Fall 2022
W200 Intro to Programming
Project 2 Report

Team Members:
- Victor Yin
- Pascual Eley
- Erik Sambrailo

Overview

For our project we have chosen to explore datasets available from Berkeley Earth. The datasets consist of time series temperature data for various regions. We have chosen to focus specifically on countries for regions. From our initial review, it appeared Berkeley Earth contained data for 237 countries with potentially 192 years of data. *(The term country will be used loosely as some regions are not technically recognized as countries.)*

We also wanted to explore country greenhouse gas emissions for comparison. As a supplemental dataset, we included UN data on various countries' greenhouse gas emissions. For comparing countries by geographic location, we additionally pulled in a dataset from Kaggle with average longitude and latitude for countries.

Below are the questions that we are interested in answering with these datasets:

Questions

- How have temperatures changed globally over the past couple centuries?
    - How does this differ by country?
    - Does a country's proximity to the equator have any effect on its temperature changes?
- How have greenhouse gas emissions changed over time?
    - How does this differ by country?
    - How do the various greenhouse gas emissions compare to each other?
    - Does a country's emissions have any correlation to its change in temperature?

Compiling Datasets

When reviewing the Berkeley Earth site we found a directory of region specific .txt files that contained temperature data. We built a scraper that iterated through the list of countries we wanted and downloaded their corresponding file. We were successfully able to download 233 countries from our list of 237 using this method.

After reviewing a couple of the country .txt files, we found that the files contained a combination of summary text and data points, but also that the formatting of the files was consistent. With that, we were able to iterate through each of the .txt files, pull out the relevant data, and convert those to dataframes. *(At first we developed a dictionary of "country: dataframe" pairs, but found later in our exploration that it was easier to navigate one consolidated dataframe.)*

Initial Review and Cleaning of Data

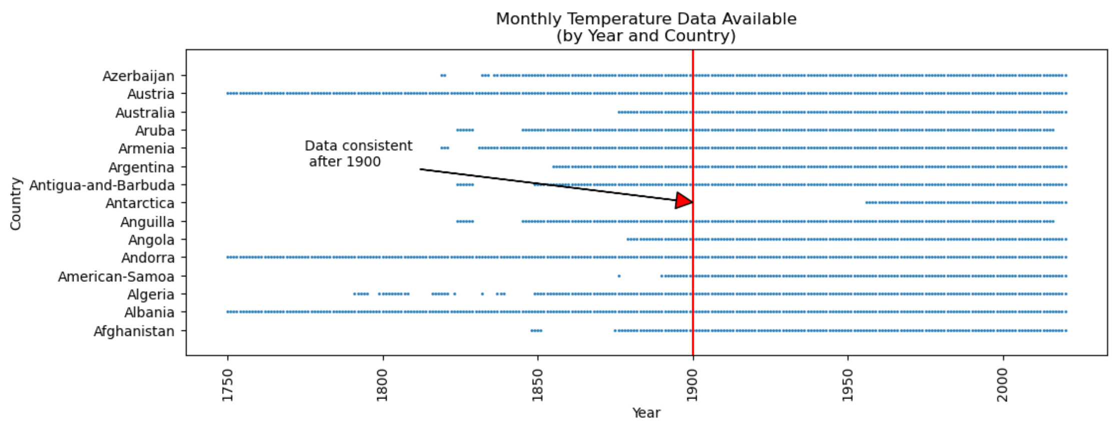
After compiling, we reviewed the data that we had available. Each country in the Berkeley Earth dataset was anchored to a baseline average temperature that was derived from its temperatures from Jan 1951-Dec 1980. This baseline was used to compare to time series temperatures, which were represented as average anomalies of centered moving averages with periods of month, annual, 5 year, 10 year & 20 year. These were also accompanied with uncertainty values.

We chose to make a couple assumptions for our analysis. First, we assumed that the baseline temperatures for each country are accurate representations and provide a sound basis for which to compare temperature anomalies between countries. Second, we are choosing to assume that the anomaly temperatures are accurate and with that we will not be reviewing uncertainty values.

