Using the Product of Ranks to Calculate Disadvantage from CalEnviroScreen 2.0 Data

Bay Area Air Quality Management District
Contact: David Holstius, Ph.D. (dholstius@baaqmd.gov)

This document describes an alternative to Methods 1–5[1], called the "product of ranks", that increases the representation of tracts having high scores on a subset of indicators. Figures and maps illustrate its application to the 19 indicators defined by CalEnviroScreen 2.0. Overall scores are equal to **Pollution Burden × Population Characteristics**, defined as:

Pollution = Ozone × PM2.5 × DieselPM × DrinkWat × Pesticides × TRI × Traffic × Cleanup × GndWat × HazWaste × ImpWat × SolidWaste

Population = Age × LBW × Asthma × Education × LingIso × Poverty × Unemp

As with Methods 1–5, various thresholds may be applied to the overall scores to identify disadvantaged communities (Figs 1–3).

Figure 1. Using the product of ranks to identify disadvantaged census tracts

Figure 1 Legend

- Red dots are in the top 15% of cumulative scores
- Green dots have cumulative scores in the top 15 to 20%
- Orange dots have cumulative scores in the top 20 to 25%

BAY AREA AIR QUALITY MANAGEMENT DISTRICT

As in Methods 1–5, raw values are first transformed to percentiles (by ranking and dividing by N), with the highest rank denoted by 1 rather than N. Thus, the top score for a given indicator will be 1/N. When a tract is scored near the top on many indicators, the contribution of these small fractions to the overall product will drive its final score up. Figures 1 and 2 illustrate that census tracts with very high Pollution scores are always identified as disadvantaged, regardless of Population score—and vice versa. Intermediate cases are smoothly interpolated.[2,3]

Advantages of this method include:

- It increases the representation of tracts with top scores on a few indicators;
- It has been recommended by scientists for screening large datasets[4]; and
- It is simple and straightforward to calculate.

