The Effects of Carbon Dioxide on Global Warming

Mark Bakishiyev, Eric Wu

Professor: Suzanne McIntosh, Ann Malavet, New York University

Abstract

Greenhouse gas has been known to be a contributing factor to global temperature increases. Of the numerous greenhouse gases emitted by human activities, Carbon Dioxide (CO2) has been the most prominent, due to humans' dependency on fossil fuels. Our project further cements analysis presented in the research paper by Daniel A. Lashof and Dilip R. Ahuja by utilizing data sets of global temperature and global carbon dioxide emissions. Our project enabled us to form an analysis specifically addressing the issue of carbon dioxide and how it relates to global warming. Using global carbon dioxide emissions and global temperature data sets, we used Hadoop tools such as MapReduce and Hive to create an analysis and Tableau to create a visualization. We observed that as global carbon dioxide emissions exponentially increased, global temperature correlated with a linear growth. We also looked into the biggest contributors towards global carbon dioxide emissions and reported that, historically, the United States has outputted the most carbon dioxide. While the global carbon footprint is distributed among every nations' CO2 output, we must remember that individuals make up a nation. Therefore, we hope that this study helps to potentially facilitate more discussion on a bolder effort to address climate change.

1. Introduction

The idea of climate change, specifically the rising global temperatures due to greenhouse gases, is not a new issue. Numerous sources have given evidence on the dangerous effects associated with climate change. Furthermore, given the global scale, the eventual effects won't discriminate, but afflict everyone. The issue, however, is that the immediate detrimental effects are less-so concrete compared to the benefits associated with transportation, electricity, industry, etc., thus lowering greenhouse gas emissions is often put on the back-burner for more immediate issues. However, as we continue to delay large-scale action to address the problem, we're digging ourselves into a hole. While we don't disagree that there have been strides towards lowering the global carbon footprint such as multiple nations' pledge to reduce carbon dioxide emissions, it's now no longer an issue of if we can make progress, but rather if we can make progress fast enough. According to NASA, the global average temperature can rise anywhere between 1.39 degrees to 5.55 degrees Celsius within the next century (NASA, 2020). To keep the temperature change on the lower end of the spectrum, its necessary that everyone is aware of the possible dangers associated with rising global carbon dioxide emissions and plays their part. Particularly, we believe that raising awareness to the issue at hand, must be more consistently addressed. Developments such as the months of drought in Australia breeding massive bushfires, the melting polar ice caps in the Arctics, and ocean acidification are premonitions to larger and more unprecedented disasters to come.

Our goal isn't to discuss whether or not global warming exists, but rather to quantify the correlation of the greatest players in greenhouse gas emissions (carbon dioxide) and global temperature increase. Utilizing greenhouse gas emissions and global temperature data sets, this paper presents a comprehensive analysis of the historic trends relating the Earth's temperature to Carbon Dioxide, and seeks to understand if these trends can be conceptualized. We will be using the design diagram as seen in Figure 1.

We started by taking our two data sets and downloading them onto our local computers in the form of CSVs. Given that both files were relatively small, we transferred the file to the Dumbo Login Node and then put the file onto HDFS. With the data distributed across HDFS, we then ran our cleaning/profiling MapReduce code and then used HiveQL queries to extract specific analytics we wanted to look into. We then compared our analytics with that of outside research mentioned in Section 3.

2. Motivation

One of our motivations behind this topic was to enlighten our own understanding of Climate Change, and specifically, we were curious at the moment when modern global warming came into being. Our global temperature data actually stretches all the way back to 1750, and we wanted to see how sea/land temperatures rose over the years. Does the issue stretch back to some of the earliest, recorded years of global temperature data according to our data set? This is one of the ideas we wanted to explore. Yet, perhaps our most important motivation for this project is to raise awareness to the issue of climate change. The issue consistently gets eclipsed by other world events, and this is worrisome because even if we took aggressive action to combat it now, it has been reported that there is a historical 'lag' in final temperature rise and atmospheric concentration (Ritchie et al., 2017). This means that even after stabilizing atmospheric condi-

