Global Data Covid
INFO584_50SU24
Professor Serhat Simsek, Ph.D.
Coach Jennifer Turner
INFO584_50SU24 - Project Group 2
Krupa Gor-Vanessa Hartkopf
Angelique Hellriegel-Krystal Yurnet
August 25, 2024

Table of Contents

Abstract

The Covid business case study analyzed global data on Covid deaths and prevalence, focusing on the top ten countries with the highest mortality rates. The analysis concluded that mortality rates were generally higher in densely populated countries, though recovery rates varied independently of population size. Additionally, the number of documented tests conducted did not consistently correlate with the reported cases, highlighting significant inconsistencies. These discrepancies may be attributed to external factors such as the quality of healthcare systems, the availability of medical supplies, public education on virus prevention, and varying living conditions. Visual representations were employed to illustrate these findings, providing a comprehensive view of the pandemic's impact across different regions.

Introduction/Purpose

The global data on the Covid dataset provides the latest statistical global data on Covid cases as of June 30, 2024. The data includes total cases, death counts, and recovered number of patients in a country format. "Tests" were included in the data set to show which area received testing for Covid. The continents in this study include North America, South America, Africa, Asia, Europe, and Oceania, also known as Australia.

Data Set

The Covid dataset from Kaggle provides a comprehensive view of the pandemic's global impact, tracking confirmed cases, deaths, recoveries, and testing across various countries. To provide a complete picture of the effects of Covid globally, we used multiple visual representations, explicitly targeting the top ten countries most impacted by the virus with the highest mortality rates. These visualizations help illustrate the factors influencing deaths versus recoveries, such as the number of cases and tests performed.

By drilling down further, we focused on the top ten countries with the highest death rates, analyzing the relationship between the number of tests conducted, the number of confirmed cases, and the corresponding outcomes regarding deaths and recoveries. For example, Exhibit #7 illustrates the number of Covid cases and deaths in these top ten countries, while Exhibit #8 compares deaths versus recoveries within the same countries. Organized by country, the dataset enables a nuanced comparative analysis, allowing us to evaluate how different national responses influenced the pandemic's outcomes.

While the dataset is invaluable for understanding the pandemic's progression and informing public health strategies, it's essential to recognize the variations in data quality due to differences in healthcare infrastructure and reporting standards across regions, which can affect accuracy. Despite these limitations, the dataset remains an essential tool for analyzing the impact of Covid and guiding future responses to global health crises.

Analysis

The Covid data analysis shows the pandemic's impact across different regions, explicitly showing data on cases, deaths, testings, and recoveries. The data shows the virus being spread across all continents, highlighting that no region was unaffected, as shown in our map visualizations. The bubble chart indicates that Europe and Asia were the most affected, with Europe having the most cases. It also shows smaller bubbles representing regions like Ociana and Africa, with fewer cases. Our bar chart shows that mortality rates also varied significantly, with Europe having the highest death toll, followed by North America and Asia. The heat map shows a correlation between high population and increased death rates. Testing data is shown in our heat map, which shows that the United States and India conducted the most tests, which reflects their effort to manage the virus. However, testing varied globally, leading to uneven responses. The comparison of cases, deaths, and recoveries in different countries reveals gaps between these metrics. For example, the United States had the highest number of cases and deaths but also the highest recovery rate. At the same time, Australia had higher recovery rates and fewer deaths. The analysis needs to show more limitations in the data, which may be due to underreporting. This can lead to an underestimation of the pandemic's real impact.

Results of Global Impact

The graphs below represent the global impact of Covid-19 and the outcome for each country. The information is represented visually in the graphs, maps, and charts below and can be accessed in Tableau Public: [Final Project Group 2 | Tableau Public](#)

The following countries in Exhibit #1 have been affected by the Covid-19 virus.

