Introduction

This is an international data exploration report focusing in and around primarily Covid 19 and Cardiovasc. The aim of this report is to provide clear and concise statistical analyses and visualisations of the aforementioned. Firstly, we will compare covid cases among eight select countries; Next, we will utilise the GDP per capita to compare and contrast the differing trend/s covid infections and deaths may have on low GDP, average GDP and high-GDP countries. Finally, we will firstly explore the Express Mortality, then we will investigate any potential correlation/s between Hospitalized Patient and Express Mortality.

Methodology

To conduct this data exploration report we will be utilising the following data variables so supply our data and data modification, to morph our data into a format which is more interoperable with other data values and variables.

Data used

- The main data set is called owid-covid-data.csv, this file contains all data which will be needed for this exploration
- The country data was taken from the “location” variable, locations were provided by Our World in Data
- The data within the “gdp_per_capita” varible was provided by World Bank World Development Indicators, source from World Bank, International Comparison Program database
- To get covid cases we use the “total_cases_per_million” variable whose data was provided by COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University
- To get covid deaths we use the “total_deaths_per_million” variable whose data was provided by COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University
- To get the cardiovasc death rate we need to use the “cardiovasc_death_rate” whose data was provided by Global Burden of Disease Collaborative Network, Global Burden of Disease Study 2017 Results
- In order to get the mortality we used two variables provided by Human Mortality Database (2021), World Mortality Dataset (2021) , they are “excess_mortality_cumulative_per_million” and “excess_mortality”
- “Hosp_patients_per_million” by National government reports and European CDC, was used to get the number of hospitalised patients per million

Data modification

- Because every row of the owid-covid-data.csv has at least one column which has one missing value we can not drop all N/A values; Hence we must instead replace them with 0. There are some problems with doing this however for this analysis we do not encounter them.
- In order to find he coloration we used pearson’s correlation, coupled with a regression plot.
- In order to change days and weeks into a month formation we have decided to sum all values for that particular month.

