Analysis on GreenHouse Gases and Regional Emission Based on Industry Activities from 2017 to 2021

Yihan Cao, Kayla Katakis, Hung Nguyen

Author contributions:

Yihan Cao contributed Abstract, Tables and Graphs, and Formatting
Kayla Katakis contributed writing for the Materials and Methods, Results, and Discussion sections
Hung Nguyen contributed the topic of interest, writing the background and aims, and graphs and tables

Abstract:

Greenhouse gas emissions problems have been one of the most controversial topics in the past twenty years or so. Addressing the problem is a crucial part of environmental protection. In this project, we will use the GHG dataset from World Bank Open Data to forward the discussion and find possible relationships between regions and different gases emissions or between industrial activities and gas emissions. Our finding indicates that the main industry activity that devotes the most overall GHG emissions is *Electricity, Gas, Steam and Air Conditioning Supply*, and Asia has the largest emission worldwide which contributes up to 50% of the world's emission.

Introduction:

Background:

Greenhouse gases (GHGs) are gases that trap heat in the Earth's atmosphere and contribute to the greenhouse effect. The most significant GHGs are carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), and fluorinated gases. These gases are released into the atmosphere through human activities such as burning fossil fuels, deforestation, and industrial processes.

The buildup of GHGs in the atmosphere has led to a warming of the Earth's surface, which is causing climate change. The consequences of climate change are far-reaching and affect all aspects of society, including food production, water availability, health, and infrastructure.

In recent years, there has been an increasing recognition of the urgency of addressing this global issue. The international community has come together to set targets for reducing GHG emissions. Many countries, cities, and companies are taking steps to reduce their carbon footprint and transition to renewable energy sources. Addressing GHG emissions is crucial for the future of our planet and society.

Aims:

In this project, we focused on two relationships: first, the relationship between an industry of human activities and the emissions of GHGs, and second, the relationship between regions and the emissions of GHGs. In other words, questions such as "is there an industry that is principally responsible for the release of a certain GHG," "which industry is responsible for most of GHGs' emission globally," and "is there a region that is principally responsible for the release of a certain GHG" are within our scope of interest.

Materials and Methods:

Datasets:

We will be working with the Annual Greenhouse Gas (GHG) dataset from *World Bank Open Data*. We obtained this dataset by navigating through the World Bank website to their data catalog where we found the GHG dataset. This dataset had many variables that were irrelevant to the process of answering our research questions, so we narrowed it down to 4 of our most important variables across 8 columns. Here is a table detailing the final variables that we chose:

Variable Name	Description	Variable Type	Units
Country	The name of the continent or subregion	Character	None
Industry	The activities causing the GHG	Character	None
Gas_Type	Type of GHG gas	Character	None
F2017 - F2021	Measurement of GHG gases in the corresponding year	Numeric	Million metric tons of the carbon dioxide equivalent

Table 1. Variable description of selected variables in the GHG dataset

We also performed some data tidying by renaming some of the variables and removing observations that weren't pertinent to our research questions. We renamed the *Country* column to *Region* to more accurately reflect the values that this variable contains and to maintain clarity. Additionally, some of the *regions* didn't represent countries or some continents. Rather, they represented types of economies or geographic areas that likely encompassed multiple countries. Since these types of regions are not important, we will remove them from our dataset. Our final dataset has 939 observations across 8 columns. Here is a preview of the data that we will be working with:

	Region	Industry	Gas Type	F2017	F2018	F2019	F2020	F2021
50	Africa	Agriculture, Forestry and Fishing	Carbon dioxide	8.598177	8.885628	9.193573	9.519654	9.869859
51	Africa	Agriculture, Forestry and Fishing	Greenhouse gas	801.551149	820.263383	841.909644	859.337790	875.720444
52	Africa	Agriculture, Forestry and Fishing	Methane	533.778860	548.656038	563.256921	574.852827	584.961495
53	Africa	Agriculture, Forestry and Fishing	Nitrous oxide	259.174112	262.721717	269.459149	274.965308	280.889091
54	Africa	Construction	Carbon dioxide	93.987951	95.613220	96.690936	87.940823	93.301037

Table 2. Preview of dataset

Methods:

We mainly want to explore relationships between emissions and our other variables. Specifically, how these relationships change over time.

Given how our data is set up and the questions we want to answer, our research leans more towards visualization rather than modeling. We conducted some initial research to look at how our variables were distributed, and then we created some more in depth plots to visualize the relationships discussed in our research questions.

