

Environmental Justice Methodological Review

Environmental Protection Agency

The Environmental Protection Agency (EPA) has developed and employs EPA EJSCREEN, an environmental justice mapping and screening tool that combines publicly available environmental data and demographic indicators for display and analysis. The tool began development in 2010, and the agency released the first public version in 2015 (Environmental Protection Agency, 2016). The metadata states that the agency's goal was to update the tool annually; however, at the time of this research, the most current version available was Version 2018.

The tool maps and analyzes eleven spatial datasets of environmental indicators measuring air quality, automobile traffic, housing stock age as a proxy for lead paint exposure, wastewater discharges, and proximity to risk management plan sites (RMP), hazardous waste facilities, and national priorities list sites (NPL) (Environmental Protection Agency, 2018).

To assess local populations' susceptibility to the mapped environmental hazards, they mapped six demographics from the 2016 American Community Survey 5-Year Estimates at the census block group level (Environmental Protection Agency, 2016).

1. Percent low income: household income is less than or equal to twice the federal poverty level
2. Percent minority: individuals who list their racial status as not white alone and/or list their ethnicity as Hispanic or Latino
3. Less than high school education: percent of individuals aged 25 or older whose education is short of a high school diploma
4. Linguistic isolation: percent of people living in linguistically isolated households defined by those who speak a non-English language and speak English less than "very well"
5. Individuals under age 5: percent of individuals under the age of 5
6. Individuals over age 64: percent of individuals older than 64

Each of the six demographic indicators can be viewed separately in the web application or in generated reports. In addition, the application produces a composite demographic index that synthesizes multiple demographic indicators. The index uses two demographic indicators explicitly named in Executive Order 12898: low income and racial minority status. The demographic index is a straightforward average of the percentages of racial minorities and low income, using the formula: $(\% \text{ minority} + \% \text{ low-income})/2$ (Environmental Protection Agency, 2017). Although combining income and racial demographics provides more information than examining either indicator alone, the index excludes additional factors such as age, education, and language isolation that may help identify socially vulnerable populations.

Centers for Disease Control

The Centers for Disease Control and Prevention (CDC) has produced a social vulnerability index to help emergency response planners and public health officials identify communities that may need additional support in the event of a disaster (Centers for Disease Control and Prevention, 2019). The CDC views social vulnerability as a key factor in overall disaster risk. Identical hazards can affect different populations in predictable ways that can and should guide resource allocation (Flanagan, Gregory, Hallisey, Heitgerd, & Lewis, 2011).

At the time of this research, the most recent CDC index used 2016 American Community Survey 5-Year Estimates demographics mapped at the census tract level and aggregated at the county level. The analysis grouped specific demographic indicators into one of four vulnerability themes: socioeconomic status, household composition & disability, minority status & language, and housing & transportation. The index scores each theme based on multiple ACS demographics and produces an overall vulnerability score by summing the individual theme scores.

For each census tract, the CDC generated percentile rankings for 15 demographic variables, ranging from 0.0 to 1.0, with higher values indicating greater vulnerability. The individual percentile rankings were combined to produce scores for four vulnerability themes and an overall Social Vulnerability Index.

Below is an outline of the demographic themes as well as the individual demographic indicators used to score the themes and overall social vulnerability.

1. Socioeconomic Status
 - a. Below Poverty: percentage of individuals with incomes below the poverty level
 - b. Unemployed: percentage of civilians over 16 who are unemployed
 - c. Income: per capita income by census tract
 - d. No High School Diploma: percentage of people aged 25 or older with no high school diploma
2. Household Composition & Disability
 - a. Aged 65 or Older: percentage of elderly individuals
 - b. Aged 17 or Younger: percentage of children
 - c. Civilian with a Disability: percentage of civilians in the noninstitutionalized population with a disability
 - d. Single-Parent Households: percentage of households headed by a single parent with children under 18
3. Minority Status & Language
 - a. Minority: percentage of individuals who are not white, non-Hispanic
 - b. Speak English “Less than Well”: percentage of individuals aged 5 or older who speak English “less than well”
4. Housing & Transportation
 - a. Multi-Unit Structures: percentage of housing units in multi-unit structures of 10 or more
 - b. Mobile Homes: percentage of housing units classified as mobile homes
 - c. Crowding: percentage of households with more people than rooms
 - d. No Vehicle: percentage of households with no vehicle available
 - e. Group Quarters: percentage of people who reside in institutionalized group quarters

The CDC also developed a supplementary binary flagging system to identify census tracts with particularly high values for individual vulnerability indicators. Variables at or above the 90th percentile received a flag of 1, while those below that threshold received 0. These flags can then be summed to identify the number of indicators for which a census tract exhibits particularly high vulnerability.

