

ENGRD 302W New Data

Veronica

2025-01-20

Importing Libraries

```
library(ggplot2)
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.3
```

```
## — Attaching core tidyverse packages ————— tidyverse 2.0.0 —
## ✓ dplyr      1.1.3      ✓ readr      2.1.4
## ✓ forcats    1.0.0      ✓ stringr   1.5.0
## ✓ lubridate  1.9.2      ✓ tibble    3.2.1
## ✓ purrr      1.0.2      ✓ tidyr     1.3.0
## — Conflicts ————— tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggthemes)
```

```
## Warning: package 'ggthemes' was built under R version 4.3.3
```

```
library(dplyr)
library(maps)
```

```
## Warning: package 'maps' was built under R version 4.3.3
```

```
##
## Attaching package: 'maps'
##
## The following object is masked from 'package:purrr':
##
##     map
```

```
library(ggmap)
```

```
## Warning: package 'ggmap' was built under R version 4.3.3
```

```
## i Google's Terms of Service: <https://mapsplatform.google.com>
##   Stadia Maps' Terms of Service: <https://stadiamaps.com/terms-of-service/>
##   OpenStreetMap's Tile Usage Policy: <https://operations.osmfoundation.org/policies/tiles/>
## i Please cite ggmap if you use it! Use `citation("ggmap")` for details.
```

```
library(plotly)
```

```
## Warning: package 'plotly' was built under R version 4.3.3
```

```
##
## Attaching package: 'plotly'
##
## The following object is masked from 'package:ggmap':
##
##   wind
##
## The following object is masked from 'package:ggplot2':
##
##   last_plot
##
## The following object is masked from 'package:stats':
##
##   filter
##
## The following object is masked from 'package:graphics':
##
##   layout
```

```
library(stringr)
library(lubridate)
library(ggrepel)
```

```
## Warning: package 'ggrepel' was built under R version 4.3.3
```

```
## setting work directory
setwd("C:/Users/13015/OneDrive - Emory University/Documents/Emory/Fall_23/ENGRD 302W/engrd302wfa
112023")
```

Cleaning County Code

```
## Loading GA emission data
GAED <- read.csv("C:/Users/13015/OneDrive - Emory University/Documents/Emory/Fall_23/ENGRD 302W/
engrd302wfall2023/Data/GAED.csv")

## Loading GA county data
GA_counties <- read.csv("C:/Users/13015/OneDrive - Emory University/Documents/Emory/Fall_23/ENGR
D 302W/engrd302wfall2023/Data/GA_counties.csv")

## standardizing county code
GA_counties$county_code <- as.character(GA_counties$county_code)
GA_counties$county_name <- tolower(GA_counties$county_name)

GAED <- GAED %>%
  mutate(county_code = gsub("g", "", geocode)) %>%
  select(-geocode)

GAED$county_code <-
  as.character(GAED$county_code)
head(GAED)
```

```
##      ldate sector  datavalue    varname county_code
## 1 1/1/2005    ac  12395.564 emi.to.ac.mt      13001
## 2 1/1/2005    cc   3331.363 emi.to.cc.mt      13001
## 3 1/1/2005    fc -20512.363 emi.to.fc.mt      13001
## 4 1/1/2005    gc   4077.993 emi.to.gc.mt      13001
## 5 1/1/2005    ic   1026.483 emi.to.ic.mt      13001
## 6 1/1/2005    rc   6685.857 emi.to.rc.mt      13001
```