We began by simply grouping the data to examine relationships numerically, which will be discussed further in detail in our results section. Next, we created some barplots to visualize how each of our categorical variables is distributed in order to see whether they might or might not skew our results later on.

Then, we get to our deeper analysis. We used the *altair* package to create more detailed charts that give us insight into our research questions. We primarily focused on emissions, and we looked at how each industry and region influences these measurements as well as how they change over time.

Results

Initial Observations:

We first took a look at how *Region* is distributed. Here is both a list view and a simple barplot.

```
World                             50
Northern Europe                   50
Asia                              50
Australia and New Zealand         50
Western Europe                    50
Western Asia                      50
Eastern Europe                    50
Europe                            50
Southern Europe                   50
Oceania                           50
Northern America                  45
Americas                          45
Latin America and the Caribbean   45
Eastern Asia                      45
Southern Asia                     44
Northern Africa                   43
South-eastern Asia                43
Sub-Saharan Africa                43
Central Asia                      43
Africa                            43
Name: Region, dtype: int64
```

Table 3. Counts of region appeared in the dataset

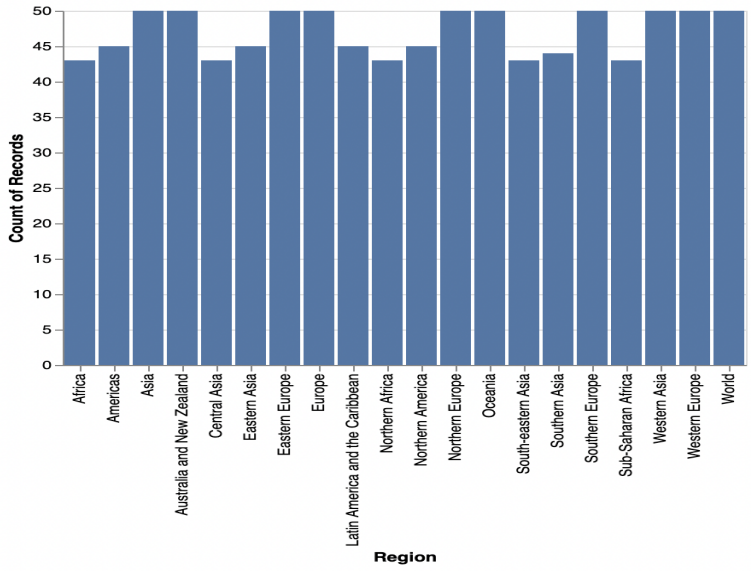


Figure 1. Barplot showing the distribution of region in the dataset

Region is relatively evenly distributed, with the countries with the most and least representation differing by only 7 observations. We also looked at how *Industry* and *Gas Type* were distributed. And the results were similar!