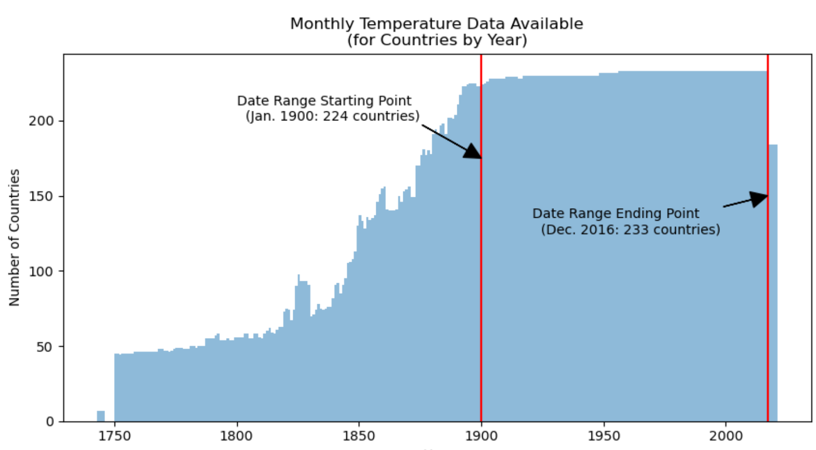
We also found various naming conventions for countries across our datasets. We had to ensure that country names were consistent across our emissions data, country geospatial data, and emissions data. We noticed that when we iterated through files to create the countries, some had dashes while others had spaces, and some used various abbreviations. We had to make these consistent as we compared the various data frames we were using. We automated alignment where possible, but manually corrected parts like country names from our emissions data to match since it was a much smaller list of countries.

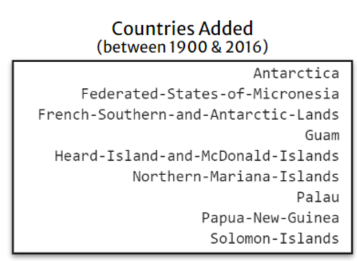
Temperature Data Availability for each Country by Year

As part of our EDA we reviewed what temperature data was available for each country by year. For this we focused on 'Monthly_anomaly'. Per the below snapshot (full chart can be found in code) we discovered that while several countries had date ranges starting as early as the 1700's, the availability of the data was sporadic up until around the year 1900.



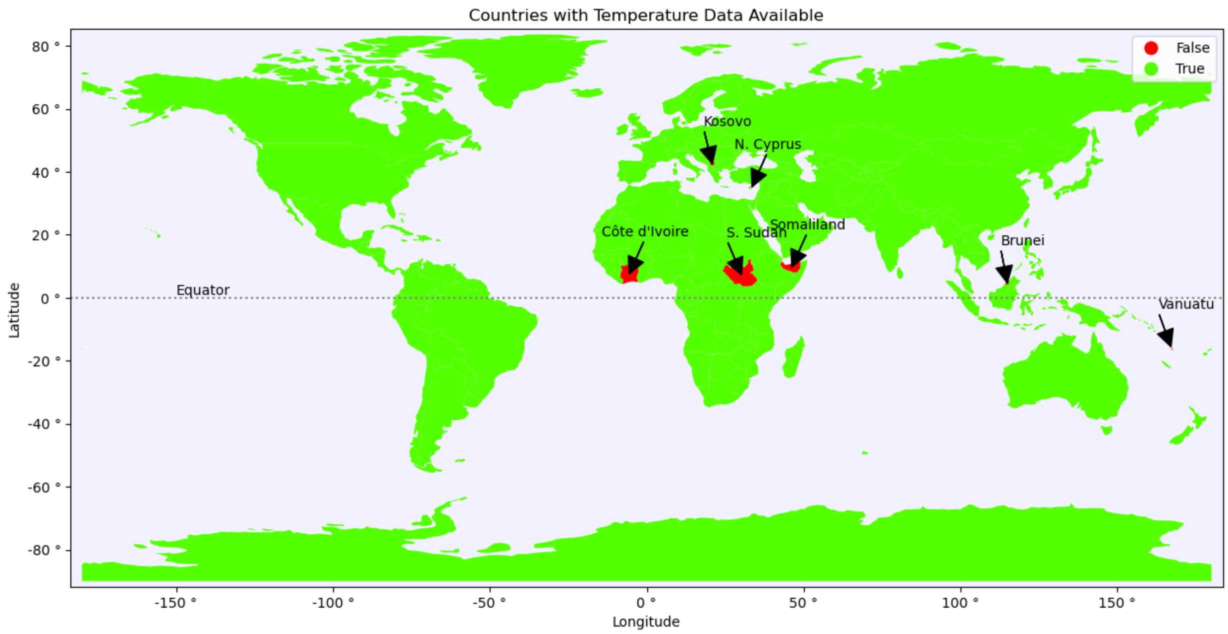
To better understand the available dates of all countries, the below plot was created.
From this we confirmed our theory about the temperature data before 1900, and also found that several countries did not have data beyond 2016. For this reason we chose to select a date range of 1900 - 2016 for our analysis of the Berkeley Earth data.