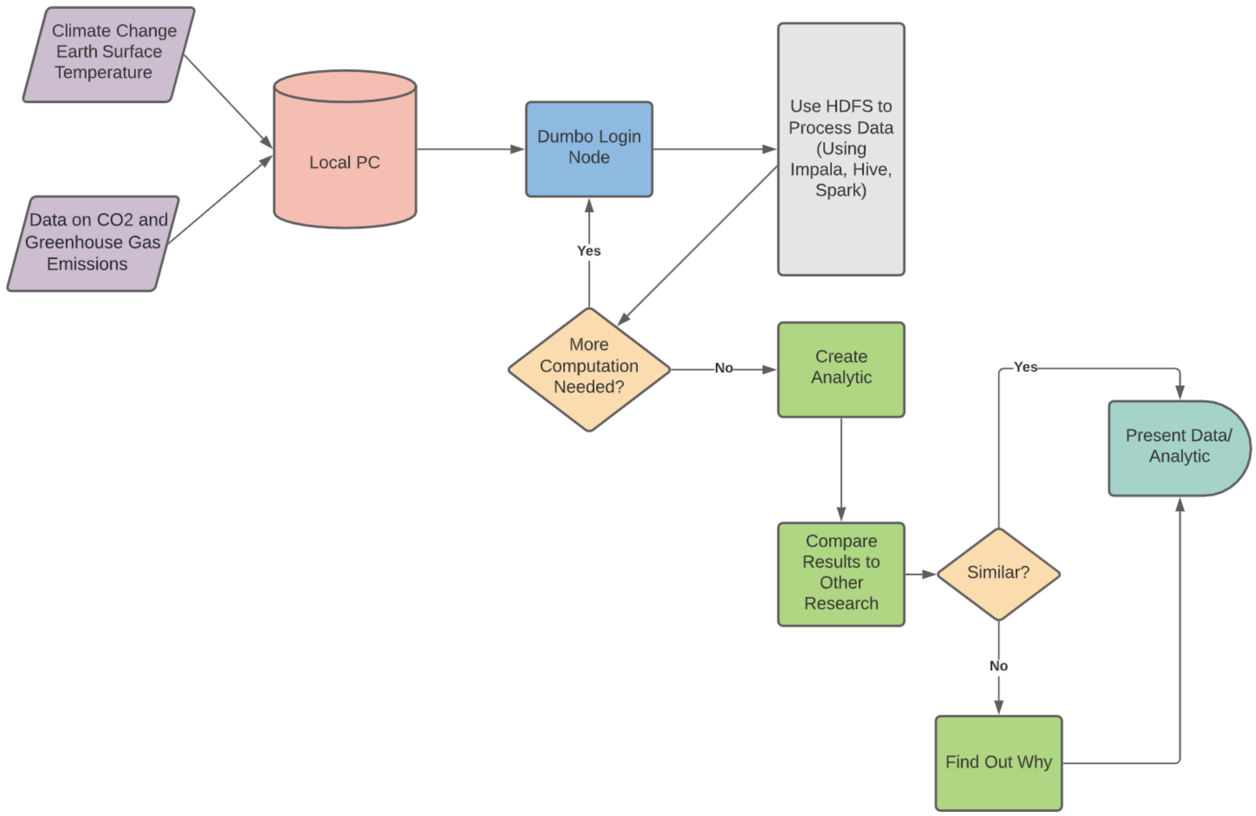

Fig. 1.— The Data Flow Diagram

tions, temperatures could continue to rise from years to decades.

With all this in mind, climate change is a crisis. It is a crisis that effects everyone, and while we hope our analysis will benefit the general population, we hope that our work specifically benefits those who have been more severely impacted by more recent developments attributed to the changing climate. These groups may be people who have been displaced by wildfires, rising sea levels, and respiratory issues (due to decline in air quality). This will benefit them because it will add to the already large body of work that stresses how much the climate crisis needs to be tackled, and how there will ultimately be more suffering if it is not dealt with soon.

3. Related Works

As mentioned in our introduction, this idea of Greenhouse gases (specifically carbon dioxide) emissions in our atmosphere correlating to the rise in global temperature is not a novel idea. We structured much of our research on that of Daniel A. Lashof and Dilip R. Ahuja article where they compared greenhouse gases and their effects on global warming (Lashof et al., 1990). A notable