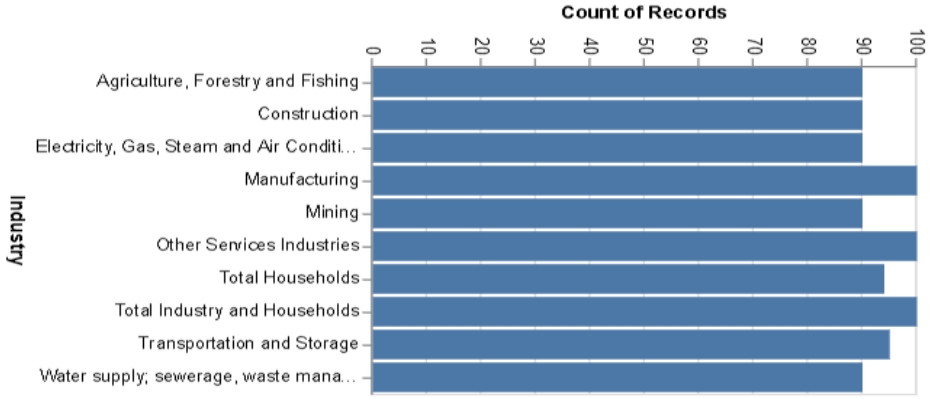


Figure 2. Horizontal barplot showing distribution of industry activities

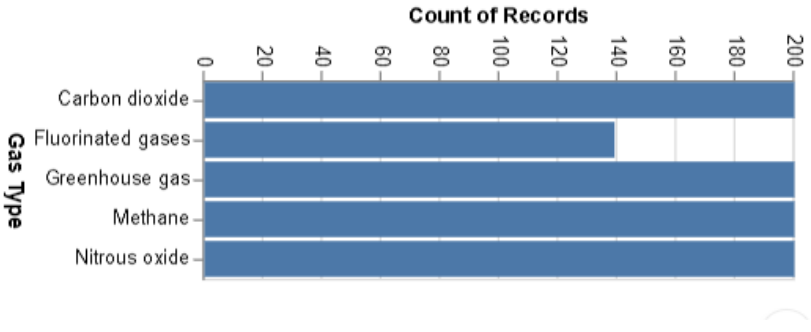


Figure 3. Horizontal barplot showing distribution of the different types of GHGs

The most and least represented industries differ by only 10 observations. Gas types all have 200 observations except for fluorinated gases, which are slightly lower at 140. However, due to the high number of total observations, we assume that they are evenly distributed enough for our research purposes. The even distribution of variables helps ensure that our results are accurate and not skewed by imbalanced proportions.

Deeper Analysis:

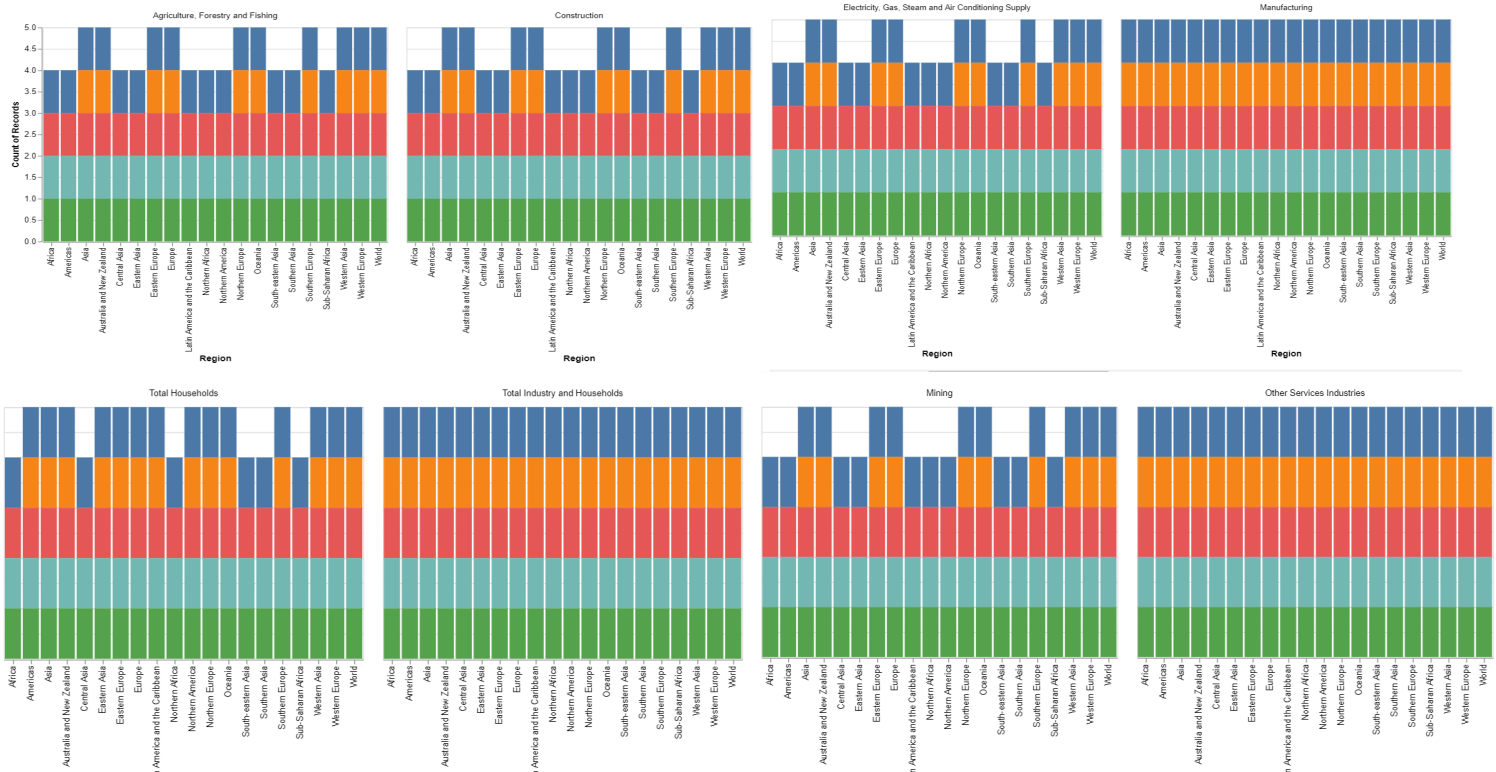
With some basic observations about our data out of the way, we can move on to deeper analysis with more detailed visualizations. Below is a table of the average emissions of each gas type in each region per year.

