Final Project (Hidden Code Chunks)

Justin Mai, Edrin Elmi, Sangmin Lee

2023-12-09

Final Project: Analyzing Green House Gas Emissions and its Impact on Air Pollution and Wildfires

1. Introduction
2. Explaining the Data

Research Question 1:

- 3. Methods
- 4. Results

Research Question 2:

- 3. Methods
- 4. Results

Research Question 3:

- 3. Methods
- 4. Results

5. Discussion: What do these results mean?
6. Summary

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Data:

```
## Rows: 1608 Columns: 23
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (11): Reporter, Sector, Subsector, NAICS Definition, Parent Company, Cit...
## dbl (12): Year, Primary NAICS Code, Total Emissions (MTCO2e), Biogenic Carbo...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 1150 Columns: 12
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (7): UNITID, AGENCY, FIRENAME, FIRENUM, STARTDATE, PERIMDATE, CAUSE
## dbl (5): OBJECTID, ACRES, YEAR, SHAPEAREA, SHAPELEN
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## New names:
## Rows: 665414 Columns: 22
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr   (4): Address, State, County, City
## dbl  (17): ...1, O3 Mean, O3 1st Max Value, O3 1st Max Hour, O3 AQI, CO Mean...
## date  (1): Date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

1. Introduction

Context:

Since the *Industrial Age*, humans globally have increased their total greenhouse gas emissions year by year. As we reached the 2000s, we realized that these emissions are directly correlated with global warming and other unbalances in the environment.

The timeline is as follows...

1st Assessment Report of the **Intergovernmental Panel on Climate Change**, 1990, believed that climate change was due largely to natural circumstances and had little to do with human influence.

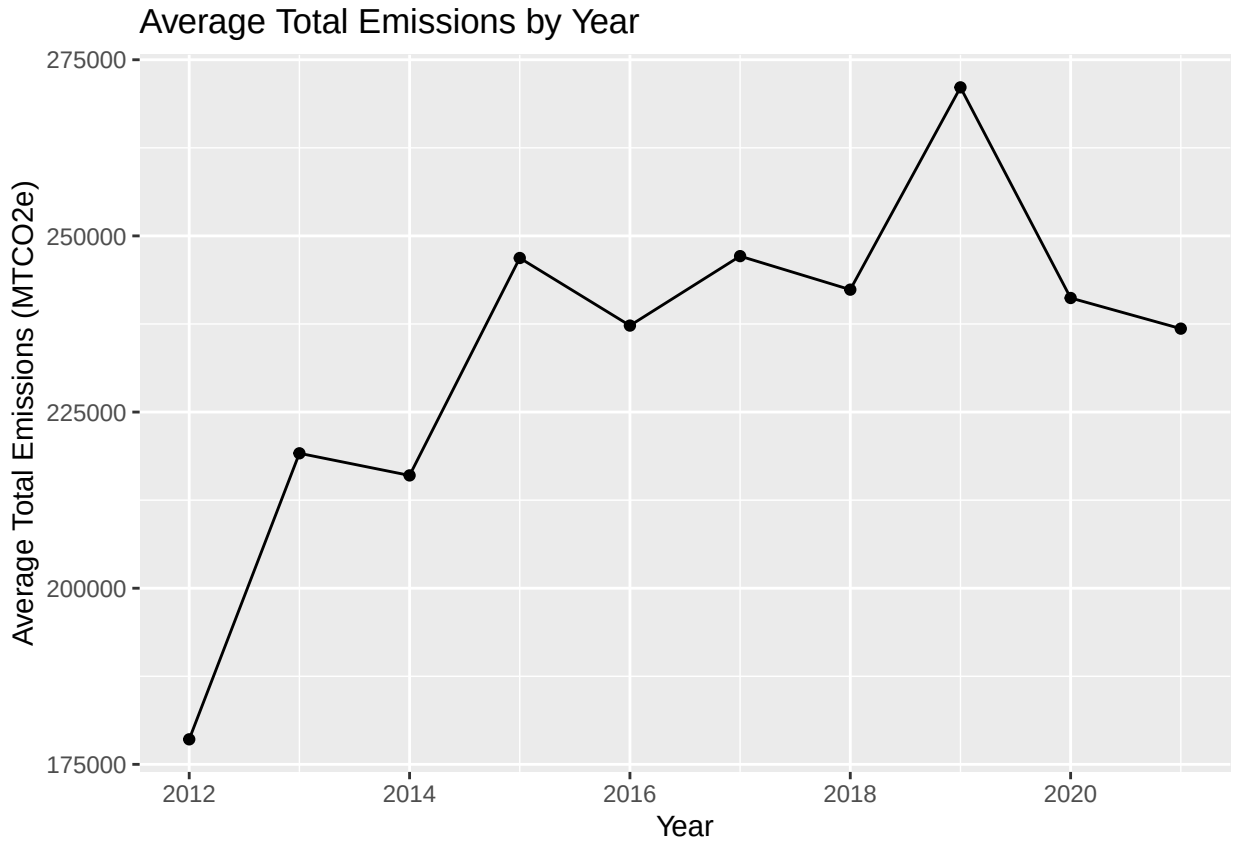
4th Assessment Report of the **IPCC**, 2007, they observed that climate change is very likely due to human actions, with evidence of an increase in average temperature and greenhouse gas concentrations released by humans.

6th Assessment Report of the **IPCC**, 2023, they have concluded that climate change is undisputedly caused by human actions and influence.

We now have glaring evidence that these abnormalities in global warming and natural effects are due to humans, so we are responsible for reversing our impact on the Earth. This is especially important to Washington residents, policymakers, and scientists because the state is largely dependent on our natural resources for food and shelter. In Washington, fisheries and shellfish harvesting is a large part of indigenous culture and residential culture. It's also creates numerous jobs for our residents. Washington is also home to many natural sights that our residents take pride in and certainly want to preserve. Mountain range sights

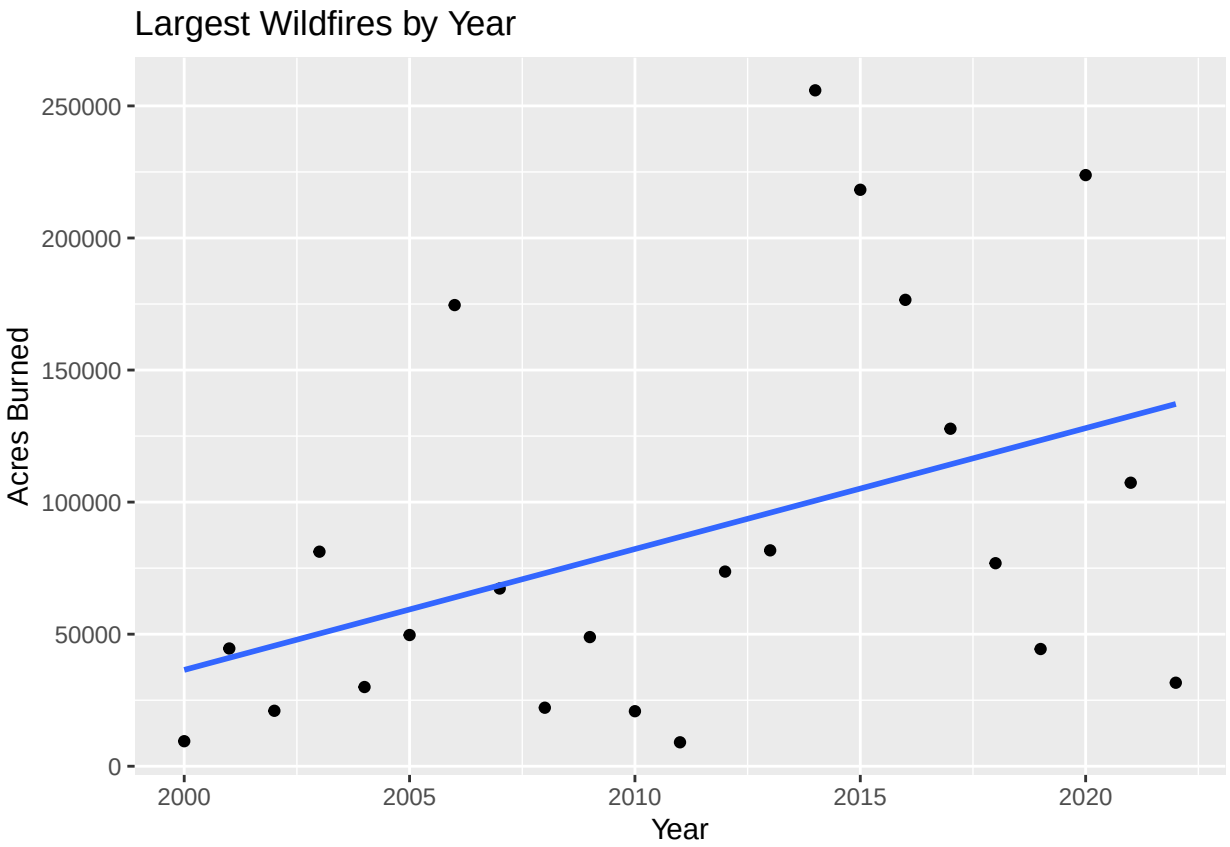
such as *Mount Rainer*, beaches, farms, and varying species of vegetation are all being negatively impacted due to our greenhouse emissions. Therefore, with greenhouse gas emissions impacting ocean temperatures and other abnormalities as much as atmospheric temperatures, pollution, and vegetation above water, its important to take steps to reverse these effects. The first step for many people is to understand the situation and visually seeing how we are impacting the Washington State.

To visualize our issue of excess greenhouse emissions, we will be *analyzing greenhouse gas emissions by different variables (sectors, corporation, location, time, etc.) and its impact on air pollution and wildfires.*



This graph demonstrates our total greenhouse gas emissions which increased from 2012 through 2019 but is moving in a decreasing trend. Even though it seems that our average total emissions is decreasing we should still be wary. One reason why is its correlation to wildfires across Washington.

