DSC465 Final Project

Inna Mykoliyuk	Aditya Patrick Moses	Sparshika Ajmaan Dinesh Kumar	Raghukula Tilak Bolli	Karan Bhuva

Introduction: 3

Visualizations: 3

Visualization – 1 3

Global Spread of Covid 19 From January 2020 – June 2020 3

Visualization – 2 4

Spread of COVID-19: A Global Overview Through a Choropleth Map 4

Visualization – 3 5

Top 10 Countries by Case fatality Rate 5

Visualization – 4 6

Top 10 Countries by Case Fatality Rate Over Time 6

Visualization – 5 7

Top 10 and Bottom 10 Countries by recovery rate 7

Visualization – 6 8

Top 10 Countries by Confirmed Cases and Deaths 8

Graph Generation Process 9

Key Learnings 9

Conclusion: 10

Appendices 11

Appendix 1: Individual Report by Inna Mykoliuk 11

Appendix 2: Individual Report by Raghukula Tilak Bolli 14

Appendix 3: Individual Report by Aditya Patrick Moses 15

Appendix 4: Individual Report by Sparshika Ajmaan Dinesh Kumar 16

Appendix 5: Individual Report by Karan Bhuva 19

Introduction: 3
Visualizations: 3
Visualization – 1 3
Global Spread of Covid 19 From January 2020 – June 2020 3
Visualization – 2 4
Spread of COVID-19: A Global Overview Through a Choropleth Map 4
Visualization – 3 5
Top 10 Countries by Case fatality Rate 5
Visualization – 4 6
Top 10 Countries by Case Fatality Rate Over Time 6
Visualization – 5 7
Top 10 and Bottom 10 Countries by recovery rate 7
Visualization – 6 8
Top 10 Countries by Confirmed Cases and Deaths 8
Graph Generation Process 9
Key Learnings 9
Conclusion: 10
Appendices 11
Appendix 1: Individual Report by Inna Mykoliuk 11
Appendix 2: Individual Report by Raghukula Tilak Bolli 14
Appendix 3: Individual Report by Aditya Patrick Moses 15
Appendix 4: Individual Report by Sparshika Ajmaan Dinesh Kumar 16

Appendix 5: Individual Report by Karan Bhuva 19

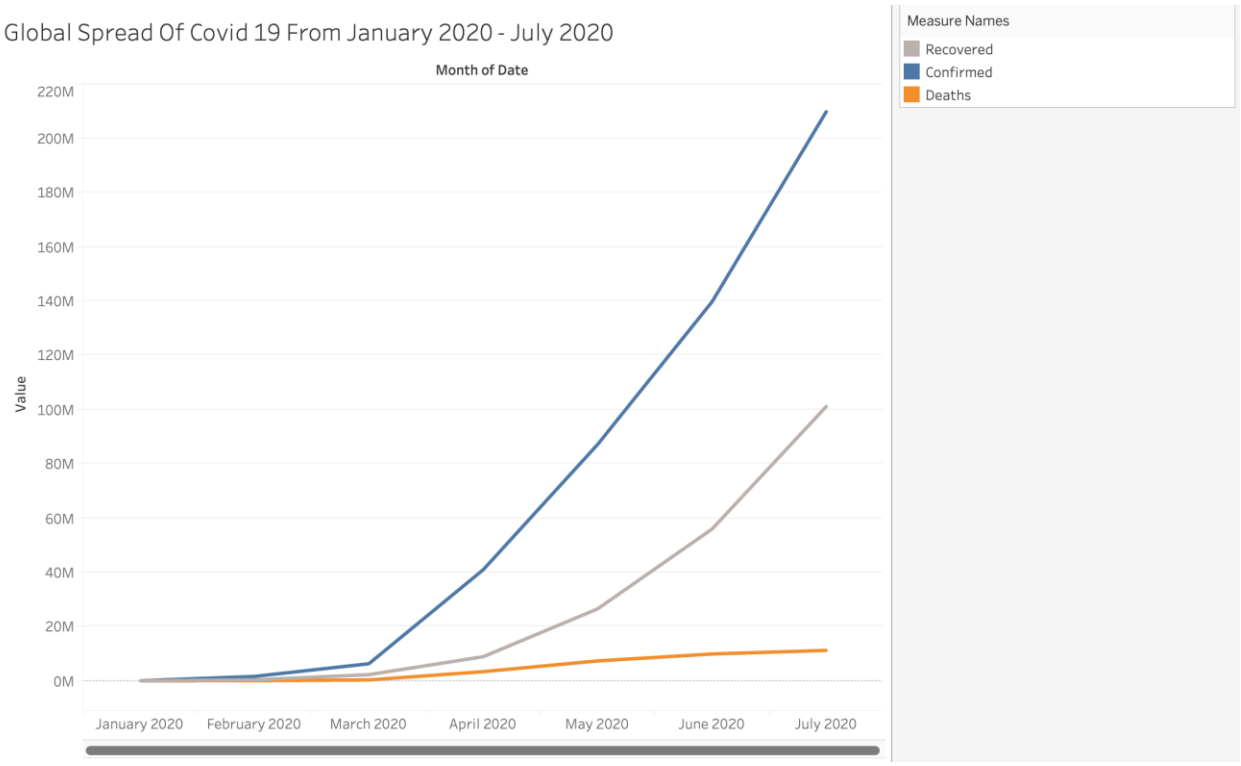
Introduction:

