA)	4	Consolidated Metropolitan Statistical Area
6	Geography (5 strata with CMSA)	5	State
7	Geography (3 strata)	1	Census tract
7	Geography (3 strata)	2	Place
7	Geography (3 strata)	3	County
8	Mode of transportation to work	1	Works at home
8	Mode of transportation to work	2	Bicycle

Strata ID	Strata Name	Strata level ID	Strata Level Name
8	Mode of transportation to work	3	Car
8	Mode of transportation to work	4	Carpool
8	Mode of transportation to work	5	Car total
8	Mode of transportation to work	6	Public transportation
8	Mode of transportation to work	7	Walking
9	Victim mode of transportation	1	All modes
9	Victim mode of transportation	2	Bicyclist
9	Victim mode of transportation	3	Bus
9	Victim mode of transportation	4	Car/pickup
9	Victim mode of transportation	5	Motorcycle
9	Victim mode of transportation	6	Pedestrian
9	Victim mode of transportation	7	Trucks
9	Victim mode of transportation	8	Vehicles
10	Mode (miles traveled)	1	Bicycle
10	Mode (miles traveled)	2	Pedestrian
10	Mode (miles traveled)	3	Bicycle
11	Mode (walk/bike)	1	Walk
11	Mode (walk/bike)	2	Bicycle
12	Housing cost burden	1	> 30%
12	Housing cost burden	2	> 50%
12	Housing cost burden	3	≥ 30%
12	Housing cost burden	4	≥ 50%
13	Family type	1	Adult1
13	Family type	2	Adult1_Child1
13	Family type	3	Adult1_Children2
13	Family type	4	Adult1_Children3
13	Family type	5	Adult1_Children
13	Family type	6	Adults2
13	Family type	7	Adults2_Child1
13	Family type	8	Adults2_Children2
13	Family type	9	Adults2_Children3
13	Family type	10	Adults2_Children
13	Family type	11	MarriedCouple2Children
13	Family type	12	SingleMother2Children
14	Industry	1	All Industry
14	Industry	2	Agriculture, Forestry, Fishing and Hunting
14	Industry	3	Mining
14	Industry	4	Construction
14	Industry	5	Manufacturing
14	Industry	6	Wholesale Trade

Strata ID	Strata Name	Strata level ID	Strata Level Name
14	Industry	7	Retail Trade
15	Injury level	1	Killed
15	Injury level	2	Severe injury
16	Housing tenure	1	Renter-occupied households
16	Housing tenure	2	Owner-occupied households
16	Housing tenure	3	Total households
17	Household crowding	1	Overcrowding (> 1.0 person per room)
17	Household crowding	2	Severe overcrowding (> 1.5 persons per room)
18	Data source for housing	1	American Community Survey (ACS)
18	Data source for housing	2	Comprehensive Housing Affordability Strategy(CHAS)
19	Income level	1	Monthly household income at ≤ 30% of HAMFI
19	Income level	2	Monthly household income at all level of HAMFI
20	Type of participation	1	registered/eligible
20	Type of participation	2	voted/registered
21	Geography (6 strata with County division)	1	Census tract
21	Geography (6 strata with County division)	2	Place
21	Geography (6 strata with County division)	3	County division
21	Geography (6 strata with County division)	4	County
21	Geography (6 strata with County division)	5	Region
21	Geography (6 strata with County division)	6	State
22	Violation category	1	Treatment technique (TT)
22	Violation category	2	MCL Total Coliform Rule-Bacteriologic (tcrb)
22	Violation category	3	Arsenic
22	Violation category	4	Nitrates
22	Violation category	5	All other MCL
23	Type of alcohol outlet sales	1	On-sale alcohol outlet
23	Type of alcohol outlet sales	2	Off-sale alcohol outlet
23	Type of alcohol outlet sales	3	All sales alcohol outlet

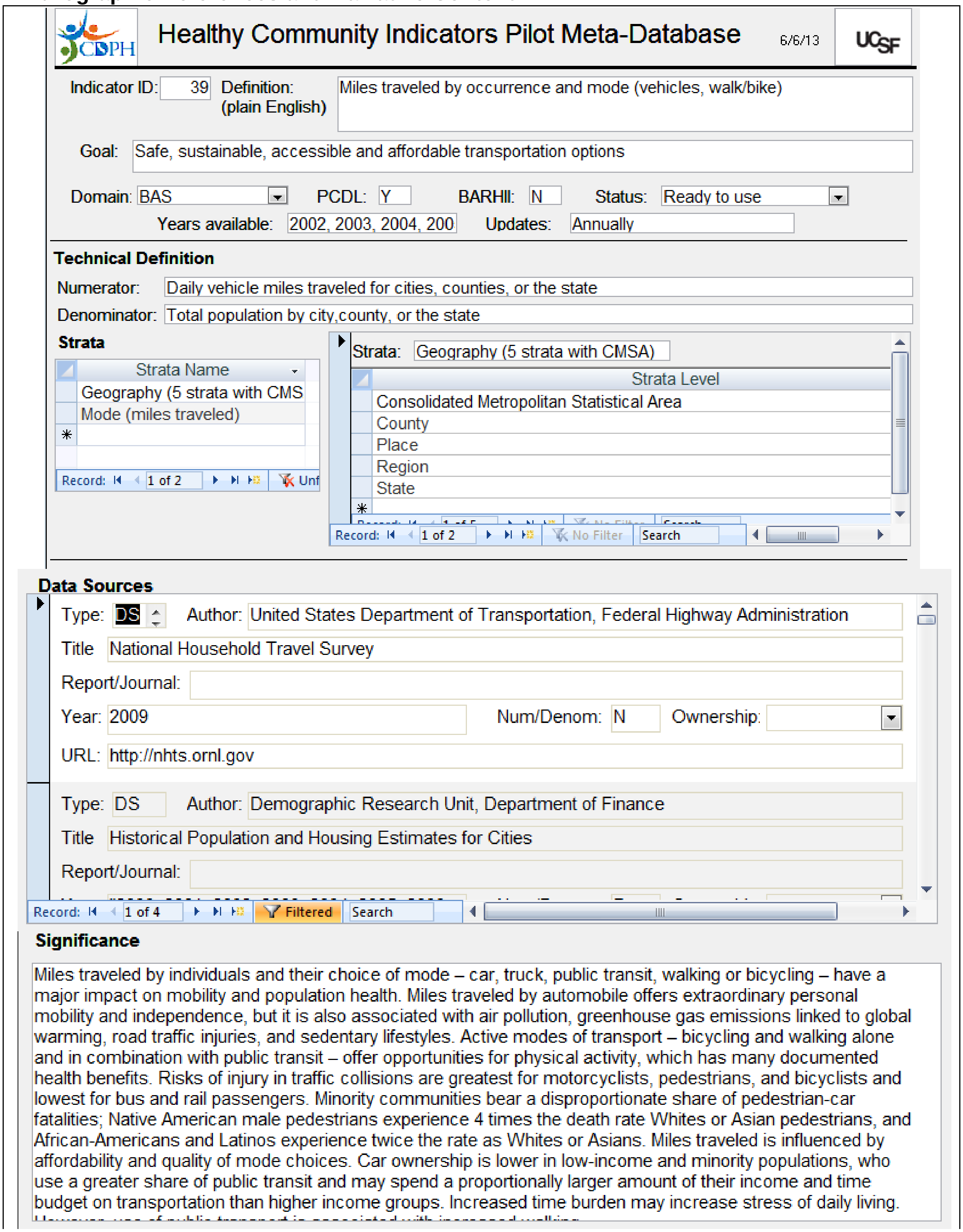
Table 40. Meta-Database Queries, Forms, and Reports

Object	Query Name	Description	SQL Syntax
Queries	PCDL_Source_Type	Totals query to generate list of indicators by count of references for data sources, evidence, indicator report to determine whether some categories have no references	TRANSFORM Count(Biblio_updated.reftype) AS CountOfreftype SELECT Indicators.ind_id, Indicators.ind_definition FROM Indicators LEFT JOIN Biblio_updated ON Indicators.ind_id = Biblio_updated.ind_id WHERE (((Indicators.PCDL)="Y")) GROUP BY Indicators.ind_id, Indicators.ind_definition PIVOT Biblio_updated.reftype;
	biblio_delete_rec	Deletes references from 'Biblio_updated table' prior to a user-supplied date	DELETE Biblio_updated.*, Biblio_updated.endnote_import_date FROM Biblio_updated WHERE (((Biblio_updated.endnote_import_date)<[biblio_updated]));
	Biblio_import_tabdl_delete_blank_records	Deletes blank rows references from 'Biblio_updated table' prior to a user-supplied date	DELETE Biblio_import_tabdl.*, Biblio_import_tabdl.[rec-number] FROM Biblio_import_tabdl WHERE (((Biblio_import_tabdl.[rec-number]) Is Null));
	Evidence_references	Select query to list references for evidence	SELECT Biblio_updated.ind_id, Biblio_updated.reftype, IIf([city] Is Null," ",[city]) AS city2, IIf([institute] Is Null," ",[institute]) AS institute2, IIf([pages] Is Null,"",[pages]) AS pages2, +[author]+". "+[title]+". "+[publisher]+"; "+[year]+". "+[URL] AS Reference, IIf([journal_name] Is Not Null,[journal_name]+" "+[volume]+": "+[pages2],[city2]+": "+[institute2]) AS publisher FROM Biblio_updated WHERE (((Biblio_updated.reftype)="EV"));
	follow-up_status	Select query of indicators by status of follow-up activities	SELECT Followup.ind_id, Indicators.ind_definition, Followup.issue_id, Followup.person_id, Followup.date_assigned, Followup.issue_description, Followup.proposed_follow_up, Followup.date_due, Followup.date_done, IIf([date_done] Is Not Null,"Done","Pending or Late") AS Status FROM Indicators LEFT JOIN Followup ON Indicators.ind_id=Followup.ind_id WHERE (((Followup.issue_id)>0)) ORDER BY Followup.ind_id, Followup.issue_id, IIf([date_done] Is Not Null,"Done","Pending or Late");

Object	Query Name	Description	SQL Syntax
	ind_id_definition	Select query that lists indicators	SELECT Indicators.ind_id, Indicators.ind_definition FROM Indicators WHERE (((Indicators.PCDL)="y")) ORDER BY Indicators.ind_id;
	Ind_status	Select query of indicators by stage of completion	SELECT Indicators.ind_definition, Indicators.ind_status FROM Indicators WHERE (((Indicators.ind_status) Is Not Null));
	Indicator_Projects	Select query to list references for indicator projects	SELECT Biblio_updated.ind_id, [author]+". "+[title]+". "+[institute]+"; "+[Year]+". "+[URL] AS Reference, Biblio_updated.reftype FROM Biblio_updated WHERE (((Biblio_updated.reftype)="IR"));
	Indicators_strata_list	Select query to link indicators to their strata andstrata levels	SELECT Indicators.ind_id, Indicators.ind_definition, indicators_strata.strata_id, indicators_strata.strata_name, Indicators_strata_level.strata_level_id, Indicators_strata_level.strata_level_name FROM Indicators LEFT JOIN (indicators_strata LEFT JOIN Indicators_strata_level ON indicators_strata.strata_id = Indicators_strata_level.strata_id) ON Indicators.ind_id = indicators_strata.ind_id;
	PCDL_no_data_source	Select query to identify indicators that have no references for data sources	SELECT PCDL_Source_Type.ind_id, PCDL_Source_Type.ind_definition FROM PCDL_Source_Type WHERE (((PCDL_Source_Type.DS) Is Null));
	PCDL_no_evidence	Select query to identify indicators that have no references for evidence	SELECT PCDL_Source_Type.ind_id, PCDL_Source_Type.ind_definition FROM PCDL_Source_Type WHERE (((PCDL_Source_Type.EV) Is Null));
	PCDL_no_ind_report	Select query to identify indicators that have no references for indicator projects	SELECT PCDL_Source_Type.ind_id, PCDL_Source_Type.ind_definition FROM PCDL_Source_Type WHERE (((PCDL_Source_Type.IR) Is Null));

Object	Query Name	Description	SQL Syntax
	Update_bibliography	Update query that updates Biblio_updates' with new references from 'Biblio_import_tabdl'	UPDATE Biblio_import_tabdl LEFT JOIN Biblio_updated ON Biblio_import_tabdl.[rec-number] = Biblio_updated.[rec-number] SET Biblio_updated.[rec-number] = [Biblio_import_tabdl].[rec-number], Biblio_updated.type = [Biblio_import_tabdl].[type], Biblio_updated.author = [Biblio_import_tabdl].[author], Biblio_updated.[Year] = [Biblio_import_tabdl].[year], Biblio_updated.title = [Biblio_import_tabdl].[title], Biblio_updated.journal_name = [Biblio_import_tabdl].[journal_name], Biblio_updated.city = [Biblio_import_tabdl].[city], Biblio_updated.institute = [Biblio_import_tabdl].[institute], Biblio_updated.volume = [Biblio_import_tabdl].[volume], Biblio_updated.[number] = [Biblio_import_tabdl].[number], Biblio_updated.pages = [Biblio_import_tabdl].[pages], Biblio_updated.reftype = [Biblio_import_tabdl].[reftype], Biblio_updated.ind_id = [Biblio_import_tabdl].[ind_id], Biblio_updated.[N/D] = [Biblio_import_tabdl].[N/D], Biblio_updated.notes = [Biblio_import_tabdl].[notes], Biblio_updated.URL = [Biblio_import_tabdl].[URL], Biblio_updated.endnote_import_date = [Biblio_import_tabdl].[endnote_import_date];
Forms	hci_summary form	Main form that displays indicator data	
	Evidence_references subform	Displays data from Evidence_references query	
	Indicator_Projects subform	Displays data from Indicator Project query	
	Indicator_strata_level subform	Displays data from 'Indicators_strata_level' table	
	Indicators_strata subform	Displays data from 'Indicators_strata' table	
	Indicators_strata subform1	Displays data from 'Indicators_strata' table	
	hci_summary_follow-up form	Main form that displays followup information on indicators	
	Followup subform	Displays data from 'Followup' table	
Reports	Followup	Displays data from follow-up_status query	

Figure 11. The 'hci_summary form' for the Key Entry of Meta Data, Display of Bibliographic References and Narrative Content



CDPH Healthy Community Indicators Pilot Meta-Database 6/6/13 UCSF

Indicator ID: 39 Definition: (plain English) Miles traveled by occurrence and mode (vehicles, walk/bike)

Goal: Safe, sustainable, accessible and affordable transportation options

Domain: BAS PCDL: Y BARHII: N Status: Ready to use

Years available: 2002, 2003, 2004, 200 Updates: Annually

Technical Definition

Numerator: Daily vehicle miles traveled for cities, counties, or the state

Denominator: Total population by city,county, or the state

Strata

Strata Name
Geography (5 strata with CMS
Mode (miles traveled)

Record: 1 of 2 Unf

Strata: Geography (5 strata with CMSA)

Strata Level
Consolidated Metropolitan Statistical Area
County
Place
Region
State

Record: 1 of 2 No Filter Search

Data Sources

Type: DS Author: United States Department of Transportation, Federal Highway Administration

Title National Household Travel Survey

Report/Journal:

Year: 2009 Num/Denom: N Ownership:

URL: http://nhts.ornl.gov

Type: DS Author: Demographic Research Unit, Department of Finance

Title Historical Population and Housing Estimates for Cities

Report/Journal:

Record: 1 of 4 Filtered Search

Significance

Miles traveled by individuals and their choice of mode – car, truck, public transit, walking or bicycling – have a major impact on mobility and population health. Miles traveled by automobile offers extraordinary personal mobility and independence, but it is also associated with air pollution, greenhouse gas emissions linked to global warming, road traffic injuries, and sedentary lifestyles. Active modes of transport – bicycling and walking alone and in combination with public transit – offer opportunities for physical activity, which has many documented health benefits. Risks of injury in traffic collisions are greatest for motorcyclists, pedestrians, and bicyclists and lowest for bus and rail passengers. Minority communities bear a disproportionate share of pedestrian-car fatalities; Native American male pedestrians experience 4 times the death rate Whites or Asian pedestrians, and African-Americans and Latinos experience twice the rate as Whites or Asians. Miles traveled is influenced by affordability and quality of mode choices. Car ownership is lower in low-income and minority populations, who use a greater share of public transit and may spend a proportionally larger amount of their income and time budget on transportation than higher income groups. Increased time burden may increase stress of daily living.