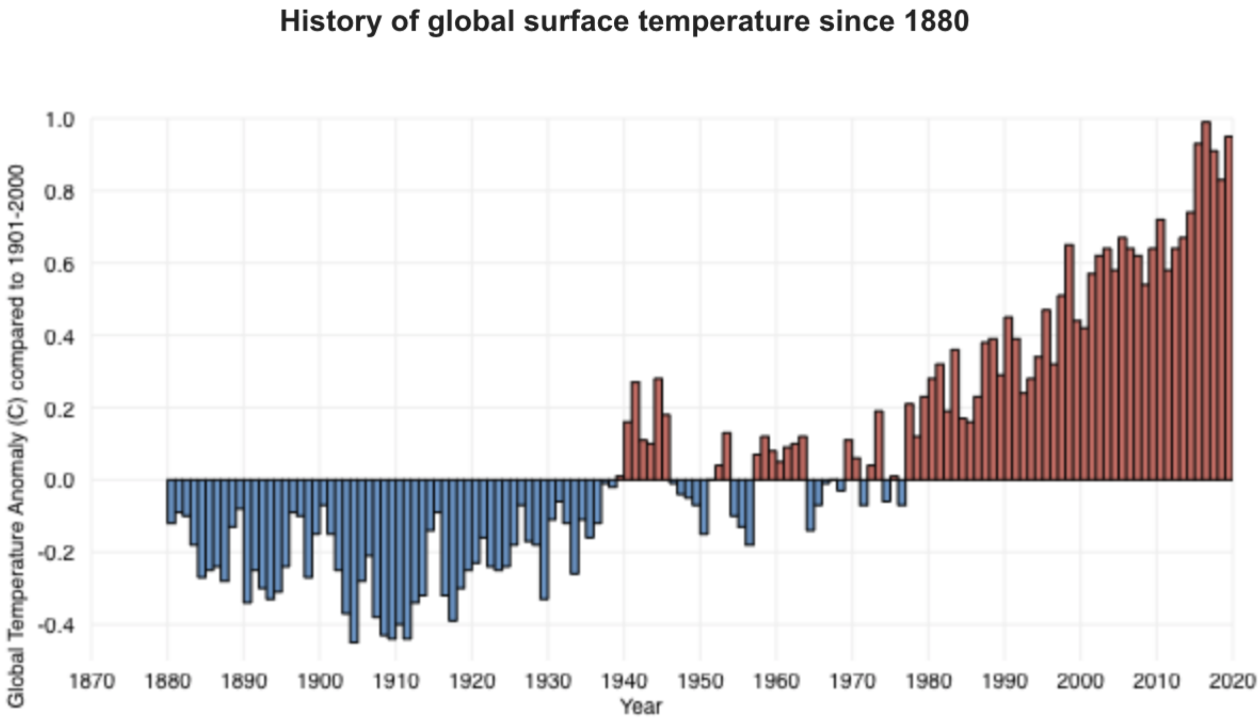

Fig. 2.— Bulletin of the American Meteorological Society Graph on global surface temperature increase. The graph shows average annual global temperatures since 1880 compared to the long-term average (NOAA, 2020)

statistic that was brought up in their research was that carbon dioxide emissions account for 80% of greenhouse gas emissions, which is why we ended up centralizing specifically on it. They had purposed their papers as a forewarning to governments to look into the most cost-effective emissions policies towards lowering the effects on global warming at both the national and international levels.

Using this information, we wanted to further solidify this correlation between carbon dioxide and global temperature increase. After creating our temperature analysis, we compared the accuracy of our results to Bulletin of the American Meteorological Society's article where they discussed the unprecedented event of global warming (Sanchez-Lugo et al., 2018). It turned out much of their research on temperature had fell in-line with our analysis, often pulling similar results on their data as our own. Something interesting to also note from this article was that it uses a "zero-base" global average temperature which established the temperature to be around 14°C based on the long-term average global temperature. Doing so, it was easier to visualize temperature change over time as seen in figure 2. One important takeaway from this article was that it accounted for data starting past the 1880s, which was a smaller range than our initial graph of land temperatures. This was important in terms of analyzing our visualization because it led us to believe that there was high uncertainty in the earlier years on measuring global temperatures due to the large fluctuations as well as the not-so-visible trend between 1750s to 1880. We ended up incorporating this in our second analysis visualization. In terms of specific analytics, the article had ended up reporting that there existed a definite increase of about 2°C in global average surface temperature that has occurred since the pre-industrial era to current times. In a similar consensus, in our data we also had seen a similar decisive increase in temperature, increasing in a more dramatic fashion after the early 1960s.

Next, we wanted to affirm our analysis on the relationship between carbon dioxide emissions and global average temperature increase. Although evidenced by our analysis and by sources such as Friedlingstein's research on the persistent growth of carbon dioxide emissions, we couldn't yet prove that correlation between the two attributes meant causation (Friedlingstein et al., 2014). In order to further push this narrative of the proclaimed greenhouse effect, we explored outside research by the National Centers for Environmental Information, where they looked into underlying studies and the cause and effect between carbon dioxide and the climate. Using paleoclimate proxies such as changes in the glacial climate including altered vegetation, land surface characteristics, and ice sheet extent, they were able to determine the relationship between carbon dioxide and temperature change from as far back as 800,000 years ago. As seen in figure 3, there is a clear interrelationship between the two, NCEI even noting that carbon dioxide to be the "extremely likely" dominant cause of the observed warming throughout history.

