Data Exploration

Megan Hessel

Load Libraries

```
library(tidyverse)
library(ggplot2)
library(tmap)
library(sf)
library(dplyr)
library(patchwork)
library(lubridate)
library(janitor)
```

Load in Data

```
plumes <- read.csv(here::here("data", "plumes.csv")) # plume

world <- read_sf(here::here("data", "gadm_410.gdb")) %>%
  clean_names()

countries <- spData::world # World Data
```

Cleaning

```
# Make sf
plumes <- st_as_sf(plumes, coords = c('plume_longitude', 'plume_latitude'), crs = "EPSG:4326

# Change to Datetime
plumes$datetime <- as_date(plumes$datetime)
```

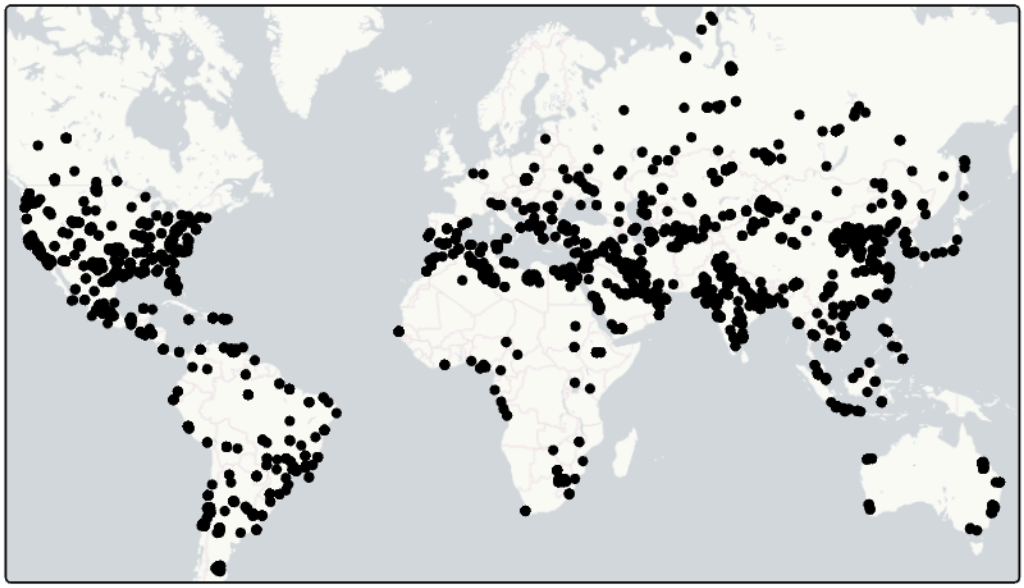
Wrangling & Subsetting

```
# Subset for CO2 and CH2's AVG and SUM emissions per sector
plume_sector_gas <- plumes %>%
  st_drop_geometry() %>%
  group_by(ipcc_sector, gas) %>%
  summarise(mean = mean(emission_auto,
                        na.rm = TRUE),
            sum = sum(emission_auto,
                        na.rm = TRUE),
            .groups = "drop")



 #..................Geospatial Wrangling + Subseting.....................

# Joining country data to plumes df
plume_countries <- st_join(countries, plumes) # join dfs

# New df: countries total emissions and population
countries_tot_emissions <- plume_countries %>%
  group_by(name_long) %>%
  summarise(total_emission = sum(emission_auto, na.rm = TRUE),
            pop = sum(pop, na.rm = TRUE),
            .groups = "drop") %>%
  mutate(weighted_total_emission = total_emission/pop)
```

1) Geospatial Mapping

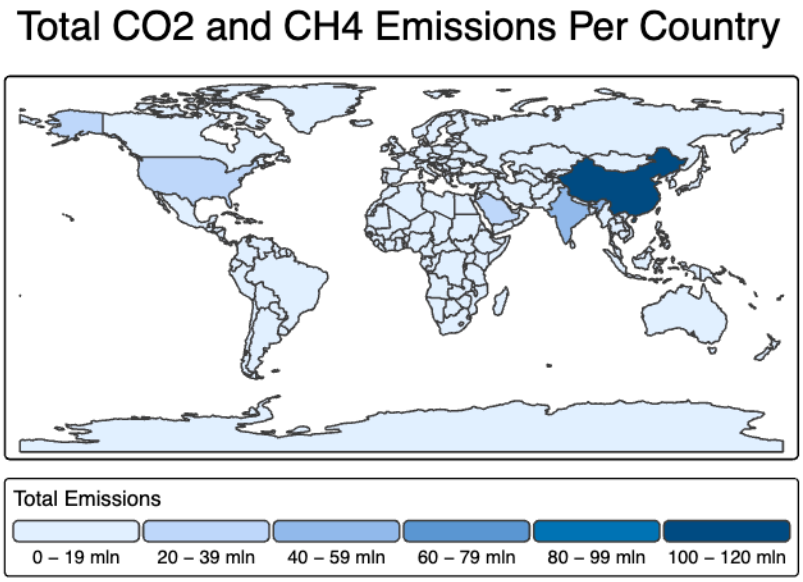
1a) Mapping Plumes

```
tm_shape(plumes) +
  tm_dots() +
  tm_basemap("CartoDB.PositronNoLabels")
```

