Index of Concentration at the Extremes Data and Methods Narrative

Objective

This document will describe the methods and analyses used to inform and develop the racial, ethnic, and economic polarization dashboards as tools to measure and monitor spatial social polarization in California.

Background

Social polarization is defined as "the segregation that emerges when factors such as income inequality, real-estate fluctuations, and economic displacement result in the differentiation of social groups. It is a state and/or tendency denoting the growth of groups at the extremities of the social hierarchy and the parallel shrinking of groups around its middle. [1]" Spatial social polarization is the geographic separation of populations by various sociodemographic characteristics, including racial and ethnic identity and economic class among other social determinants of health. This results in unequal distribution of resources and representation, and persistent polarization over time.

Historic policies and practices partly set the stage for present day social polarization. One example is the 1930s practice of redlining, or grading neighborhoods as most to least favorable for mortgage lending programs [2]. Not coincidentally, least favorable neighborhoods were predominantly areas where Black residents lived [2]. Overtime neighborhoods that were assigned higher grades saw investments and development while "redlined" neighborhoods were neglected and lacked resources [2]. Today, historically redlined neighborhoods are associated with worse health outcomes and health disparities compared to nearby neighborhoods that were deemed worthy of investment [2-5].

While the harmful legacies of historic practices, like redlining, are often referenced in the health equity literature, there remains a need to understand and measure social polarization today. The extensive literature documenting overlap in historically redlined neighborhoods and presently observed health disparities, highlights the importance of visualizing where populations are concentrated and how concentrations increase or disperse over time. Here, we introduce the Index of Concentration at the Extremes ("index") as a tool to monitor spatial social polarization in California. The Index measures the extent to which the population in a given area is concentrated by characteristics that are advantageous or disadvantageous [6]. The index is calculated using the below formula where *i* is a geographic area or unit (e.g., county, census tract), A_i is the number of residents in the advantaged group, P_i is the number of residents in the disadvantaged group, and T_i is the total population in the geographic area.

$$Index\ of\ Concentration\ at\ the\ Extremes_i = \frac{A_i - P_i}{T_i}$$

Index values range from -1 to 1, where -1 indicates that the entire population meets the criteria for disadvantage and a value of 1 indicates the entire population meets the criteria of advantage. Researchers applying the index will select the characteristics to be measured and define disadvantage and advantage for the population of interest. In public health research, the index has been applied to race and ethnicity, and various social determinants of health (SDoH) including household income and education [4, 7-9]. Indices have also been evaluated for associations with health outcomes, such as preterm birth, premature mortality, infant mortality, and diabetes mortality [4, 7-10]. Based on these studies, populations polarized toward disadvantage tend to experience worse health outcomes.

Additionally, researchers evaluated the index at different geographic levels. Specifically, Krieger et. al. examined how well the index for poverty predicted premature mortality at the census tract and neighborhood levels in Boston, compared to the ability of the poverty rate to predict the outcome [8]. They found that the index for poverty accurately predicted premature mortality rates in 84% of census tracts compared to 78% when using the poverty rate [8]. Furthermore, premature mortality rates predicted using the index for poverty were closer to the observed rates at the census tract level compared to the neighborhood level (as defined by the Boston Public Health Commission), supporting utilization of the index at smaller geographic areas [8].

Based on this research, we estimated indices for racial, ethnic, and economic polarization for California counties and census tracts.

Methods

Definitions for California indices

We developed three indices for California using demographic data from the American Community Survey (ACS) 2019 5-year population estimates files. We measured racial and ethnic polarization using race and ethnicity data reported in ACS Table DP05, and economic polarization using poverty data reported in ACS Table S1701. For both race and ethnicity indices, the advantaged group was defined as non-Hispanic White alone (not in combination with other race groups). Disadvantaged was defined as Black/African American alone or in combination with other races for the race index, and Hispanic/Latino (any race) for the ethnicity index. These criteria were based on historic and ongoing discrimination and bias against Black or African American residents across the US, leading to disproportionately higher rates of premature mortality and adverse health outcomes in this population. Additionally, Hispanic and Latino residents make up a large population in California that also faces discrimination and marginalization. In the interest of capturing Californians' lived experiences and based on literature that shows multiracial individuals of African American descent experience individual and institutional racism that affects health outcomes [11], we included multi-race residents in our disadvantaged group.

