

Particulate Matter (PM2.5): State

Ricky Truong

June 2025

Load libraries and data

```
# Delete everything in environment
rm(list = ls())

# Load libraries
library(tidyverse)
library(readxl)
library(readr)

# Load EIA Total CO2 Emission, 1970-2022
co2_1970_2022 <- read_excel("CO2.xlsx")

# Load EPA State PM2.5 Emission, 1990-2024
pm2.5_state_1990_2024 <- read_excel("state_tier1_21feb2005_ktons.xlsx",
                                     sheet = "State Trends")
```

EIA total CO2 emission, 1970-2022

Wrangle data

```
# Delete unnecessary last four columns
co2_1970_2022 <- subset(co2_1970_2022, select = -c(...55 : ...58))

# Rename columns/variables using the fourth row
names(co2_1970_2022) <- as.character(unlist(co2_1970_2022[4,]))

# Delete unnecessary rows
co2_1970_2022 <- co2_1970_2022[-c(1:4, 57),]

# Select states and years only
co2_1970_2022 <- co2_1970_2022 %>%
  select(`1970`:`2022`, State)

# Convert data to long/tidy format
```

```

co2_1970_2022 <- co2_1970_2022 %>%
  pivot_longer(cols = -State, names_to = "Year", values_to = "CO2") %>%
  mutate(Year = as.integer(Year),
         CO2 = as.numeric(CO2))

# Use state abbreviations for consistency ("Total of states" becomes NA)
co2_1970_2022$State <- state.abb[match(co2_1970_2022$State, state.name)]

```

EPA state PM2.5 emission, 1990-2024

Wrangle data

```

# Rename columns/variables using the first row
names(pm2.5_state_1990_2024) <- as.character(unlist(pm2.5_state_1990_2024[1,]))
pm2.5_state_1990_2024 <- pm2.5_state_1990_2024[-1,]

# Filter for PM2.5 only
pm2.5_state_1990_2024 <- pm2.5_state_1990_2024 %>%
  filter(Pollutant == "PM25-PRI")

# Select state and emissions columns only
emission_cols <- grep("^emissions\\d{4}$", names(pm2.5_state_1990_2024), value = TRUE)
pm2.5_state_1990_2024 <- pm2.5_state_1990_2024 %>%
  select(State, all_of(emission_cols))

# Convert data to long/tidy format (with Year, State, Emissions for each source)
pm2.5_state_1990_2024 <- pm2.5_state_1990_2024 %>%
  pivot_longer(cols = starts_with("emissions"),
               names_to = "Year",
               names_prefix = "emissions",
               values_to = "PM2.5") %>%
  mutate(Year = as.integer(Year),
         PM2.5 = as.numeric(PM2.5))

# Group by Year and State to get total PM2.5 per state and year
pm2.5_state_1990_2024 <- pm2.5_state_1990_2024 %>%
  group_by(Year, State) %>%
  summarize(PM2.5 = sum(PM2.5, na.rm = TRUE), .groups = "drop") %>%
  select(State, Year, PM2.5) %>%
  arrange(State)

```

Plot total across states per year

```

pm2.5_state_1990_2024_total <- pm2.5_state_1990_2024 %>%
  group_by(Year) %>%
  summarize(total = sum(PM2.5, na.rm = FALSE)) %>%

