

King Khalid University
Computer Science College
IS Department
MSc in Data Science



FUNDAMENTALS OF DATA SCIENCE

Assignment - #3

Basmah Misfer Saad AlQahtani

ID: 443800986

Section No.: 404

Data Visualization

Disease Outbreak Analytics

- Coronavirus -

Coronavirus disease (COVID-19) is a contagious infection caused by a newly discovered coronavirus. The majority of patients infected with the COVID-19 virus will develop mild to moderate respiratory illness and will recover without the need for specific treatment. People over the age of 65, as well as those with underlying medical conditions such as cardiovascular disease, diabetes, chronic respiratory disease, and cancer, are more likely to acquire serious illnesses. (WHO)

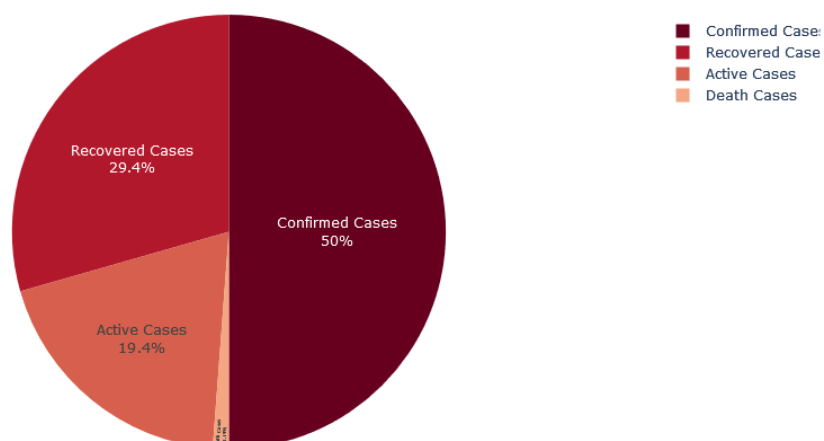
After nearly two years, we can now examine Covid-19 data, and in this report, I will depict the number of new confirmed cases, active cases, recovery cases, and deaths in many countries.

We used the dataset file (covid_19_data.csv), which provides data about countries and states around the world, the observation date, and the daily number of confirmed cases, active cases, recoveries, and deaths. We can evaluate the data over time and determine when cases are increasing or decreasing, as well as which countries are most affected by this infection.

As we all know, the importance of data visualization aids in understanding the problem, facilitates data reading, and helps in deriving conclusions after analysis through visualizing. As a result, in this analysis, I employed a variety of visualizations to better comprehend and examine the situation.

First, we can know the percentage of each cases in the world (i.e. confirmed, active, recoveries, and deaths) by using the pie graph:

The percentage of Covid-19 Cases in the world



From this graph:

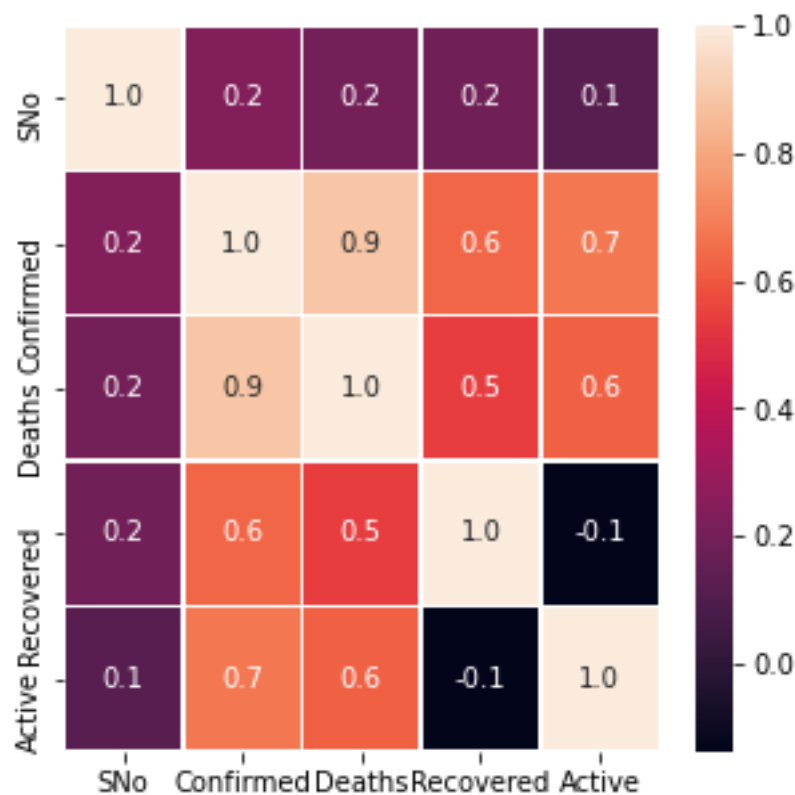
- 50% of the observations are confirmed cases greatest percentage.
- 29.4% of the observations are recovered cases.
- 19.4% of the observations are active cases.
- In this data set, we can see that the deaths rate has decreased significantly by 1.2%.