Region	Gas Type	F2017	F2018	F2019	F2020	F2021
Africa	Carbon dioxide	294.771920	301.757536	305.917024	282.313748	298.213724
	Fluorinated gases	36.848115	41.231802	45.130290	49.100145	53.456377
	Greenhouse gas	624.226028	639.697797	652.554296	628.945105	651.841588
	Methane	252.421211	258.726723	264.583885	262.581504	266.748683
	Nitrous oxide	65.978463	66.843997	68.514300	69.319810	70.842268
...	...	...	...	...	...	...
World	Carbon dioxide	7349.200712	7505.287384	7522.587904	7152.564764	7523.279165
	Fluorinated gases	186.924313	199.767228	208.482461	218.620274	230.696594
	Greenhouse gas	9885.437494	10082.535706	10118.368100	9783.678030	10192.504839
	Methane	1779.833461	1803.300437	1813.184431	1842.909536	1856.816331
	Nitrous oxide	569.479009	574.180656	574.113305	569.583456	581.712750

Table 4. Regional gas emissions with gas types from 2017-2021

We can see in the table that there is a gradual increase in the level of greenhouse gas emissions from 2017 to 2021, both within each region and on a global scale. There are, however, some nuances with these results that will be discussed in the next section.

Removing the time aspect, we look at how observations are distributed across each region based on the industry.



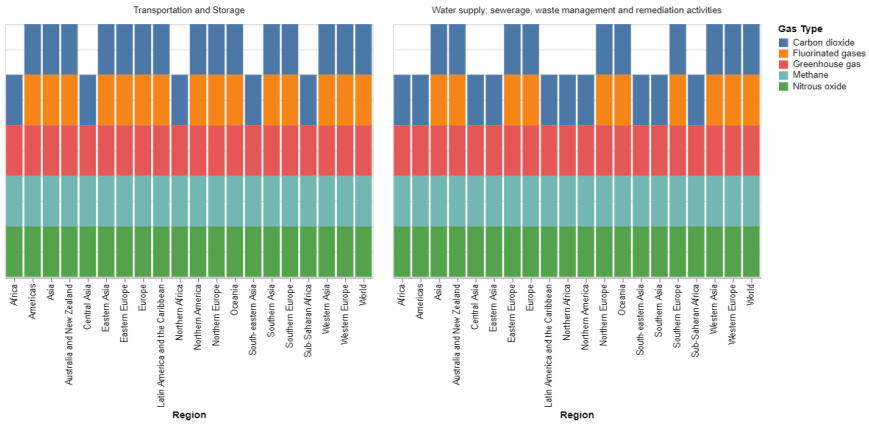


Figure 4. Barplot of regional gas emissions colored by gas type faceted on industry activities

This further shows that fluorinated gases have fewer records of emissions. Notably, certain industries had fewer records of fluorinated gases. Furthermore, certain regions were found to have fewer records of the emissions of fluorinated gases within these industries.

Next, we took a closer look at how emissions change over time. Faceting by year, we created barplots of the emission levels in each industry, color-coded by gas type.