Table 1: Characteristic-specific definitions of disadvantage and advantage

	Disadvantaged group	Advantaged group
Race	Black or African American alone or in combination with other races	Non-Hispanic White
Ethnicity	Hispanic or Latino, any race	Non-Hispanic White
Income	<200% federal poverty level	≥500% federal poverty level

For the income index, advantage was defined as all adults earning ≥500% of the federal poverty level[1] (FPL) and disadvantage was defined as individuals earning <200% FPL. Prior studies calculated an index based on household income. We opted to develop an index for poverty based on FPL instead of household income because FPL accounts for household size and allowed us to accurately categorize experiences of large households that would appear to have high earnings based on household income alone. The FPL threshold for disadvantaged (<200% FPL) was based on the most inclusive eligibility criteria for social services in California. Programs such as Women, Infants and Children (WIC) program resources, Medicaid (Medi-Cal), and CalFresh have differing income thresholds for eligibility, ranging from 100% to 200% FPL [12-14]. CalFresh applies the most inclusive eligibility criteria, <200% FPL based on gross income. Due to the range of income eligibility criteria and varying thresholds used in research, we compared the selected threshold with a more conservative threshold (<125% FPL). The latter resulted in 11% of census tracts polarized toward poverty compared to 49% based on the more liberal threshold (<200% FPL). Table 2 shows the number and percentage of census tracts in each category of the poverty index based on the two different thresholds of FPL. We opted to apply the more liberal threshold (<200% FPL) to calculate our polarization measure since the conservative threshold (<125% FPL) would exclude a large number of residents who would be eligible for social services, which intend to support “low-income” individuals and families. Excluding these residents from our definition of disadvantaged may not accurately capture the lived experiences of Californians earning ≥125% FPL and <200%FPL.

[1] In 2019, the federal poverty threshold for an individual was an annual income of $14,580. (https://aspe.hhs.gov/topics/poverty-economic-mobility/poverty-guidelines)

Table 2: Comparison of census tract poverty index values based on different thresholds of the federal poverty level (FPL) to define disadvantage (N = 7,958 CTs)

Poverty Index	<200% FPL N (%)	<125% FPL N (%)
-1 ≤ index < -0.5	814 (10%)	100 (1%)
-0.5 ≤ index < 0	3,143 (39%)	2,772 (10%)
index = 0	0	0
0 < index ≤ 0.5	3,051 (38%)	3,869 (48%)
0.5 < index ≤ 1	950 (12%)	1,217 (15%)

Indices compared to proportion measures

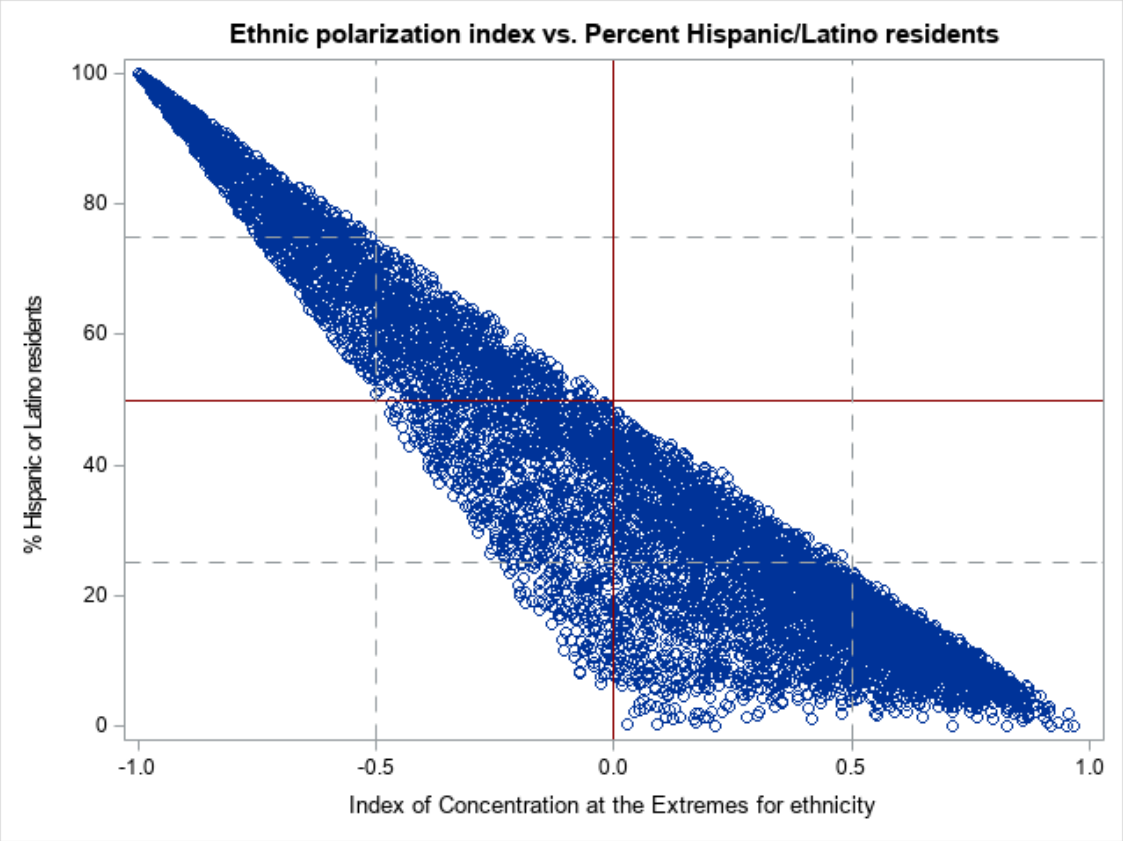
We compared each of the indices described above to their respective proportion measures. That is, we compared the ethnicity index with the proportion of residents who identified as Hispanic or Latino, race index with the proportion of residents who identified as Black or African American, and poverty index with the proportion of residents who earned <200% FPL. First, we visually examined scatter plots, with the index on the x-axis and the proportion on the y-axis, for outliers and deviation from the expected inverse relationship between index and proportion measures. As the proportion of the disadvantaged population decreases, we expected the index value to increase.