This implies that the deaths rate from Covid-19 is low in comparison to the rate of its spread, and death may occur due to the presence of numerous health symptoms, as previously stated by the World Health Organization.

In the below table, we can check out the total number of cases in the world:

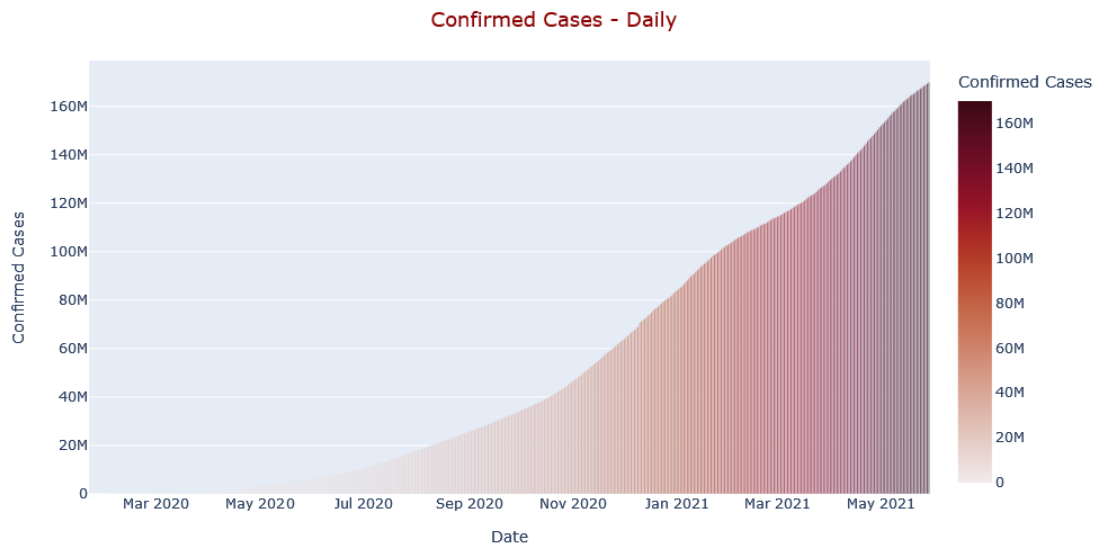
	Confirmed	Deaths:	Recovered :	Active
0	26252051758	624013017	15450237912	10177800829

Second, to study the correlation between the features of this dataset we can use correlation heatmap:



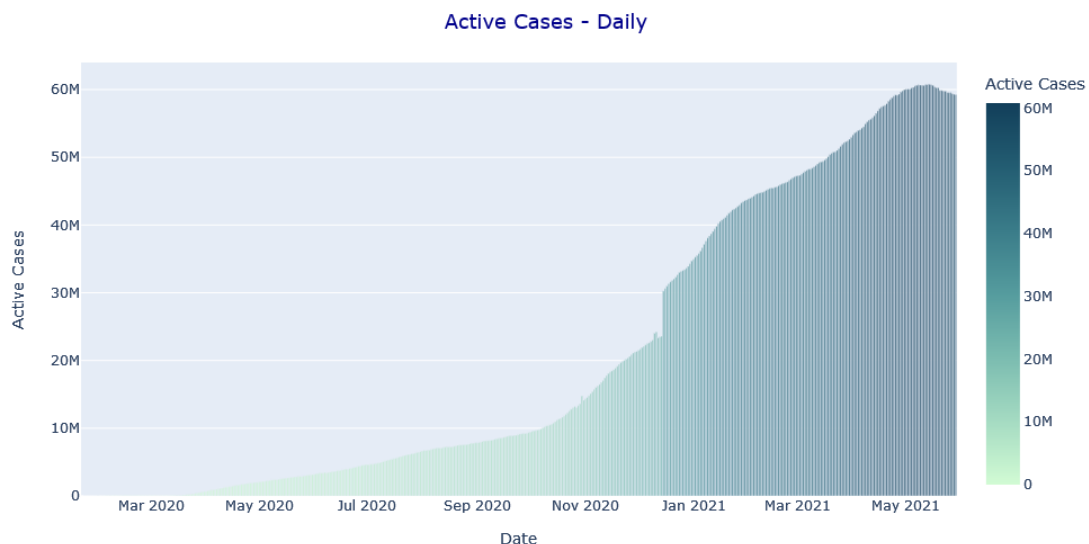
- There is a strong correlation between confirmed & deaths cases.
- There is a moderate correlation between confirmed & active cases.
- There is a moderate correlation between deaths & active cases.
- There is a moderate correlation between deaths & recovered cases.
- There is a negative weak correlation between recovered & active cases.