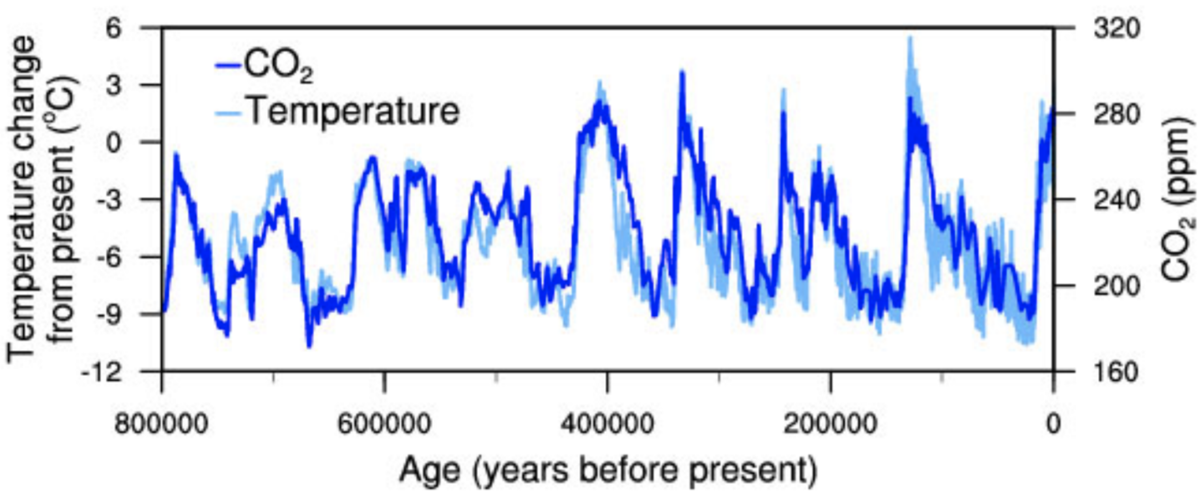

Fig. 3.— Temperature change (light blue) and carbon dioxide change (dark blue) measured from the EPICA Dome C ice core in Antarctica

4. Data Sets

4.1 Earth Surface Temperature Data

Our first data set was provided by Berkeley Earth, and it contained global land and ocean-land data of earth surface temperatures. The data points included: date, avg. land temperature, avg. land temperature uncertainty, land max temperature, land max temperature uncertainty, land min. temperature, land min. temperature uncertainty, land-ocean avg. temperature, and land-ocean avg. temperature uncertainty. The schemas that we used in our analysis were defined as follows:

- **Year** - Double
- **Avg. Land Temperature** - Double

and (next bullet list)

- **Year** - Double
- **Avg. Ocean-Land Temperature** - Double

The data for average surface land temperature actually dates back to 1750, but the ocean-land data only started to be recorded by the mid-1800s. We thought it was important to include two different analytics relating the temperature data to the CO2 emissions because there is more of an aspect of uncertainty that is associated with the earlier years of the average, global land temperature data. Yet, the ocean-land data has smaller degrees of uncertainty when compared with solely just the land data. The size of the data set was 201 KB.

4.2 CO2 and Greenhouse Gases Emissions Data

Our second data set contained a collection of key metrics provided by Our World in Data, and some of the points included were CO2 emissions, greenhouse gases, and energy mix. The data further brokedown specifics on each of the energy mix, which ended up being 32 columns. However, he schema that we used in our analysis for this data set is described below:

- **Country code** - String
- **Country** - String
- **year** - Double
- **CO2** - Double
- **Share of Global CO2** - Double

Originally before cleaning the data set, there existed a lot of information entailing other relevant data points besides just CO2 emissions throughout history. We chose not to include these other points because we specifically wanted to explore the relationship between global temperature and CO2 emissions directly. The size of the data set was 3.1 MB.

5. Analytics Stages

5.1 Ingestion

Due to the relatively small size of our data, the ingestion process was rather simple. First, we downloaded the data sets onto our local machines as CSV files. Then, we used the following command to upload the CSV's to Dumbo:

scp (insert file name here) dumbo:

After uploading our data files properly to Dumbo, we were able to put them on HDFS with the following command:

hdfs dfs -put /path/to/dir/(insert file name here)

This completed the ingestion process. Next, we focused on cleaning.

5.2 Cleaning

For both data sets, the schema we intended to use for analysis (Section 4) meant that we had to drop certain data points when writing our MapReduce programs. However, we had not anticipated that there were other hurdles that we had to deal with. Specifically, there were some records in our data sets that had incomplete or missing information which made writing Mapper code a little cumbersome because we had to consider the empty fields in the CSV files.

