INFORMATION VISUALISATION

Group 55

Aditya Babu Kambalath Cherukovilakath- 2767285K
Kapil Arora- 2754963A
Laharika Tutica- 2797248T
Yashika Rastogi- 2802810R
Priyadarshani Chimanchode- 2759607C

School of Computing Science
Sir Alwyn Williams Building
University of Glasgow
G12 8RZ

Global COVID-19 Tracker:
A Dataset of Confirmed Cases and Fatalities

Youtube Video Link: https://youtu.be/b3XC_nyMN88

Task 1:

Link of the Dataset used:
https://www.kaggle.com/datasets/themrityunjaypathak/covid-cases-and-deaths-worldwide

The title of the dataset that we chose is "Global COVID-19 Tracker: A Dataset of Confirmed Cases and Fatalities". The information included in the data pertains to the global COVID-19 cases. It is an offline or statically available dataset with ordered attributes consisting of the quantitative values of various statistics of the novel pandemic cases.

This covers the number of confirmed cases in relation to the total number of tests conducted. Additionally, it presents the current active cases, the number of individuals who have recovered, and the total fatalities across all 231 countries. It also includes information on the population of each country, which can be useful in understanding the spread of the virus.

The variables in this dataset are:

1. Country: The country where the COVID-19 cases and deaths were reported.
2. Total Cases: The total number of confirmed COVID-19 cases in the country.
3. Total Deaths: The total number of deaths due to COVID-19 in the country.
4. Total Recovered: The total number of people who have recovered from COVID-19 in the country.
5. Active cases: The total number of active cases of COVID-19 in the country.
6. Total Test: The total number of tests performed in the country to detect the virus.
7. Population: The total population of each country.

Task 2:

Users might want to visualise this data in the form of bar charts, pie charts, choropleth maps or even treemaps, to gain insights into the trends and patterns of the pandemic. There are several actions that users might want to do when exploring this COVID-19 dataset. Some of them are:

1. Conducting statistical analysis:
   Users might want to conduct statistical analysis to gain a deeper understanding of the provided data. Statistical analysis involves visualising the insights in a way that it is clear and informative by extracting trends and patterns from the data. For example, the users might want to analyse the percentage of population affected or the rate of recovery and deaths based on the total number of cases in various countries.

2. Interactive Visualisations of the data:

As an advancement to the basic visualisation techniques that have been mentioned above, interactive visualisations for users can be made which allow the user to interact with the data in a more intuitive and engaging way, using various tools and platforms such as Plotly. For example, they might want to use hovers or click on the elements and gain insights in detail.

3. Filtering the data:
   Users might want to filter the data to gain insights on specific requirements from the entire dataset such as a subset of data values, greater or lesser than a particular number.

4. Comparing the data:
   Users might want to compare the data across different countries or regions to identify trends and patterns. For example, they might want to compare the total number of cases and deaths in the United States to those in China or Europe.

Task 3:

For the visualisation of Covid cases across the world, three core visualisation systems have been implemented which are as follows:

1. Bar chart - They are used to represent values of individual categories. In the graph, rectangular bars are used to illustrate the quantitative data, wherein each bar's height corresponds to its respective value corresponding on the y-axis..
2. Choropleth map - These maps are designed to visually display data patterns and facilitate comparison between various geographic regions using colour shading or patterns.
3. Treemap- The main objective of a treemap is to display the comparative magnitude of every element in a hierarchical structure and their connection with each other, it is generally used for large amounts of data.