Exhibit #1
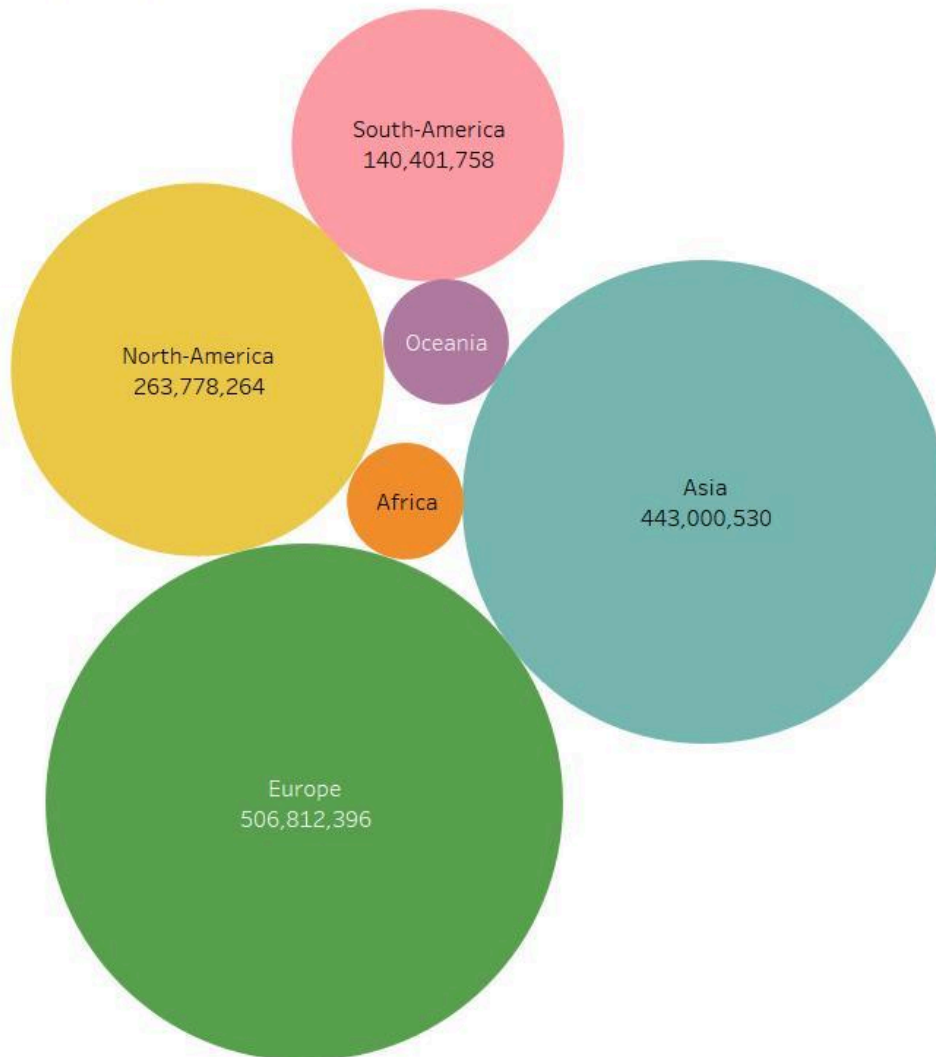

Countries Affected By Covid-19
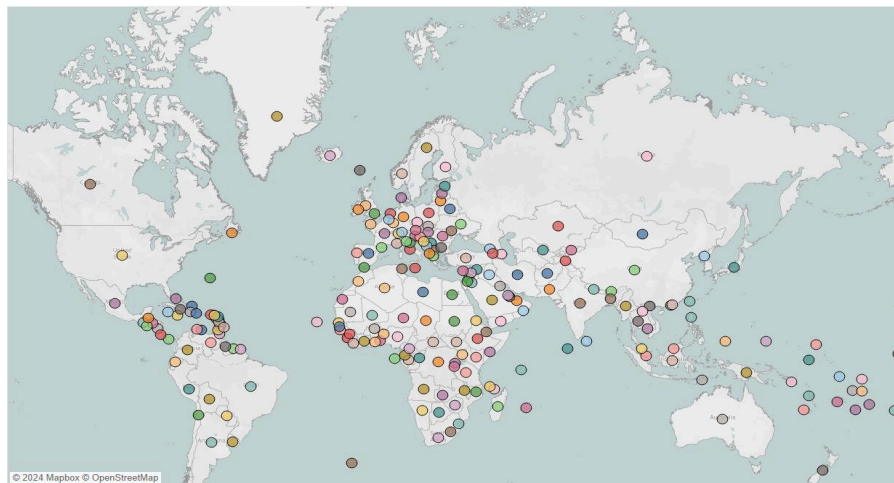
This map visualization displays the spread of Covid across various countries. The colored circles represent every country that has been infected with Covid. The spread is represented on all continents. This visualization conveys the global reach of the pandemic, showing that no region was left unaffected. It can be used to compare the spread of cases across the globe and understand the geographic distribution at a glance. This visualization is interactive on our Tableau file, showing the number of cases, deaths, recoveries, tests, and population per country when you hover or click on a circle for a more detailed outlook.

