data cleaning

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
library(haven)

setwd("~/Desktop/STAT 156 Project")

# Step 1: Load and filter the main dataset
data <- read_dta("AER20090377_FinalData.dta") %>%
  filter(year < 2004, valid >= 9, month %in% c(6, 7, 8))  # Filter year, valid days, and summer months
```

```
# Step 2: Keep monitor-years with at least 69 valid days
summer_list <- data %>%
  group_by(fips, site_id, year) %>%
  mutate(NumDays = n()) %>%  # Count valid days
  filter(NumDays >= 69) %>%  # Keep monitor-years with >= 69 days
  select(fips, site_id, year) %>%
  distinct()  # Keep unique rows
write_dta(summer_list, "SummerList.dta")  # Save the summer list
```

```
# Step 3: Merge main dataset with summer list
data <- read_dta("AER20090377_FinalData.dta") %>%
  filter(valid >= 9) %>%
  inner_join(summer_list, by = c("fips", "site_id", "year"))
```

```
# Step 4: Drop untreated counties bordering treated counties
neighbor_data <- read_dta("AER20090377_NeighborData.dta")
data <- data %>%
  inner_join(neighbor_data, by = "fips") %>%  # Merge datasets
  filter(complete.cases(.)) %>%  # Filter rows where no column is missing
  mutate(treat_rfg = ifelse(treat_CARB != 0, 0, treat_rfg))  # Turn off RFG when CARB is on

write_dta(data, "temp.dta")
```

```
# Step 5: Create weather variables and interactions
library(dplyr)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:data.table':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```r
library(haven)

# Ensure TempMax and TempMin are numeric
data$TempMax <- as.numeric(data$TempMax)
data$TempMin <- as.numeric(data$TempMin)

data <- data %>%
  # Day of Week and Day of Year calculations
  mutate(
    DOW = wday(Date),   # Use default numeric output
    DOY = yday(Date),
    month = month(Date)
  ) %>%

  # Generate temperature polynomials and interactions
  mutate(
    TempMax1 = TempMax,
    TempMax2 = TempMax^2,
    TempMax3 = TempMax^3,
    TempMin1 = TempMin,
    TempMin2 = TempMin^2,
    TempMin3 = TempMin^3,
    TempMaxMin = TempMax * TempMin,
    Rain1 = Rain,
    Rain2 = Rain^2,
    Snow1 = Snow,
    Snow2 = Snow^2,
    RainTempMax = Rain * TempMax
  ) %>%

  # Sort data by fips, site_id, and Date for lagged calculations
  group_by(fips, site_id) %>%
  arrange(Date, .by_group = TRUE) %>%

  # Generate lagged variables (force use of dplyr::lag)
  mutate(
    tracker = row_number(),
    TempMaxL1 = dplyr::lag(TempMax),
    TempMinL1 = dplyr::lag(TempMin),
    TempMaxMaxL1 = TempMax * dplyr::lag(TempMax),
    TempMaxMinL1 = TempMax * dplyr::lag(TempMin)
  ) %>%

  # Replace lagged values with NA for the first row in each group
  mutate(
    TempMaxL1 = ifelse(tracker == 1, NA, TempMaxL1),
    TempMinL1 = ifelse(tracker == 1, NA, TempMinL1)
  ) %>%
  ungroup() %>%

  # Generate DOY interactions with temperature-related variables
  mutate(
```

```
    DOY_TempMax1 = DOY * TempMax1,
    DOY_TempMax2 = DOY * TempMax2,
    DOY_TempMax3 = DOY * TempMax3,
    DOY_TempMin1 = DOY * TempMin1,
    DOY_TempMin2 = DOY * TempMin2,
    DOY_TempMin3 = DOY * TempMin3,
    DOY_TempMaxMinL1 = DOY * TempMaxMinL1
  ) %>%

  # Filter for summer months (June, July, August)
  filter(month %in% c(6, 7, 8))

write_dta(data, "temp.dta")
```

```
# Keep only summer months (June-August)
data <- data %>%
  filter(month %in% c(6, 7, 8))
write_dta(data, "temp.dta")

# Step 6: Add income and log-transform ozone variables
income_data <- read_dta("AER20090377_IncomeData.dta")
data <- data %>%
  inner_join(income_data, by = c("state_code", "county_code", "year")) %>%
  filter(ozone_max > 0, epa_8hr > 0, income > 0) %>%
  mutate(
    lozone_max = log(ozone_max),
    lozone_8hr = log(epa_8hr)
  )

write_dta(data, "temp.dta")
```

```
# Step 7: Create census region dummies
region_mapping <- list(
  Region1 = c(9, 23, 25, 33, 34, 36, 42, 44, 50),  # Region 1
  Region2 = c(17, 18, 19, 20, 26, 27, 29, 31, 38, 39, 46, 55), # Region 2
  Region3 = c(1, 5, 10, 11, 12, 13, 21, 22, 24, 28, 37, 40, 45, 47, 48, 51, 54), # Regio
n 3
  Region4 = c(2, 4, 6, 8, 15, 16, 30, 32, 35, 41, 49, 53, 56) # Region 4
)


data <- data %>%
  mutate(
    region = case_when(
      state_code %in% region_mapping$Region1 ~ 1,
      state_code %in% region_mapping$Region2 ~ 2,
      state_code %in% region_mapping$Region3 ~ 3,
      state_code %in% region_mapping$Region4 ~ 4
    )
  )
```

```
# Step 8: Create time trends and interactions
data <- data %>%
  mutate(
    DateS = as.numeric(Date) / 365,  # Time trend in years
    DateS2 = DateS^2,  # Quadratic trend
    TrendRVP = ifelse(RVPCty == 1, DateS, 0),
    QTrendRVP = ifelse(RVPCty == 1, DateS2, 0)
    # Repeat for RFG, CARB, etc. with similar logic
  )

write_dta(data, "DD_AnalysisDataset_NYR.dta")
```

```
# Step 9: First-difference dataset for panel analysis
data <- read_dta("DD_AnalysisDataset_NYR.dta") %>%
  group_by(panelid) %>%
  mutate(across(c(starts_with("lozone_"), starts_with("treat"), starts_with("Temp"),
                starts_with("Rain"), starts_with("Snow"), starts_with("DOY"),
                starts_with("Trend"), starts_with("QTrend"), DateS, income),
              ~ . - mean(., na.rm = TRUE),
              .names = "{.col}D")) %>%
  ungroup()

write_dta(data, "DD_AnalysisDataset_Diffed_NYR.dta")
```

AirQualityFigures

2024-11-19

Data Loading

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(haven)
data <- read_dta("AER20090377_FinalData.dta")
head(data)
```

```
## # A tibble: 6 × 64
##   state_code county_code site_id valid epa_8hr ozone_max   day month  year
##        <dbl>       <dbl>   <dbl> <dbl>   <dbl>     <dbl> <dbl> <dbl> <dbl>
## 1          1           1       3     7  0.0543    0.0680     1     9  1989
## 2          1           1       3     3  0.0205    0.0360     2     9  1989
## 3          1           1       3    12  0.0701    0.0790     3     9  1989
## 4          1           1       3    12  0.0539    0.0590     4     9  1989
## 5          1           1       3    12  0.0591    0.0650     5     9  1989
## 6          1           1       3     9  0.0367    0.0490     6     9  1989
## # ℹ 55 more variables: Date <date>, fips <dbl>, state <chr>, county <chr>,
## #   partial <chr>, partialinfo <chr>, regtype <chr>, fedvssip <chr>, psi <dbl>,
## #   sulfur <chr>, sulfurppm <dbl>, rfgtype <chr>, noxeffect <chr>,
## #   RVPStart <date>, RVPEnd <date>, RFGStart <date>, RFGEnd <date>,
## #   RegFlag <dbl>, RVPI <dbl>, treat_rvpII <dbl>, treat_rfg <dbl>,
## #   treat_rvpI <dbl>, treat_CARB <dbl>, TreatRFG <dbl>, panelid <dbl>,
## #   RFGStart2 <dbl>, RFGEnd2 <dbl>, TreatRVPII <dbl>, RVPStart2 <dbl>, …
```

```
colnames(data)
```

```
##  [1] "state_code"          "county_code"         "site_id"
##  [4] "valid"               "epa_8hr"             "ozone_max"
##  [7] "day"                 "month"               "year"
## [10] "Date"                "fips"                "state"
## [13] "county"              "partial"             "partialinfo"
## [16] "regtype"             "fedvssip"            "psi"
## [19] "sulfur"              "sulfurppm"           "rfgtype"
## [22] "noxeffect"           "RVPStart"            "RVPEnd"
## [25] "RFGStart"            "RFGEnd"              "RegFlag"
## [28] "RVPI"                "treat_rvpII"         "treat_rfg"
## [31] "treat_rvpI"          "treat_CARB"          "TreatRFG"
## [34] "panelid"             "RFGStart2"           "RFGEnd2"
## [37] "TreatRVPII"          "RVPStart2"           "RVPEnd2"
## [40] "TreatCARB"           "TreatRFGCA"          "RVPCty"
## [43] "RFGCty"              "CARBCty"             "TreatRVPca"
## [46] "_merge2"             "_merge3"             "SiteObs"
## [49] "TempMax"             "TempMin"             "EstTempFlag"
## [52] "NOtherStation"       "SiteObsprcp"         "Rain"
## [55] "Snow"                "EstTempFlagprcp"     "NumOffMax"
## [58] "NumOffMin"           "NumOff1Max"          "NumOff1Min"
## [61] "NOtherStationprcp" "_merge"              "urban"
## [64] "_mergeurb"
```

Figures

```
figure <- read_dta("DD_AnalysisDataset.dta")

# Create a county type variable
figure <- figure %>%
  mutate(CountyType = case_when(
    RVPCty == 1 ~ 1,
    RFGCty == 1 | RVPRFGCty == 1 ~ 2,
    CARBCty == 1 | CARBRFGCty == 1 ~ 3,
    TRUE ~ 0
  ))
```

```
# Collapse the dataset by CountyType and year for plotting
collapsed_figure <- figure %>%
  group_by(CountyType, year) %>%
  summarise(
    ozone_max = mean(ozone_max, na.rm = TRUE),
    TempMax = mean(TempMax, na.rm = TRUE)
  ) %>%
  ungroup()
```

```
## `summarise()` has grouped output by 'CountyType'. You can override using the
## `.groups` argument.
```

```
write_dta(collapsed_figure, "GraphData.dta")
```

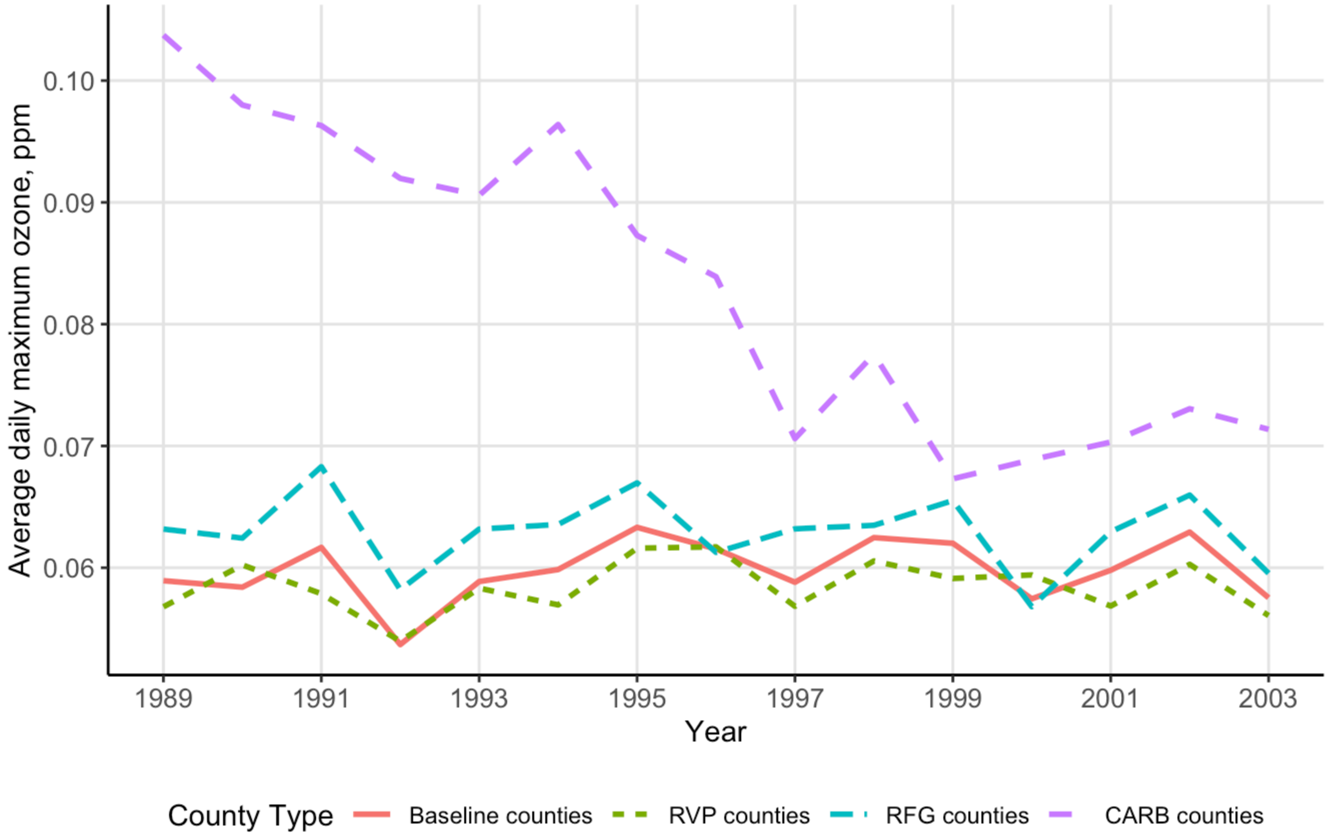
```
ggplot(collapsed_figure, aes(x = year, y = ozone_max, group = CountyType, linetype = as.
factor(CountyType))) +
  geom_line(aes(color = as.factor(CountyType)), size = 1) +
  scale_linetype_manual(values = c("solid", "dashed", "solid", "dashed")) +
  scale_color_manual(values = c("gray20", "gray40", "gray60", "gray80")) +
  labs(
    title = "Figure 3. Mean Summer Ozone Concentrations",
    subtitle = "(Broken out by regulation type)",
    y = "Average daily maximum ozone, ppm",
    x = "Year",
    color = "County Type",
    linetype = "County Type"
  ) +
  scale_x_continuous(breaks = seq(1989, 2003, by = 2)) +
  theme_classic() +
  theme(
    legend.position = "bottom",
    legend.background = element_rect(fill = "white", color = NA),
    legend.key = element_blank(),
    plot.title = element_text(face = "bold", size = 12, hjust = 0.5),
    plot.subtitle = element_text(size = 10, hjust = 0.5),
    axis.title = element_text(size = 11),
    axis.text = element_text(size = 10),
    panel.grid.major = element_line(color = "gray90")
  ) +
  scale_linetype_discrete(labels = c("Baseline counties", "RVP counties", "RFG countie
s", "CARB counties")) +
  scale_color_discrete(labels = c("Baseline counties", "RVP counties", "RFG counties",
"CARB counties"))
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Scale for linetype is already present.
## Adding another scale for linetype, which will replace the existing scale.
## Scale for colour is already present.
## Adding another scale for colour, which will replace the existing scale.
```

Figure 3. Mean Summer Ozone Concentrations

(Broken out by regulation type)




```
ggsave("OzoneSummerAverages.png", width = 10, height = 6)
```