In addition, two separate MapReduce programs were written for temperatures to isolate land temperatures and land-ocean temperatures for the purposes of running analysis. The programs themselves were not too different from one another except for the fact that we were changing the relevant data fields to gather. The Reducer code utilized one reducer to take the input from the mapper and outputted the results to /path/to/dir/**part-r-00000** on HDFS. This was then extracted from HDFS for analysis using Hive.

In contrast, only one MapReduce program was written to extract relevant data for the CO2 emissions data set. Records that did not have to do with our analysis were ultimately dropped, and the Reducer code used only one reducer to take input from the mapper and outputted results to /path/to/dir/**part-r-00000** on HDFS. This was also then extracted from HDFS for analysis using Hive.

5.3 Profiling

The profiling code for both data sets was essentially the same in that both MapReduce programs written for this simply summed the number of entries that were contained in each data set. The Mapper code simply assigned the same key to each entry in the data set with an IntWritable value of 1 where it was then passed to the Reducer where a variable counted the sum total entries. The number of entries for global land temperature dataset went from 3193 to 3180. On the other hand, the global land-sea temperature dataset 3193 to 1992. For the Carbon Dioxide dataset, we cleaned the number of entries from 24017 to 19149.

5.4 Analytics Code

To produce the analytics for our data sets, we used Hive to query the necessary data. Our approach for using hive was also rather simple. First, we created tables for the cleaned output files that we received from our MapReduce jobs. The query to create the temperature table is given by:

create external table temperaturetable (year int, temperature double) row format delimited fields terminated by ',' location '/path/to/mapreduce/temperature/output';

and the CO2 table creation query is given by:

create external table co2table (year int, emission double) row format delimited fields terminated by ',' location '/path/to/mapreduce/CO2/output';

Since we wanted to only have one measurement for every year of recorded global temperature, we chose to average all the temperatures in a given year. We used the avg() function provided in hive to do this, and the following hive query was used to create the table:

create table avgtemperature as select temperaturetable.year, temperaturetable.temperature from (select year, avg(temperature) as temperature from temperaturetable group by year) temperaturetable;

The global CO2 table was a little different in that we wanted to extract the global sum of carbon dioxide emissions for each year. We did this by using the sum() function provided in hive, and the table was created with the query below:

create table co2data2 as select co2data.year, co2data.emission from (select year, sum(emission) as emission from co2data group by year)co2data;

After creating the two tables that we needed for Analysis, we proceeded to combine them. We utilized a join operation to produce our analytics for land temperature vs CO2 emissions and sea-land temperature vs CO2 emissions. The join operation is listed below:

create table combinedData as select avgtemp.year, avgtemp.avgtemperature, co2data2.emission from avgtemp inner join co2data2 on avgtemp.year = co2data2.year;

As an extension, we also ran another hive query to explore each nation's contribution to global carbon dioxide emissions by grouping the table by country. This was given by:

create table co2data3 as select

co2data.country, co2data.emission from (select country, sum(emission) as emission from co2data group by country)co2data;

Finally, with our combined CO2-temperature tables, we wanted to see if there was an observable trend among the hive tables, and we utilized Tableau for visualization purposes as is laid out in the next section.

6. Visual Representations of Analytics

The visualization tool we used was Tableau, which we had connected to Hadoop using SSH port forwarding. With this, we imported our Hive tables onto Tableau to visualize extracted analytics. Our first visualization as seen in Figure 4 represents our initial approach of looking solely at land surface temperatures. We noted how the rise in global carbon dioxide emissions was exponential, given by our carbon dioxide growth percentage; therefore, we logged it to get a linear representation. We then layered this on top of the Land Average temperature graph to see any correlation between the two trends. The important takeaway is that while the earlier years fluctuated massively, after early 1850s, both graphs trend upwards to signify a close relationship. The uncertainty in the data of earlier years may have accounted for the volatility of the land avg temperatures.

In order to counteract the uncertainty, we decided to look at some options to extract a clearer analysis. We ended up looking into two approaches, one of which was taking both land and sea average temperatures, which encapsulates the global scale in a better form than solely land temperatures. In addition, the aforementioned uncertainty of the earlier years was addressed by limiting the visualization to years past 1850, further cutting down on the earlier years of fluctuation. As seen in Figure 5, our result showed a much clearer increase in average temperature, specifically post 1960s. When comparing this to works like that of NOAA in Figure 2, we see that they too saw a dramatic increase in temperature during this time. The steady exponential growth on CO2 emissions and the linear growth of temperature data meant that this graph quantified the relationship between the two variables better than Figure 4.