```
## `geom_smooth()` using formula = 'y ~ x'
```



This graph demonstrates that the largest wildfires recorded in Washington state is within the recent decade more so then in the past. This correlates with our total emissions data, showing that there is a positive correlation between the two variables as time moves on. Since wildfires can impact human life and wildlife alike, we must take steps towards mitigating its effects.

We will also be looking at air pollution data which is important in impacting our way of life. We will be looking to identify a correlation between different types of GHG emissions and air pollution in Washington State throughout the years.

Beyond these introductory graphs demonstrating the importance of our topic of global emissions and its correlation to disasters today and the land of the future, we are presenting this project to further understand GHG effects and to inform our audience of Washington residents.

Some of the Research questions that we set out to answer are...

Q1 **What are the total greenhouse gas emissions based on *sector, location, and company?***

Q2 **Is there a correlation between greenhouse gas emissions and wildfire *amounts and sizes* in Washington State?**

Q3: **Is there a correlation between GHG emissions and air pollution?**

All of these research questions will help us narrow down the causes of these emissions and identify the locations that we must focus on. Narrowing down the scope by *sector, location, and/or company* is important to alleviating the issue and simplifying down the root causes and where we need to pay the most attention. We would also like to identify if the location of these emissions is related to larger wildfires. Overall, this report will give climate action groups and volunteers areas we need to focus on and demonstrate potential risks that excess emissions can cause.

2. Explaining the Data

```
## # A tibble: 6 x 23
##   Reporter         Year Sector Subsector ‘Primary NAICS Code‘ ‘NAICS Definition‘
##   <chr>           <dbl> <chr>  <chr>                   <dbl> <chr>
## 1 Nippon Dynawav~  2016 Pulp ~ Kraft Mi~               322121 Paper (except New~
## 2 Boulder Park G~  2017 Power~ Other Po~               221112 Fossil Fuel Elect~
## 3 Puget Sound En~  2016 Power~ Natural ~               221112 Fossil Fuel Elect~
## 4 REG Grays Harb~  2013 Petro~ Other Pe~               325199 All Other Basic O~
## 5 Darigold - Che~  2017 Food ~ Dairy Pr~               311514 Dry, Condensed, a~
## 6 Northwest Hard~  2017 Wood ~ Lumber M~               321113 Sawmills
## # i 17 more variables: ‘Parent Company‘ <chr>, City <chr>, County <chr>,
## #   Jurisdiction <chr>, ‘Total Emissions (MTCO2e)‘ <dbl>,
## #   ‘Biogenic Carbon Dioxide (MTCO2e)‘ <dbl>, ‘Carbon Dioxide (MTCO2e)‘ <dbl>,
## #   ‘Methane (MTCO2e)‘ <dbl>, ‘Nitrous Oxide (MTCO2e)‘ <dbl>,
## #   ‘HFCs (MTCO2e)‘ <dbl>, ‘PFCs (MTCO2e)‘ <dbl>,
## #   ‘Sulfur Hexafluoride (MTCO2e)‘ <dbl>, ‘Fluorinated - Other (MTCO2e)‘ <dbl>,
## #   Notes <chr>, ‘Point Size‘ <dbl>, Location <chr>, ...
```

Washington Greenhouse Gas Emission Dataset:

This dataset originates from the open government data source for Washington state from *data.wa.gov*. This data is collected by government and federal agencies through sources from their metadata inventory. Because this data is collected based of federal agencies, it is a well trusted source, especially for collecting political and environmental data such as climate change. While government data can be restricted from the public, all datasets from this source is open for public use without restriction to access. There weren't any data quality issues that we encountered because the dataset is constantly being updated with recent time data of Washington State.

Why this Data?: This dataset will act as our central dataset to compare with to both air pollution data and wildfire data. It was important to us that our central dataset for the story is high quality data, and we found *data.wa.gov* to be a trust worthy source. This dataset is abundant in information, giving us data for CO2 emissions, methane emissions, and many other greenhouse gas emissions that we will be looking at.

Dataset Description:

```
##                             Year                           Sector
##                                0                                0
##                        Subsector                 NAICS Definition
##                                0                                7
##                     Parent Company                             City
##                               71                                0
##                           County          Total Emissions (MTCO2e)
##                                0                                0
## Biogenic Carbon Dioxide (MTCO2e)           Carbon Dioxide (MTCO2e)
##                                0                                0
##                  Methane (MTCO2e)            Nitrous Oxide (MTCO2e)
##                                0                                0
```

Each row represents and institution or company from a specified location during a specified year. We are working with 1608 observations in this dataset. The relevant variables are the ones that we have selected to keep, coded using the *dplyr package*. There are a total of 79 missing values, but most are from the *Parent Company* variable which isn't concerning. There are some companies that release very little emission so some of the emission variables are noted as 0 which is a quality problem of this dataset.

```
## # A tibble: 6 x 12
##   OBJECTID UNITID AGENCY      FIRENAME FIRENUM STARTDATE PERIMDATE   ACRES  YEAR
##      <dbl> <chr>  <chr>       <chr>    <chr>   <chr>     <chr>       <dbl> <dbl>
## 1     1292 WAPSC  " "         BAKER R~ 000198  2022/10/~ 2022/10/~   1      2022
## 2      301 WAOWF  "OKAN-WEN ~ OLD BLE~ 597     2012/09/~ 2012/09/~  22.4    2012
## 3      305 WAOWF  "OKAN-WEN ~ TIMBER ~ 644     2012/09/~ 2012/09/~ 153.     2012
## 4      990 WANES  "WA DNR - ~ NORTH C~ 001428  2021/04/~ 2022/01/~   0.181  2021
## 5      389 WAOWF  "OKAN-WEN ~ SOUTH C~ 621     2014/08/~ 2014/08/~ 888.     2014
## 6      605 WACOF  "COLVILLE ~ KELLY M~ 1894    2018/08/~ 2018/08/~  38.0    2018
## # i 3 more variables: CAUSE <chr>, SHAPEAREA <dbl>, SHAPELEN <dbl>
```

Washington Wildfire Data:

This dataset originates from the *Washington Geospatial Open Data Portal*, through the website *geo.wa.gov*. This data is regularly updated and corrected by different Washington state departments. Because the Washington State Departments collected this data, it is coming from a reliable source. The Wildfire data is made for public viewing and public use without any restrictions stated in the license. This dataset consists of records of large Wildfires from 1973 to 2022 data, which is the population. We narrow down the sample to 2000-2022 data to only analyze patterns from recent times. The only data quality issue that we encountered is that many of the values from the *CAUSES* variable are *unknown*, however this isn't an issue because we are looking at size and location of the wildfires in comparison to GHG emission patterns.

Why this dataset?: We are using this dataset because it is published by a trusted source that provides us data on wildfire times and size.

Dataset Description:

```
##    UNITID  FIRENAME   FIRENUM STARTDATE PERIMDATE     ACRES      YEAR
##         0         0         0         0         1         0         0
```

Each row represents a wildfire in Washington State. There are 1150 observations in this dataset. All of the variables from the original dataset are going to be used for our analysis except for *object ID and agency* which isn't needed for merging with other datsets. There aren't many missing values and all of the values seem relivant and reasonable.

```
## # A tibble: 6 x 22
##     ...1 Date       Address      State County City  `O3 Mean` `O3 1st Max Value`
##    <dbl> <date>     <chr>        <chr> <chr>  <chr>     <dbl>              <dbl>
## 1 138509 2005-05-11 4103 BEACON~ Wash~ King   Seat~   0.00971              0.018
## 2 138564 2005-09-09 4103 BEACON~ Wash~ King   Seat~   0.0139               0.018
## 3 649289 2022-08-16 4103 BEACON~ Wash~ King   Seat~   0.0232               0.032
## 4  93552 2003-07-10 4103 BEACON~ Wash~ King   Seat~   0.0292               0.053
## 5 588874 2020-10-20 4103 BEACON~ Wash~ King   Seat~   0.0143               0.022
## 6  70995 2002-05-09 4103 BEACON~ Wash~ King   Seat~   0.0221               0.028
## # i 14 more variables: `O3 1st Max Hour` <dbl>, `O3 AQI` <dbl>,
## #   `CO Mean` <dbl>, `CO 1st Max Value` <dbl>, `CO 1st Max Hour` <dbl>,
## #   `CO AQI` <dbl>, `SO2 Mean` <dbl>, `SO2 1st Max Value` <dbl>,
## #   `SO2 1st Max Hour` <dbl>, `SO2 AQI` <dbl>, `NO2 Mean` <dbl>,
## #   `NO2 1st Max Value` <dbl>, `NO2 1st Max Hour` <dbl>, `NO2 AQI` <dbl>
```

Washington Air Pollution Data:

We got this data off *Kaggle.com* but it is stated to be sourced by the U.S. Environmental Protection Agency (EPA). Since this is a Kaggle dataset and is published as a public dataset, it is open for all to use without any restrictions. This dataset was originally data about air pollution in the United States from 2000 to 2023,

however, we filtered the data to only be on Washington state data. So, the population of this data is the United States but we are narrowing down the sample to be Washington state data. A quality issue for this data is that all of the observations are from the Greater Seattle Area instead of representing all of Seattle. However, because some of our highest emissions reside around the Seattle area, we are able to compare this data with GHG data to answer our question of correlation between GHG emissions and air pollution.

Why this dataset?: Because this data was sourced and used by the U.S. Environmental Protection Agency, we believe this data to be reliable and accurate to what we require to answer our research question. We also found variables such as *average ozone level, AQI, and carbon levels to be important in measuring air pollution.

Dataset Description:

```
##     Date    State   County     City  O3 Mean   O3 AQI  CO Mean   CO AQI
##        0        0        0        0        0        0        0        0
## SO2 Mean  SO2 AQI NO2 Mean  NO2 AQI
##        0        0        0        0
```

Each row represents data for a specific location during a specific date. We narrowed down our sample data to 3274 observations. The relevant variables would be time, location, AQI, O3, SO2, NO2 and CO data. The logistical data will be used to view correlation and the location and time data will be used to merge and clean with the other datasets. There aren't any missing values and all of the values are reasonable (no quality issues).

Research Question 1: What are the total greenhouse gas emissions based on *sector, location, and company?*:

3. Q1 Methods

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

Which variables did you include, why?

- The variables I included were Sector, Parent Company, City, and Total_Emissions because the research question was based on only those variables.

Sector - Types of Company

Parent Company - Company overseeing smaller company

City

Total Emissions - Sum of all GHG emissions (CO, NO, S, etc.), in MTCO2e (metric tons of CO2 equivalent)

How did you process your data, why?

- I used group_by and summarize functions from the dplyr package to organize the information. I grouped the data by different categories like sectors, cities, and parent companies, and then added up the total greenhouse gas emissions for each group. Because it helps us see the big picture of how much greenhouse gases are being released in various areas and industries.

How did you select your sample?

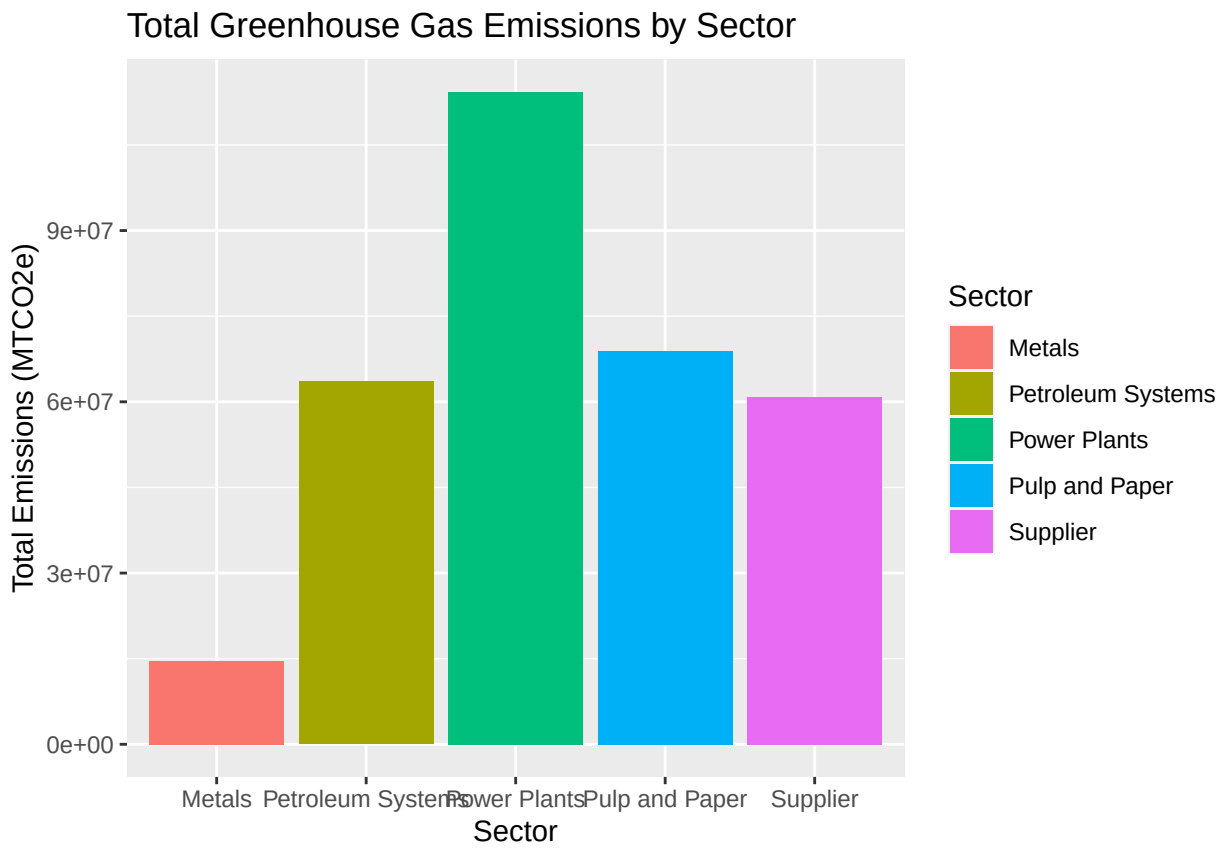
- How we selected our sample was we looked for the most credibile and most resourceful dataset on greenhouse gas emissions and this dataset was the best option which is why we selected it.

Any technical decisions you did, such as either to remove missings, or replace missing values with another one?

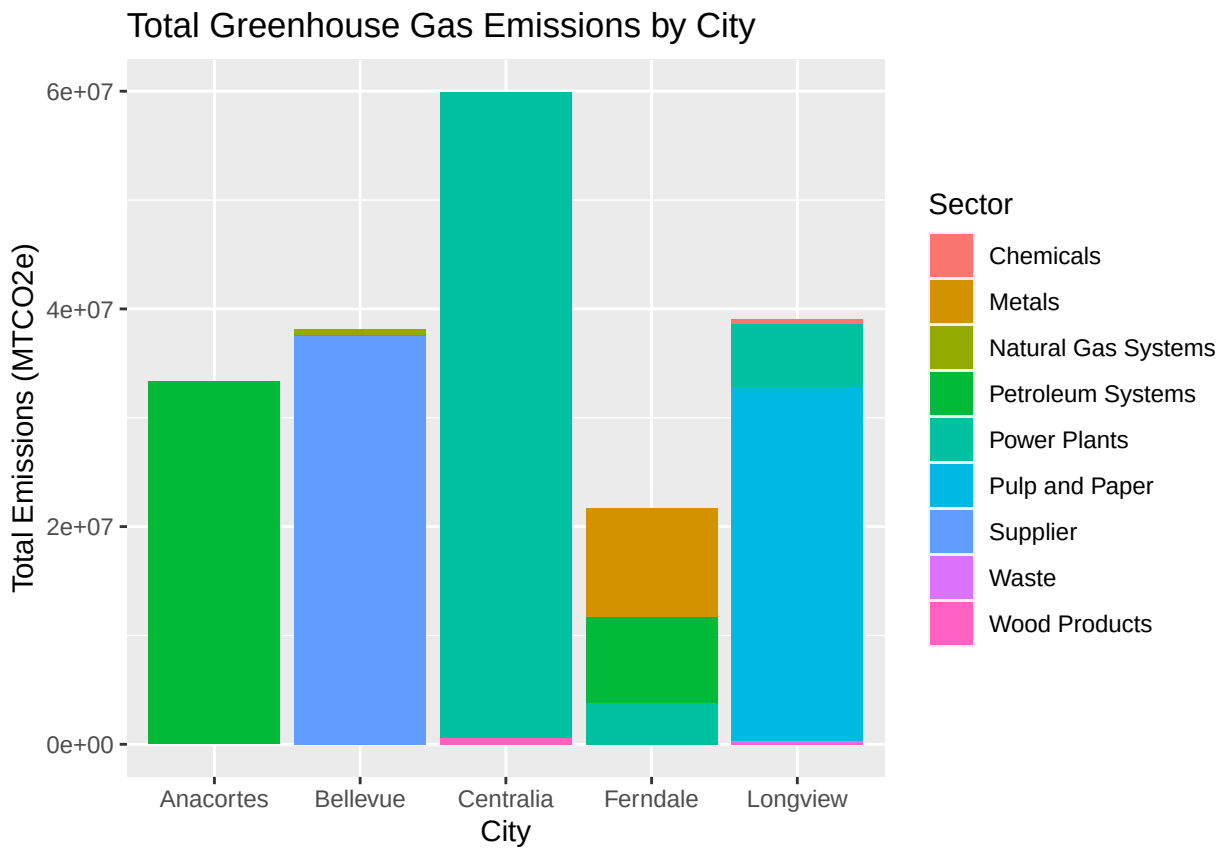
- There were no technical decision I had to make, I didn't need to remove or replace missing values
- One missing value was for parent companies but we opted to keep the data to show since it is a significant part of the total emissions

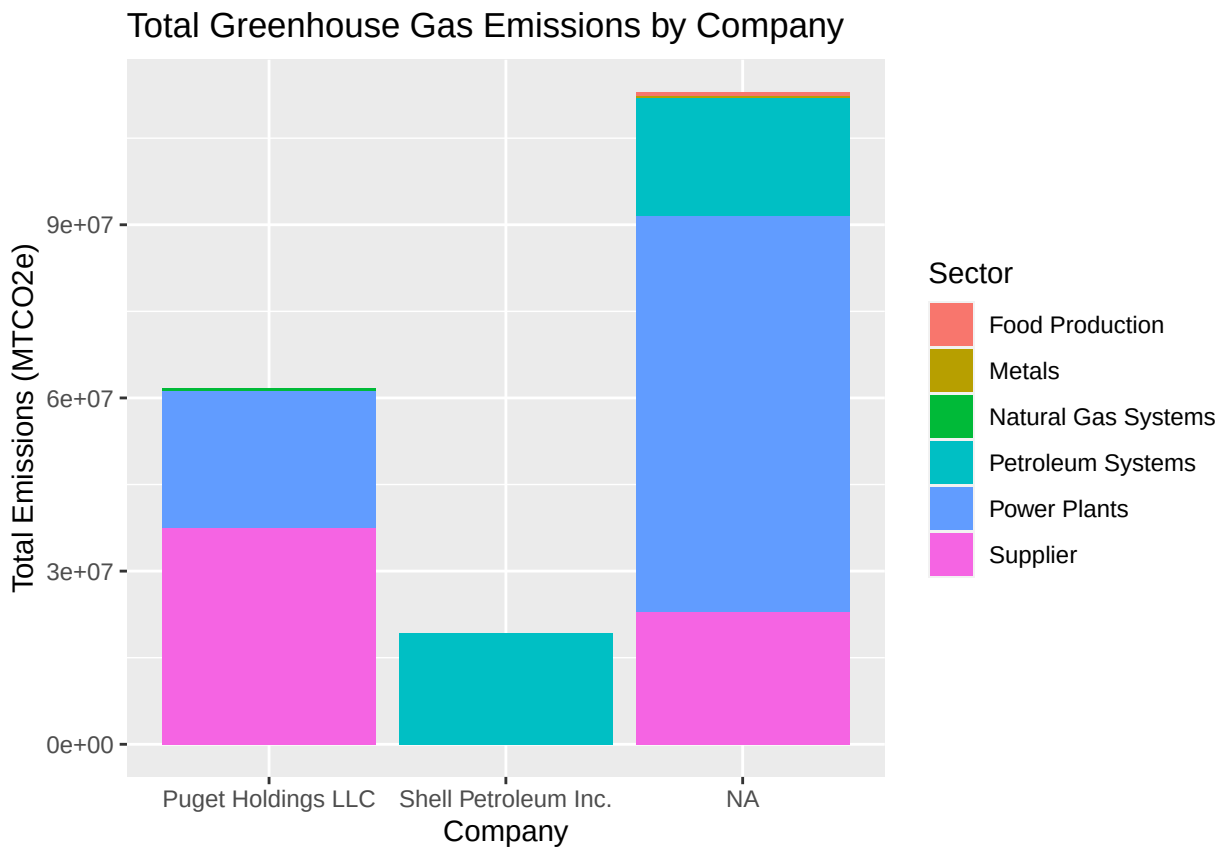
4. Q1 Results

