Data Analysis and viz

Faustin Kambale

2025-03-13

Insert libraries

```r
library (here)
```

```
## here() starts at /Users/faustinkambale/Library/CloudStorage/OneDrive-DukeUniversity/Duke/Classes/Spr
```

```r
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
```

```r
library(readxl)
library (ggplot2)
library(readr)
library(tidyr)
```

insert Datasets

```r
setwd('/Users/faustinkambale/Library/CloudStorage/OneDrive-DukeUniversity/Duke/Classes/Spring 2025/Stat
getwd ()
```

```
## [1] "/Users/faustinkambale/Library/CloudStorage/OneDrive-DukeUniversity/Duke/Classes/Spring 2025/Stat
```

```r
ev_by_state <- read.csv("ev_by_state.csv")
emmission_station <- read.csv('/Users/faustinkambale/Library/CloudStorage/OneDrive-DukeUniversity/Duke/C
ten_states <- read.csv('/Users/faustinkambale/Library/CloudStorage/OneDrive-DukeUniversity/Duke/Classes/
```

Some plots

```r
## Data viz for Alabama (can reproduce for other states)

alabama <- emmission_station %>%
  select ("year", "alab_emm", "alabama_st")

# Create a scaling factor for the second axis
scale_factor <- max(alabama$alab_emm) / max(alabama$alabama_st)

## 1. Emission trends overtime

alabama_trends <- ggplot(alabama, aes(x = year)) +
  geom_line(aes(y = alab_emm, color = "Emissions"), size = 1) +
  geom_point(aes(y = alab_emm, color = "Emissions"), size = 2) +
  geom_line(aes(y = alabama_st * scale_factor, color = "Stations"), size = 1) +
  geom_point(aes(y = alabama_st * scale_factor, color = "Stations"), size = 2) +
  scale_y_continuous(
    name = "Emissions",
    sec.axis = sec_axis(~ . / scale_factor, name = "Number of Stations")
  ) +
  scale_color_manual(values = c("Emissions" = "darkblue", "Stations" = "red")) +
  labs(title = "Emissions and Stations in Alabama (2007-2022)",
       x = "Year",
       color = "Metric") +
  theme_minimal() +
  theme(legend.position = "bottom")
```
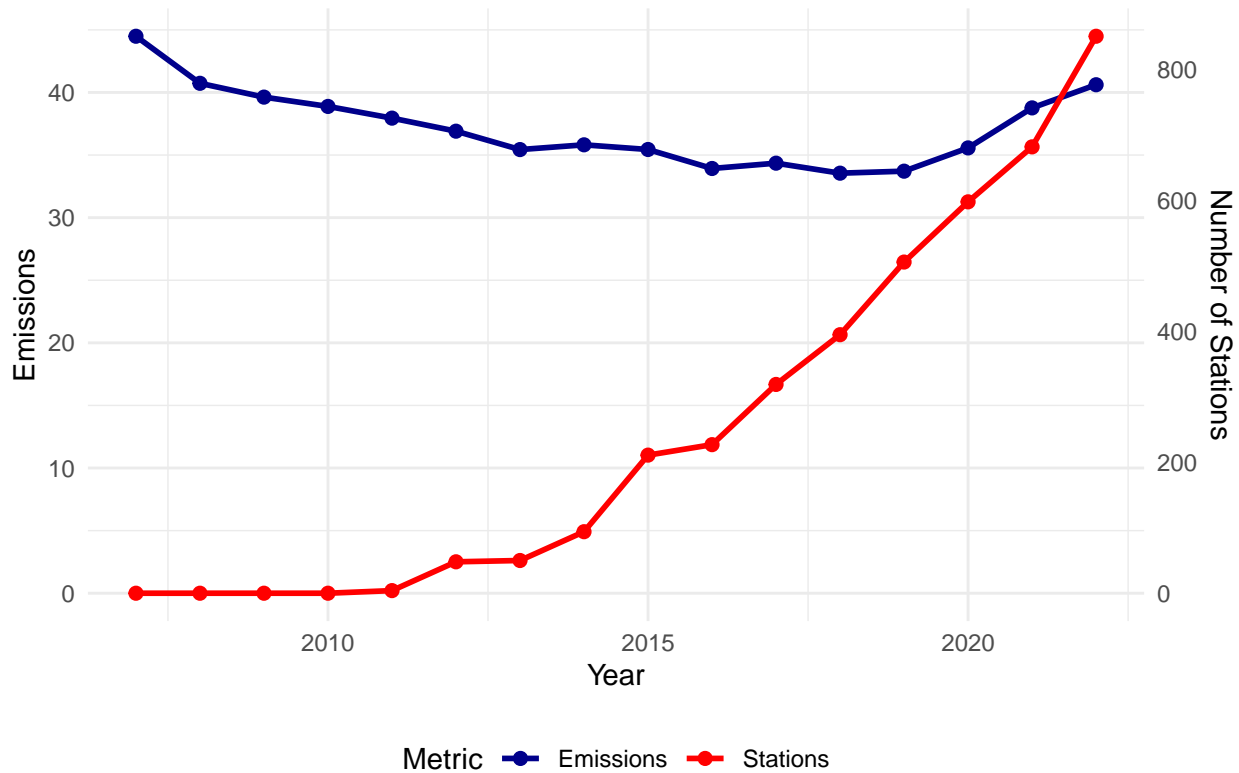
```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```r
alabama_trends
```