Summary of Evidence

Emissions from motor vehicles powered by fossil fuels are proportional to vehicle miles traveled and account for approximately one-third of California's annual emissions of air pollutants such as fine particulates and precursors of ozone. These air pollutants have established links to increased mortality, hospital admissions, and other adverse health effects. Numerous epidemiological studies have documented that physical activity, including that related to walking and bicycling, decrease risks of cardiovascular disease and stroke, colon and breast cancer, and dementia and depression. Miles traveled is also associated with road traffic injuries, although injury rates of bicyclists and pedestrians tend to level off as their miles traveled and mode share increases.

References

Reference
Woodcock J,Edwards P,Tonne C,Armstrong BG,Ashiru O,Banister D,Beevers S,Chalabi Z,Chowdhury Z,Cohen A,Franco OH,Haines A,Hickman R,Lindsay G,Mittal I,. Public health benefits of strategies to reduce
California Air Resources Board. Estimated Annual Average Emissions. "Sacramento, CA": California Air Resources Board; 2008. http://www.arb.ca.gov/app/emsinv/emssumcat.php
de Hartog JJ, Boogaard H, Nijland H, Hoek G. Do the health benefits of cycling outweigh the risks?. Environ Health Persp 118: 1109-16; 2010. http://ehp03.niehs.nih.gov/article/info:doi%2F10.1289%2Fehp.0901747

Record: 1 of 9 No Filter Search

Data Description and Analysis

Detailed Definition: Annual miles traveled by occurrence 1) per capita, 2) per square mile.

Numerator: Data on the daily vehicle miles traveled for cities, counties, and the state was obtained from CPR. Vehicles included cars, trucks, buses, and motorcycles. Daily miles traveled were multiplied by 365 to estimate annual miles. Annual miles traveled by bicyclists and pedestrians for CMSAs (county clusters for major metropolitan areas) and California was estimated between 2002 and 2010 from the NHTS by applying the annual rate of change between 2001 and 2009. Denominator 1: Annual population counts from 2002 to 2010 for California's 481 incorporated cities and 58 counties were abstracted from DOF reports. July-centered population counts were obtained by averaging January counts for consecutive years, except for 2010, which used an April

Indicator Projects

Reference
San Francisco Department of Public Health. Healthy Development Measuring Tool (HDMT). San Francisco Department of Public Health; 2011. http://www.thehdmt.org/
Stair P, Wooten H, Raimi M. "How to Create and Implement Healthy General Plans: A Toolkit for Building Healthy, Vibrant Communities Through Land Use Policy Change". Public Health Law & Policy; 2012. http://www.phlpnet.org/healthy-planning/create_implement_gp

Record: 1 No Filter Search

SAS file location

T:\HCI\Data\HCI_MTpC_39_9-6-13.sas

Limitations:

Data were not available to disaggregate different motorized modes (car, bus, truck, motorcycle), stratify by race/ethnicity, or to calculate standard errors. Data were not statistically stable for pedestrians and bicyclist at city geographies. City-specific estimates do not include miles traveled in unincorporated areas of counties. The annual miles traveled data is by occurrence and the population data is by residence; because of this discrepancy, it is not recommended to construct this indicator for cities since it could result in misleading outcomes (for example, cities with small populations but high through mileage could overestimate annual miles per capita).

Comments: 192 : the more vehicle miles traveled (VMT) in a community, the worse the air pollution

Figure 12. Form ‘hci_summary_followup’ for management of follow-up activities of entering metadata for each indicator

Healthy Community Indicators Pilot Meta-Database

DRAFT 8/27/12

UCSF

Indicator ID: 42 Definition: (plain English) Percent of residents mode of transportation to work

Goal: Safe, sustainable, accessible and affordable transportation options

Domain: BAS PCDL: Y BARHII: Status: Ready to use

Followup subform

Issue ID	Staff	Date As	Description	Proposed F/U	Date due	Date done	Comm
18	PA	9/4/2012	Bibliography	Missing evidence	9/30/2012	9/6/2012	
(New)							

Figure 13. ‘Followup’ Report Describing Follow-up Activities by Indicator

Follow-up Items for Healthy Community Indicators Project

UCSF

ID	Issue	Staff	Indicator	Date In	Issue	Proposed Follow-up	Due	Done	Status
42	18	PA	Percent of residents mode of transportation to	09/04/12	Bibliography	Missing evidence	09/30/12	09/06/12	Done
51	3	NM	Percent of population residing within ½ mile of	09/04/12	Data acquisition	Contact Caroline Rodier at ULTRANS	09/07/12	09/07/12	Done
58	19	PA	Percent of household income spent on travel	09/04/12	Bibliography	Missing evidence	09/30/12		Pending or Late
72	10	PA	Percent of households within 1/2 mile full-	09/04/12	Bibliography	Missing data source	09/14/12		Pending or Late
72	22	PA	Percent of households within 1/2 mile full-	09/04/12	Bibliography	Missing evidence	09/30/12	09/05/12	Done
85	23	PA	Percent of adults who consume five or more	09/04/12	Bibliography	Missing evidence	09/30/12	09/05/12	Done
187	28	PA	Percent of adults aged 18 - 64 without health	09/04/12	Bibliography	Missing evidence	09/30/12	09/11/12	Done
297	20	PA	Number and rate of fatal and nonfatal occupational	09/04/12	Bibliography	Missing evidence	09/30/12		Pending or Late
341	49	PA	Mean score of Academic Performance Index (API)	09/04/12	Bibliography	Missing evidence	09/30/12		Pending or Late

To facilitate quality assurance, quality improvement and comparisons efforts, HCDIP maintains a variety of standards in the reporting and visualization of its data. The Project standards, discussed below, are applied to every level and process in the generation of HCDIP products.

Geographical Lookup Files

Each indicator obtains its corresponding geographic labels from tables called lookup files. These lookup tables contain the most recent and complete list of geographical names and identifiers within each geographic level of California. HCDIP has a total of four geographical lookup files. The following are samples of lookup files:

- A list of the 58 counties in California and their FIPS codes contained in the lookup file *CA_County_FIPS2010.xls*.

	A	B	C
1	County_FIPS	County_Name	CTY
2	06001	Alameda	001
3	06003	Alpine	002
4	06005	Amador	003

- A list of the 14 regions in California is located in the lookup file *CA_regions2010.xls*.

	A	B
1	Region_Code	Region_name
2	01	Bay Area
3	02	Butte
4	03	Central/Southeast Sierra

- A list of all the census tracts, places, counties, regions, metropolitan statistical areas and state are provided in the file *Look_up_table_CT_PL_CO_RE_CA.xls* (for data sourced from the U.S. Census) The U.S. Census Bureau American Community Survey uses standard census tracts (defined as census tract summary level 140; n=8,057).

	A	B	C	D	E	F
1	Geotype	Geotypevalue	Geoname	County_FIPS	County_Name	Region_Code
2	CA	06	California			
3	RE	01	Bay Area			01
4	RE	02	Butte			02
5	RE	03	Central/Southeast Sierra			03

- A list of all the census tracts, places, counties, regions, and state are provided in the file *CHAS_LookUpTable_CT_PL_CO_RE_CA.xlsx* (for data sources from the U.S. HUD Consolidated Planning Comprehensive Housing Affordability Strategy, CHAS). CHAS data use split census tracts (defined as summary level 080; n=12,883), it is important to use CHAS lookup files with the CHAS data.

CHAS_LookUpTable_CT_PL_CO_RE_CA.xlsx - Microsoft Excel

	A	B	C	D	E	F	G	H
1	geoid	geotype	geotypevalue	geoname	county_fips	county_name	region_code	region_name
2	04000US06	CA	06	California				
3	05000US06001	CO	06001	Alameda	06001	Alameda	01	Bay Area
4	05000US06003	CO	06003	Alpine	06003	Alpine	03	Central/Southeast Sierra
5	05000US06005	CO	06005	Amador	06005	Amador	03	Central/Southeast Sierra
6	05000US06007	CO	06007	Butte	06007	Butte	02	Butte
7	05000US06009	CO	06009	Calaveras	06009	Calaveras	03	Central/Southeast Sierra

An indicator having missing values for a local area suggests that the data were not released or available for that area at the time of the data analysis. Values of the lookup tables are automatically read into the SAS program and joined with the indicator in the resulting SAS dataset (discussed below).

Race/ethnicities

The HCDIP reports on race/ethnicity for the indicators for which such information is available. The standard for the race/ethnicity strata comes from the U.S. Census. The HCDIP reports on nine groups that derive from the U.S. Census classifications (Table 41).

Table 41. Race and Ethnic Groups from the U.S. Census and Corresponding Race/Ethnicity Categories used in the HCDIP

U.S. Census Race and Ethnic Groups	HCDIP race/ethnicities
Total (001)	Total
Hispanic or Latino (of any race) (400)	Latino
White alone, not Hispanic or Latino (451)	White
Black or African American alone, not Hispanic or Latino (453)	African American
American Indian and Alaska Native alone, not Hispanic or Latino (455)	AIAN
Asian alone, not Hispanic or Latino (457)	Asian
Native Hawaiian and Other Pacific Islander, not Hispanic or Latino (459)	NHOPI
Some other race alone, not Hispanic or Latino (461)	Other
Two or more races, not Hispanic or Latino (463)	Multiple

Races are defined to be mutually exclusive. For example, “White” is defined as white alone and no other race. “African American” indicates African American or black only and so forth. “Other” is defined as a single race not captured in the other race categories. “Multiple races” include respondents who identify as two or more of the above races. The U.S. Census collects data on ethnicity, which is categorized as Hispanic and non-Hispanic and is distinct from race. This means respondents identifying as either Hispanic or Latina/o are defined as having the ethnicity, “Hispanic.” Ethnic subgroups can be of any race (white, African American, Asian, etc.). More information about ethnicity can be found at: http://factfinder2.census.gov/help/en/american_factfinder_help.htm#advanced_search/search_using_population_groups.htm#AS_Basic_Groups.

Other data sources different to the U.S. Census might not report in all of the same race/ethnicity categories. For instance, CHAS data use seven of the above race/ethnicity designations (“Other” does not apply) and Total. Values of the race/ethnicity lookup table are automatically read by the SAS program and joined to the indicator data. Remember to match the race/ethnicity designations to the pertinent data source (e.g., CHAS data with CHAS race/ethnicity groups) when performing the data analysis.

A list of all the race/ethnicity classifications used in HCDIP is contained in the lookup file *race_ethnicity.xls*:

race_eth_	race_eth_name
1	AIAN
2	Asian
3	AfricanAm
4	Latino
5	NHOPI
6	White
7	Multiple
8	Other
9	Total

SAS Program Structure

To produce each indicator, the various raw sources of data underwent standardized processing using the statistical software, Statistical Analysis System (SAS) version 9.3. The program to create each indicator is contained in a .sas file. Each file follows a general form designed to extract the raw data, clean the data, create the analytic dataset, process the data to meet HCDIP standards and output the data into a .xls or .xlsx Excel file. In the .sas file, the program will have a general form comprised of nine blocks. Block 1 reads in race/ethnicity, city, county, place, region and state lookup tables (and population and household weights) and indicator definition files. Block 2 extracts the raw data, keeping the estimate, margin of error (estimate), percent, margin of error (percent) and creates the indicator and its standard error for each geographical level, race/ethnicity subgroup and report year. This block is repeated or looped for each survey period. Block 3 calculates the indicator and its standard error at the regional level based on a county-weighted average and county-weighted standard error. This block is looped for each survey period. Block 4 computes deciles for census tracts and places. This block is looped for each survey period. Block 5 calculates the ratio of the local estimate to state estimate (i.e., the relative risk). Block 6 concatenates the various datasets and computes the 95% confidence interval, relative standard error and relative risk for an indicator. Block 7 attaches geographical labels to the master dataset. Block 8 adds indicator labels (ID number, ID definition) and date-time stamps, assesses the data stability and suppresses the unstable data, if present. Block 9 exports the final SAS dataset as a single, Excel file. This general form will vary somewhat, depending on the indicator. Some indicators will have published percentages and the corresponding margin of errors while other indicators may not. Race/ethnicity data may not be available for certain indicators. Others may only have income-level data. Consequently, the SAS program structure for each indicator will show variations of the general form.

<u>Skeleton of general form and its nine blocks:</u>

Block 1 reads in City, County, Region Look-up tables (and population and household weights) and indicator definition files.

Block 2. CT, Place, County, State files (all years + Race/ethnicity)
Loop1
  Read file
  Tag report year
   Save geo_id, race/ethnicity, N, D, SE
End loop1
List of files (f1, f2, f3 …f14)

Block 3. Regional calculation
Loop2 (for years)
   Read sas file & year
      By race/ethnicity?
     County-weighted average
     County weighted SE
 End loop2
List of years (2006-2010 2008-2010 2005-2007 2000)

Block 4. Deciles
Loop3
Read file (2006-2010)
  Compute Deciles
  Add column to file
End loop3
List of files (CT places)

Block 5. CA value for RR calculation
Symput (CA2000, CA2005-7, CA2008-2010, 2006-2010

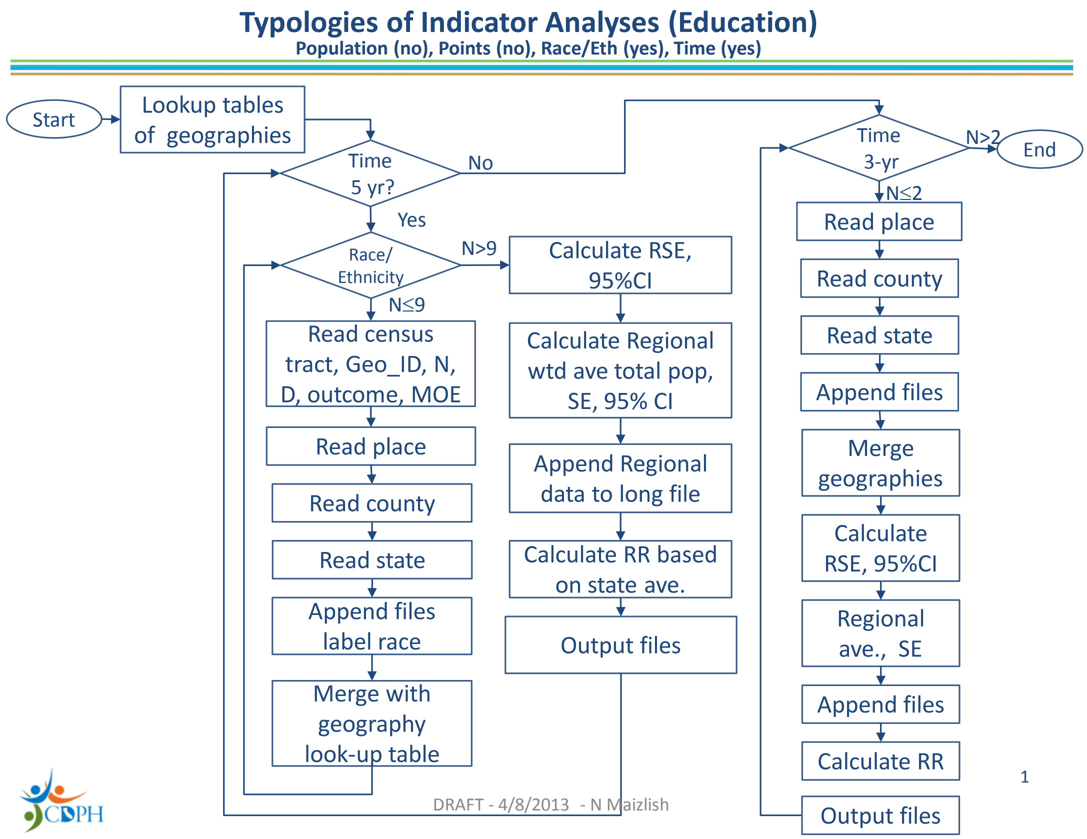
Block 6. Computed fields
Concatenate files (set f1 f2 f3 …)
Compute RSE, 95%CI, RSE, CA_RR

Block 7. Merge with geography look-up tables and title

Block 8. Label with other info (ID_NO, race_labels). Assess data stability. Suppress data if needed.

