

Analysis of Covid-19 Fatality Rate (2020)

The dataset was gathered from the Our World in Data online publication. It consists of all the daily cases, deaths for every country that had recorded cases.

Pre-Processing of Data

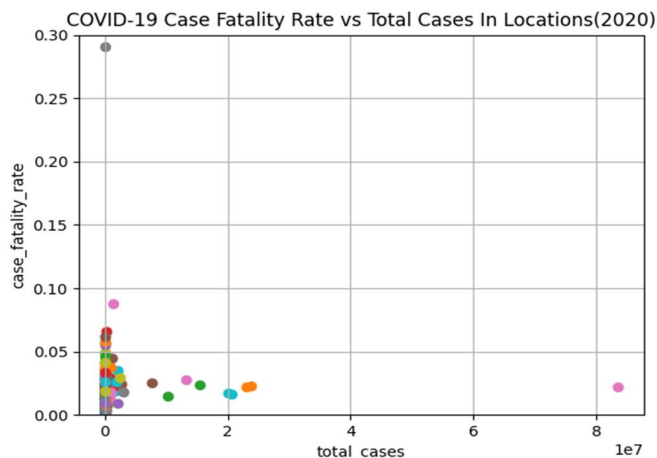
The Data was processed using Python following these steps:

1. The data was loaded in a data frame.
2. Only relevant columns were kept, i.e., Location, date, total cases, new cases, total deaths and new deaths. The total cases represent the total number of cases recorded until that date similarly for total deaths as well. The new cases refer to the cases that was recorded that day and similarly goes for new deaths.
3. Since we are interested in the data of 2020 only all the data for 2019 and 2021 was discarded.
4. Data with null values were discarded.
5. Then the data was grouped by location aggregating by Maximum total cases, Sum of new cases which should be like that of total cases, Maximum total deaths and sum of new deaths which would be equal to that of Total deaths.
6. The cases fatality rate for each location was calculated by the following equation: total deaths/ total cases for each location.

Limitations

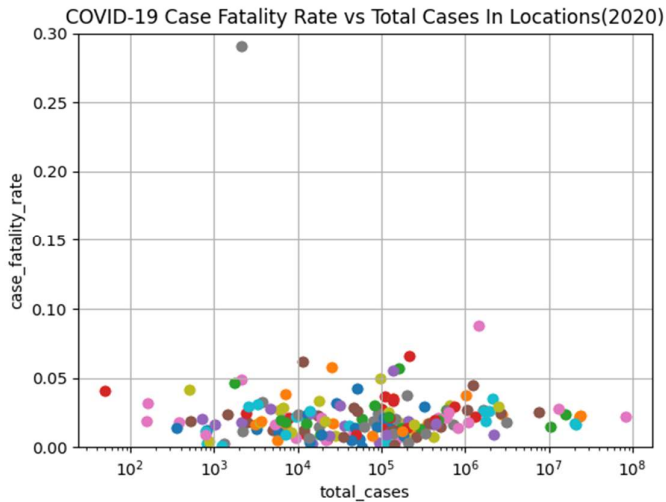
1. The case fatality rate using the above equation for each location means that if a country gets 10 cases and 2 of those did it makes a case fatality rate of 20% which is not the real case fatality rate of covid-19.
2. It combines all the age group for the fatality rate. Since elderly people and young children are more vulnerable to covid-19 they tend to bias the data and should be excluded and have their own separate analysis.
3. The number of cases is somewhat proportional to the population of the location. The fatality rate should rather be viewed globally.

Visual Analysis



In this scatter plot, most of the data points lies below the 0.05 mark which implies that the case fatality rate should lie below that value. The values above that mark could be explain by limitation 1. However, we cannot determine the mode rate since the data is not spread enough due to the horizontal axis being on an exponential scale. Furthermore, points that are

further to the right would represent countries with a very large population or continents since our data also contained data for the continents as well.



In this scatter plot more detailed can be extracted. First, most of the data points now lies below the 0.025 mark and it can be inferred that the case fatality rate of covid-19 is below 0.025. Furthermore, we can see that even with country having very high number of cases, the fatality rate is low.

Contrasting Discussion

The first scatter plot gives us a range of where the fatality might be. It does not give detailed rate for the number of total cases but a general overview.

Using a log scale for the population in the second plot, we can now have a narrower range for the rate and be more accurate. We also see the mode ratee for each power of 10 of the total cases.

Hence if we just want a general idea of where the rate is we use the first plot but if a more detailed analysis is required, the second plot must be used.