Next, we assessed associations of the indices with years of life lost (YLL) using linear regression models, where each index was the independent variable and YLL rate per 100,000 population was the dependent variable. For these analyses, we categorized the indices by quintile, similar to prior research on the index. Quintile 1 included census tracts with the lowest index values or greatest polarization toward disadvantage and Quintile 5 included census tracts with the largest index values or greatest polarization toward advantage. We compared these associations and model fit with associations of proportion measures with YLL rate. We selected years of life lost rate per 100,000 population as an appropriate outcome because it accurately reflects disparities in life expectancy and due to the availability of YLL data at the census tract level [15, 16]. YLL rate data from 2017 to 2021 were downloaded from the California Community Burden of Disease Engine website [17]. Across all models, YLL rate was included as a continuous variable. We ran bivariate models followed by models adjusted for sex, race, and urbanicity. Sex and race was downloaded from the ACS 5-year estimates in Table DP05. Urbanicity data was downloaded from the 2010 Decennial Census, which defined urban areas as those with a population less than 2,500 and at least 1,500 people residing outside institutional group quarters, and all other areas as rural areas [18].

Results

Proportion versus index scatter plots

Figure 1: Ethnicity polarization index compared to proportion of Hispanic/Latino residents per census tract (N = 8,057)



Based on ACS 5-year estimates (Table DP05), 2015-2019.

Figure 1 shows the distribution of the index for ethnicity against the proportion of residents who identified as Hispanic or Latino in each census tract. If the proportion and index were describing the population in a similar manner, we would expect to see a diagonal line from the top left to the bottom right. As the proportion of Hispanic or Latino residents decreases, we see increasing range in the index for ethnicity values. This indicates that census tracts with a small proportion of Hispanic or Latino residents are not necessarily polarized toward advantage or with residents who identify as non-Hispanic white only. For instance, census tracts with <25% Hispanic or Latino residents, range in index values from -0.25 to 1. Therefore, the index accounts for residents who neither meet the criteria for advantage nor disadvantage. If we used the proportion measure alone to categorize census tracts, we might group all census tracts with small proportions of Hispanic or Latino residents as advantaged, assuming that these census tracts are made up of largely non-Hispanic white residents. Table 3 shows the distribution of race and ethnicity among census tracts with less than 25% residents identifying as Hispanic or Latino. Census tracts with index values less than 0, indicating polarization toward disadvantage, have a large proportion of residents who identify as Asian (mean = 53.1%). As the index value increases, indicating polarization toward advantage, we see the proportion of Asian residents become smaller.