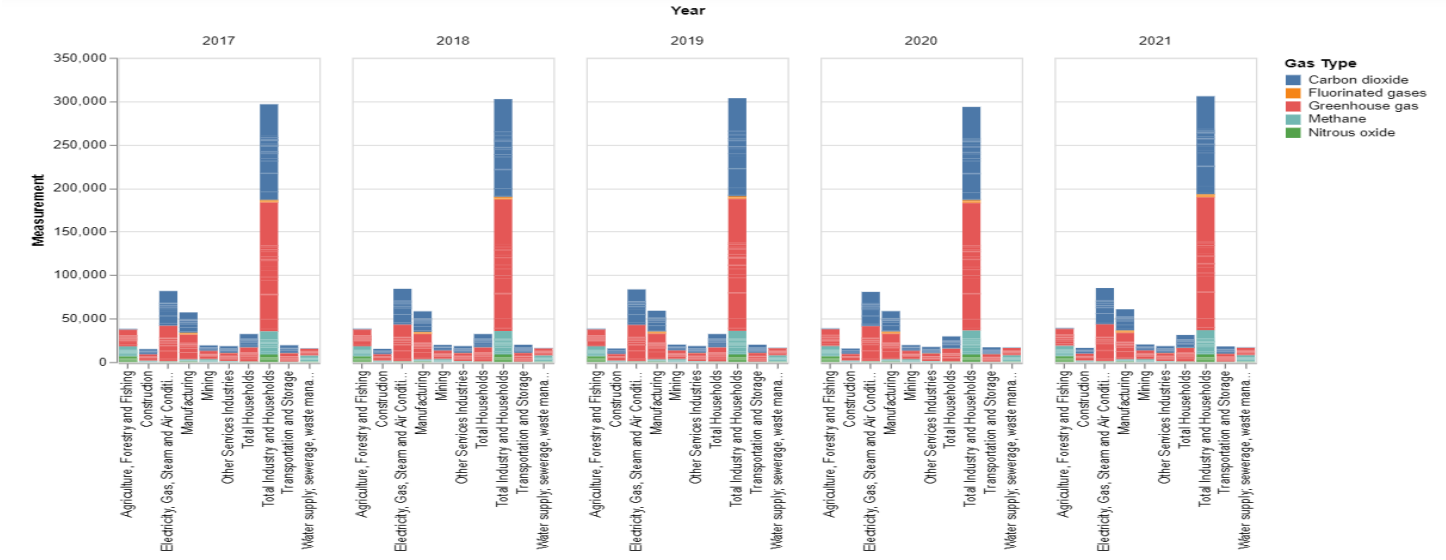


Figure 5: Barplot of GHG emission by industry, color-coded by Gas Type, faceted by Year

The red portion represents all GHG so we can ignore it because it is irrelevant to the discussion. Overall, there is a slight increase in the emissions of GHGs besides a slight decrease in 2020. Although the increase is not noticeable in the barplots, it is apparent in Table 4. Across all regions, the *Electricity Gas, Steam, and Air Conditioning* industry produced the most GHGs

in all 5 years. This is understandable as electricity is extremely prevalent in not only humans' daily lives but also in technological and industrial processes.

In regards to time, we also looked at how each region affects emission levels. Again, we have color-coded our plot by gas type.

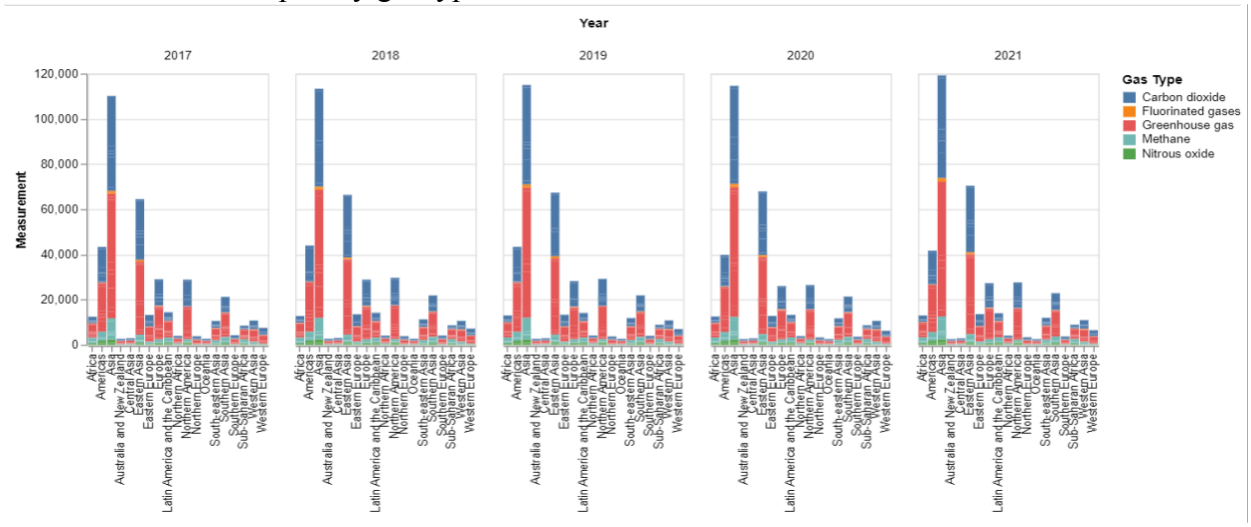


Figure 6: Barplot of GHG emission by region, color-coded by Gas Type, faceted by Year

