CSC110 Project: Relationship between Climate Change, Pollution, and Productivity of a Nation

Jason Sastra

Friday, December 11, 2020

Problem Description and Research Question

In today's age of global warming, it is widely believed that one of the factors that heavily affect global warming is pollution, specifically carbon dioxide emission. In my project, I will try to explore the relationship between carbon emission, climate change, and productivity of a nation. I would like to take in multiple data sets including the GDP of a nation, the amount of CO_2 emission, along with the change in temperature in order to attempt to compute predictions of climate change through changes in pollution and productivity. I would do this by finding the relationships between these different data sets and compute another graph which are predicted from the appropriate graph from these data sets. My final goal would be to compare my predictions with recent data sets in order to compare it to real life scenarios. I would like to create a linear regression and predict the change in global warming based on CO_2 commission and country GDP and see if it holds true. I will do this on plot.ly, allowing me to create graphs of the predicted increase in average temperature based on average increase in CO_2. I would also like to know whether or not the weather of the country shows how hard it is affected by climate change. Does climate change hurt areas with higher temperature or lower temperature? **What is the relationship between a countries productivity, pollution levels, and climate change and how could this relationship be used to limit climate change.**

Dataset Description

My first data set is a data set in csv format from the UN which describes the amount of CO_2 emission produced for several countries in the year of 1990 to 2017. It describes the amount of CO_2 emission in kilotonne CO_2. From this dataset, I can gather the yearly amount of CO_2 emission in several countries, which I would later use in order to graph these emissions, and also compare the amount of CO_2 emission from one country to another. This dataset is from this source, "http://data.un.org/Data.aspx?d=GHGf=seriesID%3aCO2" and it is named "Carbon dioxide (CO_2) Emissions without Land Use, Land-Use Change and Forestry (LULUCF), in kilotonne CO_2 equivalent".

The second data set is a data set in csv format from kaggle, provided by Berkeley Earth which provides the average temperature of multiple countries from the year 1743 until 2013. In this data set, I use the average temperature and date in order to compute the change in average temperature per year and also to graph the average temperature with respect to date. This dataset is from this source, "https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data?select=GlobalLandTemperaturesByCountry.csv" and it is named, "GlobalLandTemperaturesByCountry.csv".

The third data set is also a data set in csv format from kaggle, provided by Berkeley Earth. This time, dataset provides information on the global average temperature. It does not give information on it by country but simply the whole world's temperature. It provides temperature from the year 1750 to the year 2015 and is divided into temperatures based on the month. I used this data set to compute graphs of the average global temperature with respect to the date. This dataset is from this source, "https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data?select=GlobalTemperatures.csv", and it is named, "GlobalTemperatures.csv".

The fourth data set is a data set in csv format of the GDP ranking of all countries in the year 2019. This data set comes from the world bank. It shows the ranking of a country's GDP in comparison to countries and it also gives its GDP amount based on millions of USD. I use this dataset in order to compare the average change in temperature and

the change in CO2 for country's along with where it lies in the GDP ranking. For example, USA has an extremely high GDP and high CO2 emission, what exactly is its change in temperature. This dataset is from this source, "https://datacatalog.worldbank.org/dataset/gdp-ranking" and it is named, "GDP.csv"

The fifth data set is a data set in csv format of the change in greenhouse gas from 1983 to 2008. This dataset comes from kaggle and it comes from ESRL/NOAA Physical Sciences Division. From this dataset, I would measure the global change in CO2 and compare it to the global change in temperature. By comparing it, I would like to create a linear regression of the change in temperature versus the change in CO2. I will only use the CO2 part of it. The CO2 measurement is measured in parts per million by volume. This data set comes from the link, "https://www.kaggle.com/econdata/climate-change?select=climate%5C_change.csv" and it is titled "climate_change.csv"

Computational Plan

First Step: Transforming The Data Sets
The first thing I did was to transform all the csv files into dataclasses that are more convenient to handle. I do this by creating dataclasses which I filled with the data from the data sets. An example of this is the dataclass Country-Temperature. This dataclass three instance attributes, the country, the temperature, and the date. I then converted the data from GlobalLandTemperaturesByCountry.csv by running the function "convert_global_temperatures()". By doing so, I created a list of CountryTemperatures which contains all the information from the data set. I run this in the beginning of my file so that I could conveniently use the data in the dataclass instead of opening up the csv file every single time. I do this to all 4 of the dataset, creating the variables, country_data, global_data, co2_data, and gdp_data.
One thing that I noticed while running my functions was that the United States had two different country name. in gdp_data, it is 'United States of America' while in everything else, it is 'United States'. Just for this data, I mutated the name in co2_data because I wanted to have data on the United States. Finally, I had two function which allowed me to identify the countries in country_data and the countries in all the datasets. The function "list_of_countries_in_all()" helped me identify which countries I could compare to each other. The function returns a set of countries that are in all of the datasets. Also, since the data from pollution is yearly data, I also decided to create a helper function in order to convert the monthly data in country_temperature and global_temperature into yearly data. I named this function, "create_country_yearly_data()" In order to better compare the datasets. This function first filters out the country only data with a comprehension. Then, it averages all the temperatures in a given year by appending the temperatures into a list and using statistics.mean into it. In order to find out the year, I use the function date.strftime("%Y"). With this, data from the csv files are now sorted and I would start to use it to make various models.