Table 3: Racial and ethnic distribution in census tracts with <25% Hispanic or Latino residents, by ethnic polarization index category

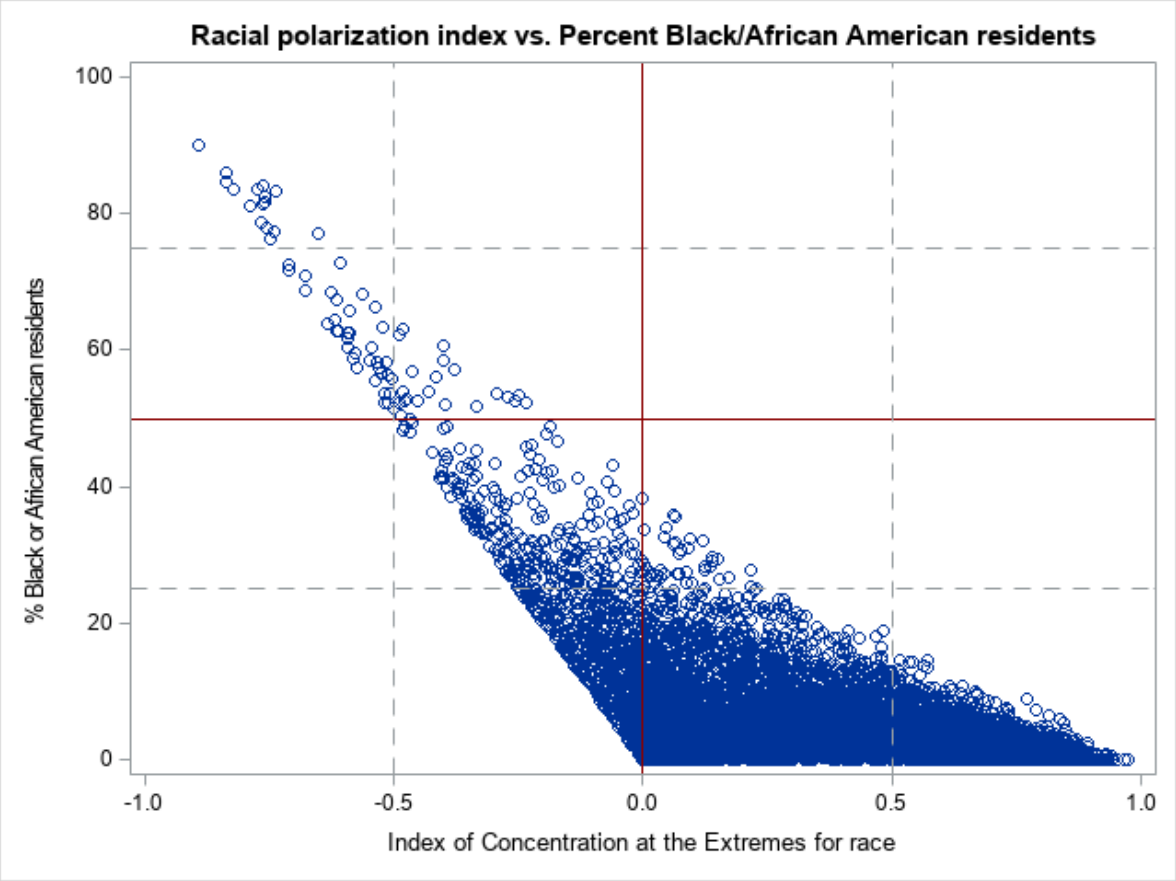
	index < 0 (N = 168 CT)		0 < index ≤ 0.5 (N = 1,633 CT)		index > 0.5 (N = 1,528 CT)	
Race and Ethnicity*	**Mean %**	**Median %**	**Mean %**	**Median %**	**Mean %**	**Median %**
White	11	11	48	51	76	76
Black or African American	18	6	7	5	3	2
AIAN	1	1	2	1	2	1
Asian	53	61	29	25	10	8
NHOPI	2	1	1	1	1	0
Some other race	10	9	6	5	3	2
Hispanic/Latino	18	19	16	17	11	11

CT = census tracts

*White includes individuals who identify as non-Hispanic white alone. All other race categories are inclusive of multiracial individuals ([race] alone or in combination with other races. Hispanic or Latino includes individuals identifying with any race.