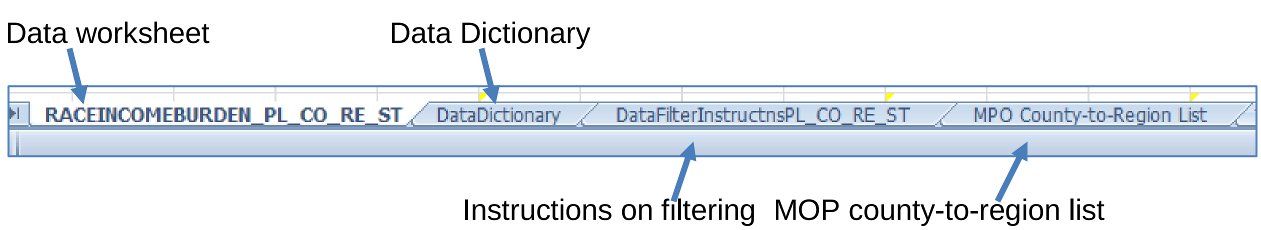
Block 9. Export final, SAS dataset as a single, Excel file.

<u>Flow chart displaying block groups illustrated by the education indicator.</u>



Standardized Output File Format

The production of each indicator in SAS is outputted as a standardized Excel file (.xlsx or .xls file) having four worksheets. The first sheet is the data worksheet containing the indicator, stratification variables and other statistical measures. The second sheet is a data dictionary listing the fields contained in the data worksheet and its attributes (see section below on this topic). The third sheet is instructions on filtering the data. The fourth and final sheet is the MPO county-to-region (derived) list. All indicator Excel files follow this form; however, some indicators may contain more fields or unique conditions. In these cases, the Excel file will have additional sheets.



Data Dictionary

As mentioned in an earlier section, the data dictionary is a worksheet in the indicator Excel file. It has a core format comprised of five columns: *Name, Definition, Type, Width* and *Coding* (Table 42). The *Name* column lists the name of each field contained in the data worksheet. The *Definition* column provides the meaning of a field. The *Type* column indicates whether the field type is numeric (N), text (T) or date (D). The *Width* column states the length of a field. The *Coding* column lists the specific values a field will take and other relevant notices. The data dictionary will have a minimum of 22 fields. Some indicators will have additional fields. In these cases, the data dictionary will be an elaboration of the general format.

Table 42.   Format of the Data Dictionary

Name	Definition	Type	Width	Coding
ind_id	HCI ID/key of indicator	N	3	
ind_definition	Definition of indicator in plain language	T	255	Free text
reportyear	Year(s) that the indicator was reported	T	9	CCYY-CCYY
race_eth_code	numeric code for a race/ethnicity group	T	1	1=American Indian or Alaska Native; 2=Asian; 3=Black or African American; 4=Hispanic or Latino; 5=Native Hawaiian or Other Pacific Islander (NHOPI); 6=White;7= Multiple; 8=Other;9=Total
race_eth_name	Name of race/ethnic group	T	25	AIAN; Asian; AfricanAm; Latino; NHOPI; White;Multiple; Other; Total
geotype	Type of geographic unit	T	2	CT,Census Tract; N,Neighborhood; PL,Place; ZC,Zipcode; CO, County; RE, region; CA, State
geotypevalue	Value of geographic unit	T	See specifics under coding	FIPS ID; place code (5-characters in length), CT code (11-characters in length starting with 06), zip code, region id
geoname	Name of geographic unit	T	25	place, county, region name se code list
county_name	Name of county that geotype is in	T	25	county name
county_fips	FIPS code of county that geotype is in	N	5	See MPO county-to-region list
region_name	Name of region	T		See MPO county-to-region list
regioncode	Numeric code of region	T		See MPO county-to-region list
numerator	Numeric value of numerator of indicator	N	double prec.	
denominator	Numeric value of denominator	N	double prec.	
percent	Numerator divided by denominator for basis	N	double prec.	
LL_95CI	Lower limit of 95% confidence interval	N	double prec.	
UL_95CI	Upper limit of 95% confidence interval	N	double prec.	
SE	Standard error of percent	N	double prec.	
RSE	Relative standard error expressed as a percent	N	double prec.	=(SE/percent) * 100
CA_decile	Statewide decile ranking	N	integer	
CA_RR	Ratio of indicator in geographic area to state average	N	double prec.	
version	Date/time stamp of version of data	D	1	mm/DD/CCYY hh:mm:ss

Data Presentation Standards

The HCDIP produced a series of narrative files with examples of data use. The narrative files followed standards for the presentations of titles, maps, graphs, tables and references.

<u>Title Standards</u>

All titles in maps, bar charts, line series graphs and tables follow a standard naming convention. A title addresses the question of what/whom, where and when. A title appears bolded and designates the exhibit type and number, indicator title, the stratification level, geographic unit, time period and if applicable, footnote designations. For example, "**Map 3. Percent of Household Overcrowding, Households at All Income Levels, Places, Los Angeles County, CA, 2006-2010***."

<u>Mapping Standards</u>

HCDIP maps were produced using software, ESRI ArcGIS Desktop version 10.2. Shapefiles for each of the geographical levels in the project were obtained from the U.S. Census (see Toolbox section, Downloading Census TIGER Shapefiles and Reprojecting in California Teale Albers Coordinate System chapter), and reprojected to the NAD 1983 California Teale Albers coordinate system.

Each indicator map produced adhered to specific mapping standards such as title, cut points relative to the state average, Arial font and context-specific font sizing, color schemes, legend attributes, inset of the State and local area, scale and North bar. The GIS and mapping specifications appear in **Table 43**.

Table 43. GIS and Mapping Specifications

<u>Projection</u> Standard projection for HCDIP maps at all geographic levels is NAD 1983 California (Teale) Albers (Meters) projection. This is a projected coordinate system (as opposed to a geographic coordinate system).
<u>Census Tracts</u> For county maps, label all cities from place shapefile (LSAD=city). Cities are labeled, but not CDPs and towns. Times New Roman 11 Black. Add census tract boundaries from census tract shapefile using no fill color and 0.4 widths, gray 30% color for the border. Add city boundaries from place shapefile using no fill color and a 2 width, white color border. Add mask to city borders for places lying outside the county (Definition Query).
<u>City Labels</u> For regional maps, label all cities over 100,000. In addition, label largest cities in counties without any larger cities (cities over 100,000). Times New Roman 12 Black For county maps, label all cities from place shapefile (LSAD = city). Cities are labeled, but not CDPs and towns. Times New Roman 11 Black
<u>County Labels</u> Arial Black Bold Italic small caps 14, for county and region. Arial 10 Bold small caps for state, using name field. State map uses manual labelling for more problematic county labels--for these counties NAME field is blank.
<u>Color Scheme</u> Lowest (best) range = HCDIP Green (RGB = 186, 219, 115) 2nd range = HCDIP Yellow (RGB = 255, 212, 69) 3rd range = HCDIP Orange (RGB = 255, 145, 56) Highest (worst) range = HCDIP Red (RGB = 242, 32, 0)
<u>Water (e.g. Bay or Pacific Ocean)</u> Change background of map to standard color for water (RGB = 165, 191, 221)
<u>Data Source</u> lower left, 'Source: xyz' , Arial 8, include data source and year include 'Analysis by CDPH and UCSF' in lower right, Arial 8
<u>North Arrow</u> ESRI North 21, Size = 52
<u>Unreliable Data</u> Add a duplicate of data layer, map RSE, separating into 2 ranges (RSE<30 and RSE>30) Use simple 10% cross hatching (black), decrease separation of hatching to 2.5 for areas such as census tracts or places in regions To avoid big gap in legend, put a "dummy" layer in the legend instead, labeled "Unreliable data (RSE ≥ 30)".

Data not reported or applicable
background layer showing gray for areas with no data, extra county layer can be used, gray (20%) with no outline

Legend
background color = RGB = (255, 234, 190)
show 4 ranges w/ ratio to state/region
green patch 24x12 x.x - y.y (lower than the state average of y.y)
yellow patch 24x12 y.y - z.z (1 to 11/2 times the state average) (use 1/2 symbol)
orange patch 24x12 z.z - a.a (1 1/2 to 2 times the state average) (use 1/2 symbol)
red patch 24x12 a.a - b.b (2+ times the state average)
cross hatching 24x12 Unreliable data
gray patch 24x12 Data not reported or applicable

Inset of CA
To be applied to maps showing county or subcounty areas.
Inset size: 1.339 in. (width) x 1.6643 in. (height).
Inset position: 0.6558 in. (X), 1.2552 in. (Y).
Make slight adjustments to size and position if necessary.
Frame border color = black
Frame border width = 2.0.
Frame border background color = white
For county-level and regional-level maps: use the statewide county shapefile. Outline color is invisible. Fill color is 30% gray. The size and position the State should span the entire data frame, but does not touch the edges of the frame border.
For county-level maps: use the individual county analyzed (e.g, shapefile of Napa Co.). Outline color is invisible. Fill color is Mars Red.
For regional-level maps: use the individual region being mapped (e.g, shapefile of the Bay Area). Outline color is invisible. Fill color is Mars Red.
For local maps (city/town level): use shapefile for the individual county that the city/town is located in. For example, if the map is of Napa city, insert the shapefile for Napa County. Outline color is invisible. Fill color is 30% gray. Label the county using Arial Black, 10, Bold, Italic, all caps. Then use shapefile for the individual city/town being mapped (e.g., shapefile of Napa City). Outline color is invisible. Fill color is Mars Red. Change shape of data frame as necessary to accommodate shapes of different counties.

Scale
lower left, width = 3.0141 height = .3036, Arial 12, 1 midpoint, both ends and midpoint labeled in miles

Title
Arial Black 14, top centered, outside map frame
Use standard naming format
Map #: Indicator (what is being mapped), by geographic area (census tract), geographic area being mapped (e.g. Bay Area Region), Year(s) (e.g., 2010)

Source
Cite the source of the data and time period(s) at the bottom of the map.

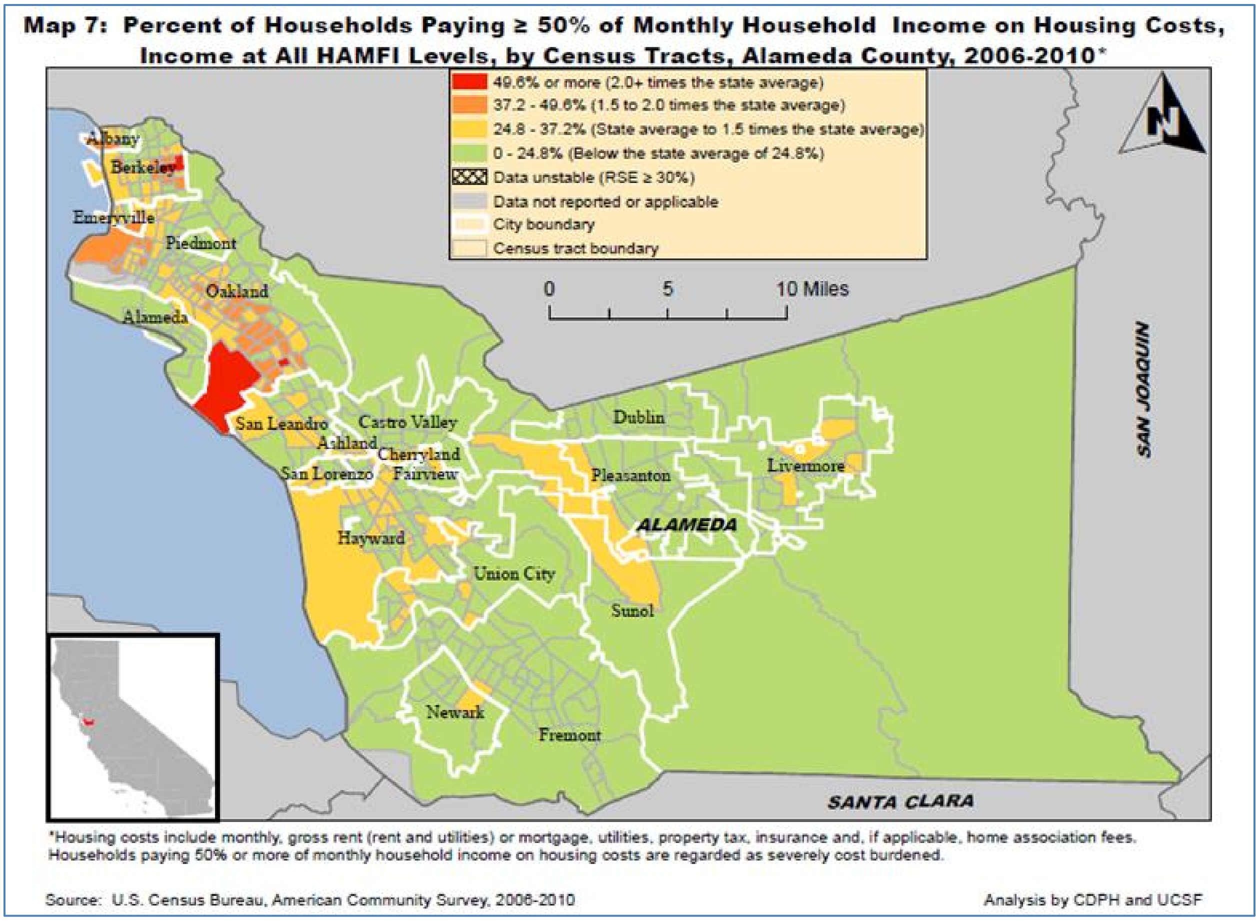


Figure 14. An Example of an Indicator Map Meeting HCDIP Standards

<u>Graphing Standards</u>

Whenever possible, an indicator is visualized through bar charts and lines series graphs. The standard bar chart displays a bolded, descriptive title and how the indicator is distributed among the race/ethnicity subgroups and the total. Values are arranged in descending order (from high to low). The “Total” group gets lighter shading. The Y-axis always starts at 0 to facilitate comparisons to other charts. Major tick marks are shown only for the vertical axis and in light gray. Cite the source of the data and time period(s) at the bottom of the chart.