Although this supplementary flagging method helps identify extreme values, it does not distinguish between demographic conditions below the 90th percentile or the degree to which a value exceeds that threshold. My research used a different approach, applying Jenks natural breaks classification to assign graduated vulnerability scores based on the statewide distribution of individual demographic indicators. This allowed the scoring system to account for varying degrees of vulnerability within Minnesota's census block groups while also incorporating selected demographic weights.

MnDOT Past Precedents

As part of the 20-Year State Highway Investment Plan 2018-2037 (Minnesota Department of Transportation, 2017), there was an environmental justice analysis as an appendix (Minnesota Department of Transportation, 2017) to the main report. The analysis relied on statewide census block group demographics from the United States Census Bureau 2014 American Community Survey 5-Year Estimates (2010-2014). The analysis focused on identifying environmental justice populations based on high concentrations of racial minorities, Hispanics, households in poverty, individuals with limited English skills, individuals 17 years of age or younger, individuals 65 years of age or older, and households without access to their own automobile.

The analysis mapped each of those seven demographics statewide as choropleth maps, showing whether each census block group exceeded the Minnesota average for the indicator in question. It mapped high concentrations of demographic indicators within census block groups in darker shades to denote magnitudes above the statewide averages.

To synthesize the indicators, the analysis produced a choropleth map by overlaying census block groups with any above-state-average environmental justice population in any of the seven American Community Survey demographics. The overlay dataset was then mapped to show the intensity of multiple environmental justice populations within census block groups, ranging from zero to four, across the state of Minnesota.

While this earlier analysis provided considerable value, I identified opportunities to improve its methodology and incorporated them into my research. First, using statistical classification methods such as Jenks natural breaks, rather than simply measuring how far demographic values exceeded statewide averages, allowed for a more nuanced analysis of varying data distributions. Second, the previous composite index considered only how many demographic indicators exceeded statewide averages. My approach introduced a more sophisticated scoring system that accounted for both the number of indicators and the degree to which individual demographic values reflected social vulnerability.

Social Vulnerability Index (SoVI) Creation

To create the social vulnerability index, I classified individual demographic values for each census block group within the full statistical distribution of the statewide database of 4,107 block groups. All statistical classification of demographic datasets was conducted using the Jenks natural breaks algorithm in Esri's ArcMap 10.5.1 GIS software. Jenks natural breaks determines the optimal data classification that minimizes variance within classification subsets, while maximizing the variance between classed data subsets. It was deemed the best classification option for handling varying skew factors in the demographic data while creating the most pertinent data groupings.

Most demographic attributes used to compile the SoVI were right-skewed to varying degrees. In most cases, census block groups clustered around relatively low demographic percentages, while a much smaller number exhibited unusually high values. These right-skewed distributions made Jenks natural breaks particularly useful for distinguishing degrees of social vulnerability.

For the thirteen demographic percentage indicators, I used seven Jenks natural breaks classes, assigning the three lowest classes a score of zero. For most indicators, classes four through seven received progressively higher vulnerability scores of 1, 2, 3, and 4. Four individual racial demographic indicators (Black, Native American, Asian, and Hispanic) received lower weights, with corresponding scores of 0.25, 0.50, 0.75, and 1.0. I assigned lower weights to these four indicators to reflect the additional socioeconomic vulnerabilities associated with the historical and structural marginalization of specific racial and ethnic communities, while ensuring their combined maximum score would not exceed the broader nonwhite population indicator's maximum score. I then summed these scores across all fourteen demographic indicators to create a composite SoVI score for each census block group.

