Analysis of International Gross Domestic Product and Carbon Emissions

Liam Orcutt, Kody Politza, Tai Kitazato, Qiutong Xiao, Zejie Yan

12/08/2022

Abstract

Using Pearson and Spearman correlation tests, autocorrelation tests, as well as graph analysis, we found several interesting trends demonstrating the effects of economic fluctuations and national/international recessions on carbon emissions. Graphs were made and statistical analyses were performed for each country over a time interval of 1995 to 2017. This time range had the most consistent reporting of values for both GDP and emissions, thus was assumed to give the most accurate results possible given our data. We found several interesting relationships that met our predictions, as well as some that completely subverted expectations. With the statistical analyses performed, we were able to research socioeconomic and environmental factors related to the relationship in an attempt to answer the questions that arose after seeing high amounts of positive or negative correlation for certain countries.

Introduction

Climate change is considered by many to be the biggest threat to our generation. Solving such a problem requires us to determine and list its many causes. The changing climate is primarily caused by the carbon footprint left behind by humanity. However, this value ebbs and flows over time. For example, the SARS-CoV-2 “coronavirus” pandemic saw drastic decreases in carbon emissions during the lockdown period (Ram et. al., 2022). This begs the question, if an international event like the pandemic can affect our emission output, what other factors play a role in deciding how much pollution we generate?

This idea was what spurred our research as we began to look for potential significant factors. After some deliberation we decided to research the relationship between gross domestic product (GDP, US dollars) per capita and CO_2 emissions (metric tons) per capita for each country. Other research has looked into similar hypotheses in the past, thus we had a solid basis to work from. One such group of individuals found that of all the countries in the EU, “only in Luxemburg the GDP strongly decoupled from CO_2 emissions from public energy and heat production [sector] for almost the entire period 1990-2014” (Beykaliyev, et. al., 2021).

Initially, we predicted that there would be a largely positive correlation between the two on an international level. Next, we aimed to narrow our scope to particular regions and compare the results of testing across these groups. Lastly, we began to investigate whether there are outlying countries with very low GDP and relatively high CO_2 emissions and vice versa, in addition to comparing different country's plots of GDP v.s emissions against their correlation coefficients.

Data & Materials

For our statistical analysis, we used two data sets from the World Bank: CO_2 emissions in metric tons per capita and GDP in U.S dollars per capita. The emissions dataset does not contain data for any country before the year 1990, whereas the GDP dataset contains data for most countries beginning in 1960.

Some countries in the datasets do not contain any data at all. This could be because those countries are territories of other countries, for example American Samoa and the British Virgin Islands. Another important thing to note is that some countries do not exist anymore, and some new countries are there as well. For example, South Sudan split from Sudan in 2011, therefore South Sudan does not have any data until that year.

Due to the variance in the data, we ended up losing around half of our data. This was primarily due to missing values, as well as performing a correlation test on two variables requires both matrices to be of identical shape. To reach this result of equivalent dimensions when storing our data, we used Python dataframes. Thankfully the World Bank formats all of their csv downloads similarly and over the same time frame, so the data files for both GDP and emissions ranged from 1960 to 2021, albeit large amounts were missing as explained prior.

In terms of preprocessing and data cleaning, the task was relatively streamlined. As alluded to, the World Bank formats their files in the same way, so they had the same amount of values from the start. However, the initial dataframe was not in a tidy format, so we aimed to change this by using the Pandas *melt* function on the two dataframes. We merged on the "Country Name" column, which formatted the data to have each country be the observation and each year to be a variable. Thus, we were left with a tidy data set consisting of a record for each country for each year. We also dropped several columns that were irrelevant to the data and were primarily metadata.

Finally, we needed only the countries that provided values for both GDP and emissions each year, thus the *dropna* function was used. This is what resulted in a large portion of our data being removed, however we were still left with a good amount. Of the original 16429 rows in the frame, after preprocessing was complete, we had 6934 rows of complete data. On closer inspection however, this loss was not as significant as we initially thought, as several thousand of those rows consisted of null values for the years 1960 to around 1990 depending on the country. Ultimately, we were left with more than enough data to work with for the majority of countries and were able to begin analysis. The time frame we chose to analyze was from 1995 to 2017.

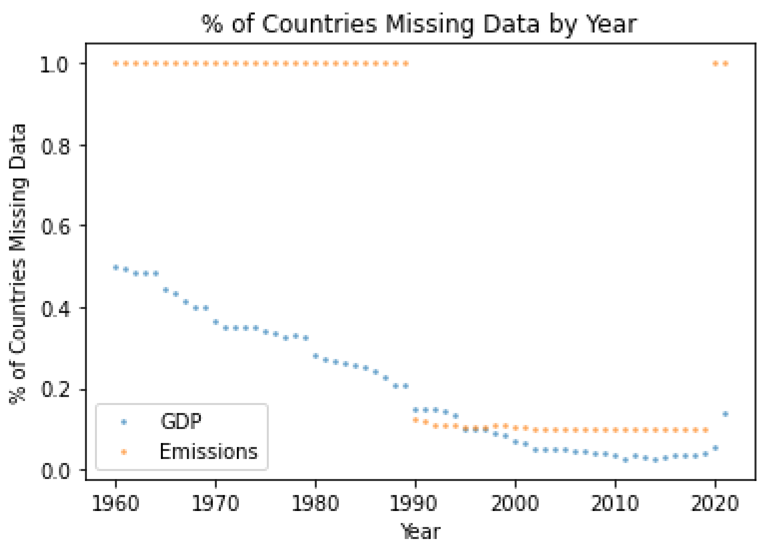

Figure 1. *Scatterplot showing the proportion of missing values for GDP per capita and* CO_2 *emissions per capita by year for all countries.*

Note: *pearson_spearman.py is used for calculating coefficients and storing them in text files, graphs.py is used for generating a GDP v.s Emissions graph for any specified country or region, autocorrelation.py is used for generating autocorrelation plots for each country.*

Results

The first question we asked was whether there was a largely positive correlation between GDP per capita and CO_2 emissions by country across the globe. Our goal was to see if one of the variables led to the other. We aimed to investigate whether we can form a conclusion of one variable while looking at the trends in the other from before. We pondered if an increased GDP led to an increased emission rate, and conducted our initial testing and analyses with this question in mind.

