# Draft template for or Outline of a standard EJ analysis

## Abstract

## Executive Summary

## Introduction (context, purpose, what is this)

## Methods

### Scope (what facilities/points/places, etc.)

### Identifying populations at each site

*Spatial resolution of data – explain, for data on residential locations, demographics, and environmental indicators.*

*Analytic method for buffering, and tools used to implement that method – The analysis used EJAM, or etc., for buffering (identifying locations to analyze and residents to characterize). e.g. Same as EJScreen in principle, i.e., Percent of each block group’s population that is estimated as inside a buffer is based on which Census block internal points are included in the buffer and using a block weight that is the C2020 block pop as a fraction of the parent block group’s C2020 pop (not the ACS pop). That block weight is a fraction of the parent block group. BUT may use a different method to identify which block points are inside the buffer than EJScreen does.*

### Summarizing within each buffer

(e.g., population weighted mean vs counts and formulas applied to sums of counts; percentiles; rounding?; etc.)

### Data sources and vintages

#### Demographics

EJScreen and user defined and ? – what years, how to cite? block group data from ACS 5-yr summary file dated \_\_\_ and possibly users, etc.

Asdf

Asdf

Asdf

asdf

#### Environmental indicators

EJScreen and user defined and ? – what years, how to cite?

Asdf

Asdf

Asdf

adsf

#### Data on locations of residents

EJScreen and user defined and ? – what years, how to cite? Census 2020 block points, or other data used to estimate where residents are within each block group. Or raster-based, dasymetric, whatever.

#### Other data

## Results – Basic information about the locations analyzed and number of people nearby

* *Explain whether the sites are to be analyzed as a whole or in some type of subgroups, such as 2 different source categories, or large vs small facilities, or some other categories we will use to compare all these stats?*
* Count of locations
* Locations missing data
* Where are the sites (regions, states, cities?, urban/rural?)
* Clustering (are they near each other? How close? Which ones are in clusters, maybe?)
* Total population near any of the sites (count of unique residents) and some summary of site-specific Population sizes and population density nearby (what % of all people are at x% of the sites? Range and Interquartile range of pop counts nearby? etc.). Make clear to what extent some people in the overall summary stats are near more than a single one of these sites.

## Results – Demographics of populations in these locations

*Side note, on statistical significance, for demographic and environmental findings: if a small number of sites is analyzed, like <100, do we care about statistical significance of differing from US averages? That would tell us whether this might have arisen by chance in this sector if they were randomly located in the US, but we do not really care about that question here, I would argue – if three sites are in this rule and all three are in low-income areas, we would not ignore that fact when looking at the rule. We do not care about why this rule will affect vulnerable populations – we only care whether it will in fact affect vulnerable populations*.

### Demographic stats on the overall population

#### Overall summary statement of some kind

#### Key important D summary indicator(s) based on some policy or default

* stats on EJScreen’s Demographic Indicator.
* stats on user-specified Demographic Indicator(s) ?

#### Key D group(s) based on policy or default

* Sdf
* Asdf
* Asdf
* asdf

#### Key D group(s) based on observed magnitude of disparities here

Largest disparity in presence of any group – which group was most over-represented here?

* Relative disparity vs US: Which group(s) had the largest ratio of local % (for the overall set of sites) to US % overall?
* Absolute disparity vs US: which group(s) had

“Large” disparities: which groups were “very” over-represented here?

## Results – Environmental conditions in these locations

Overall summary statement of some kind. E.g., a “\_\_\_very large” number of all the (12?) envt indicators were “\_\_\_very high” at a “\_\_\_very large” share of the sites or preferably for a “\_\_\_very large share of the people”? Or, how many of 12 indicators were “high” for the average person and/or site overall?

Which environmental …

## Results – Combination of demographic and environmental conditions in these locations

Summary of Envt and Demog across all indicators? In other words, a single score like the combination of 12 EJ indexes as a threshold approach summary? e.g. the average person had 5 or more of the 12 EJ indexes at least at the 95th percentile nationwide?

Individual EJ indicator of 12 that generally had highest overall percentiles? On average highest vs the one where at least a few sites had a very high number? Other?

User specified EJ index(es)?