Similar to the previous plot, there is an apparent increase in GHG emissions for all regions across all industries. For instance, Asia as a whole produced approximately 110,000 million metric tons of carbon dioxide equivalent of GHGs in 2017. In 2021, the continent reached 120,000. Furthermore, in comparison to other subregions, Eastern Asia emits significantly more GHG. Similarly, in comparison to other continents, Asia as a whole emits more than twice the level of emissions of their closest competitor, the Americas. This can be explained using the figure below (Figure 7) which illustrates the emissions of each region in 2021 color-coded by the industry.

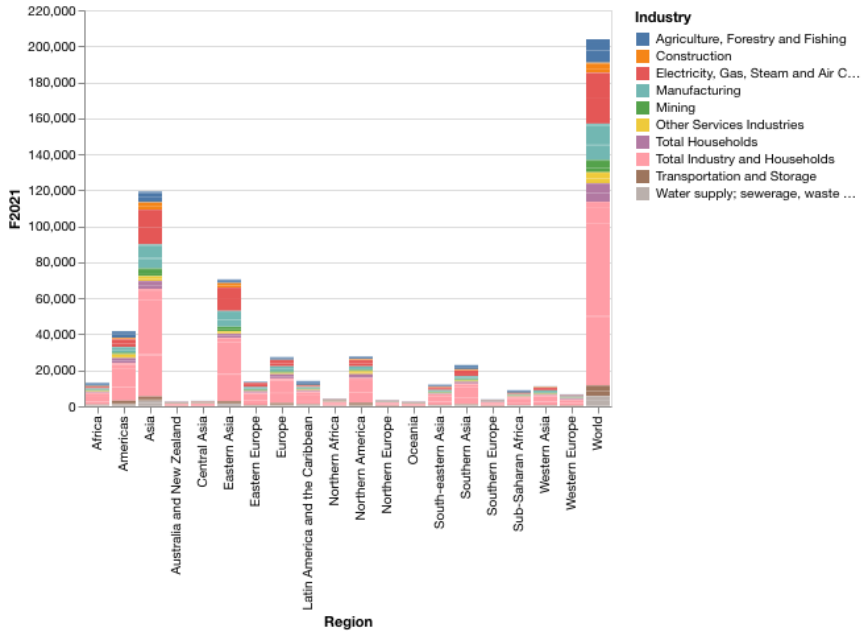


Figure 7: Barplot of GHG emissions in 2021 by region, color-coded by industry

This figure visualizes how each industry contributes to the emission of GHG in each region. The pink portion represents Total Industry and Households so it is not relevant in our analysis. From Figure 5, we learned that the Electric, Gas, Steam, and Air Conditioning industry produces the most GHG. Evidently, this industry in Asia is the most prominent in comparison to other continents which explains why Asia emits the most GHGs.

Discussion:

Overall, our visual analysis provided us with some interesting insight. As shown in our results section, each region sees a gradual increase in emission level from year to year with an interesting drop in 2020 for most gases in most regions. Interestingly enough, the only gas that did not follow this trend was fluorinated gas, which also sees less representation in the dataset. Though it was most plausibly caused by the global pandemic of COVID-19, or potentially poor data collection, the true reason behind this fluke in the data would be a possible venture for further research. In addition, certain regions emit far less GHGs than others, particularly Asia as a whole and its subregions that tower over their closest competitors, such as the Americas, in their emission levels. Again since the reasoning behind these levels was not investigated in this project, further research would be required; however, we speculate that these stark differences lie in the industries that these regions control.

In conclusion, visual analysis was adequate in answering our research questions. Carbon dioxide is the GHG that is emitted the most regardless of region and industry with methane being close by in second place. Over time, these levels may not change drastically, but industry and region do play a role in how these levels are distributed. The electricity, gas, steam, and air conditioning industry produces the highest levels, and, as discussed before, Asia is the biggest perpetrator.

With these results, our next steps are clear. Further research could involve investigating the reasoning behind these emissions, and potentially how they can be lowered. As we are all aware, excessive emissions of GHG threaten the future of our planet. By narrowing down what influences these emissions and how they change, we can research sustainable ways to maintain the industries that produce them.