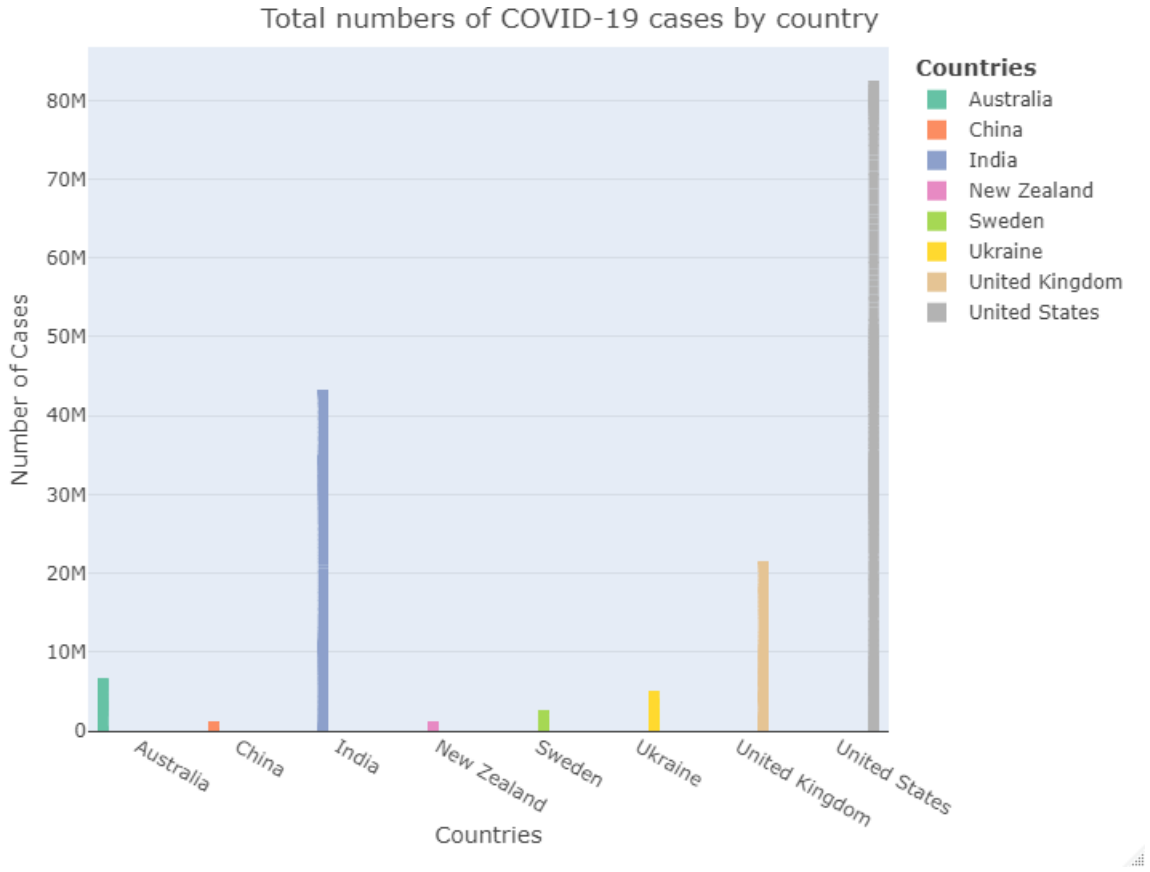
Covid 19 Cases

Numeric analysis of total cases via new cases

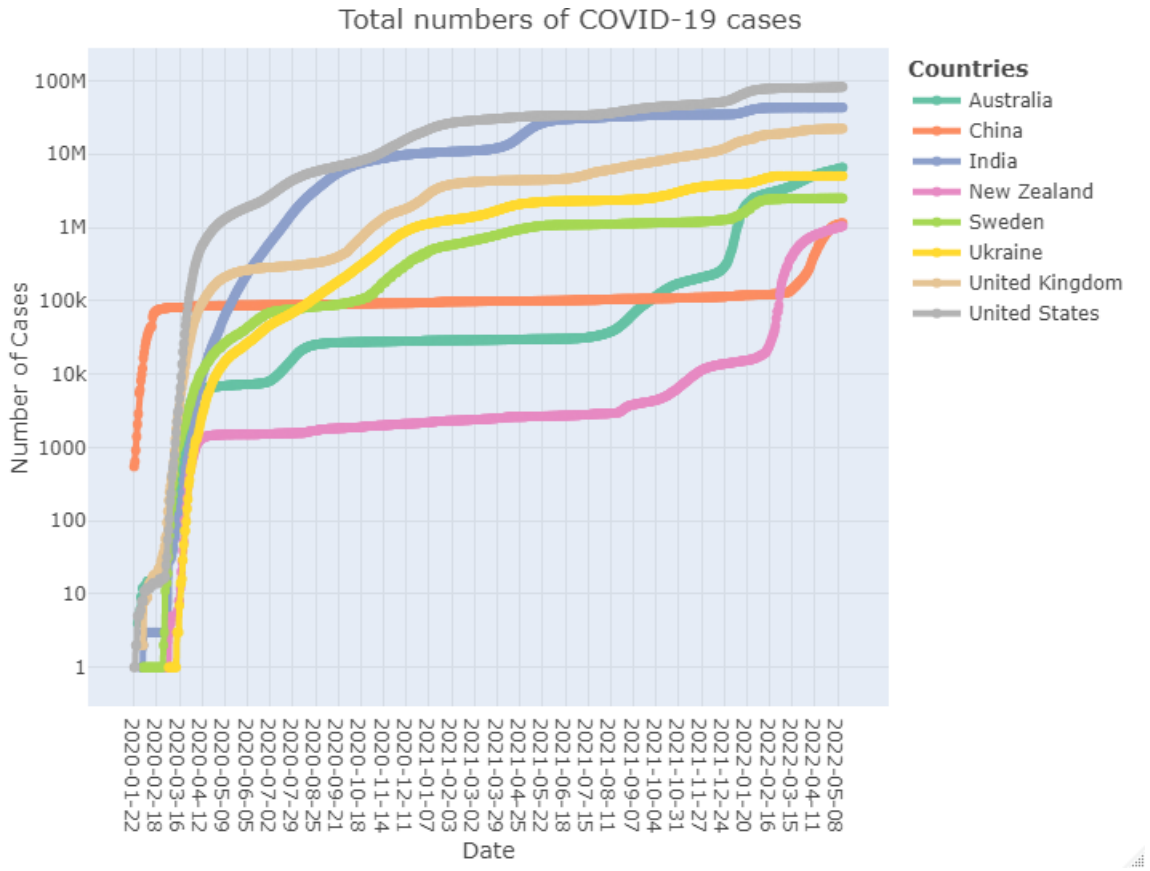
new_cases	Min. : 0
new_cases	1st Qu.: 23
new_cases	Median : 1348
new_cases	Mean : 24586
new_cases	3rd Qu.: 22801
new_cases	Max. :1383896