## Appendices

### Appendix 1

**DETAILED TABLES, STATS FOR EACH SITE**

**e.g. how many of 12 are >x? does it have any high scores for any E? any D? etc.**

### Appendix 2

**DETAILED TABLES. STATS FOR EACH INDICATOR** (demographic, environmental, and other like pop, EJ indexes??, etc.):

* **What is Score of people/sites at key cutoffs: What is the X (useful) percentile value, for this indicator (as %ile of people nearby or of sites)?**
  + min
  + min other than zero?
  + mean
  + 25th %ile of these sites or people (25% have lower than this, and 75% have higher than this – those are same if no ties with this value, but can differ if multiple places have this same exact indicator score)
  + median (half of these sites or people have a score that is >= this, and half have <= this)
  + 75th %ile of these sites or people (if a tied value, would want both >75 and <25%)
  + max
* **How many people/sites have score > key cutoffs: What % & what # of sites & people have this indicator score (raw or percentile) >= x (useful cutoff)?**
  + For percentiles
    - pctile 80.0 % of sites w data where >=80
    - pctile 90.0 % of sites w data where >=90
    - pctile 95.0 % of sites w data where >=95
    - pctile 95.0 # of sites where >=95
  + for ratios
    - %of sites w data where >1 not=1
    - ratios 2.0 % of sites w data where >=2
    - ratios 3.0 % of sites w data where >=3
    - ratios 5.0 % of sites w data where >=5
    - ratios 10.0 % of sites w data where >=10
    - ratios 10.0 # of sites w data where >=10
  + for 12 EJ indexes
    - n/12 1.0 # of sites w data where >=1
    - n/12 1.0 # of sites w data where >1 not=1
    - n/12 4.0 # of sites w data where >=4
    - n/12 7.0 # of sites w data where >=7

### Appendix 3

# HOW TO REFER TO THE PLACES STUDIED (near these facilities vs more general language)

Note: this could be a “proximity analysis” in cases where it relies on circular buffers around facility points to define buffers that include residents within a fixed distance from one or more facilities/sites. But it more generally could be an EJ analysis that describes environmental and demographic conditions among residents in any specified places, such as all the places where air quality modeling suggests risk is currently above 1 in 1 million, for example. So the language should be flexible and refer to something like this:

FOR NON-PROXIMITY ANALYSIS, GENERALLY ANY KIND OF BUFFERS/PLACES ANALYZED:

The demographics of residents in …

The demographic / environmental indicators in …

Percent low income among residents in …

The environmental conditions in …

The environmental indicators for the average resident in …

The PM2.5 levels in …

Residents in …

… these locations

… these places

… these areas

… the analyzed locations

FOR PROXIMITY ANALYSIS SPECIFICALLY, EASIER TO SAY one of these:

Residents / conditions …

… within x miles of these sites…

… within x miles of any of these sites…

… nearby

… near these sites

… near any of these sites

# Notes on methods text

|  |
| --- |
| **Methods** |
|  |
| These tables include polymers and resins 1 and 2. |
| The starting point was a spreadsheet of lat/lon and facility name/ID information was Chem Sector Facility List\_HON\_PEPO\_PR I\_ PR II\_07-14-22 |
| EJ stats here are based on ejscreenapi shiny R app for EJAM that uses the EJScreen API, obtained ca. 8/16/22. |
| This is based on using version 2.0 EJScreen. Version 2.1 will be out by early September 2022. |
| The tool was used to get a report on conditions within 1 or 3 miles of each point. |
| There were several sites where EJScreen did not return any stats, typically (or always) because of low population density nearby (within 1 mile in this case). |
| Summary stats were added to the raw outputs of that tool, using Excel in this case. |
|  |
| Note the summary stats focus on summarizing conditions at each site, mostly (i.e., within one buffer). |
| A few summary stats look across these facilities, but unfortunately one cannot summarize conditions for the |
| average person near one or more sites overall here. As explained here, this dataset cannot account for |
| double-counting people who live near multiple sites: |
| <https://rstudio-connect.dmap-stage.aws.epa.gov/content/374e6403-c660-4692-b62f-61139d1fef69/_w_cb386ee3/README.html> |
| A GIS analysis using ArcPro or the OW or OAR EJ tool can do that, and EJAM will also be able to when it launches. |
|  |
| **Spreadsheet** |
|  |
| There are some maps and graphics on added tabs, and the data is all in a tab called EJ stats. |
| You can sort on any column(s) or use Filtering to find or filter which sites are shown and in what order. |
| The spreadsheet has many hidden columns, so not all of the raw data in particular has been shown. |
| Color coding is used to create heatmaps where colors of percentile scores match colors used in EJScreen maps (80-90 yellow, 90-95 orange, 95+ red). |
| Ratio are color coded to highlight larger ratios, such as >1x, 2x, or 3x the US average, and <1/2 the average (blue). |
| Some notable columns are indicated by red highlighting above the header. |
|  |
| The spreadsheet of EJ stats has these groups of columns: |
|  |
| 1. All the data from the original site list (e.g., facility IDs, names, etc.) |
|  |
| 2. all the demographic, environmental, and EJ Index indicators from an EJScreen standard report |
| for a buffer analysis of conditions within 1 mile of each site, including: |
| raw numbers (e.g., population count, NATA cancer risk, % low income, etc.) |
| those expressed as percentiles (e.g., 75th percentile nationwide or in state for given indicator, as in EJScreen standard reports) |
| national and state averages, for reference. |
|  |
| 3. Added summary stats such as |
| how many indicators are above a given threshold |
| the max (highest) number of a given type found at that site, |
| ratio of indicator score near the site to indicator score in the US overall (e.g., 1.5 means it is 1.5 times the US number), |
| summary numbers showing how high a percentile can be seen for at least N of the 12 Envt indicators and also in the overall Demog indicator at that site |
| (N can be customized by changing the number above the table). |
| There are also links in the last 3 columns that lead to more detailed reports on individual sites. |
| Formulas and conditional formatting were used for summary stats, but formulas were replaced with values, and formulas saved in a single row above the table. |
|  |
|  |