The COVID-19 pandemic, which emerged in late 2019, rapidly became a global crisis. This project visualizes the global spread and impact of COVID-19 from January to June 2020 using various charts and maps. These visualizations provide insights into patterns of confirmed cases, fatalities, and recoveries across different regions, highlighting the pandemic's progression and the varied responses and outcomes of different nations.

Visualizations:

Visualization – 1

Global Spread of Covid 19 From January 2020 – June 2020



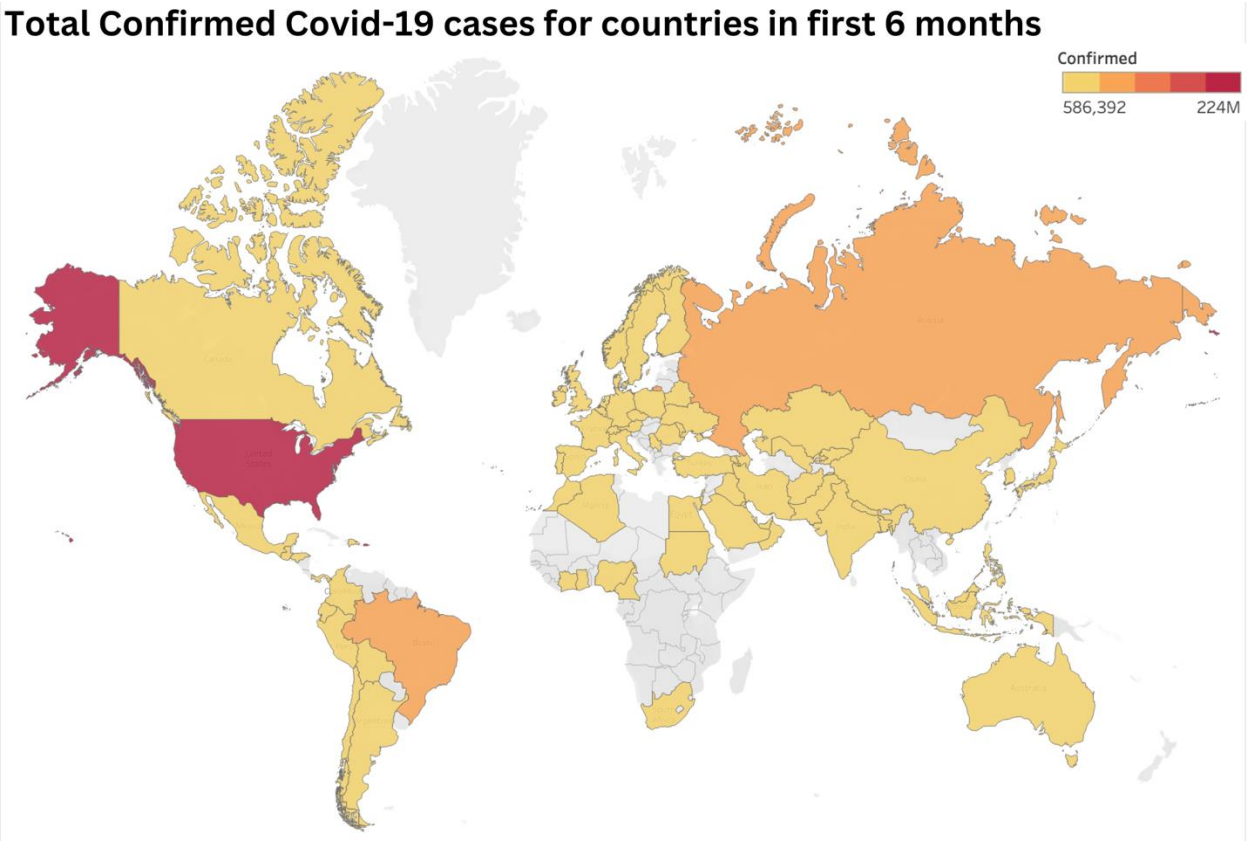
The above visualization is a line chart showing the global spread of COVID-19 from January 2020 to July 2020. The chart includes three lines representing "Confirmed," Recovered and "Deaths" cases.

