

Final Project EDA

Final Project Check In

We are using the 2022 Environmental Justice Index dataset published by the [CDC](#) with data from the U.S. Census Bureau, the U.S. Environmental Protection Agency, the U.S. Mine Safety and Health Administration, and the U.S. Centers for Disease Control and Prevention. The data is filtered to Massachusetts (1474 observations) where each row is a census tract.

Question: Which pollutants correspond with higher rates of asthma in Massachusetts? Does this differ by poverty rates? (/percentage of people living under 200% of the federal poverty level?)

Data Dictionary for EDA

https://ej.cdc.gov/Documents/Data/2022/EJI_2022_Data_Dictionary_508.pdf

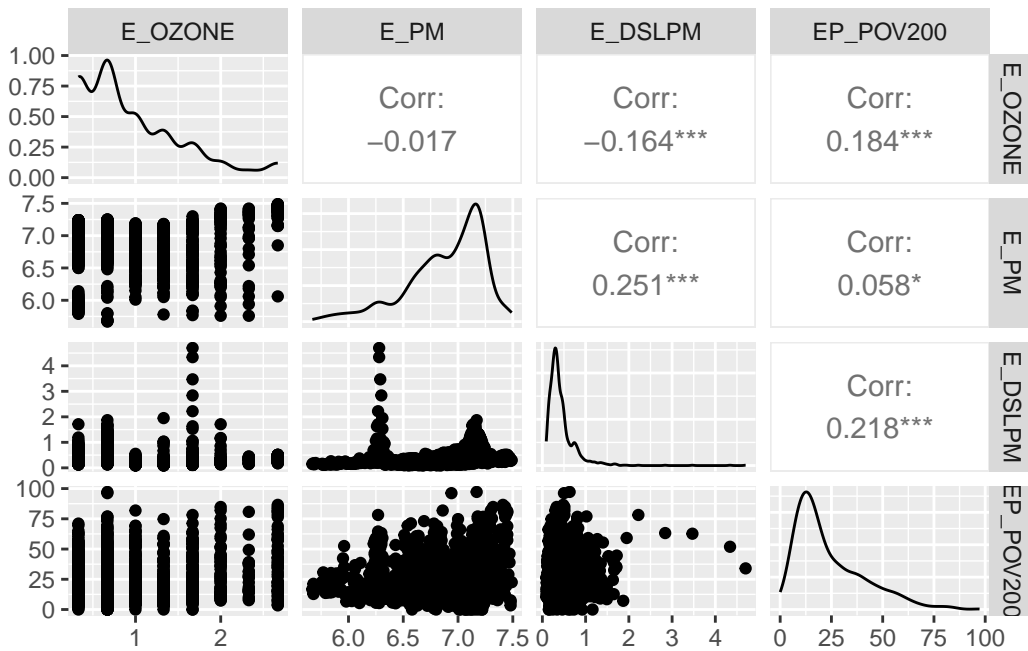
Data to Download

<https://www.atsdr.cdc.gov/placeandhealth/eji/eji-data-download.html>

```
us <- read.csv("United States.csv") |>
  dplyr::filter(StateDesc == "Massachusetts") |>
  dplyr::filter(!is.na(E_OZONE & E_PM & E_DSLPM & EP_ASTHMA & EP_POV200))

us |>
  dplyr::select(E_OZONE, E_PM, E_DSLPM, EP_POV200) |>
  GGally::ggpairs()
```

```
Registered S3 method overwritten by 'GGally':
  method from
+.gg      ggplot2
```



```
library(ggplot2)
ggplot(data = us, aes(x = E_OZONE, y = EP_ASTHMA)) +
  geom_point(position = "jitter") +
  labs(title = "EDA Visual #2", x = "Mean Days Above Ozone Standard", y = "Percentage with")
theme_minimal()
```

Warning: Removed 11 rows containing missing values (`geom_point()`).

EDA Visual #2

