EJScreen

Environmental Justice Mapping and Screening Tool

EJScreen Technical Documentation

October 2022

U.S. Environmental Protection Agency
Office of Environmental Justice and External Civil Rights
Washington, D.C. 20460

Suggested citation:

U.S. Environmental Protection Agency (EPA), 2022. EJScreen Technical Documentation.

For more information:
www.epa.gov/EJScreen

Contents

1 Introduction

EJScreen is EPA's environmental justice (EJ) screening and mapping tool that utilizes standard and nationally-consistent data to highlight places that may have higher environmental burdens and vulnerable populations. The tool offers EJ indexes by combining environmental and demographic indicators in basic geographic units of Census block groups. The tool also provides a variety of powerful data and mapping capabilities that enable users to access environmental and demographic information across the entire country, at high geographic resolution, displayed in color-coded maps and standard data reports.

This technical document provides details on the data and methods used to select the indicators and create the indexes in EJScreen. EPA annually updates EJScreen with the newest datasets available and improvements to the interface. For more information on the updates to EJScreen over time, visit the EJScreen Change Log.

Geographic Framework

- The geographic framework for EJScreen was built from 2020 Census TIGER/Line data for all 50 states, the District of Columbia, and Puerto Rico.
- The socioeconomic data source is U.S. Census Bureau's American Community Survey (ACS) 2016-2020 5-Year Estimates (ACS 2020).
- There are a total 242,335 block groups.
- The application defines the spatial reference as "WGS 1984 Web Mercator (Auxiliary Sphere)."
- The U.S. Territories of American Samoa, Commonwealth of the Northern Mariana Islands (CNMI), Guam, Puerto Rico, and the U.S. Virgin Islands are included in the application. However, Puerto Rico is included with all 50 states and the District of Columbia. The term "territories" in this document refers to the four territories excluding Puerto Rico.
- Puerto Rico uses ACS 2020. American Samoa, CNMI, and Guam use 2013 Census Place boundaries. The U.S. Virgin Islands use 2013 Census Estate boundaries. The territories data for American Samoa, CNMI, Guam, and the U.S. Virgin Islands provide additional 605 records.
- The Census Place boundaries and Census Estate boundaries for the territories are appended to the block groups for all 50 states and the District of Columbia and Puerto Rico (see the Details on U.S. Territories section for details).
- Census block centroids with population from 2020 Decennial Census P.L. 94-171 Redistricting data are used to provide population-weights for the application.

2 Overview of Data in EJScreen

This section describes the environmental and socioeconomic indicators featured in the application, describes why they are included, and the data used to derive them.

Socioeconomic Indicators in EJScreen

In the current release:

- All demographic indicators are from Census Bureau's ACS 2016-2020 5-year Summary.
- All territories' socioeconomic data are from the Census 2010 Demographic Profile Summary File for each territory, using the Place summary level for American Samoa, CNMI, and Guam, and Estates summary level for U.S. Virgin Islands. The Demographic Profiles were published in 2014 by the U.S. Census Bureau.

Summary Overview of Socioeconomic Indicators Featured in EJScreen

EJScreen uses socioeconomics indicators as very general indicators of a community's potential susceptibility to the types of environmental factors included in EJScreen. There are seven socioeconomic indicators featured in EJScreen. These indicators form the basis for both the demographic index and the supplemental demographic index:

1. **People of color:**
    - The percent of individuals in a block group who list their racial status as a race other than white alone and/or list their ethnicity as Hispanic or Latino. That is, all people other than non-Hispanic white-alone individuals. The word "alone" in this case indicates that the person is of a single race, not multiracial.
2. **Low-income:**
    - The percent of a block group's population in households where the household income is less than or equal to twice the federal "poverty level."
3. **Unemployment rate:**
    - The percent of a block group's population that did not have a job at all during the reporting period, made at least one specific active effort to find a job during the prior four weeks, and were available for work (unless temporarily ill).
4. **Limited English speaking household**:
    - A "limited English speaking household" is one in which no member 14 years old and over (1) speaks only English or (2) speaks a non-English language and speaks English "very well." In other words, all members 14 years old and over have at least some difficulty with English.
5. **Less than high school education:**
    - Percent of people age 25 or older in a block group whose education is short of a high school diploma.
6. **Under age 5:**
    - Percent of people in a block group under the age of 5.
7. **Over age 64:**
    - Percent of people in a block group over the age of 64.

Detailed Descriptions of Socioeconomic Indicators

People of Color

People of color are individuals who list their racial status as a race other than white alone and/or list their ethnicity as Hispanic or Latino. That is, all people other than non-Hispanic white-alone individuals. The word "alone" in this case indicates that the person is of a single race, not multiracial.

The U.S. Census Bureau classifies races based on the Office of Management and Budget (OMB) standards on race and ethnicity and is determined based on individuals' self-identification with one or more groups. These groups are:

- White
- Black or African American
- American Indian or Alaska Native
- Asian
- Native Hawaiian or Other Pacific Islander

Furthermore, ethnicity is defined as to whether an individual is of Hispanic origin or not. Individuals of any race can report as Hispanic.

How does EJScreen determine percent people of color?

- The ACS information on people of color is captured in the table *Hispanic or Latino Origin by Race* (ACS Table ID: B03002). The ACS divides race and Hispanic status into 21 categories, four of which are shown in Table 1.

Table 1: ACS Hispanic or Latino Origin by Race Table Element Reference

Table Element	Hispanic Status/Race Population
B03002.001	Total Population: All races/ethnicities
B03002.002	Total Population: Non-Hispanic
B03002.003	Total Population: Non-Hispanic, White Alone
B03002.012	Total Population: Hispanic

Note: The data can be downloaded from the US Census Bureau's FTP Server.

- To calculate percent people of color, two elements from Table 1 are used in the following equation:

$$\%\ People\ of\ Color = \frac{B03002.001 - B03002.003}{B03002.001}$$

- In EJScreen, the raw values for the people of color indicator range from 0% to 100%. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the people of color indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 31%. If a raw value is higher than 31%, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the

people of color indicator raw value for a block group is 74%, which is higher than 31%, that block group will be placed in the 80th percentile (which is higher than the 50th percentile).

Low-Income

Low-income is defined as a household whose income is less than or equal to twice the poverty level. For example, a household of four with a reported $40,000 total annual income is lower than twice the poverty threshold of $52,992 ($26,496 is the poverty threshold defined by the U.S. Census Bureau for 2020). This household will fall into the category of "low income" in EJScreen. The poverty level is updated by the U.S. Census Bureau annually and varies by family size and composition.

The poverty level is a national number and the same across all geographic regions. To accommodate differences in the varying costs of living across the United States and other factors, EJScreen uses twice the poverty level to capture low income households especially in high cost areas. The rationale for using twice the poverty threshold rather than just the poverty threshold includes the following considerations:

- The effects of income on baseline health and probably on other aspects of susceptibility are not limited to those below the poverty thresholds.
- Many studies in various fields use 2x poverty.
- When using twice the poverty threshold, the number or percent low income happens to roughly equal the number or percent people of color in the United States.

How does EJScreen determine percent low income?

- The ACS low income information is captured in the table *Ratio of Income to Poverty Level in the Past 12 Months* (ACS Table ID: C17002). The ACS divides the ratio of income to poverty into seven categories, as shown in Table 2.

Table 2. ACS Income/Poverty Level Table Element Reference

Table Element	Income/Poverty Level
C17002.001	Total Population Whose Poverty Status is Known
C17002.002	People with Ratio of Income to Poverty under .50
C17002.003	People with Ratio of Income to Poverty from .50 to .99
C17002.004	People with Ratio of Income to Poverty from 1.00 to 1.24
C17002.005	People with Ratio of Income to Poverty from 1.25 to 1.49
C17002.006	People with Ratio of Income to Poverty from 1.50 to 1.84
C17002.007	People with Ratio of Income to Poverty from 1.85 to 1.99
C17002.008	People with Ratio of Income to Poverty from 2.00 and over

Note: The data can be downloaded from the US Census Bureau's FTP Server.

- To calculate percent low income, two elements from Table 2 are used in the following equation:

$$\%\ Low\ Income = \frac{C17002.001 - C17002.008}{C17002.001}$$

- In EJScreen, the raw values for the low income indicator range from 0% to 100%. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the low income indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 27%. If a raw value is higher than 27%, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the low income indicator raw value for a block group is 49%, which is higher than 27%, that block group will be placed in the 80th percentile (which is higher than the 50th percentile).

Unemployment

Unemployment refers all those who did not have a job at all during the reporting period, made at least one specific active effort to find a job during the prior four weeks, and were available for work (unless temporarily ill). The U.S. Census Bureau determines unemployment through a series of questions in the ACS that establish several labor force elements.

The first parameter is whether or not an individual is age 16 or older. This is necessary because, under federal law, individuals under the age of 16 cannot fully participate in the labor force.

The second parameter is labor force participation. Individuals over the age of 16 who are not classified as members of the labor force include students, stay-at-home parents, retired workers, and others who are not actively searching for employment. An individual can only be considered unemployed if they are part of the labor force.

The last parameter is employment status. An individual is considered unemployed if they were neither "at work" nor "with a job but not at work" during the reference week (the week the survey was conducted), were actively looking for work during the last four weeks, and were available to accept a job. This also includes individuals who did not work at all during the reference week, were waiting to be called back to a job from which they had been laid off, and were available for work except for temporary illness.

How does EJScreen determine unemployment?

- The ACS unemployment information is captured in the table *Employment Status for the Population 16 Years and Over* (Table ID: B23025). The ACS divides the population age 16 or older into seven categories, which are shown in Table 3. EJScreen uses the Civilian Labor Force numbers.

Table 3: ACS Employment Status Table Element Reference

Table Element	Employment Status
B23025.001	Total Population Age 16 Years and Over
B23025.002	Total Labor Force
B23025.003	Total Civilian Labor Force
B23025.004	Civilian Labor Force: Employed
B23025.005	Civilian Labor Force: Unemployed
B23025.006	In Armed Forces

B23025.007	Total Not In Labor Force

Note: These data can be downloaded from the US Census Bureau's FTP Server.

- To calculate the percentage of unemployment in the civilian labor force, two elements from Table 3 are used in the following equation:

$$\%\ Unemployment = \frac{B23025.005}{B23025.003}$$

- In EJScreen, the raw values for the unemployment indicator range from 0% to 100%. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the unemployment indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 4.2%. If a raw value is higher than 4.2%, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the unemployment indicator raw value for a block group is 10%, which is higher than 4.2%, that block group will be placed in the 85th percentile (which is higher than the 50th percentile).