Visualization Details:

- **X-axis:** Months (January 2020 - July 2020)
- **Y-axis:** Cases (in millions)
- **Lines Representing:**
    - **Confirmed Cases (blue):** Steep increase from March 2020, reaching over 200 million by July 2020.
    - **Deaths (orange):** Steady, slower increase, reaching around 20 million by July 2020.
    - **Recovered Cases (gray):** Significant rise from March 2020, reaching about 140 million by July 2020.

Visualization – 2

Spread of COVID-19: A Global Overview Through a Choropleth Map



The provided choropleth map visualizes the global spread of COVID-19, using a sequential color scheme from gold to red to represent confirmed cases, ranging from 586 thousand to 224 million, divided into five parts for clarity.

Key Observations:

1. **General Trends:** Most countries fall within the lower range (shades of gold), indicating fewer than 40 million cases, showing a relatively lower impact.
2. **Significant Outliers:**
    a. **Russia and Brazil:** Both countries show higher case counts with darker shades. Russia has approximately 45.5 million cases, while Brazil has 89 million cases.
    b. **United States:** The USA is marked in the deepest red, with the highest number of cases at 224 million, highlighting its severe impact.

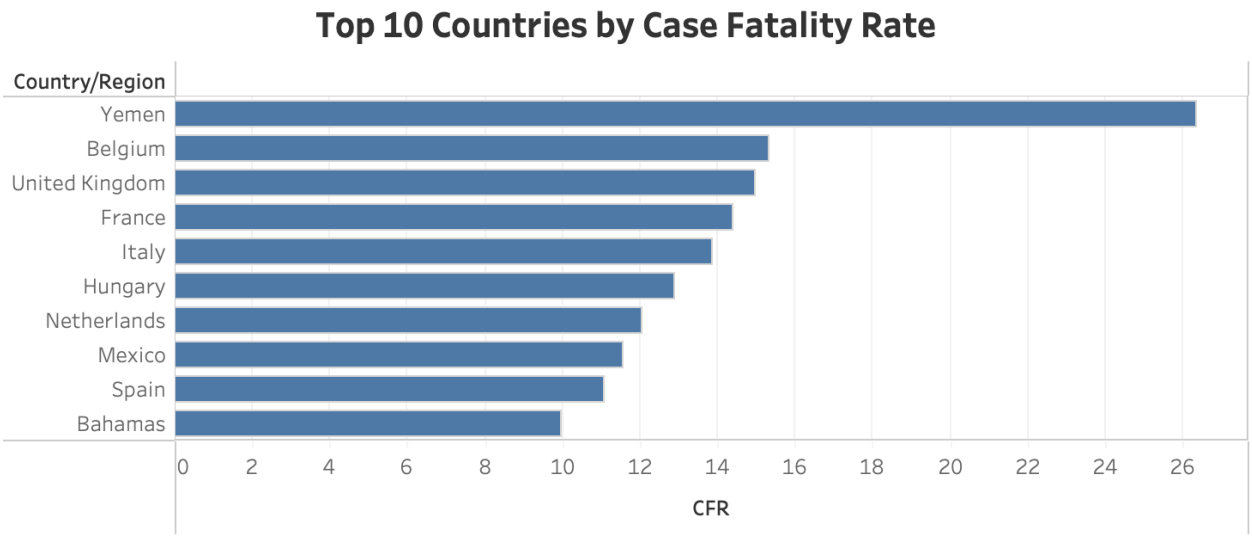
Interpretation and Insights:

The choropleth map provides a clear visual of COVID-19's distribution and intensity worldwide. It differentiates countries with lower case numbers from those with severe outbreaks, offering a snapshot of the pandemic's reach and aiding in identifying regions particularly hard-hit for guiding public health responses and resource allocation.

In conclusion, the map effectively communicates the uneven global impact of COVID-19, highlighting areas with significant outbreaks.

Visualization – 3

Top 10 Countries by Case fatality Rate



The horizontal bar chart visualizes the top 10 countries by Case Fatality Rate (CFR) during the first six months of the pandemic. Countries/regions are listed on the y-axis, while the x-axis represents CFR in percentages. Each bar extends rightward, indicating its CFR value,

with Yemen having the highest CFR. This allows for a quick comparison of the severity of impact across nations.