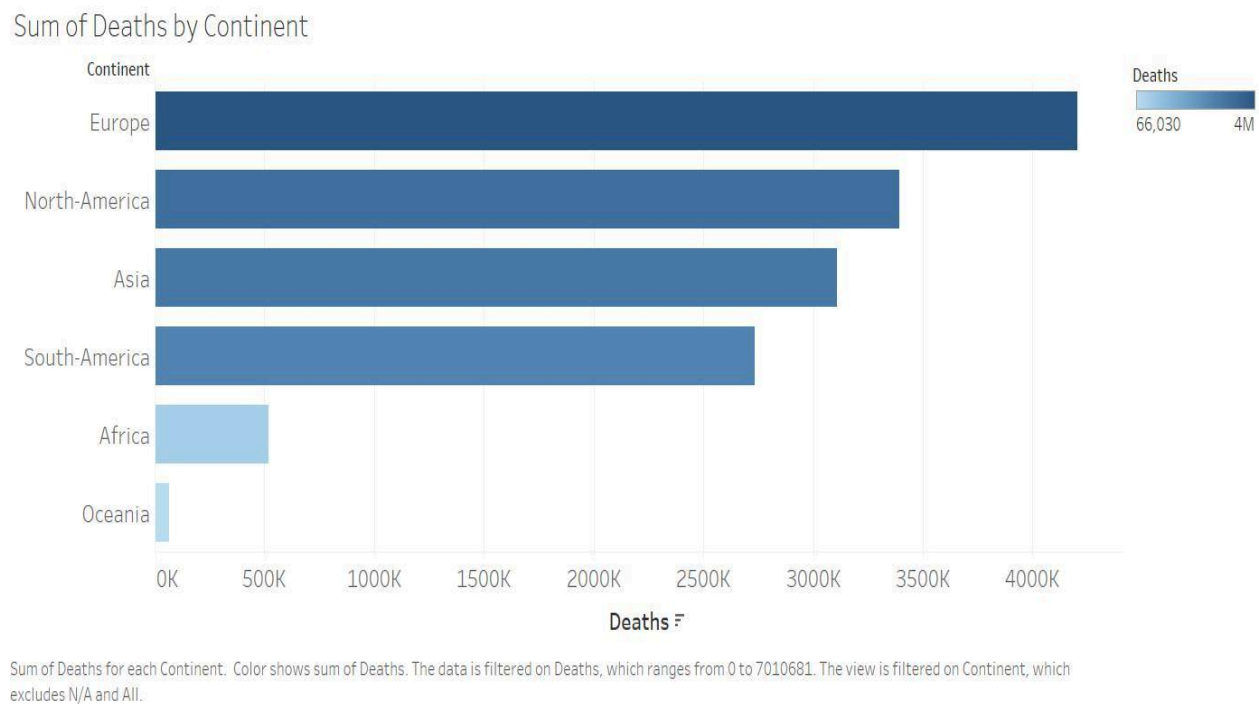The Bubble graph, Exhibit #2 shows the sum of cases by continent.

Exhibit #2

Sum of Cases by Continent

South-America
140,401,758

North-America
263,778,264

Oceania

Asia
443,000,530

Africa

Europe
506,812,396

This bubble chart compares Covid cases across the globe. The size of each bubble represents the total number of cases in each region. Larger bubbles indicate regions with more cases, while smaller bubbles represent regions with fewer cases. The different sizes help quickly convey the pandemic's impact across regions. Looking at the bubble chart, you see that Europe is the largest bubble. This indicates that it has the highest reported cases among all areas, with almost 507 million cases. Asia is closely followed by about 443 million cases. North America is in the middle range with a smaller bubble than Asia and Europe. South America, Oceania, and Africa have smaller bubbles, indicating lower case numbers. This chart effectively uses a vial hierarchy to emphasize regions with the most impact.

The chart, Exhibit #3, represents the sum of deaths by continent.

Exhibit #3

Sum of Deaths for each Continent. Color shows sum of Deaths. The data is filtered on Deaths, which ranges from 0 to 7010681. The view is filtered on Continent, which excludes N/A and All.

The bar chart vividly illustrates the significant disparities in the sum of deaths across different continents, with Europe emerging as the most affected region, recording over 4 million deaths. This staggering figure suggests that Europe experienced an exceptionally severe event or series of events that resulted in a disproportionately high mortality rate compared to other regions. North America and Asia also show substantial death tolls, with both continents reporting totals nearing 3 million and over 2.5 million, respectively. These figures indicate that these regions were similarly impacted by large-scale events, possibly pandemics or conflicts, that led to high mortality rates. In contrast, South America, while still significantly affected with over 2 million deaths, records a lower total than the regions mentioned above. Africa and Oceania show the lowest death tolls, with Africa reporting around *500,000* and Oceania fewer than *100,000*. The

lighter shading of these regions in the chart emphasizes their relatively lower impact, suggesting they were either less exposed to the deadly events or had more effective mitigation measures. The chart's color gradient effectively highlights these disparities, underlining the need for further exploration into the factors contributing to such uneven impacts across the globe, including differences in healthcare infrastructure, economic conditions, and regional responses.

The Treemap shown in Exhibit #4 below shows the sum of tests for each country.