Limited English Speaking Household

A limited English speaking household is defined as a household in which no one age 14 and over speaks only English, or speaks a non-English language and speaks English "very well" as reported in the U.S. Census Bureau's ACS. The percent of limited English speaking households is used instead of the actual number of limited English speaking households because percentages account for possible differences among population sizes and make block groups comparable.

How does EJScreen determine Limited English Speaking Households?

- The ACS limited English speaking household information is captured in the table *Household Language by Household Limited English Speaking Status* (ACS Table ID: C16002). The ACS divides limited English speaking households into four language groups as shown in Table 4.

Table 4. ACS Limited English Speaking Table Element Reference

Table Element	Language & Ability Status
C16002.001	Total Households
C16002.004	Limited English Speaking – Spanish
C16002.007	Limited English Speaking – Other Indo-European Languages
C16002.010	Limited English Speaking – Asian/Pacific Island Languages
C16002.013	Limited English Speaking – Other Languages

Note: The data can be downloaded from the US Census Bureau's FTP Server.

- To calculate percent limited English speaking, the elements from Table 4 are used in the following equation:

$$\%\ Limited\ English\ Speaking = \frac{C16002.004 + C16002.007 + C16002.010 + C16002.013}{C16002.001}$$

- In EJScreen, the raw values for the limited English speaking household indicator range from 0% to 100%. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the limited English speaking household indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 1.1%. If a raw value is higher than 1.1%, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the limited English speaking household indicator raw value for a block group is 8%, which is higher than 1.1%, that block group will be placed in the 80th percentile (which is higher than the 50th percentile).

Less than High School Education

Percent of people age 25 or older in a block group whose education is short of a high school diploma.

How does EJScreen determine less than high school education?

- The ACS education information is captured in the table *Sex by Educational Attainment for the Population 25 Years and Over* (ACS Table ID: B15002). The ACS divides the education information by grade as shown in Table 5.

Table 5. ACS Educational Attainment Table Element Reference

Table Element	Sex	Educational Attainment
B15002.001		Total Population Age ≥ 25
B15002.003	Male	No Schooling Completed
B15002.004		Nursery – 4th Grade
B15002.005		5th & 6th Grade
B15002.006		7th & 8th Grade
B15002.007		9th Grade
B15002.008		10th Grade
B15002.009		11th Grade
B15002.010		12th Grade, No Diploma
B15002.020	Female	No Schooling Completed
B15002.021		Nursery – 4th Grade
B15002.022		5th & 6th Grade
B15002.023		7th & 8th Grade
B15002.024		9th Grade
B15002.025		10th Grade
B15002.026		11th Grade
B15002.027		12th Grade, No Diploma

Note: These data can be downloaded from the US Census Bureau's FTP Server.

- To calculate the percentage of the 25 and older population with less than high school education, the elements from Table 5 are used in the following equation:

$$\% \text{ Less Than High School Education} = \frac{B15002.003 + .004 + .005 + .006 + .007 + .008 + .009 + .010 + .020 + .021 + .022 + .023 + .024 + .025 + .026}{B15002.001}$$

- In EJScreen, the raw values for the less than high school education indicator range from 0% to 100%. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the less than high school education indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 9%. If a raw value is higher than 9%, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the less than high school education indicator raw value for a block group is 20%, which is higher than 9%, that block group will be placed in the 80th percentile (which is higher than the 50th percentile).

Individuals under Age 5

Percent of people in a block group under the age of 5.

How does EJScreen determine individuals under age 5?

- The ACS information on individuals under age 5 is captured in the table *Sex by Age* (ACS Table ID: B01001). The ACS divides the population into 23 age groups for each sex. The elements that were used are shown in Table 6.

Table 6. ACS Age Table Element Reference

Table Element	Age
B01001.001	Total Population
B01001.003	Male, Under 5 Years
B01001.027	Female, Under 5 Years

Note: The data can be downloaded from the US Census Bureau's FTP Server.

- To calculate the percentage individuals under age of 5 for a block group, the elements from Table 6 are used in the following equation:

$$\% \, Individuals \; Under \; Age \; 5 = \frac{B01001.003 + B01001.027}{B01001.001}$$

- In EJScreen, the raw values for the individuals under age 5 indicator range from 0% to 50%. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the individuals under age 5 indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 5.6%. If a raw value is higher than 5.6%, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the individuals under age 5 indicator raw value for a block group is 9%, which is higher than 5.6%, that block group will be placed in the 80th percentile (which is higher than the 50th percentile).

Individuals over Age 64

Percent of people in a block group over the age of 64.

How does EJScreen determine individuals over age 64?

- The ACS information on individuals over age 64 is captured in the table *Sex by Age* (ACS Table ID: B01001). The ACS divides the population into 23 age groups for each sex. The elements that were used are shown in Table 7.

Table 7. ACS Age Table Element Reference

Table Element	Sex	Age
B01001.001		Total Population
B01001.020	Male	65 and 66 years
B01001.021		67 to 69 years
B01001.022		70 to 74 years
B01001.023		75 to 79 years
B01001.024		80 to 84 years
B01001.025		85 years and over
B01001.044	Female	65 and 66 years
B01001.045		67 to 69 years
B01001.046		70 to 74 years
B01001.047		75 to 79 years
B01001.048		80 to 84 years
B01001.049		85 years and over

Note: The data can be downloaded from the US Census Bureau's FTP Server.

- To calculate the percentage individuals over the age of 64 for a block group, the elements from Table 7 are used in the following equation:

$$\% \text{ Individuals over Age 64} = \frac{B01001.020+.021+.022+.023+.024+.025+.044+.045+.046+.047+.048+.049}{B01001.001}$$

- In EJScreen, the raw values for the individuals over age 64 indicator range from 0% to 100%. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the individuals over age 64 indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 14.2%. If a raw value is higher than 14.2%, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the individuals over age 64 indicator raw value for a block group is 22%, which is higher than 14.2%, that block group will be placed in the 80th percentile (which is higher than the 50th percentile).

Environmental Indicators in EJScreen

In the current release:

- Particulate Matter (PM) 2.5 and Ozone values are derived from 2018 source data from EPA's Office of Air Quality Planning and Standards (OAQPS), Non-attainment areas (NAA).
- Air Toxics data (Diesel PM, Cancer Risk, and Respiratory HI) are derived from 2017 source data from EPA's OAQPS.
- Traffic proximity source is derived from the 2019 Highway Performance Monitoring System (HPMS).
- Lead paint has been upgraded to Census Bureau's ACS 2016-2020 5-year Summary.
- Superfund proximity source is derived from Superfund Enterprise Management System (SEMS) database on April 26, 2022.
- Risk Management Plan (RMP) facility proximity source is derived from EPA's Facility Registry Service (FRS) by selecting facilities included in the RMP National Program System on April 26, 2022.
- Hazardous waste proximity sources are derived from operating Treatment, Storage, and Disposal Facilities (TSDFs) from RCRAInfo and Large Quantity Generators (LQGs) from the 2019 Biennial Reports (BR) on April 26, 2022.
- Underground Storage Tanks source provided by EPA's Office of Underground Storage Tanks on July 7, 2022.
- Wastewater discharge source provided by EPA's Office of Pollution Prevention and Toxics (OPPT) on March 15, 2021 from 2019 Risk-Screening Environmental Indicators (RSEI) modeled results.

Summary Overview of Environmental Indicators Featured in EJScreen

This section describes the environmental indicator data used in EJScreen. Some of these environmental indicators quantify proximity to and the numbers of certain types of potential sources of exposure to environmental pollutants, such as nearby hazardous waste sites or traffic. The lead paint indicator indicates the presence of older housing, which often, but not always, indicates the presence of lead paint, and therefore the possibility of exposure. In some cases, the term "exposure" is used very broadly here to refer to the potential for exposure. Other indicators in EJScreen are estimates of ambient levels of air pollutants, such as PM 2.5, ozone, and diesel PM. Still others are actual estimates of air toxics-related cancer risk or a hazard index (HI), which summarizes the ratios of ambient air toxics levels to health-based reference concentrations. In other words, these environmental indicators vary widely in what they indicate. EJScreen contains these 12 environmental indicators:

1. **PM 2.5**
    - PM 2.5 levels in air measured using an annual average.
2. **Ozone**
    - Ozone summer seasonal average of daily maximum 8-hour concentration in air.
3. **Diesel PM**
    - Diesel PM level in air.
4. **Air toxics cancer risk**
    - Lifetime cancer risk from inhalation of air toxics.
5. **Air toxics respiratory HI**
    - Ratio of exposure concentration to health-based reference concentration.
6. **Traffic proximity and volume**

- Count of vehicles (annual average daily traffic [AADT]) at major roads within 500 meters, divided by distance in meters (not kilometers [km]).

7. **Lead paint**
   - Percent of housing units built before 1960.
8. **Superfund proximity**
   - Count of proposed and listed NPL sites within 5 km (or nearest one beyond 5 km), each divided by distance in km. Count excludes deleted sites.
9. **RMP facility proximity**
   - Count of RMP (potential chemical accident management plan) facilities within 5 km (or nearest one beyond 5 km), each divided by distance in km.
10. **Hazardous waste proximity**
    - Count of hazardous waste management facilities (TSDFs and LQGs) within 5 km (or nearest one beyond 5 km), each divided by distance in km.
11. **Underground storage tanks (UST) and leaking UST (LUST)**
    - Count of LUSTs (multiplied by a factor of 7.7) and the number of USTs within a 1,500-foot buffered block group.
12. **Wastewater discharge**
    - RSEI modeled Toxic Concentrations at stream segments within 500 meters, divided by distance in km.

Detailed Descriptions of Environmental Indicators

Environmental Indicator—PM 2.5

What is the PM 2.5 indicator?

The PM 2.5 indicator is a measure of potential exposure to inhalable particles that are 2.5 micrometers or smaller. This is measured in terms of annual average concentration in air measured in micrograms per cubic meter. PM 2.5 information included in EJScreen highlights areas across the U.S. that are not meeting the national ambient air quality standard for PM 2.5. In other words, the levels of PM 2.5 present in these areas are deemed harmful to human health.

The PM 2.5 indicator in EJScreen is a measure of potential exposure but not a measure of risk. The raw PM 2.5 data is compiled by census tract which is supplied for use in the tool by EPA's OAQPS. For air toxics risk measures (as opposed to exposure) users can turn to EJScreen's other three indicators: cancer risk, respiratory HI, and diesel PM.

How does EJScreen determine the PM 2.5 indicator?