GA County & Time Data

```
## creating GA df
GA_join <- left_join(GAED, GA_counties, by = c("county_code" = "county_code"))

## rename columnm
GA_join <- GA_join %>%
  select(-sector, -varname) %>%
  rename("date" = "ldate")

## create date variables
GA_join$date <- as.Date(GA_join$date, "%m/%d/%Y")
GA_join$year <- year(ymd(GA_join$date))
GA_join$month <- month(ymd(GA_join$date))
GA_join$day <- day(ymd(GA_join$date))

## subsetting data
GA_join <- GA_join %>%
  select(-date) %>%
  rename("emission_value" = "datavalue") %>%
  filter(year %in% c("2013", "2014", "2015", "2016", "2017", "2018", "2019", "2020", "2021", "2022", "2023"))

## fixing dekalb county
GA_join$county_name <- replace(GA_join$county_name, GA_join$county_name=="de kalb", "dekalb")

head(GA_join)
```

```
##   emission_value county_code county_name year month day
## 1      9791.7624      13001    applying 2013     1    1
## 2      2188.6839      13001    applying 2013     1    1
## 3     -16372.6834      13001    applying 2013     1    1
## 4       5111.4737      13001    applying 2013     1    1
## 5        811.5281      13001    applying 2013     1    1
## 6       5834.9156      13001    applying 2013     1    1
```

```
summary(GA_join)
```

```
## emission_value      county_code      county_name      year
## Min.      :-100171.4 Length:138012 Length:138012 Min.      :2013
## 1st Qu.:   197.8 Class :character Class :character 1st Qu.:2015
## Median :   3432.1 Mode  :character Mode  :character Median :2018
## Mean    :   16281.2                      Mean    :2018
## 3rd Qu.:   12758.3                      3rd Qu.:2020
## Max.    :  1785955.0                      Max.    :2023
##      month          day
## Min.    : 1.000 Min.    :1
## 1st Qu.: 3.000 1st Qu.:1
## Median : 6.000 Median :1
## Mean    : 6.371 Mean    :1
## 3rd Qu.: 9.000 3rd Qu.:1
## Max.    :12.000 Max.    :1
```

GA Geographical Data

```
## creating geographical data frame
state_df <- map_data("state") %>%
  filter(region == "georgia")

## creating county data frame
county_df <- map_data("county") %>%
  filter(region == "georgia")

## Loading GA population data
GA_population <- read.csv("C:/Users/13015/OneDrive - Emory University/Documents/Emory/Fall_23/ENGRD 302W/engrd302wfall2023/Data/GA_population.csv")

## cleaning population data
GA_population <- GA_population %>%
  select(-X) %>%
  rename("county_name" = "County")

GA_population$county_name <- tolower(GA_population$county_name)

head(GA_population)
```

```
## county_name Population
## 1    appling      18236
## 2    atkinson     8375
## 3     bacon      11096
## 4     baker       3451
## 5    baldwin     45720
## 6     banks      18395
```

```
## fixing dekalb county
county_df$subregion <- replace(county_df$subregion, county_df$subregion == "de kalb", "dekalb")

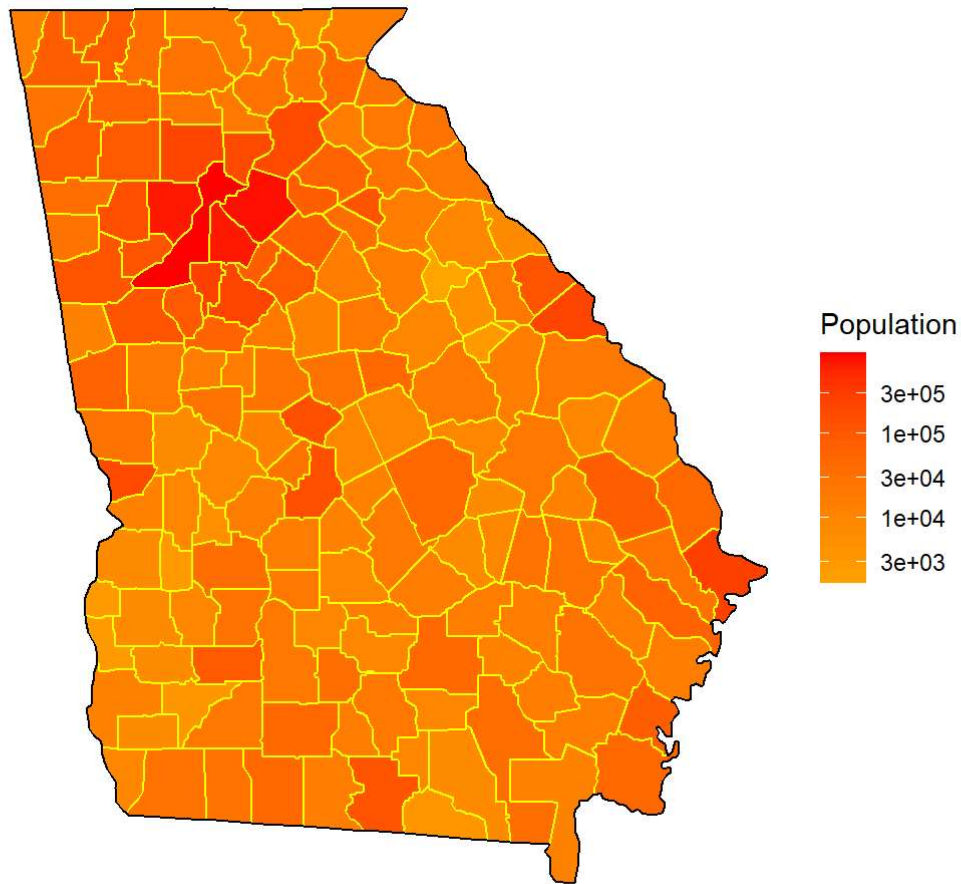
## joining county and population data
GA_geographical <- left_join(GA_population, county_df, by = c("county_name" = "subregion"))

head(GA_geographical)
```

```
##   county_name Population      long      lat group order region
## 1   appling      18236 -82.44862 31.94813   358 14727 georgia
## 2   appling      18236 -82.42570 31.94813   358 14728 georgia
## 3   appling      18236 -82.40852 31.94240   358 14729 georgia
## 4   appling      18236 -82.39706 31.94240   358 14730 georgia
## 5   appling      18236 -82.38560 31.93094   358 14731 georgia
## 6   appling      18236 -82.35122 31.91948   358 14732 georgia
```

