Project 1

The dataset used for this project is the COVID-19 Global Dataset, which includes two subset datasets downloaded from Kaggle. After cleaning both subsets by removing null values, the data was prepared for visualization using Tableau.

For this project, my main goals were to:

1. Understand geographic spread of confirmed cases across countries,
2. Compare totals and rankings between countries based on case numbers,
3. Explore possible relationships between COVID-19 testing and death rates, and
4. Observe trends over time in daily new cases for specific countries.

To visualize these tasks effectively, I used four different chart types: world map, bar chart, scatter plot and line chart; each selected to match the nature of the task. Each visualization was styled intentionally using colors, fonts, and layout choices that enhance clarity and meaning. Overall, these visualizations help communicate the global impact of COVID-19 and allow for easier comparison and understanding of key patterns within the data.

Visualization 1:

Geographical Map showing Global Distribution of Total Confirmed COVID-19 Cases

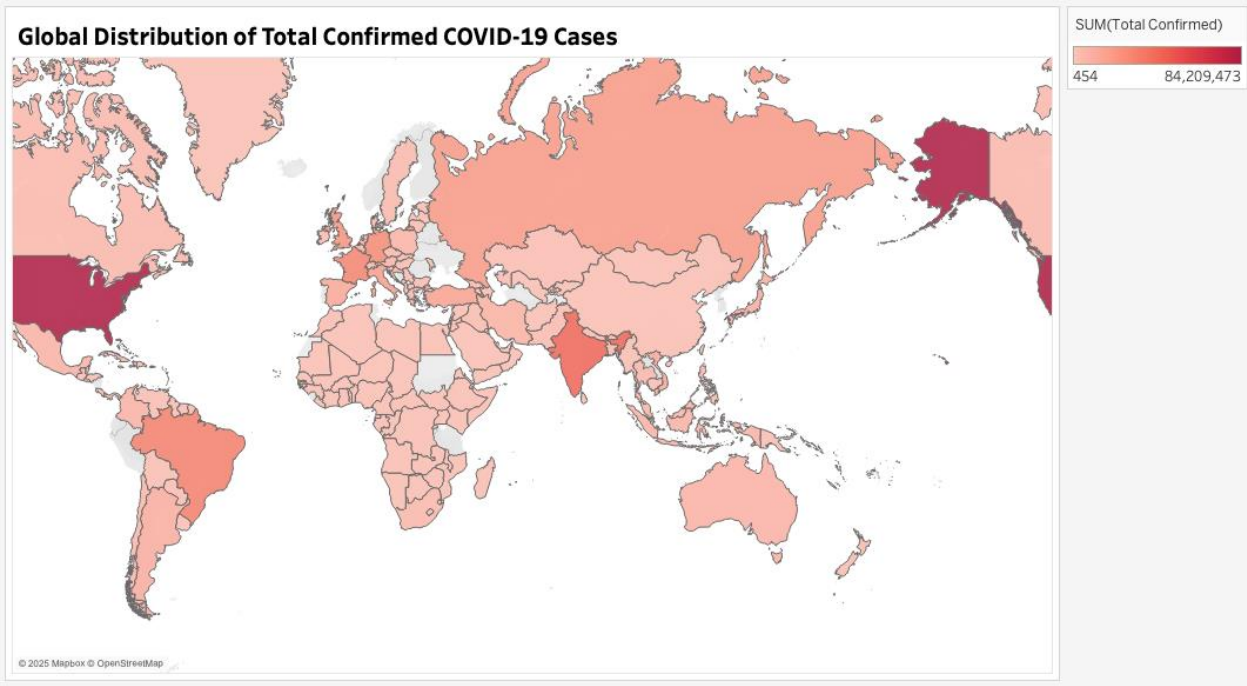

Figure 1: Global Distribution of Total Confirmed Covid-19 Cases
Link: https://public.tableau.com/views/Project1Visualization1_17520100809420/Viz1?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Reason for Choosing the Chart Type:

I chose a map because it is the most effective way to visually display global data by country. Unlike bar charts or line graphs, a map makes it immediately clear where the most and least affected areas are located. It also helps viewers understand the geographic spread of COVID-19, which is harder to grasp from tables or regular charts. A map gives the data a sense of scale and location, making it easier to interpret the impact at a glance.

Purpose of the Visualization:

The purpose of this visualization is to show how COVID-19 confirmed cases were distributed across the world by the end of 2022. It helps viewers identify countries with the highest total case counts, compare regions to see which continents were more affected and understand how the virus spread globally.