```

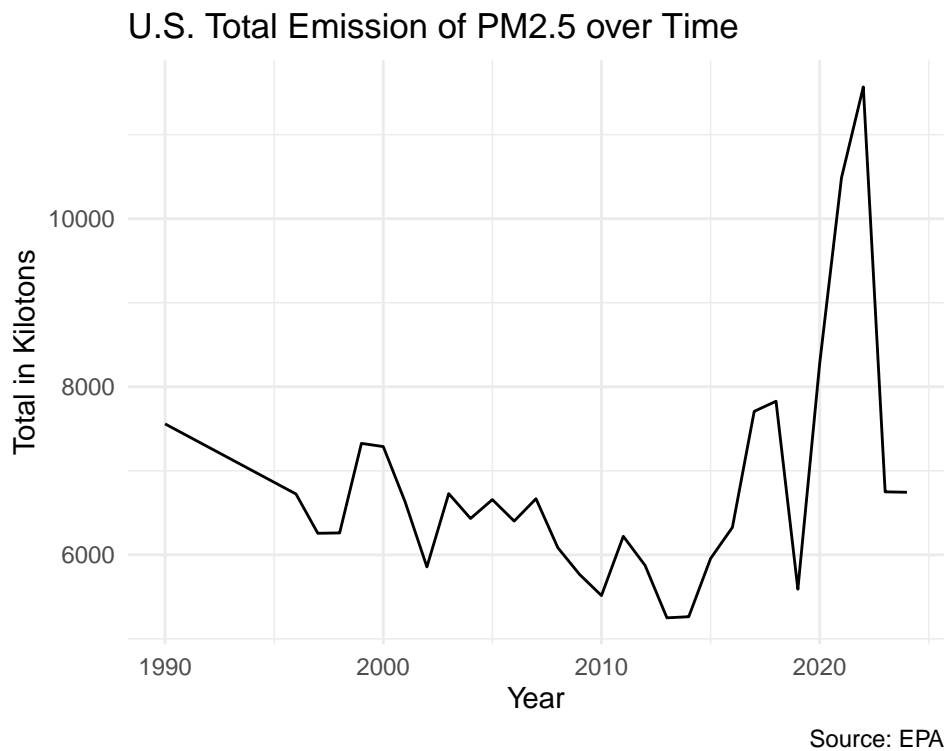
```

arrange(Year)

ggplot(pm2.5_state_1990_2024_total, aes(x = Year,
                                         y = total)) +

  geom_line() +
  labs(x = "Year",
       y = "Total in Kilotons",
       title = "U.S. Total Emission of PM2.5 over Time",
       caption = "Source: EPA") +
  theme_minimal()

```



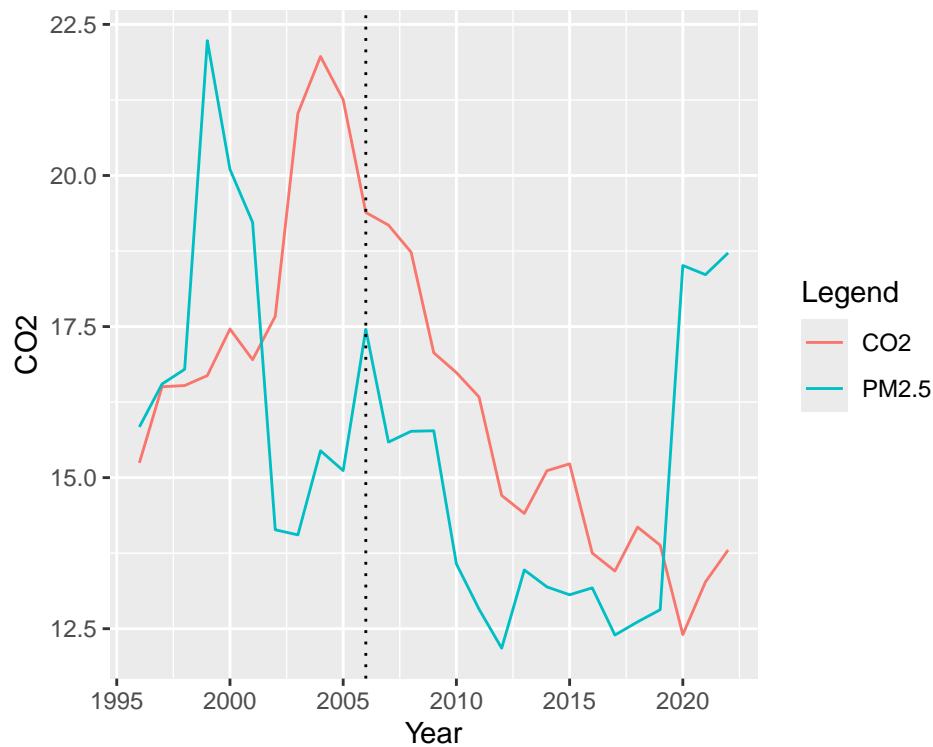
Plot CO2 emissions and PM2.5 emissions in NH

```

nh <- full_join(co2_1970_2022, pm2.5_state_1990_2024,
                by = c("State", "Year")) %>%
  filter(State == "NH", Year >= 1996, Year <= 2022)

ggplot(nh, aes(x = Year)) +
  geom_line(aes(y = CO2, color = "CO2")) +
  geom_line(aes(y = PM2.5, color = "PM2.5")) +
  geom_vline(xintercept = 2006, linetype = "dotted") +
  labs(color = "Legend")

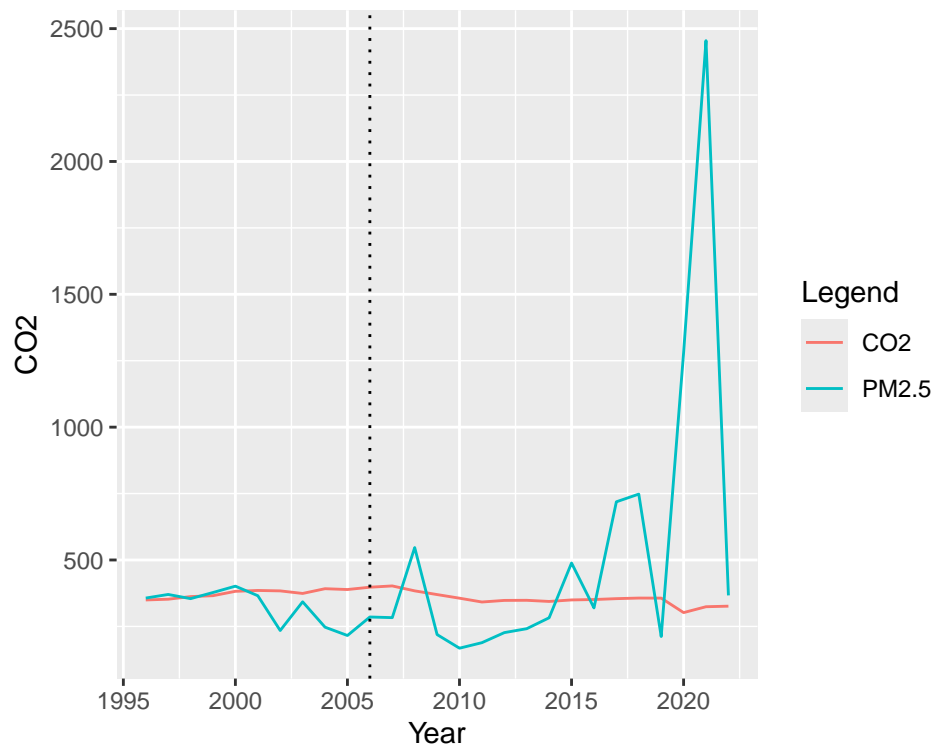
```