To work towards this goal, we first needed to visualize the data we were working with across all years. Our preliminary figure is shown here:

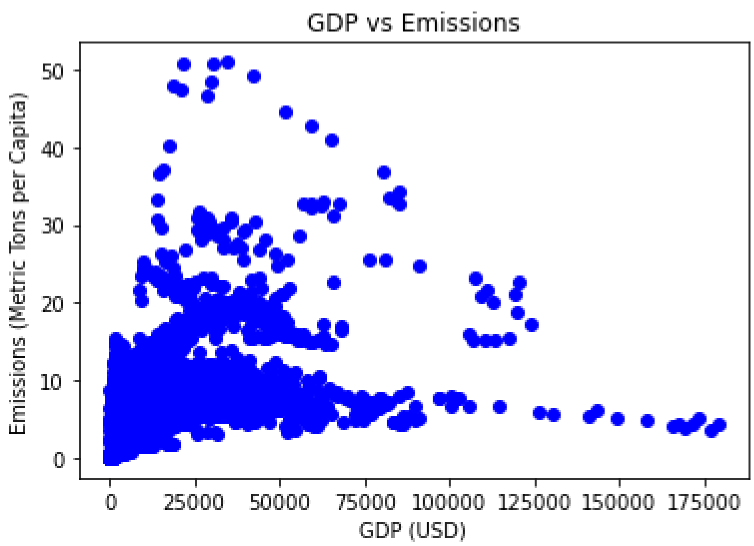

Figure 2. *GDP v.s Emissions plotted for all countries, across all years.*

While this plot is rudimentary and outliers could potentially be the same country plotted over multiple years, immediately we began to glean important information and draw minor conclusions. For one, no country has extremely high emissions and high GDP at the same time. Additionally, judging by the spread of data, there appears to be a general relationship between the variables, as shown by the grouping and clustering. As such, we predicted there may be some general correlation over all of the years.

To confirm this, we performed both Pearson and Spearman correlation tests on the collection of international data.

```
Pearson Correlation between  GDP  and  Emissions across all countries is:  0.572
Spearman Correlation between  GDP  and  Emissions across all countries is:  0.85
```

Figure 3. *Terminal output of the results of both tests. Function found on line 45 of pearson_spearman.py*

As shown above, our *corr_testing* function proved our hypothesis correct, revealing a moderately correlated linear relationship and a significantly correlated monotonic relationship between GDP and emissions on an international level.

In an effort to verify whether or not our results were consistent with others in the community, we researched different works to compare and contrast our results. One such group also performed research on GDP and emissions, through which they showed similar results: “The correlation is positive, which suggests growing per capita GDP leads to increasing carbon dioxide emissions. No turning point is found at which emissions start to decrease when reaching a high enough GDP…” (Cederborg, 2016). We concluded that since we reached a similar result, our measurements are accurate.

Another variable we can observe on an international level is autocorrelation. Autocorrelation is the comparison of a single variable with respect to time. In this instance, we looked at the GDP of each country from between 1995 and 2017, and calculated the correlation with respect to the GDP in 2017. One thing we observed in the autocorrelation graphs is the fact that most countries follow a very similar trend, which includes the graph approaching 0 at 7/8 years of lag, and it approaches a minima at around 15 years of lag. One thing we observed was the fact that although most countries followed a similar trend, a few countries with extreme real world issues (such as a coup or an economic disaster) don’t follow the trend that the other countries follow. Some of these outlying countries are covered in a later section.

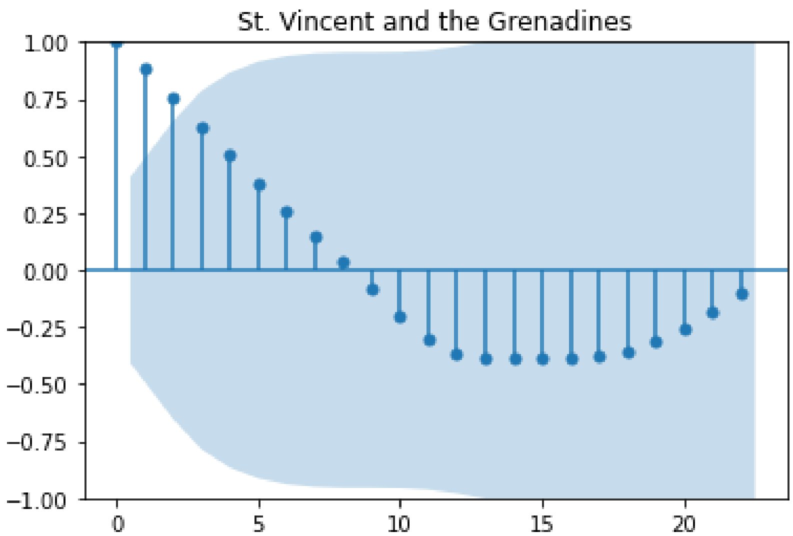

Figure 4. *Example of the general autocorrelation relationship. The x axis is the lag (by year) difference between the initial value(2017) and the observed(2017 - Lag). The y axis shows the autocorrelation coefficient.*

Second, narrowing down from strictly international analysis, we also aimed to investigate whether there was a general correlation between regions. We considered whether there were some regions that had higher or lower correlations and what factors might be influencing this. We also considered what GDP per capita is commonly associated with a high CO_2 emission and if there is a common trend in correlation among these regions.