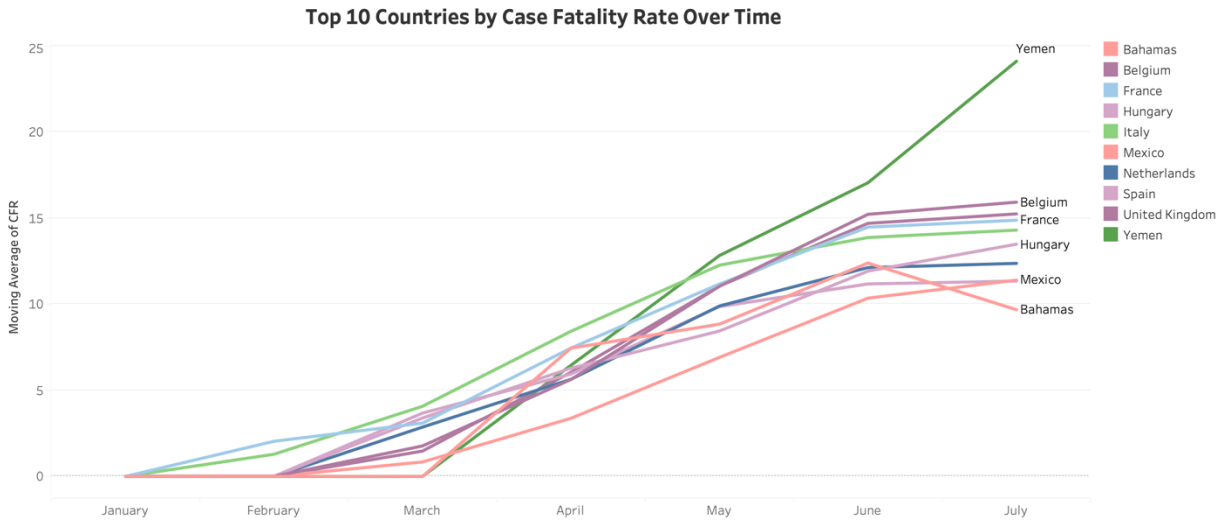
Two variables are depicted:

1. **Categorical (independent) variable**: Countries/regions on the y-axis.
2. **Numerical (dependent) variable**: CFR on the x-axis, ranging from 0 to approximately 26%.

Countries are sorted in descending order of CFR, facilitating immediate recognition of the highest rates and easy comparison. The horizontal layout aids readability, with bar lengths providing a clear visual comparison. This chart offers insights into which countries had higher CFRs and highlights the significant variation between them. The top 10 countries span different WHO (World Health Organization) regions, showing no specific regional pattern.

Visualization – 4

Top 10 Countries by Case Fatality Rate Over Time



The line graph illustrates the moving average of Case Fatality Rates (CFRs) for ten countries from January to July.

Variables:

- **X-axis (Months):** January to July, representing time.
- **Y-axis (Case Fatality Rate):** CFRs for each month.
- **Lines (Countries):** Each colored line represents a different country.

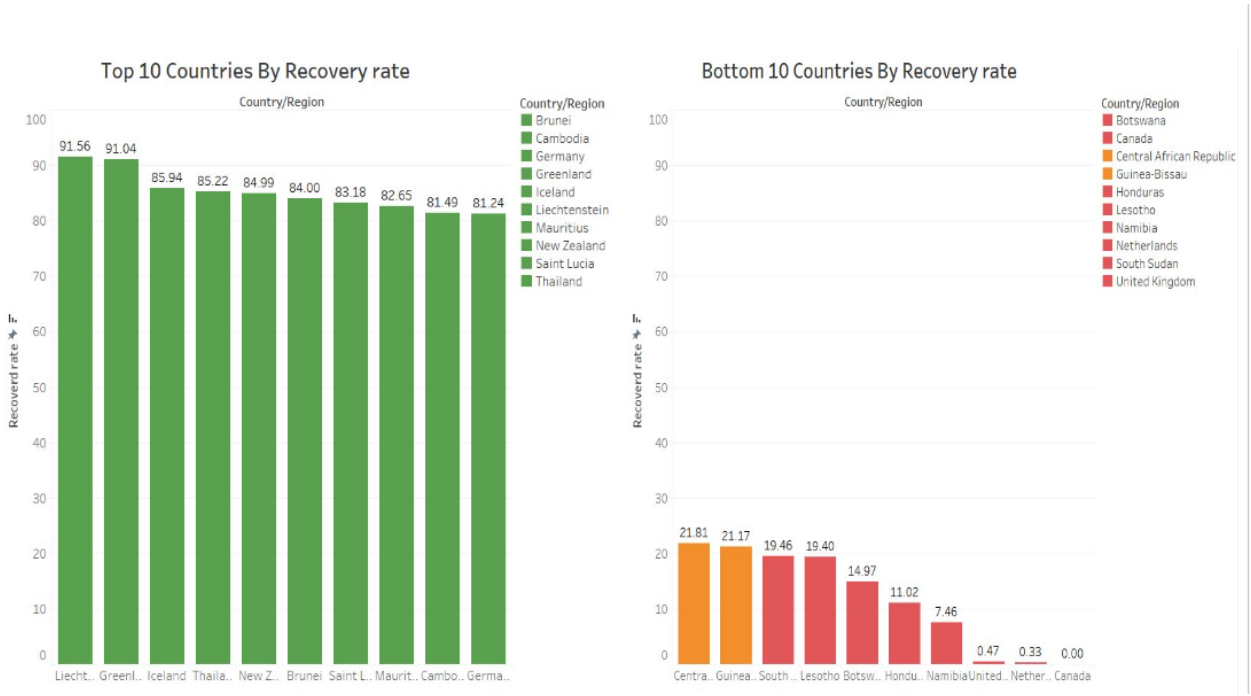
Graph Creation Process:

- Create a calculated field for CFR data for each country.
- Plot raw data points for each country over time.
- Apply a moving average to smooth lines and highlight trends.
- Choose distinct colours for each country's line.

This graph visually tracks how CFRs in different countries changed over several months, showing the pandemic's progression. A decreasing CFR trend indicates effective containment and treatment, while an increasing trend suggests difficulties.

Visualization – 5

Top 10 and Bottom 10 Countries by recovery rate



The visualization effectively contrasts the COVID-19 recovery rates of the top ten and bottom ten countries using two separate bar charts:

- Top 10 Countries by Recovery Rate: This bar chart displays the recovery rates of the countries with the highest success, represented by green bars to symbolize health and recovery. The x-axis lists the countries, while the y-axis shows their recovery rates as percentages.
- Bottom 10 Countries by Recovery Rate: In a similar format, this chart shows the countries with the lowest recovery rates, using red to orange bars to indicate varying levels of concern based on the severity of the situation.

Key Design Elements:

- Uniformity: Both charts maintain a consistent design for easy comparison, using a clear font for legibility.
- Color Coding: Green indicates high recovery rates, while red to orange signifies lower rates, adding a visual layer of interpretation.
- Comparative Layout: Placing these charts side by side allows for immediate visual comparison between the countries with the most and least successful recovery outcomes.

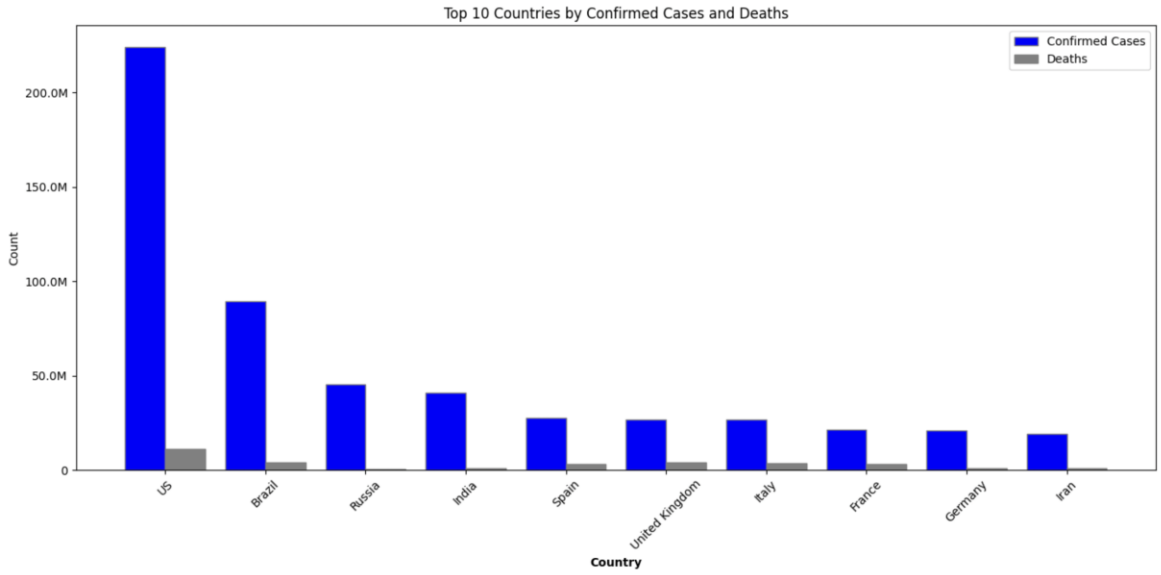
Areas for Improvement:

- Additional Data: Including information about the total number of cases or the context of data collection could provide deeper insights into each country's situation.