In Figure 2, we compared the racial polarization index and the proportion of residents who identified as Black or African American by census tract. Here, we see 95% (n = 7,635) of census tracts have less than 25% of residents who identified as Black or African American. This reflects the relatively small population of Black or African American residents in the state. However, similar to ethnic polarization, we also observe a wide range of racial polarization index values in census tracts that appear to have a small proportion of Black or African American residents, which is again accounting for residents of other racial identities in these census tracts.

Figure 2: Racial polarization index compared to proportion of Black or African American residents per census tract (N = 8,057)



Based on ACS 5-year estimates (Table DP05), 2015-2019.

Census tracts with less than 25% of residents identifying as Black or African American and racial polarization index values less than zero, are made up of higher proportions of residents who identified as some other race or Hispanic or Latino ethnicity compared to census tracts with racial polarization index values closer to 1 (Table 4).

Table 4: Racial and ethnic distribution in census tracts with <25% Black or African American residents, by racial polarization index category

	index < 0 (N = 670 CT)		0 < index ≤ 0.5 (N = 4,412 CT)		index > 0.5 (N = 2,530 CT)	
Race and Ethnicity*	**Mean %**	**Median %**	**Mean %**	**Median %**	**Mean %**	**Median %**
White	7	5	28	28	70	69
Black or African American	13	13	6	4	3	2
AIAN	2	1	2	1	2	1

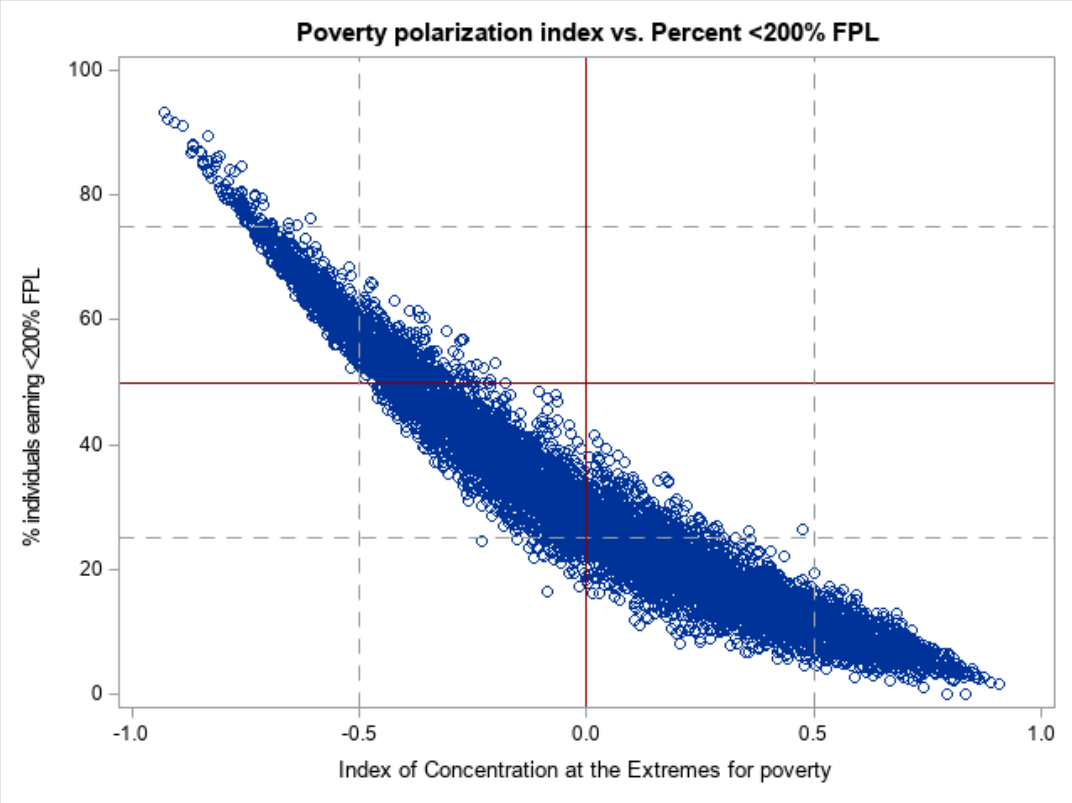
Asian	13	6	10	14	11	9
NHOPI	1	0	1	0	1	0
Some other race	31	31	18	14	4	3
Hispanic/Latino	68	71	46	44	16	11

CT = census tracts

*White includes individuals who identify as non-Hispanic white alone. All other race categories are inclusive of multiracial individuals ([race] alone or in combination with other races). Hispanic or Latino includes individuals identifying with any race.