For the countries Australia, China, India, New Zealand, Sweden, Ukraine, United Kingdom, United States, the mean number of new cases per day is 24,586 for all countries. This means each day we can assume a country may get an average of 24,586 cases.

Graphical analysis of total cases



Above we can see a bar graph of eight countries and their total number of covid cases. The y-axis shows the number of cases and the x-axis shows the country. As we can see there is a huge gap between the united states and other nations. We can assume this disparity is due to the method and realibility of the data collection method in some countries such as china as it does not make logical sense for a country with a population of 1.3billion to have less cases than a country with 24million people.

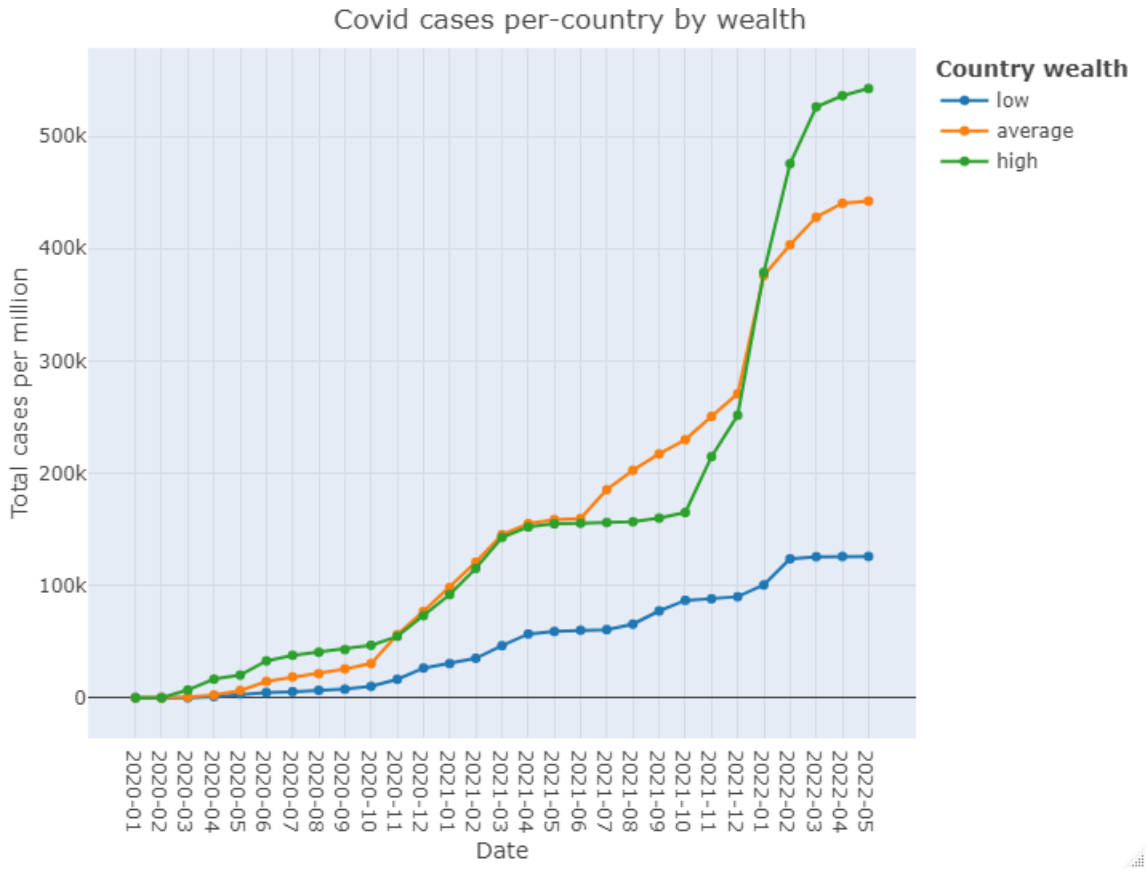


The line plot above shows the number of covid cases plotted over time, logarithmically. As we can see at the start of the pandemic, early 2020; We have a very sharp spike in the number of initial cases, followed by a leveling of he number of cases as the years go by. Interesting the number of cases recorded by China plateau's in an comparatively straight line which further goes to suggests these numbers are being artificially deflated from February 2020 – February 2022.