Reason for Choosing Style and Colors:

I used a shades-of-red color scale, where darker red indicates a higher number of confirmed cases. Red is often associated with health alerts, warnings, and emergencies, making it a fitting choice for a pandemic map.

The background was kept white to keep the focus on the countries themselves, and the title was written in Tableau Bold to clearly state what the viewer is seeing. I chose not to add labels directly on the map to avoid clutter, especially since many countries are small and close together. Additionally, the tooltip and color legend were included to make the map both interactive and easy to interpret.

Visualization 2:

Horizontal Bar Chart showing Top 10 Countries by Cumulative COVID-19 Cases

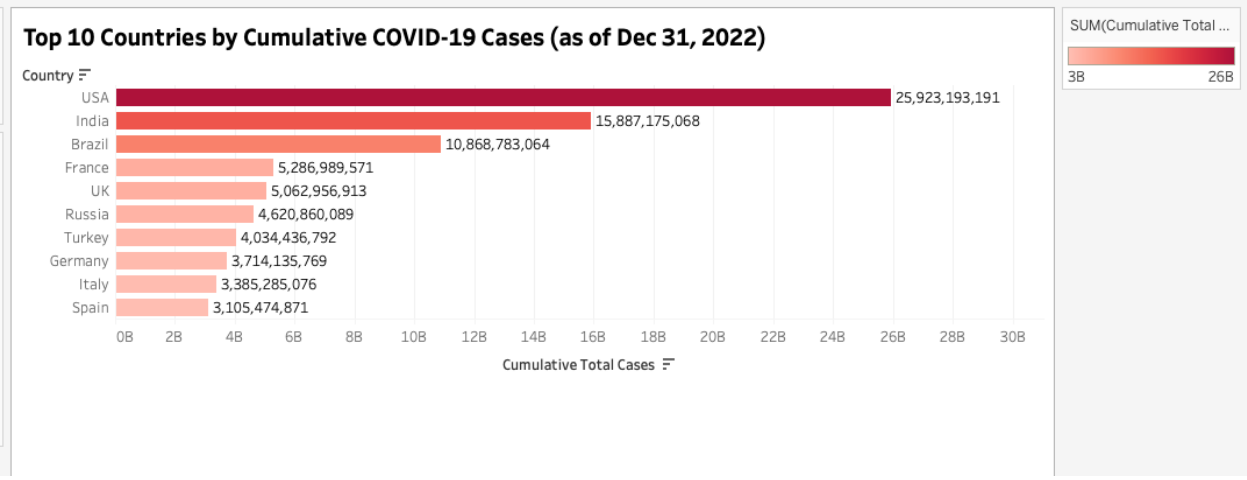

Figure 2: Top 10 Countries by Cumulative COVID-19 Cases
Link: https://public.tableau.com/views/Project1Visualization2/Viz2?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Reason for Choosing the Chart Type:

Bar charts are commonly used to compare numerical values visually. I chose a horizontal bar chart for this visualization instead of other chart types like pie charts or line charts because it is much better suited for comparing exact values across categories. Since my goal was to highlight the top 10 most affected countries, a bar chart is the ideal choice for presenting a clear ranking.

The horizontal layout was chosen specifically because it makes long country names easier to read and provides better use of space. It also prevents clutter and keeps the chart clean and easy to understand.

Purpose of the Visualization:

The purpose of this visualization is to show which countries had the highest cumulative number of COVID-19 cases as of December 31, 2022. This allows viewers to quickly compare how different countries were impacted by the pandemic. By focusing on only the top 10, the chart avoids information overload and highlights the most critical data.

Reason for Choosing Style and Colors:

I used shades of red in this chart because red is commonly associated with warnings, danger, and health-related emergencies. This makes it visually appropriate for showing the severity of a global pandemic. Darker red shades represent countries with higher case counts, reinforcing the impact visually.

For the title, I used a bold font (Tableau Bold) to draw attention, while the rest of the text was kept in Tableau's default font for simplicity and consistency.

I also added labels on each bar to display the exact number of confirmed cases, so viewers don't need to hover or estimate the values. This makes the chart more informative and user-friendly.