Data Visualization

```
## mapping GA population data
GA_population_map <- ggplot(GA_geographical, aes(x=long, y=lat)) +
  geom_polygon(aes(group=group, fill = Population), color="yellow") +
  scale_fill_gradient(low="orange", high="red", trans = "log10") +
  geom_polygon(data = state_df, aes(group=group), color = "black", fill = NA) +
  theme_void()+ coord_quickmap()
print(GA_population_map)
```



```
ggsave("GA Population Map.png", GA_population_map)
```

```
## Saving 7 x 5 in image
```

```
## subsetting 2013
GA_2013 <- GA_join %>%
  filter(year %in% c("2013"))
head(GA_2013)
```

```
##   emission_value county_code county_name year month day
## 1      9791.7624      13001    appling 2013     1    1
## 2      2188.6839      13001    appling 2013     1    1
## 3     -16372.6834      13001    appling 2013     1    1
## 4       5111.4737      13001    appling 2013     1    1
## 5        811.5281      13001    appling 2013     1    1
## 6       5834.9156      13001    appling 2013     1    1
```

```
summary(GA_2013)
```

```
## emission_value      county_code      county_name      year
## Min.   : -93435.6    Length:13356      Length:13356      Min.   :2013
## 1st Qu.:  222.2     Class :character  Class :character  1st Qu.:2013
## Median :  3713.3    Mode  :character  Mode  :character  Median :2013
## Mean   :  16419.3                                Mean   :2013
## 3rd Qu.:  13398.3                                3rd Qu.:2013
## Max.   :1462678.5                                Max.   :2013
##      month          day
## Min.   : 1.00      Min.   :1
## 1st Qu.: 3.75      1st Qu.:1
## Median : 6.50      Median :1
## Mean   : 6.50      Mean   :1
## 3rd Qu.: 9.25      3rd Qu.:1
## Max.   :12.00      Max.   :1
```

```
## subsetting 2023
GA_2023 <- GA_join %>%
  filter(year %in% c("2023"))
head(GA_2023)
```

```
## emission_value county_code county_name year month day
## 1      10920.1165      13001      appling 2023      1      1
## 2      1634.9650      13001      appling 2023      1      1
## 3     -19926.8318      13001      appling 2023      1      1
## 4      4839.6277      13001      appling 2023      1      1
## 5       683.1551      13001      appling 2023      1      1
## 6      3304.8247      13001      appling 2023      1      1
```

```
summary(GA_2023)
```

```
## emission_value      county_code      county_name      year
## Min.   : -85554.2    Length:4452      Length:4452      Min.   :2023
## 1st Qu.:  170.4     Class :character  Class :character  1st Qu.:2023
## Median :  2622.3    Mode  :character  Mode  :character  Median :2023
## Mean   :  15548.0                                Mean   :2023
## 3rd Qu.:  11364.2                                3rd Qu.:2023
## Max.   :1343732.0                                Max.   :2023
##      month          day
## Min.   :1.00      Min.   :1
## 1st Qu.:1.75      1st Qu.:1
## Median :2.50      Median :1
## Mean   :2.50      Mean   :1
## 3rd Qu.:3.25      3rd Qu.:1
## Max.   :4.00      Max.   :1
```