One of the regions we were particularly curious about was Asia, particular the East Asia and Pacific geographic region. Considering that this region encompasses several individual countries with moderate to significant correlation, and a few with little to no correlation, we predicted that the region as a whole would present a strong positive relationship.

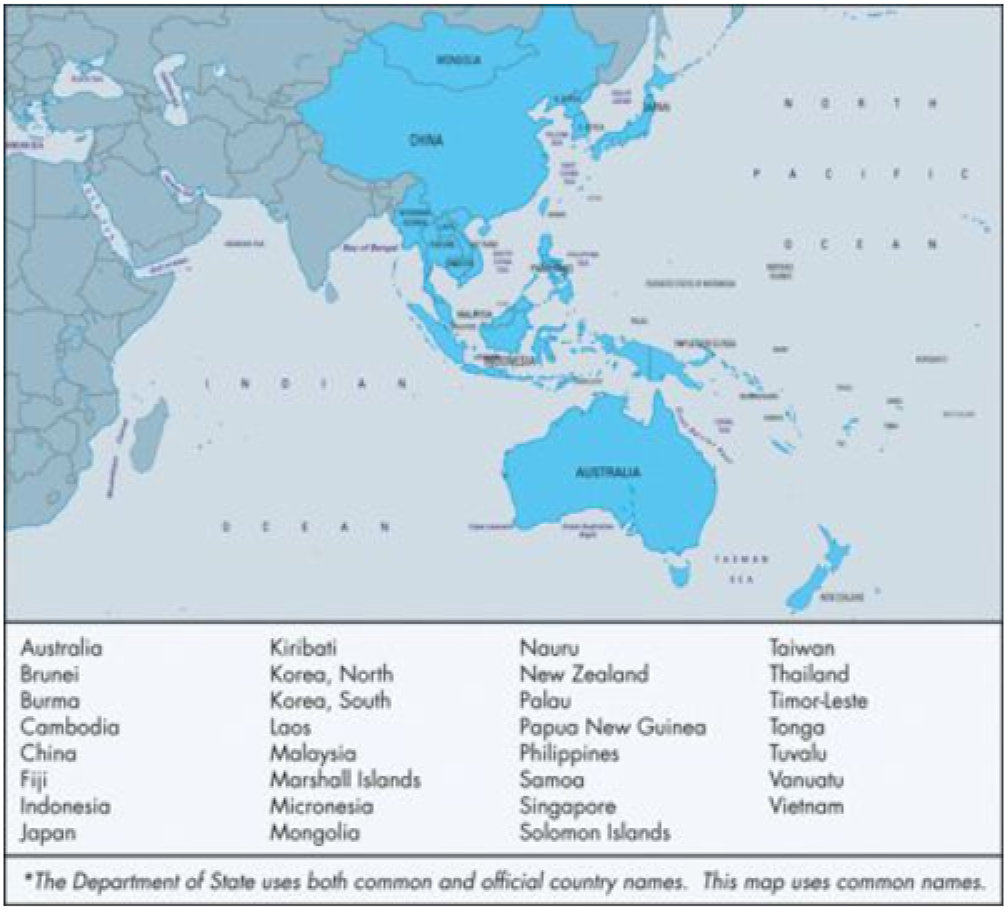

Figure 5. *East Asia and Pacific Region (EAP).*

According to the Bureau of Resource Management – a subsidiary of the U.S Department of State – as of 2011, the EAP is responsible for 25 percent of the international per capita GDP. This value is likely to have increased as shown by the recent data, GDP trend graphs, and correlation coefficients. Currently, the EAP has a Pearson coefficient of 0.963 which demonstrates very significant levels of linear correlation, in addition to a Spearman value of 0.933 which shows high levels of monotonicity. Our results concluded that the EAP fell into the top three regions with the strongest relationship (Pearson test) between rising GDP and increased emissions, trailing South Asia (0.991) and the Middle East (0.97). The following figure visualizes the strong relationship:

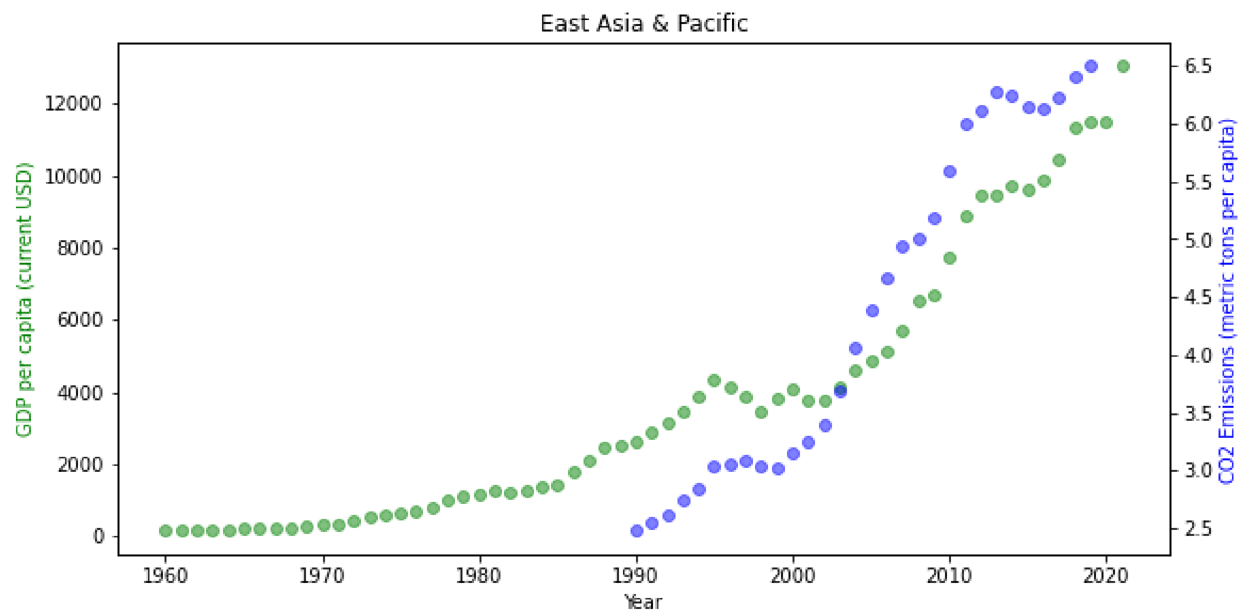

Figure 6. *Scatterplot of the EAP region's GDP and emissions by year.*

On the opposite end of the spectrum, we found interesting and unexpected results with the North American region. Initially, we hypothesized that North America would demonstrate a very strong positive linear correlation as well as high monotonicity reaching back far into the 1900s due to the extreme amounts of industrialization and motorization over the past several decades.

Figure 7. *North America Region (NA).*

Surprisingly, the NA region was found to have a significant negative relationship, meaning that since 1990, carbon emissions have been decreasing as per capita GDP has increased. Our testing showed that the NA region had a linear coefficient of -0.883 and a monotonic coefficient of -0.869.

These results can be explained primarily due to the steady decrease of coal usage since the 1900s. Due to advances in eco-technology and renewable energy sources, fossil fuels and other greenhouse gas producers have been in steady decline in recent years (Rivera, 2021). Other sources indicate that the NA region has been steadily decreasing emissions in areas other than coal and fossil fuel usage as well: “For the United States, during the period from 1990 to 2020 methane emissions decreased by 17 percent, as reduced emissions from landfills, coal mines, and natural gas systems more than offset increases in emissions from activities such as livestock production.” (Environmental Protection Agency, 2022).

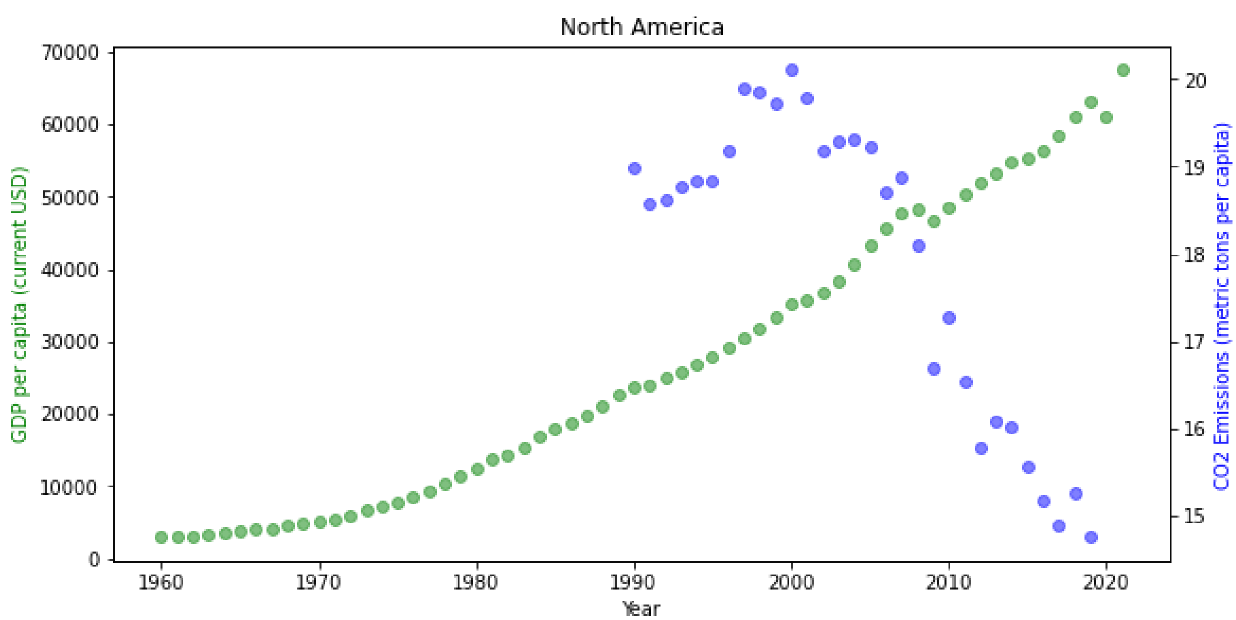

Figure 8. *Scatterplot of NA region's GDP and emissions by year.*

Despite our expectations being subverted, it was encouraging to see the results of the testing. Our results and current research show that the NA region has seen a significant decrease in emissions while per capita GDP has been on a constant rise, save for a small dip caused by the economic recession in the 2000s-2010s. With all of the evidence proposed, the strong negative linear/monotonic relationships are explained and the NA region finds itself leading the pack in terms of a negative correlation, with the European Union following at -0.683.