```
## # A tibble: 57 x 4
## # Groups:   Sector, City [51]
##    Sector            City           `Parent Company`            Total_Emissions
##    <chr>             <chr>          <chr>                                 <dbl>
##  1 Metals            Ferndale       Alcoa Corp.                         10053702
##  2 Metals            Kalama         <NA>                                  233702
##  3 Metals            Malaga         Alcoa Corp.                          1309513
##  4 Metals            Moses Lake     REC Silicon                           688200
##  5 Metals            Seattle        Nucor Corporation                    1045902
##  6 Metals            Spokane Valley Kaiser Aluminum Corp.                1176731
##  7 Metals            Tukwila        Jorgensen Forge Corp.                  73387
##  8 Petroleum Systems Anacortes      Air Liquide USA Inc.                  640264
##  9 Petroleum Systems Anacortes      Linde Gas North America LLC           509055
## 10 Petroleum Systems Anacortes      Marathon Petroleum Corp             13001913
## # i 47 more rows
```




```
## # A tibble: 20 x 4
## # Groups:   Sector, City [14]
##    Sector              City      `Parent Company`                Total_Emissions
##    <chr>               <chr>     <chr>                                     <dbl>
##  1 Chemicals           Longview  Solvay Chemicals Inc.                    481547
##  2 Metals              Ferndale  Alcoa Corp.                            10053702
##  3 Natural Gas Systems Bellevue  Puget Holdings LLC                      602466
##  4 Petroleum Systems   Anacortes Air Liquide USA Inc.                     640264
##  5 Petroleum Systems   Anacortes Linde Gas North America LLC              509055
##  6 Petroleum Systems   Anacortes Marathon Petroleum Corp                13001913
##  7 Petroleum Systems   Anacortes Shell Petroleum Inc.                   19235767
##  8 Petroleum Systems   Ferndale  Phillips 66 Company                     7880228
##  9 Power Plants        Bellevue  Puget Holdings LLC                        21055
## 10 Power Plants        Centralia <NA>                                   59338269
## 11 Power Plants        Ferndale  Puget Holdings LLC                      3781208
## 12 Power Plants        Longview  Puget Holdings LLC                      5784628
## 13 Pulp and Paper      Longview  Kapstone Paper & Packaging Cor~        16317134
## 14 Pulp and Paper      Longview  NP Paper Investor Holdings LLC           394474
## 15 Pulp and Paper      Longview  Weyerhaeuser Co.                        15779251
## 16 Supplier            Bellevue  Puget Holdings LLC                      37504945
## 17 Waste               Longview  Cowlitz County                            167364
## 18 Wood Products       Centralia Littlejohn & Co. LLC                      113780
## 19 Wood Products       Centralia Sierra Pacific Industries                 445694
## 20 Wood Products       Longview  American Industrial Partners (~           136581
```


Total Greenhouse Gas Emissions by City
Total Emissions (MTCO2e)
6e+07
4e+07
2e+07
0e+00
Anacortes
Bellevue
Centralia
Ferndale
Longview
City
Sector
Chemicals
Metals
Natural Gas Systems
Petroleum Systems
Power Plants
Pulp and Paper
Supplier
Waste
Wood Products



The data tells me about the total greenhouse gas emissions by sector, location, and companies and what it told me was that power plants emit the most green house gases out of all the sectors, and that the least is livestock, the city that emits the most greenhouse gases is Centralia and the least is Plymouth, the Company that emits the most green house gases is Puget holdings and the least is Kimberly Clark Corporation.

Research question 2. Is there a correlation between greenhouse gas emissions and wildfire *amounts and sizes* in Washington State?

3. Q2 Methods

Variables Include in final dataset (**merged_data**)

YEAR

avg_size - Average wildfire size (in acres) by year

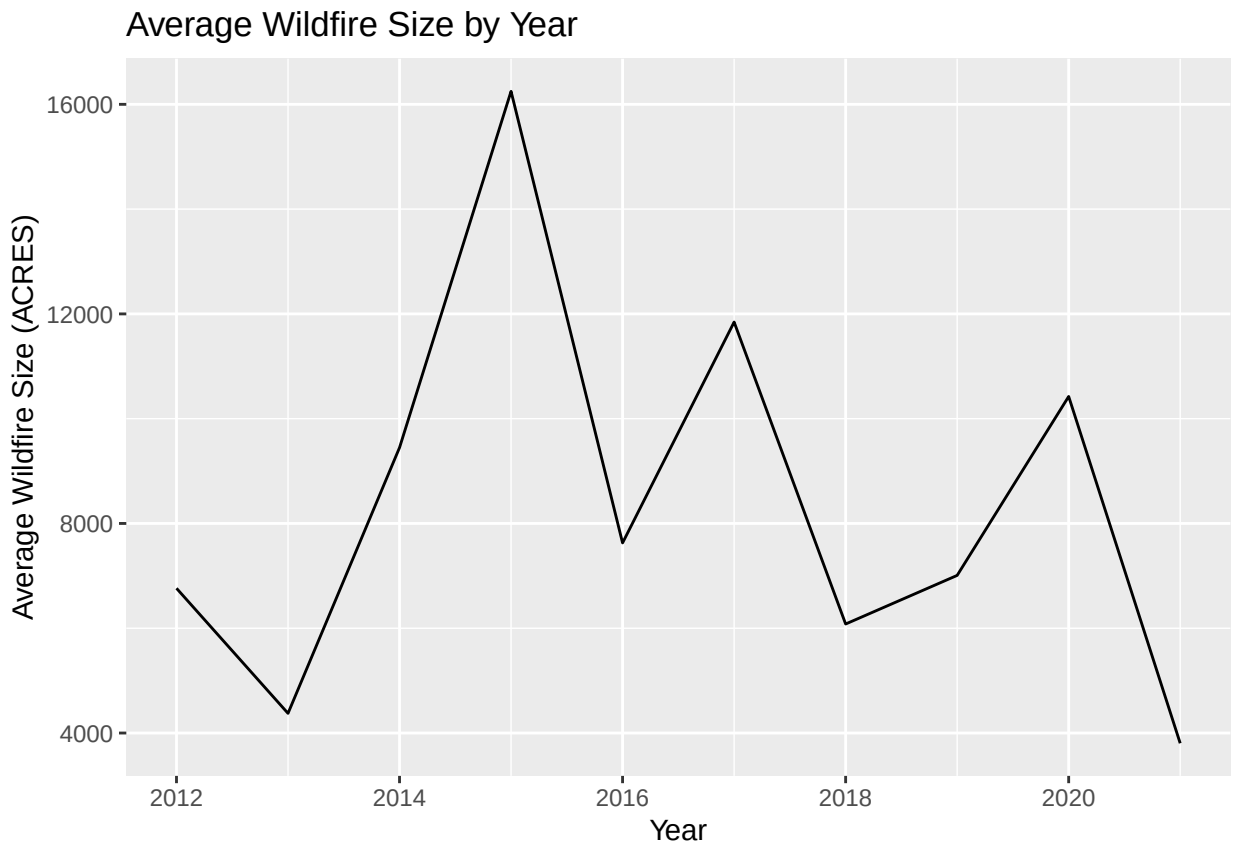
Wildfires_count - Number of wildfires for that year

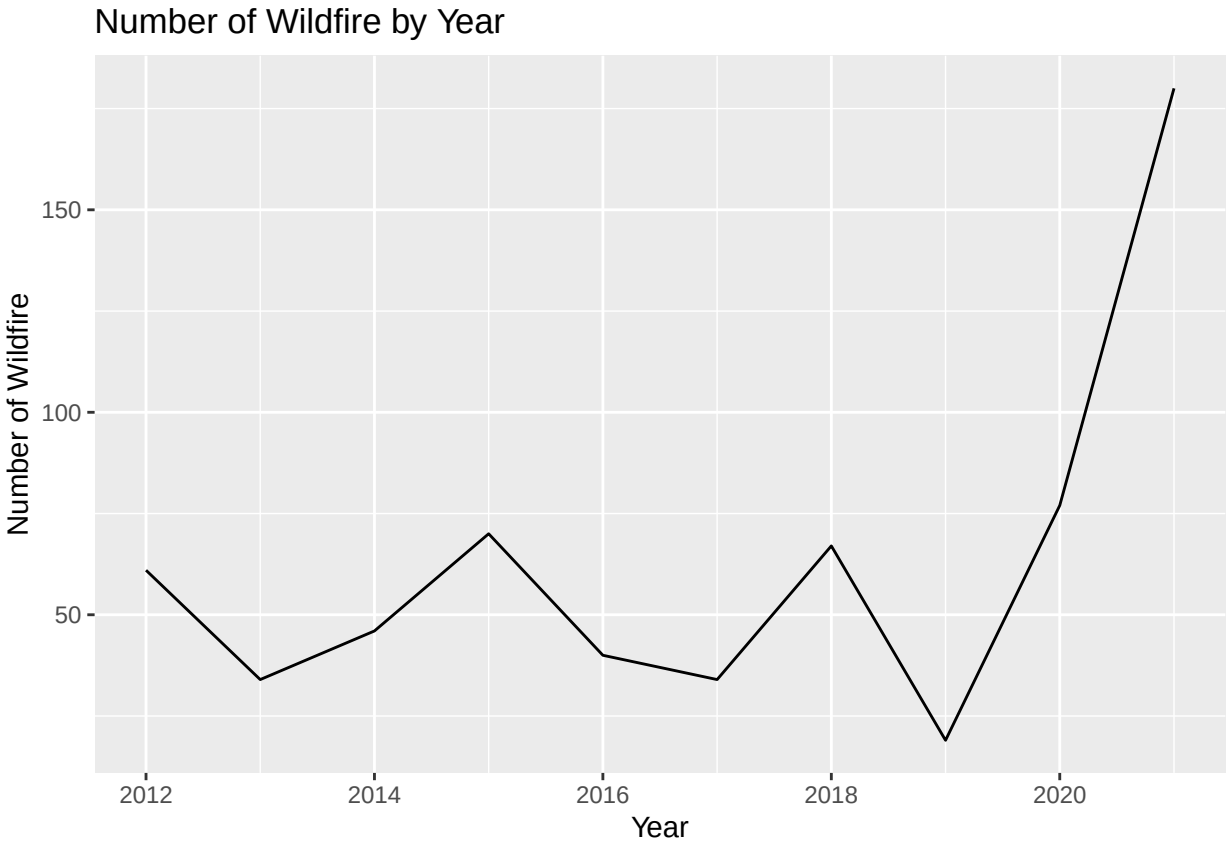
Total Emissions - Sum of all GHG emissions (CO, NO, S, etc.), in MTCO2e (metric tons of CO2 equivalent)

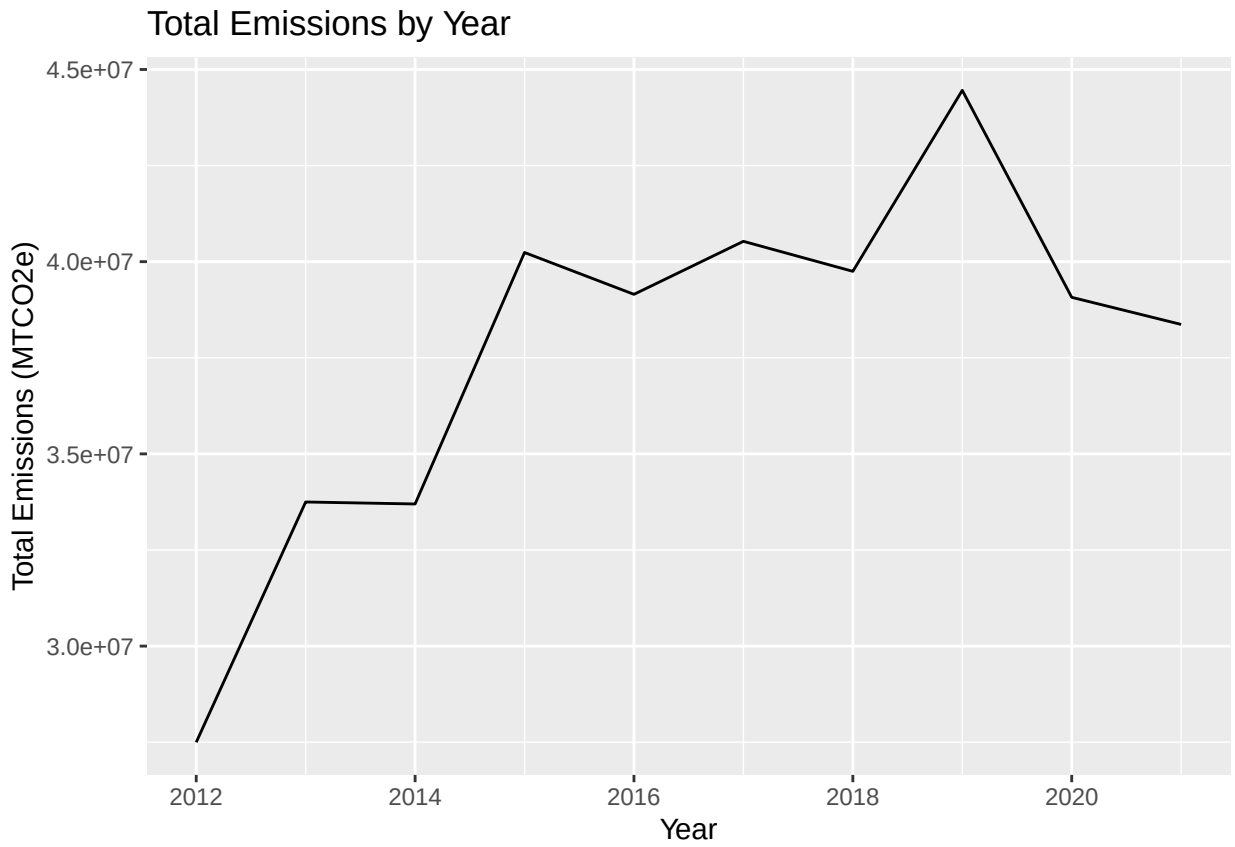
- We created **avg_size** and **Wildfires_count** to help us compare the results to **Total Emissions**, analyzing if there's a correlation or not.
- We processed the data by first using group_by() and summarize() functions to get the **avg_size**, **Wildfires_count**, and **Total Emissions**. We then joined the wildfire and GHG emissions data together using right_join().

- We selected our sample by grouping our data by year.
- We didn't make any technical decisions to replace NA's
- We used line graphs to represent the pattern of each graph, giving us a clear visual of increasing and decrasing patterns.

4. Q2 Results




Number of Wildfire by Year
Number of Wildfire
150
100
50
2012
2014
2016
2018
2020
Year



The average wildfires size by year had rapidly increased at 2015, but it has steadily decreased since 2015.

The number of wildfires had remained about 50, but during the recent 5 years, the number of wildfires has rapidly increased.

The total emissions had increased since 2012 to 2019, and has decreased since 2019 to the presemt.

Correlation between the size of wildfires and the total gas emissions. By comparing the first and the third graph, we can say that the size of wildfires and the total gas emissions are related. The rise in the number of significant wildfires coincided with a subsequent increase in the total gas emissions.

Correlation between the number of wildfires and the total gas emissions. Examining the second and third graphs, they exhibit contrasting characteristics.While the number of wildfires is rapidly increasing from 2019, the total gas emissions has decreased from 2019.

Analyzing the graph above, the correlation between wildfire size and total gas emissions appears to follow a cause-and-effect pattern, whereas the correlation between the number of wildfires and total gas emissions appears to exhibit an opposite trend.

Research Question 3: Is there a correlation between GHG emissions and air pollution?

3. Q3 Methods

To answer my question I used *dplyr* functions to merge and reshape my data. I used the **Washington GHG Emissions** data and the **United States Air Pollution** data. An issue that I would like to note is that after I filtered the Air Pollution data to Washington state only, I found that the dataset only considered Seattle

air pollution. Because there weren't any better dataset available, I made due by filtering my Washington GHG Data to the *King County* area to highlight the Greater Seattle Area. This will allow me to compare the two data since all the sectors are now around the same area and could possible impact Air pollution in Seattle.

I first selected the logistical data that I need from the GHG data and got the averages for all the variables by year to prepare for merging and reshaping.