Scoring per capita income for the SoVI required a different strategy than for the thirteen demographic percentages, whose high values were considered indicators of vulnerability. Like population demographics, per capita income was right-skewed with high-value outliers, but unlike population demographics, low per capita incomes signaled social vulnerability and needed to be identified and scored. When I applied a seven-category Jenks classification to per capita income values for the full study area of census block groups, the four most socially vulnerable categories were \$45,203 and under and included 3,457 (84%) of the 4,107 census block groups. All other scoring metrics assigned 0 values to at least 50% of census block groups, so to identify high social-vulnerability outliers, it was best to keep that basic approach for scoring individual indicators. Therefore, I increased the number of classes for the Jenks algorithm for per capita income until fewer than 50% of census block groups fell into the lowest four categories. This was achieved with a twelve-class Jenks grouping, where the cutoff for a non-zero score was per capita income below \$30,094, which 44.3% of census block groups fell into.

Table 2: SoVI Data Scoring Bounds for Individual Demographics

Demographic	Mean	Range: Mid	Range: High	Range: Higher	Range: Highest	N: Mid	N: High	N: Higher	N: Highest	Vul Total	% CBGs
Poverty	0.077	.135-.222	.223-.347	.348-.532	.541-1.0	306	249	87	41	683	16.6%
Per Capita Income	34,206	25584-30093	20834-25558	15266-20808	4217-15191	804	590	290	137	1,821	44.3%
Race: Non-white	0.155	.226-.348	.352-.502	.503-.689	.692-1.0	406	236	169	120	931	22.7%
Race: Black	0.057	.152-.244	.245-.367	.369-.538	.545-.930	235	137	79	49	500	12.2%
Race: Native	0.011	.093-.185	.201-.323	.348-.547	.608-.978	48	18	17	6	89	2.2%
Race: Asian	0.041	.103-.169	.170-.261	.266-.401	.408-.733	226	129	80	39	474	11.5%
Race: Hispanic	0.052	.117-.196	.197-.309	.312-.456	.466-.730	248	154	59	31	492	12.0%
Age: Under 5	0.061	.058-.078	.078-.104	.105-.149	.150-.301	915	701	351	69	2036	49.6%
Age: Over 64	0.158	.155-.205	.205-.271	.271-.373	.375-.672	862	644	298	81	1885	45.9%
No HS Diploma	0.078	.101-.153	.154-.235	.236-.364	.365-.626	579	303	138	48	1068	26.0%
No Veh Access HH	0.068	.098-.162	.163-.257	.258-.405	.414-.793	464	271	137	62	934	22.7%
ESL HH	0.023	.079-.133	.135-.220	.224-.345	.370-.675	207	128	44	16	395	9.6%
Single Parent HH	0.293	.289-.412	.413-.564	.565-.775	.776-1	671	506	372	166	1715	41.8%
Disability Status HH	0.223	.217-.277	.278-.351	.352-.457	.458-.768	971	657	317	80	2025	49.3%

Note: Demographic values are expressed as proportions, while per capita income is reported in dollars.

Table 2 details the data breaks used for index scoring and how many census block groups fell into vulnerable categories for the fourteen demographic indicators that were components of the SoVI. The first data column shows the statewide mean value of each demographic distribution. The next four range columns show the data bounds that classify individual census block group values into increasingly severe vulnerability categories. The next four N columns show the number of census block groups in each of the four vulnerability categories for each demographic, followed by a column that sums the total number of census block groups considered statistically vulnerable for each variable. The final column gives the percentage of the 4,107 statewide census block groups that scored as vulnerable for each demographic indicator.

As shown in the final column of Table 2, the proportion of census block groups classified as vulnerable in a single demographic varied widely depending on the indicator's statewide distribution. On the common end of vulnerability were indicators of children under 5 and households with a disabled member that were just below a majority at 49+%. The least frequently flagged demographic indicator was the proportion of Native American residents, with only 89 (2.2%) of Minnesota's 4,107 census block groups receiving a nonzero score for this measure. This reflects the relatively small statewide population share of Native Americans and their geographic concentration in particular communities. Nevertheless, the composite SoVI captured vulnerability in these communities through additional demographic and socioeconomic measures. Notably, the six highest-scoring rural census block groups all had majority Native American populations, with their elevated vulnerability scores reflecting multiple indicators beyond racial demographics alone.