Figure 2. Identification of disadvantaged census tracts, by region

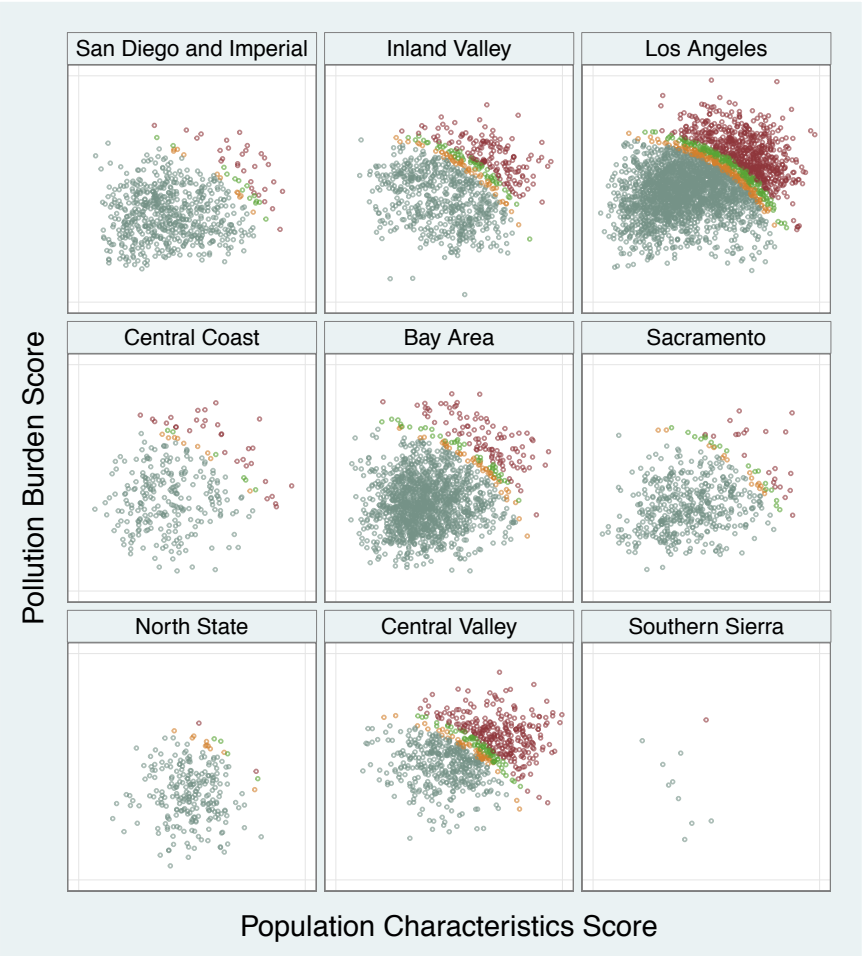


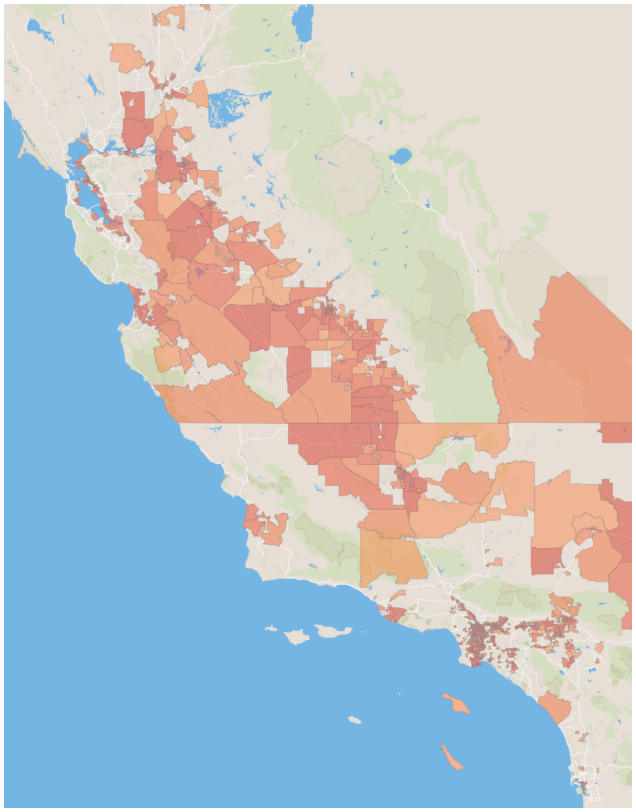
Figure 2 Legend

- Red dots are in the top 15% of cumulative scores
- Green dots have cumulative scores in the top 15 to 20%
- Orange dots have cumulative scores in the top 20 to 25%

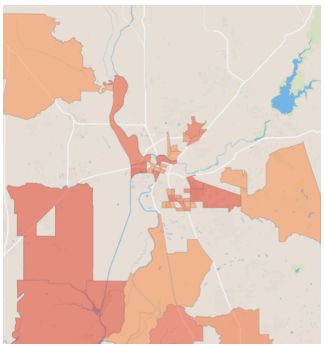
BAY AREA AIR QUALITY MANAGEMENT DISTRICT

Figure 3. Census tracts scoring in the top 5% (red) to 20% (orange)

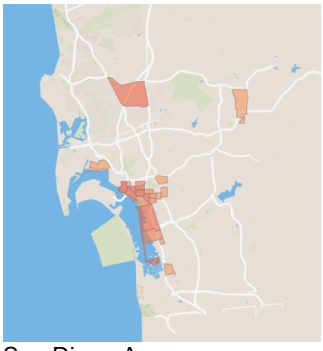
(Interactive online version: http://tiny.cc/CES-Method6)



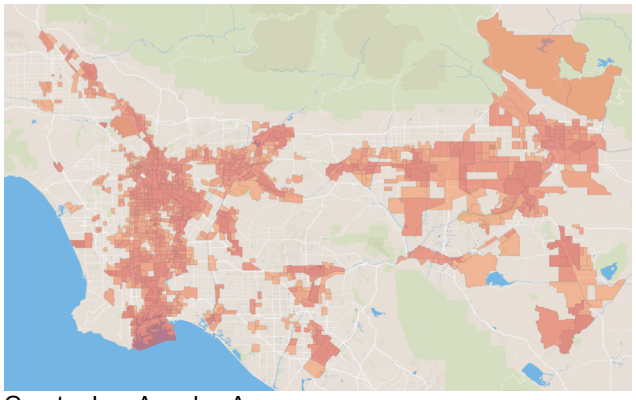
Central and Southern California



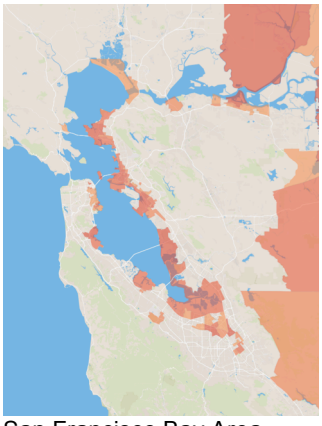
Sacramento Area



San Diego Area



Greater Los Angeles Area



San Francisco Bay Area

[1] OEHHA. Approaches to Identifying Disadvantaged Communities. http://oehha.ca.gov/ej/ces2.html, August 2014.

[2] The likelihood of being ranked at the top of one list purely by chance is 1/N. The relative likelihood of having 19 rankings equal to or greater than a tract's actual rankings can be compared to the product of its 19 scores.

[3] Missing data can be accommodated via the application of geometric means, analogous to the arithmetic means used in Methods 1–5. Further details are available in Breitling (2004, above). See http://rpubs.com/holstius/CES-Method6.

[4] Breitling R, Armengaud P, Amtmann A, Herzyk P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.* 2004 Aug 27;573(1-3):83-92.