Visualization 3:

Scatter Plot showing COVID-19 Testing vs. Death Rate by Country

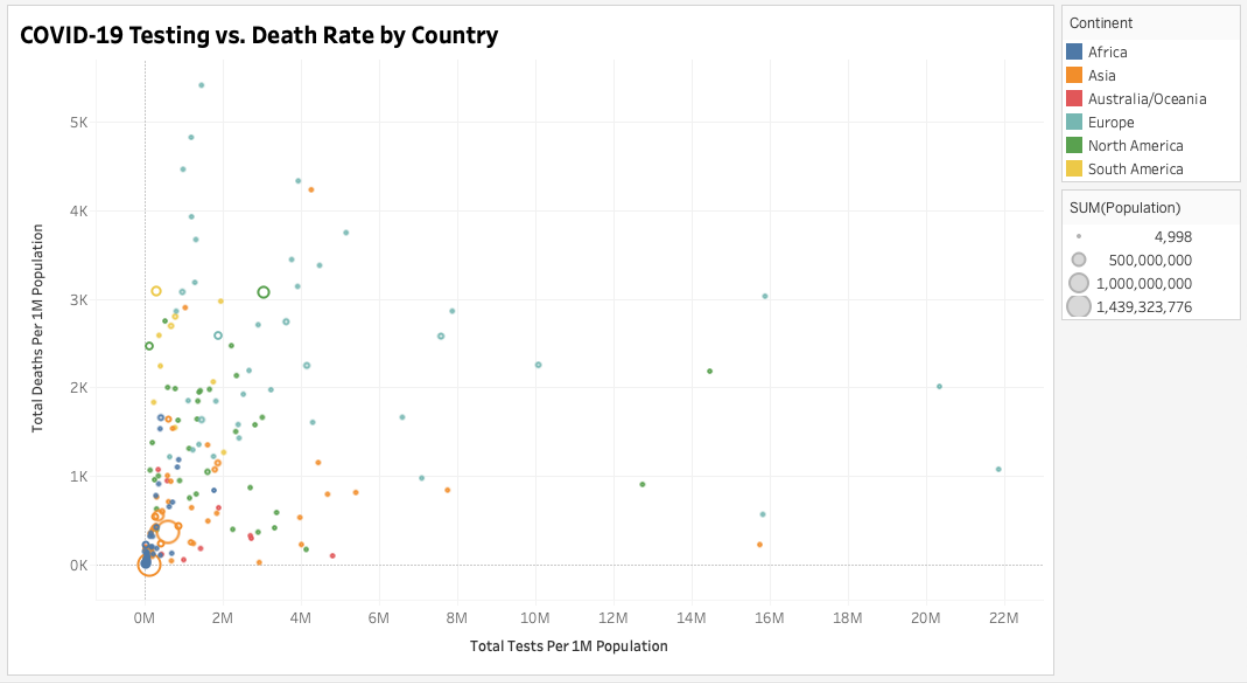

Figure 3: COVID-19 Testing vs. Death Rate by Country
Link: https://public.tableau.com/views/Project1Visualization3/Viz3?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Reason for Choosing the Chart Type:

Scatter plots are useful for exploring relationships between two numerical variables. I chose a scatter plot instead of bar charts or pie charts because those are better for showing totals or comparisons, not how two variables interact.

Since the goal of this visualization is to find out whether more COVID-19 testing is linked to fewer deaths, a scatter plot is the best fit. It makes it easy to spot trends, clusters, or outliers, and each point on the plot represents a single country which makes a scatter plot perfect for this type of analysis.

Purpose of the Visualization:

The purpose of this visualization is to understand if there is a correlation between testing rates and death rates across different countries. By plotting total tests per 1 million population on the x-axis and total deaths per 1 million population on the y-axis, we can see if countries that tested more had fewer deaths, spot countries that had high death rates despite high testing (or vice versa) and notice how countries from the same continent behaved during the pandemic. Overall, this visualization helps reveal how testing may have influenced health outcomes or death rates.

Another important feature of this visualization is its interactivity. When we hover over any data point, we can see a tooltip showing the country's name and continent, number of tests per million, deaths per million, and population size. This makes it easy to explore specific countries and compare them directly.

Reason for Choosing Style and Colors:

I used different colors for each continent to make it easier to compare regions. This way, we can quickly see if countries in the same part of the world show similar results.

The size of each circle is based on population, so larger bubbles represent countries with more people. This helps viewers understand the scale of each country in relation to its testing and death rates.

The background is plain white to keep the focus on the data. I used a bold title font (Tableau Bold) to make the chart topic stand out and left the rest of the text in Tableau's default style for simplicity and consistency.