Country by GDP

Poor Country	Average Country	Rich Country
661.24	4449.898 - 12951.84	116935.6 +

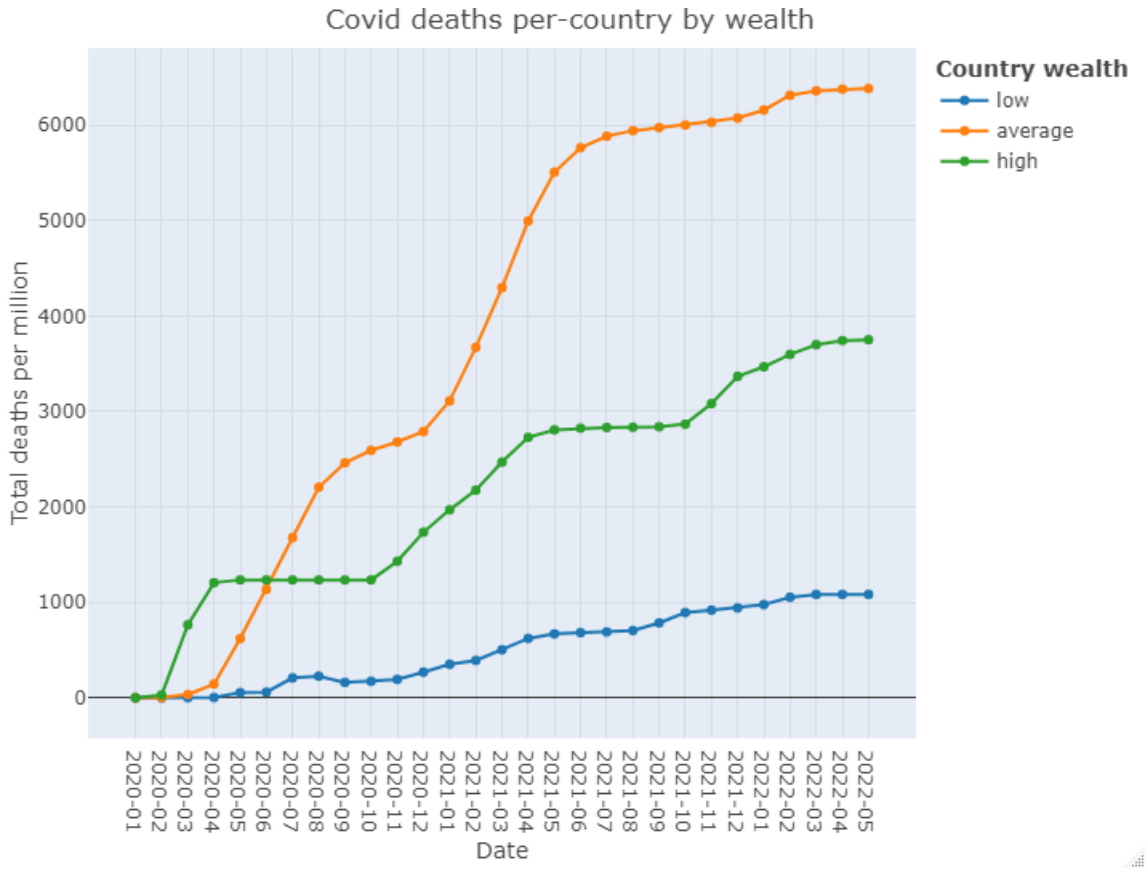
The table above shoes the quantiles for each group of country by GDP. Poor countries are those with a GDP of less than 661.24, Average countries are those with a GDP of 4449.898 - 12951.84 and GDPy countries are those with a GDP of above 116935 and above. To get these values we convert all countries GDP into a standard deviation then we extract the quantile values from said standard deviation.