Exhibit #4

Sum of Tests by Country
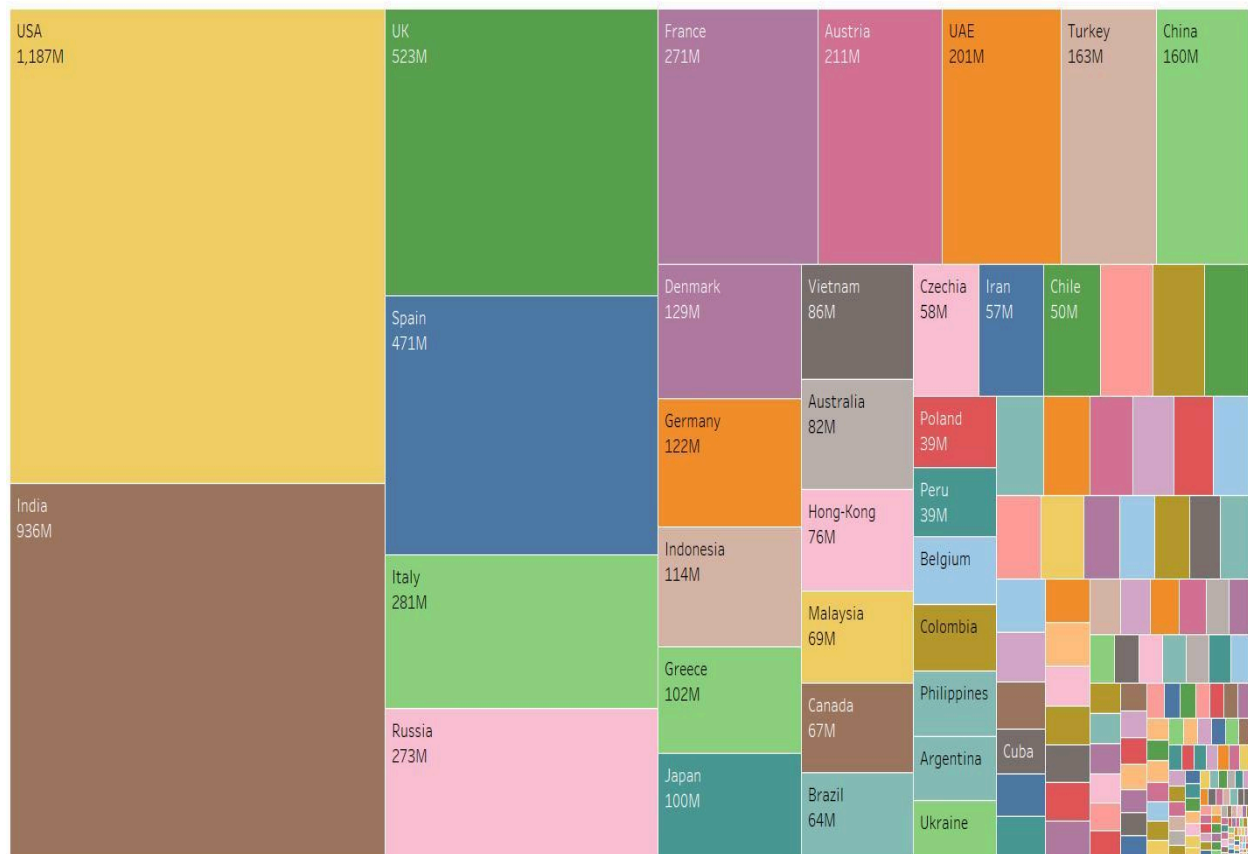
Country and sum of Tests. Color shows details about Country. Size shows sum of Tests. The marks are labeled by Country and sum of Tests.