The PM 2.5 indicator data was provided by EPA's OAQPS using a fusion of monitor data and Community Multiscale Air Quality (CMAQ) air quality modeling. For more information about the methods used, see EPA Report EPA-454/S-15-001. This is provided to EJScreen as a spreadsheet compiled by Census tracts. The tract values are re-assigned to each block group, so all block groups within each tract have the same PM 2.5 value as for the tract.

In EJScreen, the raw values for the PM 2.5 indicator range from 3.93 to 17.75. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the PM 2.5 indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 8.6. If a raw value is higher than 8.6, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the PM 2.5 indicator raw value for a block group is 9.78, which is higher than 8.6, that block group will be placed in the 80th percentile (which is higher than the 50th percentile).

Where to find more information on PM 2.5?

For more information about PM, please visit EPA's website on PM. For real-time and forecasted information about all standard EPA air quality criteria, please visit EPA's AirNow website.

Environmental Indicator—Ozone

What is the ozone indicator?

The ozone indicator in EJScreen reflects potential ozone exposure measured in terms of summer seasonal daily average maximum concentrations in an 8-hour period measured in parts per billion. Ozone information included in EJScreen highlights areas across the U.S. that are not meeting the national ambient air quality standard for ozone. In other words, the levels of ozone present in these areas are deemed harmful to human health.

The ozone indicator in EJScreen is a measure of potential exposure but not a measure of risk. The raw ozone data is compiled by census tract, which is supplied for use in the tool by EPA's OAQPS. For air toxics risk measures (as opposed to exposure) users can turn to EJScreen's other three indicators: cancer risk, respiratory HI, and diesel PM.

How does EJScreen determine the ozone indicator?

The ozone indicator data was provided as a spreadsheet by EPA's OAQPS using a fusion of monitor data and CMAQ air quality modeling. For more information about the methods used, see EPA Report EPA-454/S-15-001. This is provided to EJScreen as a spreadsheet compiled by Census tracts. The tract values are assigned to each block group, so all block groups within each tract have the same ozone value as for the tract.

In EJScreen, the raw values for the ozone indicator range from 24.6 to 74.4. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the ozone indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 42.2. If a raw value is higher than 42.2, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the ozone indicator raw value for a block group is 43.6, which is higher than 42.2, that block group will be placed in the 64th percentile (which is higher than the 50th percentile).

Where to find more information on ozone?

To learn more, visit EPA's website on ozone. For real-time and forecasted information about all standard EPA air quality criteria, please visit EPA's AirNow website.

Environmental Indicator—Diesel PM

What is the Diesel PM indicator?

The Diesel PM indicator is the estimated concentration of Diesel PM as provided by the 2017 Air Toxics update. The value of the indicator is in µg /m^3, and reported at the Census tract level. Block group level values are assigned by repeating each parent tract level value.

How does EJScreen determine the Diesel PM indicator?

Diesel PM concentrations are provided by a spreadsheet from EPA OAQPS. The source data is compiled by Census tract. The tract values are repeated for each Census block group.

In EJScreen, the raw values for the Diesel PM indicator range from 2.213 $X10^{-6}$ to 1.927. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the Diesel PM indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 0.242. If a raw value is higher than 0.242, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the Diesel PM indicator raw value for a block group is 0.306, which is higher than 0.242, that block group will be placed in the 64th percentile (which is higher than the 50th percentile).

Where to find more information on Diesel PM?

Diesel PM comes from EPA's Air Toxics Data Update. The Air Toxics Data Update is the Agency's ongoing, comprehensive evaluation of air toxics in the United States. This effort aims to prioritize air toxics, emission sources, and locations of interest for further study. It is important to remember that the air toxics data presented here provide broad estimates of health risks over geographic areas of the country, not definitive risks to specific individuals or locations. More information on the Air Toxics Data Update can be found at: https://www.epa.gov/haps.

Environmental Indicator—Air Toxics Cancer Risk

What is the air toxics cancer risk indicator?

The air toxics cancer risk indicator is the estimated lifetime inhalation cancer risk from the analyzed carcinogens in ambient outdoor air, as provided by the 2017 Air Toxics data Update. The value of the indicator is persons per million lifetime. The data is reported at the Census tract level. Block group level values are assigned by repeating each parent tract level value.

How does EJScreen determine the air toxics cancer risk indicator?

- The indicator data for air toxics cancer risk per million are provided by a spreadsheet from EPA OAQPS. The source data is compiled by Census tract.
- The tract values are repeated for each Census block group.

- In EJScreen, the raw values for the air toxics cancer risk indicator range from 8.0 to 2000.0. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the air toxics cancer risk indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 23.46. If a raw value is higher than 23.46, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the air toxics cancer risk indicator raw value for a block group is 26.18, which is higher than 23.46, that block group will be placed in the 64th percentile (which is higher than the 50th percentile).
- To estimate air toxics cancer risks, the results of cancer dose-response assessments for a given chemical were converted to a unit risk estimate (URE). That URE was then multiplied by the estimated inhalation exposure concentration to obtain an estimate of individual lifetime cancer risk.

Where to find more information on air toxics cancer risk?

Air toxics cancer risk comes from EPA's Air Toxics Data Update. The Air Toxics Data Update is the Agency's ongoing, comprehensive evaluation of air toxics in the United States. This effort aims to prioritize air toxics, emission sources, and locations of interest for further study. It is important to remember that the air toxics data presented here provide broad estimates of health risks over geographic areas of the country, not definitive risks to specific individuals or locations. More information on the Air Toxics Data Update can be found at: https://www.epa.gov/haps.

Environmental Indicator—Air Toxics Respiratory HI

What is the air toxics respiratory HI indicator?

The air toxics respiratory HI indicator is the respiratory HI from the analyzed carcinogens in ambient outdoor air, as provided by the 2017 Air Toxics Data Update. The data is reported at the Census tract level. Block group level values are assigned by repeating each parent tract level value.

How does EJScreen determine the air toxics respiratory HI indicator?

- The indicator data for air toxics respiratory HI are provided by a spreadsheet from EPA OAQPS. The source data is compiled by Census tract.
- The tract values are repeated for each Census block group.
- In EJScreen, the raw values for the air toxics respiratory HI indicator range from 0.0 to 5.0. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the air toxics respiratory HI indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 0.307. If a raw value is higher than 0.307, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the air toxics respiratory HI indicator raw value for a block group is 0.348, which is higher than 0.307, that block group will be placed in the 64th percentile (which is higher than the 50th percentile).
- Air toxics estimated chronic noncancer hazards for multiple air toxics by summing chronic noncancer hazard quotients (HQs) for individual air toxics that cause similar adverse health

effects. The result is a HI. Aggregation in this way produces a target-organ-specific HI, defined as a sum of HQs for individual air toxics that affect the same organ or organ system.

Where to find more information on the air toxics respiratory HI indicator?

Air toxics respiratory HI comes from EPA's Air Toxics Data Update. The Air Toxics Data Update is the Agency's ongoing, comprehensive evaluation of air toxics in the United States. This effort aims to prioritize air toxics, emission sources, and locations of interest for further study. It is important to remember that the air toxics data presented here provide broad estimates of health risks over geographic areas of the country, not definitive risks to specific individuals or locations. More information on the Air Toxics Data Update can be found at: https://www.epa.gov/haps.

Environmental Indicator—Traffic Proximity

What is the traffic proximity indicator?

The traffic proximity indicator is based on AADT count divided by distance in meters from the Census block centroid. The proximity score is based on the traffic within a search radius of 500 meters (or further if none is found in that radius). This distance was selected to be large enough to capture the great majority of road segments (with traffic data) that could have a significant impact on the local residents, balanced against the need to limit the scope due to computational constraints. The closest traffic is given more weight, and the distant traffic is given less weight, through inverse distance weighting. For example, traffic 500 meters away is given only one tenth as much weight as traffic 50 meters away.

Why is traffic proximity an indicator in EJScreen?

- Proximity to roads can provide access to jobs, health care, food, recreational opportunities, and other benefits. However, in EJScreen, the indicator is designed to screen for the negative aspects of very close proximity to very high volumes of traffic, which include asthma and cardiovascular and heart disease, among others.
- Residential proximity to traffic has been associated with various health impacts, particularly asthma exacerbation and possibly onset of asthma, as well as mortality rates. Proximity to traffic has also been associated with subclinical atherosclerosis (a key pathology underlying cardiovascular disease [CVD]), prevalence of CVD and coronary heart disease (CHD), incidence of myocardial infarction, and CVD mortality.

How does EJScreen determine the traffic proximity indicator?

- Highway segments are from the HPMS lines and AADT counts are from the 2019 HPMS release, Federal Highway Administration, U.S. Department of Transportation (DOT). Proximity scores are calculated by assigning inverse distance weighted scores to Census blocks (distance between block centroids and highway segments). Note that blocks outside a 3-km cutoff are set to *null* due to computational constraints. The weighted scores are then multiplied by AADT to produce the final block scores. The results are aggregated to the parent block group using the population weight for each block within the block group.

- Information about HPMS can be found on this FHWA website. The 2019 spatial HPMS data is available by state from this DOT-hosted ArcGIS server: https://geo.dot.gov/server/services. EJScreen processing uses a subset of highways that include:
    - Interstates
    - Principal Arterial—Other Freeways and Expressways
    - Principal Arterial—Other
    - Minor Arterial in urban areas

- In EJScreen, the raw values for traffic proximity range from 0.00006 to 56,126. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for traffic proximity range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 222.0. If a raw value is higher than 222.0, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the traffic proximity raw value for a block group is 406.0, which is higher than 222.0, that block group will be placed in the 64th percentile (which is higher than the 50th percentile).

Environmental Indicator—Lead Paint Indicator

What is the lead paint indicator?

The lead paint indicator is the percentage of occupied housing units built before 1960, calculated from the U.S. Census Bureau's ACS 5-year summary estimates on age of housing stock. EJScreen uses age of housing stock as a surrogate for potential lead exposure as regulations banning lead-based residential paint in 1978 led to the reduction and finally an end to the use of such paint in housing. The percentage of older housing units is a proxy for potential exposure to lead paint and lead-containing dust that accumulates indoors, in homes, or in other buildings where lead paint was used. EJScreen uses housing units built before 1960.

How does EJScreen determine the lead paint indicator?

- The data are derived from the ACS Summary, block group-level estimates.
- The ACS information on lead paint is captured in the table *Year Structure Built* (ACS Table ID: B25034). The elements that were used are shown in Table 8.