Plot CO2 emissions and PM2.5 emissions in CA

```
ca <- full_join(co2_1970_2022, pm2.5_state_1990_2024,
               by = c("State", "Year")) %>%
  filter(State == "CA", Year >= 1996, Year <= 2022)

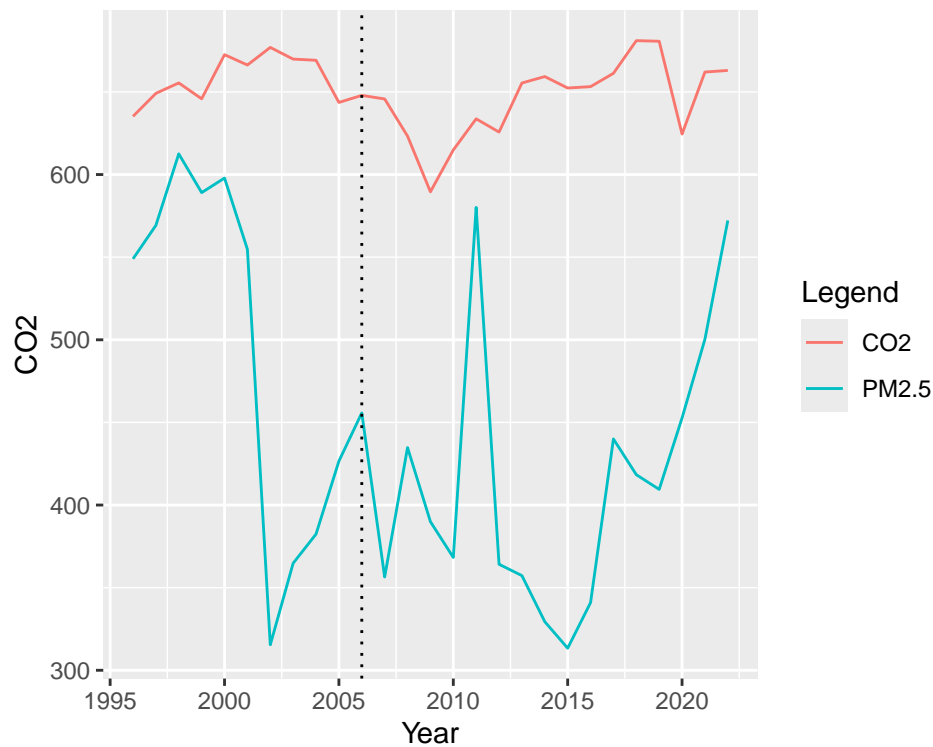
ggplot(ca, aes(x = Year)) +
  geom_line(aes(y = CO2, color = "CO2")) +
  geom_line(aes(y = PM2.5, color = "PM2.5")) +
  geom_vline(xintercept = 2006, linetype = "dotted") +
  labs(color = "Legend")
```