Finally, we wanted an empirical list of what countries of the world had the strongest relationship between GDP and emissions, and what outside information we could collect to explain this phenomenon. We also noticed some countries defied the concept of a general relationship, and we pursued outside reasoning to discover potential causes for this. By researching the socioeconomic factors behind each country's values, perhaps we could glean some knowledge on how to stem the trend moving into the future.

The country we found to have the highest Pearson correlation was India with a 0.989. India is one of the world's fastest growing economies, growing at a rate of 8-9% annually. Industrialization has grown tremendously there in the past two decades, leading to more energy consumption and opportunity for economic growth. At the same time, around 80-90 percent of India's commercial energy is being produced by fossil fuels such as oil, gas, and coal (Inumula, 2017).

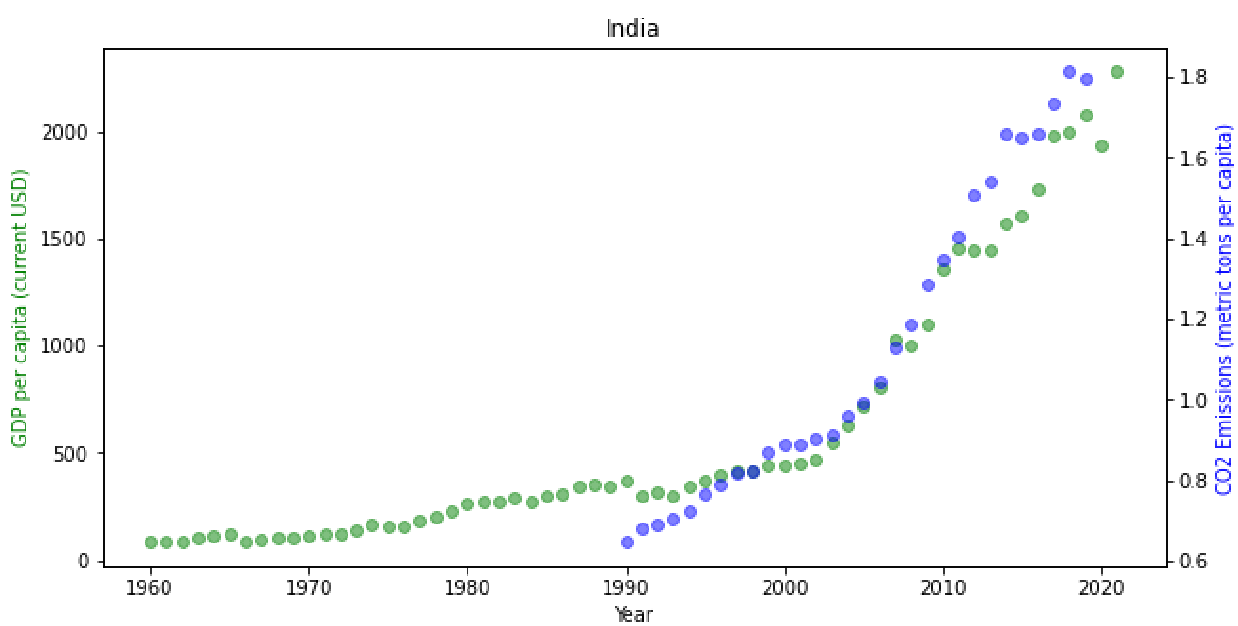

Figure 9. *Scatterplot of India's GDP and CO_2 emissions by year.*

The graph above demonstrates what was mentioned prior. It is shown that GDP per capita has been overall increasing over the years, while their emissions follow the same trend. Fossil fuels are the main culprit of producing CO2 emissions. The increase in emissions reflect the increasing economic activities, and structural reforms in India's economy.

China is one of the world's leading economies and has been experiencing rapid growth in the past decades. The country's growth has mainly relied on its use of fossil fuels that generate CO_2 emissions. As a result of this, China is currently the world leader in carbon dioxide emissions. This makes China a particularly interesting country to investigate regarding the relationship between GDP and CO_2 emissions. (Caporale, 2021)