Then, by using the bar chart, we can identify to know the growth of the virus daily:



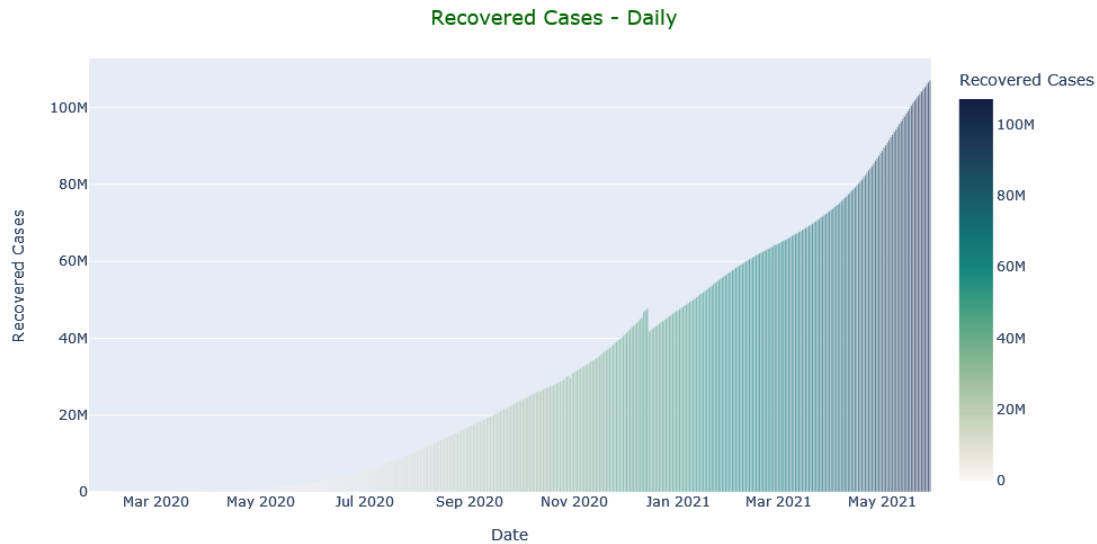
From this graph:

- We can say the start date of the virus at 27 March 2020 with 593447 confirmed case in the world.
- With the arrival of 2021, the number of cases begins to rise until it exceeds 160M at May 29, 2021.
- We can expect an increase in cases in the months following May, but we cannot be confident that this prediction is true due to the rapid development of numerous vaccinations within that period.