```
## Warning: There were 50 warnings in `summarise()`.
## The first warning was:
## i In argument: `Sector = (function (x, ...) ...`.
## i In group 1: `Year = "2012"`.
## Caused by warning in `mean.default()`:
## ! argument is not numeric or logical: returning NA
## i Run `dplyr::last_dplyr_warnings()` to see the 49 remaining warnings.
```

```
## # A tibble: 6 x 7
##   Year  `Total Emissions (MTCO2e)` Biogenic Carbon Diox~1 Carbon Dioxide (MTCO~2
##   <chr>                      <dbl>                  <dbl>                  <dbl>
## 1 2012                      58312.                  1392.                  47792.
## 2 2013                      60827.                  2315.                  51220.
## 3 2014                      71392.                  2918.                  61072.
## 4 2015                     293849.                  1550.                 285172.
## 5 2016                     307510.                  1368.                 298948.
## 6 2017                     355963.                  1412.                 347190.
## # i abbreviated names: 1: `Biogenic Carbon Dioxide (MTCO2e)`,
## #   2: `Carbon Dioxide (MTCO2e)`
## # i 3 more variables: `Methane (MTCO2e)` <dbl>, `Nitrous Oxide (MTCO2e)` <dbl>,
## #   County <chr>
```

I then formatted the Air pollution data to year only so I can merge it by year with the GHG data. I also summarized the averages for all values by the year.

```
## # A tibble: 6 x 9
##   Year  `O3 Mean` `O3 AQI` `CO Mean` `CO AQI` `SO2 Mean` `SO2 AQI` `NO2 Mean`
##   <chr>     <dbl>    <dbl>     <dbl>    <dbl>      <dbl>     <dbl>      <dbl>
## 1 2000     0.0189     24.2     0.358     5.65       2.99      12.7       18.0
## 2 2001     0.0183     24.1     0.310     5.07       3.79      14.6       17.5
## 3 2002     0.0202     25.8     0.298     4.95       2.66      12.9       17.9
## 4 2003     0.0234     29.4     0.345     5.30       2.63      12.8       16.9
## 5 2004     0.0209     26.3     0.528     7.31       3.02      12.2       16.0
## 6 2005     0.0182     23.3     0.555     7.58       2.99      12.9       15.5
## # i 1 more variable: `NO2 AQI` <dbl>
```

I then joined the two datasets by year to get and idea of what all of their variables and values look like together.

```
## # A tibble: 6 x 15
##   Year  `Total Emissions (MTCO2e)` Biogenic Carbon Diox~1 Carbon Dioxide (MTCO~2
##   <chr>                      <dbl>                  <dbl>                  <dbl>
## 1 2013                      60827.                  2315.                  51220.
## 2 2014                      71392.                  2918.                  61072.
## 3 2015                     293849.                  1550.                 285172.
```