Table 8. ACS Year Structure Built Table

Table Element	Year Structure Built
B25034.001	Total Housing Units
B25034.002	Housing Units Built 2014 or later
B25034.003	Housing Units Built 2010 to 2013
B25034.004	Housing Units Built 2000 to 2009
B25034.005	Housing Units Built 1990 to 1999
B25034.006	Housing Units Built 1980 to 1989
B25034.007	Housing Units Built 1970 to 1979

B25034.008	Housing Units Built 1960 to 1969
B25034.009	Housing Units Built 1950 to 1959
B25034.010	Housing Units Built 1940 to 1949
B25034.011	Housing Units 1939 or earlier

Note: The data can be downloaded from the US Census Bureau's FTP Server.

- To calculate the lead paint indicator, which is an indicator for percent pre-1960 housing units, four elements from Table 8 are used in the following equation:

$$\%\ Lead\ Paint = \frac{(B25034.009 + B25034.010 + B25034.011)}{B25034.001}$$

- In EJScreen, the raw values for the lead paint indicator range from 0 to 100%. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the lead paint indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 17%. If a raw value is higher than 17%, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the lead paint indicator raw value for a block group is 30%, which is higher than 17%, that block group will be placed in the 64th percentile (which is higher than the 50th percentile).

Environmental Indicator—Superfund Proximity

What is the Superfund proximity indicator?

The Superfund proximity indicator is reflective of the total count of sites proposed and listed (final) on the National Priorities List (NPL) in each block group within 5 km of the average resident in a block group, divided by distance, calculated as the population-weighted average of blocks in each block group.

How does EJScreen determine the Superfund proximity indicator?

- Final and proposed NPL sites are downloaded from the SEMS website. Proximity scores are calculated by assigning distance-weighted scores to 2010 Census blocks (distance between block centroids and facilities). The results are aggregated to the parent block group using the population weight for each block within the block group.
- In EJScreen, the raw values for the Superfund proximity indicator range from 0 to 8.988. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the Superfund proximity indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 0.061. If a raw value is higher than 0.061, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the Superfund proximity indicator raw value for a block group is 0.092, which is higher than 0.061, that block group will be placed in the 64th percentile (which is higher than the 50th percentile).
- More information about NPL sites can be found on this SEMS website.

Environmental Indicator—RMP Facility Proximity

What is the RMP facility proximity indicator?

The RMP facility proximity is reflective of the total count of RMP facilities in each block group within 5 km of the average resident in a block group, divided by distance, calculated as the population-weighted average of blocks in each block group.

How does EJScreen determine the RMP facility proximity indicator?

- RMP facilities are pulled from EPA's FRS by selecting facilities included in the RMP National Program System.
- Proximity scores are calculated by assigning distance-weighted scores to Census blocks (distance between block centroids and facilities). The results are assigned to block groups through population-weighted block to block group assignments.
- RMP facilities are queried by using the FRS Query website and selecting RMP in the program system list.
- In EJScreen, the raw values for the RMP facility proximity indicator range is from 0 to 18.45. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the RMP facility proximity indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 0.314. If a raw value is higher than 0.314, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the RMP facility proximity indicator raw value for a block group is 0.573, which is higher than 0.314, that block group will be placed in the 64th percentile (which is higher than the 50th percentile).

Environmental Indicator—Hazardous Waste Proximity

What is the hazardous waste proximity indicator?

The hazardous waste proximity indicator is reflective of the total count of hazardous waste facilities in each block group within 5 km of the average resident in a block group, divided by distance, calculated as the population-weighted average of blocks in each block group. Hazardous waste facilities are defined as Resource Conservation and Recovery Act (RCRA) handlers that are either operating TSDFs from RCRA or reporting LQGs from the 2019 BR.

How does EJScreen determine the hazardous waste proximity?

- Proximity scores are calculated by assigning distance weighted scores to Census blocks (distance between block centroids and facilities). The results are assigned to block groups through population-weighted block to block group assignments.
- TSDFs are collected by using the RCRAInfo Search website and selecting TSDF Handler Universe.
- 2019 BR LQGs are collected by using the BR Search website.
- In EJScreen, the raw values for the hazardous waste proximity indicator range is from 0 to 61.57. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the hazardous waste proximity indicator range from 0 to 100% with a median at the 50th

percentile, which corresponds to the raw value of 0.702. If a raw value is higher than 0.702, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the hazardous waste proximity indicator raw value for a block group is 1.440, which is higher than 0.702, that block group will be placed in the 64th percentile (which is higher than the 50th percentile).

Environmental Indicator—Underground Storage Tanks

What is the UST indicator?

The UST indicator quantifies the relative risk of being affected by a LUST for a block group. The indicator is derived by the weighted sum of active LUSTs and sum of active and temporarily out of service USTs within a certain distance from a block group.

How does EJScreen determine the UST indicator?

- The UST indicator scores are provided by the EPA Office of Underground Storage Tanks.
- EJScreen's UST indicator is calculated using the sum of LUSTs (multiplied by a factor of 7.7) and the number of USTs within a 1,500-foot buffered block group. That number is then divided by the area of the buffered block group in km². The 7.7 multiplier is derived from the average number of active USTs divided by the average number of LUSTs in the U.S. backlog (cleanups remaining) from 2011-2020. A 1,500-foot buffer is used as a radius of influence for the Benzene plume migration to encompass USTs/LUSTs near block groups that could potentially be affected by a release. See EPA's website on underground and leaking underground storage tanks for more information.
- In EJScreen, the raw values for the UST indicator range from 0 to 174.15. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the UST indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 1.25. If a raw value is higher than 1.25, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the UST indicator raw value for a block group is 5.6, which is higher than 1.25, that block group will be placed in the 80th percentile (which is higher than the 50th percentile).
- To calculate the UST indicator for a block group, the following equation is used:

$$UST\ Indicator = \frac{(\#\ of\ LUSTs\ \times 7.7) + (\#\ of\ Active\ USTs)}{Area\ of\ bufferd\ block\ group_{1500ft-buffer}}$$

Environmental Indicator—Wastewater Discharge

What is the wastewater discharge indicator?

The wastewater discharge indicator quantifies a block group's relative risk of exposure to pollutants in downstream water bodies. This is achieved using toxicity-weighted concentrations in stream reach segments within 500 meters of a block centroid, divided by distance in meters, presented as the population-weighted average of blocks in each block group.

How does EJScreen determine the wastewater discharge indicator?

- The wastewater discharge indicator utilizes pollutant loadings from the Discharge Monitoring Report (DMR) Loading Tool along with the RSEI model to estimate concentrations of pollutants in downstream water bodies and derive a toxicity-weighted concentration.
- To place higher emphasis on stream reaches with higher toxicity-weighted pollutant concentrations, the toxicity-weighted value for all stream reaches within 500 meters of a census block centroid is divided by the distance in meters to the census block centroid to create a weighted proximity value indicating a block's risk of exposure to pollutants in the stream reaches.
- The results are aggregated to the parent block group using the population weight for each block within the block group. The population weights come from the 2010 Census. Minor adjustments are needed to crosswalk Census 2010 blocks and 2019 blocks.
- The data was provided by EPA as a polyline feature class in a file geodatabase.
- In EJScreen, the raw values for the wastewater discharge indicator range from 0 to 63,257.7. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the wastewater discharge indicator range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 0.00104. If a raw value is higher than 0.00104, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the wastewater discharge indicator raw value for a block group is 0.045, which is higher than 0.00104, that block group will be placed in the 80th percentile (which is higher than the 50th percentile).

Considerations for Selection of Environmental Indicators in EJScreen

A variety of considerations has informed the selection of these environmental indicators; in general, the selected indicators exhibit the following characteristics:

- Resolution: Screening level data are available (or could be readily developed) at the block group level (or at least close to this resolution).
- Coverage: Screening level data are available (or could be readily developed) for the entire United States (or with nearly complete coverage).
- Relevance to EJ: Pollutants or impacts are relevant to EJ (e.g., differences between groups have been indicated in exposures, susceptibility, or health endpoints associated with the exposures).
- Public health significance: Pollutants or impacts are potentially important in the United States (e.g., notable impacts estimated or significant concerns have been expressed, at least locally, or exposure has been linked to health endpoints with substantial impacts nationwide).

3 Overview of Indexes in EJScreen

This section describes the environmental and socioeconomic indexes featured in the tool, why they are included, and how they are derived.

Demographic Indexes Included in EJScreen

EJScreen includes two demographic indexes based on different variations of the socioeconomic indicators. The two demographic indexes featured in EJScreen are:

- **Demographic Index** is based on the average of two demographic indicators: percent low-income and percent people of color.
- **Supplemental Demographic Index** is based on the average of five socioeconomic indicators: percent low-income, percent limited English speaking, percent less than high school education, percent unemployed, and low life expectancy.

Demographic Index

What is the Demographic Index in EJScreen?

The Demographic Index in EJScreen is a combination of percent low-income and percent people of color. These are the two demographic factors explicitly named in Executive Order 12898 on Environmental Justice. For each Census block group, these two numbers are simply averaged together. The formula is as follows:

$$Demographic\ Index = \frac{\%\ Low\ Income + \%\ People\ of\ Color}{2}$$

For example, if a Census block group has a low income indicator value of 25% and a people of color indicator value of 75%, the Demographic Index value would be 50%.

How does EJScreen determine the Demographic Index?

EJScreen uses these two demographic indicators:

1. The ACS low income information is captured in the table *Ratio of Income to Poverty Level in the Past 12 Months* (ACS Table ID: C17002). The ACS divides the ratio of income to poverty into seven categories, as shown in Table 9.

Table 9. ACS Income/Poverty Level Table Element Reference

Table Element	Income/Poverty Level
C17002.001	Total Population Whose Poverty Status is Known
C17002.002	People with Ratio of Income to Poverty under .50
C17002.003	People with Ratio of Income to Poverty from .50 to .99
C17002.004	People with Ratio of Income to Poverty from 1.00 to 1.24
C17002.005	People with Ratio of Income to Poverty from 1.25 to 1.49

C17002.006	People with Ratio of Income to Poverty from 1.50 to 1.84
C17002.007	People with Ratio of Income to Poverty from 1.85 to 1.99
C17002.008	People with Ratio of Income to Poverty from 2.00 and over

Note: The data can be downloaded from the US Census Bureau's FTP Server.

- To calculate percent low income, two elements from Table 9 are used in the following equation:

$$\%\ Low\ Income = \frac{C17002.001 - C17002.008}{C17002.001}$$

2. The ACS people of color information is captured in the table *Hispanic or Latino Origin by Race* (ACS Table ID: B03002). The ACS divides race and Hispanic status into 21 categories, four of which are shown in Table 10.