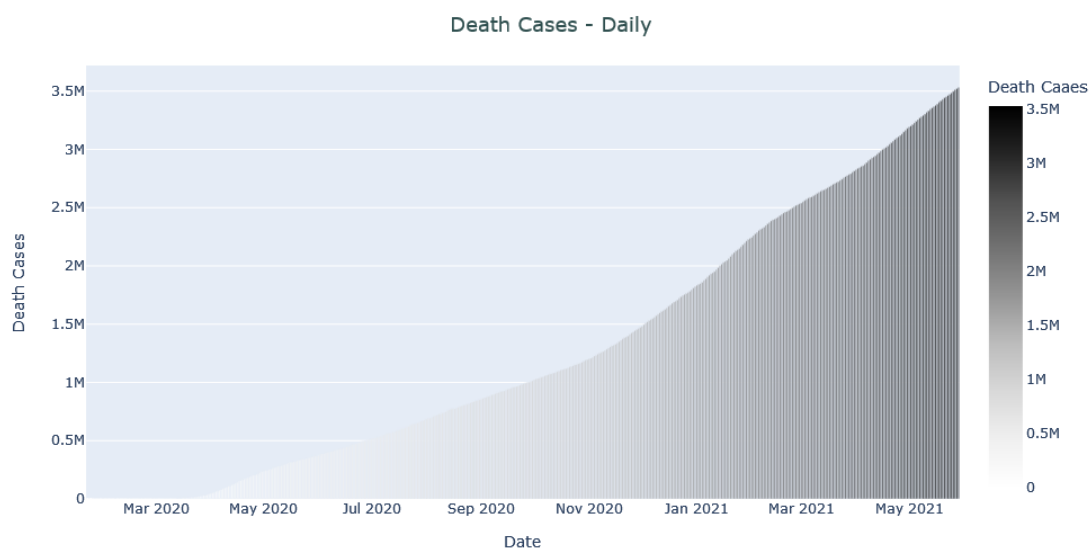
From this graph:

- With mid-December, there is a clear increase in the number of active cases, which continues until it reaches more than 60 million cases around the world in mid-May and then gradually begins to fall towards the end of May.



From this graph:

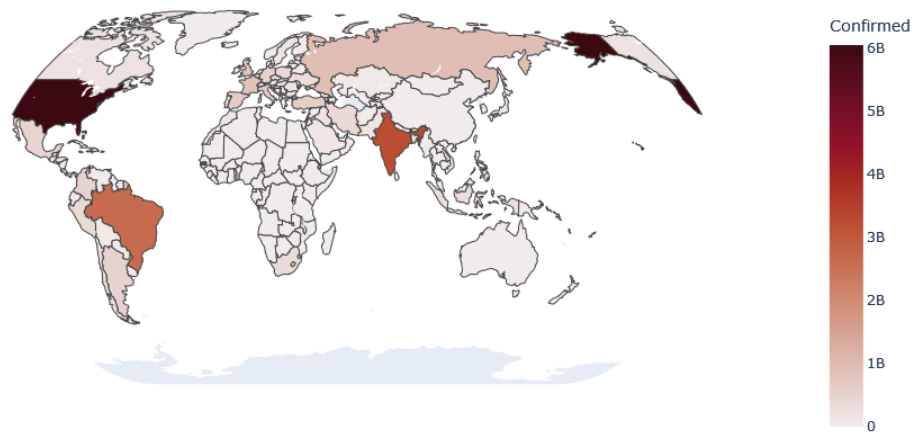
- Since the beginning of the pandemic in May 2020 until mid-December 2020, there is a steady increase in the number of recovery cases, followed by a sharp drop on December 15, following which the recovery begins to grow normally until the end of May 2021.



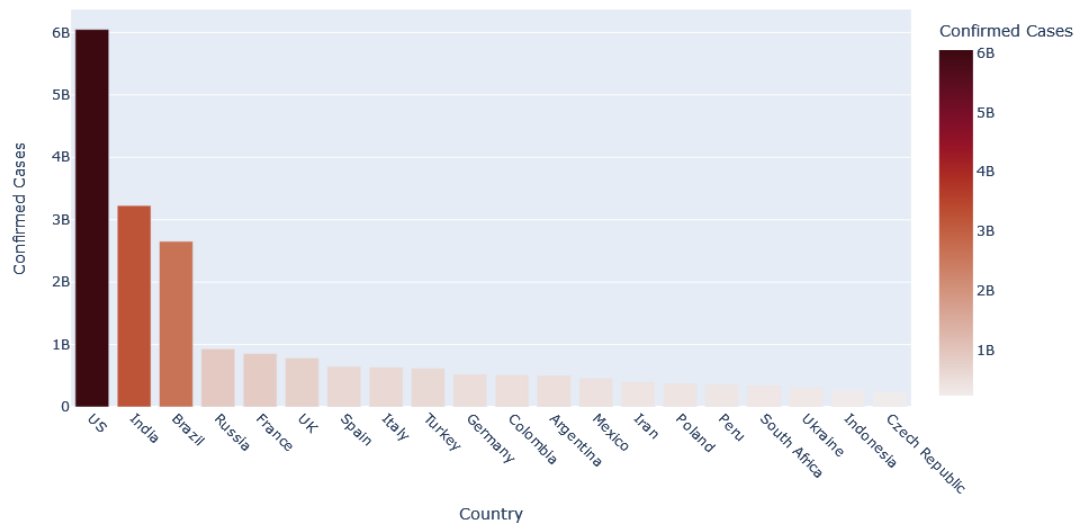
From this graph:

- Since the beginning of the pandemic in May 2020 until May 29, 2021, there is a steady increase in the number of deaths cases until it exceeds 3.5 M.
- We can expect an increase in deaths cases in the months following May 2021, but we cannot be confident that this prediction is true due to the rapid development of numerous vaccinations within that period.

World - Covid-19 Cases (confirmed)



Top 20 Countries with the most Confirmed Cases



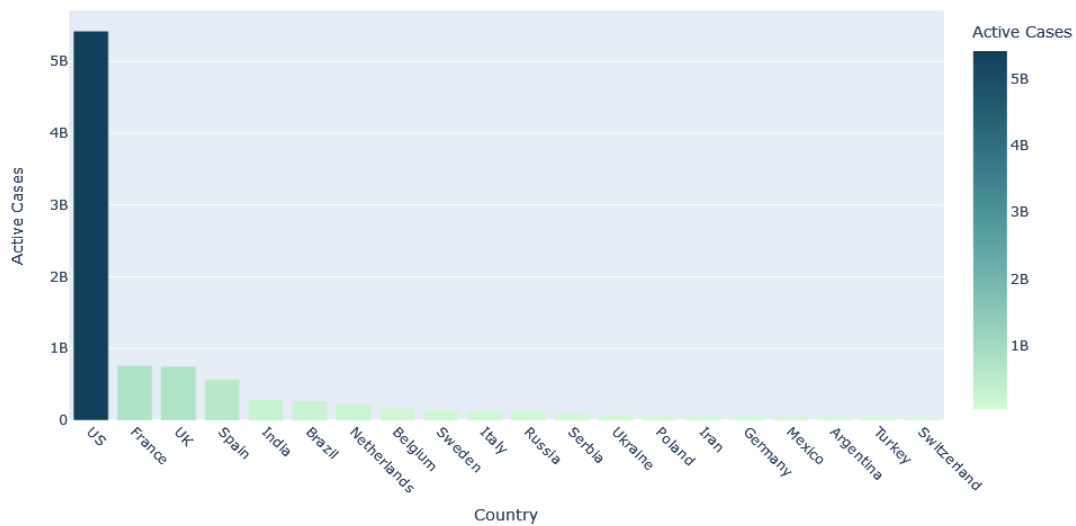
From the bar chart and the Choropleth Maps:

- We can observe that the top countries with the most confirmed cases are those with a huge population, such as the United States, India, and Brazil.
- The number of confirmed cases in the United States is clearly incomparable to any other country with any confirmed cases.

World - Covid-19 Cases(Active)



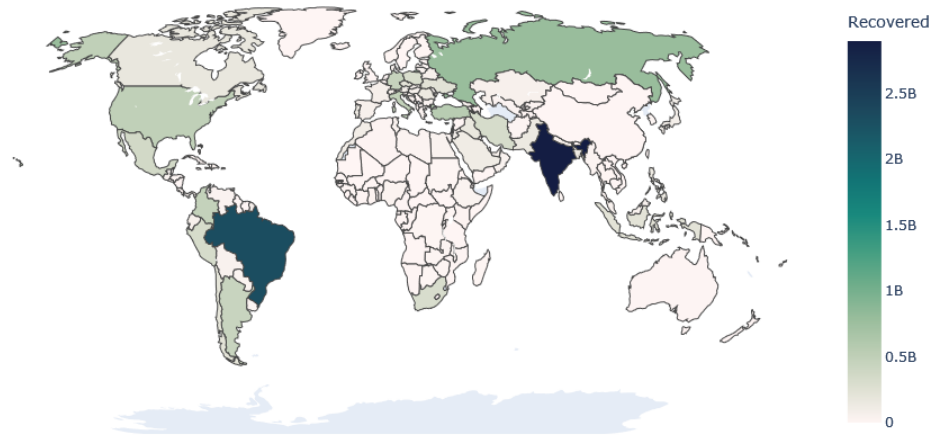
Top 20 Countries with the most Active Cases