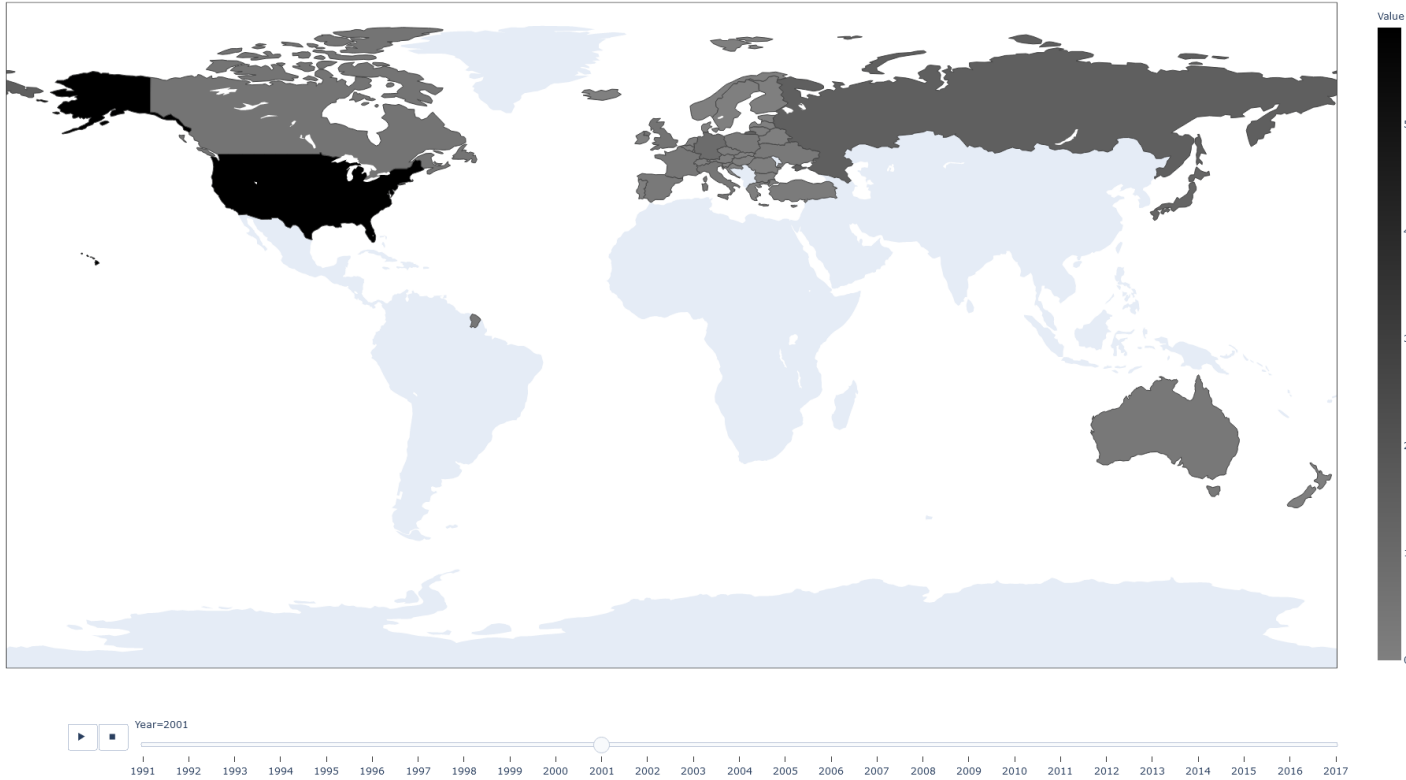**Second Step: Rough Observations based on Various Graphs**
After creating all the datasets, I decided to do a dot plot of how high each countries change of temperature is in relation to its GDP and its change in pollution in comparison to other countries within the dataset. I was curious on whether or not the highest GDP or being the highest in producing pollution means that it would have the highest change of temperature. In order to do this, I ranked the countries that are available for comparison from the set of countries in all the datasets. I ranked them based on who has the highest change in temperature, the highest change in pollution, and the highest GDP in 2019. The ranking was done through this process. First, I calculated the average increase in temperature with the function "average_yearly_increase_temperature()". Then, I ranked it with the function, "temperature_increase_ranking()" I ranked it by simply sorting the data from the average increase and ranking it based on descending order, in which the highest change in temperature is given rank 1. Similarly, the CO2 emission data was also ranked with its respective function. Then, I graphed it on a dot plot. This dot plot is the graph seen below. This graph can be generated through the function, "draw_ranking_comparison(list_of_countries_in_all, 1990)".

Difference in Ranking for GDP, temperature, and CO2 emission from the year 1990
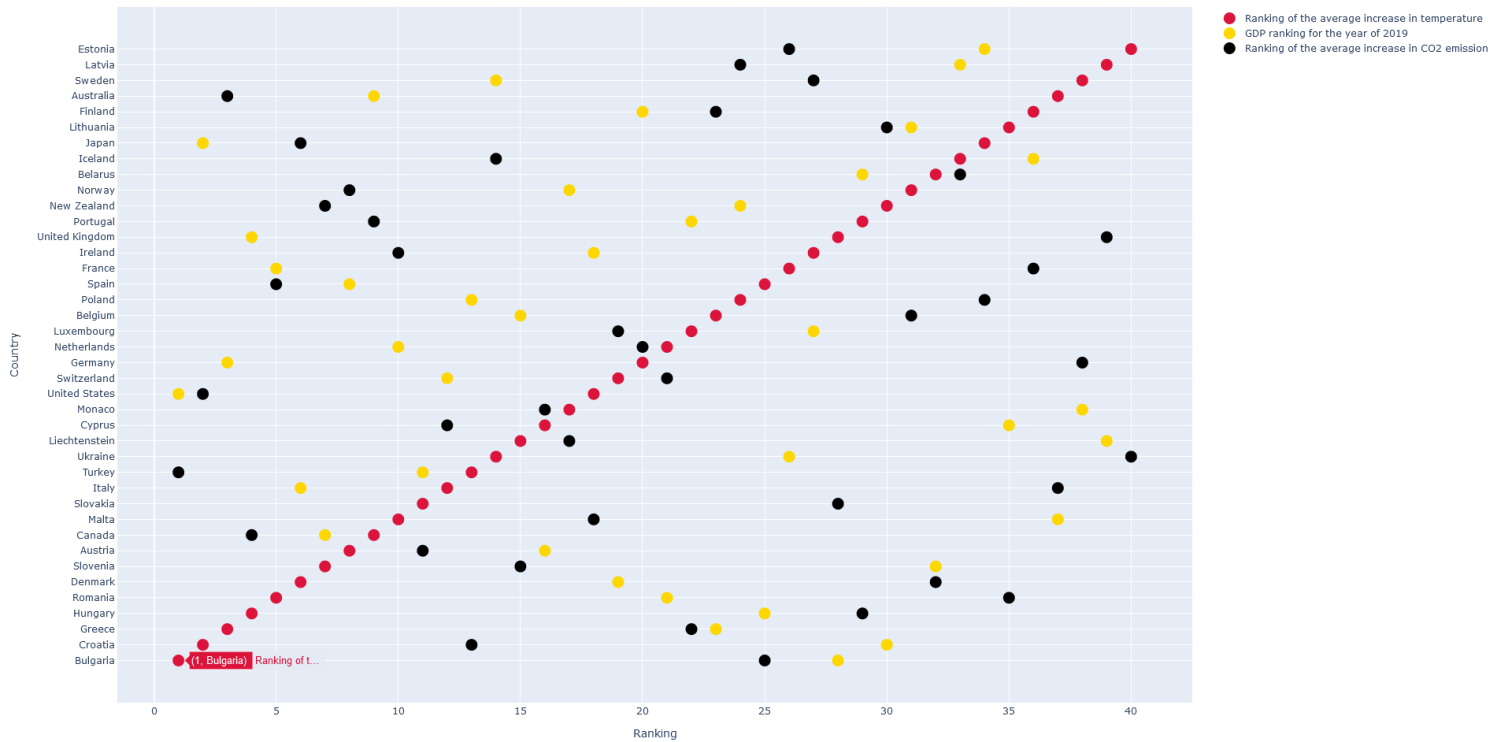
Whats interesting about this rough dot plot is that countries with high gdp and high pollution in comparison to the other countries in the data does not exactly have the highest increase in temperature. For example, the United States is number 1 for GDP, and number 2 for increase in pollution but it is only 18th in terms of the increase in temperature. On the other hand, Bulgaria, the number 1 in increase of temperature does not have the highest GDP or increase in CO2 emission. Therefore, nothing much could be concluded from the eye test from this graph. We cannot rashly conclude any conclusion yet, but this gives an interesting starting point to start with.