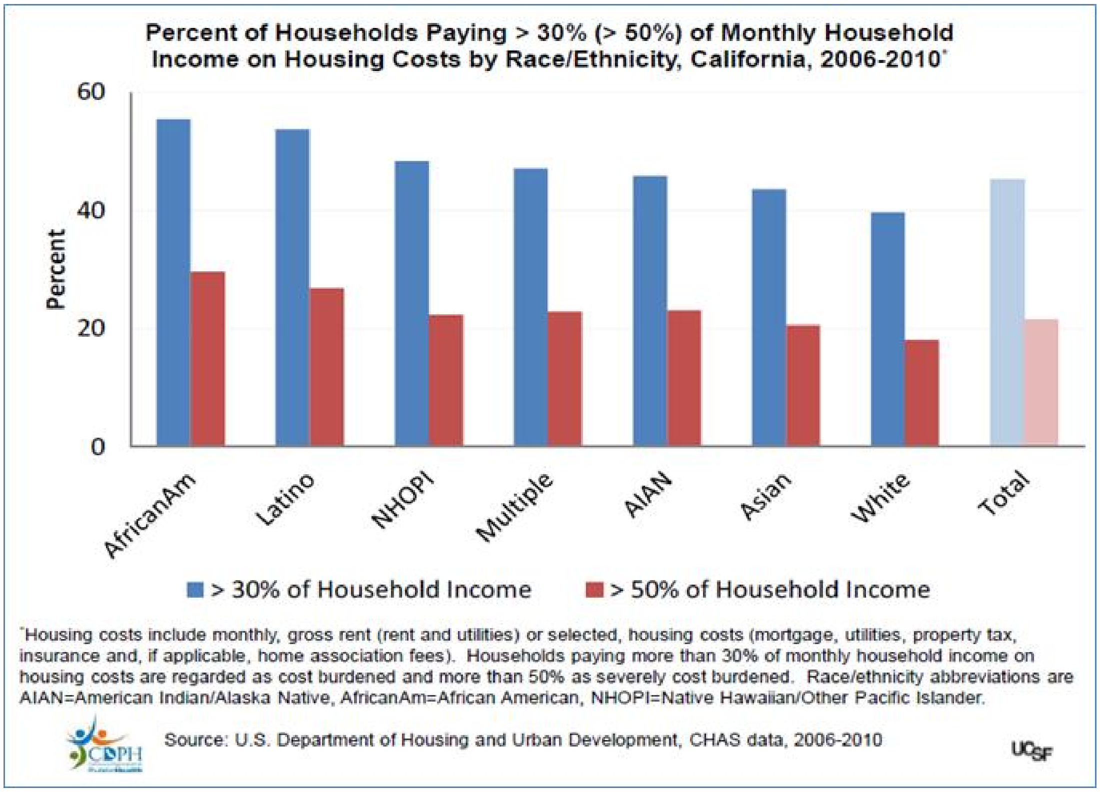


Figure 15. An Example of a Bar Graph Meeting HCDIP Standards

The standard lines series graph displays the title and indicator across time (also called time series graphs). An indicator may have single years, three-year estimates and five-year estimates available. In this case, single years should be graphed with other single years. Three-year estimates should be graphed with other three-year estimates and so on. If a mix of single and multiple years is available, non-overlapping years should be used in the graph. Cite the source of the data and time period(s) at the bottom of the graph. HCDIP charts and graphs were produced using MS Power Point 2010. A finalized line series graph is shown below.

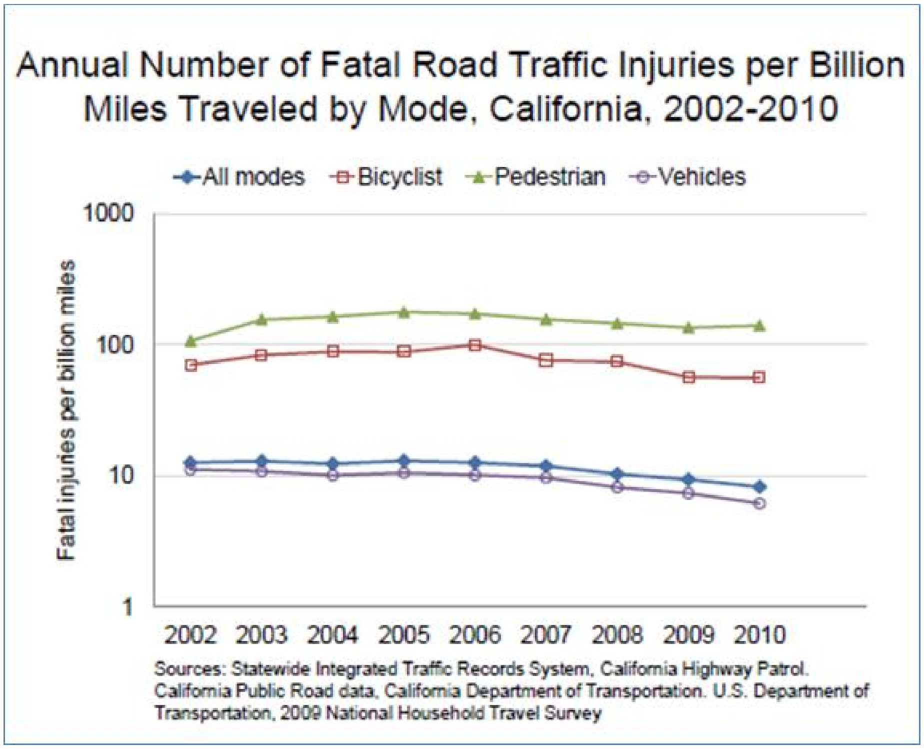


Figure 16. An Example of a Line Series Graph Meeting HCDIP Standards

<u>Table Standards</u>

HCDIP tables follow the American Medical Association Style Guide (Last reviewed 2009): http://www.google.com/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=10&cad=rja&uact=8&ved=0CFoQFjAJ&url=http%3A%2F%2Fwww.docstyles.com%2Flibrary%2Famastat.pdf&ei=U8sAVPy4D6nziwKTyYDAAw&usg=AFQjCNFYJNFS2gpgmm8rGdtSlJfpRrQJWQ. HCDIP tables were produced using MS Word 2010.

A table should have only a minimal amount of borders. The title is bolded and has the type of exhibit and its number. The title should also address the what, who, where and when. In the table, the first column containing names, of places for instance, should be left-justified. Put separator lines at the top, bottom and highest grouping level. Center column headers. Right-justify numbers. For large numbers, use a 1000 separator (","). For small numbers between 0 and 1, make sure the decimals line up in a column. At the bottom of the table, list the source for the data and the time period. When available, the table should show an example of the data for the numerator, denominator, and the indicator.

Table 1. Percent of the Population by Race/Ethnicity within ¼ Mile of Alcohol Outlets by Type of Establishment Sales, California, 2014.

Race/Ethnicity	Total Population (n)	Population within ¼ Mile - Off Sale Alcohol Outlets (n)	Percent Population within ¼ Mile - Off Sale Alcohol Outlets (%)	Population within ¼ Mile - On Sale Alcohol Outlets (n)	Percent Population within ¼ Mile - On Sale Alcohol Outlets (%)	Population within ¼ Mile - On and Off Sale Alcohol Outlets (n)	Percent Population within ¼ Mile - On and Off Sale Alcohol Outlets (%)
African American	2,163,804	1,102,027	51	720,867	33	1,197,915	55
American Indian Alaska Native	162,250	56,423	35	45,283	28	64,394	40
Asian	4,775,070	2,121,318	44	1,858,012	39	2,410,073	50
Latino	14,013,719	7,503,996	54	5,304,968	38	8,100,882	58
Multiple	968,696	383,791	40	325,960	34	439,097	45
Native Hawaiian/Other Pacific Islander	128,577	58,585	46	44,565	35	65,140	51
Other	85,587	38,448	45	31,009	36	42,855	50
White	14,956,253	5,053,477	34	4,525,755	30	5,917,756	40
Total	37,253,956	16,318,065	44	12,856,419	35	18,238,112	49

Sources: California Department of Alcohol and Beverage Control, Raw Data, April 2014. California Department of Finance, Demographic Research Unit, Redistricting Files 2010.

Figure 17. An example of a Table that Meets HCDIP Standards

References Standards

Citations used in the references sections are based on the style used by the *American Journal of Public Health*.

- HCDIP citation format for a journal article

[List all authors by Last Name and First/Middle Name Initials]. [Title of the article]. [Journal Title, in italics]. [Year of publication]. [Volume(Issue):Pages]

Example. Rose G. Sick individuals and sick populations. *Int J Epidemiol*. 1985; 14(1):32---38.

- HCDIP citation format for a report

Option 1.
[List all authors by Last Name and First/Middle Name Initials or if none, lList the affiliated organization(s) or institution(s)]. [Title of the report]. [Place], [Country of Origin]: [Year of publication]. URL: [insert web address] ([insert date that link was accessed]).

Example.
Office of the Deputy Prime Minister. The Impact of Overcrowding on Health & Education: A Review of Evidence and Literature. Wetherby, UK; 2004. URL: http://dera.ioe.ac.uk/id/eprint/5073 (accessed 8-28-2014).

Option 2.
[List all authors by Last Name and First/Middle Name Initials or if none, lList the affiliated organization(s) or institution(s)]. [Title of the report]. [Place], [Country of Origin]: [Year of publication].

Then convert citation to a hyperlink so that pressing Ctrl+Click on the citation takes the user to the website. The citation should appear in blue and be underlined.

Example.
Office of the Deputy Prime Minister. The Impact of Overcrowding on Health & Education: A Review of Evidence and Literature. Wetherby, UK; 2004.

Once maps, graphs, tables and references meet the above HCDIP standards, the exhibits are considered "publication ready." For more information about creating such tables, graphs and maps, consult the HCDIP How-To Illustrative Guide [Reference].

Appendix A: U.S. Census and American Community Survey Products Commonly Used in the HCDIP

Table Type	Abbreviation	Example
Demographic Profile or Data Profile A profile includes tables that provide various demographic, social, economic, and housing characteristics for the U.S., regions, divisions, states, counties, county subdivisions, places, metropolitan areas, American Indian and Alaska Native areas, Hawaiian home lands and congressional districts. For the 2010 Census it includes 100-percent data and for Census 2000, it includes 100-percent and sample data from the decennial censuses. The data profiles from the American Community Survey provide annual estimates for these same social, economic, housing and demographic characteristics.	DP	DP-1, DP03
Quick Tables Predefined tables with frequently requested information for a single geographic area. The information includes numerical data and derived measures (e.g., percent distributions, medians). Users may choose more than one geographic area and more than one table that display and print in a scrolling list.	QT-P	QT-P20
Detailed Table Provide the most detailed data on all topics and geographic areas from the decennial population and the American Community Survey. American Community Survey (ACS) detailed tables begin with the letters "B" for base tables, and "C" for collapsed tables. The "collapsed" tables cover the same topics as the base table, but with fewer details. 'PCT' are population tables that cover geographies to the census tract level.	PCT B	PCT055 B08301
Subject Tables Similar to data profiles but include more detailed ACS data, classified by subject.	S	S1501

To learn more about Census tables and products visit:
Census 2000: https://www.census.gov/population/www/censusdata/c2kproducts.html
FactFinder: http://factfinder2.census.gov/help/en/tables/tables_overview.htm
Census Products: http://www.census.gov/acs/www/data_documentation/product_descriptions/