The above graph shows the total number of covid cases by GDP from 2020 to 2022. The y-axis shows the number of cases per million and the x-axis shows the date by month.

We can observe an upwards trend in the number of cases which is to be expected within a pandemic. Another interesting observation is that in November-December 2021 the number of cases within high GDP countries surpassed those in average GDP countries. With this the number of cases within low GDP countries never surpassed those in high GDP countries nor average GDP countries which is interesting as intuitively we would expect low GDP countries to have the most cases.

Looking closer at the overall trend of the graph we can see that there was another point where high GDP countries and average GDP countries intersected; early 2021 was the firs time average countries over took high GDP countries in the number of cases, they maintained a small lead and even managed to increase it in June 2021 as high GDP countries cases dipped, up until the forementioned date where high GDP countries over took them.

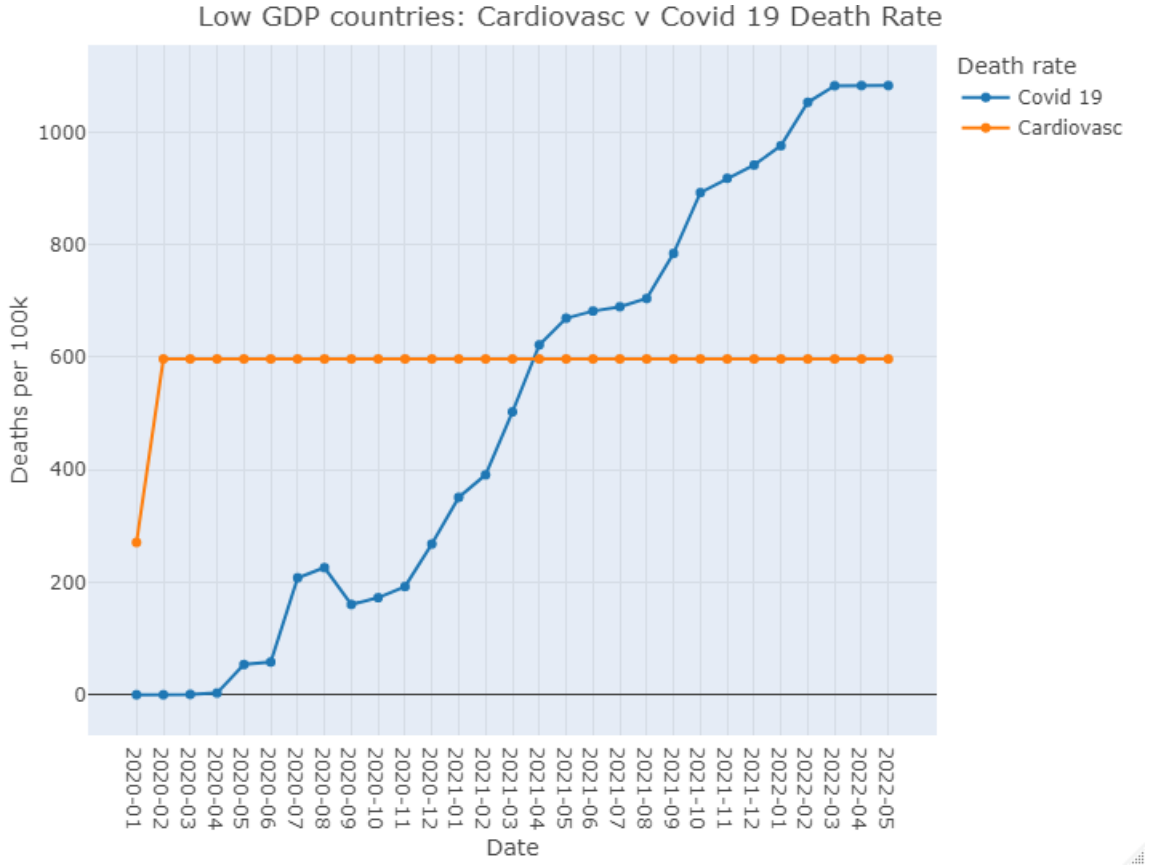


The above line graph shows the number of covid 19 related deaths from 2020- 2022, separated by a country's GDP level i.e. GDP. Overall, we can see a strong upwards trend. Unlike with covid cases the difference between countries by GDP is extremely visible here. Countries with an average GDP have not only the highest Deathrate per million people but also the fastest growing death rate per million people. After May 2020 their Death rate rose by ~137% by the end of middle 2022.