Review of Countries Geospatial Mapping Availability

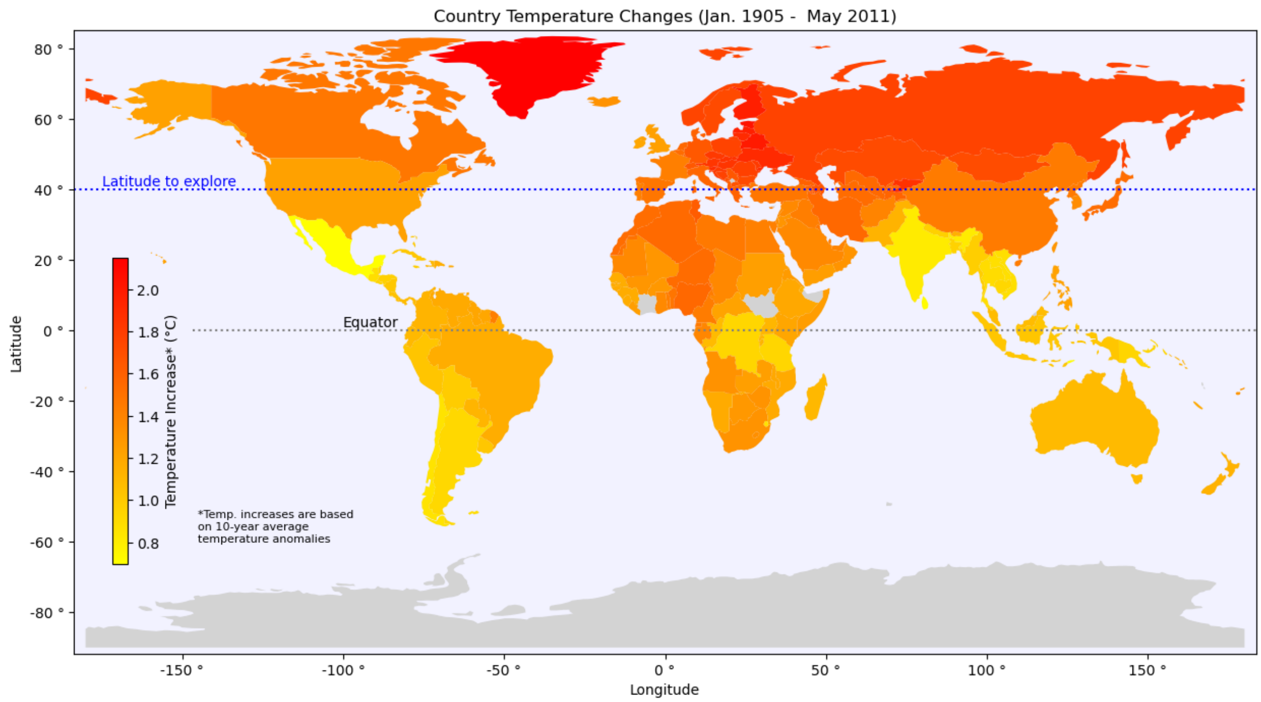
As we intend to explore geospatial mapping as part of our analysis we wanted to explore what countries are currently available for mapping. We downloaded the GeoPandas dataset 'naturalearth_lowres' which consists of 177 countries. After cross-referencing our temperature datasets, we found that we have temperature data for all but seven countries. Below shows the countries missing:



We also found that we have 63 countries for which we have temperature data that are not referenced geographically in this dataset. So any geospatial visualizations would not include those countries. For our analysis, we felt that the 170 countries available for mapping would be sufficient to glean insights and are not concerned about the absence of the additional 63 countries.