Appendix B: SAS Code for Poverty Rate Indicator

```
/*HCI_PovertyRate_754_12-11-13 – D Bustamante -
The purpose of this program is to calculate the percent of people and
children that live below the poverty line at the census tract, place, county,
regions and state geographies for years 2000, 2005-2007, 2008-2010, and 2006-
2010 with race/ethnicity stratification.

The ACS provides data on the percent of people in poverty.  Data on the total
population was obtained from different files and used as denominator; the
total population and the percent were used to calculate the numerator: number
of people in poverty.  SEs, 95% CIs, and RSEs were also calculated.

Concentrated poverty was calculated as the proportion of poor people living
in census tracts where overall poverty is >=40%. Census tracts where
50% or more of the population were enrolled in college or graduate school,
where population was below 500 people, 50% or more people living;
in group quarters, and where state or federal prisons are located, were
excluded from the calculation of concentrated poverty

This program was revised on 10-8-14 to fix the addition of empty rows for
geographies without data; the section revised comprises the lines of code
1121 to 1170.
*/

**********Reading look up tables;

* Read indicator titles file;
proc import datafile= 'T:\HCI\Data\ind_id_definition.xls' dbms=xls out=
ind_id_definition replace;
getnames=yes;
run;
proc sort; by ind_id; run;

* Read geographic lookup table;
proc import datafile= 'T:\HCI\Data\Look_up_table_CT_PL_CO_RE_CA.xls' dbms=xls
out= look_up_table_ct_pl_co_re replace;
getnames=yes;
run;

* Removing CMSA fromm lookup table, geography not available for this
indicator;
data look_up_table_ct_pl_co_re (drop=race_eth_code households2010 pop2010
pop_ge25 cmsa_name cmsa_code);
set look_up_table_ct_pl_co_re;
if geotype NE "R4";
run;
proc sort; by geotypevalue; run;

* Read race ethnicity codes and names lookup-table;
proc import datafile=
'T:\HCI\Data\PovertyRate_754\RawData\race_ethnicity.xls' dbms=xls out=
race_eth replace;
getnames=yes;
run;
```

```
* Read report years for this indicator, lookup table;
proc import datafile= 'T:\HCI\Data\PovertyRate_754\RawData\report_years.xls'
dbms=xls out= report_years replace;
getnames=yes;
run;

* Read poverty types lookup-table;
proc import datafile= 'T:\HCI\Data\PovertyRate_754\RawData\poverty.xlsx'
dbms=EXCEL out= poverty_types replace;
getnames=yes;
run;

* Read census tracts to places 2000 lookup table file (for concentrated
poverty);
proc import datafile= 'T:\HCI\Data\PovertyRate_754\RawData\CT00toPL00.xls'
dbms=xls out= CTtoPL00 replace;
getnames=yes;
run;

* Read census tracts to places 2010 lookup table file (for concentrated
poverty);
proc import datafile= 'T:\HCI\Data\PovertyRate_754\RawData\CT10toPL10.xls'
dbms=xls out= CTtoPL10 replace;
getnames=yes;
run;

*Creating a single Census Tract to Place lookup table for the two report
years that have CT data;
data CTtoPL00; set CTtoPL00; reportyear="2000-2000"; run;
data CTtoPL10; set CTtoPL10; reportyear="2006-2010"; run;

data CTtoPL; set CTtoPL00 CTtoPL10;
key=compress(reportyear)||compress(geotypevalue); run;

* Read county FIPS and regions lookup table;
proc import datafile= 'T:\HCI\Data\CA_County_FIPS2Region2010.xls' dbms=xls
out=county_regions replace;
getnames=yes;
run;
proc sort; by county_fips; run;

**********Reading raw data ACS and Census 2000 tables;
* Read in all the Geography-Year files DP03 ACS format;
* This format for years 2008-2010, 2006-2010;

%macro loop1/parmbuff;
   %let num=1;
   %let ext=.csv;
   %let dsname=%scan(&syspbuff,&num);
   %do %while(&dsname ne);

* Read files for processing;
* Each race/ethnicity has a numeric index as a file suffix, except at;
* the CT level;
filename rf "&dsname&ext";
proc import datafile= rf dbms=DLM out=DP03geography_year&num replace;
delimiter=",";
```

```
getnames=yes;
DATAROW= 2;
run;

* Capture the report year and geotype from each file name (filenames;
* have 70 characters without the term .csv);
data DP03geography_year&num (keep = geotype geotypevalue county_fips
reportyear race_eth_code HC03_VC166 HC04_VC166
HC03_VC167 HC04_VC167);
set DP03geography_year&num;
reportyear = substr("&dsname",53,9);
reportyear = translate(reportyear,"-","_");
geotype = substr("&dsname",62,2);
if geotype ="CO" then geotypevalue = substr(geo_id,10,5);
if geotype ="PL" then geotypevalue = substr(geo_id,12,5);
if geotype ="CA" then geotypevalue = substr(geo_id,10,2);
if geotype ="CT" then geotypevalue = substr(geo_id,10,11);
if geotype ="CO" then county_fips = geotypevalue;

* Recode census race_ethnicity codes into HCI codes;
* Replace census race-ethnicity codes and names with HCI standard;
* US Census: Total (001), White (451), Hispanic (400), AfricanAm (453),;
* Asian (457), AIAN (455), NHOPI (459), Multiple (463), Other (461);
* HCI standard: 1=AIAN, 2=Asian, 3=AfricanAm, 4= Latino, 5=NHOPI,;
* 6=White,7=Multiple, 8=Other,9=Total;

if popgroup_id NE . then race_eth_code= popgroup_id;

race_eth_code = tranwrd(race_eth_code,1,9);
race_eth_code = tranwrd(race_eth_code,455,1);
race_eth_code = tranwrd(race_eth_code,457,2);
race_eth_code = tranwrd(race_eth_code,453,3);
race_eth_code = tranwrd(race_eth_code,400,4);
race_eth_code = tranwrd(race_eth_code,459,5);
race_eth_code = tranwrd(race_eth_code,451,6);
race_eth_code = tranwrd(race_eth_code,463,7);
race_eth_code = tranwrd(race_eth_code,461,8);

* 3-year files do not have race_ethnicity, so total population is used;
if popgroup_id = . then race_eth_code = 9;
run;

   %let num=%eval(&num+1);
   %let dsname=%scan(&syspbuff,&num);

   %end;
%mend loop1;

%loop1
(T:\HCI\Data\PovertyRate_754\RawData\ACS_10_SF4_DP03_2006_2010CT
T:\HCI\Data\PovertyRate_754\RawData\ACS_10_SF4_DP03_2006_2010PL
T:\HCI\Data\PovertyRate_754\RawData\ACS_10_SF4_DP03_2006_2010CO
T:\HCI\Data\PovertyRate_754\RawData\ACS_10_SF4_DP03_2006_2010CA
T:\HCI\Data\PovertyRate_754\RawData\ACS_10_3YR_DP03_2008_2010PL
T:\HCI\Data\PovertyRate_754\RawData\ACS_10_3YR_DP03_2008_2010CO
T:\HCI\Data\PovertyRate_754\RawData\ACS_10_3YR_DP03_2008_2010CA);
```

```
* Read in all the Geography-Year files DP3 Census 2000 format;
* This format will be used for year 2000;

%macro loop2/parmbuff;
   %let num=1;
   %let ext=.csv;
   %let dsname=%scan(&syspbuff,&num);
   %do %while(&dsname ne);

* Read files for processing;
* Each race/ethnicity has a numeric index as a file suffix;
filename rf "&dsname&ext";
proc import datafile= rf dbms=DLM out=DP3geography_year&num replace;
delimiter=",";
getnames=yes;
DATAROW= 2;
run;

* Capture the report year and geotype from each file name;
data DP3geography_year&num (keep = geotype geotypevalue county_fips
reportyear race_eth_code HC02_VC105 HC02_VC111);
set DP3geography_year&num;
reportyear = substr("&dsname",53,9);
reportyear = translate(reportyear,"-","_");
geotype = substr("&dsname",62,2);
if geotype ="CO" then geotypevalue = substr(geo_id,10,5);
if geotype ="PL" then geotypevalue = substr(geo_id,12,5);
if geotype ="CA" then geotypevalue = substr(geo_id,10,2);
if geotype ="CT" then geotypevalue = substr(geo_id,10,11);
if geotype NE "CA" then county_fips = geotypevalue;

* Recode census race_ethnicity codes into HCI codes;
* Replace census race-ethnicity codes and names with HCI standard;
* US Census: Total (001), White (451), Hispanic (400), AfricanAm (453),;
* Asian (457), AIAN (455), NHOPI (459), Multiple (463), Other (461);
* HCI standard: 1=AIAN, 2=Asian, 3=AfricanAm, 4= Latino, 5=NHOPI,;
* 6=White,7=Multiple, 8=Other,9=Total;

if popgroup_id NE . then race_eth_code= popgroup_id;

race_eth_code = tranwrd(race_eth_code,1,9);
race_eth_code = tranwrd(race_eth_code,455,1);
race_eth_code = tranwrd(race_eth_code,457,2);
race_eth_code = tranwrd(race_eth_code,453,3);
race_eth_code = tranwrd(race_eth_code,400,4);
race_eth_code = tranwrd(race_eth_code,459,5);
race_eth_code = tranwrd(race_eth_code,451,6);
race_eth_code = tranwrd(race_eth_code,463,7);
race_eth_code = tranwrd(race_eth_code,461,8);

* 3-year files do not have race_ethnicity, so total population is used;
if popgroup_id = . then race_eth_code = 9;

run;

%let num=%eval(&num+1);
%let dsname=%scan(&syspbuff,&num);
```

```
  %end;
%mend loop2;

%loop2
(T:\HCI\Data\PovertyRate_754\RawData\DEC_00_SF4_DP_3_2000_2000CT
T:\HCI\Data\PovertyRate_754\RawData\DEC_00_SF4_DP_3_2000_2000PL
T:\HCI\Data\PovertyRate_754\RawData\DEC_00_SF4_DP_3_2000_2000CO
T:\HCI\Data\PovertyRate_754\RawData\DEC_00_SF4_DP_3_2000_2000CA);

* Read in all the Geography-Year files DP03 B ACS format;
* This format will be used for year 2005-2007;

%macro loop3/parmbuff;
   %let num=1;
   %let ext=.csv;
   %let dsname=%scan(&syspbuff,&num);
   %do %while(&dsname ne);

* Read files for processing;
* Each race/ethnicity has a numeric index as a file suffix, except at;
* the CT level;
filename rf "&dsname&ext";
proc import datafile= rf dbms=DLM out=DP03Bgeography_year&num replace;
delimiter=",";
getnames=yes;
DATAROW= 2;
run;

* Capture the report year and geotype from each file name ;
data DP03Bgeography_year&num (keep = geotype geotypevalue county_fips
reportyear race_eth_code HC01_EST_VC112
HC01_MOE_VC112 HC01_EST_VC113 HC01_MOE_VC113);
set DP03Bgeography_year&num;
reportyear = substr("&dsname",53,9);
reportyear = translate(reportyear,"-","_");
geotype = substr("&dsname",62,2);
if geotype ="CO" then geotypevalue = substr(geo_id,10,5);
if geotype ="PL" then geotypevalue = substr(geo_id,12,5);
if geotype ="CA" then geotypevalue = substr(geo_id,10,2);
if geotype ="CT" then geotypevalue = substr(geo_id,10,11);
if geotype NE "CA" then county_fips = geotypevalue;

* Recode census race_ethnicity codes into HCI codes;
* Replace census race-ethnicity codes and names with HCI standard;
* US Census: Total (001), White (451), Hispanic (400), AfricanAm (453),;
* Asian (457), AIAN (455), NHOPI (459), Multiple (463), Other (461);
* HCI standard: 1=AIAN, 2=Asian, 3=AfricanAm, 4= Latino, 5=NHOPI,;
* 6=White,7=Multiple, 8=Other,9=Total;

if popgroup_id NE . then race_eth_code= popgroup_id;

race_eth_code = tranwrd(race_eth_code,1,9);
race_eth_code = tranwrd(race_eth_code,455,1);
race_eth_code = tranwrd(race_eth_code,457,2);
race_eth_code = tranwrd(race_eth_code,453,3);
race_eth_code = tranwrd(race_eth_code,400,4);
```

```
race_eth_code = tranwrd(race_eth_code,459,5);
race_eth_code = tranwrd(race_eth_code,451,6);
race_eth_code = tranwrd(race_eth_code,463,7);
race_eth_code = tranwrd(race_eth_code,461,8);

* 3-year files do not have race_ethnicity, so total population is used;
if popgroup_id = . then race_eth_code = 9;
run;

   %let num=%eval(&num+1);
   %let dsname=%scan(&syspbuff,&num);

   %end;
%mend loop3;

%loop3
(T:\HCI\Data\PovertyRate_754\RawData\ACS_07_3YR_DP03_2005_2007PL
T:\HCI\Data\PovertyRate_754\RawData\ACS_07_3YR_DP03_2005_2007CO
T:\HCI\Data\PovertyRate_754\RawData\ACS_07_3YR_DP03_2005_2007CA) ;

**********Creating Master files with all years and geographies;
data MasterDP03;
 set DP03geography_year1 DP03geography_year2 DP03geography_year3
DP03geography_year4 DP03geography_year5
 DP03geography_year6 DP03geography_year7 ;
run;

data MasterDP03B;
 set DP03Bgeography_year1 DP03Bgeography_year2 DP03Bgeography_year3;
run;

data MasterDP3;
 set DP3geography_year1 DP3geography_year2 DP3geography_year3
DP3geography_year4;
run;

proc contents data=MasterDP03; run;
proc contents data=MasterDP03B; run;
proc contents data=MasterDP3; run;

**********Renaming variables of interest;
data MasterDP03B;
rename HC01_EST_VC112=PerOverallPov HC01_MOE_VC112=PerOverallPov_MOE
HC01_EST_VC113=PerChildPov
HC01_MOE_VC113=PerChildPov_MOE;
set MasterDP03B;
run;

data MasterDP03;
rename HC03_VC166=PerOverallPov HC04_VC166=PerOverallPov_MOE
HC03_VC167=PerChildPov HC04_VC167=PerChildPov_MOE;
set MasterDP03;
run;

data MasterDP3;
rename HC02_VC105=PerOverallPov HC02_VC111=PerChildPov;
set MasterDP3;
```

```
run;

**********Creating a single master file for Poverty;
**********Replacing MOE for SE;

data Poverty (drop=PerOverallPov_MOE PerChildPov_MOE);
set MasterDP03 MasterDP03B MasterDP3;
PerOverallPov_SE=PerOverallPov_MOE/1.645;
PerChildPov_SE=PerChildPov_MOE/1.645;
run;

proc sort data=Poverty; by geotype reportyear geotypevalue race_eth_code;
run;

**********Reading total population files to calculate denominator, numerator
and concentrated poverty;

********** Read in all the Geography-Year files B01001 format;
********** This format will be used for years 2005-2007,2008-2010, 2006-2010;

%macro loop4/parmbuff;
   %let num=1;
   %let ext=.csv;
   %let dsname=%scan(&syspbuff,&num);
   %do %while(&dsname ne);

* Read files for processing;
* Each race/ethnicity has a numeric index as a file suffix, except at;
* the CT level;
filename rf "&dsname&ext";
proc import datafile= rf dbms=DLM out=B01001geography_year&num replace;
delimiter=",";
getnames=yes;
DATAROW= 2;
run;

* Capture the report year and geotype from each file name (filenames;
* have 70 characters without the term .csv);
data B01001geography_year&num (keep = geotype geotypevalue county_fips
reportyear race_eth_code
TotPop TotPopChild);
*HD01_VD01 HD01_VD03 HD01_VD04 HD01_VD05 HD01_VD06 HD01_VD27 HD01_VD28
HD01_VD29 HD01_VD30;
set B01001geography_year&num;
reportyear = substr("&dsname",72,9);
reportyear = translate(reportyear,"-","_");
geotype = substr("&dsname",81,2);
if geotype ="CO" then geotypevalue = substr(geo_id,10,5);
if geotype ="PL" then geotypevalue = substr(geo_id,12,5);
if geotype ="CA" then geotypevalue = substr(geo_id,10,2);
if geotype ="CT" then geotypevalue = substr(geo_id,10,11);
if geotype ="CO" then county_fips = geotypevalue;

* Recode census race_ethnicity codes into HCI codes;
* Replace census race-ethnicity codes and names with HCI standard;
* US Census: Total (001), White (451), Hispanic (400), AfricanAm (453),;
* Asian (457), AIAN (455), NHOPI (459), Multiple (463), Other (461);
```

```
* HCI standard: 1=AIAN, 2=Asian, 3=AfricanAm, 4= Latino, 5=NHOPI,;
* 6=White,7=Multiple, 8=Other,9=Total;

if popgroup_id NE . then race_eth_code= popgroup_id;

race_eth_code = tranwrd(race_eth_code,1,9);
race_eth_code = tranwrd(race_eth_code,455,1);
race_eth_code = tranwrd(race_eth_code,457,2);
race_eth_code = tranwrd(race_eth_code,453,3);
race_eth_code = tranwrd(race_eth_code,400,4);
race_eth_code = tranwrd(race_eth_code,459,5);
race_eth_code = tranwrd(race_eth_code,451,6);
race_eth_code = tranwrd(race_eth_code,463,7);
race_eth_code = tranwrd(race_eth_code,461,8);

* 3-year files do not have race_ethnicity, so total population is used;
if popgroup_id = . then race_eth_code = 9;

TotPop=HD01_VD01;
*Adding Estimate (children under 18 years): - Under 5 years Estimate: - 5 to
9 years Estimate: - 10 to 14 years Estimate: - 15 to 17 years both male and
female;
TotPopChild=HD01_VD03+HD01_VD04+HD01_VD05+HD01_VD06+HD01_VD27+HD01_VD28+HD01_
VD29+HD01_VD30;

run;

   %let num=%eval(&num+1);
   %let dsname=%scan(&syspbuff,&num);

   %end;
%mend loop4;

%loop4
(T:\HCI\Data\PovertyRate_754\RawData\Total_population\ACS_07_3YR_B01001_2005_
2007CA
T:\HCI\Data\PovertyRate_754\RawData\Total_population\ACS_07_3YR_B01001_2005_2
007CO
T:\HCI\Data\PovertyRate_754\RawData\Total_population\ACS_07_3YR_B01001_2005_2
007PL
T:\HCI\Data\PovertyRate_754\RawData\Total_population\ACS_10_3YR_B01001_2008_2
010CA
T:\HCI\Data\PovertyRate_754\RawData\Total_population\ACS_10_3YR_B01001_2008_2
010CO
T:\HCI\Data\PovertyRate_754\RawData\Total_population\ACS_10_3YR_B01001_2008_2
010PL
T:\HCI\Data\PovertyRate_754\RawData\Total_population\ACS_10_SF4_B01001_2006_2
010CA
T:\HCI\Data\PovertyRate_754\RawData\Total_population\ACS_10_SF4_B01001_2006_2
010CO
T:\HCI\Data\PovertyRate_754\RawData\Total_population\ACS_10_SF4_B01001_2006_2
010PL
T:\HCI\Data\PovertyRate_754\RawData\Total_population\ACS_10_SF4_B01001_2006_2
010CT);