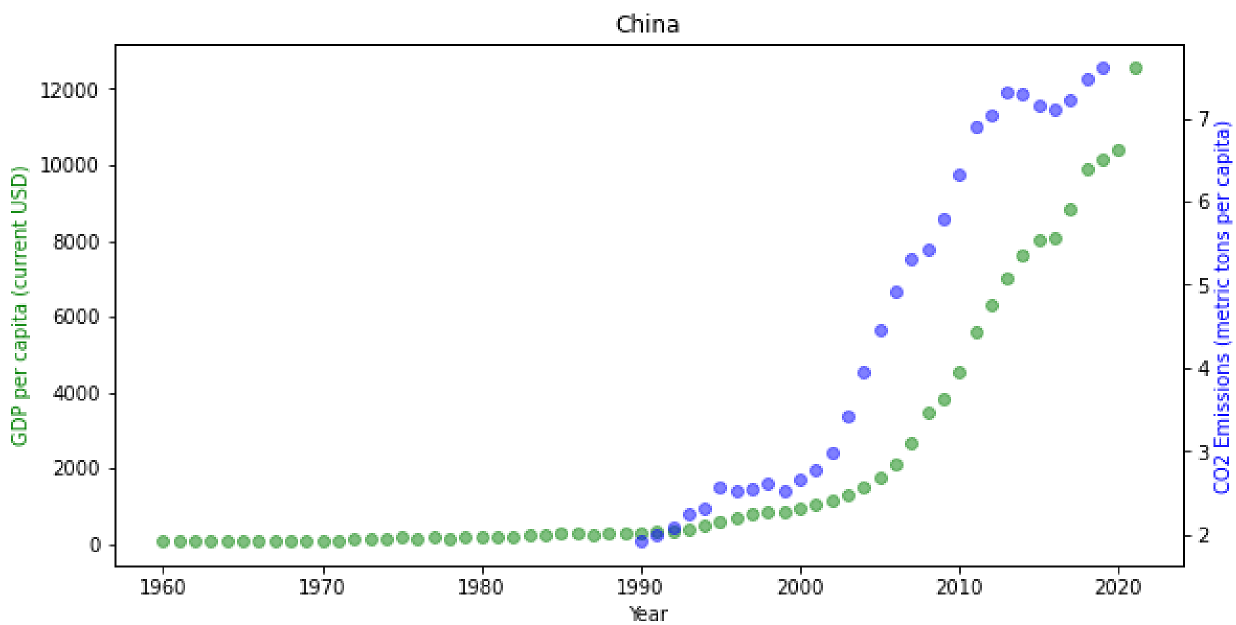

Figure 10. *Scatterplot of China's GDP and CO_2 emissions by year.*

Looking at the graph above, it is easy to see that carbon dioxide emissions increase at the same times that GDP increases. China and India seem to be in a similar situation as they are both one of the world's most populous countries and experiencing fast industrial and economic growth. This may lead people to believe that there would be a strong correlation between GDP and emissions for all countries. However, this is not the case. It was found that many other countries had a reversed correlation, as discussed below.

Unlike the previous two examples, the United States has a negative correlation between emissions and GDP at -0.883. This negative correlation can be explained by the fact that the U.S started outsourcing their manufacturing to other countries, such as China and Vietnam. Also, green energy has become more and more economically viable.

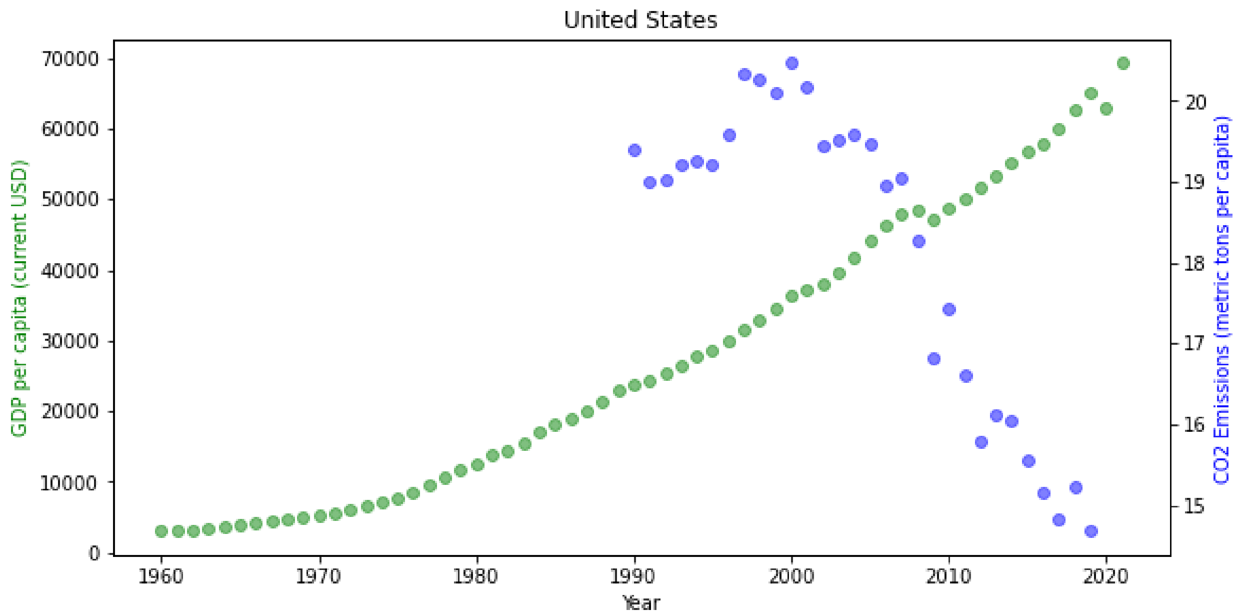

Figure 11. *Scatterplot of United States GDP and CO_2 emissions by year.*

Looking at the graph, we can see that although the US has a negative trend of emissions, it is still extremely high at around 15 tons per capita. In comparison, China peaks at 7 and India caps at around 1.8. This could be explained by the fact that the US already had enormous manufacturing capacity before 1990, and they have been working in more recent times to lower their emissions. Individual consumers in the US emit more CO2 than any other country, and this can be explained by our spending habits, which includes eating beef and driving frequently (The Nature Conservatory, 2022).

Another country with a negative correlation is Qatar (-0.722). Although Qatar is a large exporter of fossil fuels, they are on a decline in CO2 emissions. The large emission per capita metric can be explained by the fact that only around 11% of the people living in Qatar are citizens.