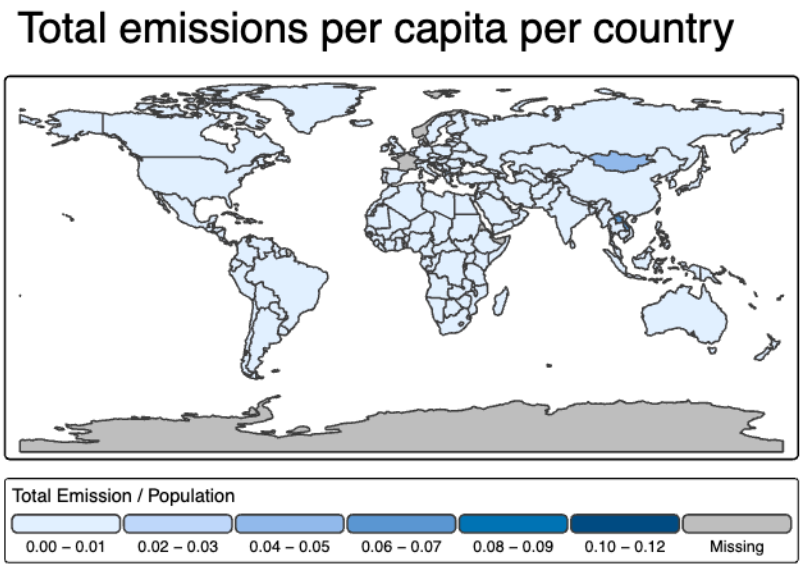


1b) Per Country - Mapping

```
# Total emissions per Country
tm_shape(countries_tot_emissions) +
  tm_polygons(fill = 'total_emission',
              fill.legend = tm_legend(title = "Total Emissions",
                                      orientation = "landscape"
                                      )) +
  tm_title(text = "Total CO2 and CH4 Emissions Per Country")
```



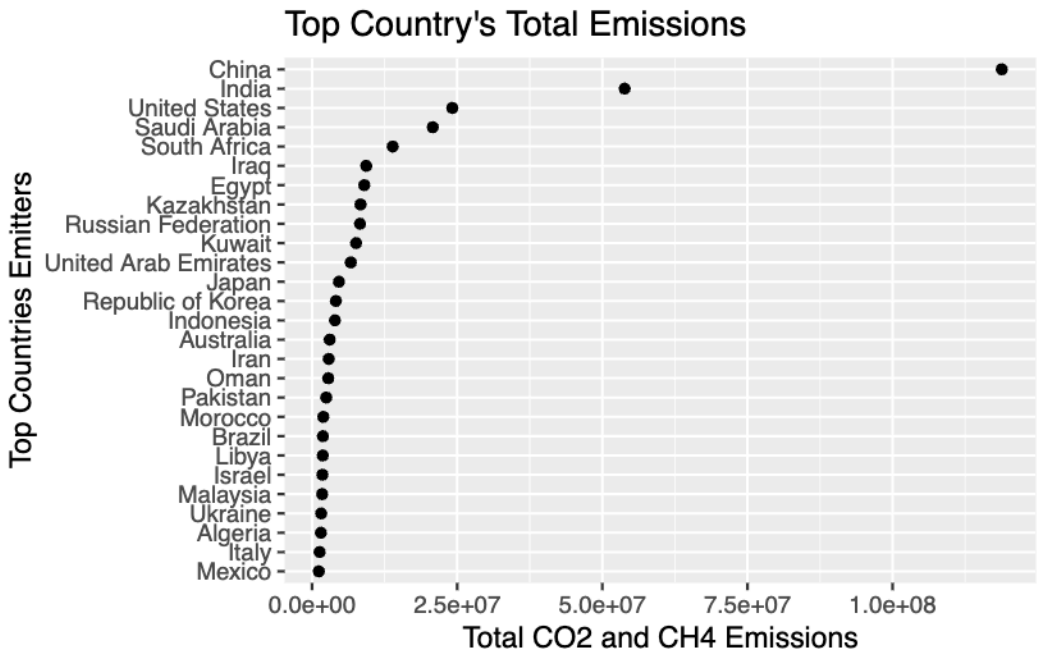

```
# weighted / Per capita Emissions per Country
tm_shape(countries_tot_emissions) +
  tm_polygons(fill = 'weighted_total_emission',
              fill.legend = tm_legend(title = "Total Emission / Population",
                                      orientation = "landscape"
                                      )) +
  tm_title(text = "Total emissions per capita per country")
```