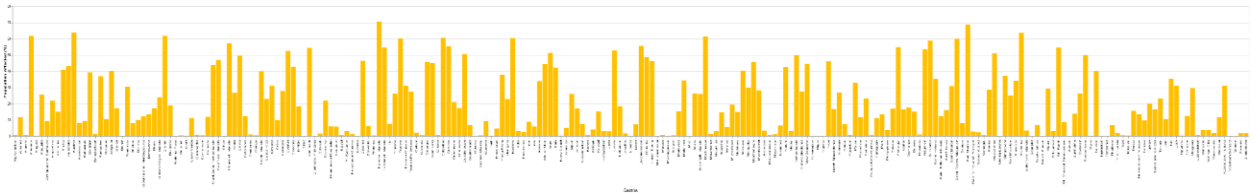


Figure 1: Data Visualisation using Bar chart

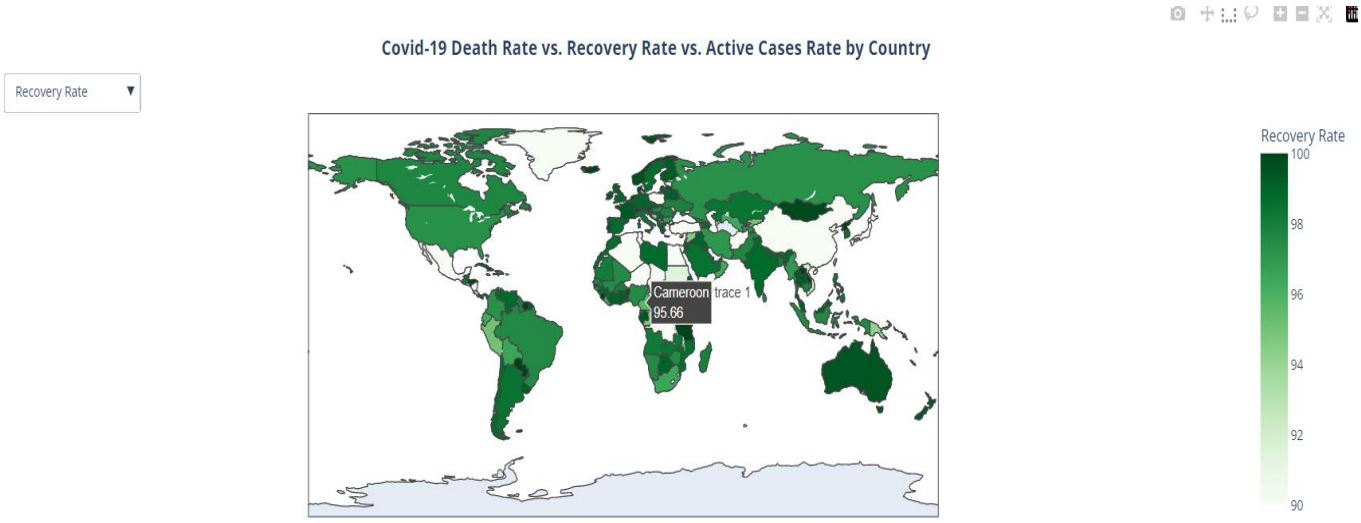


Figure 2: Data Visualisation using Choropleth

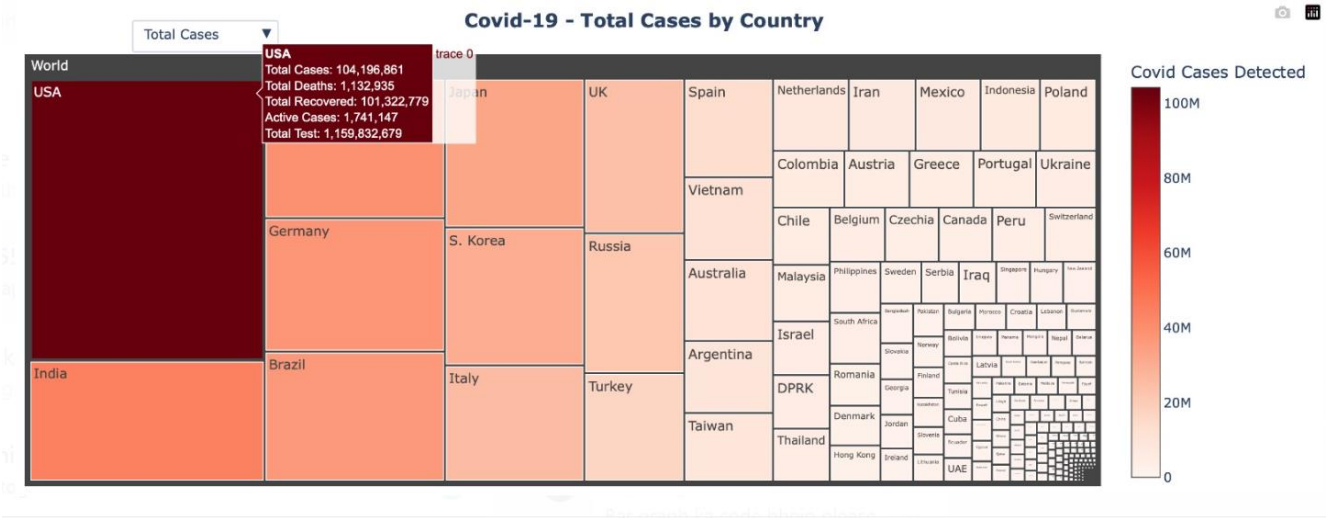


Figure 3: Data Visualisation using Treemap

Task 4: Generalised selection

System A: Bar Charts

The semantic structure of the proposal comprised four levels of abstraction: country, total population, total cases and then a breakdown of total cases into various subcategories like Active cases, Total Recovered and Total Deaths. As a part of the traversal policy, a country is selected, a total type, and a sub category. To enable interactive selection of data, the core system functionality has been extended by using Python and pandas library to read, manipulate, and visualise the data. Users have been able to select different sub categories, or filters and views using Plotly's interactive charts. The users have been enabled to perform generalised selection of data, based on the proposed traversal policy and semantic structure.