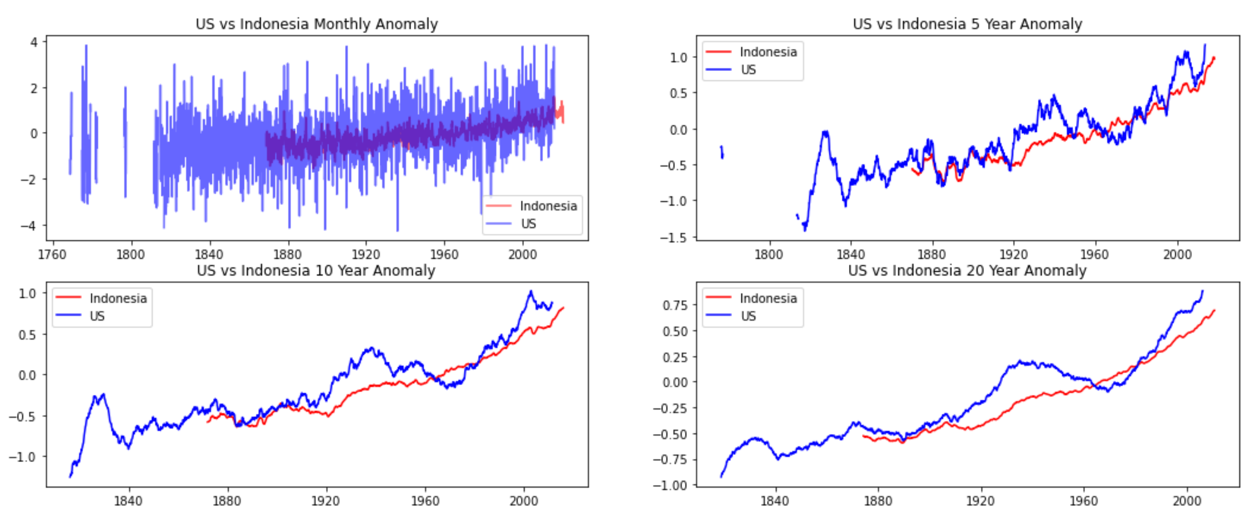
Temperature Differences Over Time

We wanted to review what temperature increases countries experienced over time and from a geographical perspective. We chose to look at the 10-year average temperature data and found that it would be best to limit our time range to Jan 1905 through May 2011. The below plot is showing the temperature increases experienced in that time frame. (Gray countries did not have enough data available for this analysis.)



Comparing Temperature Anomaly: Month vs 5 Year vs 10 Year vs 20 Year

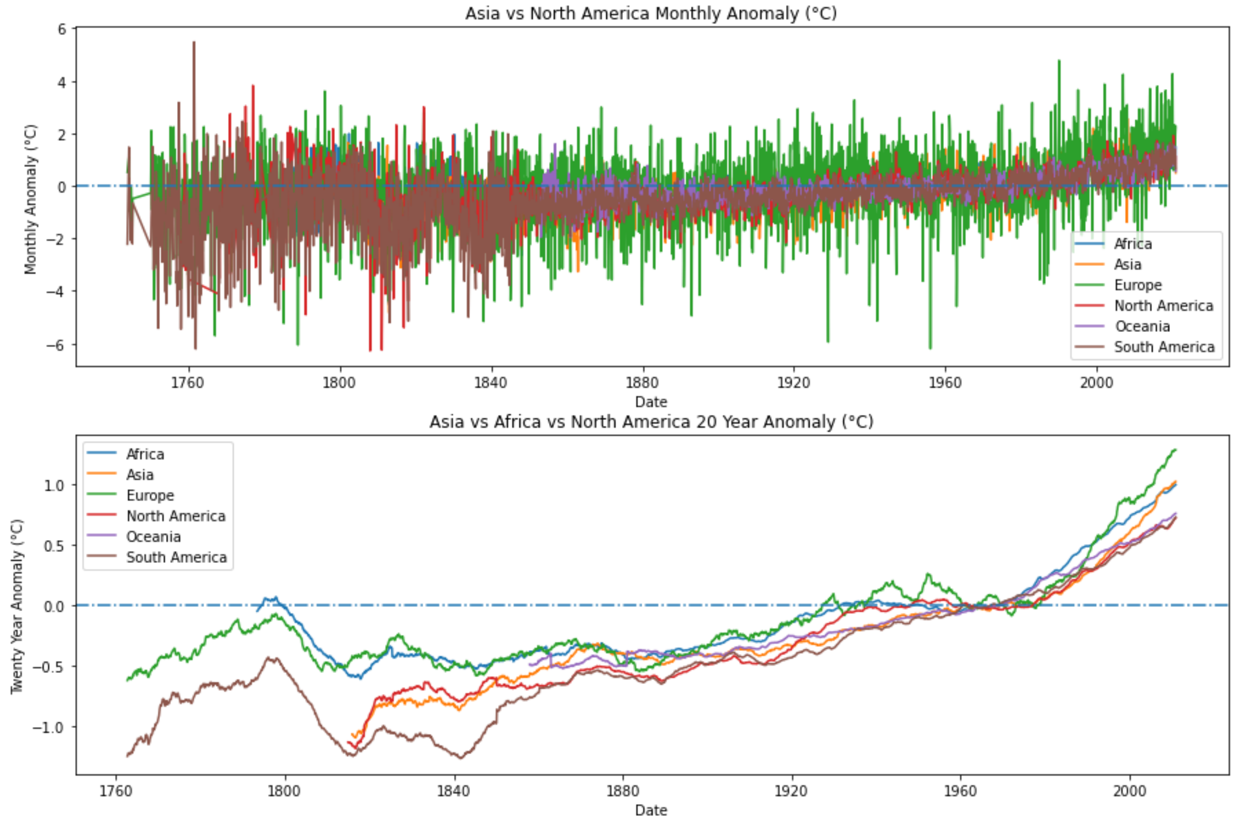
- We used a comparison to two countries (Indonesia and US) at monthly, 5 year, 10 year and 20 year anomaly scales to see what the data would look like. The variability of month-to-month anomaly makes comparing this data between countries extremely difficult. However, the 10 year and 20 year anomaly leads to a more even line that better shows a trend over time to compare countries.
- In the case of these two specific countries, a major difference doesn't seem to jump out when looking at all the data in 20-year anomaly.