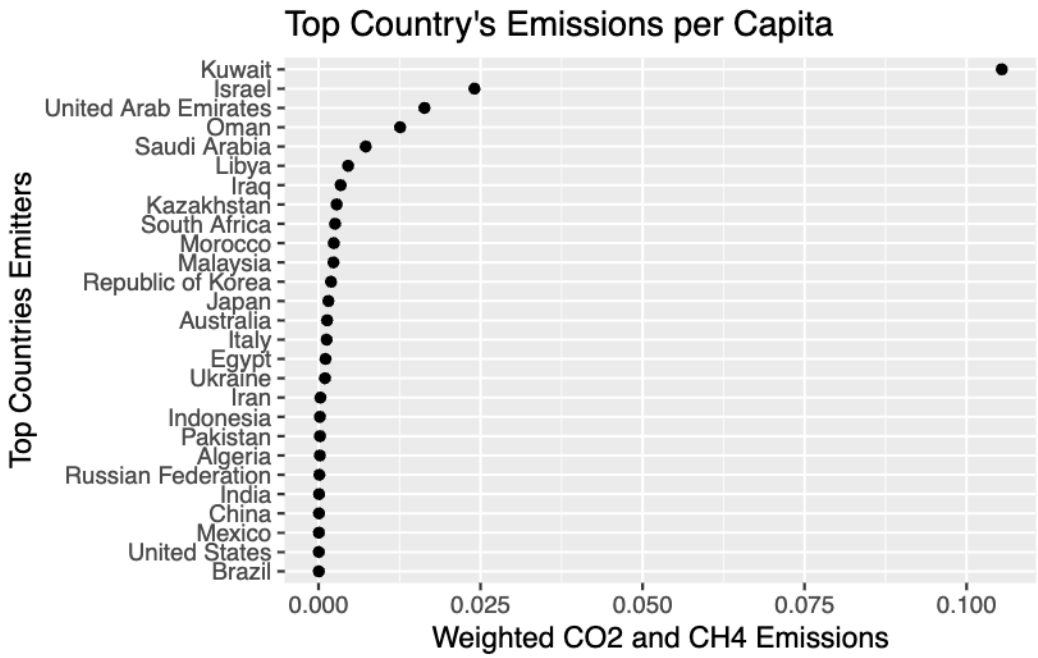
1c) Per Country - Plots

Emissions of the Top Countries

```
# GGplot of total emissions per country
countries_tot_emissions %>%
  filter(total_emission >= 1000000) %>%
  ggplot() +
  geom_point(aes(y = reorder(name_long, total_emission),
                 x = total_emission)) +
  labs(title = "Top Country's Total Emissions",
       y = "Top Countries Emitters",
       x = "Total CO2 and CH4 Emissions")
```




```
# GGplot of emissions per capita per country
countries_tot_emissions %>%
  filter(total_emission >= 1000000) %>%
  ggplot() +
  geom_point(aes(y = reorder(name_long, weighted_total_emission),
                 x = weighted_total_emission)) +
  labs(title = "Top Country's Emissions per Capita",
       y = "Top Countries Emitters",
       x = "Weighted CO2 and CH4 Emissions")
```