Another interesting graph to look at is the the change in the world itself as global warming is occuring. To do this, I created a Choropeth graph in order to see the change in temperature and the change in CO2 emission throughout a certain time period. I ran the function, "draw_changing_co2_graph(1990, 2017)", and also, "draw_changing_global_graph_temperature('1900-01-01', '2013-01-01')". In order to produce two interactive graphs, one for temperature and one for CO2 emission. One of the graph is the graph seen below

CO2 Emission: Change in Emission Throughout the Years

Year=2001

1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017

In this graph, one interesting observation is the degree of pollution that the United States. Yet these graphs do not give much analysis but is an interesting starting point to start with, therefore, we will begin to make more in-depth graph to make deeper observations regarding global warming

Third Step: Relationship between Global Pollution and Global Temperature
I will begin by graphing the yearly average global temperature using the function "draw_global_graph_year()". I do this by first averaging the temperatures in the original dataset since it is initially temperature per month. By adding up all the temperature in the single year and using statistics.mean on it I am able to find the average temperature in terms of years. The graph is the graph that can be seen below.
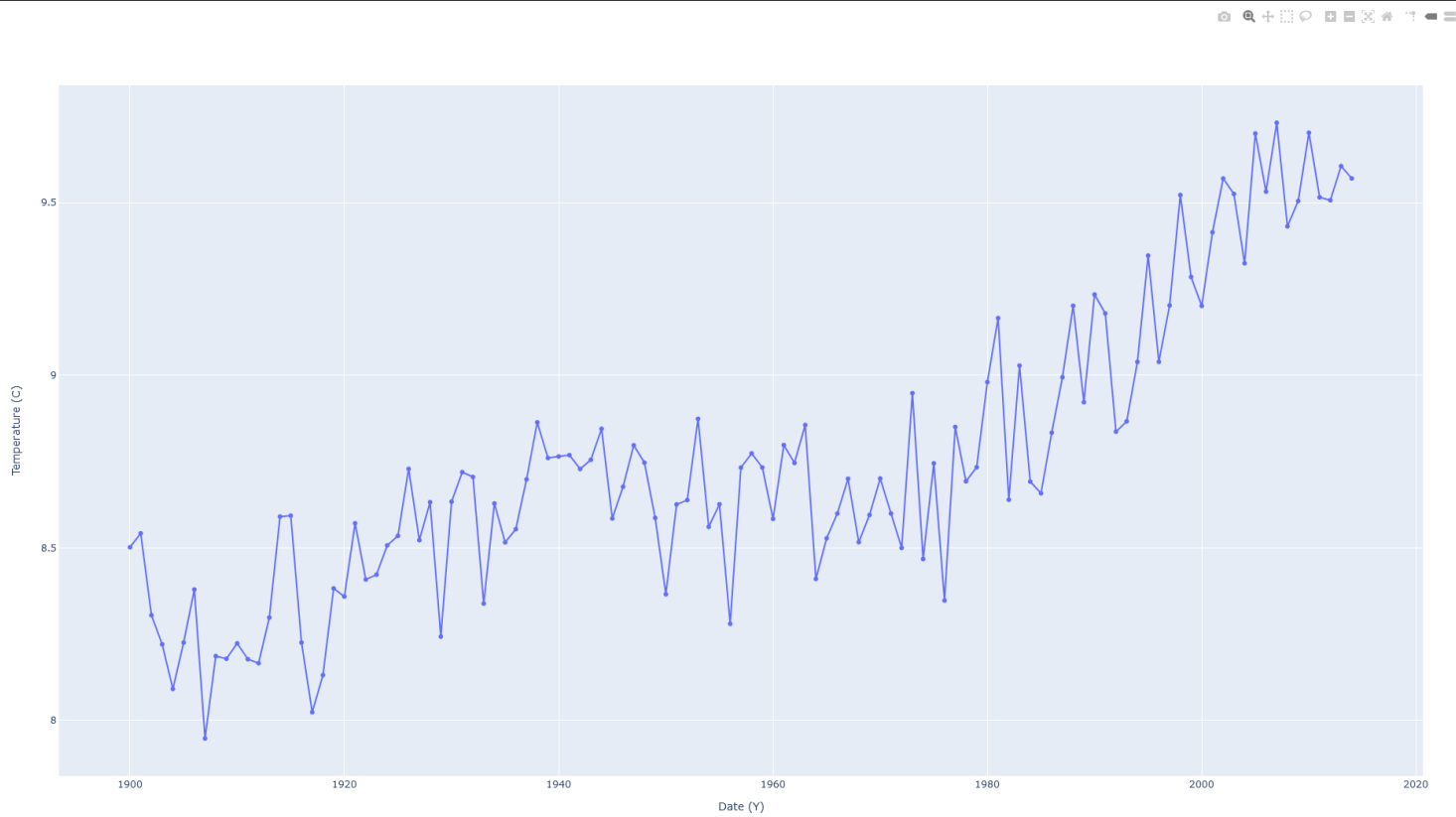
Next I would graph the global CO2 emission levels in terms of ppmv. Similarly, I also average the value so that it will be yearly instead of in monthly intervals. By having the same intervals, it will be easier to work with. The graph will be produced through the function, "draw_global_graph_year()" and will produce the graph below
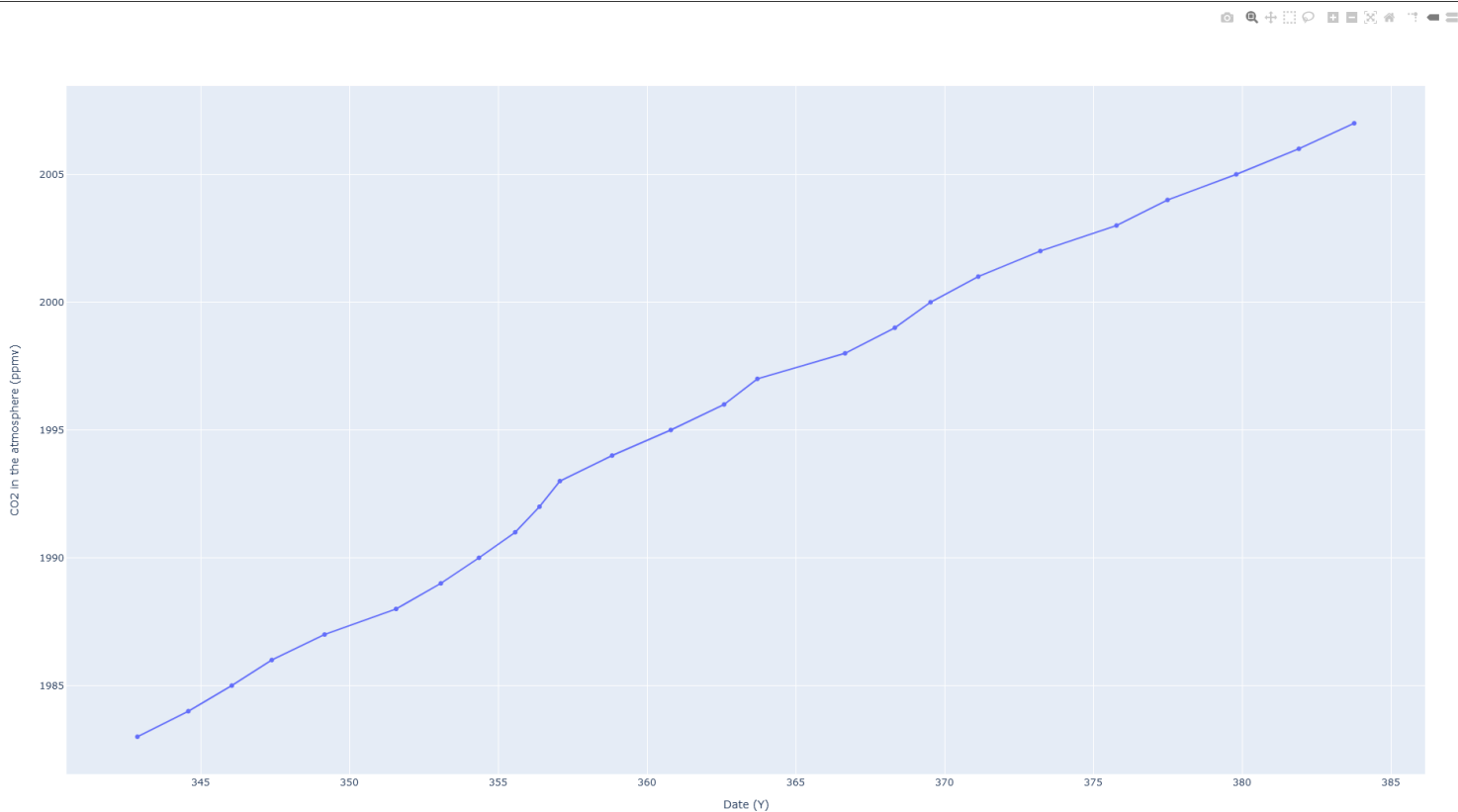
Interestingly, these two graphs both have positive correlation as the year increase, notably at years above 1976, global temperatures would start to increase heavily, which is coincidentally when CO2 in the air also starts increasing, at least in this dataset. Just looking at this graph would lead people to believe that there is some correlation between average temperature and CO2 in the air.
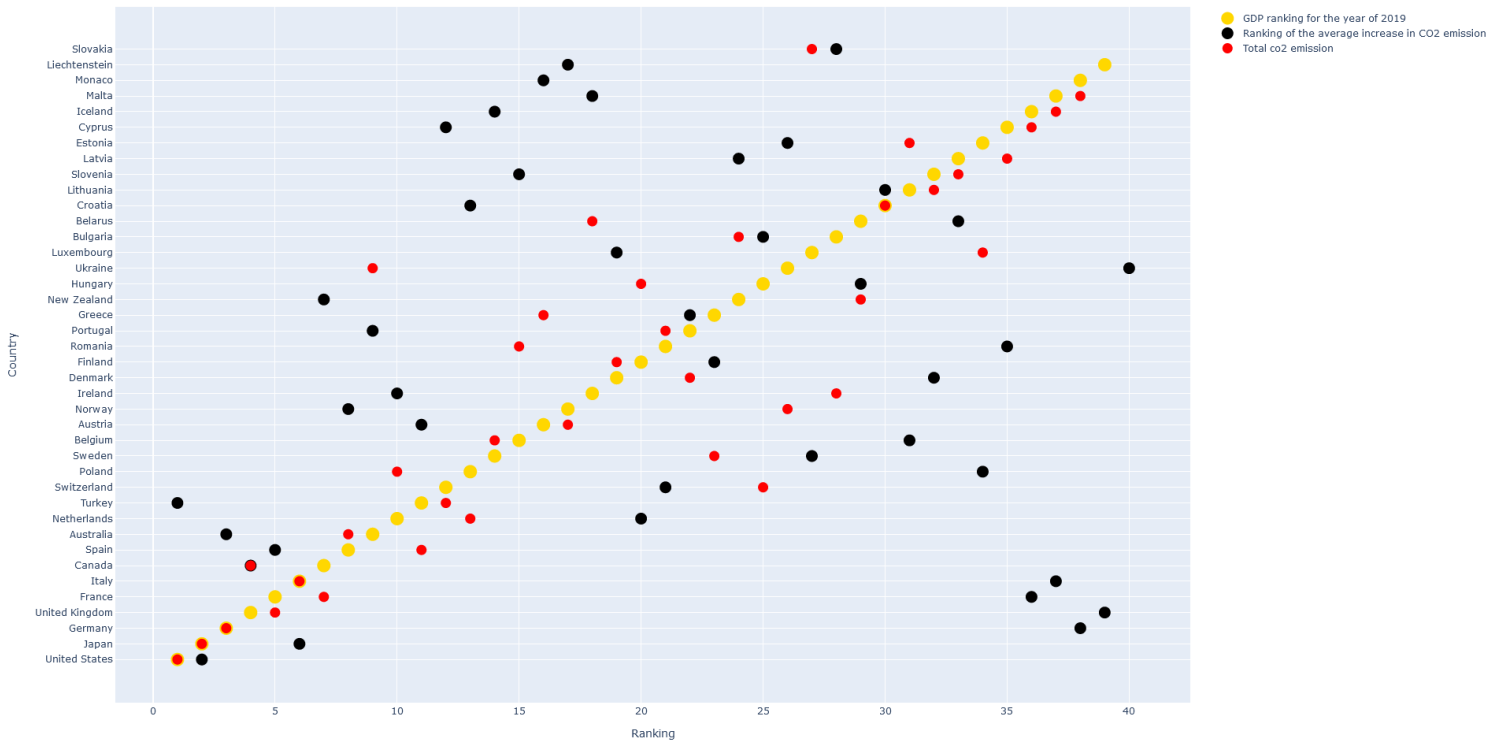
Next, I created a graph of average temperature vs CO2 emission using the year as a connecting point. The x axis would be CO2 emission and the average temperature would be the y axis. I then used linear regression and calculated the r^2 of its linear regression. The graph below is what happens when I run the function, "draw_global_co2_vs_temperature()".