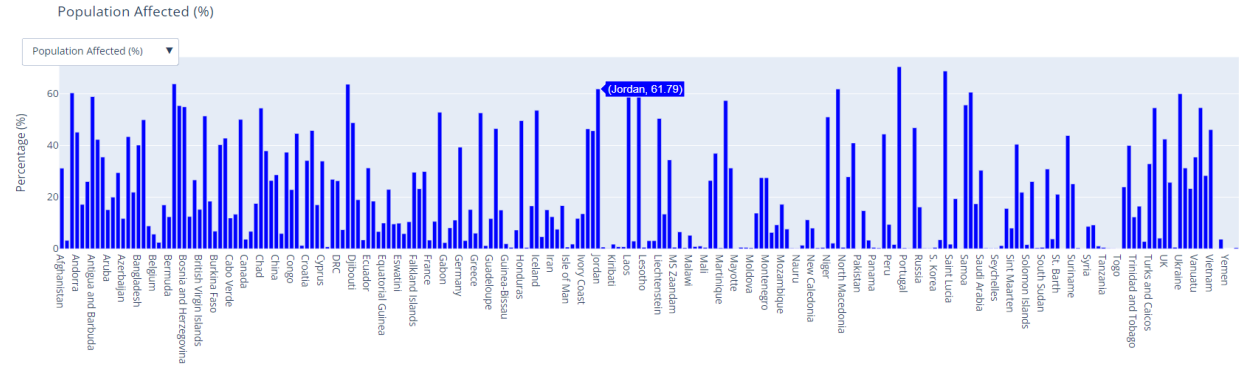


Figure: Interactive Bar chart of percentage population affected

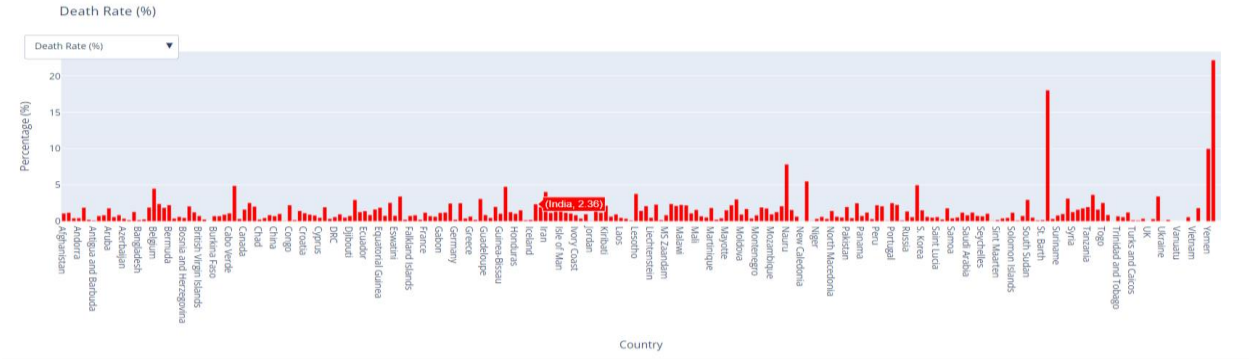


Figure: Interactive Bar chart of death rate

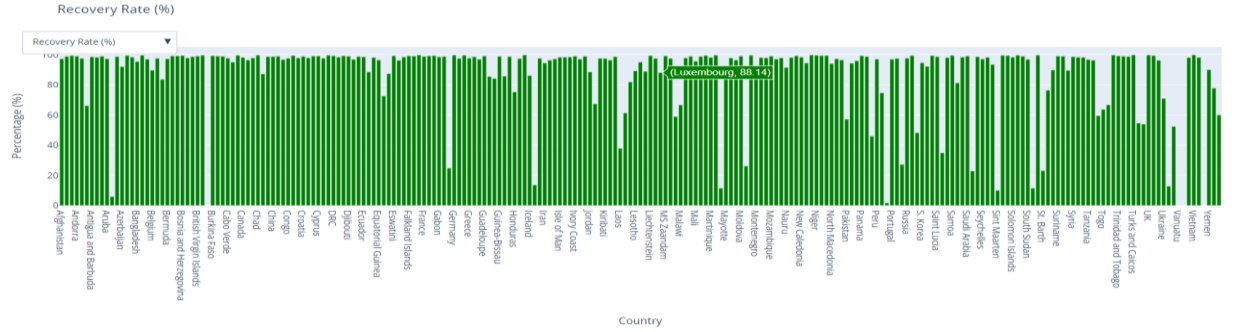


Figure: Interactive Bar chart of recovery rate

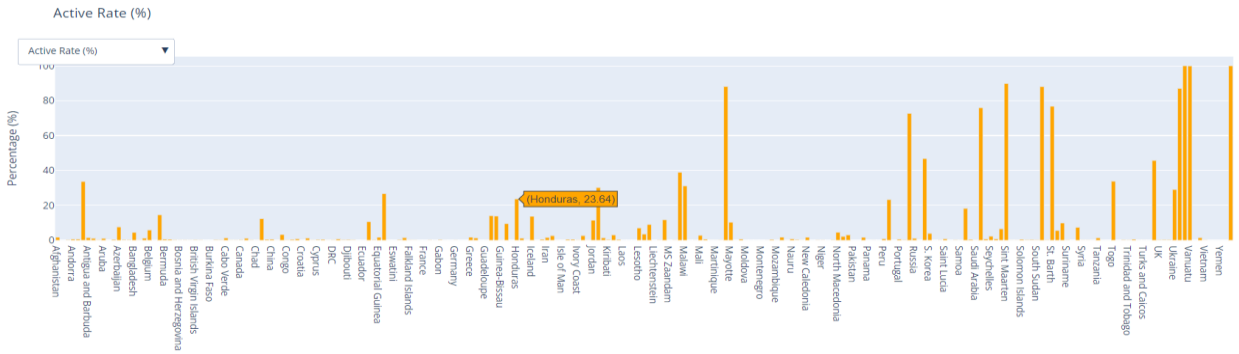


Figure: Interactive Bar chart of active cases

System B: Choropleth

Choropleth is an interactive world map tool which has been used to demonstrate the COVID-19 data for every country, based on the death rate, recovery rate, active cases rate, and affected population. The semantic structure included filters that allowed users to observe changing patterns across the mentioned four categories using aggregated data. Furthermore, multi-view filters allowed the users to view grouped data, such as regions with low recovery rates or countries with high active cases. Each country on the map has been represented by a colour that corresponds to the selected metric's data value, and the colour scale has been defined in such a way that the intensity of each colour represents a data value. The map also included a drop-down menu that allowed the users to choose the metric they wish to view and a tooltip that displayed the country and metric data.

