World Bank Analysis

Pablo Gomez

2022-11-09

Introduction

Leading into the process of writing out the report for my code I first needed to identify the data I would use for my analysis, specifically what analysis I could fulfill using the World Bank website. I have always been interested in finding out the information regarding GDP of a given country and the relation with environmental impacts. To have a *fair* or proportional analysis it is important for me to use data that included per capita, or per **Person**. Also I wanted to see how trade or how globalization has an impact on the environment. Therefore I chose the three variables of:

1. GDP Per Capita
2. CO2 Emissions Per Capita
3. Net Barter Terms Trade Index

I downloaded the relevant CSV files to read into a R script file. Inside the R script itself, the raw data itself from the *World Bank* website itself was in wide format and not in the most optimal position for data analysis. Therefore using manipulations we cleared out redundant columns in the data frame and converted the data into long format to further analyze more effectively.

Variables

1. GDP Per Capita GDP Per Capita measures the respective country GDP(*Gross Domestic Product*) per **Person**. This variable is good on the surface level to evaluate how much wealth a country contains or produces per proportional indices. The units are in USD.

2. CO2 Emissions Per Capita CO2 Emissions Per Capita measures the respective amount of Carbon Dioxide(**CO2**) emitted per person per year in the corresponding country. The units are in kilograms(**kg**).

3. Net Barter Terms Trade Index Net Barter Index is a arbitrary variable in purpose to measure the relative country's exports and import. The higher the score the less open to trade or integrated the country is.

Munging

Steps needed to properly format data for practical use:

1. Find the 3 Variables chosen to investigate and analyze more as explained earlier
    - GDP Per Capita

- CO2 Emissions Per Capita
- Net Barter Terms Trade Index

2. One of the variables needed to be either population or GDP Per Capita. **GDP Per Capita** was chosen

3. Other two chosen variables were **CO2 Emissions Per Capita** and **Net Barter Terms Trade Index**

4. Can not use lab shown variables of **electricity** and **secondary education**

5. Getting the data into a tidy format various processes were used using pipes:

   - Read CSV file using **read_csv** object with skip conditions since World Bank spreadsheets had non formatted information on the top of the spreadsheet
   - Eliminate irrelevant columns using **Select(-)**
   - Rename Columns to ease further data manipulation using **rename()** function
   - Specifically rename a pattern of years that started with 'X' using **rename_with()** function
   - Format overall data from wide format into long format using **pivot_longer()**
   - Change data type of year into integer for easier data manipulation using **mutate()**
   - eliminate any null containing data using **drop_na**

6. **Log10** transform variables to clearly display the significance of levels in magnitude. Since a lot of the data, especially containing GDP and CO2 emission calculations, is heavily skewed right, on a graph the output is not aesthetically pleasing. Therefore a log transformation does not change much fundamentally the data but helps display relevant information.

7. Center and standardize all variables for easier graph display for clear message intention. Centering and standardizing is use to transform data to further help the curve resemble a normal distribution when looking at the histogram of the variable. Therefor we use following:

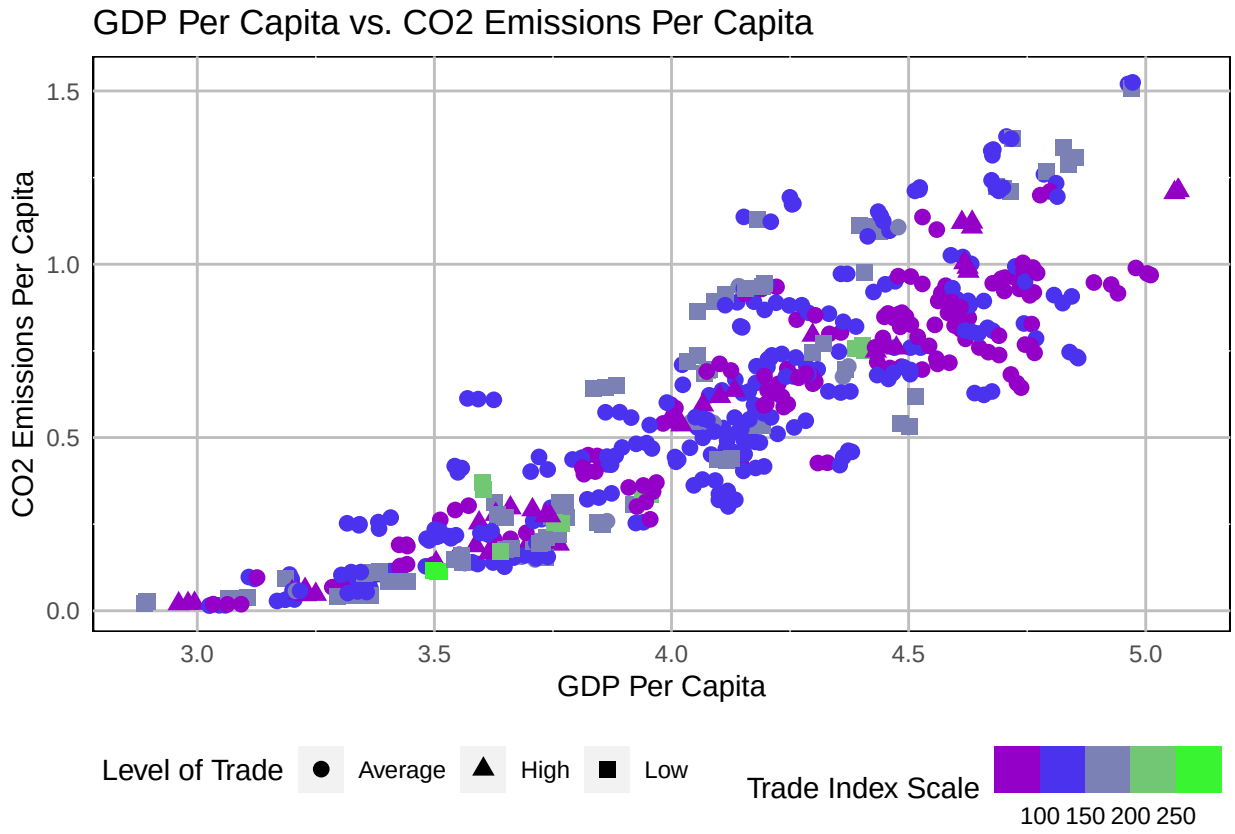
$$Z = \frac{x - \overline{x}}{\sigma}$$

   - $\overline{x}$ - Mean
   - σ - Standard Deviation

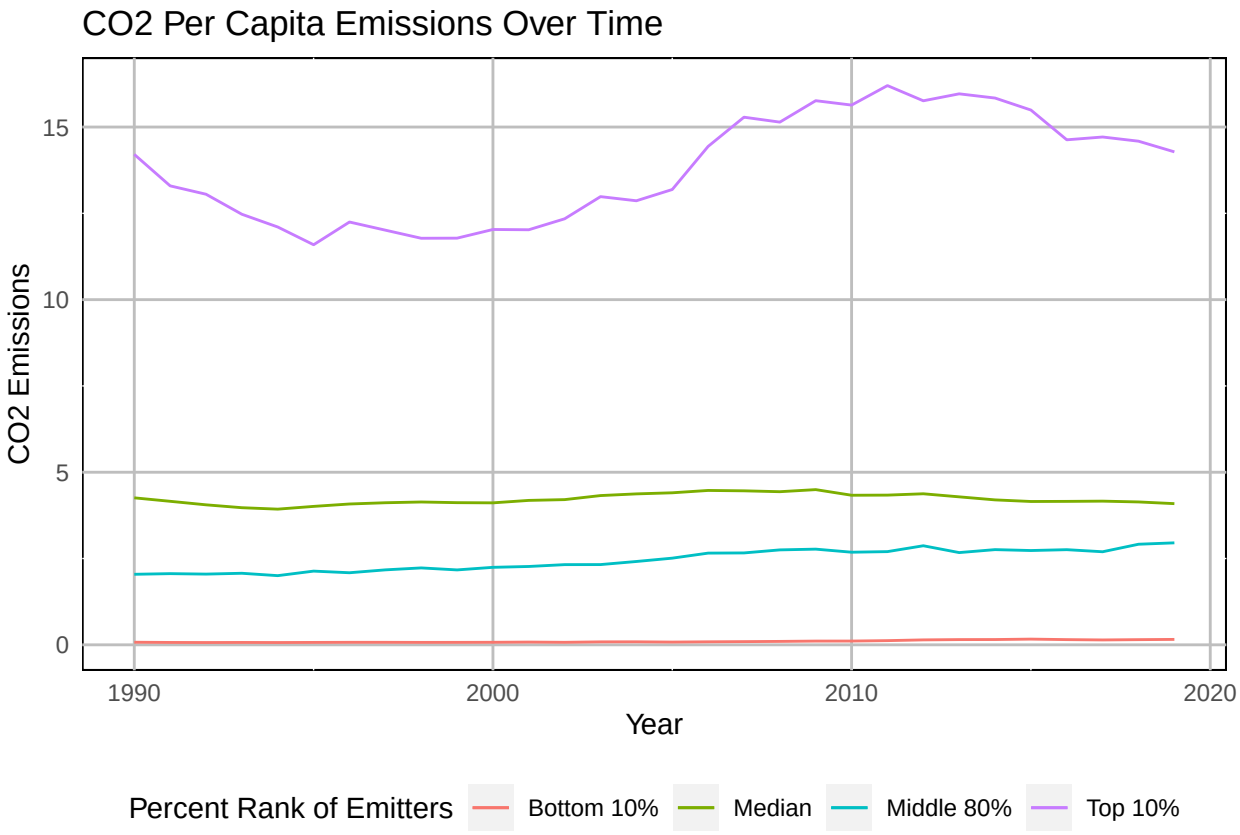
8. Merge all data frames together using **Joins** by matching one data fame with the others corresponding *Country*,*Country Code*, and *Year*.