Temperature Anomaly by Region:

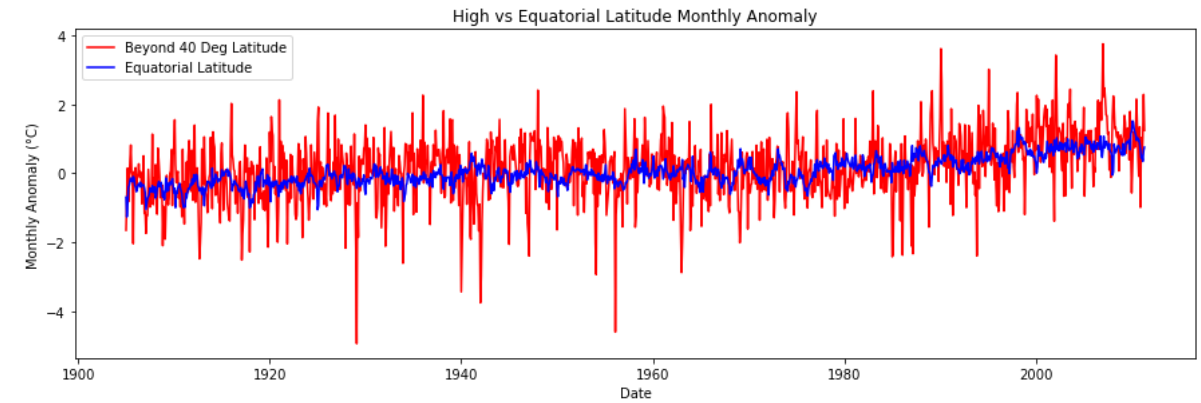
- Looked at anomaly averages broken down by region (Africa, Asia, Europe, North America, Oceania, South America).

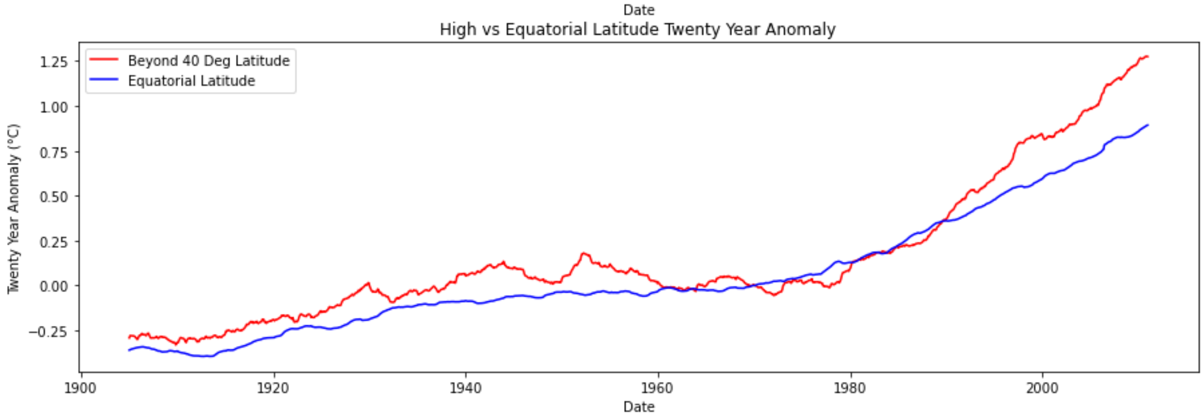
- It seems as though the monthly anomaly is more volatile for Europe than any of the other regions.
- 20 Year Anomaly seems to follow a similar trend regardless of region, but the anomaly average in recent years is higher for Europe than other regions.



High vs Equatorial Latitude Anomalies:

- Based on Europe having a more volatile anomaly and a higher recent trend, we wanted to see if countries at higher (or lower) latitudes had different average temperature anomaly patterns.
- For this analysis we used the same date range as geodata earlier in our analysis (1905-2011).
- We split the world at +/- 40 deg Latitude. This left 58 Countries with Latitude above/below 40 degrees. 152 countries between 40 and -40 (what we refer to as 'equatorial' in this analysis, although they would not necessarily be very close to the equator). (Note, there are only 3 countries with average latitudes more southerly than -40, so the data is dominated by northern countries north of 40)
- For countries with an average latitude over 40 degrees or under -40 degrees, the average monthly anomaly seems to have more variability.
- We <u>even</u> see more variability in the 20 year anomaly, with high latitude countries seeming to have a sustained higher anomaly in recent years.