The treemap chart visually represents the distribution of Covid tests conducted across various countries, with each block's size reflecting the relative volume of tests. The United States and India stand out prominently with the most significant blocks, indicating that these two countries conducted the highest number of tests, underscoring their extensive efforts to track and manage the spread of the virus. The United Kingdom, Spain, and Italy also have sizable blocks, reflecting their significant testing campaigns. In contrast, smaller blocks represent countries with lower testing volumes, highlighting a considerable global disparity in testing capacities and efforts. This chart effectively illustrates the uneven global response to Covid testing, with some countries
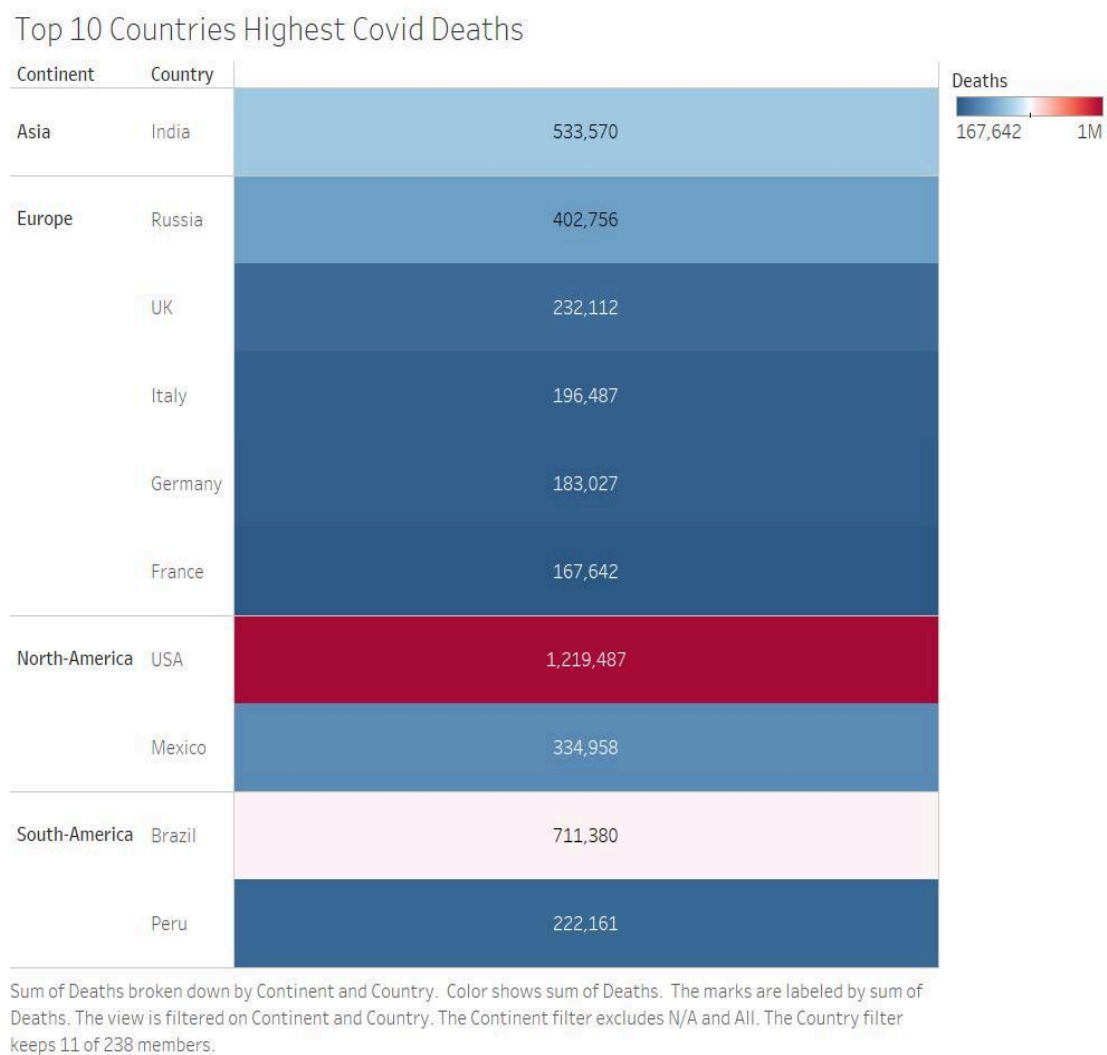
implementing large-scale testing strategies while others, possibly due to limited resources or different public health strategies, conducted far fewer tests.

Drilled Down by Country

The top ten countries with the highest Covid fatalities are listed in Exhibit #5.

Exhibit #5

Death Heat Map A

Top 10 Countries Highest Covid Deaths

Continent	Country	Deaths
Asia	India	533,570
Europe	Russia	402,756
	UK	232,112
	Italy	196,487
	Germany	183,027
	France	167,642
North-America	USA	1,219,487
	Mexico	334,958
South-America	Brazil	711,380
	Peru	222,161