This structured approach highlights disparities in global health responses during the pandemic and provides a quick, comparative view of recovery rates across different nations.

Visualization – 6

Top 10 Countries by Confirmed Cases and Deaths



Data Insights:

1. **US:** Highest number of confirmed cases and deaths.
2. **Brazil:** High confirmed cases, fewer deaths than the US.

3. **Other Countries:** Russia, India, Spain, the UK, Italy, France, Germany, and Iran show varying confirmed cases and deaths, with cases always higher than deaths.

Graph Generation Process

The graph was likely generated using Matplotlib in Python. Steps include:

1. **Data Collection:** Gather accurate COVID-19 data for the top 10 affected countries.
2. **Data Cleaning:** Fix errors, remove duplicates, and handle missing values.
3. **Data Transformation:** Organize data into a structured format, aggregating totals for each country.
4. **Data Validation:** Cross-reference with multiple sources and perform statistical checks.
5. **Final Preparation:** Format data for Matplotlib, ensuring all points are correctly labelled.

Key Learnings

1. **Data Quality and Preparation:** Accurate data collection and cleaning are crucial for reliable visualizations.
2. **Effective Data Transformation and Structuring:** Aggregating and normalizing data improves clarity and accuracy.
3. **Visualization Techniques:** Use distinct colours and proper labelling for readability.
4. **Interpretation and Real-World Applications:** Visualizations provide public health insights, inform policy, and educate the public.

Significance

1. The graph highlights the severity of the COVID-19 pandemic in different countries, with the US showing the highest impact.
2. It underscores the need for robust healthcare infrastructure and pandemic preparedness.
3. Governments can use such data to allocate resources effectively, prioritize vaccination campaigns, and implement health measures.
4. Such visualizations help in educating the public about the pandemic's impact, promoting adherence to health guidelines.

Conclusion:

This report provides a comprehensive analysis of the global spread and impact of COVID-19 through various visualizations. The data highlights significant trends, such as the uneven distribution of cases and fatalities, with notable outliers like the United States, Brazil, and Russia. The visualizations underscore the importance of robust data analysis in understanding the pandemic's progression, guiding public health responses, and informing policy decisions. By illustrating disparities in infection, recovery, and fatality rates, the report emphasizes the need for continued vigilance and effective healthcare strategies worldwide.

Appendices

Appendix 1: Individual Report by Inna Mykoliuk

Our group project focused on analyzing COVID-19 trends from January to July 2020. The project aimed to provide insights into how the pandemic spread during its early months. My contributions were centered on two key visualizations: a horizontal bar chart showing the top 10 countries by Case Fatality Rates (CFR) and a line graph representing the moving average of CFRs for ten countries.

Role and Contributions:

Worked on dataset search, offered different dataset variations for the project

Storytelling ideas

Worked on the initial explanatory analysis of different datasets on Covid-19

Line Graph “Moving Average of Case Fatality Rates (CFRs)”

For the line graph, my role involved several critical tasks to ensure the accurate and effective representation of the CFR data over time. The steps I undertook included:

I began by creating calculated fields for the CFR data for each country. This involved calculating the CFR using the formula: (SUM([Deaths]) / SUM([Confirmed])) * 100. This step was crucial for transforming raw data into a format suitable for visualization.

After calculating the CFRs, I plotted the raw data points for each country over the period from January to July. This initial plot provided a basic visualization of the trends but was too uneven to draw meaningful conclusions.

To smooth out and highlight the trends, I applied a moving average to the data points. This technique helped in reducing noise and making the trends more apparent and easier to analyze.

To distinguish between the ten countries, I selected distinct colors for each country's line. This step ensured that viewers could easily differentiate between the countries and follow their respective trends.

This line graph was helpful in our analysis by providing visual evidence of how different countries' CFRs evolved over the months. It showcased the progression of the pandemic in various locations, allowing for a comparative analysis of how the CFRs were increasing. A decreasing CFR could have indicated successful management of the pandemic, while an increasing CFR suggested ongoing challenges.

Horizontal Bar Chart, "Top 10 Countries by Case Fatality Rate"

For the horizontal bar chart, I focused on highlighting the countries with the highest CFRs. The steps involved were as follows:

The two variables in this visualization were the countries/regions (categorical) on the y-axis and the CFR (numerical) on the x-axis.

I sorted the countries in descending order of CFR. This sorting made it immediately apparent which countries had the highest fatality rates, facilitating easy comparison.

The horizontal layout was chosen to allow for easy reading of the country names and a straightforward visual comparison of the CFRs. Each bar's length corresponded to the CFR value, with longer bars indicating higher fatality rates.

