GDP Impact on COVID-19 Written Analysis

By: Jenny Bui, Hoangyen Cao, Josh Pardo, Jonathan Rivera

Introduction & Initial Hypothesis

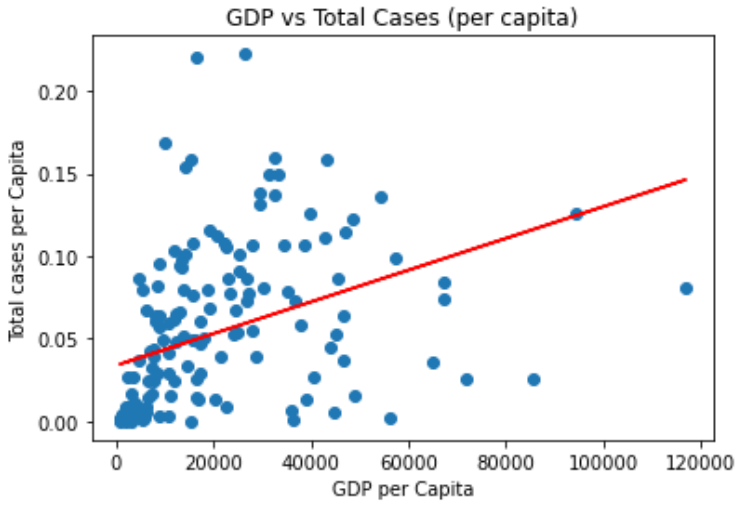
COVID-19 is a global pandemic that has truly shaken up life as we know it, and its effects can be felt to this day. Countries across the world have seen its impact and destruction, and therefore data on this virus is truly massive and widespread. Our group decided that by working with data that is so immense and dynamic, we would challenge ourselves to find relevant information that can be truly useful to both this course and to working with real-world data in general. Our main focus of this data analysis is to see if we could use COVID data to find whether a country's wealth level, measured by GDP, can impact the effects of the virus. We hypothesized that the more wealthy the country, the better equipped they'd be to deal with effects of COVID-19.

Data Sourcing

For our research, we collected COVID-19 data through a Kaggle source dataset developed by Deep Shah. We downloaded the CSV file from the dataset to run into our Jupyter Notebook, then proceeded to clean our data in order to obtain only the columns that related to our preferred analysis of GDP. We reduced our dataset from the original 65 columns to 9, then grouped our data by continent.

GDP/Capita vs Total Cases/Capita

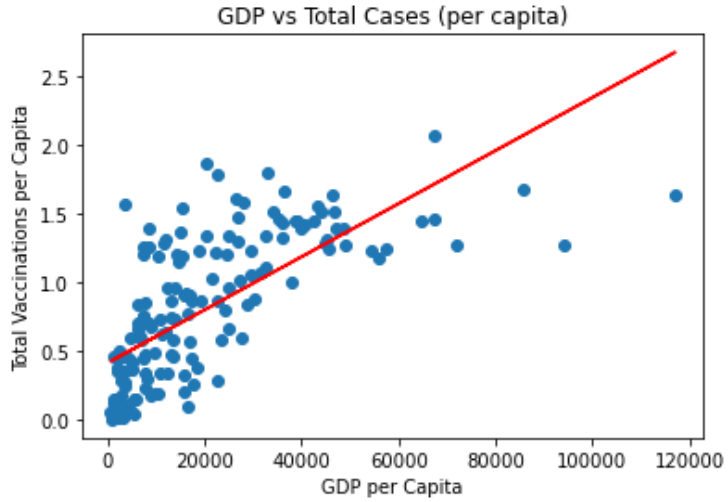
Our first research question asked whether a country's GDP per capita has an effect on total cases of COVID. To dive deeper into this question, we had to manipulate our data by first extracting the columns we needed: location, GDP per capita, total cases, and population. We then performed a regression analysis by using the "linregress" module, analyzing a possible association between GDP per capita and COVID cases per capita, dividing total cases by population for scaling purposes. We then plotted a scatter plot, with GPD per capita as our x value, and cases per capita as our y value, fitting our regression line to show the trend between the data. There did seem to be a weak association, with an r-value of 0.3865.




```
The r-value is: 0.3865399384772331
```