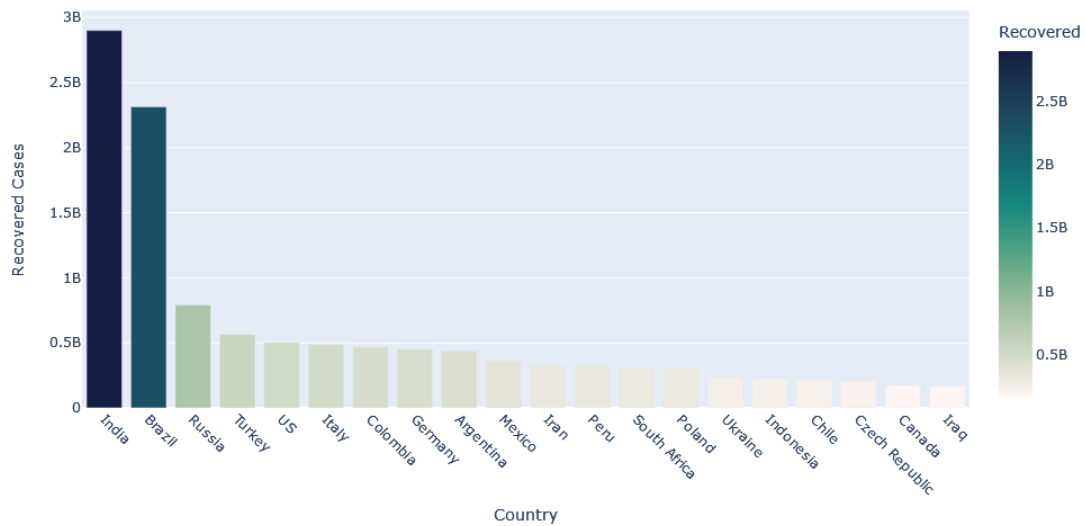
From the bar chart and the Choropleth Maps:

- The number of active cases in the United States is clearly incomparable to any other country with any active cases.
- Take note of the moderately positive correlation (0.7) between confirmed and active cases. As a result, as the number of confirmed cases increases, so will the number of active cases.

World - Covid-19 Cases(Recovered)