This bar chart provided a visualization of the countries most severely impacted by COVID-19 in terms of fatality rates. By visualizing the top 10 countries, it highlighted significant variations in CFRs and underscored the diverse impacts of the pandemic across different regions.

Key Learnings

Through this project, I gained a deeper understanding of several key aspects of data visualization:

Raw data often needs to be transformed into calculated fields or smoothed using techniques like moving averages to reveal meaningful patterns and trends.

Effective use of color and design elements, such as distinct colors for different lines in a graph, is crucial for clarity and ease of interpretation.

Visualizations such as bar charts can provide powerful insights by enabling quick comparisons across categories, in this case, different countries.

Visualization is not just about presenting data but also about telling a story. Our visualizations helped convey the narrative of how the pandemic evolved differently across countries in the first six months.

Overall Experience

The project proved the importance of clear and effective data visualization in making complex data accessible and understandable. By transforming raw COVID-19 data into insightful visual representations, we were able to contribute to a better understanding how the pandemic was spreading in its first months and how CFR was increasing with each month.

My contributions to the project involved creating a line graph of moving average CFRs and a horizontal bar chart of top CFR countries, each providing unique insights into the pandemic's trends. The project was a valuable learning experience, enhancing my skills in data transformation, visualization design, and storytelling through data as well as working in a team and overcoming teamwork challenges.

Appendix 2: Individual Report by Raghukula Tilak Bolli

Line Chart: Global Spread of COVID-19

Visualization Details:
This line chart illustrates the global spread of COVID-19 from January to July 2020. It includes three lines representing confirmed cases (blue), deaths (orange), and recovered cases (gray). The X-axis shows the months, while the Y-axis represents the number of cases in millions.

Contribution:
I compiled the relevant data, created the line chart, and annotated key trends, such as the steep rise in confirmed cases starting from March 2020 and the contrasting slower increase in death cases. The chart visually highlights the rapid increase in confirmed cases, reaching over 200 million by July 2020, the significant rise in recovered cases to about 140 million, and the relatively lower increase in death cases, reaching around 20 million.

Key Takeaways:

Visualization Techniques: Using different colors and adequate labeling, such as tilting labels for readability, increases the graph's clarity and interpretability.

Accurate Data Gathering and Cleansing: Ensuring the accuracy of data is essential for credible visualizations and avoiding misleading conclusions.

Effective Data Transformation and Structuring: Converting data into suitable forms improves clarity and accuracy in comparisons and visualizations.

Appendix 3: Individual Report by Aditya Patrick Moses

Choropleth for global spread of Covid 19

Introduction: This report analyzes the global spread of COVID-19 using a choropleth map, which employs a gold-to-red color scheme to represent confirmed cases ranging from 586 thousand to 224 million, divided into five parts for clarity.

Key Observations:

General Trends:

- Most countries are in shades of gold, indicating fewer than 40 million cases, suggesting a relatively lower impact.

Significant Outliers:

- Russia and Brazil: Represented in darker shades with approximately 45.5 million and 89 million cases, respectively.
- United States: Marked in the deepest red with the highest number of cases at 224 million, highlighting a severe impact.

Interpretation and Insights:

- The map provides a clear visual of COVID-19's distribution and intensity.
- It differentiates countries with lower case numbers from those with severe outbreaks.
- Visualization aids in identifying hard-hit regions and can guide public health responses and resource allocation.

Conclusion

The choropleth map effectively illustrates the uneven global impact of COVID-19, highlighting significant outbreaks and providing insights for public health strategies and resource distribution.

Appendix 4: Individual Report by Sparshika Ajmaan Dinesh Kumar

Introduction

Our group project focused on analyzing the trends of COVID-19 from January to July 2020. The primary aim was to provide comprehensive insights into how the pandemic spread during its early months. This analysis involved various data visualization and statistical methods to highlight key trends and patterns in the COVID-19 outbreak. My contribution to this project specifically revolves around the visualization titled "Top 10 Countries by Confirmed Cases and Deaths." This explanation aims to break down the elements of the graph, describe its creation process, and interpret its significance. We extracted the data for this analysis from Kaggle, which included more than 49,000 rows and 10 columns, ensuring a robust dataset for accurate and detailed insights.

Graph Description

Data Representation:

- Bars: Two sets of bars are used for each country:
  - Blue bars: Represent the number of confirmed cases.
  - Gray bars: Represent the number of deaths.

Data Insights:

- The US has the highest number of confirmed cases and deaths.
- Brazil follows with significantly high confirmed cases but lower deaths compared to the US.
- Russia, India, Spain, the United Kingdom, Italy, France, Germany, and Iran also show varying levels of confirmed cases and deaths, with confirmed cases always higher than deaths.