```
## 4 2016                    307510.                  1368.                298948.
## 5 2017                    355963.                  1412.                347190.
## 6 2018                    361356.                  1514.                352301.
## # i abbreviated names: 1: 'Biogenic Carbon Dioxide (MTCO2e)',
## #   2: 'Carbon Dioxide (MTCO2e)'
## # i 11 more variables: 'Methane (MTCO2e)' <dbl>,
## #   'Nitrous Oxide (MTCO2e)' <dbl>, County <chr>, 'O3 Mean' <dbl>,
## #   'O3 AQI' <dbl>, 'CO Mean' <dbl>, 'CO AQI' <dbl>, 'SO2 Mean' <dbl>,
## #   'SO2 AQI' <dbl>, 'NO2 Mean' <dbl>, 'NO2 AQI' <dbl>
```

Variables included for final dataset (**ap__ghg__3**):

Year

Total Emissions - Sum of all GHG emissions (CO, NO, S, etc.), in MTCO2e (metric tons of CO2 equivalent)

Biogenic Carbon Dioxide (MTCO2e)

Carbon Dioxide (MTCO2e)

Methane (MTCO2e)

Nitrous Oxide (MTCO2e)

County - Only used King County Data

O3 Mean - Average Ozone level for a day per year.

O3 AQI - Average Air Quality Index for Ozone per year

CO Mean - Average Carbon Monoxide level for a day per year.

CO AQI - Average Air Quality Index for Carbon Monoxide per year

SO2 Mean - Average Sulphur Dioxide level a day per year

SO2 AQI - Average Air Quality Index for Sulphur Dioxide per year

NO2 Mean - Average Nitrogen Dioxide level a day per year

NO2 AQI - Average Air Quality Index for Nitrogen Dioxide per year