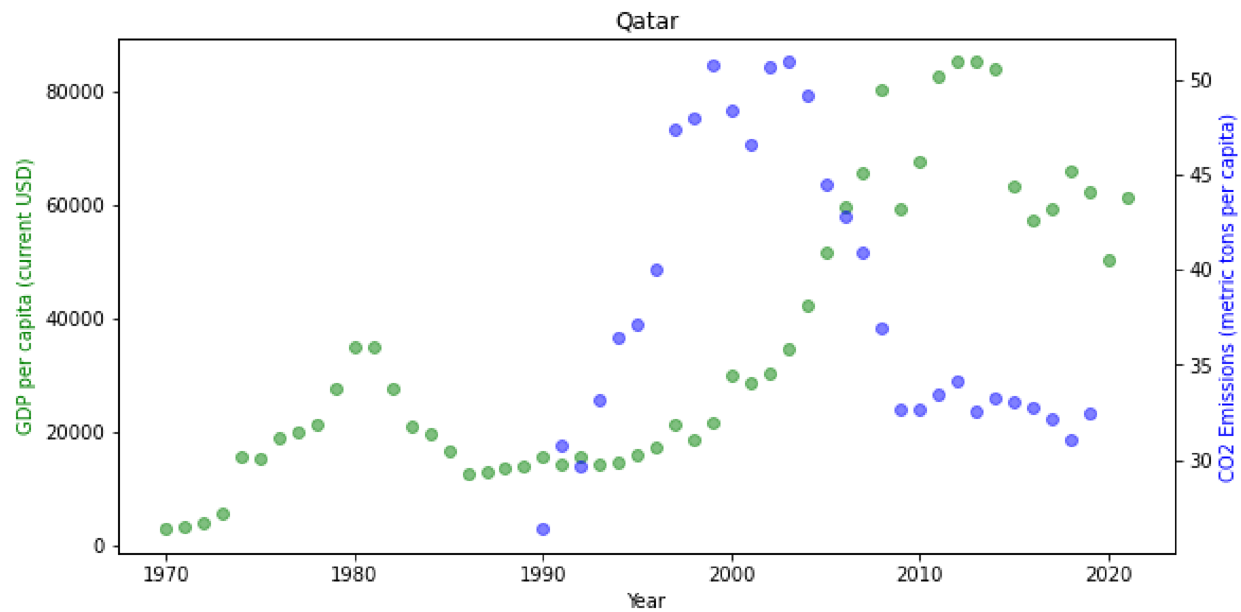

Figure 12. *Scatterplot of Qatar's GDP and CO_2 emissions by year.*

From the plot above we can see Qatar's decline in emissions. One thing that is notable is the fact that the decline of emissions was followed by a decline in GDP not very much after. This could be due to the fact that the emissions are directly correlated to GDP.

The final country we want to observe is Japan. There are a few notable things we noticed in the data, firstly in the autocorrelation. Japan is an example of a country that does not follow the standard trend for autocorrelation of GDP with respect to time. This could be explained by their numerous environmental disasters, along with the fact that the US was controlling their economy for a time after the second World War. One thing that we can note from the graph is the fact that it almost follows a similar shape, but it looks much more compressed. This could indicate that it took a while after the war for Japan to readjust to the global economy.

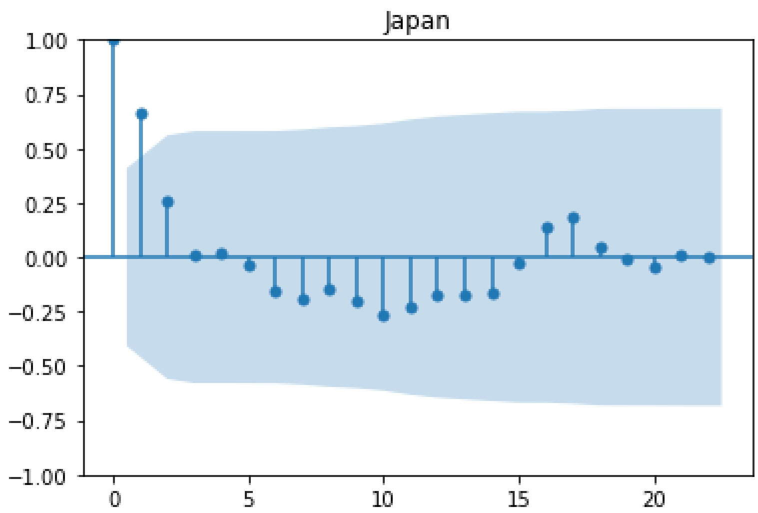

Figure 13: *Autocorrelation of Japan from 1995 - 2017*

Along with its autocorrelation plot, Japan's correlation between emissions and GDP is almost non-existent (0.181). This, like the previous graph, highlights Japan's GDP instability. This also helps support the fact that Japan was and is still struggling to recover from the war. Another interesting point on the graph is the low point of emissions in 2009, which almost matches the all time low emissions of 2019. The sharp decline in 2009 was led by their plan to reduce emissions called the Kyoto Protocol. Even though they were shifting most of their energy use from nuclear to fossil fuels (Reuters, 2011).