The linear regression is y = 1.4882313028812986 + x0.021286751086781564It also has an r^2 = 0.7379434278337981

This linear regression has an extremely high r^2 value. The value of 0.740 is ridiculously high and shows the high correlation between the CO2 in the atmosphere and the average temperature of the world. From the dot plot we cannot conclude the correlation in terms of specifics country's pollution production but the correlation between global pollution and global average temperature is clearly shown in this graph. Which means that even though a country's emission might not affect its average temperature, CO2 emission clearly affects global warming in a global scale.

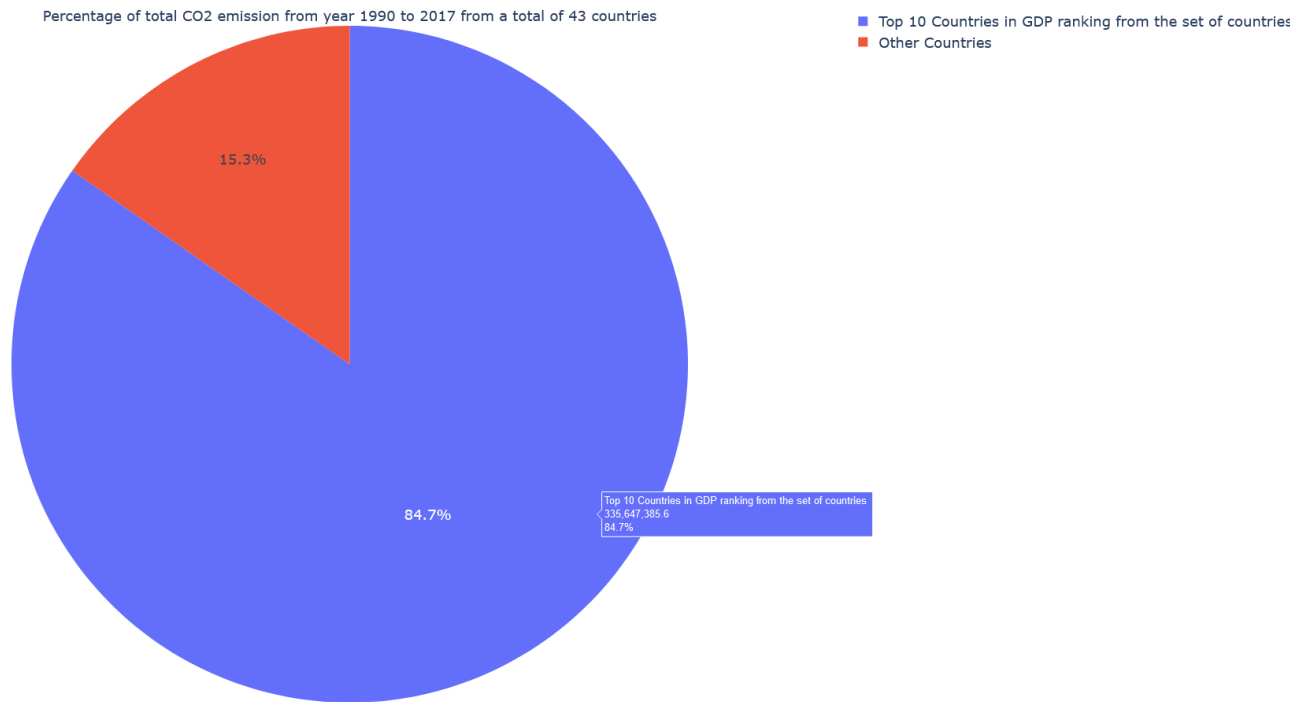**Fourth Step: Examining the relationship between GDP and CO2 emission**
Since there does not seem to be a relationship between GDP and increase in temperature from what can be seen in the dot plot, I have decided to examine the relationship between GDP and CO2 emission. After all, there is a high correlation between CO2 emission and global warming. First, I want to see the rough relationship between GDP in 2019 and CO2 emission. To do this, I repeat the ranking dot plot but without the average temperature. The picture below is what is produced by running the function "co2_gdp_ranking(list_of_countries_CO2(), 1983".

Difference in Ranking for GDP(2019), overall CO2 emission, and average increase in CO2 emission from the year 1990

As we can observe, the red dots are often close to the yellow dots in terms of its positioning. From this graph, it is possible to see the correlation between a country's GDP and the total amount of CO2 emission produced. It is most likely that country's with a high GDP would release a lot of CO2 and pollute the atmosphere. This can be seen in the first three countries, United States, Japan, and Germany which holds the top 3 in GDP and also top 3 in total CO2 emission from 1990. This brings up an interesting conclusion, from the second step we know that GDP and emission does not have that much correlation with a country's average temperature increase. But from the third step, we know that the increase in global temperature is highly correlated with the amount of CO2 in the atmosphere. This means that country's with a high GDP should be held accountable for releasing CO2. While it is not showing a significant difference in increase in their own borders, the CO2 emitted still affects global warming in a global scale.

Fifth Step: Percentage of contribution of CO2 from major countries
From the previous step, we have found out that CO2 emission does indeed have correlation with climate change and that countries with higher GDP has higher CO2 emission. But what exactly is the percentage of emission that major countries contribute to the percentage of emission? I will attempt to measure this by summing up the top 10 countries in terms of GDP that is within the co2_data. Then, I will compare it to the sum of co2 emission of the rest of the countries. I will run the function, "draw_percentage_of_pollution(helper_functions.list_of_countries_co2(), 1990, 2017, 10)" to make the graph

Percentage of total CO2 emission from year 1990 to 2017 from a total of 43 countries

15.3%

84.7%

Top 10 Countries in GDP ranking from the set of countries
335,647,385.6
84.7%

In the scenario, I sorted the countries based on its GDP and isolated the top 10 countries from the others. When looking at the pie graph, an astonishing fact is that the countries in the top 10 of the GDP ranking makes up 84.7% of the pollution in the data set. This 10 country makes up 84.7% of the emission that 43 countries make together. Of course, this data set does not contain all the countries, therefore, this conclusion must not be taken too conclusively. Yet, this pie graph shows that country's with high GDP are often major contributors in terms of CO2 emission. Therefore, major countries should be held accountable for its CO2 emission due to the sheer volume of CO2 that they produce.