As a final insight, we wanted to observe a visualization of the greatest contributors to global Carbon Dioxide emissions. Its common to hear about how China is the greatest modern-day contributor to CO2 emissions; however, we thought it was essential to also know the greatest historical emitters. Summing up all of the carbon dioxide emissions since 1750, we see that the United States actually has produced the most carbon dioxide emissions. To quantify these values, the United States were responsible for 14.81 percent of the total global carbon dioxide emissions. China is a close second at about 13.073 percent of the total Carbon footprint. While these numbers may differ by a couple percentage points, we should note the scale of our visualization in Figure 6 is in the millions of tons. So it's quite a drastic representation of the contribution the United States have added to the global carbon footprint.

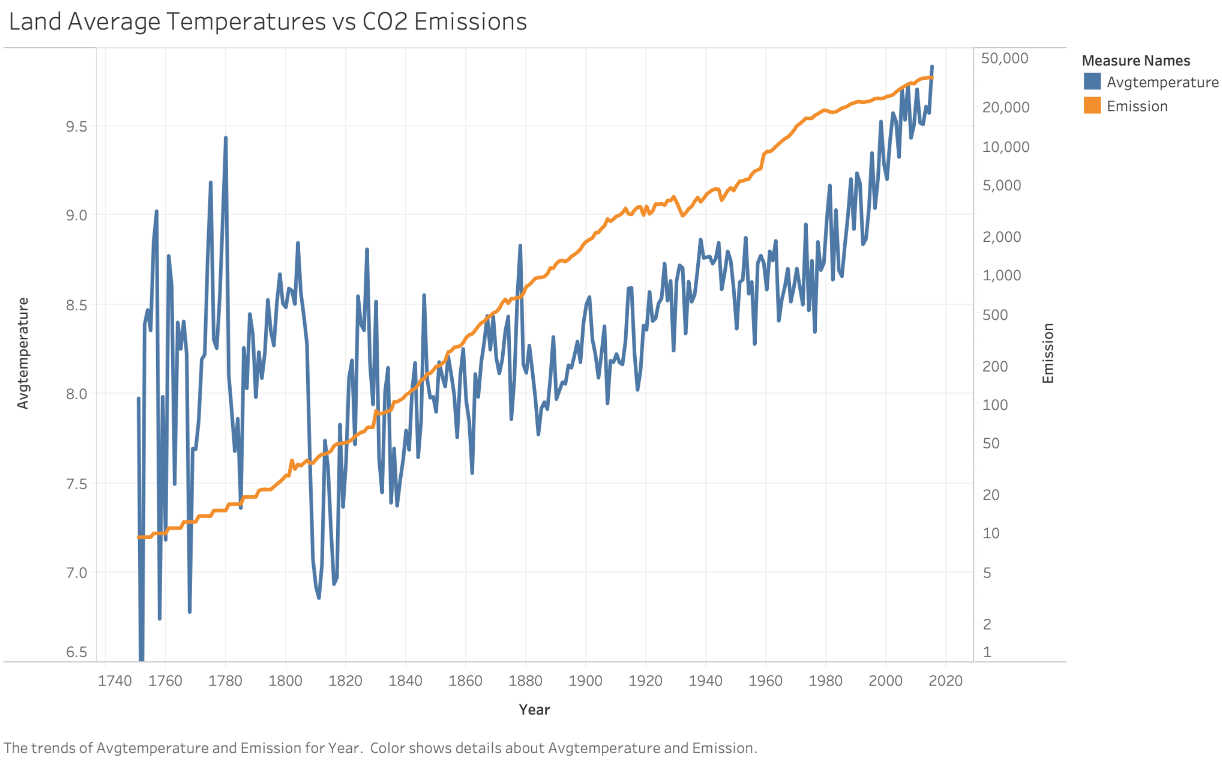

Fig. 4.— A comparison of average land temperature data and total carbon dioxide emissions