Deaths: 167,642 — 1M

Sum of Deaths broken down by Continent and Country. Color shows sum of Deaths. The marks are labeled by sum of Deaths. The view is filtered on Continent and Country. The Continent filter excludes N/A and All. The Country filter keeps 11 of 238 members.
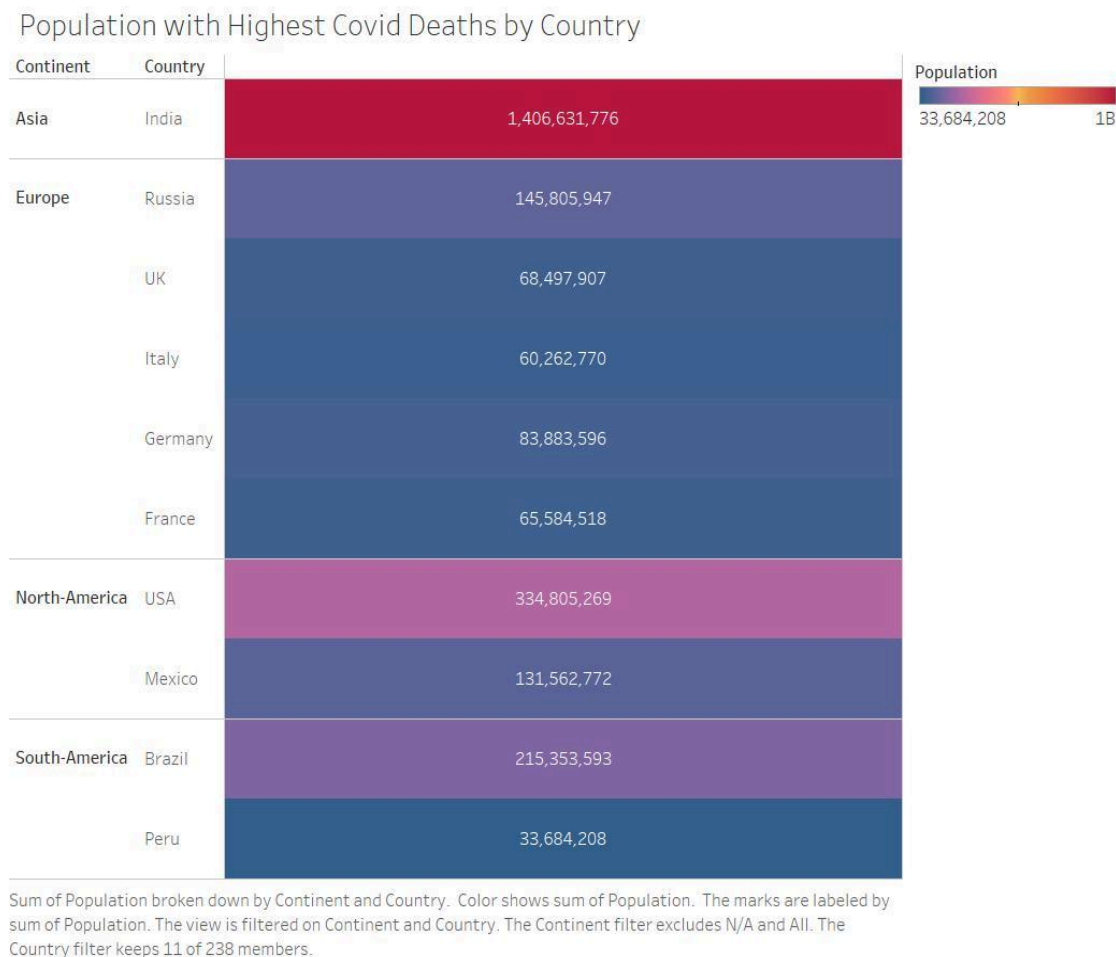
Heat Map A shows the top ten countries worldwide with the highest death rates due to Covid. As the heat map indicates, the USA (red) has the highest number of Covid-related deaths, with South America (white) coming in second. In heat map A, the darker the blue, the fewer deaths, and as the blue gets lighter, closer to white, the death rates are higher compared to the top ten.

The next map, Heat Map B features the same ten countries with the highest Covid rates, but it indicates population. You can see that India (red) has the highest population out of the top ten, with the USA (light purple) coming in second and Brazil (dark purple) coming in third. The darker the blue, the lower the population. The closer to the purple color indicates a higher population, respectfully.

Exhibit #6

The graph in Exhibit #6 represents the population for the top ten countries with the highest Covid fatalities.

Population with Highest Deaths

Population with Highest Covid Deaths by Country

Continent	Country	Population
Asia	India	1,406,631,776
Europe	Russia	145,805,947
	UK	68,497,907
	Italy	60,262,770
	Germany	83,883,596
	France	65,584,518
North-America	USA	334,805,269
	Mexico	131,562,772
South-America	Brazil	215,353,593
	Peru	33,684,208