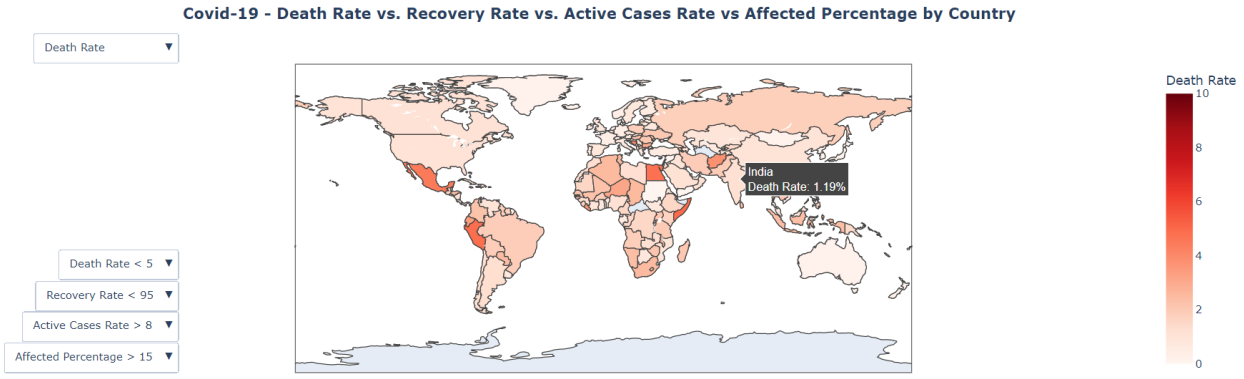


Figure: Interactive Choropleth with added filter showing death rate < 5

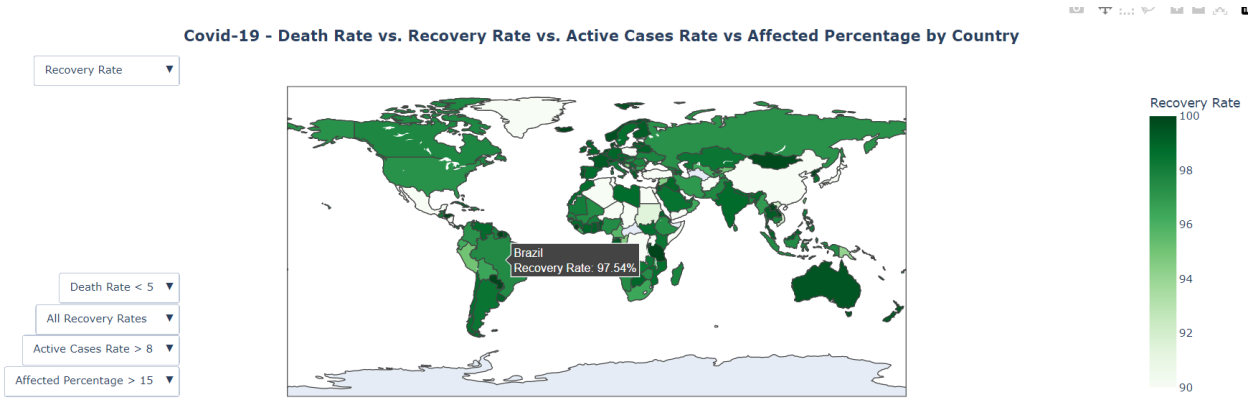


Figure: Interactive Choropleth with added filter showing all recovery rates

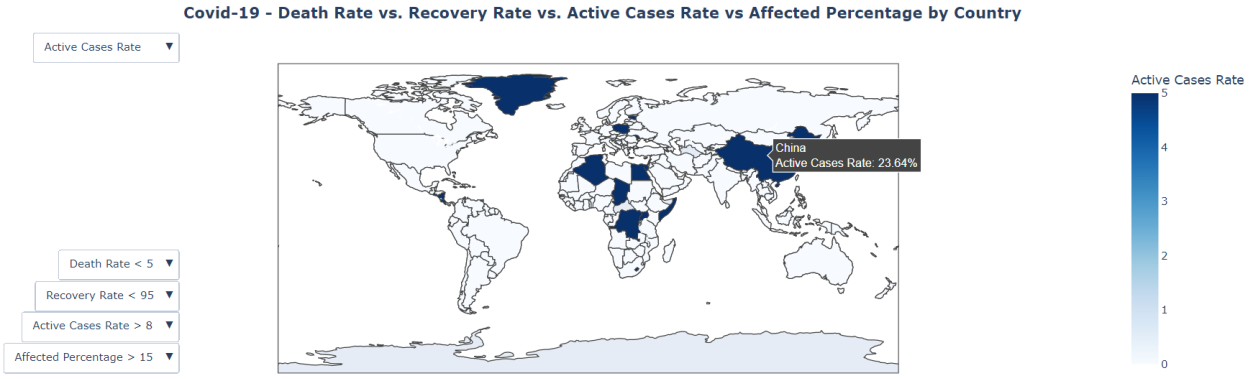


Figure: Interactive Choropleth with added filter showing active cases rate > 8

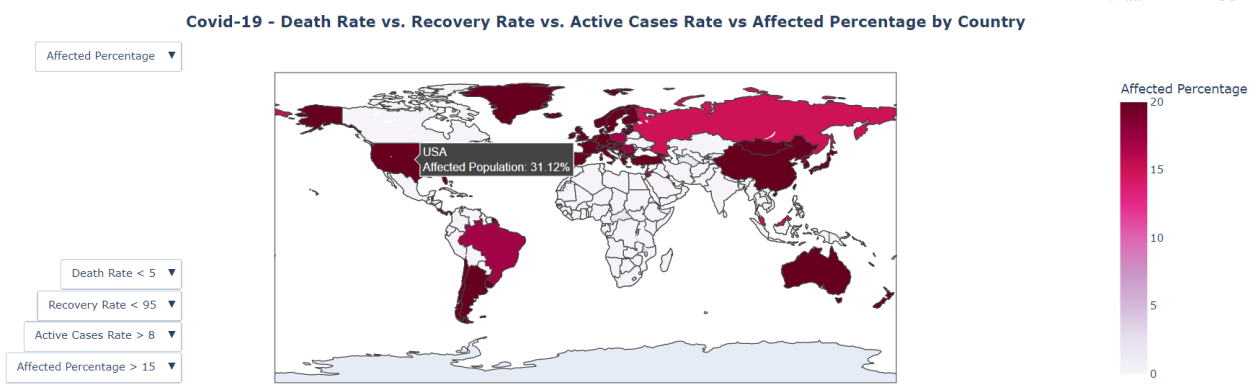


Figure: Interactive Choropleth with added filter showing affected population > 15%

System C: Treemap

The semantic structure here involves segregation of the data based on the country, filtering based on various subcategories and then changing the view according to the user's wish from a Treemap to a Sunburst diagram and vice versa. An interactive filtering treemap has been created to analyze the data by displaying the distribution of cases, deaths, or recovery rates across different countries using the tooltip. The treemap consists of rectangles that represent countries, with their size representing the corresponding metric. This allows users to quickly and easily identify the

most important values in a large dataset. In addition, this interactive treemap has been considered to change its view to a sunburst diagram of the same properties which provides a user-friendly way to analyze and visualize COVID-19 data.