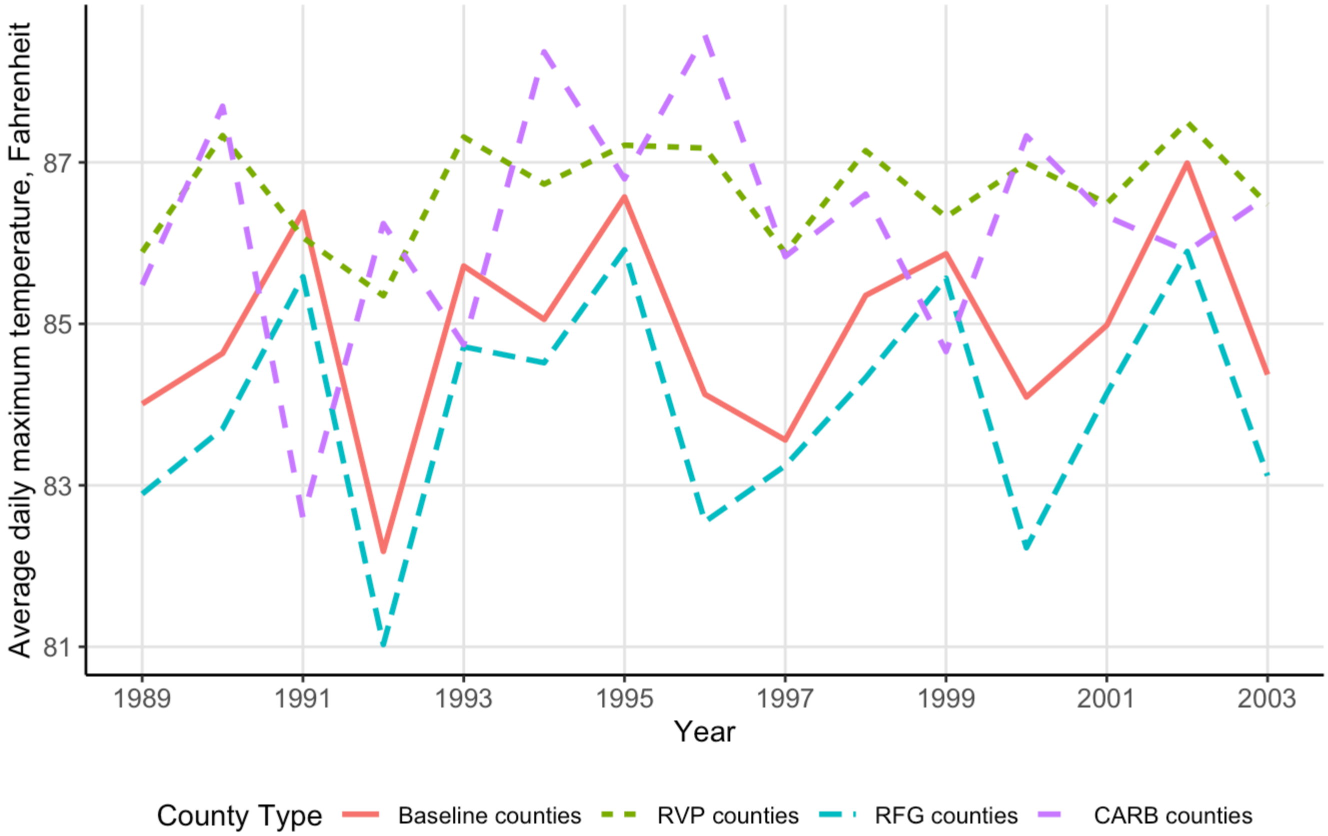
TemperatureSummerAverages

```
ggplot(collapsed_figure, aes(x = year, y = TempMax, group = CountyType, linetype = as.fa
ctor(CountyType))) +
  geom_line(aes(color = as.factor(CountyType)), size = 1) +
  scale_linetype_manual(values = c("solid", "dashed", "solid", "dashed")) +
  scale_color_manual(values = c("gray20", "gray40", "gray60", "gray80")) +
  labs(
    title = "Figure 4. Mean Summer Daily Temperature Maxima",
    subtitle = "(Broken out by regulation type)",
    y = "Average daily maximum temperature, Fahrenheit",
    x = "Year",
    color = "County Type",
    linetype = "County Type"
  ) +
  scale_x_continuous(breaks = seq(1989, 2003, by = 2)) +
  theme_classic() +
  theme(
    legend.position = "bottom",
    legend.background = element_rect(fill = "white", color = NA),
    legend.key = element_blank(),
    plot.title = element_text(face = "bold", size = 12, hjust = 0.5),
    plot.subtitle = element_text(size = 10, hjust = 0.5),
    axis.title = element_text(size = 11),
    axis.text = element_text(size = 10),
    panel.grid.major = element_line(color = "gray90")
  ) +
  scale_linetype_discrete(labels = c("Baseline counties", "RVP counties", "RFG countie
s", "CARB counties")) +
  scale_color_discrete(labels = c("Baseline counties", "RVP counties", "RFG counties",
"CARB counties"))
```

```
## Scale for linetype is already present.
## Adding another scale for linetype, which will replace the existing scale.
## Scale for colour is already present.
## Adding another scale for colour, which will replace the existing scale.
```

Figure 4. Mean Summer Daily Temperature Maxima

(Broken out by regulation type)




```
ggsave("TemperatureSummerAverages.png", width = 10, height = 6)
```

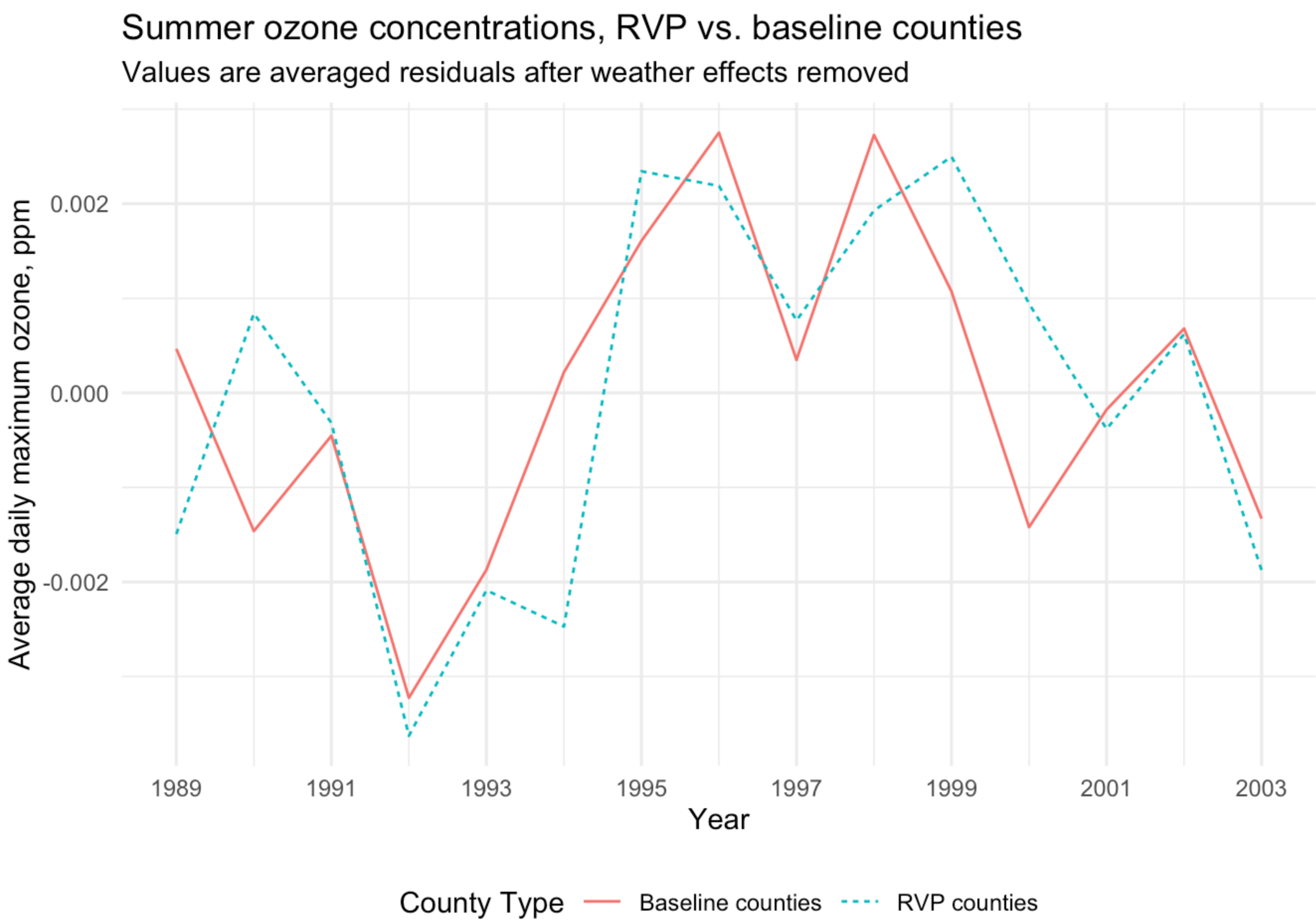
Diffed

plot 1

```
plot_data <- read_dta("GraphData_Resids.dta")

# Step 2: Plot RVP vs. Baseline Counties
ggplot(data = plot_data %>% filter(CountyType %in% c(0, 1)), aes(x = year, y = O3_Resid,
group = CountyType, linetype = as.factor(CountyType))) +
  geom_line(aes(color = as.factor(CountyType))) +
  scale_linetype_manual(values = c("dashed", "solid")) +
  scale_color_manual(values = c("black", "red")) +
  labs(
    title = "Summer ozone concentrations, RVP vs. baseline counties",
    subtitle = "Values are averaged residuals after weather effects removed",
    y = "Average daily maximum ozone, ppm",
    x = "Year",
    color = "County Type",
    linetype = "County Type"
  ) +
  scale_x_continuous(breaks = seq(1989, 2003, by = 2), labels = seq(1989, 2003, by = 2))
```

```
+
  theme_minimal() +
  theme(legend.position = "bottom") +
  scale_linetype_discrete(labels = c("Baseline counties", "RVP counties")) +
  scale_color_discrete(labels = c("Baseline counties", "RVP counties"))
```

```
## Scale for linetype is already present.
## Adding another scale for linetype, which will replace the existing scale.
## Scale for colour is already present.
## Adding another scale for colour, which will replace the existing scale.
```

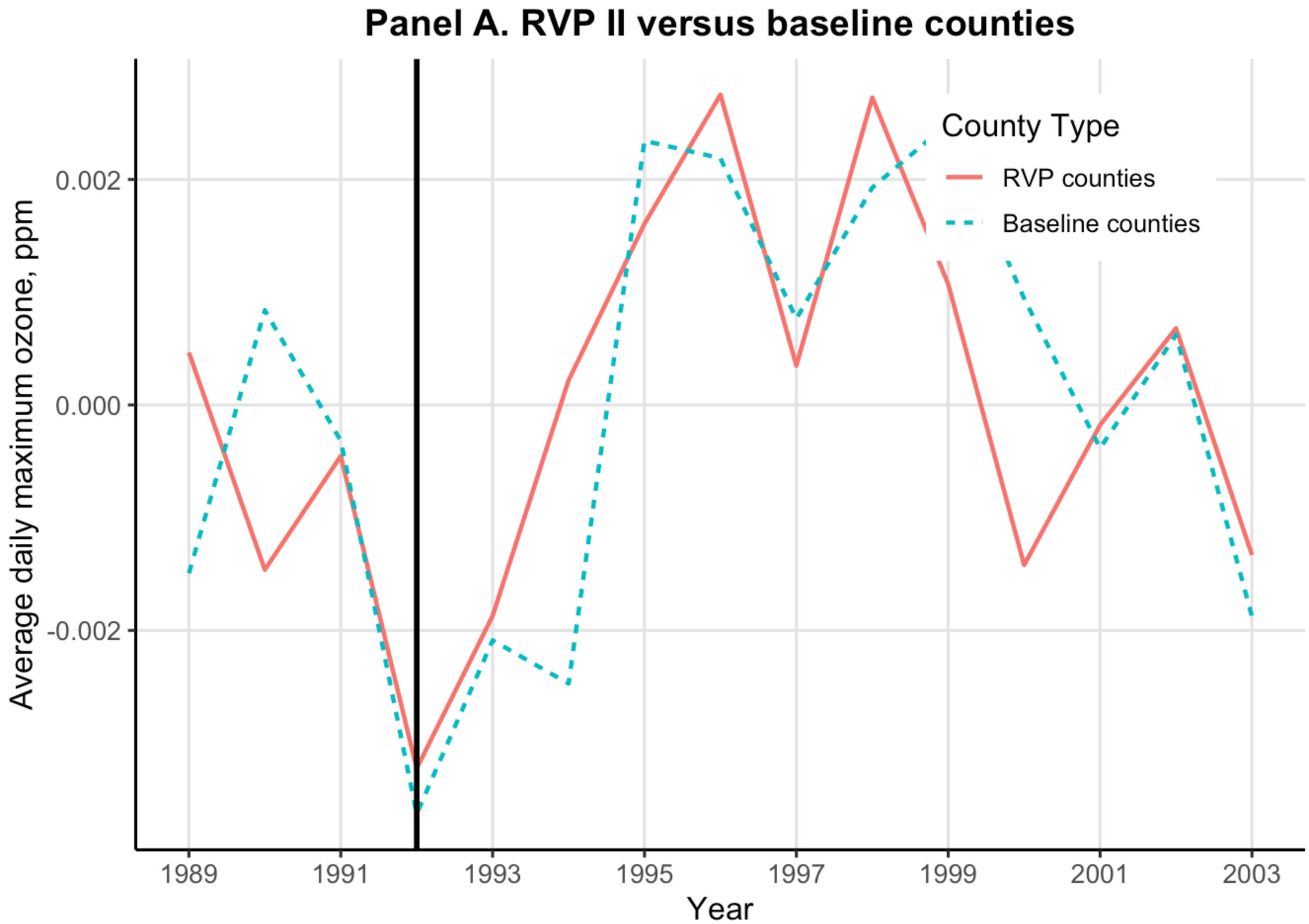



```
# Save the plot to a file
ggsave("Graph_AnnualResids_RVP.png", width = 10, height = 6)
```

```
ggplot(data = plot_data %>% filter(CountyType %in% c(0, 1)),
       aes(x = year, y = O3_Resid, group = CountyType, linetype = as.factor(CountyTyp
e))) +
  geom_line(aes(color = as.factor(CountyType)), size = 0.8) +
  geom_vline(xintercept = 1992, color = "black", size = 1) +
  scale_linetype_manual(values = c("solid", "dashed")) +
  scale_color_manual(values = c("gray40", "gray60")) +
  labs(
    title = "Panel A. RVP II versus baseline counties",
    y = "Average daily maximum ozone, ppm",
    x = "Year",
    color = "County Type",
    linetype = "County Type"
  ) +
  scale_x_continuous(breaks = seq(1989, 2003, by = 2)) +
  theme_classic(base_size = 12) +
  theme(
    legend.position = c(0.8, 0.85),
    legend.background = element_rect(fill = "white", color = NA),
    legend.key = element_blank(),
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5),
    axis.title = element_text(size = 12),
    axis.text = element_text(size = 10),
    panel.grid.major = element_line(color = "gray90"),
    panel.grid.minor = element_blank()
  ) +
  scale_linetype_discrete(labels = c("RVP counties", "Baseline counties")) +
  scale_color_discrete(labels = c("RVP counties", "Baseline counties"))
```

```
## Warning: A numeric `legend.position` argument in `theme()` was deprecated in ggplot2
## 3.5.0.
## ℹ Please use the `legend.position.inside` argument of `theme()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Scale for linetype is already present.
## Adding another scale for linetype, which will replace the existing scale.
## Scale for colour is already present.
## Adding another scale for colour, which will replace the existing scale.
```