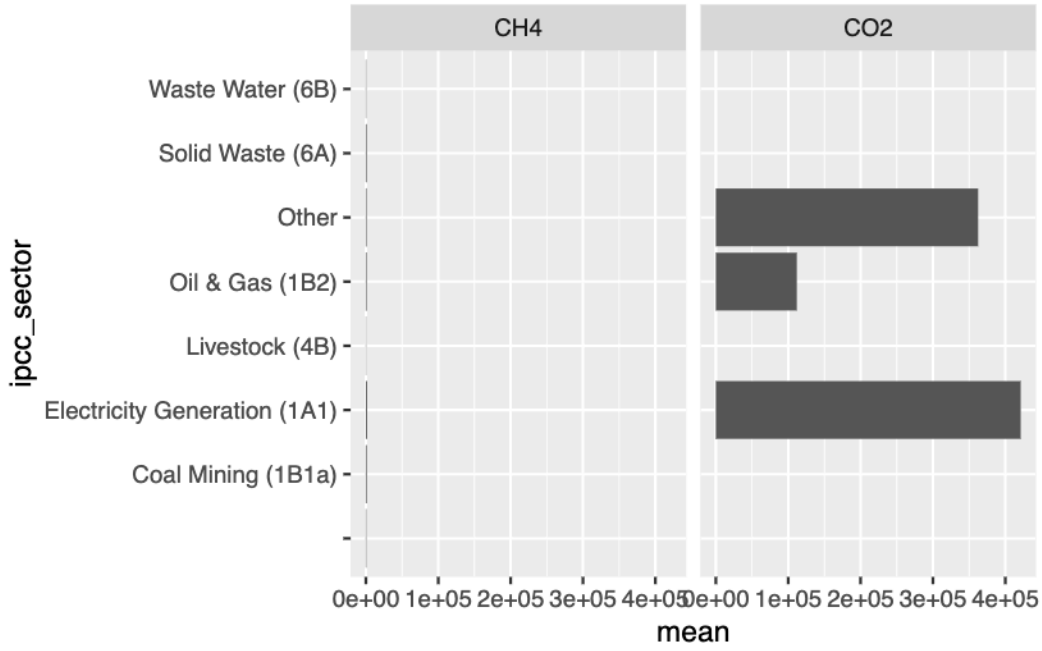


2) CO2 and CH4 emissions

2a) *AVERAGE* emission per sector

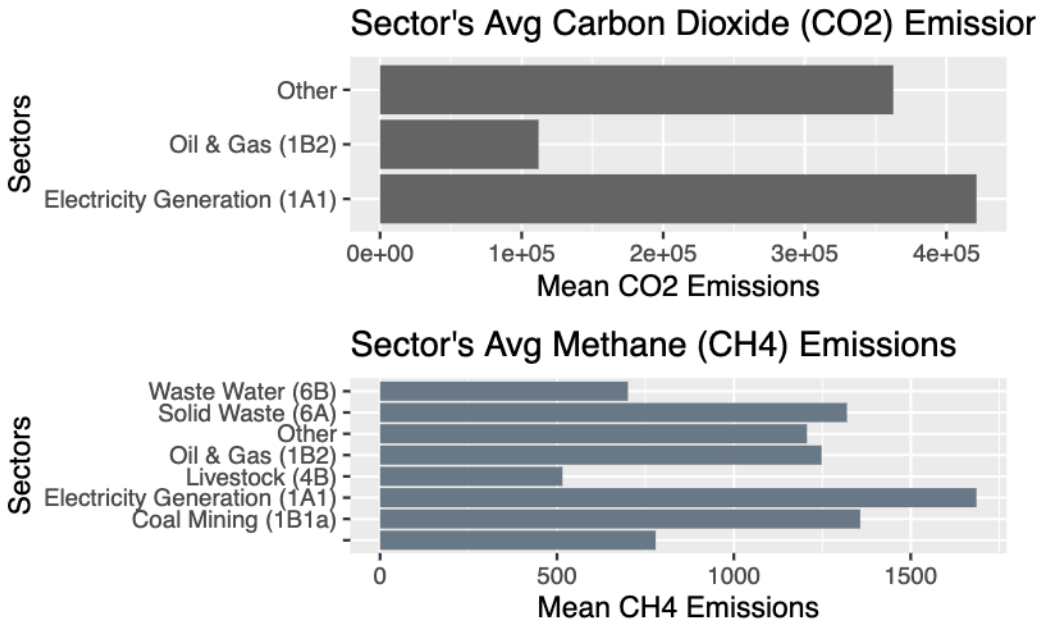
One map where the axis are the same. Another map where the axis are different.

```
# same x axis
plume_sector_gas %>%
  ggplot() +
  geom_col(aes(x = ipcc_sector, y = mean)) +
  facet_wrap(~gas) +
  coord_flip()
```




```
# Seperate x axis
avg_co2_sector <- plume_sector_gas %>%
  filter(gas == "CO2") %>%
  ggplot() +
  geom_col(aes(y = ipcc_sector,
               x = mean),
           fill = "grey40") +
  labs(title = "Sector's Avg Carbon Dioxide (CO2) Emissions",
       y = "Sectors",
       x = "Mean CO2 Emissions")


avg_ch4_sector <- plume_sector_gas %>%
  filter(gas == "CH4") %>%
  ggplot() +
  geom_col(aes(y = ipcc_sector,
               x = mean),
           fill = "slategray4") +
  labs(title = "Sector's Avg Methane (CH4) Emissions",
       y = "Sectors",
       x = "Mean CH4 Emissions")
```