********** Read in all the Geography-Year files DP1 format;
********** This format will be used for year 2000;
```

```
%macro loop5/parmbuff;
   %let num=1;
   %let ext=.csv;
   %let dsname=%scan(&syspbuff,&num);
   %do %while(&dsname ne);

* Read files for processing;
* Each race/ethnicity has a numeric index as a file suffix, except at;
* the CT level;
filename rf "&dsname&ext";
proc import datafile= rf dbms=DLM out=DP1geography_year&num replace;
delimiter=",";
getnames=yes;
DATAROW= 2;
run;

* Capture the report year and geotype from each file name (filenames;
* have 70 characters without the term .csv);
data DP1geography_year&num (keep = geotype geotypevalue county_fips
reportyear race_eth_code TotPop TotPopChild);
*HC01_VC01 HC01_VC19 Saving total population and population eighteen years
and over;
set DP1geography_year&num;
reportyear = substr("&dsname",69,9);
reportyear = translate(reportyear,"-","_");
geotype = substr("&dsname",78,2);
if geotype ="CO" then geotypevalue = substr(geo_id,10,5);
if geotype ="PL" then geotypevalue = substr(geo_id,12,5);
if geotype ="CA" then geotypevalue = substr(geo_id,10,2);
if geotype ="CT" then geotypevalue = substr(geo_id,10,11);
if geotype ="CO" then county_fips = geotypevalue;

* Recode census race_ethnicity codes into HCI codes;
* Replace census race-ethnicity codes and names with HCI standard;
* US Census: Total (001), White (451), Hispanic (400), AfricanAm (453),;
* Asian (457), AIAN (455), NHOPI (459), Multiple (463), Other (461);
* HCI standard: 1=AIAN, 2=Asian, 3=AfricanAm, 4= Latino, 5=NHOPI,;
* 6=White,7=Multiple, 8=Other,9=Total;

if popgroup_id NE . then race_eth_code= popgroup_id;

race_eth_code = tranwrd(race_eth_code,1,9);
race_eth_code = tranwrd(race_eth_code,455,1);
race_eth_code = tranwrd(race_eth_code,457,2);
race_eth_code = tranwrd(race_eth_code,453,3);
race_eth_code = tranwrd(race_eth_code,400,4);
race_eth_code = tranwrd(race_eth_code,459,5);
race_eth_code = tranwrd(race_eth_code,451,6);
race_eth_code = tranwrd(race_eth_code,463,7);
race_eth_code = tranwrd(race_eth_code,461,8);

* 3-year files do not have race_ethnicity, so total population is used;
if popgroup_id = . then race_eth_code = 9;

TotPop=HC01_VC01;
```

```
TotPopChild=HC01_VC01-HC01_VC19;  *Getting the population under 18 by
substracting the population over 18 from the total;
run;

   %let num=%eval(&num+1);
   %let dsname=%scan(&syspbuff,&num);

   %end;
%mend loop5;

%loop5
(T:\HCI\Data\PovertyRate_754\RawData\Total_population\DEC_00_SF2_DP1_2000_200
0CA
T:\HCI\Data\PovertyRate_754\RawData\Total_population\DEC_00_SF2_DP1_2000_2000
CO
T:\HCI\Data\PovertyRate_754\RawData\Total_population\DEC_00_SF2_DP1_2000_2000
PL
T:\HCI\Data\PovertyRate_754\RawData\Total_population\DEC_00_SF2_DP1_2000_2000
CT);

* Concatenating all total population files into a single file;
data Population;
format geotypevalue $CHAR12. county_fips $CHAR5.;
set B01001GEOGRAPHY_YEAR1 B01001GEOGRAPHY_YEAR2 B01001GEOGRAPHY_YEAR3
B01001GEOGRAPHY_YEAR4 B01001GEOGRAPHY_YEAR5 B01001GEOGRAPHY_YEAR6
B01001GEOGRAPHY_YEAR7 B01001GEOGRAPHY_YEAR8 B01001GEOGRAPHY_YEAR9
B01001GEOGRAPHY_YEAR10 DP1GEOGRAPHY_YEAR1 DP1GEOGRAPHY_YEAR2
DP1GEOGRAPHY_YEAR3 DP1GEOGRAPHY_YEAR4;
run;

***********************Merging poverty and population files using a unique
id;
data Poverty;
set Poverty;
ID=trim(geotype)||trim(geotypevalue)||trim(reportyear)||trim(race_eth_code);
run;

data Population;
set Population;
ID=trim(geotype)||trim(geotypevalue)||trim(reportyear)||trim(race_eth_code);
run;

proc sort data=Poverty; by id; run;
proc sort data=Population; by id; run;

PROC SQL;
CREATE TABLE Poverty_Pop(DROP=oldID)
AS
SELECT
coalesce(Poverty.oldID,Population.oldID) AS ID,
*
FROM Population (RENAME=(ID=oldID))
full join
Poverty (RENAME=(ID=oldID))
ON Population.oldID=Poverty.oldID
;
QUIT;
```

```
proc sort data=Poverty_Pop; by geotype reportyear geotypevalue race_eth_code;
run;

**************Calculating the numerator (number of poor people or number of
poor children) from denominator and percent;
data Poverty_Pop (drop=id);
set Poverty_Pop;
if PerOverallPov GE 0 then NumOverallPov=TotPop*PerOverallPov/100;
if PerChildPov GE 0 then NumChildPov=TotPopChild*PerChildPov/100;
run;

proc sort data=Poverty_Pop; by geotype reportyear geotypevalue race_eth_code;
run;

*********************************************;
***********Concentrated poverty CA ***********;
*********************************************;

*Read school enrollment population data for 2000 (table QTP19) and 2006-10 CT
(table S1401);
* This information is necessary to remove the Census tracts with >50%
population enrolled in college or graduate school;
%macro loop6/parmbuff;
   %let num=1;
   %let ext=.csv;
   %let dsname=%scan(&syspbuff,&num);
   %do %while(&dsname ne);

* Read files for processing;
* Each race/ethnicity has a numeric index as a file suffix, except at;
* the CT level;
filename rf "&dsname&ext";
proc import datafile= rf dbms=DLM out=EnrollmentGeography_year&num replace;
delimiter=",";
getnames=yes;
DATAROW= 2;
run;

* Capture the report year and geotype from each file name (filenames;
* have 70 characters without the term .csv);
data EnrollmentGeography_year&num;
set EnrollmentGeography_year&num;
reportyear = substr("&dsname",72,9);
reportyear = translate(reportyear,"-","_");
geotype = substr("&dsname",81,2);
if geotype ="CO" then geotypevalue = substr(geo_id,10,5);
if geotype ="PL" then geotypevalue = substr(geo_id,12,5);
if geotype ="CA" then geotypevalue = substr(geo_id,10,2);
if geotype ="CT" then geotypevalue = substr(geo_id,10,11);
if geotype ="CO" then county_fips = geotypevalue;

* Recode census race_ethnicity codes into HCI codes;
* Replace census race-ethnicity codes and names with HCI standard;
* US Census: Total (001), White (451), Hispanic (400), AfricanAm (453),;
* Asian (457), AIAN (455), NHOPI (459), Multiple (463), Other (461);
* HCI standard: 1=AIAN, 2=Asian, 3=AfricanAm, 4= Latino, 5=NHOPI,;
```

```
* 6=White,7=Multiple, 8=Other,9=Total;

if popgroup_id NE . then race_eth_code= popgroup_id;

race_eth_code = tranwrd(race_eth_code,1,9);
race_eth_code = tranwrd(race_eth_code,455,1);
race_eth_code = tranwrd(race_eth_code,457,2);
race_eth_code = tranwrd(race_eth_code,453,3);
race_eth_code = tranwrd(race_eth_code,400,4);
race_eth_code = tranwrd(race_eth_code,459,5);
race_eth_code = tranwrd(race_eth_code,451,6);
race_eth_code = tranwrd(race_eth_code,463,7);
race_eth_code = tranwrd(race_eth_code,461,8);

* 3-year files do not have race_ethnicity, so total population is used;
if popgroup_id = . then race_eth_code = 9;
run;

   %let num=%eval(&num+1);
   %let dsname=%scan(&syspbuff,&num);

   %end;
%mend loop6;

%loop6
(T:\HCI\Data\PovertyRate_754\RawData\School_Enrollment\ACS_10_5YR_S1401_2006_
2010CT
T:\HCI\Data\PovertyRate_754\RawData\School_Enrollment\DEC_00_SF3_QTP19_2000_2
000CT);

data enrollmentgeography_year1 (keep=geotype geotypevalue College_GS);
set enrollmentgeography_year1;
College_GS=HC01_EST_VC08+HC01_EST_VC09; *Population 3 years and over enrolled
in college or graduate school;
run;

data enrollmentgeography_year2 (keep=geotype geotypevalue College_GS);
set enrollmentgeography_year2;
College_GS=HC01_VC13+HC01_VC15; *Population 3 years and over enrolled in
college or graduate school;
run;

*Read group quarter population data for 2000 and 2006-10 CT;
* This information is necessary to remove the CT with >50% population in a
group quarter;
%macro loop7/parmbuff;
   %let num=1;
   %let ext=.csv;
   %let dsname=%scan(&syspbuff,&num);
   %do %while(&dsname ne);

* Read files for processing;
* Each race/ethnicity has a numeric index as a file suffix, except at;
* the CT level;
filename rf "&dsname&ext";
proc import datafile= rf dbms=DLM out=GroupQGeography_year&num replace;
delimiter=",";
```

```
getnames=yes;
DATAROW= 2;
run;

* Capture the report year and geotype from each file name (filenames;
* have 70 characters without the term .csv);
data GroupQGeography_year&num;
set GroupQGeography_year&num;
reportyear = substr("&dsname",69,9);
reportyear = translate(reportyear,"-","_");
geotype = substr("&dsname",78,2);
if geotype ="CO" then geotypevalue = substr(geo_id,10,5);
if geotype ="PL" then geotypevalue = substr(geo_id,12,5);
if geotype ="CA" then geotypevalue = substr(geo_id,10,2);
if geotype ="CT" then geotypevalue = substr(geo_id,10,11);
if geotype ="CO" then county_fips = geotypevalue;

* Recode census race_ethnicity codes into HCI codes;
* Replace census race-ethnicity codes and names with HCI standard;
* US Census: Total (001), White (451), Hispanic (400), AfricanAm (453),;
* Asian (457), AIAN (455), NHOPI (459), Multiple (463), Other (461);
* HCI standard: 1=AIAN, 2=Asian, 3=AfricanAm, 4= Latino, 5=NHOPI,;
* 6=White,7=Multiple, 8=Other,9=Total;

if popgroup_id NE . then race_eth_code= popgroup_id;

race_eth_code = tranwrd(race_eth_code,1,9);
race_eth_code = tranwrd(race_eth_code,455,1);
race_eth_code = tranwrd(race_eth_code,457,2);
race_eth_code = tranwrd(race_eth_code,453,3);
race_eth_code = tranwrd(race_eth_code,400,4);
race_eth_code = tranwrd(race_eth_code,459,5);
race_eth_code = tranwrd(race_eth_code,451,6);
race_eth_code = tranwrd(race_eth_code,463,7);
race_eth_code = tranwrd(race_eth_code,461,8);

* 3-year files do not have race_ethnicity, so total population is used;
if popgroup_id = . then race_eth_code = 9;
run;

   %let num=%eval(&num+1);
   %let dsname=%scan(&syspbuff,&num);

   %end;
%mend loop7;

%loop7
(T:\HCI\Data\PovertyRate_754\RawData\GroupQuarters\ACS_10_5YR_B26001_2006_201
0CT
T:\HCI\Data\PovertyRate_754\RawData\GroupQuarters\DEC_00_SF2_DP1____2000_2000
CT);

data groupqgeography_year1 (keep=geotype geotypevalue GroupQ);
set groupqgeography_year1;
GroupQ=HD01_VD01; *Population living in group quarters;
run;
```

```
data groupqgeography_year2 (keep=geotype geotypevalue GroupQ);
set groupqgeography_year2;
GroupQ=HC01_VC38; *Population living in group quarters;
run;

* Read CT with federal or state prisions or youth correctional facilities;
proc import datafile=
'T:\HCI\Data\PovertyRate_754\RawData\CA_Prisons\Prisons_in_CT_Lookup_2010.xls
x' dbms=EXCEL out=CA_prisons replace;
getnames=yes;
run;

data CA_prisons (keep= geotypevalue institutions county);
set CA_prisons;
geotypevalue = substr(geo_id,10,11);
run;

* 2006-2010 population and poverty by CT;
data CT_2006_2010_Pop_Pov(drop=TotPopChild PerChildPov PerOverallPov_SE
PerChildPov_SE NumChildPov);
set Poverty_Pop;
if geotype = "CT";
if reportyear = "2006-2010";
if race_eth_code = 9;
run;

* Merge population, poverty, prison, and school enrollment info for CT 2006-
2010;
proc sort data=CT_2006_2010_Pop_Pov; by geotypevalue; run;
proc sort data=EnrollmentGeography_YEAR1; by geotypevalue; run;
proc sort data=GroupQGeography_YEAR1; by geotypevalue; run;
proc sort data=CA_prisons; by geotypevalue; run;

* Create a Concentrated Poverty file for CT year 2006-10 corrected by
population, prisons, and school enrollment;
data CT_2006_2010(drop=GroupQ College_GS TotPop PerOverallPov Institutions
county exclude county_fips NumOverallPov race_eth_code);
merge CT_2006_2010_Pop_Pov EnrollmentGeography_YEAR1 GroupQGeography_YEAR1
CA_prisons;
by geotypevalue;
if College_GS/TotPop >= 0.5 then exclude=1;
if GroupQ/TotPop >= 0.5 then exclude = 1;
if (Institutions NE " " and GroupQ/TotPop >= 0.5) then exclude=1;
if TotPop < 500 then exclude=1;
if (PerOverallPov GE 40 and exclude NE 1) then concentrated="yes"; else
concentrated="no";
run;

* Create a Concentrated Poverty file for CT year 2000;
data CT_2000_Pop_Pov (drop=TotPopChild PerChildPov PerOverallPov_SE
PerChildPov_SE NumChildPov);
set Poverty_Pop;
if geotype = "CT";
if reportyear = "2000-2000";
if race_eth_code = 9;
run;
```

```
* Read CT with federal or state prisions or youth correctional facilities -
year 2000;
proc import datafile=
'T:\HCI\Data\PovertyRate_754\RawData\CA_Prisons\Prisons_in_CT_Lookup_2000.xls
x' dbms=EXCEL out=CA_prisons2000 replace;
getnames=yes;
run;

data CA_prisons2000 (keep= geotypevalue institutions county);
rename tract_key=geotypevalue;
set CA_prisons2000;
run;

/*The following prisons were excluded because did not exist in 2000
06071009102 Federal Correctional Institution (FCI) Victorville Medium I;
Federal Correctional Institution (FCI) Victorville Medium II: United States
Penitentiary (USP) Victorville      FCI Victorville Medium I; FCI Victorville
Medium II; USP Victorville    2010
06019008302 Federal Correctional Institution (FCI) Mendota  FCI Mendota 2011
06047000503 United States Penitentiary (USP) Atwater  USP Atwater 2001
06035040600 Federal Correctional Institution (FCI) Herlong  FCI Herlong 2005
*/

proc sort data=CT_2000_Pop_Pov; by geotypevalue; run;
proc sort data=EnrollmentGeography_YEAR2; by geotypevalue; run;
proc sort data=GroupQGeography_YEAR2; by geotypevalue; run;
proc sort data=CA_prisons2000; by geotypevalue; run;

data CT_2000 (keep=reportyear geotype geotypevalue concentrated);
merge CT_2000_Pop_Pov EnrollmentGeography_YEAR2 GroupQGeography_YEAR2
CA_prisons2000;
by geotypevalue;
if College_GS/TotPop >= 0.5 then exclude=1;
if GroupQ/TotPop >= 0.5 then exclude = 1;
if (Institutions NE " " and GroupQ/TotPop >= 0.5) then exclude=1;
if TotPop < 500 then exclude=1;
if (PerOverallPov GE 40 and exclude NE 1) then concentrated="yes"; else
concentrated="no";
reportyear="2000-2000";
race_eth_code=9;
run;

* File containing information on CT with overall poverty >=40% for 2000 and
2006-2010 (corrected) ;
data CT_00_0610;
set CT_2006_2010 CT_2000;
county_fips=substr(geotypevalue,1,5);
run;

* Merge with CT file that has # poor people by race/eth;
data Poverty_Pop_CT_preCon (drop=TotPop TotPopChild PerOverallPov PerChildPov
PerOverallPov_SE PerChildPov_SE NumChildPov);
set Poverty_Pop;
if geotype="CT";
county_fips=substr(geotypevalue,1,5);
run;
```

```
proc sort data=Poverty_Pop_CT_preCon; by reportyear geotypevalue; run;
proc sort data=CT_00_0610; by reportyear geotypevalue; run;

* File containing concentrated poverty Census Tracts 2000 and 2006-10;
data Poverty_Pop_CT;
merge Poverty_Pop_CT_preCon CT_00_0610;
by reportyear geotypevalue;
key=compress(reportyear)||compress(geotypevalue);
run;

* Concentrated poverty by place;
* Assign CT to places;
proc sort data=Poverty_Pop_CT; by key; run;
proc sort data=CTtoPL; by key; run;

PROC SQL;
CREATE TABLE Poverty_Pop_CT_PL(DROP=oldkey)
AS
SELECT
coalesce(Poverty_Pop_CT.oldkey,CTtoPL.oldkey) AS key,
*
FROM Poverty_Pop_CT (RENAME=(key=oldkey))
full join
CTtoPL (RENAME=(key=oldkey))
ON Poverty_Pop_CT.oldkey=CTtoPL.oldkey;
QUIT;

proc sort data=Poverty_Pop_CT_PL; by concentrated reportyear place_code
race_eth_code; run;

* Adding up the number of poor people in concentrated poverty tracts;
proc means data=Poverty_Pop_CT_PL noprint;
by concentrated reportyear place_code race_eth_code;
var NumOverallPov;
output out=ConcPovPL sum=NumConcPov1;
run;

* The following steps are necessary to keep the true zeroes different from
the missing values;
data ConcPovPLno (drop=_type_ _freq_ NumConcPov1);
rename place_code=geotypevalue;
set ConcPovPL;
if concentrated="no";
if NumConcPov1 NE . then NumConcPov2=0; else NumConcPov2=.;
run;

data ConcPovPLyes (drop=_type_ _freq_ );
rename place_code=geotypevalue;
set ConcPovPL;
if concentrated="yes";
run;

proc sort data=ConcPovPLno; by reportyear geotypevalue race_eth_code; run;
proc sort data=ConcPovPLyes; by reportyear geotypevalue race_eth_code; run;

* File containing concentrated poverty by place;
data ConPovPLTot (drop=concentrated NumConcPov1 NumConcPov2);
```

```
merge ConcPovPLno ConcPovPLyes;
by reportyear geotypevalue race_eth_code;
if NumConcPov1>=0 then NumConcPov=NumConcPov1;
if (NumConcPov2 NE . and NumConcPov1 = .) then NumConcPov=NumConcPov2;
if reportyear=" " then delete;
if geotypevalue=" " then delete;
if race_eth_code=" " then delete;
geotype="PL";
run;

* Concentrated poverty by county;