9. Display data using **print(data)**

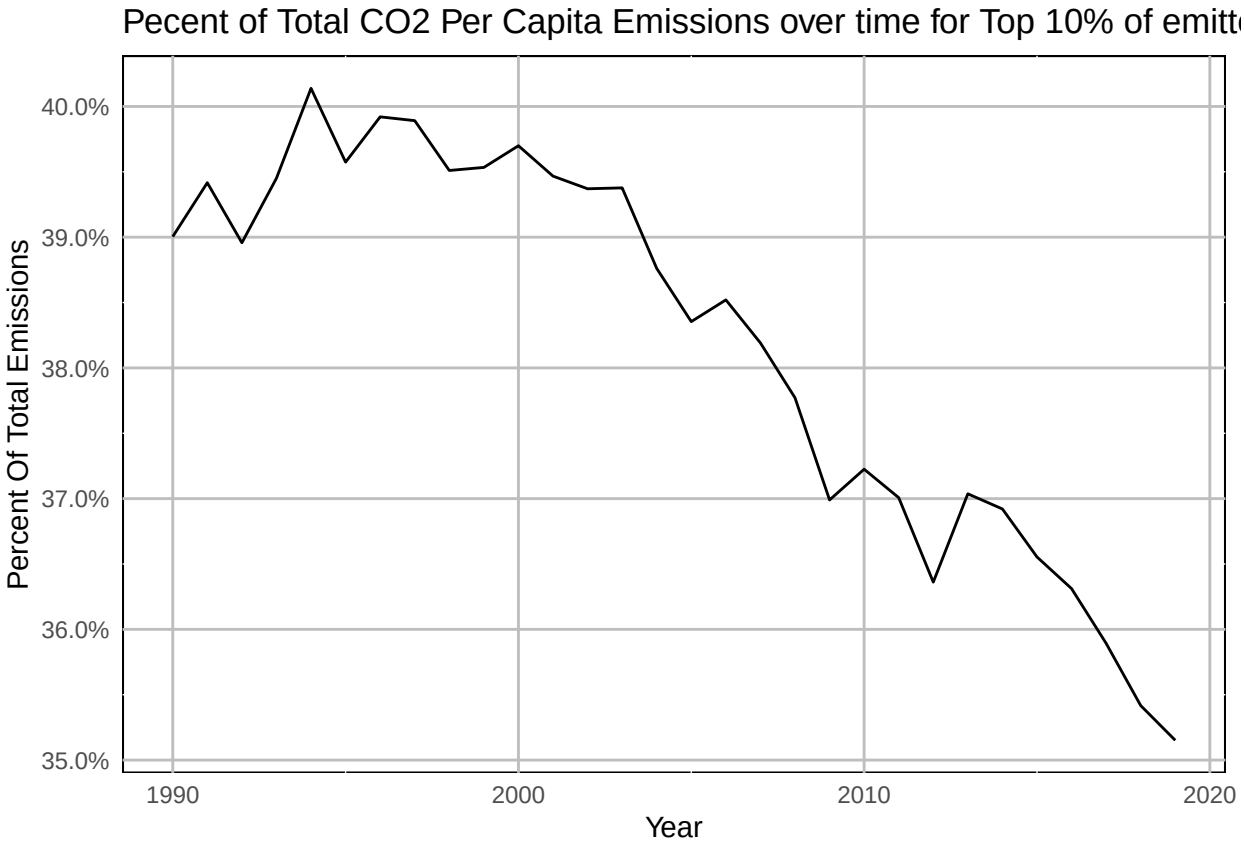
Findings



- We can clearly see a correlation between GDP and CO2 emission. With the more increase in GDP, the more in the increase in CO2 emission.
- The graph also tries to illustrate a relationship between trade with GDP and CO2 emissions. There is no clear relationship with the Trade index variable.



- We see in this visualization an illustration of CO2 emissions ordered by levels of magnitude in ranking percentile of polluters.
- We clearly see the inequality of how much the top 10% of polluters emitted CO2 compared to the others.



- We see in this visualization an illustration of the percentage of the total emittions of a given year of the top 10% of emitters over time.
- We want to identify a relationship in how much the top emitting countries contrbute to the overall CO2 global emittions.
- The trend we clearly see is that starting from the 1990's the top 10% of emitting countries have decreased in overall share of Global pollution. Noticably is that the share of the total polution is still really large as depicted in the previous graph.

Literature

Finding Scholarly Works we see the following that associate the findings of above:

1. The first article by Najid Ahmad: https://www.sciencedirect.com/science/article/pii/S0360544216319119, describes an economist theory of Kuznets curve and how that describes the relationship between income and environmental impact. The Kuznets curve argues that the increase in GDP growth increases environmental damage, but inversely overtime, there is a certain economic development point to which the more increase in GDP actually improves environmental impact. In the article, the use of Croatia as an example that supports this Kuznets curve validifies it.
2. The second article finds a second analysis of a similar relation to the Kuznets curve in Brazil during the mid to late 2000's: https://www.sciencedirect.com/science/article/pii/S0921800911000838. Luciano Charlita de Freitas found that the variables moved away from each other than in a parallel manner when the GDP of the country increased. The article contributes reforms in the energy sector as to reasons why the effect took place and to why privatization of energy sectors lead to the change.

Conclusion

In this analysis we try to find a correlation between GDP and CO2 emissions. In the pre-analysis stage it would have made sense to assume the more increase in GDP results increase to CO2 emissions. The first visualization illustrates a exponential relationship. As you look into it more the observation is not as black and white as it appears. Yes the top emitters have significant levels of magnitude to the reset of the world when it comes to CO2 emissions but a percent of the overall analysis finds the share of the total emissions is falling. Developing countries are now increasing the CO2 emissions as they develop their respective economies. Nonetheless the top emitters carry a lot of the responsibility in emissions since they clearly propagate this inequality in energy consumption. A Kuznets curve can find that after a certain level of development, the top emitters can reduce the emissions using the capital they have amassed over time.