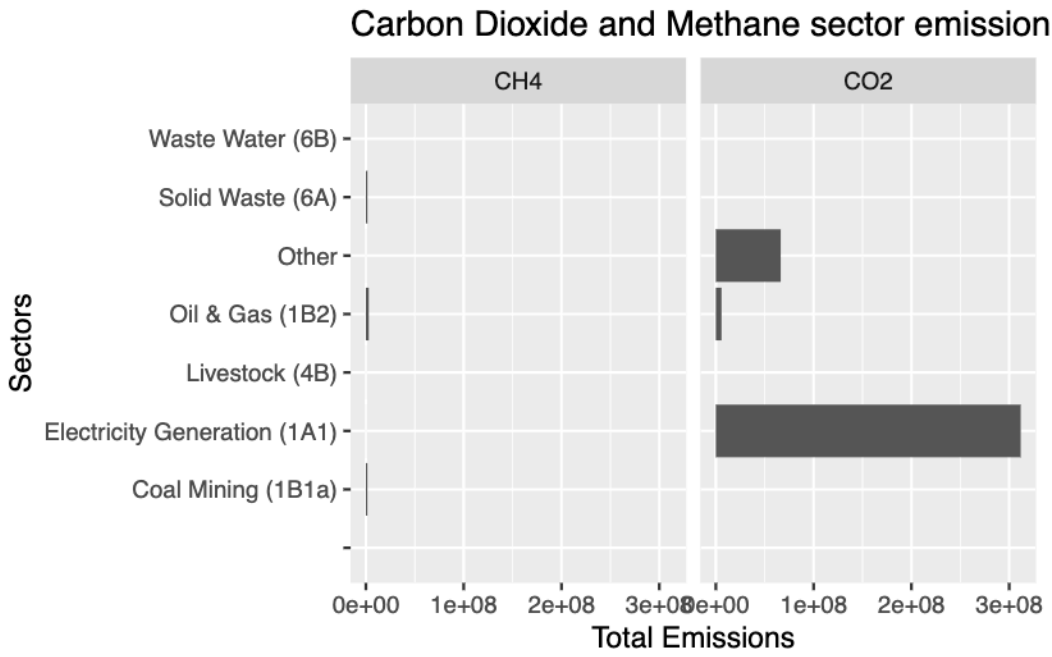
```
avg_co2_sector / avg_ch4_sector
```



2b) *TOTAL* emission per sector

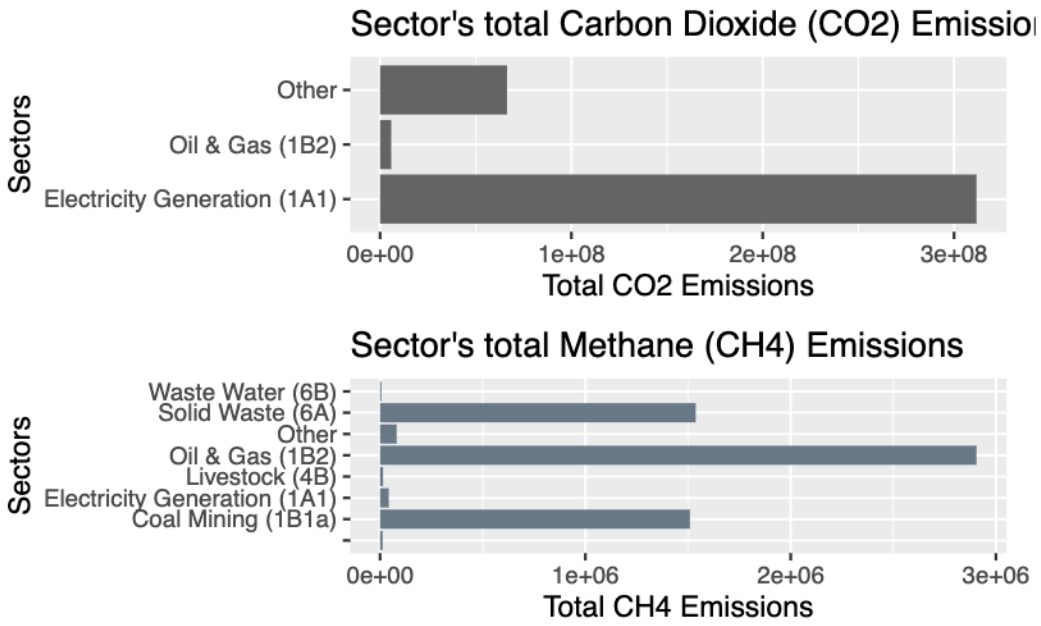
One map where the axis are the same. Another map where the axis are different.

```
# same x axis
plume_sector_gas %>%
  ggplot() +
  geom_col(aes(x = ipcc_sector, y = sum)) +
  facet_wrap(~gas) +
  coord_flip() +
  labs(title = "Carbon Dioxide and Methane sector emissions",
       x = "Sectors",
       y = "Total Emissions")
```




```
# Seperate x axis
sum_co2_sector <- plume_sector_gas %>%
  filter(gas == "CO2") %>%
  ggplot() +
  geom_col(aes(y = ipcc_sector,
               x = sum),
           fill = "grey40") +
  labs(title = "Sector's total Carbon Dioxide (CO2) Emissions",
       y = "Sectors",
       x = "Total CO2 Emissions")


sum_ch4_sector <- plume_sector_gas %>%
  filter(gas == "CH4") %>%
  ggplot() +
  geom_col(aes(y = ipcc_sector,
               x = sum),
           fill = "slategray4") +
  labs(title = "Sector's total Methane (CH4) Emissions",
       y = "Sectors",
       x = "Total CH4 Emissions")
```