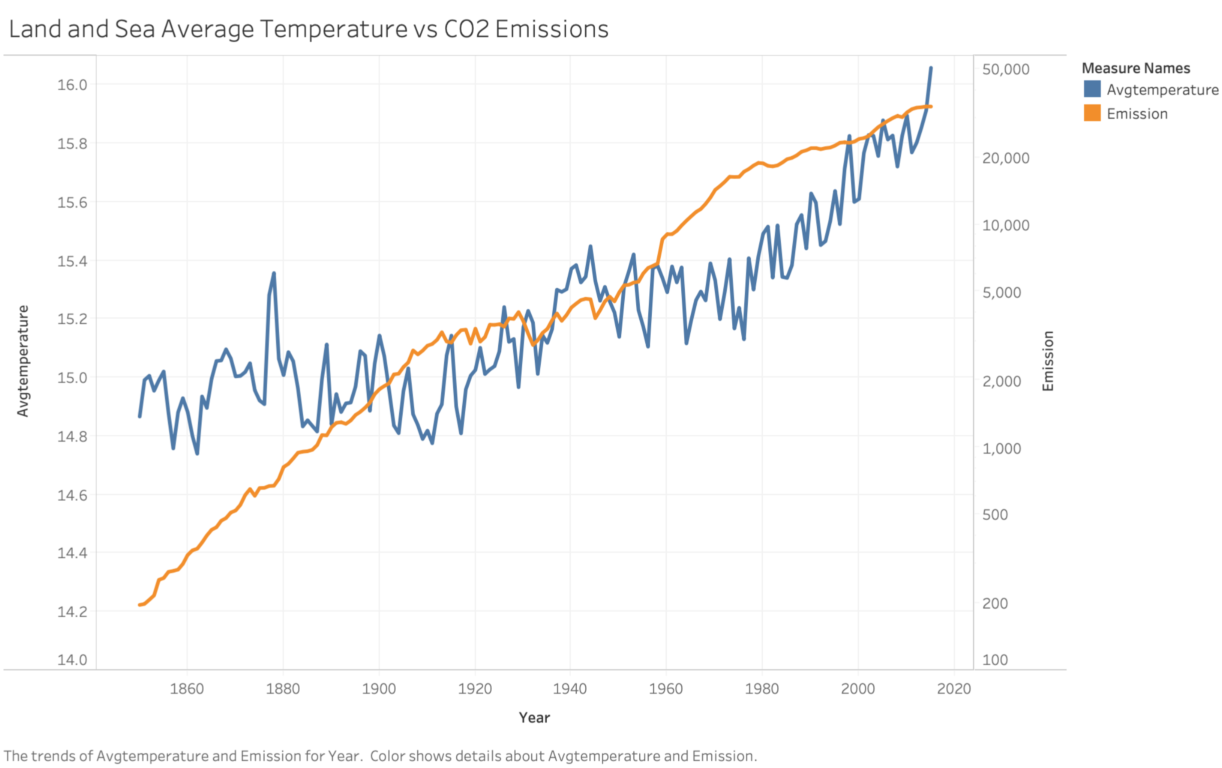

Fig. 5.— A comparison of average land and sea temperature data and total carbon dioxide emissions. Extracted data points prior to 1850 to eliminate uncertain temperature data similar to NOAA's graph.