Sixth Step: Relationship between average temperature and increase in temperature

My personal interest for this project is trying to find if the climate a country currently has is more affected or less affected by global warming. In order to do this, I decided to graph a graph for a country's average increase in temperature versus their average temperature. By doing this, I will be able to see if hotter countries are affected more by climate change or if cold countries are affected more by climate change. The process that I used to do this is to first calculate the average temperature with the function "average_country_temperature" which simply adds up all the temperature and averages it out with statistics.mean. Next, I calculated their increase with "average_yearly_increase_temperature". I do this by finding out the change in temperature in each yearly interval, adding them up, and then averaging it out. Then, I plot both values in the graph below using the function "draw_country_increase_vs_country_average" which results in the graph below

Average increase in temperature as a function of average temperature.
The linear regression has an equation y = -0.0023860786221891455 + x0.0004621799455150217. It also has an r^2 = 0.028882165358118495
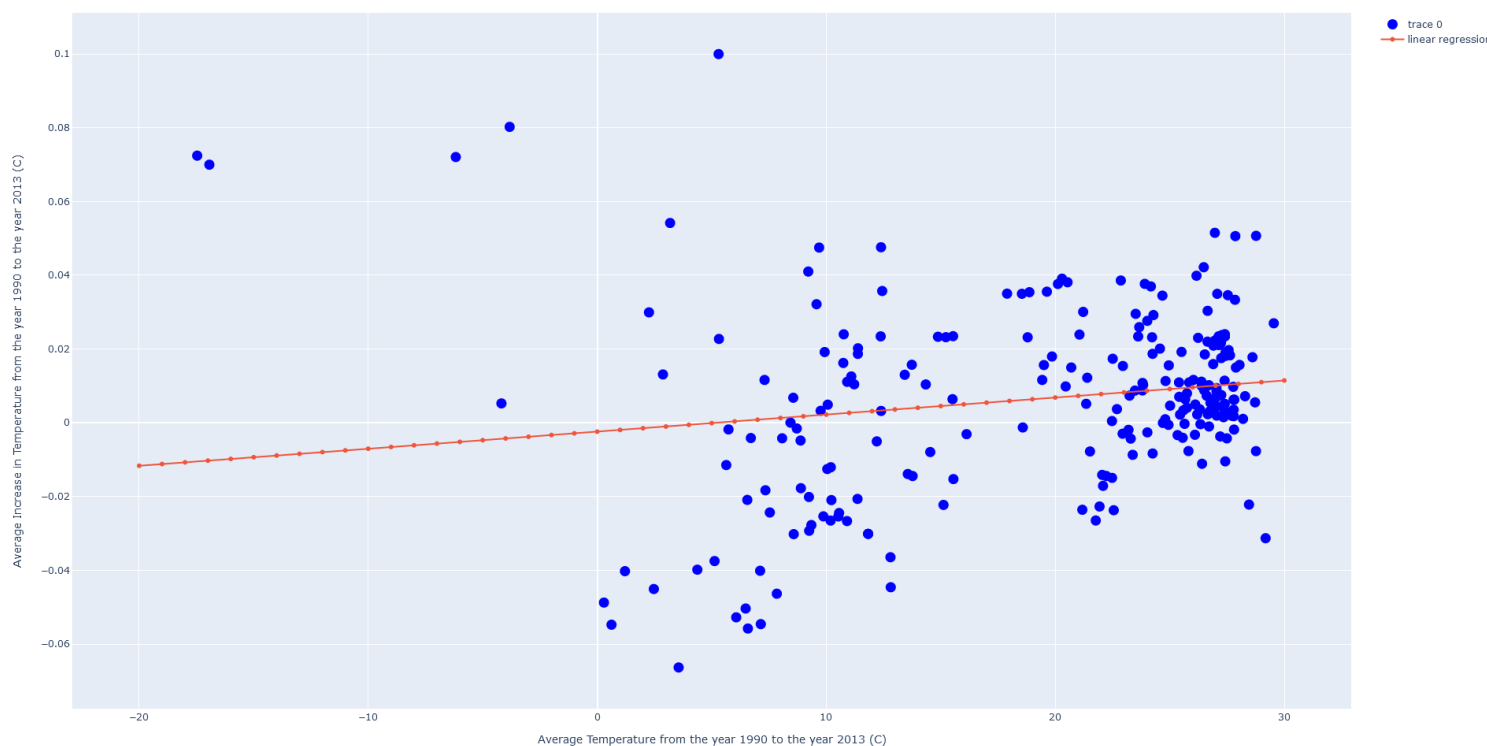
From this linear regression, there does not seem to be any relationship between a country's temperature and how badly it is affected by global warming. While there might be a slight correlation in which hotter countries have a higher increase in temperature, the correlation is pretty low as it only has an r^2 of 0.0288

Seventh Step: Analyzing the trend of increase in CO2 emission and increase in temperature in individual countries In this scenario, wrote a function in order to analyze trends of singular countries. What are their changes in temperature when they have changed their CO2 emission by a certain amount. Does their temperature increase if their emission increase? The way I calculated these values is by first limiting the range of the year to years that contain both CO2 data and temperature data. Then, I isolated a dataset with just the country that is desired. After that, I simply find the change when it changes year and then append it all to a list. This whole process is done by the helper functions. yearly_increase_temperature and yearly_increase_co2. Since the list is sorted by date, the x and y values from two function is corresponding to the same change in year. In order to produce the graph below, we run the function, "draw_graph.draw_increase_co2_increase_temperature_country(country_data, co2_data, 'United States', 1990, 2013)"