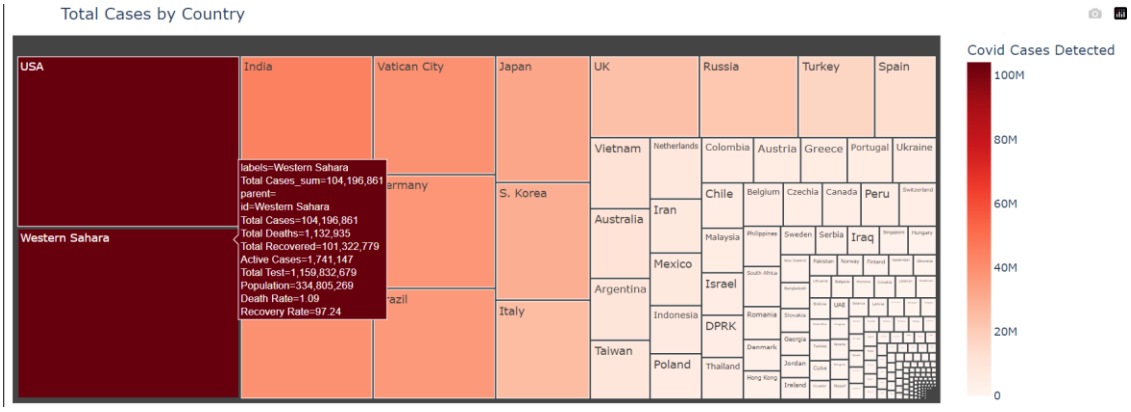


Figure: Interactive Treemap with added hover-template showing the statistics of the hovered Country

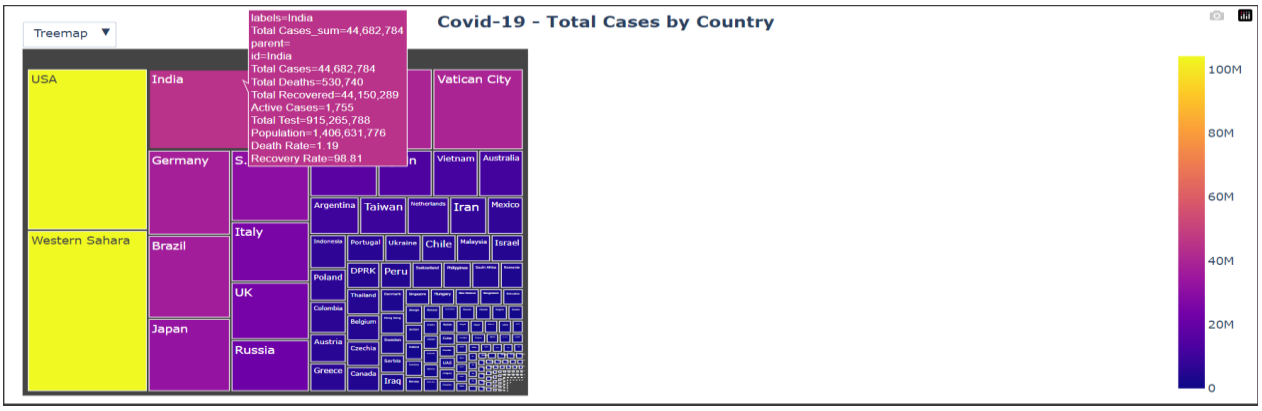


Figure: Interactive Treemap with added hover-template and option to change view showing the statistics of the hovered Country