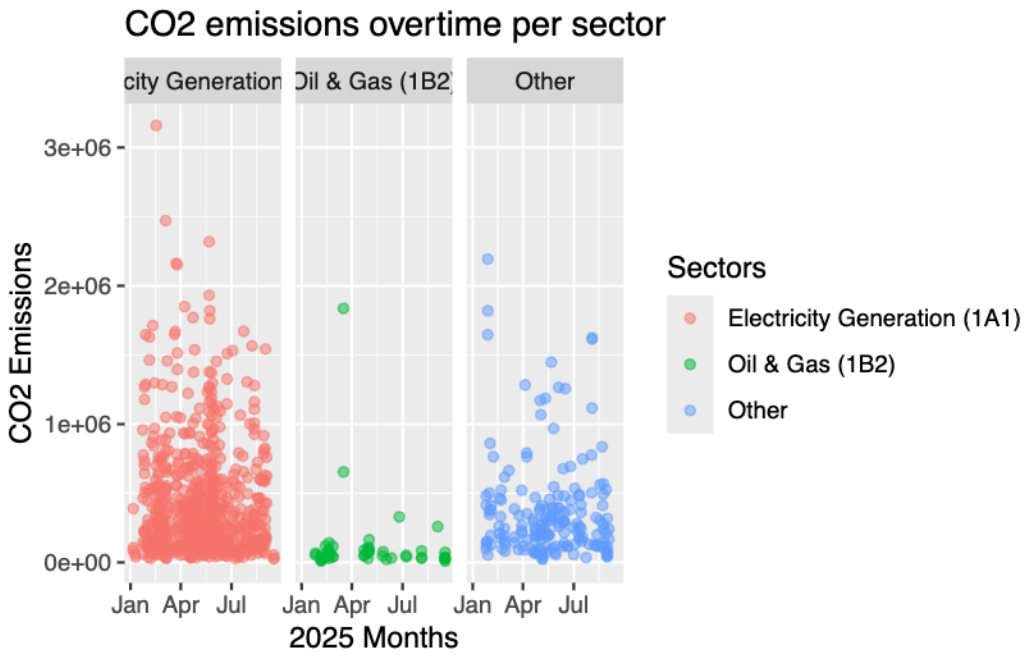
```
sum_co2_sector / sum_ch4_sector
```



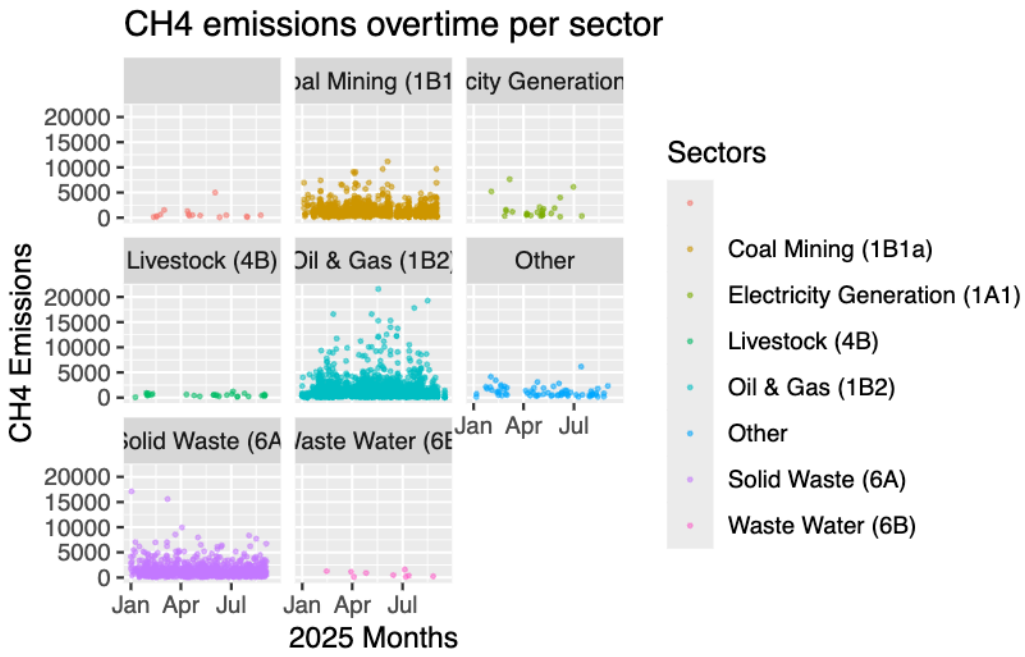
3) CH4 and CO2 Change Overtime

3a) Emissions overtime per sector

```
# sectors' CO2 emissions over time
plumes %>%
  filter(gas == "CO2") %>%
  ggplot() +
  geom_point(aes(x = datetime, y = emission_auto, col = ipcc_sector),
             alpha = 0.5) +
  labs(title = "CO2 emissions overtime per sector",
       y = "CO2 Emissions",
       x = "2025 Months",
       col = "Sectors") +
  facet_wrap(~ipcc_sector)
```