Increase in temperature as a function of increase in CO2 for the country United States in the years 1990 to 2013
The linear regression has an equation y = 0.018379522641852785 + x7.885192200411632e-07. It also has an r^2 = 0.0833463219364815
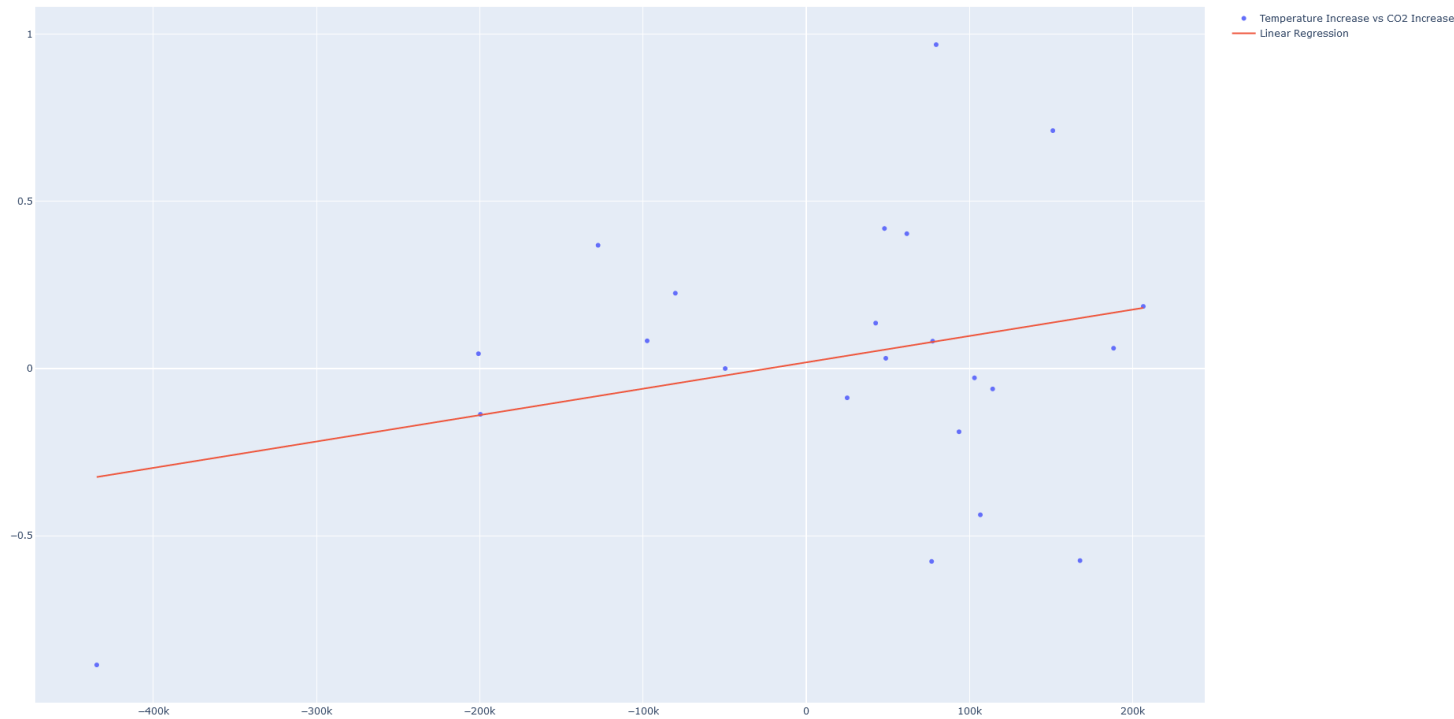
From this linear regression, we can observe that it is true indeed that global warming is linked with CO2 emission even in a country-wide scale. By running this function on different countries, it is possible to observe the differences in terms of change in temperature and change in CO2 emission. But what if we were to analyze this in a global scale?

1 Eighth Step: Analyzing the trend of increase in CO2 emission and increase in temperature in all 43 countries in the UN dataset

Instead of analyzing each and every country's trend, this time, I decided to analyze the trend of all the countries that is within the UN dataset. In this scenario, I summed up their total yearly change in CO2 emission for that specific room and I also summed up their total change in temperature. I graph this graph by running the function, "draw_graph.draw_increase_co2_increase_temperature(country_data, co2_data, helper_functions.list_of_countries_in_all(co2_data, country_data), 1990, 2013)". Then I obtain this graph.

Increase in temperature as a function of increase in CO2 in the years 1990 to 2013
The linear regression has an equation y = -0.19044112984173406 + x1.7895459358447483e-06. It also has an r^2 = 0.0007350431250782163
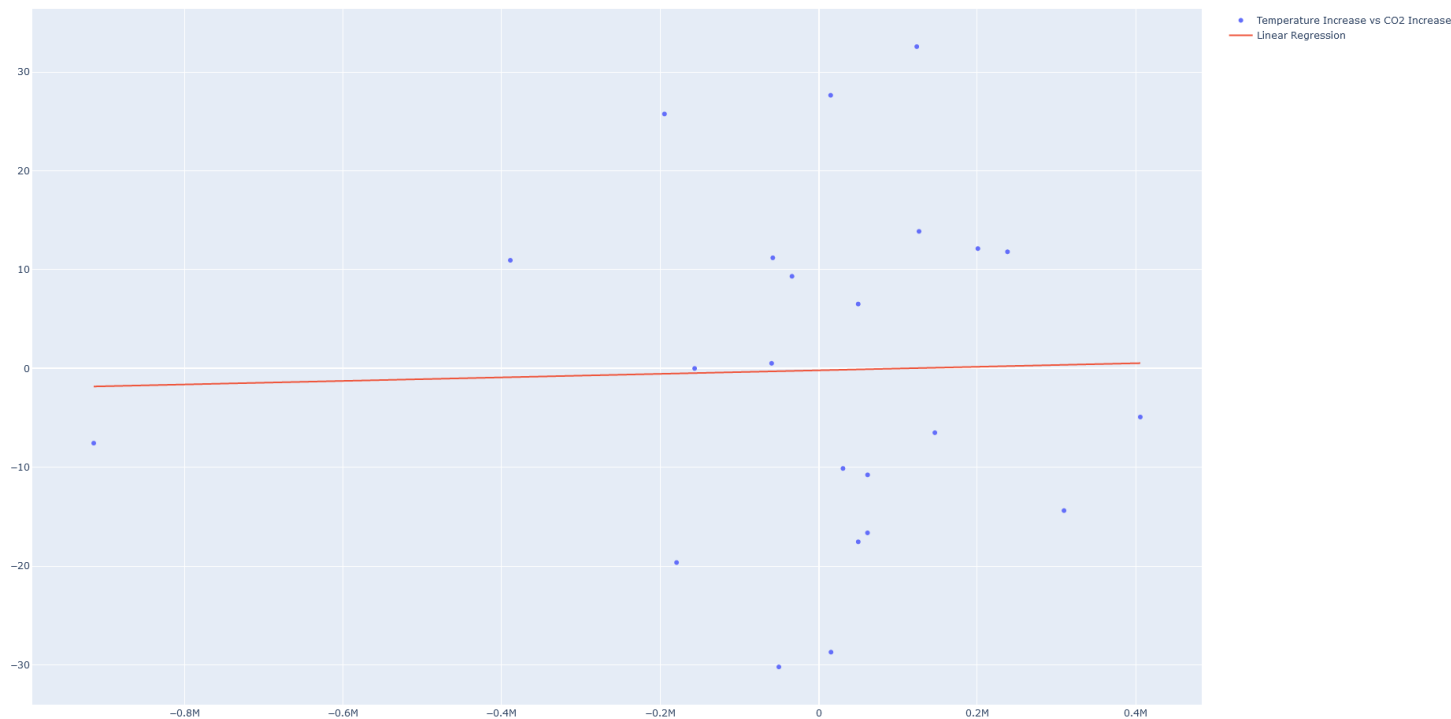
Of course, this data set is limited, after all, it only has data from a limited amount of countries, in this case, the 43 countries that are within the both the UN dataset and the CountryTemperatures dataset. From this graph, we can observe a slight correlation in CO2.