Plot CO2 emissions and PM2.5 emissions in TX

```
tx <- full_join(co2_1970_2022, pm2.5_state_1990_2024,
               by = c("State", "Year")) %>%
  filter(State == "TX", Year >= 1996, Year <= 2022)

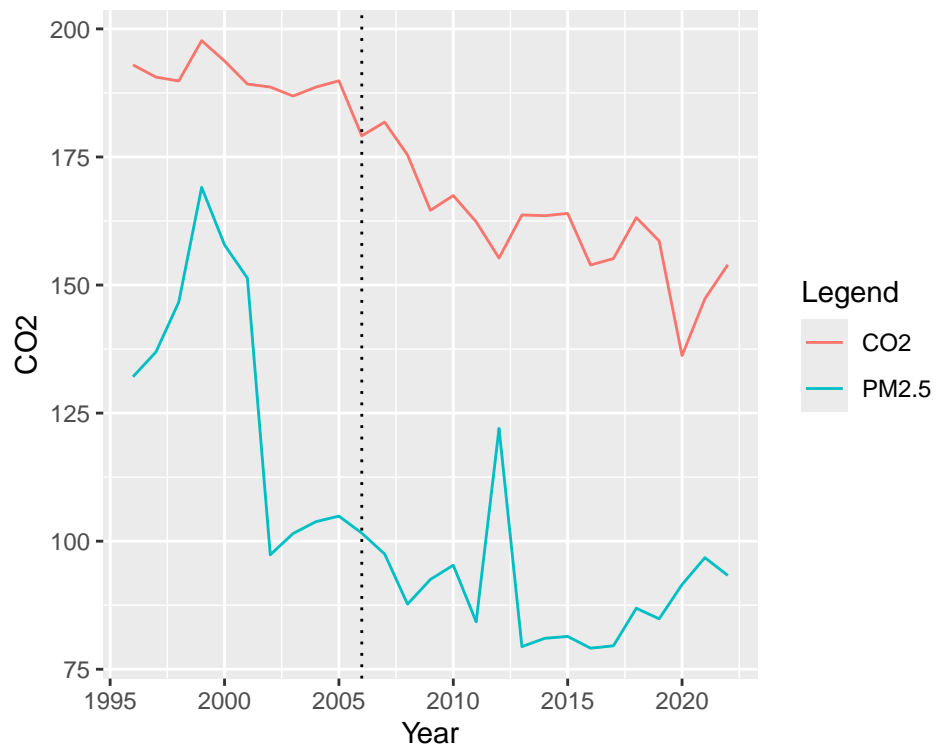
ggplot(tx, aes(x = Year)) +
  geom_line(aes(y = CO2, color = "CO2")) +
  geom_line(aes(y = PM2.5, color = "PM2.5")) +
  geom_vline(xintercept = 2006, linetype = "dotted") +
  labs(color = "Legend")
```



Plot CO2 emissions and PM2.5 emissions in MI

```
mi <- full_join(co2_1970_2022, pm2.5_state_1990_2024,
               by = c("State", "Year")) %>%
  filter(State == "MI", Year >= 1996, Year <= 2022)

ggplot(mi, aes(x = Year)) +
  geom_line(aes(y = CO2, color = "CO2")) +
  geom_line(aes(y = PM2.5, color = "PM2.5")) +
  geom_vline(xintercept = 2006, linetype = "dotted") +
  labs(color = "Legend")
```



Plot CO2 emissions and PM2.5 emissions in HI

```
hi <- full_join(co2_1970_2022, pm2.5_state_1990_2024,
               by = c("State", "Year")) %>%
  filter(State == "HI", Year >= 1996, Year <= 2022)

ggplot(hi, aes(x = Year)) +
  geom_line(aes(y = CO2, color = "CO2")) +
  geom_line(aes(y = PM2.5, color = "PM2.5")) +
  geom_vline(xintercept = 2006, linetype = "dotted") +
  labs(color = "Legend")
```