proc sort data=Poverty_Pop_CT; by reportyear concentrated county_fips
race_eth_code; run;

* Adding up the number of poor people in concentrated poverty tracts;
proc means data=Poverty_Pop_CT noprint;
by reportyear concentrated county_fips race_eth_code;
var NumOverallPov;
output out=ConcPovCO sum=NumConcPov1;
run;

* The following steps are necessary to keep the true zeroes different from
the missing values;
data ConcPovCOno (drop=_type_ _freq_ NumConcPov1);
rename county_fips=geotypevalue;
set ConcPovCO;
if concentrated="no";
if NumConcPov1 NE . then NumConcPov2=0; else NumConcPov2=.;
run;

data ConcPovCOyes (drop=_type_ _freq_ );
rename county_fips=geotypevalue;
set ConcPovCO;
if concentrated="yes";
run;

proc sort data=ConcPovCOno; by reportyear geotypevalue race_eth_code; run;
proc sort data=ConcPovCOyes; by reportyear geotypevalue race_eth_code; run;

* File containing concentrated poverty by county;
data ConPovCOTot (drop=concentrated NumConcPov1 NumConcPov2);
merge ConcPovCOno ConcPovCOyes;
by reportyear geotypevalue race_eth_code;
if NumConcPov1>=0 then NumConcPov=NumConcPov1;
if (NumConcPov2 NE . and NumConcPov1 = .) then NumConcPov=NumConcPov2;
if reportyear=" " then delete;
if geotypevalue=" " then delete;
if race_eth_code=" " then delete;
geotype="CO";
run;

* Concentrated poverty CA;
proc sort data=Poverty_Pop_CT; by reportyear concentrated race_eth_code; run;

* Adding up the number of poor people in concentrated poverty tracts;
proc means data=Poverty_Pop_CT noprint;
by reportyear concentrated race_eth_code;
```

```
var NumOverallPov;
output out=ConcPovCA sum=NumConcPov1;
run;

* The following steps are necessary to keep the true zeroes different from
the missing values;
data ConcPovCAno (drop=_type_ _freq_ NumConcPov1);
set ConcPovCA;
if concentrated="no";
if NumConcPov1 NE . then NumConcPov2=0; else NumConcPov2=.;
run;

data ConcPovCAyes (drop=_type_ _freq_ );
set ConcPovCA;
if concentrated="yes";
run;

proc sort data=ConcPovCAno; by reportyear race_eth_code; run;
proc sort data=ConcPovCAyes; by reportyear race_eth_code; run;

* File containing concentrated poverty state;
data ConPovCATot (drop=concentrated NumConcPov1 NumConcPov2);
merge ConcPovCAno ConcPovCAyes;
by reportyear race_eth_code;
if NumConcPov1>=0 then NumConcPov=NumConcPov1;
if (NumConcPov2 NE . and NumConcPov1 = .) then NumConcPov=NumConcPov2;
if reportyear=" " then delete;
if race_eth_code=" " then delete;
geotypevalue="06";
geotype="CA";
run;

* Concentrated poverty by place county and state;
data ConcPov;
set ConPovPLTot ConPovCOTot ConPovCATot;
run;

*Merging concentrated poverty with long file, calculating the proportion of
the poor living in high poverty CT;
proc sort data=Poverty_Pop; by geotype reportyear race_eth_code geotypevalue;
run;
proc sort data=ConcPov; by geotype reportyear race_eth_code geotypevalue;
run;

data Poverty_Pop_Con;
merge Poverty_Pop ConcPov;
by geotype reportyear race_eth_code geotypevalue;
if PerOverallPov=. then NumConcPov=.;
if TotPop=. then NumConcPov=.;
if NumConcPov > NumOverallPov then NumConcPov = NumOverallPov;
if NumOverallPov>0 then ConcPov=NumConcPov/NumOverallPov*100;
run;

proc sort data=Poverty_Pop_Con; by geotype reportyear race_eth_code
geotypevalue; run;

****************************************************;
```

```
***Regional calculation of outcome and se ***;
****************************************************;

%macro loop6(yrlist);
%local i yrgroup;
%let i=1;
%let yrgroup=%scan(&yrlist,&i);
%do %while (&yrgroup ne);

* Select out total population for counties;
* SAS looping macro does not recognize hyphen in arguments so;
* They are removed for year ranges;
data counties_x_regions&i;
  set Poverty_Pop_Con;
if reportyear = "2006-2010" then reportyear="20062010";
if reportyear = "2005-2007" then reportyear="20052007";
if reportyear = "2008-2010" then reportyear="20082010";
if reportyear = "2000-2000" then reportyear="20002000";
if geotype="CO" and reportyear=&yrgroup;
run;

proc sort; by county_fips; run;

* Merge with regions to create county files by mode;
data counties_x_regions&i;
 merge counties_x_regions&i county_regions;
 by county_fips;
 re_PerOverallPov_se_sq = PerOverallPov_se**2;
 re_TotPop_sq = TotPop**2;
 re_PerChildPov_se_sq = PerChildPov_se**2;
 re_TotPopChild_sq = TotPopChild**2;
run;

proc sort; by race_eth_code region_code; run;

* Calculate the county-weighted percent, se and rse for regions;
proc means noprint data = counties_x_regions&i;
by race_eth_code region_code;
var TotPop;
var TotPopChild;
var NumOverallPov;
var NumChildPov;
var NumConcPov;
var PerOverallPov / weight = TotPop;
var PerChildPov / weight = TotPopChild;
var re_PerOverallPov_se_sq / weight = re_TotPop_sq;
var re_PerChildPov_se_sq / weight = re_TotPopChild_sq;
output out = poverty_re&i sum = TotPop TotPopChild NumOverallPov NumChildPov
NumConcPov PerOverallPov_sum PerChildPov_sum re_PerOverallPov_se_sq_sum
re_PerChildPov_se_sq_sum mean=TotPop_mean TotPopChild_mean NumOverallPov_mean
NumChildPov_mean TotPopCon_mean PerOverallPov PerChildPov
re_PerOverallPov_se_sq_mean re_PerChildPov_se_sq_mean;
run;

data poverty_re&i;
format reportyear $CHAR9.0;
set poverty_re&i;
```

```
PerOverallPov_se = ((re_PerOverallPov_se_sq_sum)**.5)/TotPop;
PerChildPov_se = ((re_PerChildPov_se_sq_sum)**.5)/TotPopChild;
if NumOverallPov>0 then ConcPov=NumConcPov/NumOverallPov*100;
if NumOverallPov=0 then ConcPov=0;
reportyear = "&yrgroup";
geotype = "RE";
geotypevalue = region_code;
run;

%let i=%eval(&i+1);
%let yrgroup=%scan(&yrlist,&i);
%end;
%mend loop6;
%loop6(20002000 20052007 20082010 20062010);


*** Add back the hyphen in report year;
data poverty_re (drop=_TYPE_ _FREQ_ TotPop_mean TotPopChild_mean
NumOverallPov_mean NumChildPov_mean re_PerOverallPov_se_sq_mean
re_PerChildPov_se_sq_mean re_PerOverallPov_se_sq_sum re_PerChildPov_se_sq_sum
PerOverallPov_sum PerChildPov_sum TotPopCon_mean);
 set poverty_re1 poverty_re2 poverty_re3 poverty_re4;
if reportyear = "20062010" then reportyear="2006-2010";
if reportyear = "20052007" then reportyear="2005-2007";
if reportyear = "20082010" then reportyear="2008-2010";
if reportyear = "20002000" then reportyear="2000-2000";
if race_eth_code = . then delete;
run;

**********************Merging regional and total files;
data Poverty_Pop_Con_2 (drop=region_code);
set Poverty_Pop_Con poverty_re;
run;

proc sort data=Poverty_Pop_Con_2; by geotype reportyear race_eth_code
geotypevalue; run;

**********************Creating a stacked file with all poverty rates in the
same column;

data Poverty_Pop_Con_Ov (drop= TotPopChild NumChildPov PerChildPov
PerChildPov_SE NumConcPov ConcPov);
rename TotPop=TotalPop NumOverallPov=NumPov     PerOverallPov=percent
PerOverallPov_SE=percent_SE;
format Poverty $CHAR12.;
set Poverty_Pop_Con_2;
Poverty="Overall";
run;

data Poverty_Pop_Con_Ch (drop= TotPop NumOverallPov PerOverallPov
PerOverallPov_SE NumConcPov ConcPov);
rename TotPopChild=TotalPop NumChildPov=NumPov  PerChildPov=percent
PerChildPov_SE=percent_SE;
format Poverty $CHAR12.;
set Poverty_Pop_Con_2;
Poverty="Child";
run;
```

```
data Poverty_Pop_Con_Co (drop= TotPop PerOverallPov PerOverallPov_SE
TotPopChild NumChildPov PerChildPov PerChildPov_SE);
rename NumConcPov=NumPov ConcPov=percent NumOverallPov=TotalPop;
format Poverty $CHAR12.;
set Poverty_Pop_Con_2;
Poverty="Concentrated";
if geotype="CT" then delete;
run;

data Poverty_Pop_Con_3;
set Poverty_Pop_Con_Ov Poverty_Pop_Con_Ch Poverty_Pop_Con_Co;
run;

************* Calculate Decile of geographically resolved area all races;
data Poverty2PL0610;
set Poverty_Pop_Con_3;
if reportyear = "2006-2010" and geotype = "PL" and race_eth_code=9;
run;

************* Create statewide deciles with total race-ethnicity;
data Poverty2PL0610;
set Poverty2PL0610;
proc rank groups=10 out=Poverty2PL0610_decile descending;
var percent;
ranks place_decile;
run;

*** Rescale so top decile = 1;
data Poverty2PL0610_decile (keep = reportyear geotypevalue place_decile
race_eth_code);
set Poverty2PL0610_decile;
place_decile = place_decile + 1;
race_eth_code = 9;
run;

proc sort data = Poverty2PL0610_decile; by reportyear geotypevalue
race_eth_code;run;
proc sort data = Poverty_Pop_Con_3; by reportyear geotypevalue
race_eth_code;run;

*** Merge decile data back to master file;
data Poverty_Pop_Con_4;
merge Poverty_Pop_Con_3 Poverty2PL0610_decile;
by reportyear geotypevalue race_eth_code;
run;

********** Calculate Relative Risk for each reportyear;

* Create look-up table by mode all races combined for California;
data PovertyCA (drop = TotPop NumPov percent_SE geotypevalue geoname geotype
county_fips place_decile);
set Poverty_Pop_Con_3 (rename = (percent=percent_CA));
if geotype = "CA" and race_eth_code=9;
run;

proc sort data = PovertyCA; by Poverty reportyear ; run;
```

```
proc sort data = Poverty_Pop_Con_4; by Poverty reportyear ; run;

* Merge file with look-up table;
data Poverty_Pop_Con_5(drop = percent_CA );
merge PovertyCA Poverty_Pop_Con_4;
by Poverty reportyear;
CA_RR=percent/percent_CA;
run;

data Poverty_Pop_Con_6;
set Poverty_Pop_Con_5;
if percent_SE NE . then LL_95CI_percent=percent-1.96*percent_SE;
if percent_SE NE . then UL_95CI_percent=percent+1.96*percent_SE;
if percent > 0 then percent_RSE=percent_SE/percent*100;
if (reportyear NE "2000-2000" and LL_95CI_percent<0) then LL_95CI_percent=0;
if (reportyear NE "2000-2000" and UL_95CI_percent>100) then
UL_95CI_percent=100;
if (reportyear = "2000-2000" ) then LL_95CI_percent=.;
if (reportyear = "2000-2000" ) then UL_95CI_percent=.;
run;

* Adding a flag indicating which census tract where considered for
concentrated poverty (ConcentratedCT);
data CT_00_0610 (drop=NumOverallPov key);
set CT_00_0610;
run;

proc sort data=Poverty_Pop_Con_6; by geotype geotypevalue reportyear; run;
proc sort data=CT_00_0610; by geotype geotypevalue reportyear; run;

data Poverty_Pop_Con_6_2;
rename concentrated=ConcentratedCT;
merge Poverty_Pop_Con_6 CT_00_0610;
by geotype geotypevalue reportyear;
run;

**********Create geography by year by poverty type by race master ;

* Create dummy ID ;
data look_up_table_ct_pl_co_re; set look_up_table_ct_pl_co_re; ID = 1;  run;
data report_years; set report_years; ID = 1; run;

* Create many-to-many table of all geography x year combinations;
 proc SQL;
CREATE TABLE geography_year AS SELECT *
FROM look_up_table_ct_pl_co_re FULL JOIN report_years
ON look_up_table_ct_pl_co_re.ID=report_years.ID;

* Create many-to-many table of all geography x year x poverty types
combinations;
proc SQL;
CREATE TABLE geography_year_types AS SELECT *
FROM geography_year FULL JOIN poverty_types
ON geography_year.ID=poverty_types.ID;

* Create many-to-many table of all geography x year x poverty types x race
combinations;
```

```
proc SQL;
CREATE TABLE geography_year_types_race AS SELECT *
FROM geography_year_types FULL JOIN race_eth
ON geography_year_types.ID=race_eth.ID;

*Drop unnecessary variables and rows and create a key to match geography
master to output file;
data geography_year_types_race  (drop=id);
set geography_year_types_race ;
key=trim(geotypevalue)||trim(reportyear)||trim(geotype)||trim(Poverty)||trim(
race_eth_code);
if geotype="CT" and Poverty="Concentrated" then delete;
if geotype="CT" and reportyear="2005-2007" then delete;
if geotype="CT" and reportyear="2008-2010" then delete;
if reportyear="2005-2007" and race_eth_code NE 9 then delete;
if reportyear="2008-2010" and race_eth_code NE 9 then delete;
if reportyear="2005-2007"  and Poverty="Concentrated"  then delete;
if reportyear="2008-2010"  and Poverty="Concentrated" then delete;
run;

***Drop unnecessary variable and create a key to match output to geography
master;
data Poverty_Pop_Con_6_2 (drop=county_fips);
set Poverty_Pop_Con_6_2;
key=trim(geotypevalue)||trim(reportyear)||trim(geotype)||trim(Poverty)||trim(
race_eth_code);
run;

***Sorting geography master and output by key before merging;
proc sort data = geography_year_types_race; by key;run;
proc sort data = Poverty_Pop_Con_6_2; by key;run;

********** Create master output file that includes each geography;
********** as a row even if there no data are available;
data Poverty_Pop_Con_7;
merge geography_year_types_race(IN=In1) Poverty_Pop_Con_6_2 (IN=In2);
by key;
ind_id=754;
if In1=1 then output Poverty_Pop_Con_7;
run;

***Add race/ethnicity name to output file;
proc sort data = race_eth; by race_eth_code;run;
proc sort data = Poverty_Pop_Con_7; by race_eth_code;run;

data Poverty_Pop_Con_7 (drop=key);
merge Poverty_Pop_Con_7 race_eth;
by race_eth_code;
run;

***Add indicator Id and definition to output file;
proc sort data = Poverty_Pop_Con_7; by ind_id;run;
proc sort data = ind_id_definition; by ind_id;run;

data Poverty_Pop_Con_7;
merge Poverty_Pop_Con_7 ind_id_definition;
by ind_id;
```

```
if ind_id=754;
run;

data Poverty_Pop_Con_8  (drop=id);
retain ind_id ind_definition reportyear race_eth_code race_eth_name geotype
geotypevalue geoname
county_name county_fips region_name region_code Poverty TotalPop NumPov
percent LL_95CI_percent UL_95CI_percent
percent_SE percent_RSE place_decile CA_RR ConcentratedCT version;
set Poverty_Pop_Con_7;
if ((geotype="CT" and reportyear="2005-2007") or (geotype="CT" and
reportyear="2008-2010")) then delete;
if ((Poverty="Concentrated" and reportyear="2005-2007") or
(Poverty="Concentrated" and reportyear="2008-2010")) then delete;
if percent_se=. then LL_95CI_percent=.;
if percent_se=. then UL_95CI_percent=.;
if race_eth_code=. then race_eth_code=9;
if race_eth_code=9 then race_eth_name="Total";
if reportyear = "2000-2000" then reportyear = "2000";
version = DATETIME();
format version datetime16.0;
run;

proc sort data=Poverty_Pop_Con_8;
by geotype reportyear geoname Poverty;
run;

* Export complete file;
proc export data= Poverty_Pop_Con_8
outfile='T:\HCI\Data\PovertyRate_754\Output\HCI_PovertyRate_754_CT_PL_CO_RE_C
A.XLSX' dbms=XLSX replace;
run;

/*
*Commented out section: calculating percentage of statistically unreliable
data;
data Poverty_Pop_Con_8;
set Poverty_Pop_Con_8;
if percent_rse>= 30.5 then flag="Unreliable";
if percent_rse > 0 and percent_rse < 30.5 then flag="Reliable";
if percent_rse=0 or percent_rse = . then flag = "ND";
id=1;
run;

proc sort data=Poverty_Pop_Con_8; by geotype  reportyear race_eth_code flag
poverty; run;

proc means data=Poverty_Pop_Con_8;
var id;
var TotalPop;
by geotype reportyear race_eth_code  flag poverty;
output out=stability sum=id TotalPop;

proc sort data=stability; by geotype reportyear race_eth_code poverty; run;

proc transpose data=stability out=stability_geographies;
var id; by geotype reportyear race_eth_code poverty; id flag; run;
```

```
proc transpose data=stability out=stability_population;
var TotalPop; by geotype reportyear race_eth_code poverty; id flag; run;

data stability_geographies (drop=_name_);
set stability_geographies;
if reliable=. then reliable=0;
if unreliable=. then unreliable=0;
if ND=. then ND=0;
total_available=unreliable+reliable;
total=unreliable+ND+reliable;
percent_unreliable_av=unreliable/total_available*100;
percent_unreliable=unreliable/total*100;
version = DATETIME();
format version datetime16.0;
run;

data stability_population (drop=_name_);
set stability_population;
if reliable=. then reliable=0;
if unreliable=. then unreliable=0;
if ND=. then ND=0;
total_available=unreliable+reliable;
total=unreliable+ND+reliable;
percent_unreliable_av=unreliable/total_available*100;
percent_unreliable=unreliable/total*100;
version = DATETIME();
format version datetime16.0;
run;

proc export data= stability_geographies
outfile='T:\HCI\Data\PovertyRate_754\Output\HCI_PovertyRate_754_stability_geo
graphies.xlsx' dbms=EXCEL replace;
run;

proc export data= stability_population
outfile='T:\HCI\Data\PovertyRate_754\Output\HCI_PovertyRate_754_stability_pop
ulation.xlsx' dbms=EXCEL replace;
run;
*/
```