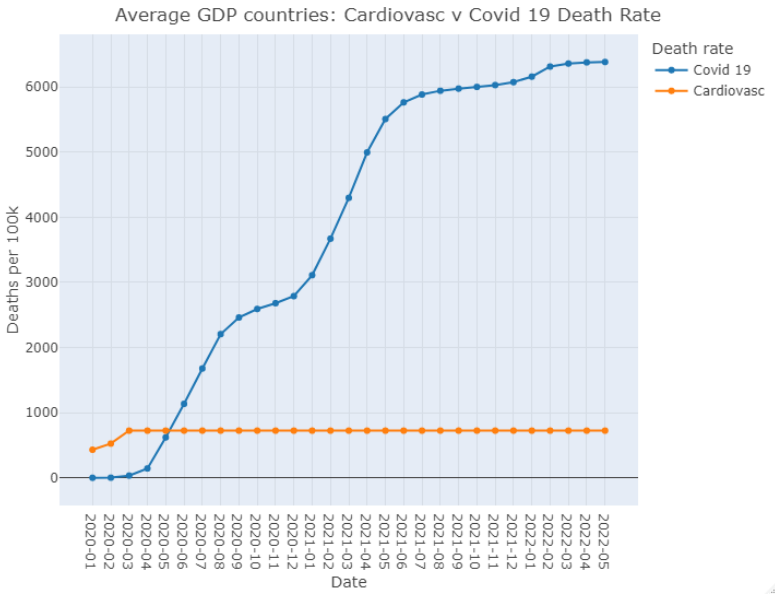
Although having comparable infection rates rich countries did significantly better than average GDP countries in terms of keeping their death rate at a relatively low number; Having ~2500, less deaths per million people.

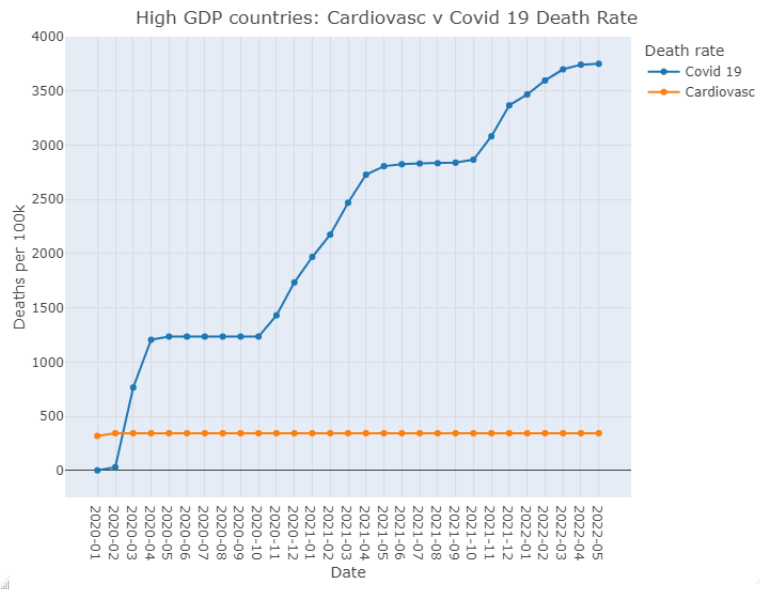
The Death rate of the low GDP countries stayed almost the same as their covid cases. Low GDP countries maintaining a low infection rate and death rate could imply that their ability to record cases is lacking.

Cardiovasc v Covid 19 Death Rate



Low-income countries have a relatively high Cardiovasc death rate compared to their covid death rate. which means of all the GDP, low GDP countries had their Covid death rate and Cardiovasc death rate intersect later into the year.