Table 10. ACS Hispanic or Latino Origin by Race Table Element Reference

Table Element	Hispanic Status/Race Population
B03002.001	Total Population: All races/ethnicities
B03002.002	Total Population: Non-Hispanic
B03002.003	Total Population: Non-Hispanic, White Alone
B03002.012	Total Population: Hispanic

Note: The data can be downloaded from the US Census Bureau's FTP Server.

- To calculate percent people of color, two elements from Table 10 are used in the following equation:

$$\%\ People\ of\ Color = \frac{B03002.001 - B03002.003}{B03002.001}$$

- To calculate the EJScreen Demographic Index, the results from the two previous calculations are averaged as follows:

$$Demographic\ Index = \frac{\%\ Low\ Income + \%\ People\ of\ Color}{2}$$

In EJScreen, the raw values for the Demographic Index range from 0% to 100%. In order to make the indicator more comparable, statistical percentiles are used. The percentiles for the Demographic Index range from 0 to 100% with a median at the 50th percentile, which corresponds to the raw value of 30%. If a raw value is higher than 30%, it would be placed above the 50th percentile; if the value is lower, it would be placed below the 50th percentile. For example, if the Demographic Index raw value for a block

group is 50%, which is higher than 30%, that block group will be placed in the 74th percentile (which is higher than the 50th percentile).

Why is the Demographic Index in EJScreen?

- These are the two demographic factors explicitly named in Executive Order 12898 on Environmental Justice.
- Low income and people of color populations often experience greater exposure to environmental burdens than the general population as a whole.
- Many studies have established that sources of environmental hazards are often located and concentrated in areas that are dominated by low income and people of color populations.

Supplemental Demographic Index

What is the Supplemental Demographic Index in EJScreen?

The Supplemental Demographic Index uses the same updated methodology and calculation as the EJ Indexes but replaces the current Demographic Index (the average percent low-income and percent people of color) with a supplemental five-factor demographic index. The five socioeconomic indicators considered are percent low life expectancy, percent low-income, percent unemployed, percent limited English speaking, and percent less than high school education.

How does EJScreen determine the Supplemental Demographic Index?

The Supplemental Demographic Index is the average of the following five indicators:

1. Low Life Expectancy—Life Expectancy at Birth from Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS):

   % Low Life Expectancy is defined as "1 – (Life Expectancy / Max Life Expectancy)"

   Note: *This is derived from the CDC life expectancy at birth data using the formula above.*

2. The ACS low income information is captured in the table *Ratio of Income to Poverty Level in the Past 12 Months* (ACS Table ID: C17002). The ACS divides the ratio of income to poverty into seven categories, as shown in Table 11.

Table 11. ACS Income/Poverty Level Table Element Reference

Table Element	Income/Poverty Level
C17002.001	Total Population Whose Poverty Status is Known
C17002.002	People with Ratio of Income to Poverty under .50
C17002.003	People with Ratio of Income to Poverty from .50 to .99
C17002.004	People with Ratio of Income to Poverty from 1.00 to 1.24
C17002.005	People with Ratio of Income to Poverty from 1.25 to 1.49
C17002.006	People with Ratio of Income to Poverty from 1.50 to 1.84

C17002.007	People with Ratio of Income to Poverty from 1.85 to 1.99
C17002.008	People with Ratio of Income to Poverty from 2.00 and over

Note: The data can be downloaded from the US Census Bureau's FTP Server.

- To calculate percent low income, two elements from Table 11 are used in the following equation:

$$\%\ Low\ Income = \frac{C17002.001 - C17002.008}{C17002.001}$$

3. The ACS unemployment information is captured in the table *Employment Status for the Population 16 Years and Over* (Table ID: B23025). The ACS divides the population age 16 or older into seven categories, which are shown in Table 12. EJScreen uses the Civilian Labor Force numbers.

Table 12: ACS Employment Status Table Element Reference

Table Element	Employment Status
B23025.001	Total Population Age 16 Years and Over
B23025.002	Total Labor Force
B23025.003	Total Civilian Labor Force
B23025.004	Civilian Labor Force: Employed
B23025.005	Civilian Labor Force: Unemployed
B23025.006	In Armed Forces
B23025.007	Total Not In Labor Force

Note: These data can be downloaded from the US Census Bureau's FTP Server.

- To calculate the percentage of unemployment in the civilian labor force, two elements from Table 12 are used in the following equation:

$$\%\ Unemployment = \frac{B23025.005}{B23025.003}$$

4. The ACS limited English speaking household information is captured in the table *Household Language by Household Limited English Speaking Status* (ACS Table ID: C16002). The ACS divides limited English speaking households into four language groups as shown in Table 13.

Table 13. ACS Limited English Speaking Table Element Reference

Table Element	Language & Ability Status
C16002.001	Total Households
C16002.004	Limited English Speaking – Spanish

C16002.007	Limited English Speaking – Other Indo-European Languages
C16002.010	Limited English Speaking – Asian/Pacific Island Languages
C16002.013	Limited English Speaking – Other Languages

Note: The data can be downloaded from the US Census Bureau's FTP Server.

5. The ACS education information is captured in the table *Sex by Educational Attainment for the Population 25 Years and Over* (ACS Table ID: B15002). The ACS divides the education information by grade as shown in Table 14.

Table 14. ACS Educational Attainment Table Element Reference

Table Element	Sex	Educational Attainment
B15002.001		Total Population Age ≥ 25
B15002.003	Male	No Schooling Completed
B15002.004		Nursery – 4th Grade
B15002.005		5th & 6th Grade
B15002.006		7th & 8th Grade
B15002.007		9th Grade
B15002.008		10th Grade
B15002.009		11th Grade
B15002.010		12th Grade, No Diploma
B15002.020	Female	No Schooling Completed
B15002.021		Nursery – 4th Grade
B15002.022		5th & 6th Grade
B15002.023		7th & 8th Grade
B15002.024		9th Grade
B15002.025		10th Grade
B15002.026		11th Grade
B15002.027		12th Grade, No Diploma

Note: These data can be downloaded from the US Census Bureau's FTP Server.

Here is the formula for computing the Supplemental Demographic Index:

Supplemental Demographic Index = (% Low Life Expectancy + % Low Income + % Unemployment Rate + % Limited English Speaking + % Less Than High School Education) / 5

Note that the CDC Life Expectancy data are available for about 90% of the country. For the areas where the Life Expectancy data are not available, the Supplemental Demographic Index becomes the average of four indicators instead of five, which is calculated as "(% Low Income + % Unemployment Rate + % Limited English Speaking + % Less Than High School Education) / 4".

Why is the Supplemental Demographic Index in EJScreen?

The Supplemental Demographic Index offers a different perspective on community-level vulnerability across the country. The Supplemental Demographic Index also provides additional functionality for use

in decision-making consideration, such as the allocation of government resources when awarding grants. For questions on the appropriate use of the EJ and Supplemental Demographic Indexes, please contact your servicing legal office.

Indexes Included in EJScreen

EJScreen features two sets of indexes—12 EJ Indexes and 12 Supplemental EJ Indexes, which are described in detail below. The indexes are a combination of environmental indicators described above and the Demographic Index, or the Supplemental Demographic Index described above.

EJ Indexes

The EJ Indexes are a combination of environmental indicators described above and the Demographic Index. EJScreen features a single EJ Index for each of the 12 environmental indicators.

EJ Index Calculations:

To calculate a single EJ Index, EJScreen combines a single environmental indicator with the Demographic Index.

The equation for EJ Index calculations is as follows:

EJ Index = Demographic Index x Normalized Environmental Indicator

where Normalized Environmental Indicator is Percentile of Environmental Indicator Source Data

Percentiles Methodology Notes:

- Percentiles calculations are unweighted.
- Percentile ties use a floor method. This produces the lowest value, for example, for an element with more than 50 percent zeros, like “Limited English Speaking”, 0 will yield 0 Percentile.

EJ Index Example:

For block group (BG) = 410510068022

% people of color for BG = 0.24898694281855
% low income for BG = 0.153534443944169

National Percentile for Environmental Indicator for Superfund Proximity = 65

Demographic Index for BG = (% people of color + % low income) / 2
Demographic Index for BG = (0.24898694281855 + 0.153534443944169) / 2
Demographic Index for BG = 0.2012606933813595

EJ Index for Superfund Proximity = National Percentile for Environmental Indicator * Demographic Index for BG
EJ Index for Superfund Proximity = 65 * 0.2012606933813595
EJ Index for Superfund Proximity = 13.082

EJ Index Percentile in USA = 54

Supplemental Indexes

The supplemental indexes are a combination of environmental indicators described above and the Supplemental Demographic Index. Similar to the EJ Indexes, EJScreen features a single supplemental index for each of the 12 environmental indicators.

Supplemental Index Calculations:

To calculate a single supplemental index, EJScreen combines a single environmental indicator with the Supplemental Demographic Index.

Supplemental Index = Supplemental Demographic Index x Normalized Environmental Indicator

where Normalized Environmental Indicator is Percentile of Environmental Indicator Source Data

Supplemental Index Example:

For block group (BG) = 410510068022

% low life expectancy for BG = 0.152820512820513
% low income for BG = 0.153534443944169
% unemployment rate for BG = 0.016653449643140
% limited English speaking for BG = 0.011891891891892
% less than high school education for BG = 0.067593177511055

National Percentile for Environmental Indicator for Superfund Proximity = 65

Supplemental Demographic Index for BG = (% low life expectancy + % low income + % unemployment rate + % limited English speaking + % less than high school education) / 5
Supplemental Demographic Index for BG = (0.152820512820513 + 0.153534443944169 + 0.016653449643140 + 0.011891891891892 + 0.067593177511055) / 5
Supplemental Demographic Index for BG = 0.0804986951621538

Supplemental Index for Superfund Proximity = National Percentile for Environmental Indicator * Supplemental Demographic Index for BG
Supplemental Index for Superfund Proximity = 65 * 0.0804986951621538
Supplemental Index for Superfund Proximity = 5.232415185539997

Supplemental Index Percentile in USA = 50

4 Technical Details on Percentiles

What a Percentile Means

EJScreen puts each indicator or index value in perspective by reporting the value as a percentile. For example, an area may show 60% of housing was built prior to 1960. It may not be obvious whether this is a relatively high or low value, compared to the rest of the nation or in the state. Therefore, EJScreen also reports that 60% pre-1960 puts this area at the 80th percentile nationwide. For a place at the 80th percentile nationwide, that means 20% of the U.S. population has a higher value.

A percentile in EJScreen tells us roughly what percent of the U.S. population lives in a block group that has a lower value (or in some cases, a tied value). This means that 100 minus the percentile tells us roughly what percent of the U.S. population has a higher value. This is generally a reasonable interpretation because for most indicators there are not many exact ties between places and not many places with missing data.