Population legend: 33,684,208 — 1B

Sum of Population broken down by Continent and Country. Color shows sum of Population. The marks are labeled by sum of Population. The view is filtered on Continent and Country. The Continent filter excludes N/A and All. The Country filter keeps 11 of 238 members.
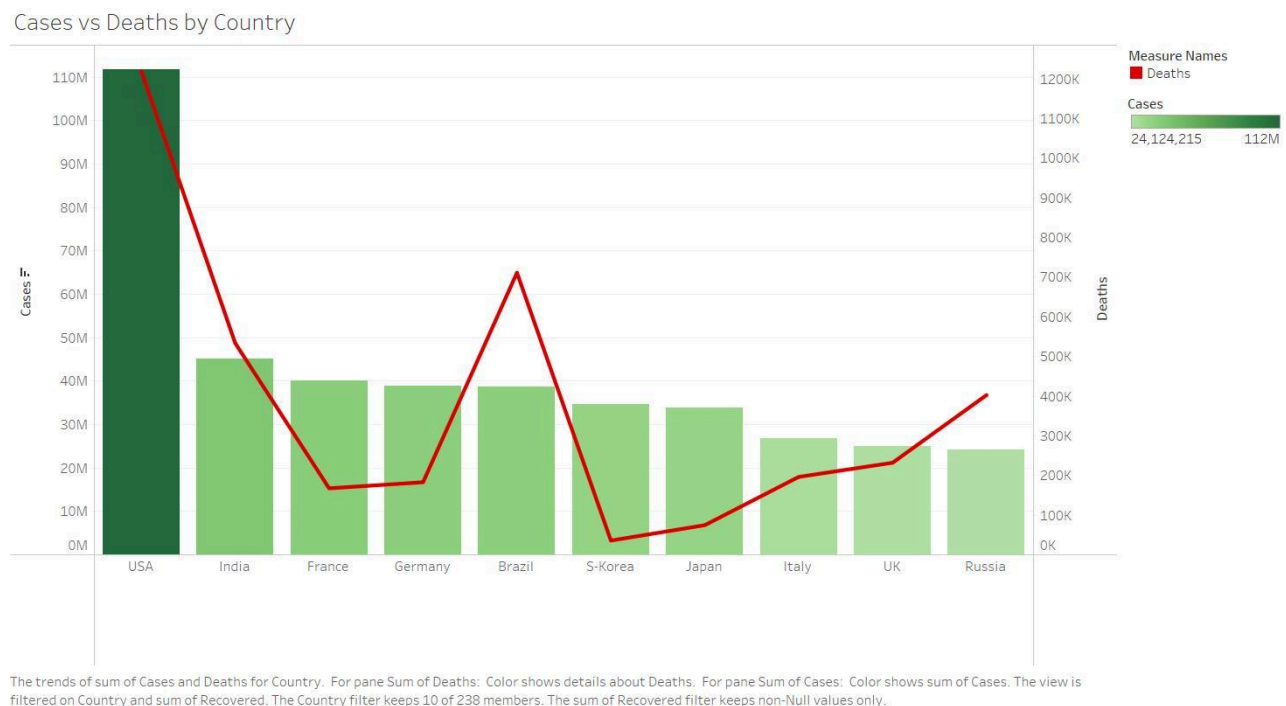
There is a correlation between the highest rate of death and the highest population within a country. The higher the population seems, the higher the death rates. Due to the nature of the virus and the fact that it's airborne when people are in close quarters, the virus can quickly spread to dense populations. According to the CDC, distance is one of the prevention or proactive actions that safeguard people against the virus. It makes sense that high populations would have a higher death rate due to Covid.

Exhibit #7

Top ten countries in Exhibit #7 with the highest deaths versus the number of cases.

The trends of sum of Cases and Deaths for Country. For pane Sum of Deaths: Color shows details about Deaths. For pane Sum of Cases: Color shows sum of Cases. The view is filtered on Country and sum of Recovered. The Country filter keeps 10 of 238 members. The sum of Recovered filter keeps non-Null values only.

The bar graph displays the number of cases by country. The darker the green bar, the higher the number of cases. The red line compares the number of deaths versus the number of cases. Countries with no recovery data (null) were not displayed in the top ten. The USA has the highest cases at *111,820,082* and the highest death rate at *1,219,487*. The distance between the cases and deaths in the US is close, unlike the rest of the top ten. Russia had more deaths, *402,756,* compared to the number of identified cases, *24,124,215*. Brazil had the most significant gap between the number of instances of *38,743,918* and fatalities of *711,380*. South Korea and Japan also had a considerable gap between cases and deaths. Several factors and outside influences could have affected the cases and deaths. According to the CDC, death rates are higher in people with underlying conditions and health conditions, so depending on the country

depending on the person, death rates may be higher for people who have underlying health conditions. Also, many countries closed their borders, and people wore masks during the pandemic to prevent the virus from spreading.

Exhibit #8

The top ten countries in Exhibit #8 had the highest recovery versus mortality rates.

Recovered vs Deaths by Country

The trends of sum of Recovered and Deaths for Country. For pane Sum of Recovered: Color shows sum of Recovered. For pane Sum of Deaths: Color shows details about Deaths. The view is filtered on Country and sum of Recovered. The Country filter keeps 10 of 238 members. The sum of Recovered filter keeps non-Null values only.

The bar graph shows the number of documented recoveries (blue bars) from Covid versus the number of deaths (red line) in that country. The darker the blue, the higher the recovery rates. The red line shows the number of deaths. Countries with no data (null) for recovery were not displayed in the top ten. The USA has the highest number of fatalities, *1,219,487,* but also the highest recovery rate, *109,814.428.*

Interestingly, Australia and South Korea had fewer deaths and more and exceeded recovery rates. Australia had 24,414 deaths and more recovered patients at 11,820,014. South Korea recovered *34,535,939,* with death rates at *35,934.* Many outside factors may have affected higher recovery and lower death rates within a particular country. Higher recovery rates may be due to the country's medical facilities or the population's health.