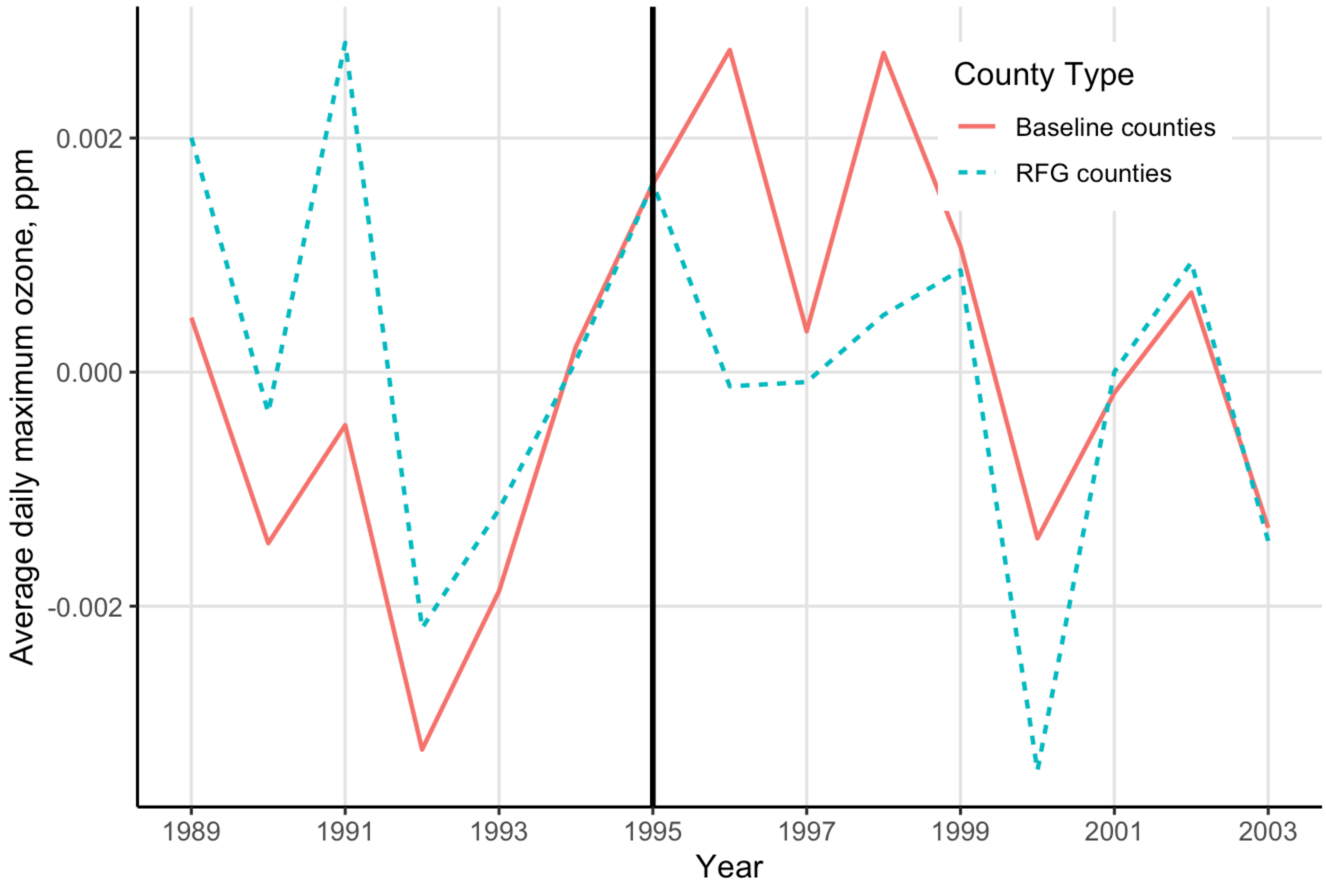

```
ggsave("Updated_Graph_AnnualResids_RVP.png", width = 10, height = 6)
```

Plot 2

```
# Step 3: Plot RFG vs. Baseline Counties
ggplot(data = plot_data %>% filter(CountyType %in% c(0, 2)),
       aes(x = year, y = O3_Resid, group = CountyType, linetype = as.factor(CountyTyp
e))) +
  geom_line(aes(color = as.factor(CountyType)), size = 0.8) +
  geom_vline(xintercept = 1995, color = "black", size = 1) +  # Vertical line to indicat
e the event in 1995
  scale_linetype_manual(values = c("solid", "dashed")) +
  scale_color_manual(values = c("gray40", "gray60")) +
  labs(
    title = "Panel B. RFG versus baseline counties",
    y = "Average daily maximum ozone, ppm",
    x = "Year",
    color = "County Type",
    linetype = "County Type"
  ) +
  scale_x_continuous(breaks = seq(1989, 2003, by = 2)) +
  theme_classic(base_size = 12) +
  theme(
    legend.position = c(0.8, 0.85),
    legend.background = element_rect(fill = "white", color = NA),
    legend.key = element_blank(),
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5),
    axis.title = element_text(size = 12),
    axis.text = element_text(size = 10),
    panel.grid.major = element_line(color = "gray90"),
    panel.grid.minor = element_blank()
  ) +
  scale_linetype_discrete(labels = c("Baseline counties", "RFG counties")) +
  scale_color_discrete(labels = c("Baseline counties", "RFG counties"))
```

```
## Scale for linetype is already present.
## Adding another scale for linetype, which will replace the existing scale.
## Scale for colour is already present.
## Adding another scale for colour, which will replace the existing scale.
```

Panel B. RFG versus baseline counties

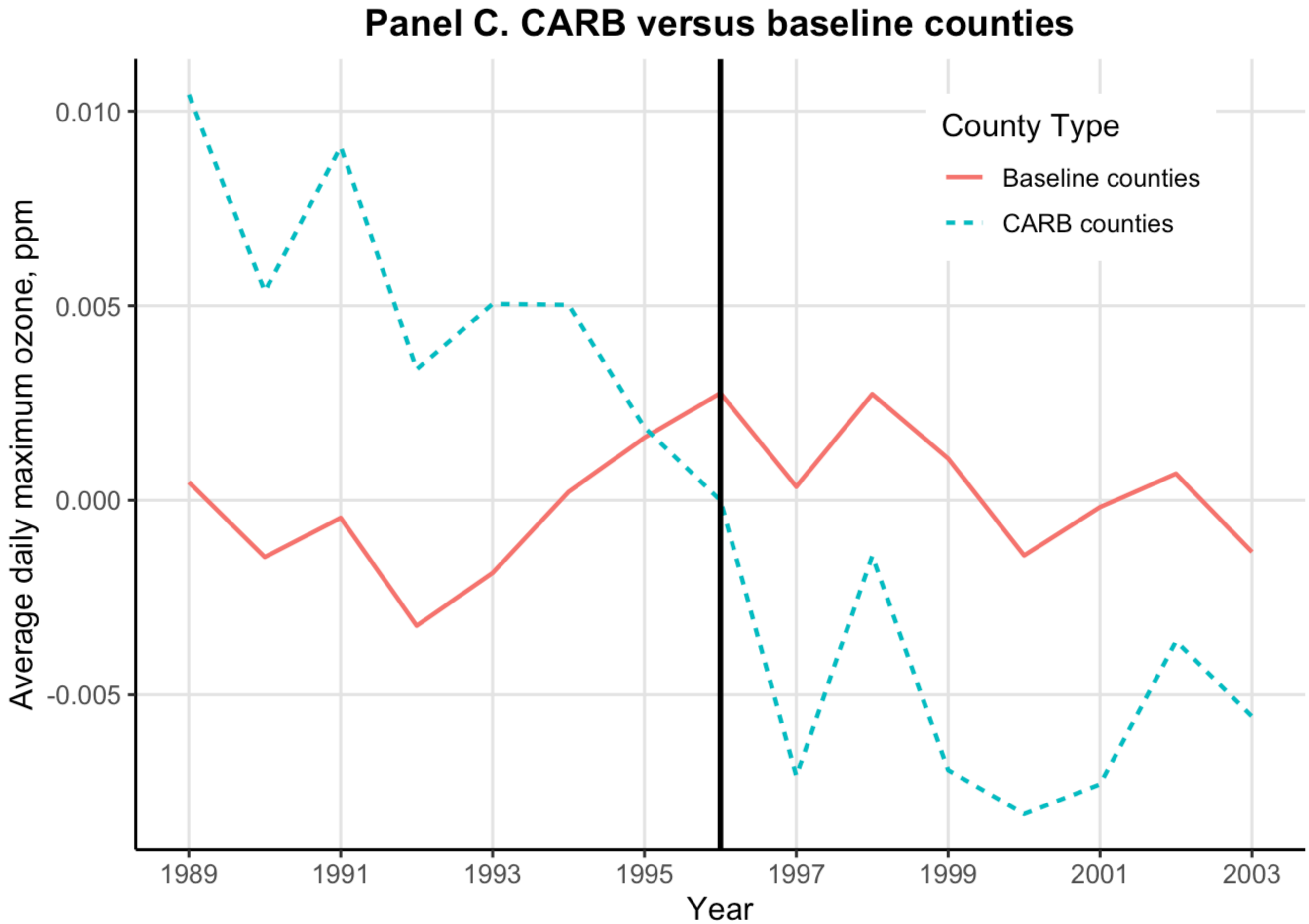



```
ggsave("Improved_Graph_AnnualResids_RFG.png", width = 10, height = 6)
```

Plot 3

```
# Step 4: Plot CARB vs. Baseline Counties
ggplot(data = plot_data %>% filter(CountyType %in% c(0, 3)),
       aes(x = year, y = O3_Resid, group = CountyType, linetype = as.factor(CountyTyp
e))) +
  geom_line(aes(color = as.factor(CountyType)), size = 0.8) +  # Thinner line for subtle
appearance
  geom_vline(xintercept = 1996, color = "black", size = 1) +  # Vertical line to indicat
e the event in 1996
  scale_linetype_manual(values = c("solid", "dashed")) +
  scale_color_manual(values = c("gray40", "gray60")) +  # Use shades of gray for underst
ated look
  labs(
    title = "Panel C. CARB versus baseline counties",
    y = "Average daily maximum ozone, ppm",
    x = "Year",
    color = "County Type",
    linetype = "County Type"
  ) +
  scale_x_continuous(breaks = seq(1989, 2003, by = 2)) +
  theme_classic(base_size = 12) +  # Use a classic theme for a clean look
  theme(
    legend.position = c(0.8, 0.85),  # Position the legend inside the plot area
    legend.background = element_rect(fill = "white", color = NA),  # White background fo
r the legend
    legend.key = element_blank(),  # Remove legend key box
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5),  # Emphasize the t
itle and center it
    axis.title = element_text(size = 12),
    axis.text = element_text(size = 10),
    panel.grid.major = element_line(color = "gray90"),  # Make grid lines very subtle
    panel.grid.minor = element_blank()
  ) +
  scale_linetype_discrete(labels = c("Baseline counties", "CARB counties")) +
  scale_color_discrete(labels = c("Baseline counties", "CARB counties"))
```

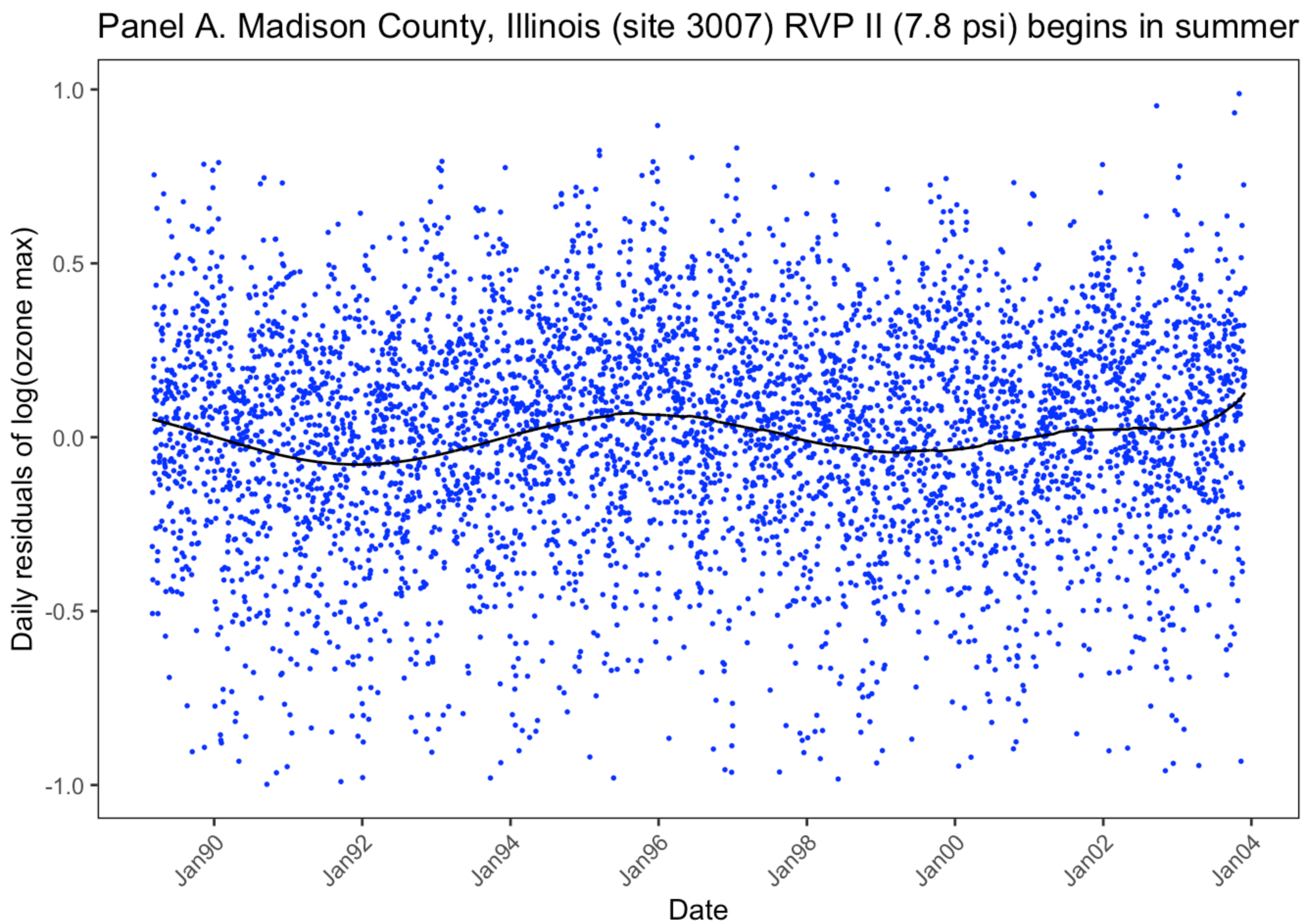
```
## Scale for linetype is already present.
## Adding another scale for linetype, which will replace the existing scale.
## Scale for colour is already present.
## Adding another scale for colour, which will replace the existing scale.
```




```
# Save the updated plot
ggsave("Improved_Graph_AnnualResids_CARB.png", width = 10, height = 6)
```

```
library(haven)
library(dplyr)
library(ggplot2)

# Step 1: Load the dataset and name it accordingly
RD_RVP_PlotData_8T <- read_dta("RD_RVP_PlotData_8T.dta")

# Step 2: Select fips and site_id of interest
filtered_data <- RD_RVP_PlotData_8T %>%
  filter(fips == 17119 & site_id == 3007)

# Step 3: Create series to plot - Residuals and treatment effects
# Calculate residuals and prepare data for plotting
filtered_data <- filtered_data %>%
  mutate(
    Resids = lozone_max - PredLogO3,
    PlotResids = Resids + BX,
    PlotLine = BX
  )

# Step 4: Normalize residuals so they center at zero
# Normalize PlotResids
filtered_data <- filtered_data %>%
  mutate(Mean_PlotResids = mean(PlotResids, na.rm = TRUE),
         PlotResids = PlotResids - Mean_PlotResids) %>%
  select(-Mean_PlotResids)

# Normalize PlotLine
filtered_data <- filtered_data %>%
  mutate(Mean_PlotLine = mean(PlotLine, na.rm = TRUE),
         PlotLine = PlotLine - Mean_PlotLine) %>%
  select(-Mean_PlotLine)

# Step 5: Filter data to only include values between -1 and 1 for plotting
plot_data <- filtered_data %>%
  filter(PlotResids <= 1 & PlotResids >= -1, PlotLine <= 1 & PlotLine >= -1)

# Step 6: Plotting with updated aesthetics to match the first figure
ggplot() +
  # Scatter plot for PlotResids
  geom_point(data = plot_data, aes(x = Date, y = PlotResids), color = "blue", size = 0.
3) +
  # Line plot for PlotLine
  geom_line(data = plot_data, aes(x = Date, y = PlotLine), color = "black") +
  labs(
    y = "Daily residuals of log(ozone max)",
    x = "Date",
    title = "Panel A. Madison County, Illinois (site 3007) RVP II (7.8 psi) begins in su
mmer 1995"
  ) +
  scale_x_date(date_labels = "%b%y", date_breaks = "2 years") +
  theme_bw() +
```