In Figure 3, we compared the poverty index and the proportion of residents earning <200% FPL. Here, we see a slight deviation from the expected diagonal, but no obvious outliers that required investigation.

Figure 3: Economic polarization index compared to proportion of individuals earning <200% Federal Poverty Level (FPL) (N = 8,057)



Based on ACS 5-year estimates (Table S1701), 2015-2019.

Next, we examined associations of polarization indices and YLL rate at the census tract level. Prior research consistently reported the worst health outcomes in census tracts polarized toward racial and economic disadvantage [4, 7, 9]. For the purposes of our work, we evaluated these association to confirm that the associations are in the expected direction for California census tracts. Table 5 shows the association

of the proportion of residents who identified as Hispanic or Latino and YLL rate as well as the association of the ethnic polarization index and YLL rate. In the former model, the referent quintile (Q1) included census tracts with the lowest proportions of Hispanic or Latino residents. Census tracts in all comparison quintiles had significantly greater YLL rates compared to census tracts with the smallest Hispanic or Latino populations. Furthermore, as the proportion of Hispanic or Latino residents increased, the effect estimate increased, indicating greater YLL rate in these census tracts. In the latter model, the referent group (Q5) included census tracts with the largest index values or strongest polarization toward advantage. Census tracts with the lowest index values (Q1) had the largest effect estimate or the greatest difference in YLL rate compared to census tracts that are polarized toward advantage. Therefore, the associations resulting from the proportion measure and the ethnic polarization index do align. We observed similar patterns when we compared models with proportion of residents earning <200% FPL as the independent variable and economic polarization index as the independent variable (Table 6).

However, the association of the proportion of residents who identified as Black or African American with YLL rate was the opposite direction than we expected. Here, we saw that census tracts in quintiles with larger proportions of Black or African American residents (Q2 – Q5) than census tracts with the smallest proportions of Black or African American residents (Q1) had lower YLL rate per 100,000 population, as indicated by the negative effect estimates. The outcome, YLL rate is based on years of life lost across the entire population in each census tract, regardless of race and ethnicity. Therefore, the negative effect estimates may be due to small populations of Black or African American residents compared to residents who identify with other race and ethnicity groups. In comparison, when the racial polarization index is the independent variable, we observed expected associations that reflect widely acknowledged racial health disparities. Census tracts that were more polarized toward disadvantage, according to the racial polarization index, had higher YLL rates compared to census tracts polarized toward advantage.

Furthermore, when we compared model fit of models built using a proportion measure as the independent variable with models built using an index measure as the independent variable, we saw that the models with index measures as independent variables consistently had lower AIC values. Thus, the indices for race, ethnicity, and poverty had similar associations with YLL rate as the more commonly used proportion measures and resulted in improved model fit.

Table 5: Multivariable associations of proportion of Hispanic/Latino residents and ethnic polarization index with years of life lost (YLL) rate

	β (95% CI)	p-value	AIC
Proportion Hispanic/Latino population		< .0001	10355615
Q2 vs. Q1	0.208 (0.207, 0.209)		
Q3 vs. Q1	0.280 (0.279, 0.281)		
Q4 vs. Q1	0.420 (0.419, 0.421)		

Q5 vs. Q1	0.325 (0.323, 0.326)		
Index, Ethnicity		< .0001	6738249
Q1 vs. Q5	0.456 (0.454, 0.457)		
Q2 vs. Q5	0.450 (0.449, 0.452)		
Q3 vs. Q5	0.317 (0.315, 0.318)		
Q4 vs. Q5	0.142 (0.141, 0.144)		

*Models are adjusted for sex (male, female), race (White, Black or African American, American Indian and Alaska Native, Asian, Native Hawaiian and Other Pacific Islander, and some other race), and urbanicity (urban, rural).
Referent quintile for the polarization measure includes census tracts with the lowest proportion of Hispanic/Latino residents. Referent quintile for the index includes census tracts with the strongest polarization toward advantage.