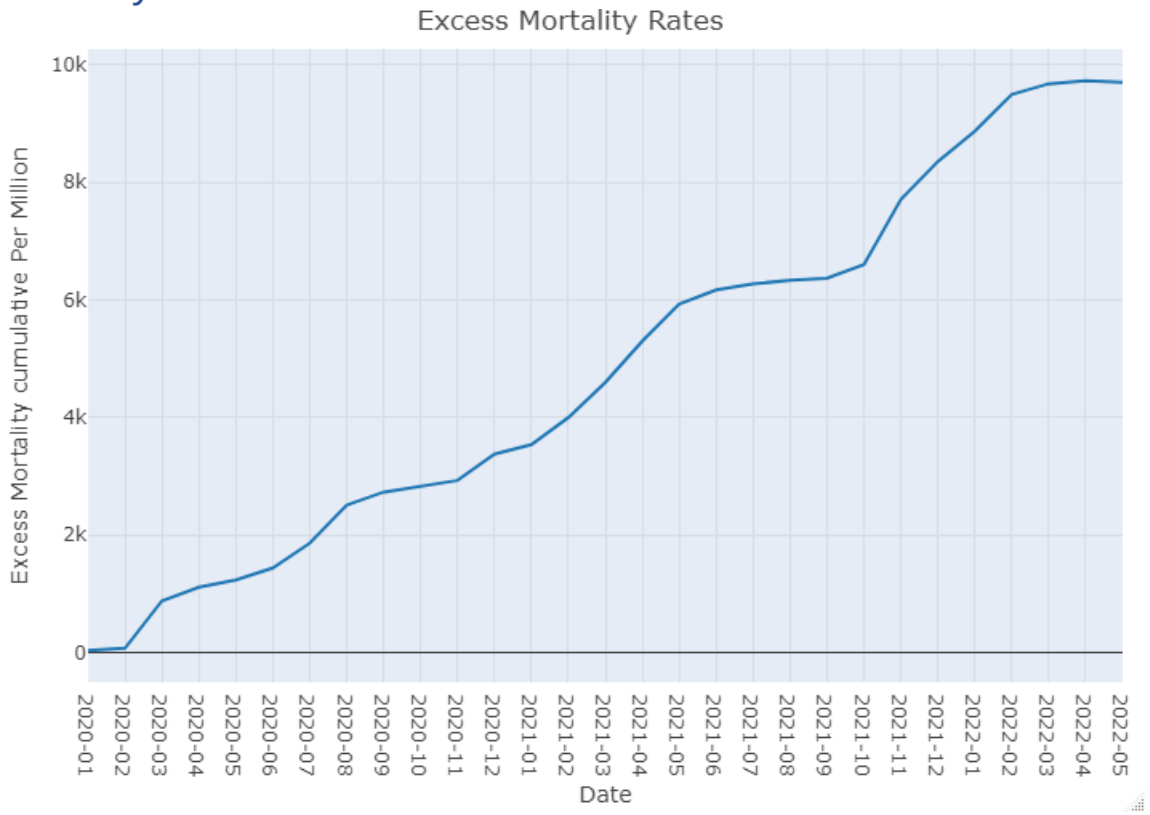




Average GDP countries and High GDP countries have a very close intersection with their Cardiovasc death rate. March – April is when they intersect. After their intersection they both rise significantly above their Cardiovasc death rate. Average GDP countries specifically have

a ~ 151% difference between their covid 19 Death rate in May 2022 as compared to the same time for Cardiovasc death rate.