Note that when there are ties, a "floor" method is used to make the assignment. For example, if an indicator value of 0 is assigned percentiles from 0 to 10, the reported percentile will be 0 instead of 10.

More precisely, the exact percentile for a given raw indicator value is calculated as the number of U.S. residents of block groups with that value or lower, divided by the total population with known indicator values. This is typically the same as or almost exactly the same as dividing by the total U.S. population, but for some indicators some locations do not have an indicator value. For example, the Air Toxics indicators are missing for only about one twentieth of 1% of the U.S. population in EJScreen. The calculated percentile would change by much, much less than 1 percentile point if calculated as a fraction of the total population instead of as a fraction of those with valid indicator values.

Color-coded High Percentile Bins

Locations at least at the 80th percentile but less than the 90th are shown in yellow on EJScreen maps, while those at the 90th percentile but less than 95th percentile are orange on the maps, and those at the 95th percentile or above are shown in red on maps and reports. These colors call attention to certain locations as a very simple way to communicate relative screening results. There is no official policy significance assigned to each individual color on the maps, but the choice of these categories or "bins" is noteworthy because it signifies that certain ranges of percentiles may merit closer attention.

Percentiles at or above the 95th percentile are shown in red on the EJScreen standard report. This is a way to call particular attention to those cases where the value is in the top 5% of the nation (or region or state). Indicator or index values in the top 5% tend to be much higher than those in the next 5-10%, so they may merit close attention. This is especially true for the indicators with highly skewed distributions, such as the traffic proximity indicator. For example, block groups in the top 5% (shown in red on maps and reports) have traffic, NPL, and TSDF proximity indicators on average that are about three times as high as in the next 5% (shown in orange on the maps). These differences are far less

extreme in the cases of PM 2.5 and lead paint indicators, which don't vary as much across block groups. In general, though, indicator or index values above the 95th percentile represent much higher demographic, environmental, or EJ Index values than those at lower percentiles.

The maps also identify areas in the 90th to 95th percentiles as orange, and those at the 80th to 90th as yellow. These additional categories highlight larger groups of locations that have indicator or index values well above the national mean or median for the given indicator or index. The actual values are lower than those in the top 5%, typically much lower, but they are still in the top 10 to 20% of values for the U.S. population overall.

A relatively high percentile means the value is relatively uncommon. However, a high percentile is not necessarily a real concern from a health or legal perspective. To understand the actual health or other implications of any screening results requires looking at the actual data the indicator represents, and also looking at other relevant data if available. Besides the percentile, other important considerations in interpreting any screening results include the following:

1. whether and to what extent the environmental data shows values above any relevant health-based or legal threshold,
2. the significance of any such thresholds, or the magnitude and severity of the health or other impacts of the given environmental concern, nationally or locally, and
3. the degree of any disparity between various groups, in exposures to the relevant environmental pollutants.

In maps, EJScreen focuses on the U.S. percentiles as a way to visualize all results in common units.

The U.S. percentile uses the U.S. population as the basis of comparison. The state percentile was calculated based on the population in a given state (or District of Columbia or Puerto Rico). The national or state mean value was calculated as the population weighted average of the block groups with data for that indicator, within the respective geographic scope.

Note that the U.S. and state percentiles both will rank block groups in exactly the same rank order within the given state. If the goal is just to rank or compare locations within a single state, it does not matter whether the U.S. or state percentile is used. The difference between state and U.S. percentiles becomes apparent mainly in two situations: when comparing places across states, or when comparing results to some pre-determined, specific reference percentile (e.g., 80th percentile).

The advantage of U.S. percentiles for an EJ Index, for example, is that a higher percentile in place A versus place B clearly indicates that the combination of the environmental indicator and Demographic Index is greater in place A than place B. In a sense, the U.S. percentile indicates how uncommon it is to have such a high level for an indicator or index.

State percentiles cannot be compared across states as easily. If two places A and B, in two different states, happen to both be at the 80th percentile for the traffic proximity EJ Index, for example, it is not clear which actually has the higher index value. It just means that A's index is just as uncommon within

that state as B's is in B's state. However, this may be useful information because an EJScreen user may want to know how high the indicator is relative to the rest of that state.

The state and U.S. percentiles will be very similar if the state and U.S. average indicator values are very similar. However, if the state average is very low compared to the U.S., the state percentile shown will be higher than U.S. percentile shown, for a given raw value of an indicator. If the state average is much higher than the U.S. average, for an indicator like the traffic proximity indicator, then a traffic score that would normally be considered fairly high nationwide, such as the 90th percentile in the U.S., would not be considered very unusual within that state, so the state percentile would be lower, and might be only 78th percentile, for example. The state percentile being lower than the U.S. percentile does not mean the indicator value is lower in the given place, it just means the state average is higher than the U.S. average.

How Percentiles are Calculated

The percentiles and lookup tables were calculated using the statistical software called R, using code written by EPA, based on wtd.quantile() and wtd.Ecdf() functions in the Hmisc package (http://cran.r-project.org/web/packages/Hmisc/index.html). The scripting language R is documented here: http://cran.r-project.org

5 Thresholds

Indexes Threshold Map Widget

The Indexes Threshold Map Widget allows users to investigate a potential hotspot with user-specified criteria based on all 12 EJ Indexes. It works with EJScreen Indexes and Supplemental Indexes.

It provides users with a new capability of specifying custom areas of concern based on Percentile bins tabulated for the 12 EJ Indexes.

- The Indexes Threshold Map Widget is a tool for users to produce custom threshold maps. The tool allows the user to select:
    - Data type—EJ Index or Supplemental Index
    - Data source–U.S. or State Percentiles
    - Index Percentile Range
    - All indexes or user-selected subset of indexes
- The tool uses data built from these EJScreen Index datasets:
    1. National Percentiles built based on the two-factor Demographic Index
    2. State Percentiles built based on the two-factor Demographic Index
    3. National Percentiles built based on the Supplemental Demographic Index
    4. State Percentiles built based on the Supplemental Demographic Index
- Datasets include these elements:
    1. 12 Calculated EJ Index Percentiles
    2. 101 Percentile bin counters (0 to 100)

Initial Filter Approach for Screening

In past screening experience, EPA has found it helpful to establish a suggested Agency starting point for the purpose of identifying geographic areas that may warrant further consideration, analysis, or outreach. The use of an initial filter promotes consistency and provides a pragmatic first step for EPA programs and regions when interpreting screening results. For early applications of EJScreen, EPA identified the 80th percentile filter as that initial starting point. In other words, an area with any of the 12 EJ Indexes at or above the 80th percentile nationally should be considered as a potential candidate for further review. Further review may include considering other factors and other sources of information such as health-based information, local knowledge, proximity and exposure to environmental hazards, susceptible populations, unique exposure pathways, and other federal, regional, state, and local data. This filter is simply a starting point, and program offices and regions should perform additional analysis before making any decisions about potential environmental justice issues. As EPA gains further experience and insight into the performance of the tool and its applicability for different uses, program offices and regions may opt to designate starting points that are more inclusive or specifically tailored to meet programmatic needs more effectively.

The 80th percentile filter in EJScreen is not intended to designate an area as an "EJ community." EJScreen provides screening level indicators, not a determination of the existence or absence of EJ concerns. Nor does the use of the 80th percentile filter suggest that all of the 12 environmental indicators are equal in terms of their impact on human health and the environment. Instead, the 80th percentile filter encourages programs to consider environmental indicators outside of their areas of concentration. The Agency may revise this approach in the future based on experience. This 80th percentile filter is for internal EPA use and is not intended to apply to States or other organizations.

6 Buffer Reports

EJScreen allows a user to define a buffer, such as the circle that includes everything within 1 mile of a specific point. Non-circular, user-defined shapes also can be defined to represent buffers of any shape. A report summarizes the demographics of residents within this buffer, as well as the environmental indicators and EJ Index values within the buffer.

The summary within a buffer is designed to represent the average resident within the buffer, and also provides an estimate of the total population residing in the buffer. For example, the traffic proximity indicator for a buffer is the population-weighted average of all the traffic indicator values in the buffer. Similarly, the percent people of color would be a weighted average, which is the same as the overall percent people of color for all residents in the buffer.

Some block groups will be partly inside and partly outside a buffer, and any buffer analysis must estimate how much of each block group's population is inside the buffer. Areal apportionment of block groups is one standard method, but it assumes that population is evenly spread throughout a block group, which may be far from the actual distribution of residents. Areal apportionment of blocks would be even more accurate but extremely computationally intensive.

To provide the most accurate counts that are currently feasible for a screening tool, EJScreen uses an approach based on decennial Census block internal points. EJScreen estimates the fraction of the Census block group population that is inside the buffer by using *block*-level population counts from the decennial Census. These blocks provide data about where residents are at a higher resolution than block groups. Each block has an internal point defined by the Census Bureau, and the entire block population is counted as inside or outside the buffer depending on whether the block internal point is inside or outside. This assumption typically introduces relatively little error because blocks are so small relative to a typical buffer, so a small fraction of the total buffer population is in blocks that span an edge of the buffer. Also, any blocks along the edge of a buffer whose populations are close to 0 or 100% inside the buffer will be well represented by this assumption.

As long as users draw buffers much larger than a local block group, this method should represent the average person inside the buffer reasonably well.

The calculation of a value for the buffer is essentially the population-weighted average of the indicator values in the blocks included in the buffer, where each block uses the indicator values of the block group containing it. A block group is weighted based on the fraction of the current ACS block group population that is considered in the buffer. That fraction is estimated as the decennial Census block population divided by the decennial Census block group population. The formula below is used to estimate the population average of a raw indicator value in a buffer. This formula is simply a population-weighted average – it sums the population-weighted raw values, and then divides that sum by the total population in the buffer.

$$Value(A) = \sum_{\forall Blk, \overline{Blk} \cap A} \frac{\frac{BlockPop}{BGPop} * BGACSPop * BG\_RawValue}{\sum_{\forall Blk, Blk \cap A} \frac{BlkPop}{BGPop} * BGACSPop}$$

"BlockPop" refers to the decennial Census block level population total (used here because the ACS does not provide block resolution), and "BG" indicates block group. "BGACSPop" is the block group estimated population count from the current ACS, which is often different than the decennial Census total for all blocks in the block group, because the ACS data used here is a composite estimate based on survey samples spanning five years, while the decennial Census is a full count at one point in time.

7 Technical Details on Proximity Indicators

Several of EJScreen's environmental indicators are direct or indirect estimates of potential exposure or health risks, such as the Air Toxics cancer risk estimates and the ozone and PM 2.5 concentration estimates. There are other aspects of an individual's or a community's environmental concerns that are less readily quantified in terms of emissions, concentrations, or risk estimates.