Downloading the Data

Datasets to be Downloaded Download all the dataset in the links given below

http://data.un.org/Data.aspx?d=GHGf=seriesID%3aCO2 (Download the UN data as a csv file)

https://datacatalog.worldbank.org/dataset/gdp-ranking

(For the world bank file, go to the data and resources and download the csv file. You need to right clik and open in new tab for the download to work)

https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data?select=GlobalTemperatures.csv

https://www.kaggle.com/econdata/climate-change?select=climate_change.csv

https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data?select=GlobalLandTemperaturesByC

Modules to be Downloaded Download Plotly with the link given below https://pypi.org/project/plotly/

Download Statistics Module https://docs.python.org/3/library/statistics.html

Download Datetime Module https://pypi.org/project/DateTime/

Download Dataclasses Module https://pypi.org/project/dataclasses/

Download Operator Module https://docs.python.org/3/library/operator.html

Downlaod Typing Module https://pypi.org/project/typing/

Download csv Module https://docs.python.org/3/library/csv.html

Download pandas module https://pandas.pydata.org/

2 Changes in the Project

While I was doing my project, I made some slight adjustments. I listened to the TA feedback and decided to not incorporate pygame in my project in contrast to my original plans. Additionally, I also decided to explore the contributions of major country towards global corruption after finding out that major countries are often linked with

higher CO2 emission. This part was seen in the fifth step on my computational overview. Other than that, I did not make any major changes in my project.

3   Discussion

As a reminder, my discussion question was **what is the relationship between a countries productivity, pollution levels, and climate change and how could this relationship be used to limit climate change.**. In my analysis of this data, I used CO2 emission levels to measure pollution level, I used GDP to measure a country's productivity, and finally, I measured the change in average temperature to show the effects of global warming. In this third step, I analyzed a linear regression between the change in global temperature and the change in CO2 emission in order to show that there does indeed exist a correlation between these two variables. Interestingly enough, when seen from the second step, country's with high change in emission does not exactly have the highest change in average temperature. This is weird, after all, there does exist a correlation between global temperature and CO2 emission. So what does this show? The first conclusion that I want to make is that while CO2 emission might not severely alter a country's temperature such that it changes in temperature more than the global change, CO2 emission still affects climate change in a global scale, and therefore country's should still take action to limit CO2 emission in order to stop climate change in a global scale.

The next variable I would like to analyze is the productivity of a nation. In this case, I used GDP in order to measure productivity. Again, as seen in step two, countries with high GDP is not necessarily the most affected by climate change. For example, Bulgaria is most heavily affected by climate change yet it is only the 28th in terms of GDP from the 41 countries that are analyzed. So how exactly can GDP be factored into this discussion? Well, through the fourth step, it looks as if countries with high GDP are also similarly highly ranked in terms of CO2 emission. On a negative note, it shows that countries productivity seems to increase with more CO2 emission. On a positive note, as seen from the four black dots in the corner of the graph in the fourth step, some countries have begun to take a step away from CO2. They have begin to decrease their CO2 emission drastically. Yet, much still needs to be done. In the pie chart in step five, it could be observed that amongst 43 countries, the top 10 in terms of GDP contributed up to 84.7% of the total CO2 emission from all these countries. Of course, this data is but a small sample size of all the countries in the world since it is only data of 43 countries. But this small bubble still shows how a majority of CO2 emission in the world is produced by major nations. The major nation's average temperature does not seem to be affected that much in comparison to the ratio of pollution that they produce in the world. In a fair world, the major nation should take up 84.7% of global warming, yet the country of Bulgaria is the one most affected by global warming. Due to this, one conclusion that I would like to make is that in terms of analyzing change in temperature, it is of importance to analyze the change in a global scale rather than in a country-sized scale. After all, while the production of CO2 is heavily weighted towards major nations, the effects of global warming is equally shared by all.

Another analysis that I did out of curiosity is whether or not a country's temperature affect the extent of temperature change. For example, whether or not a cold country like Canada is affected more by global warming than warmer countries. This analysis is done in the sixth step. it resulted in an extremely low correlation, with a very low slope of 0.00046 and an r^2 of 0.029. But from the eye test, it does seem like countries with higher temperature has a higher average change, but there are like 5 outliers in which those countries have a pretty low average temperature yet has a drastic increase on temperature. Just from this graph, there does not seem to be much conclusion to be made.

Of course, there are still issues to be discussed about in this analysis. First of all, the sample of CO2 emission is not exactly complete. The dataset from the UN only contains 43 countries. Second of all, the GDP ranking used was GDP in 2019, so all analysis regarding GDP was not taking changes into GDP into account. Finally, all these data all end in the year of 2013, when there was not much movements against climate change, and therefore we do not know to what extent countries temperature have been affected in the years after 2013.

References

UNData, "*Carbon dioxide (CO2) Emissions without Land Use, Land-Use Change and Forestry (LULUCF), in kilotonne CO2 equivalent*", Greenhouse Gas Inventory Data, January 1, 2019. Web. December 11, 2020
http://data.un.org/Data.aspx?d=GHGf=seriesID%3aCO2

Berkeley Earth. *Climate Change: Earth Surface Temperature Data.* (Version 2). Berkeley, California: Berkeley Earth, May 2, 2020. Web. December 11, 2020.
https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data

World Bank Group. *GDP* Ranking. World Bank Group, July 1, 2020. Web. December 11, 2020
https://datacatalog.worldbank.org/dataset/gdp-ranking

synergy37AI. *climate change Green House Gas Effects.* (Version 1). ESRL/NOAA Physical Science Laboratory, December 27, 2019. Web. December 11, 2020 https://www.kaggle.com/econdata/climate-change?select=climate_change.csv

Behic. "The Beginner's Guide – Building Interactive Maps in Python." *Sonsuzdesign*, Wordpress, August 1, 2020, https://sonsuzdesign.blog/2020/08/01/the-beginners-guide-building-interactive-maps-in-python/. Accessed December 13, 2020