**Emissions and Stations in Alabama (2007−2022)**

```r
# Saving the plot
ggsave("alabama_emm_stations_dual_axis.jpg", device = "jpeg", width = 8, height = 5, dpi = 300)

## 2. Alabama Bar chat

alabama_scaled <- alabama %>%
  mutate(alab_emm_scaled = alab_emm / max(alab_emm) * max(alabama_st)) %>%  # Scale emissions
  pivot_longer(cols = c(alabama_st, alab_emm_scaled), names_to = "Variable", values_to = "Value")

# Create grouped bar chart with normalized emissions
alab_bar <- ggplot(alabama_scaled, aes(x = factor(year), y = Value, fill = Variable)) +
  geom_col(position = "dodge") +  # Group bars side-by-side
  labs(title = "Comparison of EV Stations and Emissions in Alabama",
       x = "Year",
       y = "# of Stations (Private & Public)",
       fill = "Legend") +  # Legend for clarity
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
alab_bar
```

Comparison of EV Stations and Emissions in Alabama
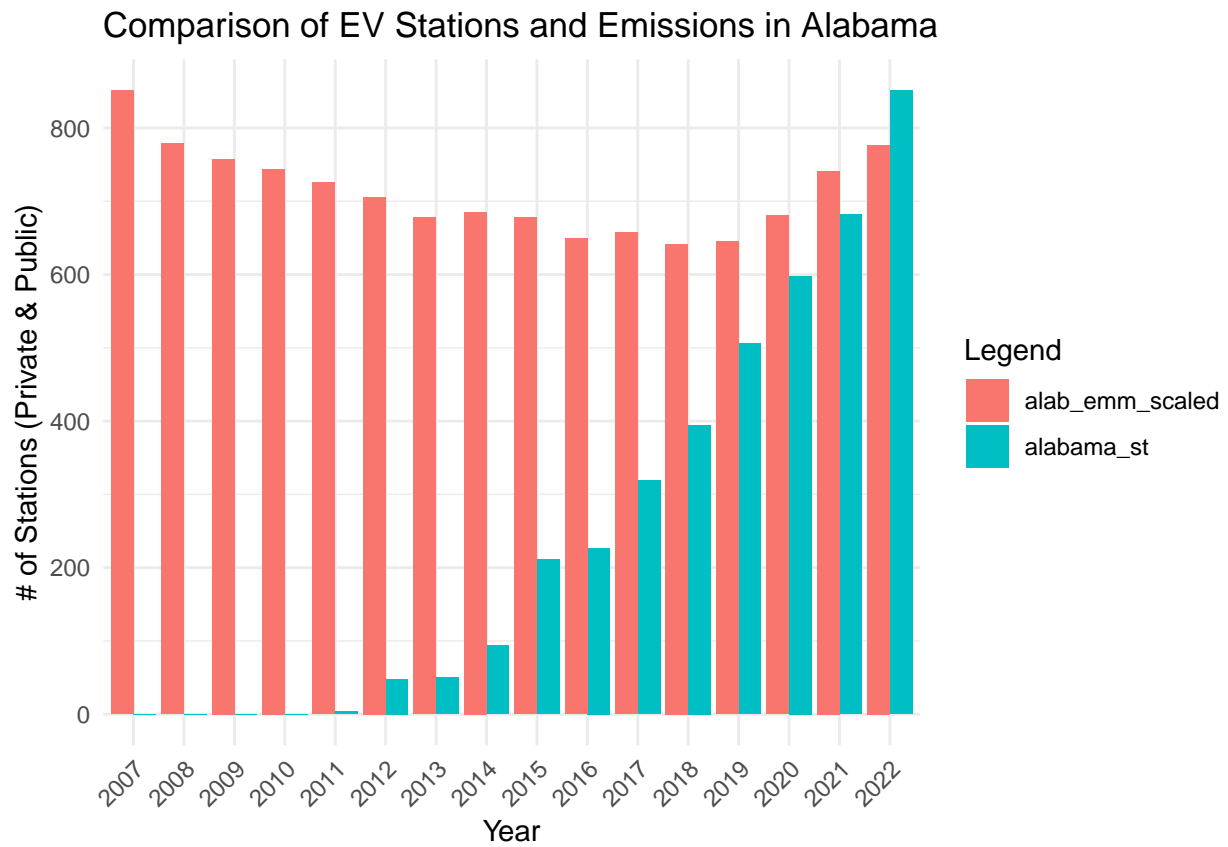

```
#saving the plot
ggsave("alabama_emm_stations.jpg", device = "jpeg", width = 8, height = 5, dpi = 300)
```