People may be concerned about living near facilities that handle hazardous substances, and other potential sources of pollution, such as highways or abandoned waste sites. Concern over "locally undesirable land uses" (LULUs) is in some cases founded on the potential for routine or episodic releases of pollutants to the air, land, or water, and the potential for such releases to cause human health or environmental adverse effects or other societal disamenities.

The purpose of the proximity measures in EJScreen is to systematically and consistently quantify different degrees of potential for these effects. We have developed a method to calculate a score that represents the relative magnitude of the proximity of the population within a block group to facilities, waste sites, or traffic surrounding it. A block group with more facilities closer to the block group's residential population will have a higher score than a block group where facilities are further away. We have applied this method to these facility or site types:

- NPL sites (a key subset of "Superfund," sites).
- Hazardous waste TSDFs, subject to regulations under the RCRA.
- RMP facilities, which are facilities that maintain greater than certain quantities of extremely hazardous substances, and are required to take certain actions, including filing risk management plans, under section 112 (r) of the Clean Air Act.

We have developed a similar approach to represent proximity to traffic volume on nearby highways and proximity to toxic concentrations on nearby water segments.

In the sections below, we will describe the general approach, in terms of facility proximity. We will then describe how it differs for traffic and wastewater proximity. Then we will discuss certain adjustments we have made, mostly to make the approach computationally efficient, and summarize the data sources and computational routine that we applied to implement this approach. We conclude with caveats and other observations.

Calculating Proximity to Facilities

Each of the 242,335 block groups for the U.S. states, District of Columbia, and Puerto Rico is made up of between one block and several hundred blocks. Most block groups nationwide are smaller than approximately 0.5 square miles, an area that if circular would have a radius of about 640 meters. In block groups of this median size, the average residence generally would be about 430 to 720 meters (or less than half a mile) away from a given point within the block group, such as a facility, as explained at the end of this section. About 20-25% of block groups covered an area smaller than a circle of radius

300-350 meters (almost one quarter of a mile). Also, a very small number of block groups are extremely large in area, in very rural locations.

All of a block group's blocks may have residential population estimated by the 5-year ACS, or only some, and some block groups have no residents at all. Blocks and block groups vary greatly in geographic area, and in population. The approach used here works first at the block level, based on measures of proximity to the facilities in or near the blocks. The block-level measures are then aggregated among all the blocks within a block group, weighted by the number of people in the different blocks.

Thus, while population is considered in aggregating the block scores, the measure does not increase or decrease for block groups with higher or lower populations. The measure is, rather, a characteristic of the residents of the block group, in the same way that cancer risk from Air Toxics or ozone concentration are estimated measures of the conditions of those places.

Let

i represent a particular facility
j represent a block within a block group
k represent a block group
d_{ij} is the distance, in km, from block j's centroid to the given location of facility i
pop_{jk} is the estimated population of block j within block group k
pop_k is the total estimated population of block group k
$f(d_{ij})$ is a function representing the proximity of facility i to block j, a declining function of the distance, d_{ij}
$BlockScore_{jk}$ is the aggregation of the proximity influences of all facilities affecting block jk
$BlockGroupScore_k$ is the population-weighted aggregation of the block group's component blocks

We have chosen to define the proximity function as

$f(d_{ij}) = 1 / d_{ij}$

That is, a facility 1 km from a block's population contributes twice the score as a facility 2 km from the same block. We note that we have made a choice in using inverse distance for this function. Air dispersion modeling for pollutants following Gaussian plume assumptions would show a generally greater drop-off in concentration, roughly with the second power to 2.5 power of one over distance. But actual concentrations around individual plants follow often-complex patterns that depend on the particular mix of stack versus fugitive emissions, characteristics of stack height, exit velocity and temperature, the presence of buildings or other land surface characteristics and meteorology. Some substances react readily with other substances in the atmosphere, or precipitate out readily. It is not uncommon for concentrations to rise for some distance from the emitting source, and then to fall from that peak concentration. The Gaussian plume model applies to gases, and emissions of particulates can drop off more quickly than gases.

Releases to land may follow extremely complex patterns of dispersion. Added to that are the very site-specific characteristics of potential human exposure via drinking water, vapor intrusion, or contact with contaminated soils, etc. For water pollution, similar complexities exist, most notably that an effluent is

carried away downstream of a running body of water, dilution can be complicated by the presence of other water entering stream segments, by volatilization, by biological and chemical interactions, and by deposition to sediments, and finally by the treatment and removal of a water pollutant sent to a publicly-owned treatment works.

We also note that researchers and others have taken varied approaches to representing the proximity of facilities to populations. The Environmental Justice Screening Method (EJSM) model of environmental justice concerns, developed for the state of California, scored facility proximity in concentric rings around a population centroid (Pastor Jr., Morello-Frosch, & Sadd, 2010; Sadd, Pastor, Morello-Frosch, Scoggins, & Jesdale, 2011). All facilities within 1 mile received a score of 3. All within the 1- to 3-mile band received a score of 2, and those between 3 and 5 miles received a score of 1. Anything beyond 5 miles received a score of zero. This step-wise scoring represents the judgment of the model developers, influenced by interactions with various stakeholders.

Finally, we note that EJScreen's measure of proximity is intended to represent more than simply real or potential human health adverse effects coming from exposure. Some parts of the environmental justice literature reflect semi-quantitative factors, such as increased psychological stress, fear, and other reactions to the presence of LULUs. This is not the forum for sorting through those factors.

However, we have made a judgment call: For the purposes of this EJScreen tool, we represent a facility's measure of proximity by the inverse of its distance from the estimated location of the average person. A block's proximity score is the sum of the inverse distances of all the facilities of a particular type.

Note that for the minority of block groups in the United States with no residential population, we take a straight average of the block scores.

The units for these measures are facilities per km. A block group could have a score of 1.0 if all residents were an average of 1 km from a single facility, and all other facilities were so distant (> 5 km) as to make no contribution to the score. Another block group could have a score of 1.0 if there were five facilities that were all exactly 5 km from the residents.

Calculating Proximity to Traffic

We have adopted essentially the same approach described above for representing proximity to highway segments – an inverse distance-weighted sum of highway segments surrounding each block, and a population-weighted sum of the individual blocks' contributions to the block group.

The highway segment database that we have used is described in section 2. These segments differ from a facility database in that they are lines on a geographic area, rather than points that represent the facilities. In our approach, we find the distance from the block centroids to the nearest part of each surrounding highway segment. The nearest point, d_{ij}, could be an end of the highway segment or some point between the ends.

We also multiplied each d_{ij} by the AADT estimate that is associated with each highway segment. This is meant to reflect the traffic intensity, and this differs from the facility approach, where we have taken

each facility within each group as having equal importance. Also, for traffic proximity, the search radius is 500 meters and the score uses distance in meters, not km.

Calculating Proximity to Toxic Weighted Wastewater Dischargers

The proximity to wastewater dischargers was calculated using the same methodology as the proximity to traffic—an inverse distance-weighted sum of water reach segments surrounding each block, and a population-weighted sum of the individual blocks' contributions to the block group. As described above in the Environmental Indicators in EJScreen section, the source data is RSEI modeled output of toxics concentrations mapped to water segments. So essentially EPA replaced highways with National Hydrography Dataset (NHD) reach water segments, and AADT counts with toxic concentrations.

The Wastewater Discharge map layer (as it is currently referred to in EJScreen) was adjusted to display all National Pollutant Discharge Elimination System (NPDES) facility latitude/longitude with pollutant loadings greater than zero. In addition, EPA provided more information in the popup box associated with each mapped facility, including a link to the detailed facility information in the DMR Pollutant Loading Tool, so that users can query the specific pollutants the facility may be discharging.

All relevant stream parameters were already included in the RSEI modeling context and were upgraded to NHD Version 2.0, which contained improved flow estimates and better connectivity.

Currently the DMR Loading Tool includes 672 reported chemicals from DMRs, of which 201 are also reported to TRI. The ideal set of loadings would be the union of the DMR and TRI datasets. For the overlap, that is, facility/chemical combinations reported to both DMR and TRI, DMR reported quantities were prioritized, as they are not subject to some of the constraints of TRI reporting and are usually more accurate, being based on actual monitoring data.

The concentrations of the modeled chemicals were weighted by their relevant toxicity weight. Only toxicity weights from RSEI were used. EPA excluded chemicals without RSEI toxicity weights or which cannot be readily extrapolated for an existing chemical. RSEI toxicity weights are based solely on chronic human health effects, and therefore, may be more suitable for an application like EJScreen that is geared toward screening for human health concerns. For water releases, the RSEI toxicity weight is calculated as the reciprocal of the reference dose (RfD) for noncarcinogens or cancer potency factor/$1x10^{-6}$ for carcinogens. Crosswalks between chemical sets (DMR and TRI) allowed EPA to apply toxicity weights and decay rates of TRI chemicals to matching DMR chemicals, so that DMR data could be incorporated into the RSEI model.

The new indicator approach ran the standard RSEI modeling data processing procedures. EPA extracted DMR and TRI data from the DMR Pollutant Loading Tool and structured the data to closely mimic the standard RSEI input tables. EPA aimed to minimize the adjustments needed to accommodate DMR data in the RSEI modeling procedures.

The new water indicator calculates proximity as the distance between the centroid of the Census block to the midpoint of any stream reaches within a 500-meter radius of the block centroid. If no reaches are

found within 500 meters, the nearest neighbor within 3 km is used. The indicator uses the output from RSEI, to give more emphasis to stream reaches with higher toxicity-weighted pollutant concentrations.

$$f(d_{ij}, W) = \frac{W}{d_{ij}}$$

Where

W is the weighting factor (described below) and $W > 0$;

d_{ij} is the distance, in km, from the Census block centroid, *j*, to the midpoint of the stream reach, *i*; and,

$f(d_{ij}, W)$ is the weighted proximity function.

The resulting block values were aggregated to Census block groups using a population-weighted method to produce block group level results. This is the same aggregation method used for the other EJScreen proximity indicators.

Calculating Proximity – Additional Details

Proximity Calculations—block group proximity scores for Traffic, Superfund, RMP, Hazardous Waste, and Wastewater Discharges are calculated with a new source. Census block centroids with population-weights were updated to new 2020 version. These block points and associated populations and block group weights were derived from 2020 Decennial Census P.L. 94-171 Redistricting data.

We address two modifications to the general method described above. The first deals with instances where a facility or highway segment location is very close to the centroid of the block. The second is an accommodation to the computational intensity of the general method.