Even though some variables are more relevant than others, I chose to incorporate all these variables so I can show the individual trends for each, giving a more accurate depiction when comparing GHG emissions and air pollution/AQI.

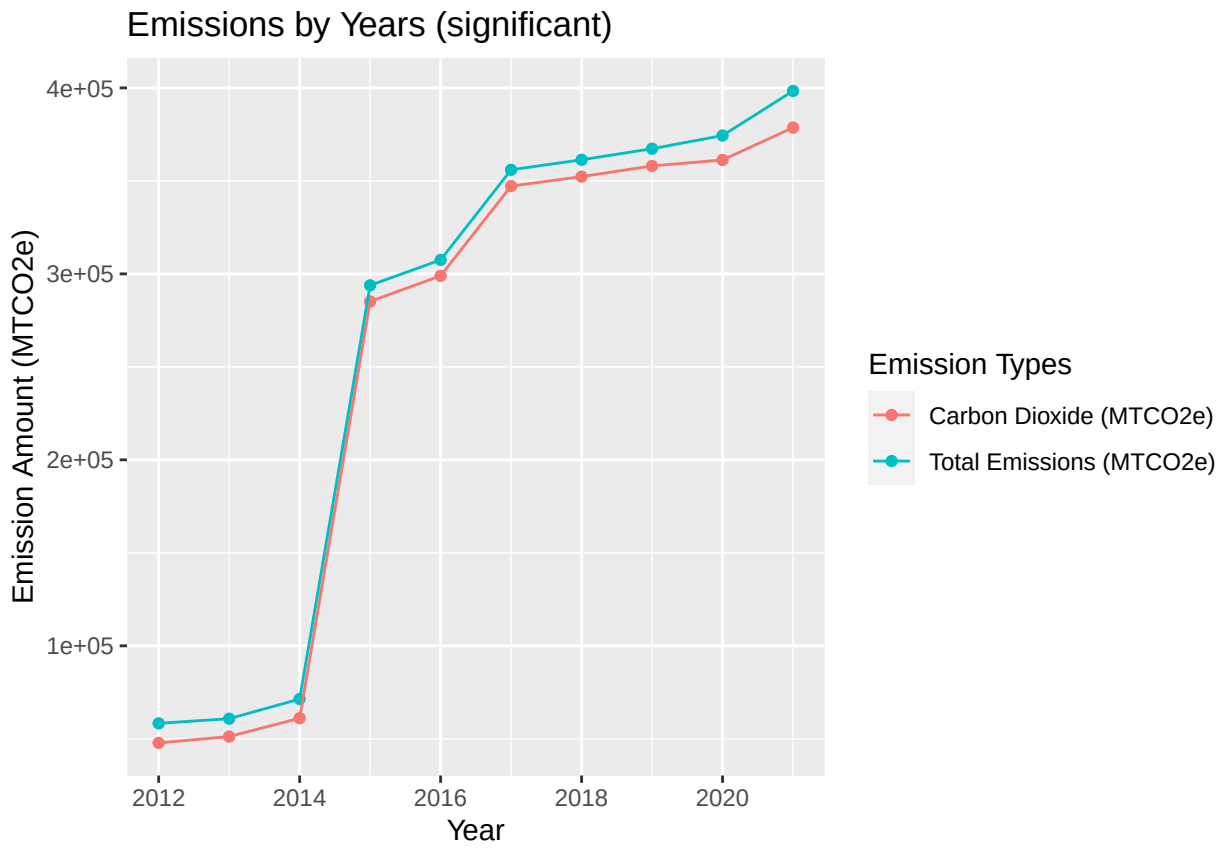
To create the graphs, I then pivoted both the Air pollution and GHG datasets to long form, making it easier to sort the points by the variables I want. For the GHG emissions data, I removed all of the non-numeric variables except for county because it isn't needed when comparing general air pollution in Seattle. For the Air Pollution data I also only kept the numeric data that I needed to graph *AQI values* and *Average Gas Levels* (which is the average of the mean gas levels per day).

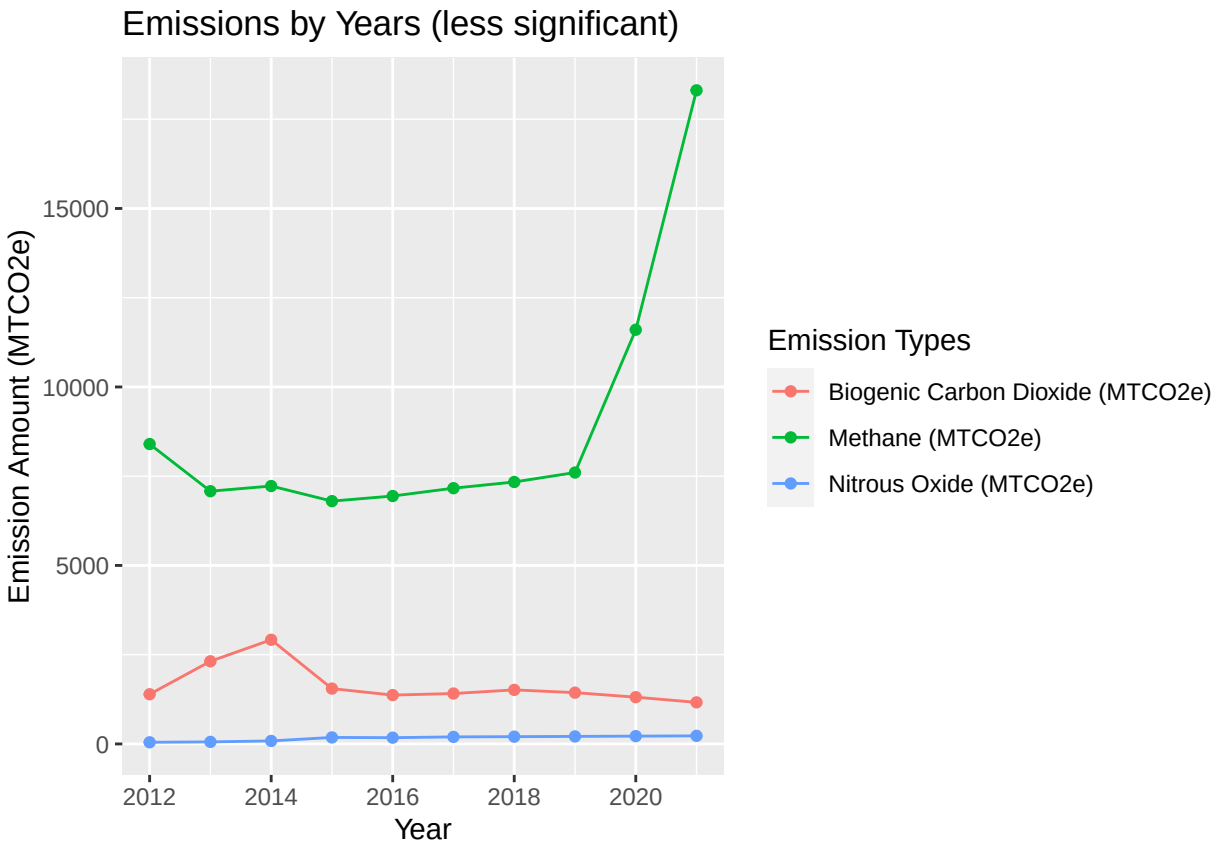
To summarize, all this preparation will be used to create various line plots that compares different factors of GHG emissions to air quality and air pollution data. I selected my sample by making sure that data in both datasets are comparable to each other, so it was important for me to narrow down the scope of location and time. I also didn't use any complex statistical methods, just graphed my data using color-cordinated line plots to demonstrate patterns.

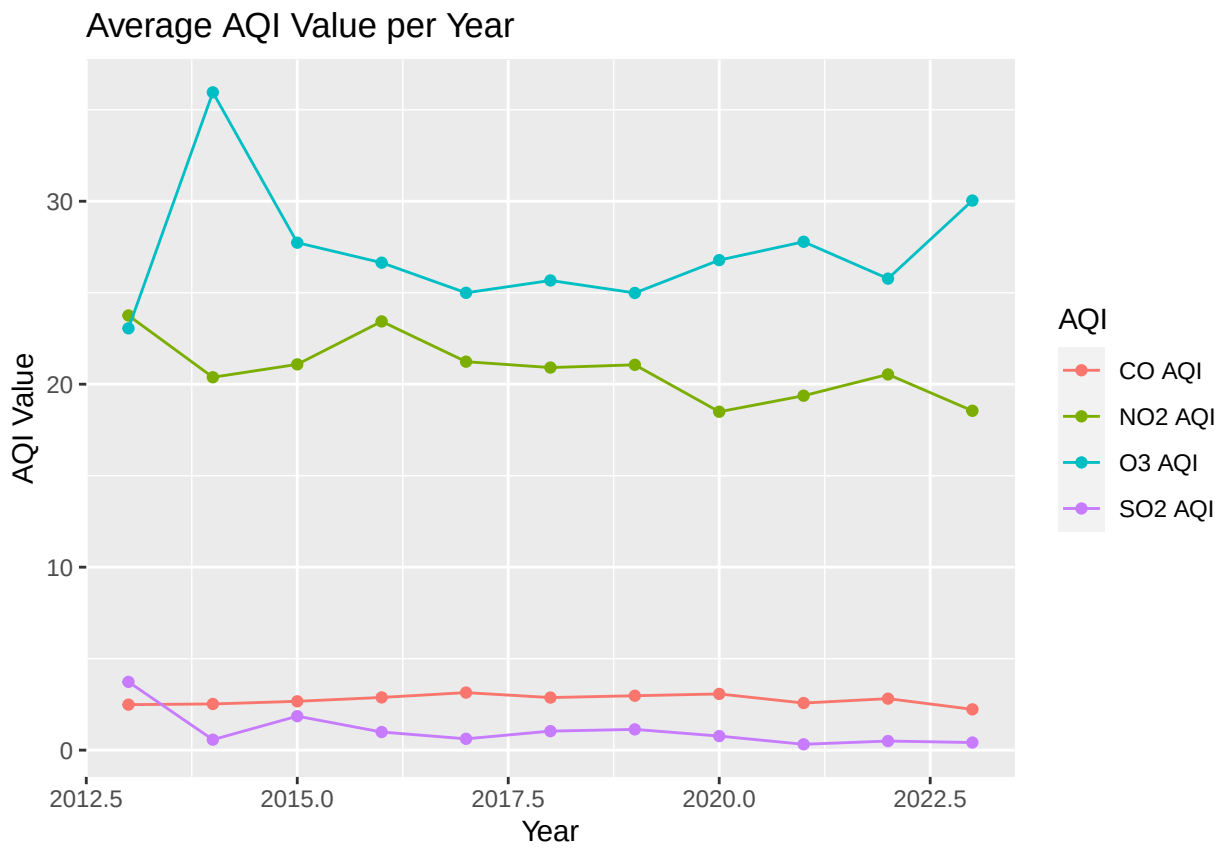
4. Q3 Results

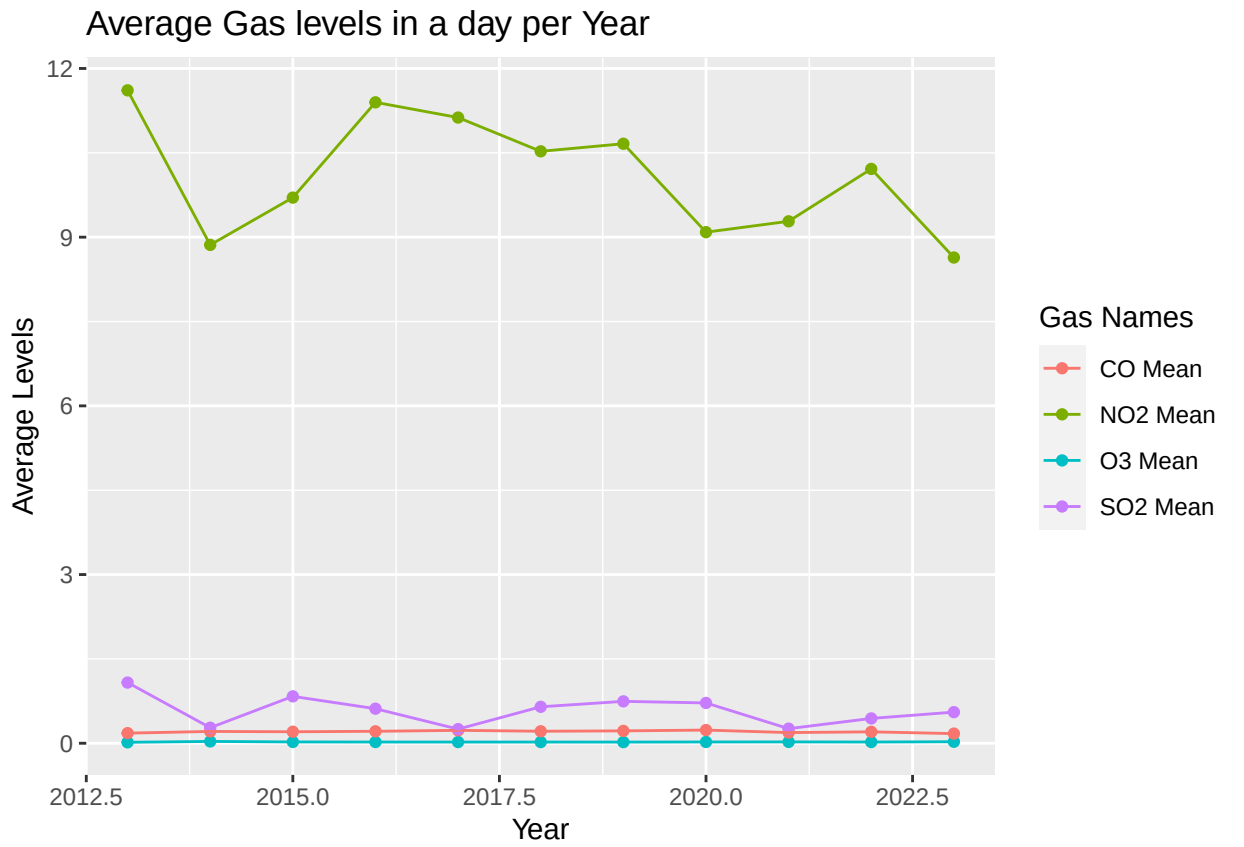
```
## # A tibble: 9 x 15
##   Year  'Total Emissions (MTCO2e)' Biogenic Carbon Diox~1 Carbon Dioxide (MTCO~2
```