Appendix C: Metropolitan Planning Organization (MPO) Regions of California

County_FIPS	County	MPO Region	MPO_Region_code
06001	Alameda	Bay Area	1
06013	Contra Costa	Bay Area	1
06041	Marin	Bay Area	1
06055	Napa	Bay Area	1
06075	San Francisco	Bay Area	1
06081	San Mateo	Bay Area	1
06085	Santa Clara	Bay Area	1
06095	Solano	Bay Area	1
06097	Sonoma	Bay Area	1
06007	Butte	Butte	2
06003	Alpine	Central/Southeast Sierra	3
06005	Amador	Central/Southeast Sierra	3
06009	Calaveras	Central/Southeast Sierra	3
06027	Inyo	Central/Southeast Sierra	3
06043	Mariposa	Central/Southeast Sierra	3
06051	Mono	Central/Southeast Sierra	3
06109	Tuolumne	Central/Southeast Sierra	3
06053	Monterey	Monterey Bay	4
06069	San Benito	Monterey Bay	4
06087	Santa Cruz	Monterey Bay	4
06015	Del Norte	North Coast	5
06023	Humboldt	North Coast	5
06033	Lake	North Coast	5
06045	Mendocino	North Coast	5
06105	Trinity	North Coast	5
06035	Lassen	Northeast Sierra	6
06049	Modoc	Northeast Sierra	6
06057	Nevada	Northeast Sierra	6
06063	Plumas	Northeast Sierra	6
06091	Sierra	Northeast Sierra	6
06093	Siskiyou	Northeast Sierra	6
06011	Colusa	Northern Sacramento Valley	7
06021	Glenn	Northern Sacramento Valley	7
06103	Tehama	Northern Sacramento Valley	7
06017	El Dorado	Sacramento Area	8
06061	Placer	Sacramento Area	8
06067	Sacramento	Sacramento Area	8
06101	Sutter	Sacramento Area	8
06113	Yolo	Sacramento Area	8
06115	Yuba	Sacramento Area	8
06073	San Diego	San Diego	9
06019	Fresno	San Joaquin Valley	10

County_FIPS	County	MPO Region	MPO_Region_code
06029	Kern	San Joaquin Valley	10
06031	Kings	San Joaquin Valley	10
06039	Madera	San Joaquin Valley	10
06047	Merced	San Joaquin Valley	10
06077	San Joaquin	San Joaquin Valley	10
06099	Stanislaus	San Joaquin Valley	10
06107	Tulare	San Joaquin Valley	10
06079	San Luis Obispo	San Luis Obispo	11
06083	Santa Barbara	Santa Barbara	12
06089	Shasta	Shasta	13
06025	Imperial	Southern California	14
06037	Los Angeles	Southern California	14
06059	Orange	Southern California	14
06065	Riverside	Southern California	14
06071	San Bernardino	Southern California	14
06111	Ventura	Southern California	14

(Halter G, Hitchcock A, Kelly T, Mortenson M. 2010 California Regional Progress Report: One State, Many Regions, and Our Future: Tracking Progress Toward Sustainability. Sacramento, CA: Applied Development Economics Inc., Collaborative Economics Inc., Information Center for the Environment at UC-Davis, CALTRANS: Division of Transportation Planning; 2010. http://dot.ca.gov/hq/tpp/offices/orip/Collaborative%20Planning/Files/CARegionalProgress_2-1-2011.pdf Accessed on 7/5/12.)

Appendix D: Road Traffic Injuries Indicator Numerator Information, SWITRS Detailed Codebook

SWITRS Collision Raw Data

Item Name	Variable Name	Description	Possible Values
Case Id	**caseid**	the unique identifier of the collision report (barcode beginning 2002; 19 digit code prior to 2002)	
Collision Year	**year**	the year when the collision occurred	
County City Location	**location**	the location code of where the collision occurred	Data may appear with no leading zero.
Caltrans County	**caltransc**		Includes blanks and nulls
City			
County			

SWITRS Party Raw Data

Case Id	**caseid**	the unique identifier of the collision report (barcode beginning 2002; 19 digit code prior to 2002)	
Party Number	**parnum**		1 to 999
Party Type	**ptype**		1 - Driver (including Hit and Run) 2 - Pedestrian 3 - Parked Vehicle 4 - Bicyclist 5 - Other - - Not Stated
Statewide Vehicle Type	**vehtype**		A - Passenger Car/Station Wagon B - Passenger Car with Trailer C - Motorcycle/Scooter D - Pickup or Panel Truck E - Pickup or Panel Truck with Trailer

			F - Truck or Truck Tractor G - Truck or Truck Tractor with Trailer H - Schoolbus I - Other Bus J - Emergency Vehicle K - Highway Construction Equipment L - Bicycle M - Other Vehicle N - Pedestrian O - Moped - or blank - Not Stated

SWITRS Victim Raw Data

Case Id	**caseid**	the unique identifier of the collision report (barcode beginning 2002; 19 digit code prior to 2002)	
Party Number	**parnum**		1 to 999
Victim Role	**vtype**		1 - Driver 2 - Passenger (includes non-operator on bicycle or any victim on/in parked vehicle or multiple victims on/in non-motor vehicle) 3 - Pedestrian 4 - Bicyclist 5 - Other (single victim on/in non-motor vehicle; e.g. ridden animal, horse-drawn carriage, train, or building) 6 - Non-Injured Party
Victim Degree of Injury	**vinjury**		1 - Killed 2 - Severe Injury 3 - Other Visible Injury 4 - Complaint of Pain 0 - No Injury
Victim Sex	**vsex**		M - Male F - Female

			- - Not Stated
Victim Age	**vage**	the age of the victim at the time of the collision	0 - 125+ (998=UNKNOWN)

Classification of HCDIP Victim Mode of Transport

SWITRS (victim) Mode (Vtype)	HCDIP injury mode
Passenger Car Pickup Truck Passenger Car & Trailer Pickup Truck & Trailer Other Vehicle	Car/Pickup
Truck or Truck Tractor Truck & Trailer Highway Construction Equipment Emergency Vehicle	Truck
Schoolbus Other Bus	Bus
Motorcycle Moped	Motorcycle
Pedestrian	Pedestrian
Bicycle	Bicyclist
Vtype = Other-non motor vehicle AND Vehtype:	
Passenger Car Pickup Truck Passenger Car & Trailer Pickup Truck & Trailer Other Vehicle - (not stated)	Car/Pickup
Truck or Truck Tractor Truck & Trailer Highway Construction Equipment Emergency Vehicle	Truck
Schoolbus Other Bus	Bus

Motorcycle Moped	Motorcycle
Pedestrian	Pedestrian

SWITRS Codes to resolve "Other-Non motor vehicle"

Other Non Motor Veh	Party Type Ptype	Vehicle type of party (vehtype)	Involv	Ped
1 - Driver 2 - Passenger (includes non-operator on bicycle or any victim on/in parked vehicle or multiple victims on/in non-motor vehicle) 3 - Pedestrian 4 - Bicyclist 5 - Other (single victim on/in non-motor vehicle; e.g. ridden animal, horse-drawn carriage, train, or building) 6 - Non-Injured Party	1 - Driver (including Hit and Run) 2 - Pedestrian 3 - Parked Vehicle 4 - Bicyclist 5 - Other - - Not Stated	Pickup truck Passenger Car Pickup Truck & Trailer Passenger Car Highway Construction Equipment Motorcycle Bicycle Pedestrian	A - Non-Collision B - Pedestrian C - Other Motor Vehicle D - Motor Vehicle on Other Roadway E - Parked Motor Vehicle F - Train G - Bicycle H - Animal I - Fixed Object J - Other Object - - Not Stated	A - No Pedestrian Involved B - Crossing in Crosswalk at Intersection C - Crossing in Crosswalk Not at Intersection D - Crossing Not in Crosswalk E - In Road, Including Shoulder F - Not in Road G - Approaching/Leaving School Bus - - Not Stated