Table 6: Multivariable associations of the proportion of residents earning <200% FPL and the economic polarization index with years of life lost (YLL) rate

	β (95% CI)	p-value	AIC
Proportion <200% FPL		< .0001	9277707
Q2 vs. Q1	0.126 (0.125, 0.127)		
Q3 vs. Q1	0.290 (0.289, 0.291)		
Q4 vs. Q1	0.407 (0.406, 0.408)		
Q5 vs. Q1	0.532 (0.531, 0.534)		
Index, Poverty		< .0001	4723251
Q1 vs. Q5	0.852 (0.850, 0.853)		
Q2 vs. Q5	0.686 (0.685, 0.687)		
Q3 vs. Q5	0.528 (0.527, 0.529)		
Q4 vs. Q5	0.288 (0.287, 0.289)		

*Models are adjusted for sex, race and ethnicity, and urbanicity.
**Referent quintile for the proportion measure includes census tracts with the lowest proportion of residents earning <200% FPL. Referent quintile for the index includes census tracts with the strongest polarization toward advantage.

Table 7: Multivariable associations of the proportion of Black or African American residents and the racial polarization index with years of life lost (YLL) rate

	β (95% CI)	p-value	AIC
Proportion African American or Black population		< .0001	106142767
Q2 vs. Q1	-0.083 (-0.084, -0.082)		
Q3 vs. Q1	-0.182 (-0.183, 0.181)		
Q4 vs. Q1	-0.122 (-0.124, -0.121)		
Q5 vs. Q1	-0.077 (-0.078, -0.076)		
Index, Race		< .0001	6815860
Q1 vs. Q5	0.316 (0.313, 0.319)		
Q2 vs. Q5	0.293 (0.290, 0.295)		
Q3 vs. Q5	0.222 (0.221, 0.224)		

Q4 vs. Q5	0.107 (0.106, 0.108)

*Models are adjusted for sex (male, female), race (White, Black or African American, American Indian and Alaska Native, Asian, Native Hawaiian and Other Pacific Islander, and some other race), and urbanicity (urban, rural).
**Referent quintile for the proportion measure includes census tracts with the lowest proportion of Black or African American residents. Referent quintile for the index includes census tracts with the strongest polarization toward advantage.

Summary of polarization indices in California

State-level

For the state of California as a whole, the ethnic polarization index value was 0.02, racial polarization index value was 0.30, and the economic polarization index value was 0.01. Polarization estimates for California were all close to zero, suggesting very little or no polarization in the state. However, as we calculated index values for smaller geographic areas, we began to discover areas polarized toward advantage and disadvantage.

County-level

Table 8 shows the distribution of index values for California counties. On average, a county's polarization index for ethnicity was 0.23, however the majority of counties had ethnic polarization index values close to zero (IQR: -0.04 – 0.56). Similarly, the economic polarization index suggested very little polarization at the county level (mean: -0.05; IQR: -0.20 – 0.09). The racial polarization index indicated that most counties in California had concentrations of non-Hispanic white residents and no county had a concentration of Black of African American residents (mean: 0.50; IQR: 0.30 – 0.69).

Table 8: Distribution of polarization indices for California counties

Index	N	Mean (SD)	Median	IQR	Min-Max
Ethnicity	58	0.23 (0.37)	0.23	-0.04 - 0.56	-0.74 - 0.77
Race	58	0.50 (0.22)	0.5	0.30 - 0.69	0.07 - 0.87
Poverty	58	-0.05 (0.20)	-0.07	-0.20 - 0.09	-0.35 - 0.43

*IQR = interquartile range

Census tract-level

We observed greatest variance in index values at the census tract level. Table 9 shows the distribution of census tract level polarization indices. At the most granular geographic level, we identified areas where almost 100% of the population met the criteria for disadvantage or advantage for each index, that is, all three indices ranged from -1 to 1, approximately.

Table 9: Distribution of polarization indices for California census tracts

Index	N	Mean (SD)	Median	IQR	Min-Max
Ethnicity	7992	0.01 (0.49)	0.06	-0.39 - 0.31	-0.93 - 0.91

Race	7992	0.32 (0.30)	0.30	-0.07 - 0.57	-1.00 - 0.97
Poverty	7958	0.01 (0.38)	0.004	-0.30 - 0.31	-0.93 - 0.91

*IQR = interquartile range

Based on these findings of the indices, our Polarization Dashboard presents data at the census tract level with county-level data provided for comparison.