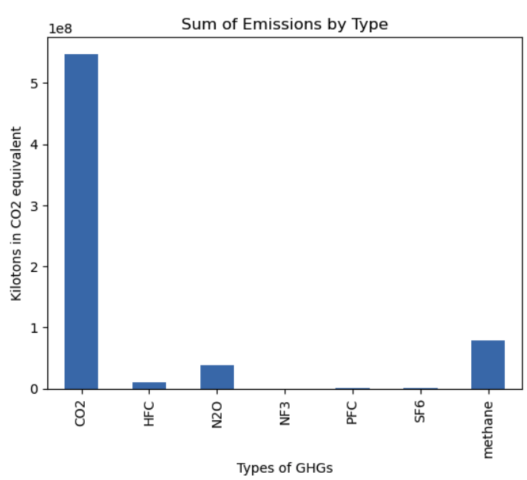




Greenhouse gas emissions

GHGs are atmospheric gasses responsible for causing global warming and climate change. The major GHGs are carbon dioxide (CO2), methane (CH4) and nitrous oxide (N20). Less prevalent --but very powerful -- greenhouse gasses are hydrofluorocarbons (HFCs), perfluorocarbons (PFCs) and sulfur hexafluoride (SF6).

For our EDA, we reviewed the UN's database of emissions for its member countries for the last 30 years. The data is classified into a countries' total emissions broken down by type for each year from 1990 to 2019. Carbon and methane emissions represent over 92% of all greenhouse gas emissions produced by the 43 countries or areas represented in the data.

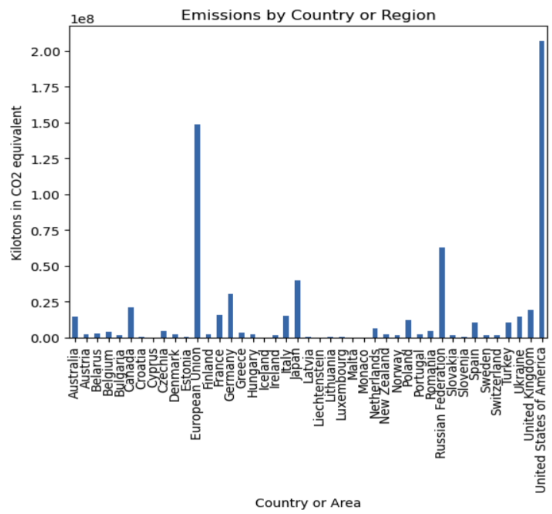



```
Type
CO2        547825489.32
HFC          9644833.36
N20         38760819.43
NF3            36928.70
PFC          1627874.26
SF6          1309967.33
methane     78257252.53
```

Approximately 500 billion kilotons of CO2 were released into the atmosphere from the UN nations over this period of time, vastly more than any other greenhouse gas. All of these gasses remain in the atmosphere long enough to become well mixed, meaning that the amount that is measured in the atmosphere is roughly the same all over the world, regardless of the source of the emissions. Although certain gasses may be more potent than others, the sheer concentration and proliferation of CO2 in the atmosphere gives it the largest effect on atmospheric/temperature change. Thus, our EDA delves into the carbon emissions of various countries.

Total emissions by Country

To gain a better understanding of the emissions, we also looked at the countries who emitted the most during this period of time. We can see from the table below that the USA, European Union, Russia, and Japan produced the most emissions out of these countries.

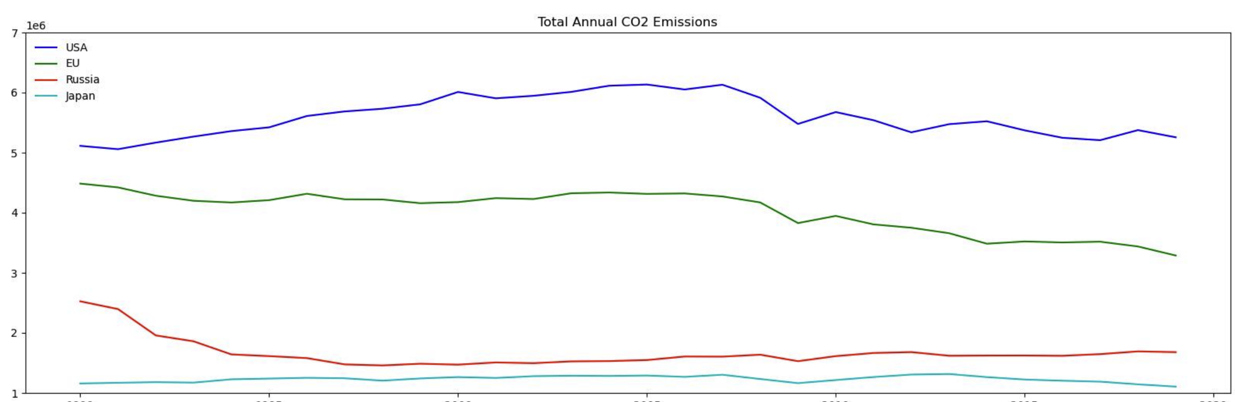


Type	Country or Area	
CO2	United States of America	167927361.97
	European Union	120823499.60
	Russian Federation	49909669.59
	Japan	36938568.58
	Germany	26251704.63
methane	United States of America	21128317.05
	European Union	16550898.53
CO2	Canada	16332753.49
	United Kingdom	15751101.26
N20	United States of America	13703510.56
CO2	Italy	13001643.35

Carbon Dioxide

Carbon Dioxide is the most prevalent greenhouse gas, accounting for 80% of all emissions from the included UN countries in the last 30 years. As seen in the below table, the top emissions are CO2 emissions from the US, European Union, Russian Federation, and Japan

From the largest emitters, we wanted to graph how their carbon emissions changed over this period of time.