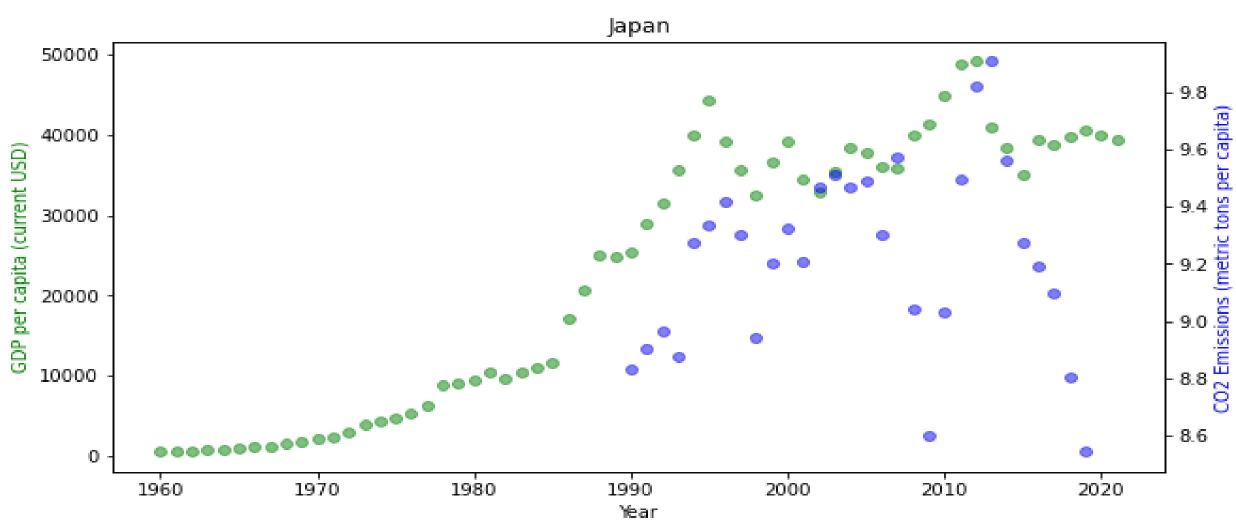

Figure 14. *Scatterplot of Japan's GDP and CO_2 emissions by year.*

Closing Discussion

Moving forward, testing could be taken even further. We could implement additional statistical tests such as covariance testing to observe the fluctuations in one variable when the other changes, or perhaps collinearity testing. An interesting concept that would have been exciting to implement would be using machine learning techniques to build models off of our data analysis and attempt to predict different results given testing data.

Throughout our research process comparing gross domestic product and carbon emissions, we explored several hypotheses on international, regional, and individual scopes, and were able to reach conclusions regarding each one. We discovered interesting relationships, some that were expected and others that surprised us. With our analysis completed, we have compiled a cache of information shedding light on a select few correlations we observed. However, climate change is a dynamic problem that is always in flux. There are dozens if not hundreds of different relationships to analyze and explore, with new data coming to the forefront each passing year.

References

Ram L. Ray, Vijay P. Singh, Sudhir K. Singh, Bharat S. Acharya, Yiping He,

What is the impact of COVID-19 pandemic on global carbon emissions?,

Science of The Total Environment, Volume 816, 2022, 151503, ISSN 0048-9697,

https://doi.org/10.1016/j.scitotenv.2021.151503.

Bekaliyev, A., Junissov, A., Kakimov, Y., & Poulopoulos, S. (2021). Evaluation of decoupling of GDP and CO2 emissions in EU-15. IOP Conference Series. Earth and Environmental Science, 899(1), IOP conference series. Earth and environmental science, 2021, Vol.899 (1).

Cederborg, Jenny, and Sara Snöbohm. *Is There a Relationship between Economic Growth and ... - Diva Portal*. 2016, https://www.diva-portal.org/smash/get/diva2:1076315/FULLTEXT01.pd.

Inumula, K. M., & Deeppa, K. (2017). Cointegration between Energy Consumption, CO2 Emissions, and GDP in India. Jindal Journal of Business Research, 6(1), 1–13. https://doi.org/10.1177/2278682116684268

Caporale, G., Claudio-Quiroga, G., & Gil-Alana, L. (2021). Analysing the relationship between CO2 emissions and GDP in China: A fractional integration and cointegration approach. Journal of Innovation and Entrepreneurship, 10(1), 1-16.

BUREAU OF RESOURCE MANAGEMENT. “East Asia and Pacific Region.” *U.S. Department of State*, U.S. Department of State, 21 Apr. 2011, https://2009-2017.state.gov/s/d/rm/rls/perfrpt/2010performancesummary/html/163043.htm.

Rivera, Alfredo. “Preliminary US Greenhouse Gas Emissions Estimates for 2021.” *Rhodium Group*, 10 Jan. 2022, https://rhg.com/research/preliminary-us-emissions-2021/#:~:text=As%20a%20result%2C%20progress%20in,2025%20and%202030%20climate%20targets.

Environmental Protection Agency. “Climate Change Indicators: U.S Greenhouse Gas Emissions.” *EPA*, Environmental Protection Agency, 2022, https://www.epa.gov/climate-indicators/climate-change-indicators-us-greenhouse-gas-emissions.

Reuters Staff. “Japan greenhouse gas emissions hit record low in 2009/2010” Reuters, April 2011, https://www.reuters.com/article/uk-climate-japan-emissions/japan-greenhouse-gas-emissions-hit-record-low-in-2009-10-idUKTRE73P0TA20110426

EIA, “What is U.S. electricity generation by energy source?” November 2022, https://www.eia.gov/tools/faqs/faq.php?id=427&t=3

The Nature Conservatory, “What is your carbon footprint”, 2022,

https://www.nature.org/en-us/get-involved/how-to-help/carbon-footprint-calculator/