Final Project EDA

Lindsay Kowal & Kathleen Bacigalupi

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?)

Our response variable will be the percentage of individuals that have asthma (EP_ASTHMA). Our explanatory variables will be the percentage below 200% poverty (EP_POV200), the annual mean days above O3 standard (E_OZONE), the annual mean days above PM 2.5 regulatory standard (E_PM), and the ambient concentrations of diesel (E_DSLPM).

Data Dictionary for EDA

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

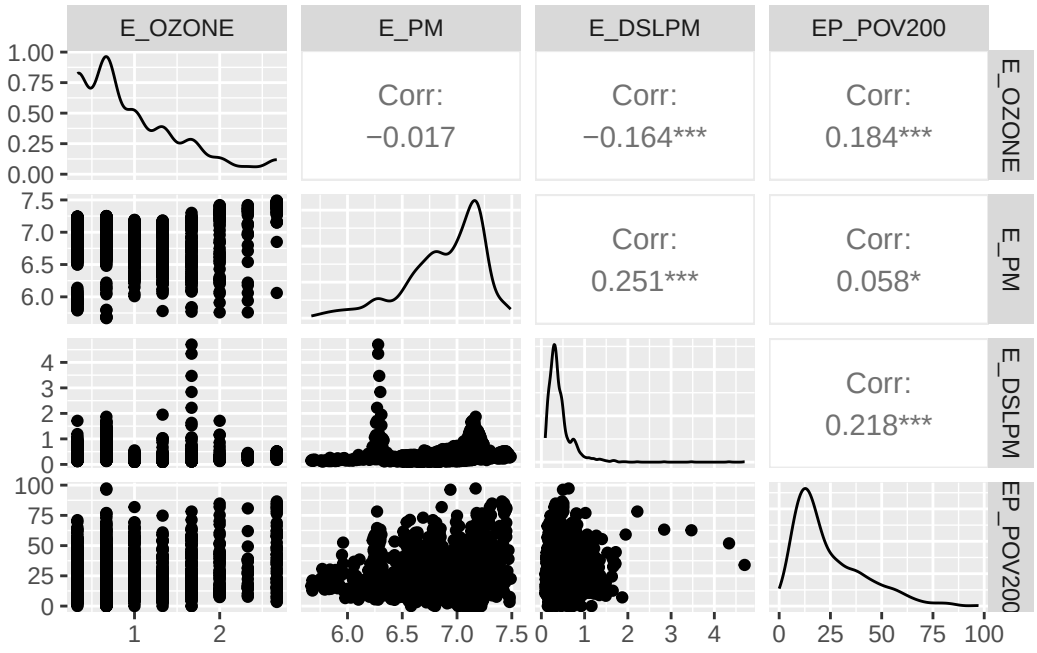
Data to Download

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))
```

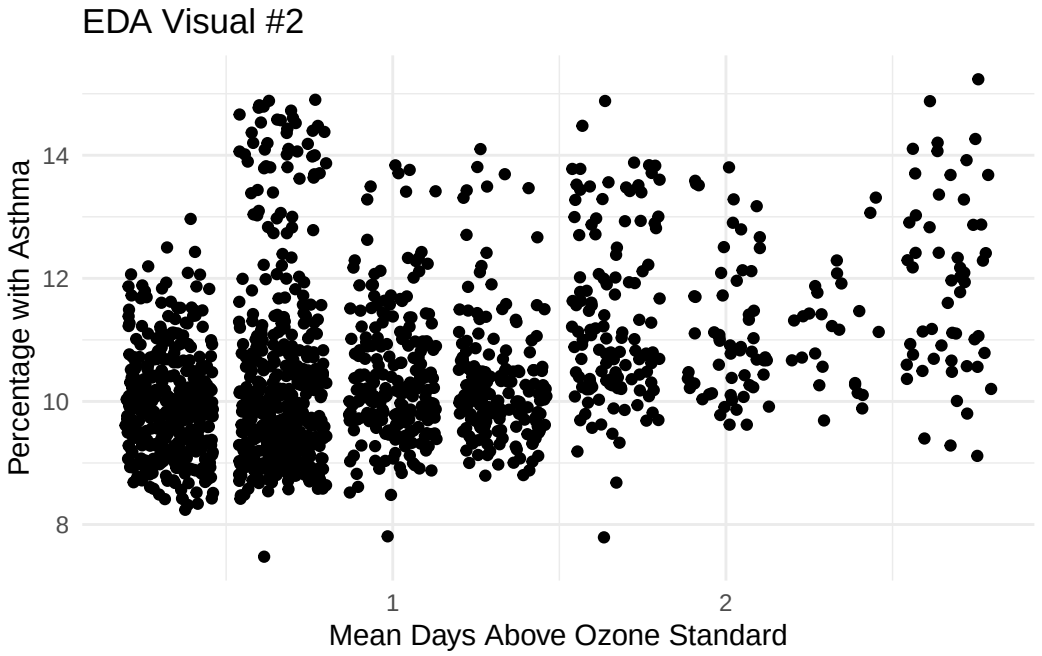
```
us |>
  dplyr::select(E_OZONE, E_PM, E_DSLPM, EP_POV200) |>
  GGally::ggpairs()
```

```
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()`).
```



Formatting Equations

$$\log(\frac{\hat{\pi}}{1-\hat{\pi}}) = \beta_0 + \beta_1(age_i) + \beta_2(sex_i) + \beta_3(age_i)(sex_i) + \beta_4(age_i^2) + \beta_5(age_i^2)(sex_i) \quad (1)$$

Then can reference equation 1) (no reference number with * after equation/equation)

New line with \, &'s are where the lines go

$$\begin{aligned}\log(\frac{\hat{\pi}}{1-\hat{\pi}}) =& \beta_0 + \beta_1(age_i) + \beta_2(sex_i) + \beta_3(age_i)(sex_i) \\ &+ \beta_4(age_i^2) + \beta_5(age_i^2)(sex_i)\end{aligned}$$

Then can reference equations as well: (1)

To make curly braces and or parenthsis large:

$$\begin{aligned}\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) =& \beta_0 + \beta_1(age_i) + \beta_2(sex_i) + \beta_3(age_i)(sex_i) \\ &+ \beta_4(age_i^2) + \beta_5(age_i^2)(sex_i)\end{aligned}$$

Curly braces:

$$\log\left\{\frac{\hat{\pi}}{1-\hat{\pi}}\right\} = \beta_0 + \beta_1(age_i) + \beta_2(sex_i) + \beta_3(age_i)(sex_i) + \beta_4(age_i^2) + \beta_5(age_i^2)(sex_i)$$