Below provides a summary of the changes in carbon emissions for each country along with research corroborating the changes:

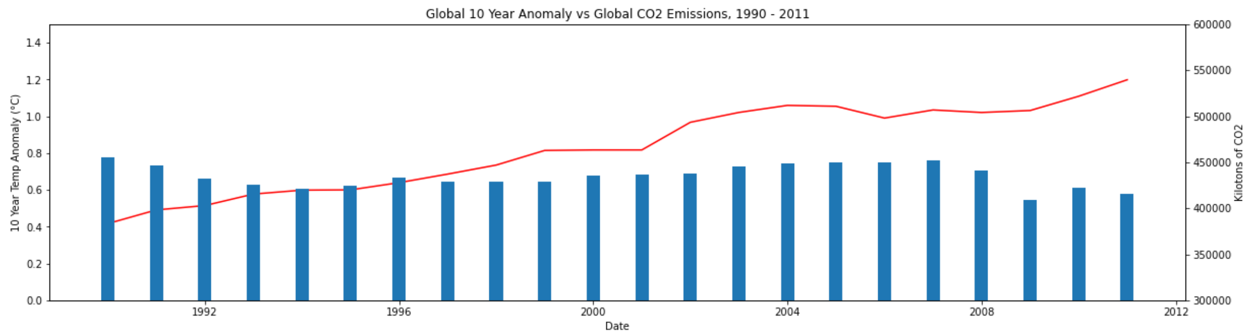
- US carbon emissions constantly increased and peaked in 2005 slowly decreasing to 1990 levels through 2019
    - The rising levels of carbon emissions from 1990 to 2005 can be attributed to a growing economy and increase in fossil fuels. The decrease post-2005 can be attributed to the financial crisis decreasing consumption and a shift to alternative fuels, such as natural gas.[1]
- EU annual carbon emissions showed roughly a 22% decrease since 1990
    - Between 1990 and 2020, the EU reduced its GHG emissions in seven out of eight sectors.
    - Energy industries cut the most (nearly 50%), followed by manufacturing and construction. In third place were households, commerce, and institutions, then industrial processes. Agriculture came in fifth place, although in recent years the rate of reduction has slowed. Sixth was fuel with waste management in seventh.[2]
- Russian carbon emissions showed a sharp 40% decrease from 1990 to roughly 2000s, remaining stable through 2019.
    - The initial sharp drop can be attributed to the fall of the Soviet Union in 1991, where the restructuring of the agricultural sector from state-sponsored to market economy led to significant decreases in meat production. The empty farmlands became a carbon sink that absorbed much of the atmospheric carbon as well. [3]
- Japan carbon emissions stayed relatively constant with few spikes, finally showing decreases in the years leading up through 2019.
    - Japan as a major developed country has very few fossil fuels and minerals and has historically relied on nuclear power as a source of energy along with coal. In line with their Kyoto Protocol and the Paris Agreement, Japan has agreed to reduce its already low GHG emission by 2030.[4]

---

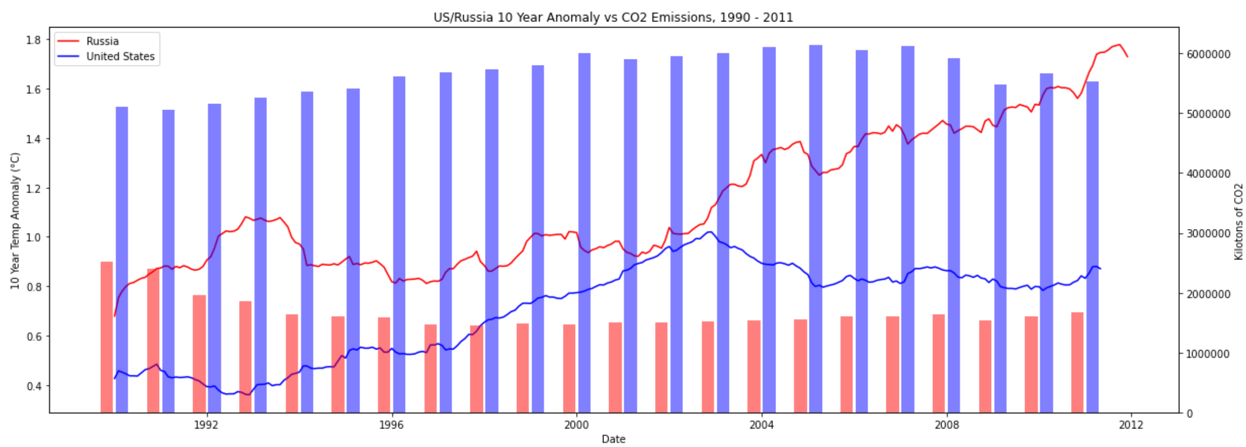
[1] US Carbon Emissions
[2] EU Carbon Emissions
[3] Russian Carbon Emissions
[4] Japan Emissions