Graph Generation Process

The graph was likely generated using a data visualization library such as Matplotlib in Python. Below is a step-by-step guide on how such a graph can be created:

Data Collection: Gather accurate information from reliable sources about COVID-19 cases and deaths in the top 10 affected countries.

Data Cleaning: Ensure accuracy by fixing errors, removing duplicates, and handling missing values in the collected data.

Data Transformation: Organize the data into a structured format suitable for analysis, including aggregating totals for confirmed cases and deaths for each country.

Data Validation: Validate the data's accuracy by cross-referencing with multiple sources and performing basic statistical checks.

Final Preparation: Organize the data into a format compatible with visualization tools like Matplotlib, ensuring all data points are correctly labeled for easy interpretation in the graph.

Key learnings:

Data Quality and Preparation: Ensuring accurate data collection and thorough cleaning is crucial for reliable visualizations and preventing misleading conclusions.

Effective Data Transformation and Structuring: Aggregating, normalizing, and structuring data into compatible formats enhances the clarity and accuracy of comparisons and visualizations.

Visualization Techniques: Using distinct colors and proper labeling, including tilting labels for readability, improves the clarity and interpretability of the graph.

Interpretation and Real-World Applications: Visualizations provide valuable insights for public health, inform policy decisions, and educate the public, promoting adherence to health guidelines and strategic resource allocation.

Significance

The graph highlights the severity of the COVID-19 pandemic in different countries, with the US showing the highest impact.

It underscores the need for robust healthcare infrastructure and pandemic preparedness.

Governments can use such data to allocate resources effectively, prioritize vaccination campaigns, and implement health measures.

Such visualizations help in educating the public about the pandemic's impact, promoting adherence to health guidelines.

Conclusion

The graph "Top 10 Countries by Confirmed Cases and Deaths" is a powerful tool for visualizing the global impact of COVID-19. By comparing confirmed cases and deaths across the ten most affected countries, it provides valuable insights into the pandemic's severity and informs public health strategies and policies. The process of creating this graph involves data collection, preparation, and visualization using tools like Matplotlib, making it a straightforward yet effective method for data presentation and analysis.

Appendix 5: Individual Report by Karan Bhuva

Introduction:

Our group effort focused on assessing COVID-19 patterns between January and July 2020. The major goal was to provide thorough insights into how the pandemic spread in the early months. This investigation used a variety of data visualization and statistical tools to identify major trends and patterns in the COVID-19 epidemic. My contribution to this project focuses on the visualization named "Top 10 and Bottom 10 Countries by Recovery Rate."

Role and Contributions:

My role involved collecting the necessary data, ensuring its accuracy, and using Tableau to design and implement these bar charts. This task required me to perform data cleaning and manipulation using Tableau to prepare the dataset for effective visualization.

Visualization Techniques:

I chose bar charts for this visualization because they clearly display comparisons across categories, in this case, countries. The dual-chart setup allows viewers to easily compare countries with the highest and lowest recovery rates from COVID-19, providing a straightforward visual summary of the data. I utilized a consistent color scheme to differentiate between the top and bottom countries, enhancing the visual appeal and readability of the charts.

Top 10 Countries by Recovery Rate:

This bar chart displays the recovery rates of the countries with the highest success, represented by green bars to symbolize health and recovery. The x-axis lists the countries, while the y-axis shows their recovery rates as percentages.

Bottom 10 Countries by Recovery Rate:

In a similar format, this chart shows the countries with the lowest recovery rates, using red to orange bars to indicate varying levels of concern based on the severity of the situation.

Reflection on Data Visualization:

Through this project, I learned the importance of visual aesthetics in conveying data insights effectively. The choice of color, scale, and the overall layout significantly impacts how the audience perceives and interprets the data. I also gained skills in data preprocessing and learned more advanced functionalities in Tableau, which I found invaluable for creating professional-level visualizations.

Technical Skills Enhanced:

My technical skills in using software tools like Tableau have substantially improved. I've become more adept at transforming raw data into informative, visually appealing charts that tell a story or highlight key insights. Moreover, working with real-world data has taught me to anticipate and overcome challenges related to data cleanliness and completeness.

Conclusion:

The contrasted picture of recovery rates highlights global disparities in pandemic response. Countries with greater recovery rates are likely to have better healthcare systems and more effective pandemic responses, whereas countries with lower rates may suffer owing to a variety of causes, including limited resources.

This research not only helped me improve my technical skills in data visualization, but it also highlighted the broader significance of data interpretation in addressing global health emergencies.