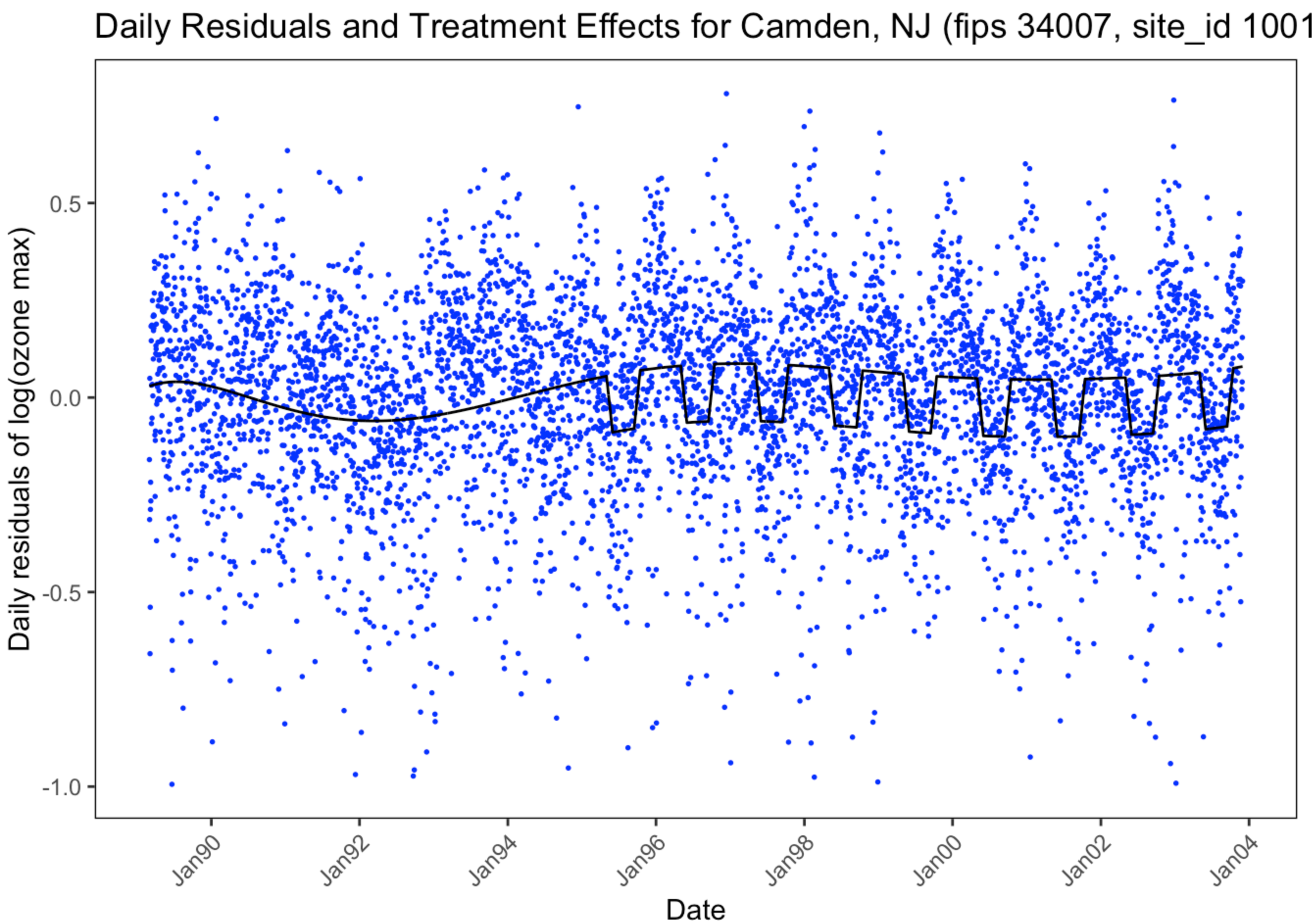
```
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    panel.background = element_blank(),
    panel.border = element_rect(color = "black")
  )
```




```
ggsave("RDPlot_RVP_Fips17119_Site3007_mean.png", width = 10, height = 6)
```

```
library(haven)
library(dplyr)
library(ggplot2)

# Step 1: Load the dataset and name it accordingly
RD_RFG_PlotData_8T <- read_dta("RD_RFG_PlotData_8T.dta")

# Step 2: Select the specific fips and site_id of interest
filtered_data <- RD_RFG_PlotData_8T %>%
  filter(fips == 34007 & site_id == 1001)

# Step 3: Create series to plot - Residuals and treatment effects
# Calculate residuals and prepare data for plotting
filtered_data <- filtered_data %>%
  mutate(
    Resids = lozone_max - PredLogO3,
    PlotResids = Resids + BX,
    PlotLine = BX
  )

# Step 4: Normalize residuals so they center at zero
# Normalize PlotResids
filtered_data <- filtered_data %>%
  mutate(Mean_PlotResids = mean(PlotResids, na.rm = TRUE),
         PlotResids = PlotResids - Mean_PlotResids) %>%
  select(-Mean_PlotResids)

# Normalize PlotLine
filtered_data <- filtered_data %>%
  mutate(Mean_PlotLine = mean(PlotLine, na.rm = TRUE),
         PlotLine = PlotLine - Mean_PlotLine) %>%
  select(-Mean_PlotLine)

# Step 5: Filter data to only include values between -1 and 1 for plotting
plot_data <- filtered_data %>%
  filter(PlotResids <= 1 & PlotResids >= -1, PlotLine <= 1 & PlotLine >= -1)

# Step 6: Plotting with updated aesthetics
ggplot() +
  # Scatter plot for PlotResids
  geom_point(data = plot_data, aes(x = Date, y = PlotResids), color = "blue", size = 0.
3) +
  # Line plot for PlotLine
  geom_line(data = plot_data, aes(x = Date, y = PlotLine), color = "black") +
  labs(
    y = "Daily residuals of log(ozone max)",
    x = "Date",
    title = "Daily Residuals and Treatment Effects for Camden, NJ (fips 34007, site_id 1
001)"
  ) +
  scale_x_date(date_labels = "%b%y", date_breaks = "2 years") +
  theme_bw() +
```

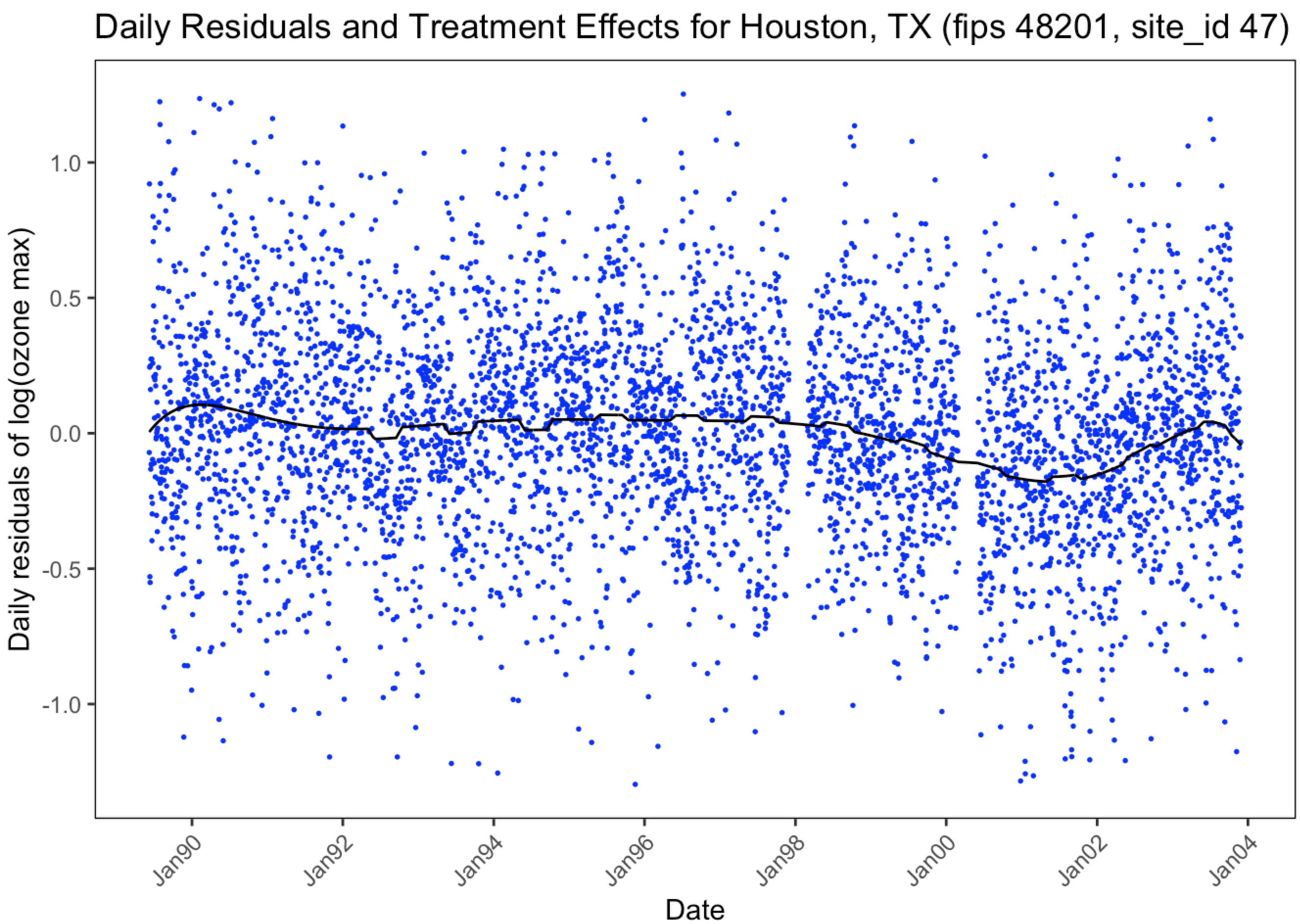
```
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid.major = element_blank(),  # Remove major grid lines
    panel.grid.minor = element_blank(),  # Remove minor grid lines
    panel.background = element_blank(),  # Remove background color
    panel.border = element_rect(color = "black")
  )
```




```
# Save the plot to a file
ggsave("RDPlot_RFG_Fips34007_Site1001_CamdenNJ.png", width = 10, height = 6)
```

```
library(haven)
library(dplyr)
library(ggplot2)

# Step 1: Load the dataset and name it accordingly
RD_RVP_RFG_PlotData_8T <- read_dta("RD_RVP_RFG_PlotData_8T.dta")

# Step 2: Select the specific fips and site_id of interest
filtered_data <- RD_RVP_RFG_PlotData_8T %>%
  filter(fips == 48201 & site_id == 47)

# Step 3: Create series to plot - Residuals and treatment effects
# Calculate residuals and prepare data for plotting
filtered_data <- filtered_data %>%
  mutate(
    Resids = lozone_max - PredLogO3,
    PlotResids = Resids + BX,
    PlotLine = BX
  )

# Step 4: Normalize residuals so they center at zero
# Normalize PlotResids
filtered_data <- filtered_data %>%
  mutate(Mean_PlotResids = mean(PlotResids, na.rm = TRUE),
         PlotResids = PlotResids - Mean_PlotResids) %>%
  select(-Mean_PlotResids)

# Normalize PlotLine
filtered_data <- filtered_data %>%
  mutate(Mean_PlotLine = mean(PlotLine, na.rm = TRUE),
         PlotLine = PlotLine - Mean_PlotLine) %>%
  select(-Mean_PlotLine)

# Step 5: Filter data to only include values between -1.3 and 1.3 for PlotResids,
# and between -1.5 and 1.5 for PlotLine to match the Stata range limits
plot_data <- filtered_data %>%
  filter(PlotResids <= 1.3 & PlotResids >= -1.3, PlotLine <= 1.5 & PlotLine >= -1.5)

# Step 6: Plotting with updated aesthetics to match the desired figure
ggplot() +
  # Scatter plot for PlotResids
  geom_point(data = plot_data, aes(x = Date, y = PlotResids), color = "blue", size = 0.
3) +
  # Line plot for PlotLine
  geom_line(data = plot_data, aes(x = Date, y = PlotLine), color = "black") +
  labs(
    y = "Daily residuals of log(ozone max)",
    x = "Date",
    title = "Daily Residuals and Treatment Effects for Houston, TX (fips 48201, site_id
47)"
  ) +
  scale_x_date(date_labels = "%b%y", date_breaks = "2 years") +
```

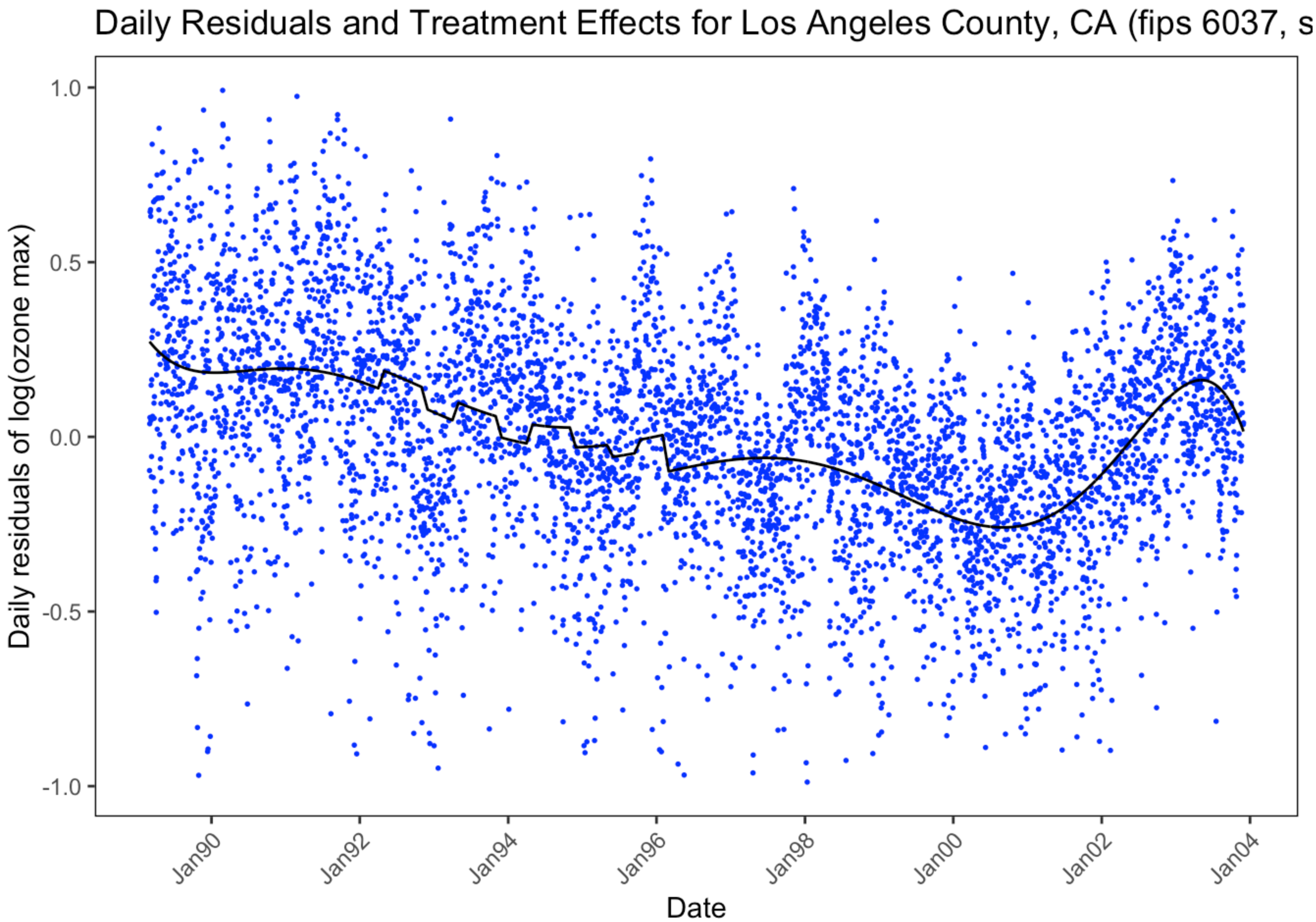
```
  scale_y_continuous(breaks = seq(-1.5, 1.5, 0.5)) +  # Set y-axis labels to match ylabel in Stata code
  theme_bw() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid.major = element_blank(),  # Remove major grid lines
    panel.grid.minor = element_blank(),  # Remove minor grid lines
    panel.background = element_blank(),  # Remove background color
    panel.border = element_rect(color = "black")
  )
```




```
# Save the plot to a file
ggsave("RDPlot_RVPRFG_Fips48201_Site47_Houston.png", width = 10, height = 6)
```

```
library(haven)
library(dplyr)
library(ggplot2)

# Step 1: Load the dataset and name it accordingly
RD_CARB_PlotData_8T <- read_dta("RD_CARB_PlotData_8T.dta")

# Step 2: Select the specific fips and site_id of interest
filtered_data <- RD_CARB_PlotData_8T %>%
  filter(fips == 6037 & site_id == 1201)

# Step 3: Create series to plot - Residuals and treatment effects
# Calculate residuals and prepare data for plotting
filtered_data <- filtered_data %>%
  mutate(
    Resids = lozone_max - PredLogO3,
    PlotResids = Resids + BX,
    PlotLine = BX
  )

# Step 4: Normalize residuals so they center at zero
# Normalize PlotResids
filtered_data <- filtered_data %>%
  mutate(Mean_PlotResids = mean(PlotResids, na.rm = TRUE),
         PlotResids = PlotResids - Mean_PlotResids) %>%
  select(-Mean_PlotResids)

# Normalize PlotLine
filtered_data <- filtered_data %>%
  mutate(Mean_PlotLine = mean(PlotLine, na.rm = TRUE),
         PlotLine = PlotLine - Mean_PlotLine) %>%
  select(-Mean_PlotLine)

# Step 5: Filter data to only include values between -1 and 1 for both PlotResids and Pl
otLine
plot_data <- filtered_data %>%
  filter(PlotResids <= 1 & PlotResids >= -1, PlotLine <= 1 & PlotLine >= -1)

# Step 6: Plotting with updated aesthetics to match the desired figure
ggplot() +
  # Scatter plot for PlotResids
  geom_point(data = plot_data, aes(x = Date, y = PlotResids), color = "blue", size = 0.
3) +
  # Line plot for PlotLine
  geom_line(data = plot_data, aes(x = Date, y = PlotLine), color = "black") +
  labs(
    y = "Daily residuals of log(ozone max)",
    x = "Date",
    title = "Daily Residuals and Treatment Effects for Los Angeles County, CA (fips 603
7, site_id 1201)"
  ) +
  scale_x_date(date_labels = "%b%y", date_breaks = "2 years") +
```