Limitations/Conclusion

Limitation:

1. **Data Inconsistencies and Reporting Delays:**
   - One of the significant limitations of the Covid datasets, including those from Kaggle and the NCBI article, is the inconsistency in data reporting across different countries. Variations in testing rates, diagnostic criteria, and the timing of data updates can lead to discrepancies in the reported number of cases and deaths. Additionally, delays in reporting, especially in regions with limited healthcare infrastructure, can result in underestimation or misrepresentation of the accurate scale of the pandemic.

2. **Underreporting and Misclassification of Deaths:**
   - The NCBI article highlights the issue of underreporting and misclassification of Covid-related deaths. In many countries, deaths from Covid may have been underreported due to lack of testing, especially in the early stages of the pandemic, or misclassified as other causes. This underreporting is particularly prevalent in regions with strained healthcare systems or limited access to testing, such as rural areas in India. As a result, the actual death toll of the pandemic might be significantly higher than what is officially recorded.

3. **Excess Mortality Data:**
   - While excess mortality is a valuable measure for understanding the broader impact of the pandemic, it is also subject to limitations. Excess mortality data can be influenced by factors unrelated to Covid, such as other epidemics, natural disasters, or changes in population demographics. Moreover, the availability and accuracy of excess mortality data vary widely across countries, making it challenging to draw direct comparisons.

4. **Socioeconomic and Demographic Factors:**
   - The datasets do not fully account for the wide range of socioeconomic and demographic factors influencing the spread of Covid and its impact. Factors such as income inequality, access to healthcare, population density, and public health infrastructure play crucial roles in determining how severely different populations are affected. Without adjusting for these variables, the data might not fully capture the disparities in health outcomes between different regions and demographic groups.

5. **Varying Healthcare Capacity:**
   - The healthcare capacity of each country, including the availability of medical resources, hospital beds, and healthcare workers, significantly affected the outcomes of Covid cases. The datasets may not sufficiently capture the strain on healthcare systems, particularly during peak waves of the pandemic. This

limitation makes it difficult to assess the effectiveness of healthcare responses purely based on reported case and death numbers.

Conclusion:

The analysis of Covid data from the Kaggle dataset and insights from the NCBI article provide a multifaceted view of the pandemic's global impact. The data reveals significant variations in how countries like the United States, India, and the United Kingdom managed the crisis, highlighting the successes and challenges each faced.

In the United States, despite extensive testing and advanced healthcare infrastructure, the country experienced one of the highest death tolls globally. This outcome underscores the importance of timely and coordinated public health responses and addressing socioeconomic disparities exacerbating health crises.

According to the NIH, India's experience was marked by a devastating second wave driven by the Delta variant. The country's challenges in testing and healthcare access, particularly in rural areas, led to significant underreporting of cases and deaths. The pandemic exposed critical weaknesses in healthcare infrastructure and highlighted the importance of equitable access to medical resources.

After initial struggles, the United Kingdom demonstrated the effectiveness of a robust vaccination campaign in controlling the virus and reducing mortality. This case emphasizes the critical role of vaccines and the importance of adaptability in public health strategies.

However, these findings must be viewed in light of the limitations of the available data. Inconsistencies in reporting, underreporting of deaths, and the influence of socioeconomic factors all suggest that the true impact of the pandemic may be even more significant than what the data reveals. Excess mortality, while a valuable metric, is also subject to limitations and may not fully capture the broader effects of the pandemic.

In conclusion, the Covid pandemic has exposed the vulnerabilities and strengths of global public health systems. It has underscored the need for accurate data, international cooperation, and resilient healthcare infrastructure to better prepare for future pandemics. As we continue to analyze and learn from this unprecedented crisis, these lessons will be crucial in shaping a more effective and equitable global health response in the future.

Data Selected for study: https://www.kaggle.com/datasets/sanaaafrine/Covid-dataset
Tableau Public Link: **Final Project Group 2 | Tableau Public**

References:

CDC. (2024b). *Search results*. Centers for Disease Control and Prevention.
https://search.cdc.gov/search/?query=Covid&dpage=1

International Monetary Fund. (2024). *International Monetary Fund* . IMF.
https://www.imf.org/external/datamapper/PPPPC@WEO/THA/CAF

Johnson, S. R. (2023, September 6). *10 countries with the Best Public Health Systems in the
World | Best Countries | U.S. news*. US News.
https://www.usnews.com/news/best-countries/slideshows/countries-with-the-most-well-develope
d-public-health-care-system

COVID-19 Excess Mortality Collaborators. (2022, April 16). *Estimating excess mortality due to
the COVID-19 pandemic: A systematic analysis of covid-19-related mortality, 2020-21*. Lancet
(London, England).
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8912932/#:~:text=Although%20reported%20C
OVID%2D19%20deaths,excess%20mortality

WHO. (2024). *Covid cases | WHO Covid Dashboard*. World Health Organization.
https://data.who.int/dashboards/Covid19/cases