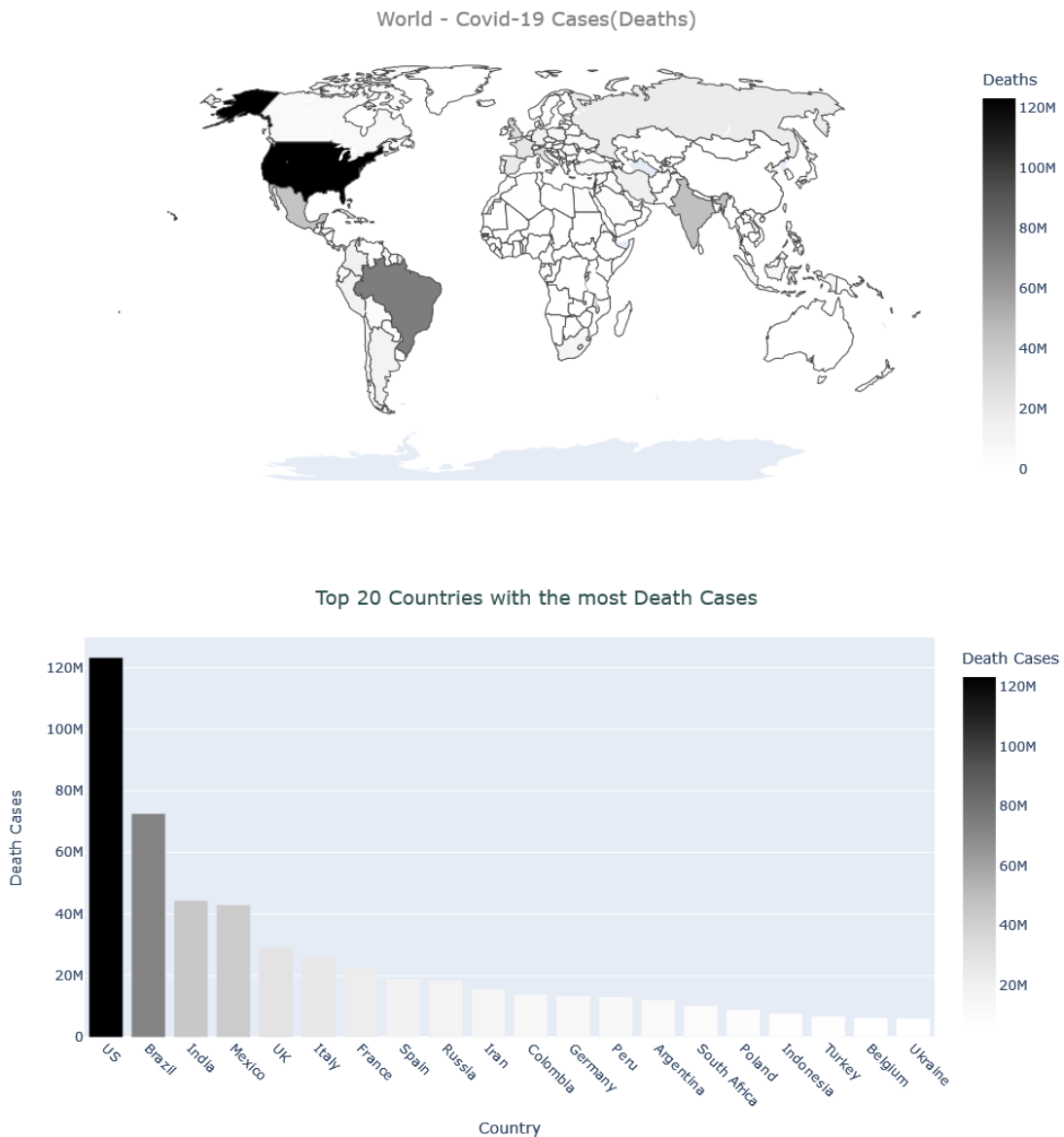


Top 20 Countries with the most Recovered Cases



From the bar chart and the Choropleth Maps:

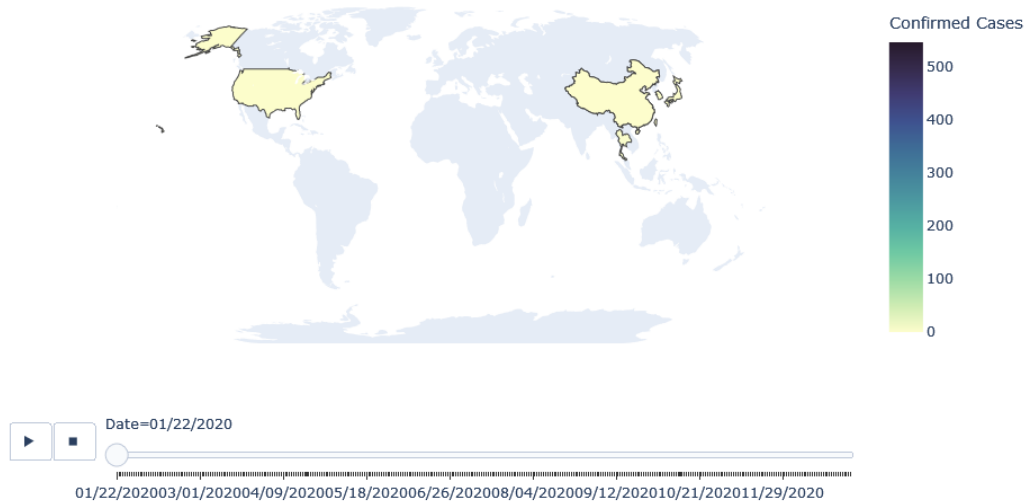
- We should emphasize that the recovery rate in India and Brazil is exceptionally high, exceeding the 2 billion case recovery threshold. In comparison, the recovery rate in the United States is average, ranking fifth.



From the bar chart and the Choropleth Maps:

- Because of the strong correlation between confirmed cases and deaths cases (0.9), we see the similar pattern here: the deaths rate rises dramatically in all three countries: The United States, India, and Brazil.
- In addition, we have noticed a dramatic rise in deaths in Mexico as compared to confirmed cases.