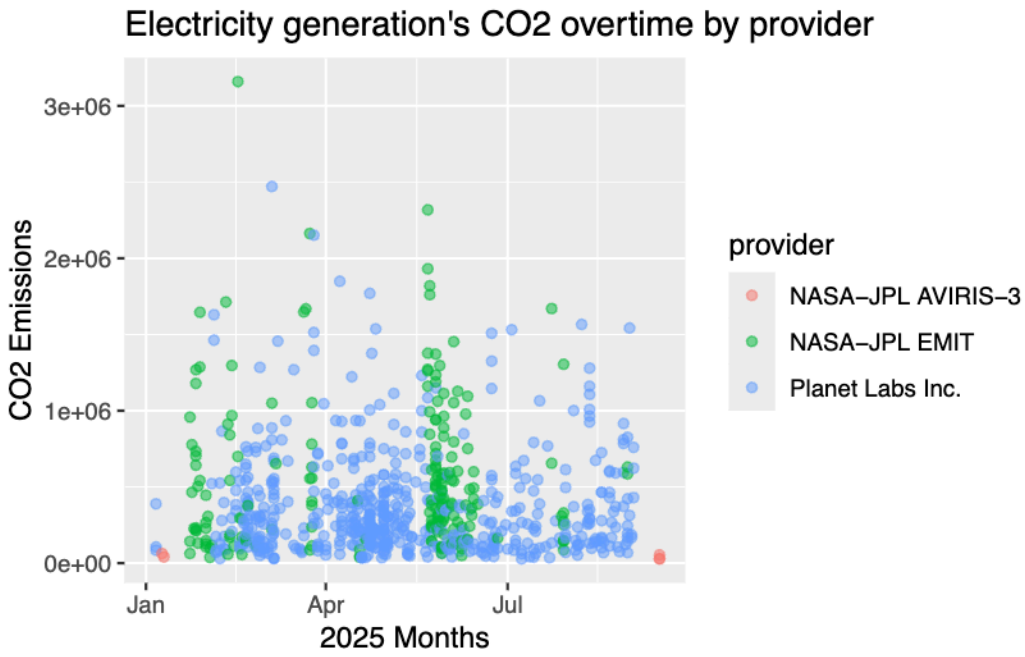

```
# sectors' CH2 emissions over time
plumes %>%
  filter(gas == "CH4") %>%
  ggplot() +
  geom_point(aes(x = datetime, y = emission_auto, col = ipcc_sector),
             alpha = 0.5,
             size = 0.5) +
  labs(title = "CH4 emissions overtime per sector",
       y = "CH4 Emissions",
       x = "2025 Months",
       col = "Sectors") +
  facet_wrap(~ipcc_sector)
```



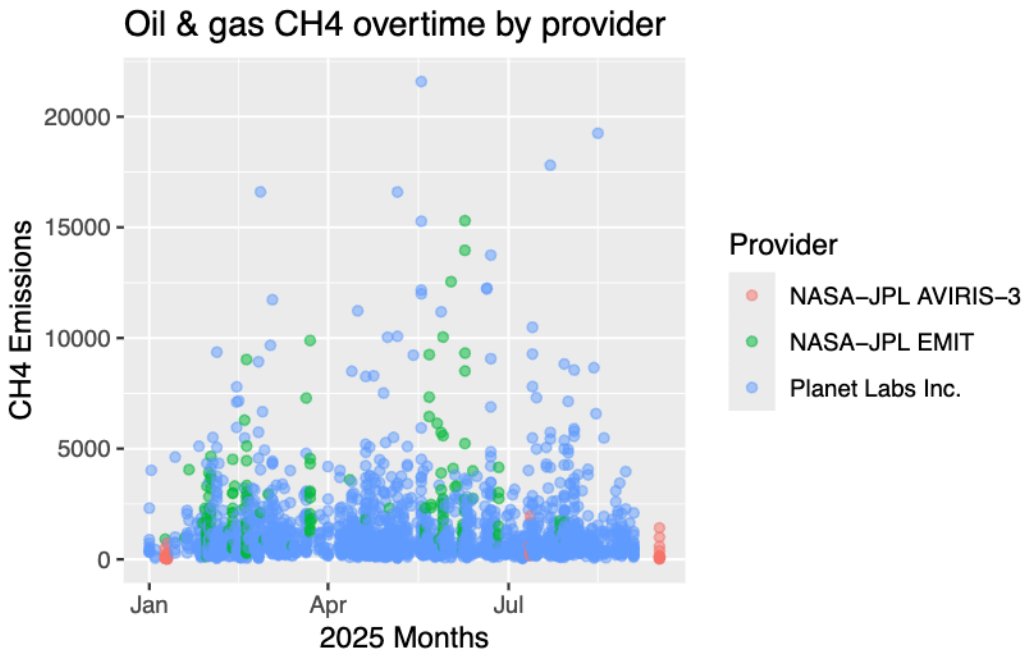
3b) Highest total emitters emissions overtime