Strengths & Limitations

We have highlighted the following strengths of the index: the index quantifies the degree of polarization toward either advantage or disadvantage as opposed to describing only the size of the disadvantaged population in a given area; the index captures social polarization in areas where the proportion of residents in the disadvantaged group may be small by accounting for residents who neither meet the criteria for advantage nor disadvantage; the index mirrors associations of parallel proportion measures and health outcomes such as YLL rate, which supports use of the index as a measure that can inform public health intervention; and the index is a tool to monitor social polarization over time.

Although informative, there are limitations to the index of concentration at the extremes. First, there are other sociodemographic characteristics by which residents may be concentrated, such as language and age. We selected three characteristics – income, race, and ethnicity – which are referenced in existing literature on the index of concentration at the extremes. Other characteristics may be considered for future indices. Similarly, we acknowledge that other groups, who are not specifically identified as “disadvantaged” in our indices, also experience systemic disadvantage. In our future work, we plan to explore spatial social polarization of other racial and ethnic groups in California.

References

1. Caves, R.W., Encyclopedia of the City. 2005: Routledge.
2. Menendian, S.G., Samir, Twenty-First Century Racial Residential Segregation in the United States, in The Roots of Structural Racism Project. 2021.
3. Adkins-Jackson, P.B., et al., Measuring Structural Racism: A guide for epidemiologists and other health researchers. Am J Epidemiol, 2021.
4. Krieger, N., et al., Using the Index of Concentration at the Extremes at multiple geographical levels to monitor health inequities in an era of growing spatial social polarization: Massachusetts, USA (2010-14). Int J Epidemiol, 2018. 47(3): p. 788-819.
5. The Digital Scholarship Lab and the National Community Reinvestment Coalition. Not Even Past: Social Vulnerability and the Legacy of Redlining. Available from: https://dsl.richmond.edu/socialvulnerability.
6. Massey DS, The Prodigal Paradigm Returns: Ecology Comes Back to Sociology. In: Booth A Crouter A editors. Does It Take a Village? Community Effects on Children, Adolescents, and Families. 2000, Psychology Press: Florence, KY.
7. Krieger, N., et al., Structural Racism, Historical Redlining, and Risk of Preterm Birth in New York City, 2013-2017. Am J Public Health, 2020. 110(7): p. 1046-1053.
8. Krieger, N., et al., Measures of Local Segregation for Monitoring Health Inequities by Local Health Departments. Am J Public Health, 2017. 107(6): p. 903-906.
9. Krieger, N., et al., Public Health Monitoring of Privilege and Deprivation With the Index of Concentration at the Extremes. Am J Public Health, 2016. 106(2): p. 256-63.
10. Feldman, J.M., et al., Police-Related Deaths and Neighborhood Economic and Racial/Ethnic Polarization, United States, 2015-2016. Am J Public Health, 2019. 109(3): p. 458-464.
11. Snyder, C.R., Navigating in murky waters: How multiracial Black individuals cope with racism. Am J Orthopsychiatry, 2016. 86(3): p. 265-76.
12. Department of Healthcare Services. Do you Quality for Medi-Cal Benefits? 2022; Available from: https://www.dhcs.ca.gov/services/medi-cal/Pages/DoYouQualifyForMedi-Cal.aspx.
13. Public Health Foundation Enterprise WIC. WIC Guidelines and Qualifications. 2023; Available from: https://www.phfewic.org/how-wic-works/who-qualifies/.
14. California Department of Social Services. Eligibility and Issuance Requirements. 2022; Available from: https://www.cdss.ca.gov/inforesources/cdss-programs/calfresh/eligibility-and-issuance-requirements#income.
15. Amiri, S., et al., Disparities in years of potential life lost among racial and ethnic groups in Washington state. Arch Public Health, 2022. 80(1): p. 211.
16. Martinez, R., et al., Reflection on modern methods: years of life lost due to premature mortality-a versatile and comprehensive measure for monitoring non-communicable disease mortality. Int J Epidemiol, 2019. 48(4): p. 1367-1376.
17. California Department of Public Health. California Community Burden of Disease Engine. 2023; Available from: https://skylab.cdph.ca.gov/communityBurden/.
18. United States Census Bureau. 2010 Census Urban and Rural Classification and Urban Area Criteria. 2021; Available from: https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/2010-urban-rural.html#:~:text=The%20Census%20Bureau%20identifies%20two,and%20less%20than%2050%2C000%20people.