GA Yearly Average County Emission

```
## cleaning GA
GA_join <- GA_join %>%
  select(emission_value, county_name, year) %>%
  group_by(year, county_name) %>%
  mutate(emission = mean(emission_value)) %>%
  select(-emission_value) %>%
  distinct()
head(GA_join)
```

```
## # A tibble: 6 × 3
## # Groups:   year, county_name [6]
##   county_name year emission
##   <chr>      <dbl>    <dbl>
## 1 appling    2013    -174.
## 2 appling    2014     716.
## 3 appling    2015     443.
## 4 appling    2016     423.
## 5 appling    2017     575.
## 6 appling    2018     496.
```

Top 10 Emission Production

```
## subsetting top 10 counties
GA_join_top10 <- GA_join %>%
  group_by(county_name) %>%
  summarize(mean(emission)) %>%
  rename("emission" = "mean(emission)") %>%
  arrange(-emission) %>%
  top_n(n = 10, wt = emission)
print(GA_join_top10)
```

```
## # A tibble: 10 × 2
##   county_name emission
##   <chr>          <dbl>
## 1 fulton        388198.
## 2 gwinnett      230317.
## 3 cobb          209371.
## 4 dekalb        208566.
## 5 richmond      120962.
## 6 chatham       119372.
## 7 clayton       82806.
## 8 forsyth       70760.
## 9 hall          63835.
## 10 henry        57941.
```