GDP/Capita vs Vaccinations/Capita

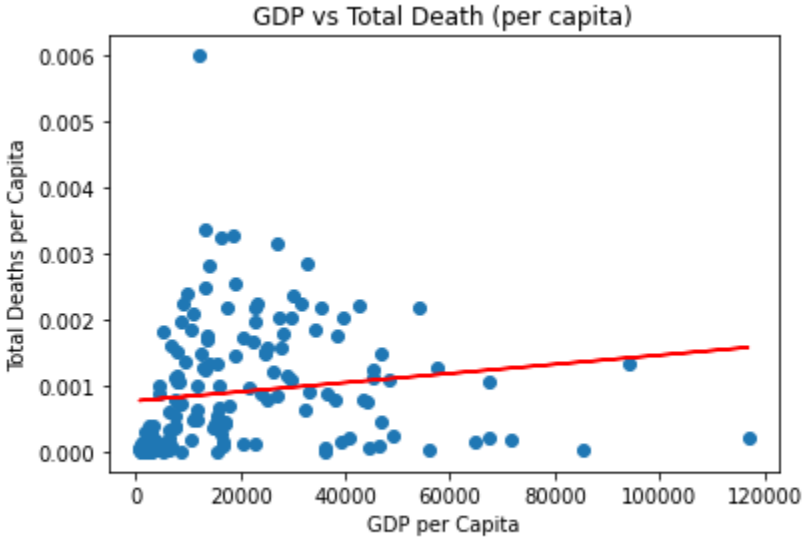
The next research question also used GDP per capita, this time trying to find any correlation between a country's GDP per capita and their vaccination rate. We used country, GDP per capita, total vaccinations, and population as our variables. Similar to our last research on GPD on cases of COVID, we plotted a scatter plot illustrating GDP per capita and total vaccinations per capita, and used a regression line to fit and show the trend line of this correlation. In the case of these two variables, we found a strong correlation between them, as countries with a higher GDP level tended to have a higher vaccination rate.




```
The r-value is: 0.7007688655670645
```

GDP/Capita vs Total Deaths/Capita

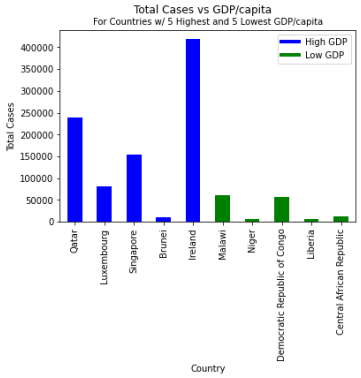
Our next study with GDP used a similar method to the previous two, this time using the total deaths column instead of cases or vaccinations. Of the three studies calculating any correlation between a country's GDP rate and COVID impact, the total deaths had the lowest correlation of them all, with an r-value of 0.1435. Therefore, it is very unlikely that GDP rate of a country has any significant impact on the death rate from COVID-19.

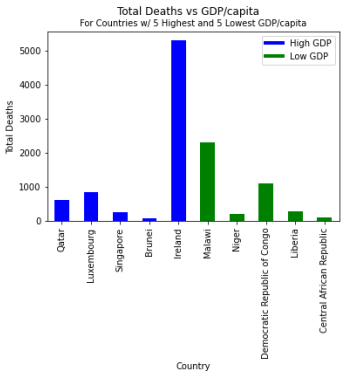


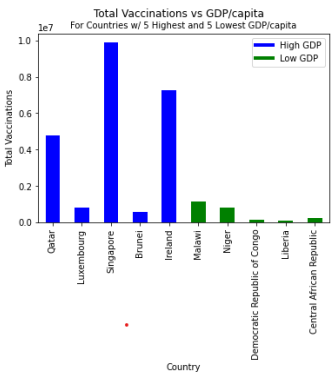

```
The r-value is: 0.14346408860619608
```

5 Highest and Lowest GPD Rates

To further break down and clean our data, we used sorting techniques to grab the 5 countries with the highest GDP rates, as well as 5 with the lowest GDP. We then plotted bar graphs illustrating the differences between COVID impact on the highest GDP's vs lowest GDP's of our dataset.







Constraints

With a dataset that is so geographically diverse and changes on a day to day basis, there are bound to be a wide possibility of constraints that could potentially hinder our data and research. One of our initial thoughts would be socioeconomic factors, and how low income areas could possibly not have the means to distribute data accurately, especially with the brutal nature of the virus. Additionally, there could be political forces within certain countries where government could alter data, or restrict certain data from being publicized. Finally, within the data itself, we realized that total vaccinations were sometimes larger than the actual population size of a country itself. We realized that this was because “total vaccinations” included both doses, with a single dose adding to the count. Therefore, data could potentially be skewed when looking at it from our GDP and vaccination visualizations.

Conclusion

Overall, we found that there definitely was an association between GDP per capita and its impact on COVID outcomes. Variables relating to COVID effects: vaccinations, cases, and deaths, were impacted by GDP, but not to the level we thought during our initial hypothesis. The strongest correlation we found was between the higher level of GDP and better vaccination rates, which we expected. However, death rates per capita compared to GDP seemed to have little association, showing that COVID deaths can not be strongly attributed to the financial well being of a country. This project was definitely challenging, but we’re hopeful that learning these skills of cleaning large datasets and manipulating data for analysis can be applied to our careers in the future.