Mortality Rate



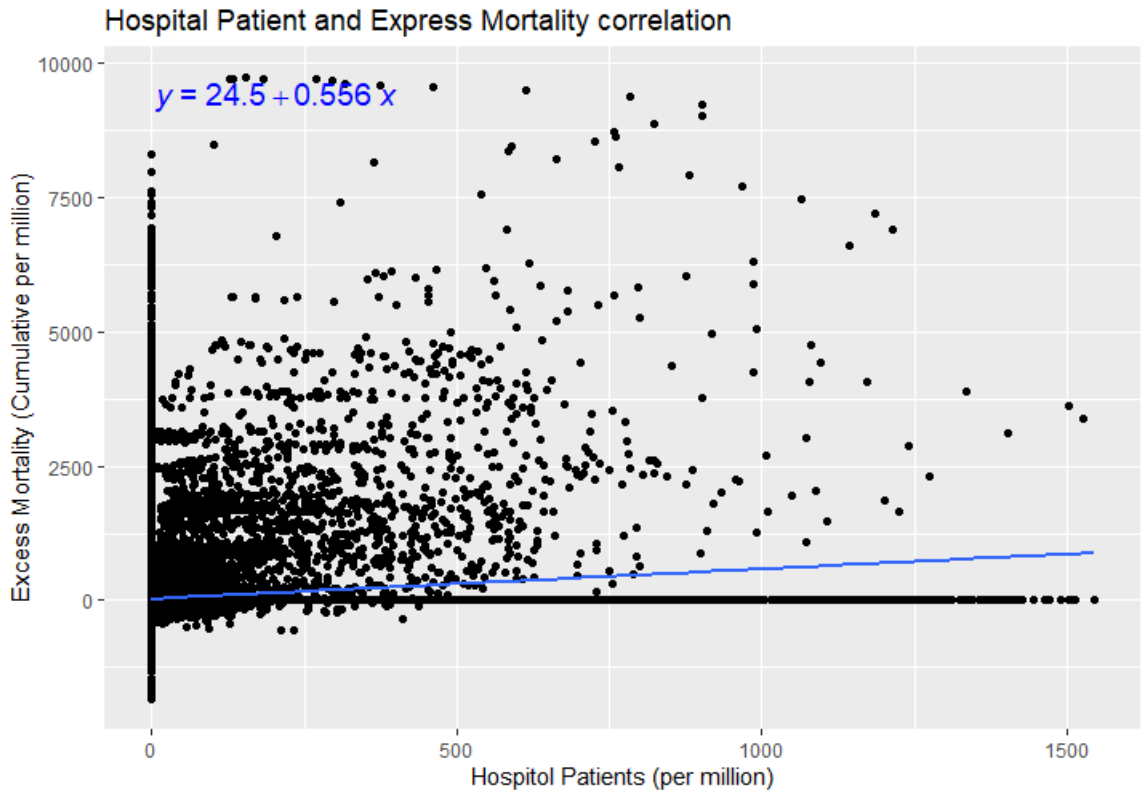
In the table above we can see the total express mortality rate per million. This graph show an almost perfect linear line, where each month from the year 2020 – 2022 the amount of deaths per million has steadily increased from 0 in 2020 to almost 10,000 in 2022.

This trend tracks well with the other graphs we have observed, as shown in the Low, Average and High GDP countries we expect an increase in the deaths from covid year over year.

Hospital Patient and Express Mortality correlation

t	P-value	95 percent confidence interval	Correlation
67.113	< 2.2e-16	0.1490787 0.1579386	0.1535117

The hosp_patients_per_million and excess_mortality_cumulative_per_million have a correlation of 0.153 which is a weak correlation; This means the two forementioned variables have a correlation however it is very weak a strong correlation would lean closer to 1.



The above graph shows the correlation between Hospital Patients (per million) and Excess Mortality (Cumulative per million). As we can see there is a weak correlation between these two variables; Meaning, as the patients in the hospital increases so does the Excess Mortality rate. The equation that governs the correlation is y = 24.5 + 0.556 x.

This relationship could be due to the fact that people already in the hospital are already at a greater chance of becoming ill from covid at the same time although a hospital may have a sizeable number of sick people it is also one of the cleanest places which may make it difficult for the virus to spread hence creating a weak positive correlation.

Conclusion

To conclude this report we have made a few key observations, firstly it seems that some of the data provided may not be as reliable as other (Chinese Covid statistics). The we surprisingly saw that countries with a high GDP had more infections nearing the end of the pandemic, how ever they also manage to keep down their death rate by a sizable percentage. Which means that there most certainly is some correlation between the number of dead and where said person resided. Next, we say that Cardiovasc is not comparable in terms of infection and lethality with covid, in a matter of months covid managed to not only infect more people than Cardiovasc, but also managed to end more lives as well. The Mortality rate trend was going up each year of the pandemic, a reason for this could have been that more people were getting infected and succumbing to the infection. Finally, we saw that there is a weak correlation between the Hospital Patients (per million) and Excess Mortality (Cumulative per million), whereas one increased so did the other.