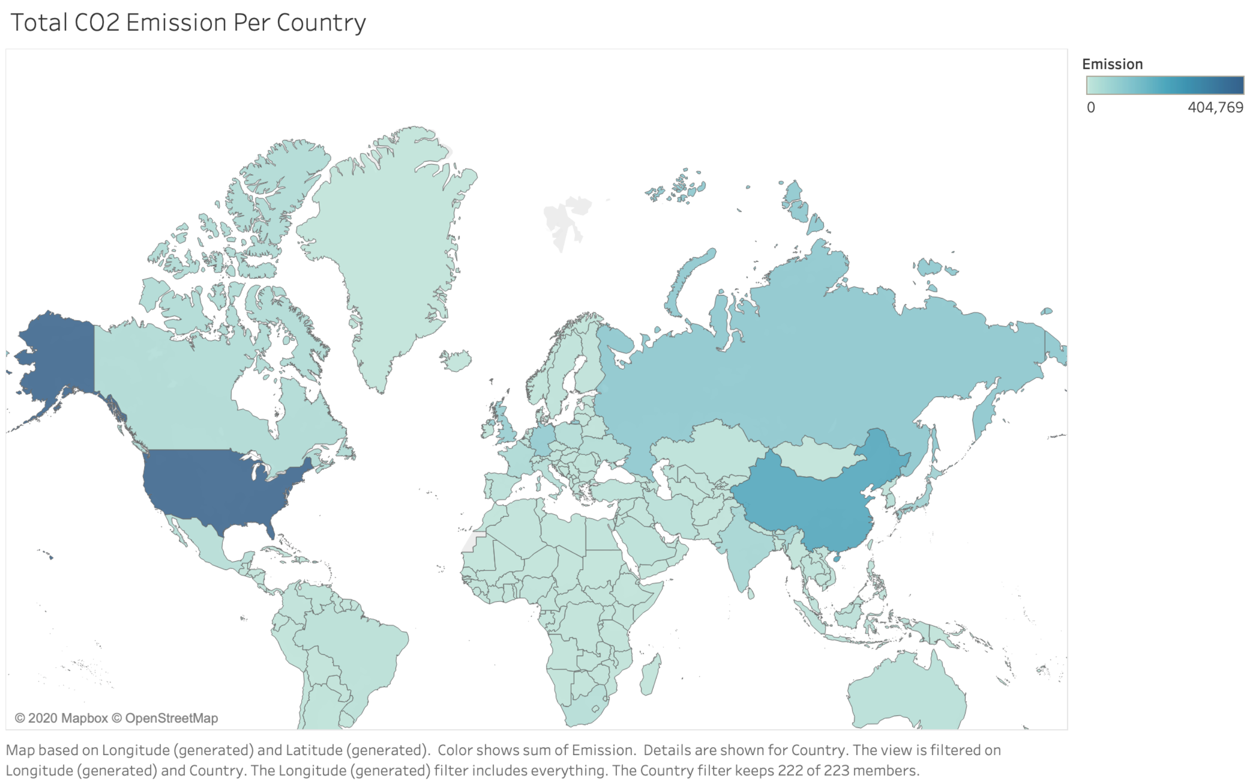

Fig. 6.— A visual representation of each country's' total carbon dioxide emissions. The data is measured in measured in million tonnes per year.

7. Conclusion

While our data fell in general consensus to the published articles that we based our arguments off of, the analytics that we ended up producing showed a more dramatic, exponential increase of CO2 emissions relative to global temperature, which expectedly, had a slower, linear growth. The implication of the analyzed correlation means that as our planet continues to emit carbon dioxide, the planet's temperature will continue to rise. This signals a troublesome trend considering each year's carbon dioxide emissions is exponentially increasing.

While globally we're becoming more and more industrialized, especially post pre-industrial era, this has led to a massive increase in anthropogenic greenhouse gas emissions due to consumption like fossil fuel use. While the overall technological growth of the planet is important, we must also measure the potential risks. In order to slow down or maybe the eventual halting of rising global temperatures, stability in CO2 concentrations and other greenhouse gases in our Earth's atmosphere in necessary. One such step is not only honoring the pledges made by each nation on controlling carbon dioxide emissions, but also furthering those ideals and taking more drastic measures to protect our planet. We want to stress how climate change is not an issue that can be put on the back-burner for our future selves, as it could lead to extremely detrimental conditions to our shared planet.

Our analysis was a mixture of verifying previous research correlating carbon dioxide emissions and global temperatures and looking into details that we weren't able to find readily available information on such as the total contribution to the global carbon footprint by each nation. Because global warming is a commonly researched topic, finding new or important analytics is difficult. However, something that intrigued us during this pandemic was the carbon emission effects of Covid-19 and whether or not this changes anything. Evidently, carbon dioxide emissions will have lowered due to the pandemic, however, with this 9-month period of lowered carbon dioxide, how much will it actually influence global temperature? Obviously, we can't know yet, but given the amount of carbon dioxide we've historically emitted into our atmosphere, it's hard to imagine too much impact on the rising global warming. Nevertheless, we feel this would definitely be an interesting topic to look further into, as it could signify the drastic efforts needed to actually lower global warming.

8. References

Friedlingstein, P., et al. “Persistent Growth of CO2 Emissions and Implications for Reaching Climate Targets.” Nature Geoscience, vol. 7, no. 10, 2014, pp. 709–715., doi:10.1038/ngeo2248.

Hannah Ritchie and Max Roser (2017) - "CO2 and Greenhouse Gas Emissions". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions'

“The Effects of Climate Change.” NASA, NASA, 21 Aug. 2020, climate.nasa.gov/effects/.

Sánchez-Lugo, A., Berrisford, P., Morice, C., and Argüez, A. (2018). Temperature [in State of the Climate in 2018]. Bulletin of the American Meteorological Society, 99(8), S11–S12.

NOAA National Centers for Environmental information, Climate at a Glance: Global Time Series, published November 2020, retrieved on December 11, 2020 from https://www.ncdc.noaa.gov/cag/

Lashof, D., Ahuja, D. Relative contributions of greenhouse gas emissions to global warming. Nature 344, 529–531 (1990). https://doi.org/10.1038/344529a0

“Human and Natural Drivers of Climate Change.” Human and Natural Drivers of Climate Change - AR4 WGI Summary for Policymakers, 2007, archive.ipcc.ch/publication_and_data/ar4/wg1/en/spmsspm-human-and.html.

Cavicchioli, Ricardo, et al. “Scientists’ Warning to Humanity: Microorganisms and Climate Change.” Nature News, Nature Publishing Group, 18 June 2019, www.nature.com/articles/s41579-019-0222-5?fbclid=IwAR3iNffbq2Kd5y7XXMBTmC22g5Jtngmg_yzO38-6IFCI7MVyc06Zai4snl0.