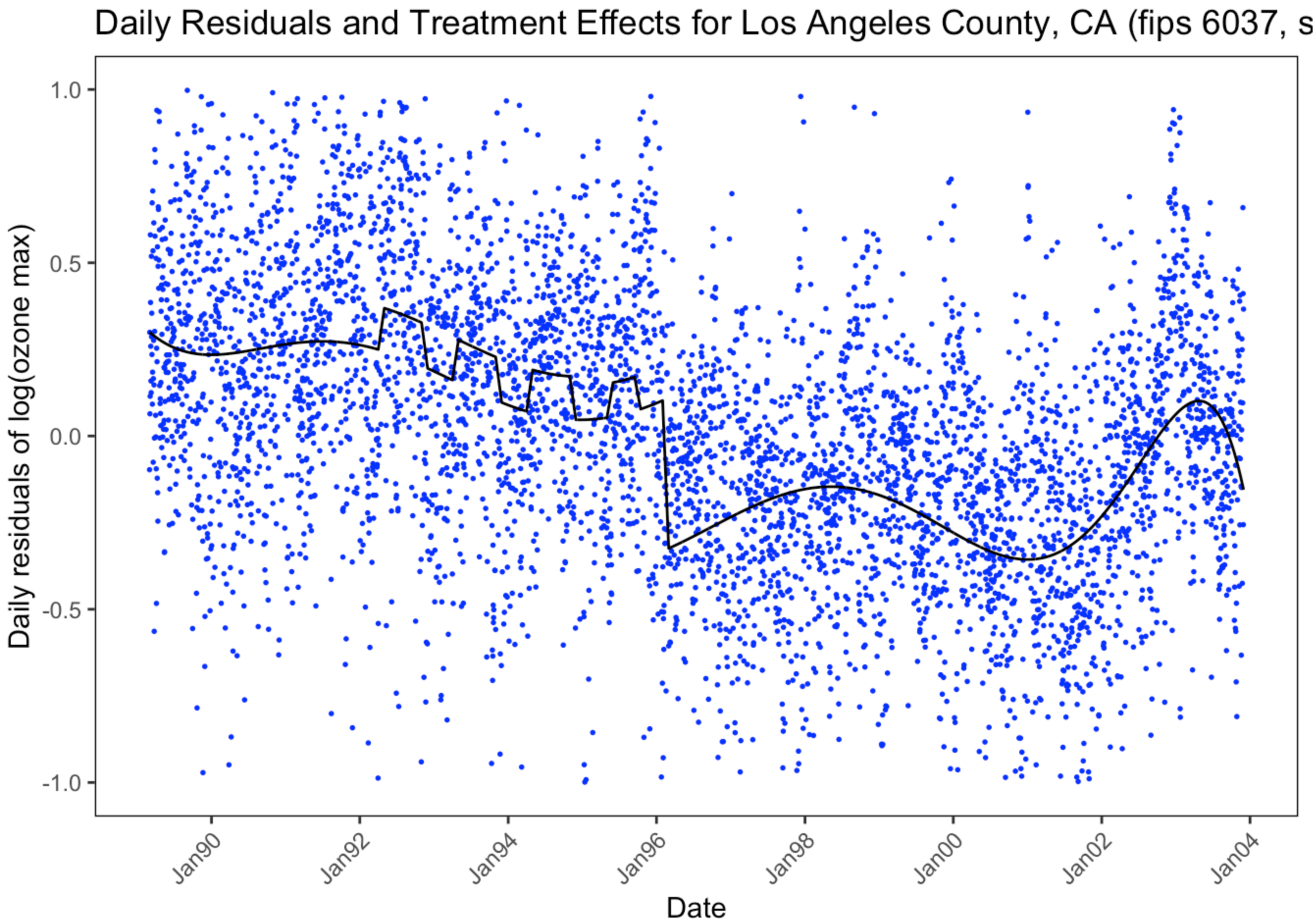
```
  theme_bw() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid.major = element_blank(),  # Remove major grid lines
    panel.grid.minor = element_blank(),  # Remove minor grid lines
    panel.background = element_blank(),  # Remove background color
    panel.border = element_rect(color = "black")
  )
```




```
# Save the plot to a file
ggsave("RDPlot_CARB_Fips6037_Site1201.png", width = 10, height = 6)
```

```
library(haven)
library(dplyr)
library(ggplot2)

# Step 1: Load the dataset and name it accordingly
RD_CARB_PlotData_8T <- read_dta("RD_CARB_PlotData_8T.dta")

# Step 2: Select the specific fips and site_id of interest
filtered_data <- RD_CARB_PlotData_8T %>%
  filter(fips == 6037 & site_id == 1701)

# Step 3: Create series to plot - Residuals and treatment effects
# Calculate residuals and prepare data for plotting
filtered_data <- filtered_data %>%
  mutate(
    Resids = lozone_max - PredLogO3,
    PlotResids = Resids + BX,
    PlotLine = BX
  )

# Step 4: Normalize residuals so they center at zero
# Normalize PlotResids
filtered_data <- filtered_data %>%
  mutate(Mean_PlotResids = mean(PlotResids, na.rm = TRUE),
         PlotResids = PlotResids - Mean_PlotResids) %>%
  select(-Mean_PlotResids)

# Normalize PlotLine
filtered_data <- filtered_data %>%
  mutate(Mean_PlotLine = mean(PlotLine, na.rm = TRUE),
         PlotLine = PlotLine - Mean_PlotLine) %>%
  select(-Mean_PlotLine)

# Step 5: Filter data to only include values between -1 and 1 for both PlotResids and Pl
otLine
plot_data <- filtered_data %>%
  filter(PlotResids <= 1 & PlotResids >= -1, PlotLine <= 1 & PlotLine >= -1)

# Step 6: Plotting with updated aesthetics to match the desired figure
ggplot() +
  # Scatter plot for PlotResids
  geom_point(data = plot_data, aes(x = Date, y = PlotResids), color = "blue", size = 0.
3) +
  # Line plot for PlotLine
  geom_line(data = plot_data, aes(x = Date, y = PlotLine), color = "black") +
  labs(
    y = "Daily residuals of log(ozone max)",
    x = "Date",
    title = "Daily Residuals and Treatment Effects for Los Angeles County, CA (fips 603
7, site_id 1701)"
  ) +
  scale_x_date(date_labels = "%b%y", date_breaks = "2 years") +
```

```
  theme_bw() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid.major = element_blank(),  # Remove major grid lines
    panel.grid.minor = element_blank(),  # Remove minor grid lines
    panel.background = element_blank(),  # Remove background color
    panel.border = element_rect(color = "black")
  )
```




```
# Save the plot to a file
ggsave("RDPlot_CARB_Fips6037_Site1701.png", width = 10, height = 6)
```

stat156

2024-12-10

Table 1 : load packages & data

```r
library(haven)
library(dplyr)

setwd("/Users/jaeeun/Desktop/STAT156_GP/")
data <- read_dta("DD_AnalysisDataset.dta")
```

table 1

```r
years <- function(yr){
  t1<- data %>%
    filter(!is.na(epa_8hr) & epa_8hr < 0.25 & month %in% c(6, 7, 8))
  sample<- t1[t1$year == yr, ]

  observation <- t1 %>%
    filter(year == yr) %>%
    nrow()

  counties <- length(unique(sample$fips))

  total <- length(unique(sample$panelid))
  urban <- length(sample %>%
    filter(urban == 1) %>%
    pull(panelid) %>%
    unique())

  rural <-length(sample %>%
    filter(urban == 3) %>%
    pull(panelid) %>%
    unique())

  RVP1 <- length(sample %>%
    filter(treat_rvpI == 1) %>%
    pull(fips) %>%
    unique())

  RVP2 <- length(sample %>%
                   filter(treat_rvpII == 1) %>%
                   pull(fips) %>%
                   unique())
  RFG95 <- length(unique_fips <- sample %>%
                    filter(treat_rfg == 1) %>%
                    pull(fips) %>%
```

```
                    unique())
  carb <- length(unique_fips <- sample %>%
                   filter(treat_CARB == 1) %>%
                   pull(fips) %>%
                   unique())

  index <- c(year = yr, observation, counties, total, urban, rural, RVP1, RVP2, RFG95, carb)
  return(index)
}
results <- lapply(1989:2003, years)
columns <- c("year", "observation", "counties", "total", "urban",

             "rural", "RVP1", "RVP2", "RFG95", "CARB")
final_df <- data.frame(do.call(rbind, results))
colnames(final_df) <- columns
```

Total & Average

```
sum(final_df$observation) #1144025
```

```
## [1] 1144025
```

```
round(sum(final_df$observation) /15) #76268
```

```
## [1] 76268
```

```
round(sum(final_df$counties) /15) #471
```

```
## [1] 471
```

```
round(sum(final_df$total) /15) #854
```

```
## [1] 854
```

```
round(sum(final_df$urban)/15) #166
```

```
## [1] 166
```

```
round(sum(final_df$rural) /15) #322
```

```
## [1] 322
```

```
table1 <- rbind(
  final_df, c("Total",sum(final_df$observation), NA, "","","","", "","",""),
  c("Average",
    round(sum(final_df$observation)/15),
    round(sum(final_df$counties) / 15),
    round(sum(final_df$total) / 15),
```

```r
    round(sum(final_df$urban) / 15),
    round(sum(final_df$rural) / 15), "", "","","",""
  )
)

print(table1)
```

```
##        year observation counties total urban rural RVP1 RVP2 RFG95 CARB
## 1      1989       63076      418   720   153   244  371    0     0    0
## 2      1990       66108      436   751   157   268  381    0     0    0
## 3      1991       69164      451   782   151   297  395    0     0    0
## 4      1992       69848      452   789   155   300    0  132     0    0
## 5      1993       72606      469   815   167   301    0  140     0    0
## 6      1994       74440      473   835   163   316    0  140     0    0
## 7      1995       77007      477   865   170   330    0  111   111    0
## 8      1996       76462      471   854   165   330    0   76   106   48
## 9      1997       78283      478   873   166   336    0   76   108   48
## 10     1998       79544      487   889   165   344    0   82   108   49
## 11     1999       80750      485   899   168   344    0   87   108   49
## 12     2000       82466      489   915   178   346    0   97   107   49
## 13     2001       83781      490   929   178   355    0   97   108   47
## 14     2002       85230      495   943   177   361    0  100   109   49
## 15     2003       85260      498   945   180   362    0  101   108   50
## 16    Total     1144025     <NA>
## 17 Average       76268      471   854   166   322
```

Table 2 : load package & data

```r
library(dplyr)
library(haven)
library(fixest)

setwd("/Users/jaeeun/Desktop/STAT156_GP/")
data <- read_dta("DD_AnalysisDataset.dta")
dif<- read_dta("DD_AnalysisDataset_Diffed.dta")
```

Table2 - (1)

```r
data <- data %>%
    select(lozone_max, lozone_8hr, treat_rvpI, treat_rvpII, treat_rfg, treat_CARB,income,

           DOW, DOY, panelid, region, year, TempMax, TempMin, Rain, Snow)

model1 <- feols(
  lozone_max ~ treat_rvpI + treat_rvpII + treat_rfg + treat_CARB | panelid + region^year,
  data = data
)

#summary(model1)
```

Table2 - (2)

```r
model2 <- feols(
  lozone_max ~ treat_rvpI + treat_rvpII + treat_rfg + treat_CARB | panelid + region^year +

    region^DOW + region^DOY + TempMax + TempMin + Rain + Snow,
  data = data
)

#summary(model2)
```

Table2 - (3)

```r
data$income <- data$income/1000000000

model3 <- feols(
    lozone_max ~ treat_rvpI + treat_rvpII + treat_rfg + treat_CARB + income | panelid + region^year +

      region^DOW + region^DOY + TempMax + TempMin + Rain + Snow,
    data = data
)

#summary(model3)
```

Table2 - (4)

```r
dif$incomeD <- dif$incomeD/1000000000

model4 <- feols(
  lozone_maxD ~
    treat_rvpID + treat_rvpIID + treat_rfgD  + treat_CARBD +

    incomeD +

    TrendRVP1D + TrendRVP2D + TrendRVP3D + TrendRVP4D +
    TrendRFG1D + TrendRFG2D + TrendRFG3D +
    TrendRVPRFG1D + TrendRVPRFG2D + TrendRVPRFG3D + TrendRVPRFG4D +
    TrendCARB4D + TrendCARBRFG4D +

    TempMinD +TempMax1D + TempMax2D +TempMax3D +
    TempMin2D +TempMin3D +
    TempMaxMinD +TempMaxL1D + TempMinL1D +
    TempMaxMaxL1D + TempMaxMinL1D +

    RainD + Rain2D +RainTempMaxD +
    SnowD + Snow2D +

    DOYD + DOYTempMax1D + DOYTempMax2D + DOYTempMax3D +
    DOYTempMin1D + DOYTempMin2D + DOYTempMin3D +
    DOYTempMaxMinD +DOYRain1D + DOYRain2D +
    DOYSnow1D +DOYSnow2D + DOYRainTempMaxD +
    DOYTempMaxL1D + DOYTempMinL1D +
    DOYTempMaxMaxL1D + DOYTempMinL1D +

    `_DMDOWXTempM_1D` +`_DMDOWXTempM_2D` + `_DMDOWXTempM_3D` +
```

```r
    `_DMDOWXTempM_4D` + `_DMDOWXTempM_5D` + `_DMDOWXTempM_6D` +
    `_DmDOWXTempM_1D` + `_DmDOWXTempM_2D` + `_DmDOWXTempM_3D` +
    `_DmDOWXTempM_4D` + `_DmDOWXTempM_5D` + `_DmDOWXTempM_6D` +

    `_RYyear_1990D` + `_RYyear_1991D` + `_RYyear_1992D` +
    `_RYyear_1993D` + `_RYyear_1994D` + `_RYyear_1995D` +
    `_RYyear_1996D` + `_RYyear_1997D` + `_RYyear_1998D` +
    `_RYyear_1999D` + `_RYyear_2000D` + `_RYyear_2001D` +
    `_RYyear_2002D` +`_RYyear_2003D` +

    `_RWDOWXreg_1_2D` + `_RWDOWXreg_1_3D` + `_RWDOWXreg_1_4D` +
    `_RWDOWXreg_2_2D` + `_RWDOWXreg_2_3D` + `_RWDOWXreg_2_4D`,

  cluster = "A_StateYear",
  data = dif
)

#summary(model4)
```