Looking at Provider information as well

```
plumes %>%
  filter(ipcc_sector == "Electricity Generation (1A1)",
         gas == "CO2") %>%
  ggplot() +
  geom_point(aes(x = datetime, y = emission_auto, col = provider),
             alpha = 0.5) +
  labs(title = "Electricity generation's CO2 overtime by provider",
       y = "CO2 Emissions",
       x = "2025 Months")
```



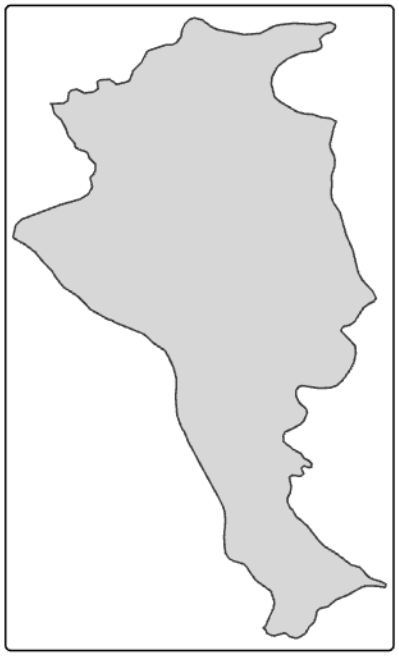

```
plumes %>%
  filter(ipcc_sector == "Oil & Gas (1B2)",
         gas == "CH4") %>%
  ggplot() +
  geom_point(aes(x = datetime, y = emission_auto, col = provider),
             alpha = 0.5) +
  labs(title = "Oil & gas CH4 overtime by provider",
       y = "CH4 Emissions",
       x = "2025 Months",
       col = "Provider")
```



4) `World` data

The `world` data is polygons of each plume recorded in 2025

```
# Just look at 1 polygon (first rows)
world[1,] %>%
  tm_shape() +
  tm_polygons()
```



Questions

1. What have you learned about your data? Have any potentially interesting patterns emerged? Point to specific visualizations that you created as you describe your findings.

- #1 Geospatial mapping: Of the 2025 plume data, China and India are the main total emitters for CO2 and CH4. But when taking population into account, Kuwait and Israel are the top emitters.
- #2 CO2 and CH4 emissions: Carbon dioxide, compared to methane, is being emitted at a much greater rate in 2025 from less sectors. The sector emitting the most carbon dioxide and methane is electricity generation. The dataset has a large portion of CO2 emissions coming from "other" which needs to be explored in the metadata.
- #3 CO2 and CH4 Change overtime: The CO2 measurements from the oil and gas sector are at weird time intervals, but the sector emits methane at a more constant rate overtime. Coal mining, oil and gas, and solid waste are regularly producing methane throughout 2025.

2. In FPM #1, you outlined some questions that you wanted to answer using these data. Have you made any strides towards answering those questions? If yes, how so? If no, what next steps do you need to take (e.g. I need to create X plot

type, I still need to track down Y data, I need to restructure existing data so that you can visualize it in Z ways, etc.)? Have any new questions emerged?

- In FPM #1, I asked: Who are the largest emitters in terms of (a) sector, (b) geographical, and (c) CO2 vs MH4 emissions? With FPM #2, I attempted to look into each of these. I made geographical plots and maps to understand who are geographically the largest emitters. Additionally, I made plots comparing CO2 and CH4 emissions between the various sectors, providers, and overtime.
- Furthermore, I asked questions about who is being affected vs consuming the benefits of the industries and which sectors could feasible reduce emissions quickest. However, I don't know how to answer those questions with the data set yet.
  - The `world` data set (which I did not explore much in this FPM) is polygons of recorded plumes overtime. That data set can help me understand the affected communities. However, since the sectors widely vary, I do not know how to examine who is benefiting from the industries in the differing countries. I will probably need to track down energy-use data.
  - Currently, I do not have enough knowledge on the industries and plume emissions to look at the data and know which sectors could feasible reduce emissions quickest/easiest.

3. What challenges do you foresee encountering with your data? These can be data wrangling and / or visualization challenges.

- I am nervous about needing more data to create an good, informative narrative with the infographic. I am also nervous about the scope of the data set. The Carbon Mapping data is global which is super insightful. However, large scoped data can be difficult to interpret and show information effectively. A lot of people can take a data set and make a ggplot. But teasing the data apart to show interesting patterns is the most educational part. Therefore, I think diving into this huge, global data set and pulling it apart to find trends and tell a story will be challenging.