Extremely Small d_{ij} Values

Our intention is to represent the proximity of facilities or highway segments to the population within each block. All facilities and each part of all highway segments fall within one block. By chance, some portion of those points fall very close to the block centroids.

We do not know how the population is geographically distributed within any block, but we assume that people are more likely to be distributed across the blocks' expanses than to be concentrated at one point, such as the centroid. In fact, for rural, suburban, and many non-high rise urban areas, people's residences are more likely to be closer to the blocks' peripheries (bounded by roads) than clustered at the centroids. Thus, when a facility location happens to be very close to the block centroid, it would result in an artificially high contribution to the block's score. This is not a hypothetical problem: We have observed d_{ij} values well below 100 meters, and some below 10 meters.

In looking for solutions to the problem, we conducted analyses and arrived at the approach we have adopted. Blocks vary widely in their total area and in their shapes. Both can be found in the Census

Bureau's Tiger shape files. Dealing explicitly with the individual block shapes would be computationally very intensive because there are over 11 million blocks. Since we cannot easily find out how the residents are actually distributed in those areas, we made two simplifying assumptions:

- residents are evenly distributed across the surface area of each block, and
- each block can be represented by a circle whose radius is [Block area / Pi]1/2 .

We call this latter value the Block Area Equivalent Radius.

Our investigations indicate that for any d_{ij} less than the Block Area Equivalent Radius, 0.9 times that value is a reasonable representation of the average distance from the facility for all residents in the block. We call this the d_{ij} corrected.

Our computational scheme determines the d_{ij} values as described above, tests for the comparison with Block Area Equivalent Radius, and substitutes d_{ij} corrected values. We found that we needed to make that correction for less than 1% of all facility/block combinations in an early testing dataset that used 2005-2009 ACS data.

Accommodating to Computational Intensity – Combine a Distance Limit with a Nearest Facility Approach

Our task is to compute a proximity score for each of the facility or site types and highway segments for each of the more than 235,000 block groups, composed of over 8 million blocks. The number of facilities nation-wide varies from hundreds of TSDFs to many thousands of RMP facilities. Computing all the combinations would require more computational time and resources than were available.

In addition, doing so would be wasteful and perhaps irrelevant. The one over distance function we have chosen to represent concerns about facilities and highways drops off greatly for most facilities beyond the nearest ones. The miniscule contribution of a facility 100 km or more from a block is not only small, compared with those that may be within 5 to 10 km, but has little common-sense meaning, in our view.

Consequently, we have followed the general approach described above only for facilities or sites within 5 km of a block's centroid, and within 500 meters for highway segments. Depending on the facility or site type, we find that 30-40% of block groups have at least one facility (RMP or TSDF) within the 5-km limit, and almost 10% have one or more NPL sites within 5 km, in EJScreen.

Of course, every block and block group has one nearest facility, even though it may be beyond the 5-km horizon, and some of those may be fairly close to that limit. We have also calculated the distance to the facility nearest to each of the blocks. For those blocks lacking anything within the 5 km, we represented the facility proximity by one over the distance to that single nearest facility.

This added computational complexity to the approach, but at far less cost than computing the full matrix of millions of blocks times thousands of facilities and sites.

This hybrid approach results in every block (and thus every block group) having a nonzero, positive proximity score. All of the resulting block proximity scores are necessarily less than the score had we computed the full matrix, but we judge that this is a reasonable and practical compromise. Counting only the single nearest facility beyond 5 km has the effect of shifting scores under 0.2 to the left, to lower scores than if all were counted, but the graphs show no major discontinuities, suggesting this limitation (counting only the nearest one) has little impact overall.

Data and Computational Scheme

Using the Census block centroids, the distance to all facilities within 5 km of all blocks (not just block groups) was determined, and distance to the nearest facility at any distance was determined if none were found within 5 km.

The d_{ij} values were compared to the Block Area Equivalent Radius and corrected values were used when necessary, before computing $1 / d_{ij}$. The $1 / d_{ij}$ values were summed for each block to compute the $BlockScore_{jk}$. These were then rolled up to the block group level, applying the population weighting described above, for the final $BlockGroupScore_{k}$.

Caveats and Observations

Several aspects of the proximity analysis approach have been mentioned above, but deserve summary here.

- We recognize that our selection of the inverse of distance is a design choice that represents our judgment of a balance among competing factors.
- We recognize that one could potentially attempt to distinguish among facilities within each facility category by quantitative or qualitative measures of importance. These could include total pounds released or toxicity-weighted releases for NPDES facilities; the number of accidental releases and/or their apparent severity for RMP facilities; some classification of the likelihood of releases for NPL sites or TSDFs; and general indications of scale for all of them. We note that CalEnviroscreen has addressed this issue to some extent, and that the RSEI tool based on TRI data may be relevant to future work on this issue. At this point, we have chosen not to develop any such potential scaling adjustments.
- We recognize that all location data are subject to potential error. While we have high confidence in the block centroid locations, we know that the facility or site or roadway location data may contain larger or smaller errors, and that for large facilities or sites, one point may not be an entirely adequate representation of the location of its releases or of neighbors' perceptions.
- We recognize that the computational accommodation we describe above results in a hybrid of measures: For some block groups, all blocks have one or more facilities within 5 km and the score is the summation of all those potentially multiple facility/block combinations; for other block groups, none of the blocks have a facility within 5 km and the score is the contribution of

the single facility closest to each block; and for some block groups, we have a mix of those situations. We believe that this is a reasonable compromise.

8 Details on U.S. Territories

EJScreen features select environmental and socioeconomic indicators for the following U.S. Territories: American Samoa, CNMI, Guam, Puerto Rico, and the U.S. Virgin Islands. American Samoa, CNMI, and Guam use 2013 Census Place boundaries. The U.S. Virgin Islands use 2013 Census Estate boundaries. However, Puerto Rico is included with all 50 states and the District of Columbia. The term "territories" in this document refers to the four territories excluding Puerto Rico. The territories data for American Samoa, CNMI, Guam, and the U.S. Virgin Islands provides an additional 605 records. These territories have State percentiles data and do not have National percentiles.

Data Sources

- Census boundaries (Places, Estates, and Counties) are from Cartographic Boundary files, 2013, 1:500K scale.
- Demographic data—2010 Demographic Profiles of Island Areas published in 2014.

Processing Steps

- The block population weight table was generated based on distributing 2010 population for Estate or Place boundaries
- Build demographic indicator tables by extracting the Place/Estate summary level data from the Demographic Profiles. Indicators include: % people of color; % low income; % limited English speaking; % less than high school education; % under age 5; and % over age 64
- Calculate Demographic Index: (% people of color + % low income) / 2
- Collect all the available environmental data sources
- Calculate EJ Indexes.

The EJ Index Availability for the Territories

	American Samoa	CNMI	Guam	U.S. Virgin Islands
PM 2.5	No	No	No	No
Ozone	No	No	No	No
Diesel PM	No	No	No	No
Air Toxics Cancer Risk	No	No	No	No
Air Toxics Respiratory HI	No	No	No	No
Traffic Proximity	No	No	No	No
Lead Paint	*Yes*	*Yes*	*Yes*	*Yes*
Superfund Proximity	No	No	*Yes*	*Yes*
RMP Facility Proximity	*Yes*	No	*Yes*	*Yes*
Hazardous Waste Proximity	No	No	*Yes*	*Yes*
Underground Storage Tanks	*Yes*	*Yes*	*Yes*	*Yes*
Wastewater Discharge	No	No	No	No

Notes about Puerto Rico

- While Puerto Rico is a U.S. Territory, it is included in the Census ACS 5-year Summary datasets, so EJScreen treats it like a state.
- It has both National and State-level percentiles and indexes.

9 Other Data Element Descriptions

This section describes additional datasets now available in EJScreen. They include Health Disparities (Low Life Expectancy, Heart Disease, and Asthma); Climate Change-related (Wildfire Risk and Flood Risk); Sensitive Communities (Colonias from U.S. Department of Housing and Urban Development [HUD], Texas, and New Mexico); and Critical Service Gaps (broadband service gaps, food deserts, and medically underserved areas).

Health Disparities Data

- **Low Life Expectancy**—Average life expectancy data developed as a collaboration between NCHS, the National Association for Public Health Statistics and Information Systems (NAPHSIS), and the Robert Wood Johnson Foundation. This data is available at the tract level. Source: U.S. Small-area Life Expectancy Estimates Project (USALEEP)
  It is derived from Life Expectancy at Birth from CDC, National Center for Health Statistics using the formula of % Low Life Expectancy is defined as "1 – (Life Expectancy / Max Life Expectancy)". The following processing steps were used to bring the data into EJScreen:
  - The source is Census 2015 Tract-level data, so it was first converted to 2010 tracts using the Census 2010 to 2015 relationship table.
  - 2010 tracts were then converted to 2020 tracts using the Census 2010 to 2020 relationship table.
  - Low Life Expectancy values were assigned to each child 2020 block group within the same Census tract.
- **Heart Disease**—Heart disease prevalence among adults aged 18 years or older. The term "heart disease" refers to several types of heart conditions. This data is available at the tract level. Source: CDC Places Data
- **Asthma**—Asthma prevalence among adults aged 18 or older. This data is available at the tract level. Source: CDC Places Data

Wildfire and Flood Risk Data from First Street

The First Street Foundation is partnering with EPA to provide climate risk data to EJScreen. First Street provided Census block group-level data, including percent of properties at risk of being affected by wildfires or by flooding based on current climate conditions as well as conditions projected to exist in 30 years. EJScreen merged the data with current EJScreen spatial data and generated national and state percentiles. EJScreen includes:

- **Wildfire Risk**—The household risk of wildfire exposure under 2022 weather conditions as modeled by the First Street Foundation. Source: https://firststreet.org/
- **Flood Risk**—The household risk of flooding under 2022 weather conditions as modeled by the First Street Foundation. Source: https://firststreet.org/

The First Street Foundation-Wildfire Model (FSF-WFM) is a 30-meter resolution model representing the wildfire exposure for any specific location in the contiguous U.S., today and with the future climate change. The risk of wildfire is derived from a series of inputs associated with fire fuels, weather, human influence, and fire movement. Bringing all of these inputs together, at a national scale, in a high-resolution, climate-adjusted model represents a first-of-its-kind property-level wildfire risk model.

The First Street Foundation Flood Model is a nationwide probabilistic flood model that shows the risk of flooding at any location in all 50 states and Puerto Rico due to rainfall (pluvial), riverine flooding (fluvial), and coastal surge flooding. While other hydraulic and hydrologic models show refined risks of flooding in certain areas, this model provides complete coverage across the United States at 3-meter resolution.