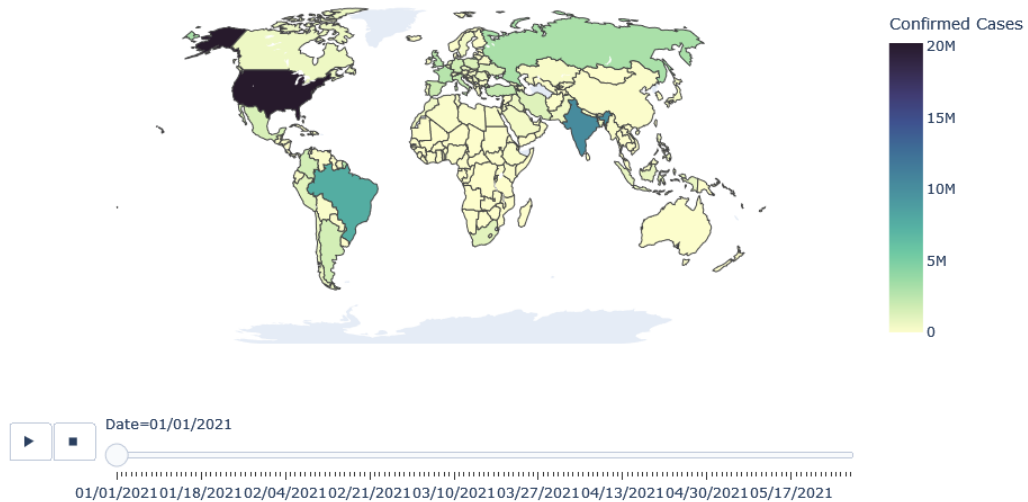
World - Confirmed Cases by date in 2020



From this Choropleth Maps (world map in 2020):

- The coronavirus made an effective spread to European countries and Iran in March 2020, most particularly Italy, Spain, and Germany.
- However, not long after, the virus spreads rapidly to the United States, which stays a hotspot of cases for the entire year.

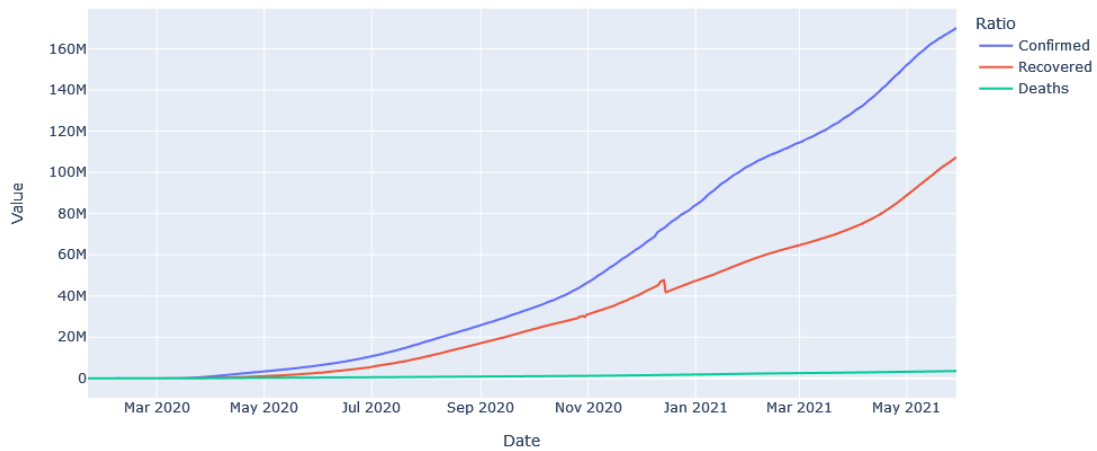
World - Confirmed Cases by date in 2021



From this Choropleth Maps (world map in 2021):

- Covid-19 is expected to spread to Brazil, India, and Russia.
- From January to May 2021, the number of confirmed cases in the United States will increase exponentially.

Confirmed cases, Recovered cases, and Death Over Time



Finally, this is a summarized representation of the previous bar graphs. We would like to see the red and blue lines intersect and pass each other.

However, by the middle of 2021, the convergence has slowed and the ratio of confirmed cases has surpassed the ratio of recovered cases.

References:

- <https://towardsdatascience.com/coronavirus-data-visualizations-using-plotly-cfbdb8fcfc3d#fc91>
- <https://informationisbeautiful.net/visualizations/covid-19-coronavirus-infographic-datapack/>
- <https://www.kaggle.com/carriech/covid-19-analysis-plotly>
- <https://plotly.com/>