Table2 - (5)

```r
model5 <- feols(
  lozone_maxD ~
    treat_rvpID + treat_rvpIID + treat_rfgD  + treat_CARBD +

    incomeD +

    TrendRVP1D + TrendRVP2D + TrendRVP3D + TrendRVP4D +
    TrendRFG1D + TrendRFG2D + TrendRFG3D +
    TrendRVPRFG1D + TrendRVPRFG2D + TrendRVPRFG3D + TrendRVPRFG4D +
    TrendCARB4D + TrendCARBRFG4D +

    QTrendRVP1D + QTrendRVP2D + QTrendRVP3D + QTrendRVP4D +
    QTrendRFG1D + QTrendRFG2D + QTrendRFG3D +
    QTrendRVPRFG1D + QTrendRVPRFG2D + QTrendRVPRFG3D + QTrendRVPRFG4D +
    QTrendCARB4D +

    TempMinD + TempMax1D + TempMax2D + TempMax3D +
    TempMin2D + TempMin3D +
    TempMaxMinD + TempMaxL1D + TempMinL1D +
    TempMaxMaxL1D + TempMaxMinL1D +

    RainD + Rain2D + RainTempMaxD +
    SnowD + Snow2D +

    DOYD + DOYTempMax1D + DOYTempMax2D + DOYTempMax3D +
    DOYTempMin1D + DOYTempMin2D + DOYTempMin3D +
    DOYTempMaxMinD + DOYRain1D + DOYRain2D +
    DOYSnow1D + DOYSnow2D + DOYRainTempMaxD +
    DOYTempMaxL1D + DOYTempMinL1D +
    DOYTempMaxMaxL1D + DOYTempMinL1D +

    `_DMDOWXTempM_1D` + `_DMDOWXTempM_2D` + `_DMDOWXTempM_3D` +
```

```
    `_DMDOWXTempM_4D` + `_DMDOWXTempM_5D` + `_DMDOWXTempM_6D` +
    `_DmDOWXTempM_1D` + `_DmDOWXTempM_2D` + `_DmDOWXTempM_3D` +
    `_DmDOWXTempM_4D` + `_DmDOWXTempM_5D` + `_DmDOWXTempM_6D` +

    `_RYyear_1990D` + `_RYyear_1991D` + `_RYyear_1992D` +
    `_RYyear_1993D` + `_RYyear_1994D` + `_RYyear_1995D` +
    `_RYyear_1996D` + `_RYyear_1997D` + `_RYyear_1998D` +
    `_RYyear_1999D` + `_RYyear_2000D` + `_RYyear_2001D` +
    `_RYyear_2002D` + `_RYyear_2003D` +

    `_RWDOWXreg_1_2D` + `_RWDOWXreg_1_3D` + `_RWDOWXreg_1_4D` +
    `_RWDOWXreg_2_2D` + `_RWDOWXreg_2_3D` + `_RWDOWXreg_2_4D`,

  cluster = "A_StateYear",

  data = dif
)

#summary(model5)
```

Table2 - (6)

```
model6 <- feols(
  lozone_8hr ~ treat_rvpI + treat_rvpII + treat_rfg + treat_CARB | panelid + region^year,
  data = data
)

#summary(model6)
```

Table2 - (7)

```
model7 <- feols(
  lozone_8hr ~ treat_rvpI + treat_rvpII + treat_rfg + treat_CARB | panelid + region^year +

    region^DOW + region^DOY + TempMax + TempMin + Rain + Snow,
  data = data
)

#summary(model7)
```

Table2 - (8)

```
model8 <- feols(
  lozone_8hrD ~
    treat_rvpID + treat_rvpIID + treat_rfgD  + treat_CARBD +

    incomeD +

    TrendRVP1D + TrendRVP2D + TrendRVP3D + TrendRVP4D +
    TrendRFG1D + TrendRFG2D + TrendRFG3D +
    TrendRVPRFG1D + TrendRVPRFG2D + TrendRVPRFG3D + TrendRVPRFG4D +
    TrendCARB4D + TrendCARBRFG4D +
```

```r
    QTrendRVP1D + QTrendRVP2D + QTrendRVP3D + QTrendRVP4D +
    QTrendRFG1D + QTrendRFG2D + QTrendRFG3D +
    QTrendRVPRFG1D + QTrendRVPRFG2D + QTrendRVPRFG3D + QTrendRVPRFG4D +
    QTrendCARB4D +

    TempMinD + TempMax1D + TempMax2D + TempMax3D +
    TempMin2D + TempMin3D +
    TempMaxMinD + TempMaxL1D + TempMinL1D +
    TempMaxMaxL1D + TempMaxMinL1D +

    RainD + Rain2D + RainTempMaxD +
    SnowD + Snow2D +

    DOYD + DOYTempMax1D + DOYTempMax2D + DOYTempMax3D +
    DOYTempMin1D + DOYTempMin2D + DOYTempMin3D +
    DOYTempMaxMinD + DOYRain1D + DOYRain2D +
    DOYSnow1D + DOYSnow2D + DOYRainTempMaxD +
    DOYTempMaxL1D + DOYTempMinL1D +
    DOYTempMaxMaxL1D + DOYTempMinL1D +

    `_DMDOWXTempM_1D` + `_DMDOWXTempM_2D` + `_DMDOWXTempM_3D` +
    `_DMDOWXTempM_4D` + `_DMDOWXTempM_5D` + `_DMDOWXTempM_6D` +
    `_DmDOWXTempM_1D` + `_DmDOWXTempM_2D` + `_DmDOWXTempM_3D` +
    `_DmDOWXTempM_4D` + `_DmDOWXTempM_5D` + `_DmDOWXTempM_6D` +

    `_RYyear_1990D` + `_RYyear_1991D` + `_RYyear_1992D` +
    `_RYyear_1993D` + `_RYyear_1994D` + `_RYyear_1995D` +
    `_RYyear_1996D` + `_RYyear_1997D` + `_RYyear_1998D` +
    `_RYyear_1999D` + `_RYyear_2000D` + `_RYyear_2001D` +
    `_RYyear_2002D` + `_RYyear_2003D` +

    `_RWDOWXreg_1_2D` + `_RWDOWXreg_1_3D` + `_RWDOWXreg_1_4D` +
    `_RWDOWXreg_2_2D` + `_RWDOWXreg_2_3D` + `_RWDOWXreg_2_4D`,

  cluster = "A_StateYear",

  data = dif
)

#summary(model8)
```

Table2 - fi

```r
coef_model1 <- coef(model1)
coef_model1["income"] <- NA

coef_model2 <- coef(model2)
coef_model2["income"] <- NA

coef_model3 <- coef(model3)

coef_model6 <- coef(model6)
coef_model6["income"] <- NA
```

```r
coef_model7 <- coef(model7)
coef_model7["income"] <- NA

selected_vars <- c("treat_rvpID", "treat_rvpIID", "treat_rfgD", "treat_CARBD", "incomeD")

coef_model4 <- sapply(selected_vars,

                      function(var) if (var %in% names(coef(model4))) coef(model4)[var] else NA)
coef_model5 <- sapply(selected_vars,

                      function(var) if (var %in% names(coef(model5))) coef(model5)[var] else NA)
coef_model8 <- sapply(selected_vars,

                      function(var) if (var %in% names(coef(model8))) coef(model8)[var] else NA)


coef_df <- data.frame(
  Model1 = coef_model1,
  Model2 = coef_model2,
  Model3 = coef_model3,
  Model4 = coef_model4,
  Model5 = coef_model5,
  Model6 = coef_model6,
  Model7 = coef_model7,
  Model8 = coef_model8
)

coef_df <- round(coef_df, 3)

print(coef_df)
```

```
##             Model1 Model2 Model3 Model4 Model5 Model6 Model7 Model8
## treat_rvpI   0.016  0.018  0.020  0.010  0.009  0.018  0.021  0.011
## treat_rvpII -0.007 -0.012 -0.008 -0.014 -0.022 -0.005 -0.010 -0.022
## treat_rfg   -0.029 -0.030 -0.018 -0.046 -0.046 -0.028 -0.029 -0.051
## treat_CARB  -0.095 -0.089 -0.077 -0.081 -0.089 -0.090 -0.085 -0.090
## income          NA     NA -1.281 -0.206 -0.213     NA     NA -0.012
```

Table3 : load package & data

```r
library(dplyr)
library(haven)
library(fixest)

setwd("/Users/jaeeun/Desktop/STAT156_GP/")
data <- read_dta("DD_AnalysisDataset.dta")
table(data$urban)
```

```
##
##      1      2      3
## 222982 490539 430504
```

```r
data$income <- data$income/1000000000
urban <- data %>% filter(urban == 1)
sub <- data %>% filter(urban == 2)
rural <- data %>% filter(urban == 3)
```

Table3 - Urban (1)

```r
urban <- urban %>% dplyr::select(lozone_max, lozone_8hr, treat_rvpI, treat_rvpII, treat_rfg,
                                  treat_CARB, income, DOW, DOY, panelid, region, year, TempMax,

                                 TempMin,
                                 Rain, Snow)

model1 <- feols(
    lozone_max ~ treat_rvpI + treat_rvpII + treat_rfg + treat_CARB + income | panelid +
      region^year + region^DOW + region^DOY + TempMax + TempMin + Rain + Snow,
    data = urban
)

#summary(model1)
```

Table3 - Urban (2)

```r
u_dif <- read_dta("urban_Diffed.dta")
model2 <- feols(
  lozone_maxD ~
    treat_rvpID + treat_rvpIID + treat_rfgD  + treat_CARBD +

    incomeD +

    TrendRVP1D + TrendRVP2D + TrendRVP3D + TrendRVP4D +
    TrendRFG1D + TrendRFG2D + TrendRFG3D +
    TrendRVPRFG1D + TrendRVPRFG2D + TrendRVPRFG3D + TrendRVPRFG4D +
    TrendCARB4D + TrendCARBRFG4D +

    TempMinD + TempMax1D + TempMax2D + TempMax3D +
    TempMin2D + TempMin3D +
    TempMaxMinD + TempMaxL1D + TempMinL1D +
    TempMaxMaxL1D + TempMaxMinL1D +

    RainD + Rain2D + RainTempMaxD +
    SnowD + Snow2D +

    DOYD + DOYTempMax1D + DOYTempMax2D + DOYTempMax3D +
    DOYTempMin1D + DOYTempMin2D + DOYTempMin3D +
    DOYTempMaxMinD + DOYRain1D + DOYRain2D +
    DOYSnow1D + DOYSnow2D + DOYRainTempMaxD +
    DOYTempMaxL1D + DOYTempMinL1D +
    DOYTempMaxMaxL1D + DOYTempMinL1D +

    `_DMDOWXTempM_1D` + `_DMDOWXTempM_2D` + `_DMDOWXTempM_3D` +
    `_DMDOWXTempM_4D` + `_DMDOWXTempM_5D` + `_DMDOWXTempM_6D` +
    `_DmDOWXTempM_1D` + `_DmDOWXTempM_2D` + `_DmDOWXTempM_3D` +
```

```
    `_DmDOWXTempM_4D` + `_DmDOWXTempM_5D` + `_DmDOWXTempM_6D` +

    `_RYyear_1990D` + `_RYyear_1991D` + `_RYyear_1992D` +
    `_RYyear_1993D` + `_RYyear_1994D` + `_RYyear_1995D` +
    `_RYyear_1996D` + `_RYyear_1997D` + `_RYyear_1998D` +
    `_RYyear_1999D` + `_RYyear_2000D` + `_RYyear_2001D` +
    `_RYyear_2002D` + `_RYyear_2003D` +

    `_RWDOWXreg_1_2D` + `_RWDOWXreg_1_3D` + `_RWDOWXreg_1_4D` +
    `_RWDOWXreg_2_2D` + `_RWDOWXreg_2_3D` + `_RWDOWXreg_2_4D`,

  cluster = "A_StateYear",
  data = u_dif
)

#summary(model2)
```

Table3 - suburban(3)

```
sub <- sub %>%
  dplyr::select(lozone_max, lozone_8hr, treat_rvpI, treat_rvpII, treat_rfg,
         treat_CARB, income, DOW, DOY, panelid, region, year,

         TempMax, TempMin, Rain, Snow)

model3 <- feols(
    lozone_max ~ treat_rvpI + treat_rvpII + treat_rfg + treat_CARB + income | panelid + region^year +

      region^DOW + region^DOY + TempMax + TempMin + Rain + Snow,
    data = sub
)

#summary(model3)
```

Table3 - suburban(4)

```
s_dif <- read_dta("sub_Diffed.dta")
model4 <- feols(
  lozone_maxD ~
    treat_rvpID + treat_rvpIID + treat_rfgD  + treat_CARBD + incomeD +

    TrendRVP1D + TrendRVP2D + TrendRVP3D + TrendRVP4D +
    TrendRFG1D + TrendRFG2D + TrendRFG3D +
    TrendRVPRFG1D + TrendRVPRFG2D + TrendRVPRFG3D + TrendRVPRFG4D +
    TrendCARB4D + TrendCARBRFG4D +

    TempMinD + TempMax1D + TempMax2D + TempMax3D +
    TempMin2D + TempMin3D +
    TempMaxMinD + TempMaxL1D + TempMinL1D +
    TempMaxMaxL1D + TempMaxMinL1D +

    RainD + Rain2D + RainTempMaxD +
    SnowD + Snow2D +
```

```
    DOYD + DOYTempMax1D + DOYTempMax2D + DOYTempMax3D +
    DOYTempMin1D + DOYTempMin2D + DOYTempMin3D +
    DOYTempMaxMinD + DOYRain1D + DOYRain2D +
    DOYSnow1D + DOYSnow2D + DOYRainTempMaxD +
    DOYTempMaxL1D + DOYTempMinL1D +
    DOYTempMaxMaxL1D + DOYTempMinL1D +

    `_DMDOWXTempM_1D` + `_DMDOWXTempM_2D` + `_DMDOWXTempM_3D` +
    `_DMDOWXTempM_4D` + `_DMDOWXTempM_5D` + `_DMDOWXTempM_6D` +
    `_DmDOWXTempM_1D` + `_DmDOWXTempM_2D` + `_DmDOWXTempM_3D` +
    `_DmDOWXTempM_4D` + `_DmDOWXTempM_5D` + `_DmDOWXTempM_6D` +

    `_RYyear_1990D` + `_RYyear_1991D` + `_RYyear_1992D` +
    `_RYyear_1993D` + `_RYyear_1994D` + `_RYyear_1995D` +
    `_RYyear_1996D` + `_RYyear_1997D` + `_RYyear_1998D` +
    `_RYyear_1999D` + `_RYyear_2000D` + `_RYyear_2001D` +
    `_RYyear_2002D` + `_RYyear_2003D` +

    `_RWDOWXreg_1_2D` + `_RWDOWXreg_1_3D` + `_RWDOWXreg_1_4D` +
    `_RWDOWXreg_2_2D` + `_RWDOWXreg_2_3D` + `_RWDOWXreg_2_4D`,

  cluster = "A_StateYear",
  data = s_dif
)

#summary(model4)
```

Table3 - rural(5)

```
rural <- rural %>%
  dplyr::select(lozone_max, lozone_8hr, treat_rvpI, treat_rvpII, treat_rfg,
         treat_CARB, income, DOW, DOY, panelid, region, year,
         TempMax, TempMin, Rain, Snow)

model5 <- feols(
    lozone_max ~ treat_rvpI + treat_rvpII + treat_rfg + treat_CARB + income | panelid + region^year +

      region^DOW + region^DOY + TempMax + TempMin + Rain + Snow,
    data = rural
)

#summary(model5)
```