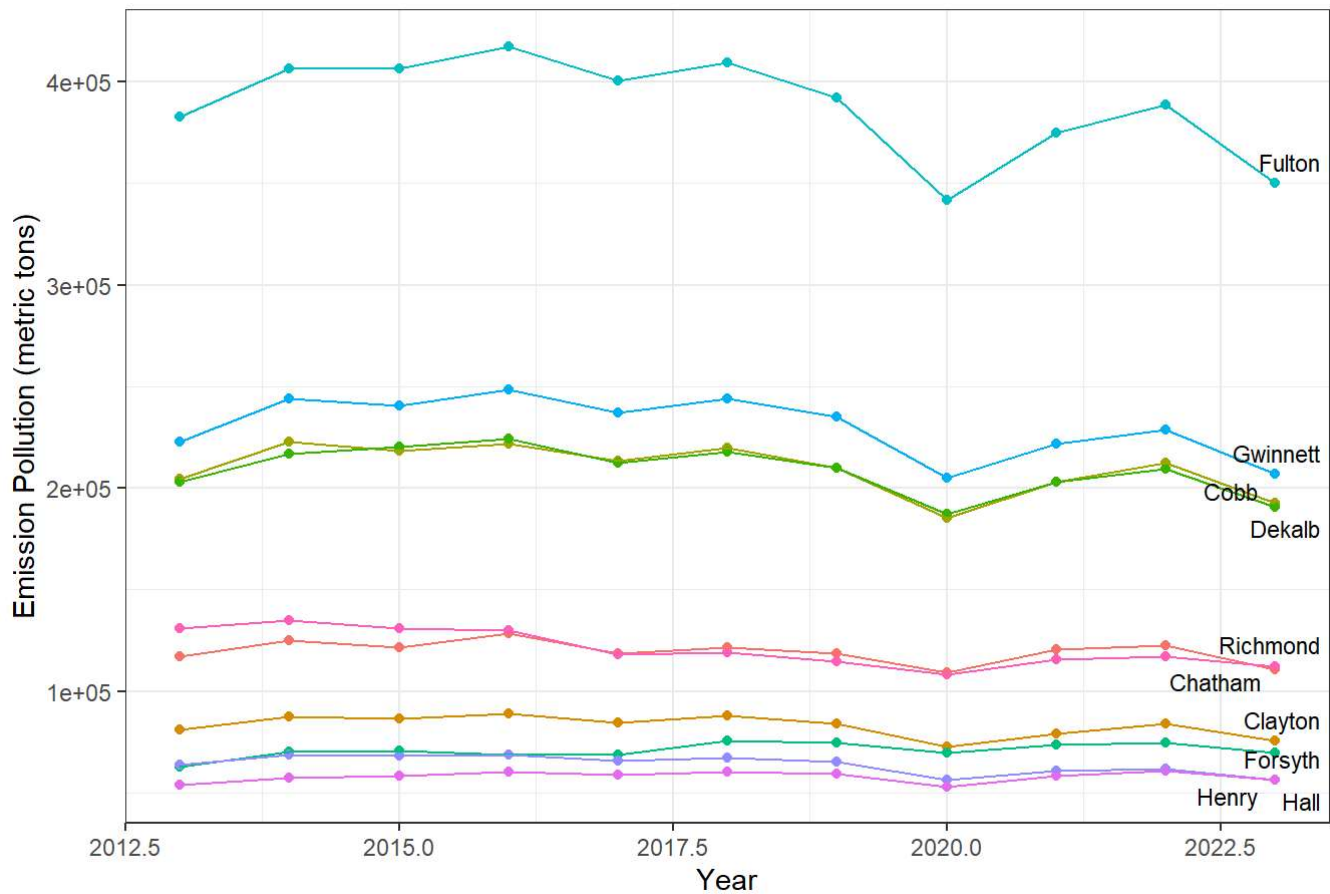
Creating & Plotting Longitudinal Data Frame

```
## subsetting longitudinal data
GA_top10_longitudinal <- GA_join %>%
  filter(county_name %in% c("fulton", "gwinnett", "cobb", "dekalb", "richmond", "chatham", "clayton", "forsyth", "hall", "henry")) %>%
  mutate(county_name = str_to_title(county_name))

## adding value to each line
df_end <- GA_top10_longitudinal %>%
  filter(year == 2023)

## plotting longitudinal df
GA_top10_plot <- ggplot(data = GA_top10_longitudinal, aes(x = year, y = emission, color = county_name)) +
  geom_point() +
  geom_line(aes(group = county_name)) +
  geom_text_repel(
    aes(label = county_name), data = df_end,
    fontface = "plain", color = "black", size = 3
  ) +
  labs(x="Year", y="Emission Pollution (metric tons)", title= "GA County Emission Pollution 2013-2023") +
  theme_bw() +
  theme(legend.position = "none")
print(GA_top10_plot)
```

GA County Emission Pollution 2013-2023



```
ggsave("GA County Emission Pollution 2013-2023 Line Graph.png", GA_top10_plot)
```

```
## Saving 7 x 5 in image
```