```
##   <chr>                      <dbl>                 <dbl>                  <dbl>
## 1 2013                      60827.                2315.                 51220.
## 2 2014                      71392.                2918.                 61072.
## 3 2015                     293849.                1550.                285172.
## 4 2016                     307510.                1368.                298948.
## 5 2017                     355963.                1412.                347190.
## 6 2018                     361356.                1514.                352301.
## 7 2019                     367291.                1437                 358041.
## 8 2020                     374389.                1310.                361258.
## 9 2021                     398349.                1164.                378650.
## # i abbreviated names: 1: 'Biogenic Carbon Dioxide (MTCO2e)',
## #   2: 'Carbon Dioxide (MTCO2e)'
## # i 11 more variables: 'Methane (MTCO2e)' <dbl>,
## #   'Nitrous Oxide (MTCO2e)' <dbl>, County <chr>, 'O3 Mean' <dbl>,
## #   'O3 AQI' <dbl>, 'CO Mean' <dbl>, 'CO AQI' <dbl>, 'SO2 Mean' <dbl>,
## #   'SO2 AQI' <dbl>, 'NO2 Mean' <dbl>, 'NO2 AQI' <dbl>
```




Emissions by Years (less significant)
Emission Amount (MTCO2e)
15000
10000
5000
0
2012
2014
2016
2018
2020
Year
Emission Types
Biogenic Carbon Dioxide (MTCO2e)
Methane (MTCO2e)
Nitrous Oxide (MTCO2e)


Average AQI Value per Year
AQI Value
30
20
10
0
2012.5
2015.0
2017.5
2020.0
2022.5
Year
AQI
CO AQI
NO2 AQI
O3 AQI
SO2 AQI



Our results here are only based on King County area data. They tell us that even though there is a clear increasing trend in *Average GHG Emissions per Year*, for significant GHG such as carbon dioxide and biogenic carbon dioxide the trends for AQI and average gas levels in a day per year seems to be stagnate, flat trend. All AQI variables demonstrate a similar trend throughout the recent period of time with some fluctuations in O3 and NO2 AQI but no visible trend. It's the same for average gas levels with fluctuation in NO2 means but no notable trend. Therefore, we can't conclude a correlation between GHG emissions and AQI or air pollution which generally makes sense for now because there isn't a notable difference between the air in the Seattle area throughout the years (only stretches of days of high AQI generally).

5. Discussion: What do these results mean?

Research Questions:

We set out and asked these three questions to help Washington State residents gain perspective and learn about the correlation between our greenhouse gas emissions and its impacts on our land and air.

Q1 **What are the total greenhouse gas emissions based on *sector, location, and company?***

We started out by asking which *sector, location, and company* are emitting the most greenhouse gases to narrow the scope of our issues. We found that Power Plants, Pulp and Paper, and Petroleum Systems are emitting most in Washington state. Power plants are the primary sector that we need to worry about. Because it these emit the most GHG we understand that the majority of our energy for electricity, heating, cooling, etc. are coming from oil, coal, and other fossil fuel based products. Therefore, the sector data tells us that we need to focus on developing tools for us to use renewable energy. We also see way to develop paper is also releasing lots of GHG and petroleum systems (which extract fuels from fossils or rocks) is also emitting GHG. We found that 2 of the top 3 emitters are fossil fueled based highlighting the importance of

developing renewable energy and contributing to that cause. The company data confirms our findings since the top companies for emissions are petroleum and energy based companies.

Our findings for cities with the most GHG emissions also identifies areas of focus for Washington. We found that the highest emitting cities are majorly along the west of Washington with cities such as **Centralia, Longview, Bellevue, and Tacoma**. With this information, people in these cities could support and notify city councils to these specific issues, spreading awareness and concern to the situation.

Our limitations with this data is that we are only looking at company based data so the total emissions aren't regarding the emissions of everyday humans as well. To bring more awareness to the situation we would need to gather information for household GHG emissions as well. This would likely bring the attention of all Washington people and not just companies, giving them more incentive to make changes to their everyday lives to reduce their GHG output.

Q2 **Is there a correlation between greenhouse gas emissions and wildfire *amounts and sizes* in Washington State?**

We then looked for a correlation between GHG emissions and wildfires, since this is a topic of concern for many people along the west coast in recent time, including Washington residents. In our introduction, we analyzed a correlation between GHG emissions and the largest wildfire sizes. This is concerning but we wanted to look into further detail, involving more instances than just the largest wildfire. Looking at the average size of wildfires (in acres) we can see that there is no true pattern. This tells us that wildfire sizes would likely vary by other circumstances and situations and is likely not majorly due to GHG emissions. However, looking at the data for number of wildfires per year, we are seeing an exponential increase in the past few years, somewhat aligning to our GHG emission patterns. This means that there is a possible that GHG emissions are increasing the number of large wildfires in Washington but its more than likely not completely causal. This data should make Washington natives aware of the possible correlation between GHG emissions and wildfire amounts, but isn't something to be overly concerned about.

The limitations of this data and story is that even though people should be made aware of a possible correlation, this isn't an experiment so we can't assume causation. This means that we haven't narrowed down the scope of the causes for these increased wildfire amounts. For our future work, we would need to understand the gases in the air that can cause wildfires to increase in size and amount and analyze the amount of that gas being emitted or is present in the air already.

Q3: **Is there a correlation between GHG emissions and air pollution?**

Finally, our last research question analyzes the correlation between GHG emissions, AQI, and air pollution. For this question, for the sake of accuracy, we only sampled the King County data since the air pollution dataset only contained data from the greater Seattle area. Nonetheless, we could still use this data to discover a correlation (if any) between the 3 variables. The GHG emission data demonstrates clear increase in gases such as CO2 and Methane, contributing to the increasing trend for total emissions, however, our results for AQI and gas levels displayed a stagnate/random trend. This means that there is likely no correlation between GHG and air pollution or AQI with our sample (GHG doesn't impact AQI). This shows Washington residents that although we are increasing in total emissions, it hasn't reached a point yet where our air quality is largely impacted, relieving some concerns.

A limitation of the data for this research question is that we are looking at data limited to the Seattle area, for the future we should gather more data all around Washington to give more accurate results for our research question. We could also gather data of when the AQI is at its highest in Washington and see if the total emissions during those specific days are increased or not.

6. Summary

Informing people of issues like excess GHG emissions is the first step to creating a solution. We are targetting Washington residents, scientists, and policymakers with this project. People generally must first visualize the issue at hand before they have the incentive to work towards the cause. With this project we are looked at

GHG emissions throughout recent history (2010-2023) to demonstrate where the issue of increased emissions is taking place and the possible impacts of these emissions on our state. We discovered a general increasing trend when it comes to total emissions and ranked the sectors, companies, and cities based off their emission totals. After we analyzed the locations of the issue we moved to how this issue could lead to issues that impact human and wildlife. We first compared this GHG data to wildfire data in Washington and discovered a possible correlation between an increase in GHG emissions throughout the years and an increase in wildfire sizes. We then looked at AQI and air pollution data to discover that on average, our air is still safe and isn't showing any signs of worsening yet even with an increase in total emissions. Overall, our concerns with an increase in GHG emissions impacting our everyday lives shouldn't be present yet, but we do need to be wary of this increase since it could possible impact our future. We need to work towards different solutions to reduce our GHG emissions for the future at hand.