Table3 - rural(6)

```
r_dif <- read_dta("rural_Diffed.dta")
model6 <- feols(
  lozone_maxD ~
    treat_rvpID + treat_rvpIID + treat_rfgD  + treat_CARBD + incomeD +

    TrendRVP1D + TrendRVP2D + TrendRVP3D + TrendRVP4D +
    TrendRFG1D + TrendRFG2D + TrendRFG3D +
```

```
    TrendRVPRFG1D + TrendRVPRFG2D + TrendRVPRFG3D + TrendRVPRFG4D +
    TrendCARB4D + TrendCARBRFG4D +

    TempMinD + TempMax1D + TempMax2D + TempMax3D +
    TempMin2D + TempMin3D +
    TempMaxMinD + TempMaxL1D + TempMinL1D +
    TempMaxMaxL1D + TempMaxMinL1D +

    RainD + Rain2D + RainTempMaxD +
    SnowD + Snow2D +

    DOYD + DOYTempMax1D + DOYTempMax2D + DOYTempMax3D +
    DOYTempMin1D + DOYTempMin2D + DOYTempMin3D +
    DOYTempMaxMinD + DOYRain1D + DOYRain2D +
    DOYSnow1D + DOYSnow2D + DOYRainTempMaxD +
    DOYTempMaxL1D + DOYTempMinL1D +
    DOYTempMaxMaxL1D + DOYTempMinL1D +

    `_DMDOWXTempM_1D` + `_DMDOWXTempM_2D` + `_DMDOWXTempM_3D` +
    `_DMDOWXTempM_4D` + `_DMDOWXTempM_5D` + `_DMDOWXTempM_6D` +
    `_DmDOWXTempM_1D` + `_DmDOWXTempM_2D` + `_DmDOWXTempM_3D` +
    `_DmDOWXTempM_4D` + `_DmDOWXTempM_5D` + `_DmDOWXTempM_6D` +

    `_RYyear_1990D` + `_RYyear_1991D` + `_RYyear_1992D` +
    `_RYyear_1993D` + `_RYyear_1994D` + `_RYyear_1995D` +
    `_RYyear_1996D` + `_RYyear_1997D` + `_RYyear_1998D` +
    `_RYyear_1999D` + `_RYyear_2000D` + `_RYyear_2001D` +
    `_RYyear_2002D` + `_RYyear_2003D` +

    `_RWDOWXreg_1_2D` + `_RWDOWXreg_1_3D` + `_RWDOWXreg_1_4D` +
    `_RWDOWXreg_2_2D` + `_RWDOWXreg_2_3D` + `_RWDOWXreg_2_4D`,

  cluster = "A_StateYear",
  data = r_dif
)

#summary(model6)
```

Table3 - fi

```
coef_model1 <- coef(model1)
coef_model3 <- coef(model3)
coef_model5 <- coef(model5)

selected_vars <- c("treat_rvpID", "treat_rvpIID", "treat_rfgD", "treat_CARBD", "incomeD")

coef_model2 <- sapply(selected_vars,
                      function(var) if (var %in% names(coef(model2))) coef(model2)[var] else NA)
coef_model4 <- sapply(selected_vars,
                      function(var) if (var %in% names(coef(model4))) coef(model4)[var] else NA)
coef_model6 <- sapply(selected_vars,
                      function(var) if (var %in% names(coef(model6))) coef(model6)[var] else NA)
```

```r
coef_df <- data.frame(
  Urban1 = coef_model1,
  Urban2 = coef_model2,
  Suburban3 = coef_model3,
  Suburban4 = coef_model4,
  Rural5 = coef_model5,
  Rural6 = coef_model6
)

coef_df <- round(coef_df, 3)

print(coef_df)
```

```
##              Urban1 Urban2 Suburban3 Suburban4 Rural5 Rural6
## treat_rvpI    0.019  0.019     0.029     0.011  0.020  0.004
## treat_rvpII   0.008  0.005    -0.009    -0.023 -0.018 -0.016
## treat_rfg    -0.005 -0.038    -0.025    -0.058 -0.025 -0.045
## treat_CARB   -0.063 -0.079    -0.105    -0.095 -0.060 -0.068
## income       -1.307  0.438    -1.513    -0.677 -1.438  0.079
```

Table 4 : load package & data

```r
library(dplyr)
library(haven)
library(fixest)

setwd("/Users/jaeeun/Desktop/STAT156_GP/")
data <- read_dta("filtered.dta")
dif<- read_dta("filtered_Diffed.dta")
```

Table4 - (1)

```r
data <- data %>%
    select(lozone_max, lozone_8hr, treat_rvpI, treat_rvpII, treat_rfg, treat_CARB,income,
           DOW, DOY, panelid, region, year, TempMax, TempMin, Rain, Snow)

model1 <- feols(
  lozone_max ~ treat_rvpI + treat_rvpII + treat_rfg + treat_CARB | panelid + region^year,
  data = data
)

#summary(model1)
```

Table4 - (2)

```r
model2 <- feols(
  lozone_max ~ treat_rvpI + treat_rvpII + treat_rfg + treat_CARB | panelid + region^year +

    region^DOW + region^DOY + TempMax + TempMin + Rain + Snow,
  data = data
)

#summary(model2)
```

Table4 - (3)

```
data$income <- data$income/1000000000

model3 <- feols(
    lozone_max ~ treat_rvpI + treat_rvpII + treat_rfg + treat_CARB + income | panelid + region^year +

      region^DOW + region^DOY + TempMax + TempMin + Rain + Snow,
    data = data
)

#summary(model3)
```

Table4 - (4)

```
dif$incomeD <- dif$incomeD/1000000000

model4 <- feols(
  lozone_maxD ~
    treat_rvpID + treat_rvpIID + treat_rfgD  + treat_CARBD +

    incomeD +

    TrendRVP1D + TrendRVP2D + TrendRVP3D + TrendRVP4D +
    TrendRFG1D + TrendRFG2D + TrendRFG3D +
    TrendRVPRFG1D + TrendRVPRFG2D + TrendRVPRFG3D + TrendRVPRFG4D +
    TrendCARB4D + TrendCARBRFG4D +

    TempMinD + TempMax1D + TempMax2D + TempMax3D +
    TempMin2D + TempMin3D +
    TempMaxMinD + TempMaxL1D + TempMinL1D +
    TempMaxMaxL1D + TempMaxMinL1D +

    RainD + Rain2D + RainTempMaxD +
    SnowD + Snow2D +

    DOYD + DOYTempMax1D + DOYTempMax2D + DOYTempMax3D +
    DOYTempMin1D + DOYTempMin2D + DOYTempMin3D +
    DOYTempMaxMinD + DOYRain1D + DOYRain2D +
    DOYSnow1D + DOYSnow2D + DOYRainTempMaxD +
    DOYTempMaxL1D + DOYTempMinL1D +
    DOYTempMaxMaxL1D + DOYTempMinL1D +

    `_DMDOWXTempM_1D` + `_DMDOWXTempM_2D` + `_DMDOWXTempM_3D` +
    `_DMDOWXTempM_4D` + `_DMDOWXTempM_5D` + `_DMDOWXTempM_6D` +
    `_DmDOWXTempM_1D` + `_DmDOWXTempM_2D` + `_DmDOWXTempM_3D` +
    `_DmDOWXTempM_4D` + `_DmDOWXTempM_5D` + `_DmDOWXTempM_6D` +

    `_RYyear_1990D` + `_RYyear_1991D` + `_RYyear_1992D` +
    `_RYyear_1993D` + `_RYyear_1994D` + `_RYyear_1995D` +
    `_RYyear_1996D` + `_RYyear_1997D` + `_RYyear_1998D` +
    `_RYyear_1999D` + `_RYyear_2000D` + `_RYyear_2001D` +
    `_RYyear_2002D` + `_RYyear_2003D` +
```

```
    `_RWDOWXreg_1_2D` + `_RWDOWXreg_1_3D` + `_RWDOWXreg_1_4D` +
    `_RWDOWXreg_2_2D` + `_RWDOWXreg_2_3D` + `_RWDOWXreg_2_4D`,

  cluster = "A_StateYear",
  data = dif
)

#summary(model4)
```

Table4 - (5)

```
model5 <- feols(
  lozone_maxD ~
    treat_rvpID + treat_rvpIID + treat_rfgD  + treat_CARBD +

    incomeD +

    TrendRVP1D + TrendRVP2D + TrendRVP3D + TrendRVP4D +
    TrendRFG1D + TrendRFG2D + TrendRFG3D +
    TrendRVPRFG1D + TrendRVPRFG2D + TrendRVPRFG3D + TrendRVPRFG4D +
    TrendCARB4D + TrendCARBRFG4D +

    QTrendRVP1D + QTrendRVP2D + QTrendRVP3D + QTrendRVP4D +
    QTrendRFG1D + QTrendRFG2D + QTrendRFG3D +
    QTrendRVPRFG1D + QTrendRVPRFG2D + QTrendRVPRFG3D + QTrendRVPRFG4D +
    QTrendCARB4D +

    TempMinD + TempMax1D + TempMax2D + TempMax3D +
    TempMin2D + TempMin3D +
    TempMaxMinD + TempMaxL1D + TempMinL1D +
    TempMaxMaxL1D + TempMaxMinL1D +

    RainD + Rain2D + RainTempMaxD +
    SnowD + Snow2D +

    DOYD + DOYTempMax1D + DOYTempMax2D + DOYTempMax3D +
    DOYTempMin1D + DOYTempMin2D + DOYTempMin3D +
    DOYTempMaxMinD + DOYRain1D + DOYRain2D +
    DOYSnow1D + DOYSnow2D + DOYRainTempMaxD +
    DOYTempMaxL1D + DOYTempMinL1D +
    DOYTempMaxMaxL1D + DOYTempMinL1D +

    `_DMDOWXTempM_1D` + `_DMDOWXTempM_2D` + `_DMDOWXTempM_3D` +
    `_DMDOWXTempM_4D` + `_DMDOWXTempM_5D` + `_DMDOWXTempM_6D` +
    `_DmDOWXTempM_1D` + `_DmDOWXTempM_2D` + `_DmDOWXTempM_3D` +
    `_DmDOWXTempM_4D` + `_DmDOWXTempM_5D` + `_DmDOWXTempM_6D` +

    `_RYyear_1990D` + `_RYyear_1991D` + `_RYyear_1992D` +
    `_RYyear_1993D` + `_RYyear_1994D` + `_RYyear_1995D` +
    `_RYyear_1996D` + `_RYyear_1997D` + `_RYyear_1998D` +
    `_RYyear_1999D` + `_RYyear_2000D` + `_RYyear_2001D` +
    `_RYyear_2002D` + `_RYyear_2003D` +
```

```
    `_RWDOWXreg_1_2D` + `_RWDOWXreg_1_3D` + `_RWDOWXreg_1_4D` +
    `_RWDOWXreg_2_2D` + `_RWDOWXreg_2_3D` + `_RWDOWXreg_2_4D`,

  cluster = "A_StateYear",

  data = dif
)

#summary(model5)
```

Table4 - (6)

```
model6 <- feols(
  lozone_8hr ~ treat_rvpI + treat_rvpII + treat_rfg + treat_CARB | panelid + region^year,
  data = data
)

#summary(model6)
```

Table4 - (7)

```
model7 <- feols(
  lozone_8hr ~ treat_rvpI + treat_rvpII + treat_rfg + treat_CARB | panelid + region^year +

    region^DOW + region^DOY + TempMax + TempMin + Rain + Snow,
  data = data
)

#summary(model7)
```

Table4 - (8)

```
model8 <- feols(
  lozone_8hrD ~
    treat_rvpID + treat_rvpIID + treat_rfgD  + treat_CARBD +

    incomeD +

    TrendRVP1D + TrendRVP2D + TrendRVP3D + TrendRVP4D +
    TrendRFG1D + TrendRFG2D + TrendRFG3D +
    TrendRVPRFG1D + TrendRVPRFG2D + TrendRVPRFG3D + TrendRVPRFG4D +
    TrendCARB4D + TrendCARBRFG4D +

    QTrendRVP1D + QTrendRVP2D + QTrendRVP3D + QTrendRVP4D +
    QTrendRFG1D + QTrendRFG2D + QTrendRFG3D +
    QTrendRVPRFG1D + QTrendRVPRFG2D + QTrendRVPRFG3D + QTrendRVPRFG4D +
    QTrendCARB4D +

    TempMinD + TempMax1D + TempMax2D + TempMax3D +
    TempMin2D + TempMin3D +
    TempMaxMinD + TempMaxL1D + TempMinL1D +
    TempMaxMaxL1D + TempMaxMinL1D +
```

```r
    RainD + Rain2D + RainTempMaxD +
    SnowD + Snow2D +

    DOYD + DOYTempMax1D + DOYTempMax2D + DOYTempMax3D +
    DOYTempMin1D + DOYTempMin2D + DOYTempMin3D +
    DOYTempMaxMinD + DOYRain1D + DOYRain2D +
    DOYSnow1D + DOYSnow2D + DOYRainTempMaxD +
    DOYTempMaxL1D + DOYTempMinL1D +
    DOYTempMaxMaxL1D + DOYTempMinL1D +

    `_DMDOWXTempM_1D` + `_DMDOWXTempM_2D` + `_DMDOWXTempM_3D` +
    `_DMDOWXTempM_4D` + `_DMDOWXTempM_5D` + `_DMDOWXTempM_6D` +
    `_DmDOWXTempM_1D` + `_DmDOWXTempM_2D` + `_DmDOWXTempM_3D` +
    `_DmDOWXTempM_4D` + `_DmDOWXTempM_5D` + `_DmDOWXTempM_6D` +

    `_RYyear_1990D` + `_RYyear_1991D` + `_RYyear_1992D` +
    `_RYyear_1993D` + `_RYyear_1994D` + `_RYyear_1995D` +
    `_RYyear_1996D` + `_RYyear_1997D` + `_RYyear_1998D` +
    `_RYyear_1999D` + `_RYyear_2000D` + `_RYyear_2001D` +
    `_RYyear_2002D` + `_RYyear_2003D` +

    `_RWDOWXreg_1_2D` + `_RWDOWXreg_1_3D` + `_RWDOWXreg_1_4D` +
    `_RWDOWXreg_2_2D` + `_RWDOWXreg_2_3D` + `_RWDOWXreg_2_4D`,

  cluster = "A_StateYear",

  data = dif
)

#summary(model8)
```

Table4 - fi

```r
coef_model1 <- coef(model1)
coef_model1["income"] <- NA

coef_model2 <- coef(model2)
coef_model2["income"] <- NA

coef_model3 <- coef(model3)

coef_model6 <- coef(model6)
coef_model6["income"] <- NA

coef_model7 <- coef(model7)
coef_model7["income"] <- NA

selected_vars <- c("treat_rvpID", "treat_rvpIID", "treat_rfgD", "treat_CARBD", "incomeD")

coef_model4 <- sapply(selected_vars,

                      function(var) if (var %in% names(coef(model4))) coef(model4)[var] else NA)
coef_model5 <- sapply(selected_vars,
```

```
                      function(var) if (var %in% names(coef(model5))) coef(model5)[var] else NA)
coef_model8 <- sapply(selected_vars,
                      function(var) if (var %in% names(coef(model8))) coef(model8)[var] else NA)

coef_df <- data.frame(
  Model1 = coef_model1,
  Model2 = coef_model2,
  Model3 = coef_model3,
  Model4 = coef_model4,
  Model5 = coef_model5,
  Model6 = coef_model6,
  Model7 = coef_model7,
  Model8 = coef_model8
)

coef_df <- round(coef_df, 3)

print(coef_df)
```

```
##              Model1 Model2 Model3 Model4 Model5 Model6 Model7 Model8
## treat_rvpI   -0.009  0.000 -0.001 -0.008 -0.007 -0.007  0.003 -0.006
## treat_rvpII -0.009 -0.016 -0.011 -0.023 -0.033 -0.009 -0.015 -0.033
## treat_rfg    -0.031 -0.036 -0.023 -0.066 -0.065 -0.031 -0.036 -0.071
## treat_CARB   -0.148 -0.132 -0.108 -0.151 -0.159 -0.139 -0.124 -0.163
## income           NA     NA -1.677 -0.233 -0.252     NA     NA -0.042
```

IPW estimators :

load packages & data

```
library(dplyr)
library(haven)
library(fixest)
library(MASS)
library(WeightIt)
library(estimatr)
library(nnet)

setwd("/Users/jaeeun/Desktop/STAT156_GP/")
data <- read_dta("DD_AnalysisDataset.dta")
data$income <- data$income/1000000000
```

IPW modeling - using filtered data in table 4

```
data <- read_dta("filtered.dta")
dif<- read_dta("filtered_Diffed.dta")

data<-data[order(data$panelid,data$Date), ]
dif<-dif[order(dif$panelid),]
#summary(data$panelid == dif$panelid)
```

```r
#summary(data$fips == dif$A_fips)
dif$incomeD <- dif$incomeD/1000000000
filtered<-cbind(dif,urban = data$urban)
```