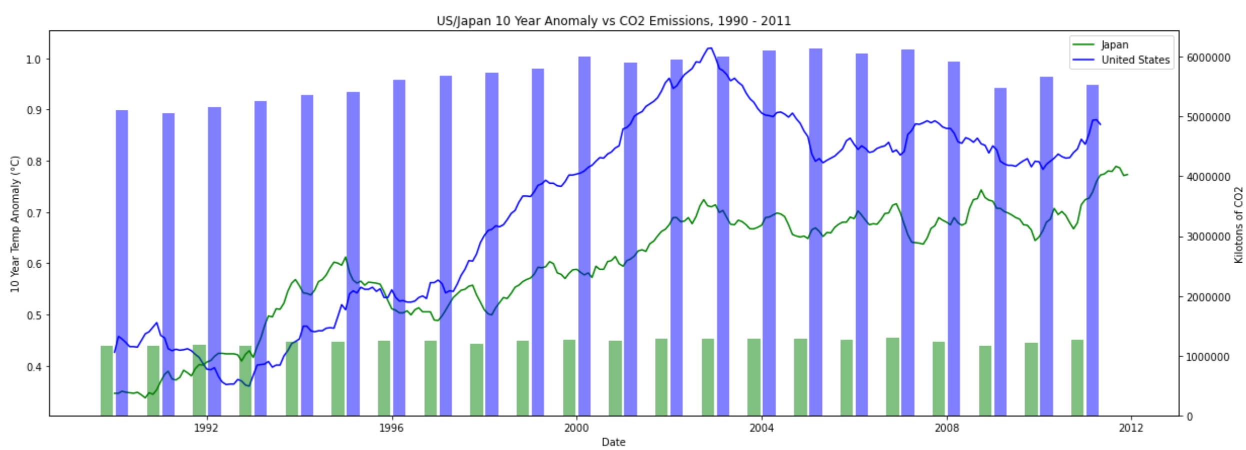
Emissions vs Temperature Anomaly:

Using a snapshot of 1990 to 2011, we evaluated the global 10 year anomaly temperature change against the global CO2 emissions of the UN nations. We observed that although emissions seem to be decreasing in this period of time, the anomaly continues to increase.



**This aggregate data only includes countries that have emissions AND anomaly data for (41 Total): ['Spain', 'Belgium', 'Australia', 'Croatia', 'Slovenia', 'Liechtenstein', 'Hungary', 'United-Kingdom', 'Austria', 'Ireland', 'Russia', 'United-States', 'Germany', 'Japan', 'Luxembourg', 'Canada', 'Latvia', 'Cyprus', 'Norway', 'Malta', 'Estonia', 'Greece', 'Ukraine', 'Denmark', 'New-Zealand', 'Bulgaria', 'Romania', 'Belarus', 'Poland', 'France', 'Iceland', 'Netherlands', 'Turkey', 'Switzerland', 'Sweden', 'Monaco', 'Finland', 'Italy', 'Lithuania', 'Slovakia', 'Portugal']*





The graphs above show the difference in carbon emissions between the US vs Japan and US vs Russia. The US and Japan show a similar temperature increase that tapers off in the after 2007. Additionally, the US shows a gradual increase in emissions until 2007 while Japan stays fairly constant.

However, when we look at the chart with Russia and the US, we see a sudden drop in emissions, but the temperature anomaly only seems to be increasing further.

Conclusion

The results lead us to suspect that there isn't a strong correlation between the amount of emissions and temperature change in a given country or region; rather, the geographic location in terms of latitude and longitude has a greater bearing on temperature change. As mentioned prior, all emissions will proliferate in Earth's atmosphere evenly, so certain regions are bound to face more adverse effects relative to their greenhouse gas emissions than others. While we haven't done any rigorous statistical analysis of the data, our explorations have pointed us to the fact that every country is part of the same planet that's getting warmer. Any one country cannot save itself through its own actions, we all have to work together to prevent climate change.

Appendix

Dataset References:

http://berkeleyearth.org/data/

http://data.un.org/Default.aspx

https://www.kaggle.com/datasets/paultimothymooney/latitude-and-longitude-for-every-country-and-state