Visualization 3:

Line Chart showing Cumulative COVID-19 Deaths Over Time of Top 3 Countries with Highest COVID-19 Cases: U.S., India, Brazil

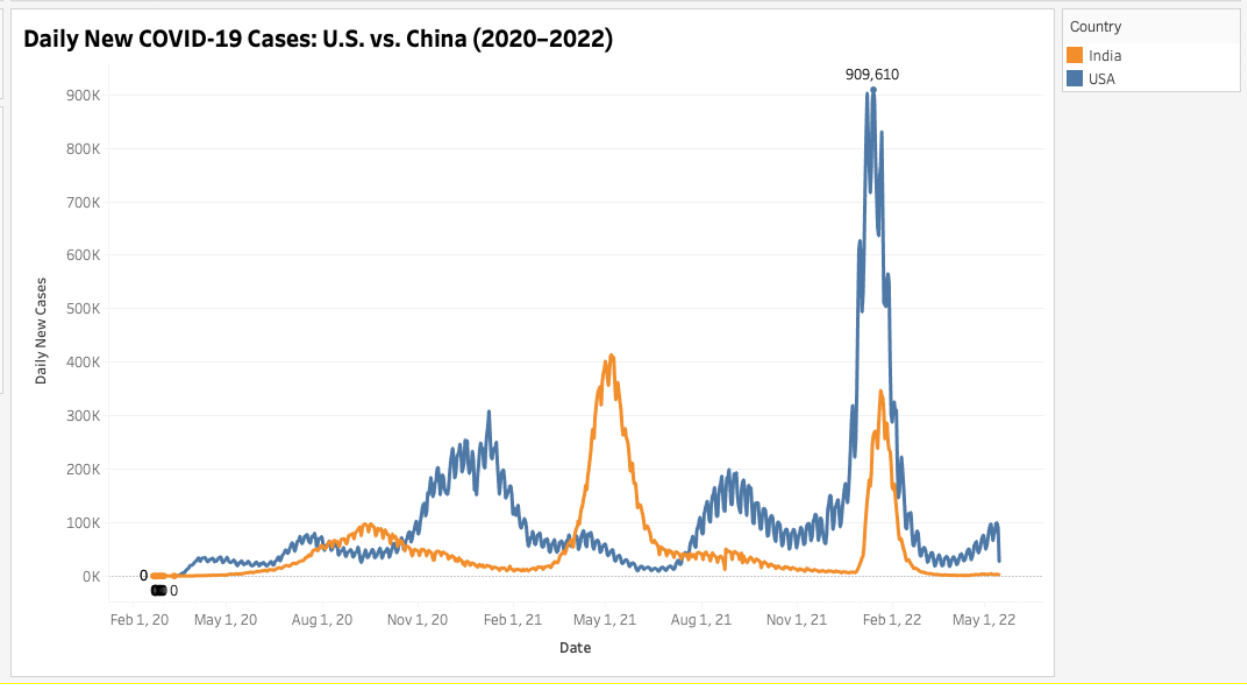

Figure 4: Daily New COVID-19 Cases: U.S. vs. India
Link: https://public.tableau.com/views/Project1Visualization4/Viz4?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Reason for Choosing the Chart Type:

I chose a line chart because it is the best way to show changes over time. This chart type helps track how daily new COVID-19 cases changed from 2020 to 2022 in the United States, India and Brazil. Other chart types like bar charts or pie charts are better for totals or comparisons at one point in time, but they don't show trends well like line chart does. A line chart clearly shows rises, drops, and spikes in cases across different time periods, making it perfect for this analysis.

Purpose of the Visualization:

The purpose of this visualization is to show how cumulative deaths from COVID-19 changed over time in the top 3 countries with the highest total cases: the United States, India, and Brazil. While these three countries had the most confirmed cases, their death rankings were different with the U.S. having the highest deaths, followed by Brazil, and then India. This helps us understand how the pandemic affected each country differently and see when each country experienced death numbers rise over time.

Reason for Choosing Style and Colors:

I used contrasting line colors to clearly separate the countries; the U.S. is shown in red while Brazil is in green, and India is shown in yellow. These colors stand out against the white background and are easy to tell apart.

I also added labels on the chart to present the highest and lowest points on the line chart to interpret the chart easily. Additionally, I added a clear title using bold title font (Tableau Bold) while the rest of the text was kept in Tableau's default font for simplicity and consistency.

Reference:

Assaker, J. (2020). *COVID-19 Global Dataset.* Kaggle.
https://www.kaggle.com/datasets/josephassaker/covid19-global-dataset

Tableau Public Links:

Visualization 1 :
https://public.tableau.com/views/Project1Visualization1_17520100809420/Viz1?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Visualization 2:

https://public.tableau.com/views/Project1Visualization2/Viz2?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Visualization 3:
https://public.tableau.com/views/Project1Visualization3/Viz3?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Visualization 4:

https://public.tableau.com/views/Project1Visualization4/Viz4?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link