Calculate propensity scores

```r
propensity_model <- multinom(urban ~ treat_rvpID + treat_rvpIID + treat_rfgD  + treat_CARBD +

    incomeD +

    TrendRVP1D + TrendRVP2D + TrendRVP3D + TrendRVP4D +
    TrendRFG1D + TrendRFG2D + TrendRFG3D +
    TrendRVPRFG1D + TrendRVPRFG2D + TrendRVPRFG3D + TrendRVPRFG4D +
    TrendCARB4D + TrendCARBRFG4D +

    QTrendRVP1D + QTrendRVP2D + QTrendRVP3D + QTrendRVP4D +
    QTrendRFG1D + QTrendRFG2D + QTrendRFG3D +
    QTrendRVPRFG1D + QTrendRVPRFG2D + QTrendRVPRFG3D + QTrendRVPRFG4D +
    QTrendCARB4D +

    TempMinD + TempMax1D + TempMax2D + TempMax3D +
    TempMin2D + TempMin3D +
    TempMaxMinD + TempMaxL1D + TempMinL1D +
    TempMaxMaxL1D + TempMaxMinL1D +

    RainD + Rain2D + RainTempMaxD +
    SnowD + Snow2D +

    DOYD + DOYTempMax1D + DOYTempMax2D + DOYTempMax3D +
    DOYTempMin1D + DOYTempMin2D + DOYTempMin3D +
    DOYTempMaxMinD + DOYRain1D + DOYRain2D +
    DOYSnow1D + DOYSnow2D + DOYRainTempMaxD +
    DOYTempMaxL1D + DOYTempMinL1D +
    DOYTempMaxMaxL1D + DOYTempMinL1D +

    `_DMDOWXTempM_1D` + `_DMDOWXTempM_2D` + `_DMDOWXTempM_3D` +
    `_DMDOWXTempM_4D` + `_DMDOWXTempM_5D` + `_DMDOWXTempM_6D` +
    `_DmDOWXTempM_1D` + `_DmDOWXTempM_2D` + `_DmDOWXTempM_3D` +
    `_DmDOWXTempM_4D` + `_DmDOWXTempM_5D` + `_DmDOWXTempM_6D` +

    `_RYyear_1990D` + `_RYyear_1991D` + `_RYyear_1992D` +
    `_RYyear_1993D` + `_RYyear_1994D` + `_RYyear_1995D` +
    `_RYyear_1996D` + `_RYyear_1997D` + `_RYyear_1998D` +
    `_RYyear_1999D` + `_RYyear_2000D` + `_RYyear_2001D` +
    `_RYyear_2002D` + `_RYyear_2003D` +

    `_RWDOWXreg_1_2D` + `_RWDOWXreg_1_3D` + `_RWDOWXreg_1_4D` +
    `_RWDOWXreg_2_2D` + `_RWDOWXreg_2_3D` + `_RWDOWXreg_2_4D`,

  cluster = "A_StateYear",

  data = filtered)
```

```
## # weights:  288 (190 variable)
```

```
## initial  value 499960.874775
## final  value 499960.874776
## converged
```

```r
filtered$pscore <- predict(propensity_model, type = "probs")[,1]
filtered$weight <- ifelse(filtered$urban == 1, 1 / filtered$pscore, 1 / (1 - filtered$pscore))

model <- feols(lozone_maxD ~
    treat_rvpID + treat_rvpIID + treat_rfgD  + treat_CARBD +

    incomeD +

    TrendRVP1D + TrendRVP2D + TrendRVP3D + TrendRVP4D +
    TrendRFG1D + TrendRFG2D + TrendRFG3D +
    TrendRVPRFG1D + TrendRVPRFG2D + TrendRVPRFG3D + TrendRVPRFG4D +
    TrendCARB4D + TrendCARBRFG4D +

    QTrendRVP1D + QTrendRVP2D + QTrendRVP3D + QTrendRVP4D +
    QTrendRFG1D + QTrendRFG2D + QTrendRFG3D +
    QTrendRVPRFG1D + QTrendRVPRFG2D + QTrendRVPRFG3D + QTrendRVPRFG4D +
    QTrendCARB4D +

    TempMinD + TempMax1D + TempMax2D + TempMax3D +
    TempMin2D + TempMin3D +
    TempMaxMinD + TempMaxL1D + TempMinL1D +
    TempMaxMaxL1D + TempMaxMinL1D +

    RainD + Rain2D + RainTempMaxD +
    SnowD + Snow2D +

    DOYD + DOYTempMax1D + DOYTempMax2D + DOYTempMax3D +
    DOYTempMin1D + DOYTempMin2D + DOYTempMin3D +
    DOYTempMaxMinD + DOYRain1D + DOYRain2D +
    DOYSnow1D + DOYSnow2D + DOYRainTempMaxD +
    DOYTempMaxL1D + DOYTempMinL1D +
    DOYTempMaxMaxL1D + DOYTempMinL1D +

    `_DMDOWXTempM_1D` + `_DMDOWXTempM_2D` + `_DMDOWXTempM_3D` +
    `_DMDOWXTempM_4D` + `_DMDOWXTempM_5D` + `_DMDOWXTempM_6D` +
    `_DmDOWXTempM_1D` + `_DmDOWXTempM_2D` + `_DmDOWXTempM_3D` +
    `_DmDOWXTempM_4D` + `_DmDOWXTempM_5D` + `_DmDOWXTempM_6D` +

    `_RYyear_1990D` + `_RYyear_1991D` + `_RYyear_1992D` +
    `_RYyear_1993D` + `_RYyear_1994D` + `_RYyear_1995D` +
    `_RYyear_1996D` + `_RYyear_1997D` + `_RYyear_1998D` +
    `_RYyear_1999D` + `_RYyear_2000D` + `_RYyear_2001D` +
    `_RYyear_2002D` + `_RYyear_2003D` +

    `_RWDOWXreg_1_2D` + `_RWDOWXreg_1_3D` + `_RWDOWXreg_1_4D` +
    `_RWDOWXreg_2_2D` + `_RWDOWXreg_2_3D` + `_RWDOWXreg_2_4D`,

  cluster = "A_StateYear",

  data =filtered)
```

```r
weighted_model <- feols(lozone_maxD ~
    treat_rvpID + treat_rvpIID + treat_rfgD  + treat_CARBD +

    incomeD +

    TrendRVP1D + TrendRVP2D + TrendRVP3D + TrendRVP4D +
    TrendRFG1D + TrendRFG2D + TrendRFG3D +
    TrendRVPRFG1D + TrendRVPRFG2D + TrendRVPRFG3D + TrendRVPRFG4D +
    TrendCARB4D + TrendCARBRFG4D +

    QTrendRVP1D + QTrendRVP2D + QTrendRVP3D + QTrendRVP4D +
    QTrendRFG1D + QTrendRFG2D + QTrendRFG3D +
    QTrendRVPRFG1D + QTrendRVPRFG2D + QTrendRVPRFG3D + QTrendRVPRFG4D +
    QTrendCARB4D +

    TempMinD + TempMax1D + TempMax2D + TempMax3D +
    TempMin2D + TempMin3D +
    TempMaxMinD + TempMaxL1D + TempMinL1D +
    TempMaxMaxL1D + TempMaxMinL1D +

    RainD + Rain2D + RainTempMaxD +
    SnowD + Snow2D +

    DOYD + DOYTempMax1D + DOYTempMax2D + DOYTempMax3D +
    DOYTempMin1D + DOYTempMin2D + DOYTempMin3D +
    DOYTempMaxMinD + DOYRain1D + DOYRain2D +
    DOYSnow1D + DOYSnow2D + DOYRainTempMaxD +
    DOYTempMaxL1D + DOYTempMinL1D +
    DOYTempMaxMaxL1D + DOYTempMinL1D +

    `_DMDOWXTempM_1D` + `_DMDOWXTempM_2D` + `_DMDOWXTempM_3D` +
    `_DMDOWXTempM_4D` + `_DMDOWXTempM_5D` + `_DMDOWXTempM_6D` +
    `_DmDOWXTempM_1D` + `_DmDOWXTempM_2D` + `_DmDOWXTempM_3D` +
    `_DmDOWXTempM_4D` + `_DmDOWXTempM_5D` + `_DmDOWXTempM_6D` +

    `_RYyear_1990D` + `_RYyear_1991D` + `_RYyear_1992D` +
    `_RYyear_1993D` + `_RYyear_1994D` + `_RYyear_1995D` +
    `_RYyear_1996D` + `_RYyear_1997D` + `_RYyear_1998D` +
    `_RYyear_1999D` + `_RYyear_2000D` + `_RYyear_2001D` +
    `_RYyear_2002D` + `_RYyear_2003D` +

    `_RWDOWXreg_1_2D` + `_RWDOWXreg_1_3D` + `_RWDOWXreg_1_4D` +
    `_RWDOWXreg_2_2D` + `_RWDOWXreg_2_3D` + `_RWDOWXreg_2_4D`,
                          weights = ~weight, cluster = "A_StateYear",
                          data =filtered)


#summary(model)
#summary(weighted_model)
model_coef <- as.data.frame(coef(model))
weight_coef <- as.data.frame(coef(weighted_model))
coef <- cbind(model_coef, weight_coef)
```

8hr

```
model_8hr <- feols(lozone_8hrD ~
    treat_rvpID + treat_rvpIID + treat_rfgD  + treat_CARBD +

    incomeD +

    TrendRVP1D + TrendRVP2D + TrendRVP3D + TrendRVP4D +
    TrendRFG1D + TrendRFG2D + TrendRFG3D +
    TrendRVPRFG1D + TrendRVPRFG2D + TrendRVPRFG3D + TrendRVPRFG4D +
    TrendCARB4D + TrendCARBRFG4D +

    QTrendRVP1D + QTrendRVP2D + QTrendRVP3D + QTrendRVP4D +
    QTrendRFG1D + QTrendRFG2D + QTrendRFG3D +
    QTrendRVPRFG1D + QTrendRVPRFG2D + QTrendRVPRFG3D + QTrendRVPRFG4D +
    QTrendCARB4D +

    TempMinD + TempMax1D + TempMax2D + TempMax3D +
    TempMin2D + TempMin3D +
    TempMaxMinD + TempMaxL1D + TempMinL1D +
    TempMaxMaxL1D + TempMaxMinL1D +

    RainD + Rain2D + RainTempMaxD +
    SnowD + Snow2D +

    DOYD + DOYTempMax1D + DOYTempMax2D + DOYTempMax3D +
    DOYTempMin1D + DOYTempMin2D + DOYTempMin3D +
    DOYTempMaxMinD + DOYRain1D + DOYRain2D +
    DOYSnow1D + DOYSnow2D + DOYRainTempMaxD +
    DOYTempMaxL1D + DOYTempMinL1D +
    DOYTempMaxMaxL1D + DOYTempMinL1D +

    `_DMDOWXTempM_1D` + `_DMDOWXTempM_2D` + `_DMDOWXTempM_3D` +
    `_DMDOWXTempM_4D` + `_DMDOWXTempM_5D` + `_DMDOWXTempM_6D` +
    `_DmDOWXTempM_1D` + `_DmDOWXTempM_2D` + `_DmDOWXTempM_3D` +
    `_DmDOWXTempM_4D` + `_DmDOWXTempM_5D` + `_DmDOWXTempM_6D` +

    `_RYyear_1990D` + `_RYyear_1991D` + `_RYyear_1992D` +
    `_RYyear_1993D` + `_RYyear_1994D` + `_RYyear_1995D` +
    `_RYyear_1996D` + `_RYyear_1997D` + `_RYyear_1998D` +
    `_RYyear_1999D` + `_RYyear_2000D` + `_RYyear_2001D` +
    `_RYyear_2002D` + `_RYyear_2003D` +

    `_RWDOWXreg_1_2D` + `_RWDOWXreg_1_3D` + `_RWDOWXreg_1_4D` +
    `_RWDOWXreg_2_2D` + `_RWDOWXreg_2_3D` + `_RWDOWXreg_2_4D`,

  cluster = "A_StateYear",

  data =filtered)

weighted_model_8hr <- feols(lozone_8hrD ~
    treat_rvpID + treat_rvpIID + treat_rfgD  + treat_CARBD +

    incomeD +
```

```
    TrendRVP1D + TrendRVP2D + TrendRVP3D + TrendRVP4D +
    TrendRFG1D + TrendRFG2D + TrendRFG3D +
    TrendRVPRFG1D + TrendRVPRFG2D + TrendRVPRFG3D + TrendRVPRFG4D +
    TrendCARB4D + TrendCARBRFG4D +

    QTrendRVP1D + QTrendRVP2D + QTrendRVP3D + QTrendRVP4D +
    QTrendRFG1D + QTrendRFG2D + QTrendRFG3D +
    QTrendRVPRFG1D + QTrendRVPRFG2D + QTrendRVPRFG3D + QTrendRVPRFG4D +
    QTrendCARB4D +

    TempMinD + TempMax1D + TempMax2D + TempMax3D +
    TempMin2D + TempMin3D +
    TempMaxMinD + TempMaxL1D + TempMinL1D +
    TempMaxMaxL1D + TempMaxMinL1D +

    RainD + Rain2D + RainTempMaxD +
    SnowD + Snow2D +

    DOYD + DOYTempMax1D + DOYTempMax2D + DOYTempMax3D +
    DOYTempMin1D + DOYTempMin2D + DOYTempMin3D +
    DOYTempMaxMinD + DOYRain1D + DOYRain2D +
    DOYSnow1D + DOYSnow2D + DOYRainTempMaxD +
    DOYTempMaxL1D + DOYTempMinL1D +
    DOYTempMaxMaxL1D + DOYTempMinL1D +

    `_DMDOWXTempM_1D` + `_DMDOWXTempM_2D` + `_DMDOWXTempM_3D` +
    `_DMDOWXTempM_4D` + `_DMDOWXTempM_5D` + `_DMDOWXTempM_6D` +
    `_DmDOWXTempM_1D` + `_DmDOWXTempM_2D` + `_DmDOWXTempM_3D` +
    `_DmDOWXTempM_4D` + `_DmDOWXTempM_5D` + `_DmDOWXTempM_6D` +

    `_RYyear_1990D` + `_RYyear_1991D` + `_RYyear_1992D` +
    `_RYyear_1993D` + `_RYyear_1994D` + `_RYyear_1995D` +
    `_RYyear_1996D` + `_RYyear_1997D` + `_RYyear_1998D` +
    `_RYyear_1999D` + `_RYyear_2000D` + `_RYyear_2001D` +
    `_RYyear_2002D` + `_RYyear_2003D` +

    `_RWDOWXreg_1_2D` + `_RWDOWXreg_1_3D` + `_RWDOWXreg_1_4D` +
    `_RWDOWXreg_2_2D` + `_RWDOWXreg_2_3D` + `_RWDOWXreg_2_4D`,
                          weights = ~weight, cluster = "A_StateYear",
                          data =filtered)

#summary(model_8hr)
#summary(weighted_model_8hr)

model_coef_8hr <- as.data.frame(coef(model_8hr))
weight_coef_8hr <- as.data.frame(coef(weighted_model_8hr))
coef_8hr <- cbind(model_coef_8hr, weight_coef_8hr)
```

plot

```
df_max <- round(coef[rownames(coef) %in% c("treat_rvpID",
                                           "treat_rvpIID","treat_rfgD","treat_CARBD","incomeD"), ],3)
```

```r
rownames(df_max) <- c("RVPI","RVPII", "RFG","CARB","Income")
colnames(df_max) <- c("Model","IPW")

df_8hr <- round(coef_8hr[rownames(coef_8hr) %in% c("treat_rvpID", "treat_rvpIID",

                                                    "treat_rfgD", "treat_CARBD","incomeD"), ],3)
rownames(df_8hr) <- c("RVPI","RVPII", "RFG","CARB","Income")
colnames(df_8hr) <- c("Model","IPW")

print(df_max)
```

```
##         Model    IPW
## RVPI   -0.007 -0.008
## RVPII  -0.033 -0.031
## RFG    -0.065 -0.063
## CARB   -0.159 -0.164
## Income -0.252  0.104
```

```r
print(df_8hr)
```

```
##         Model    IPW
## RVPI   -0.006 -0.006
## RVPII  -0.033 -0.030
## RFG    -0.071 -0.069
## CARB   -0.163 -0.165
## Income -0.042  0.322
```