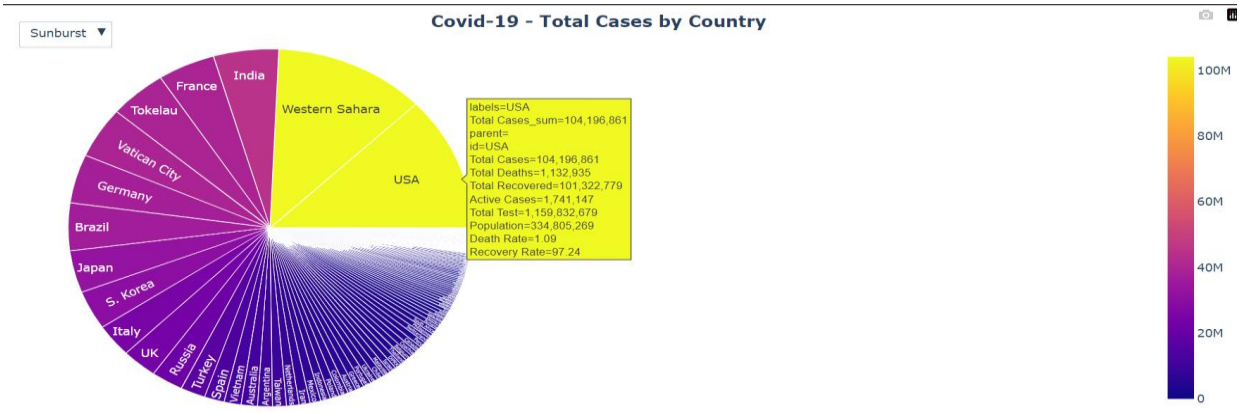


Figure: Interactive Sunburst Diagram with added hover-template and option to change view showing the statistics of the hovered Country

Task 5: Video Demonstration

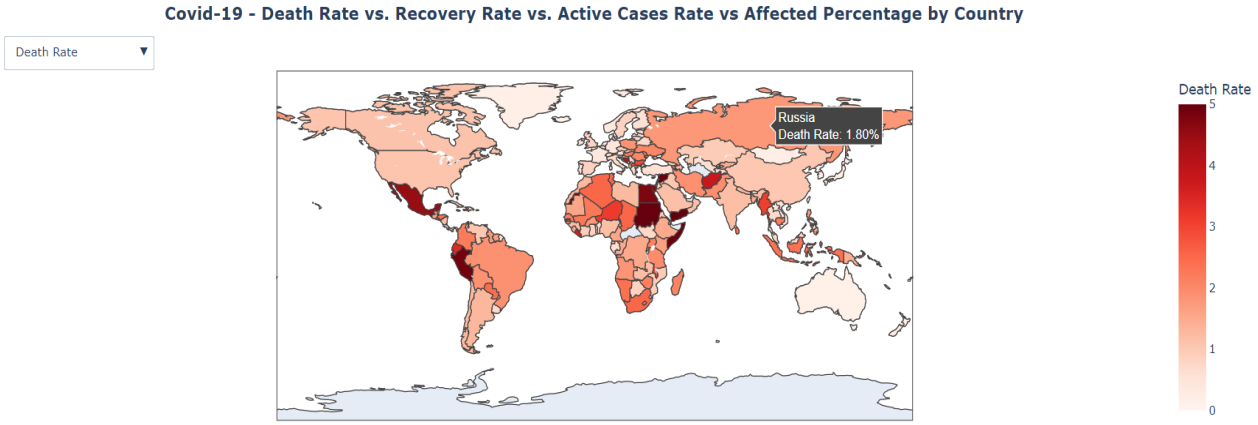
Youtube Video Link: https://youtu.be/b3XC_nyMN88

Task - 6: Design Comparison

Following is the description of six different Information visualisation design decisions:

1. Data granularity:

   The design decision on the level of data granularity displayed in the visualisation is vital in conveying the intended message. The hovered bar chart considered displaying data for comparatively fewer categories of data, whereas the choropleth map has been ideal for displaying data for all geographical regions at once. The treemap showed hierarchical data for many categories at a time, and the sunburst view displayed the same data in a radial manner. In terms of this dataset, choropleth map is being chosen to be the best choice for data granularity as this dataset relates to the geography of the world.

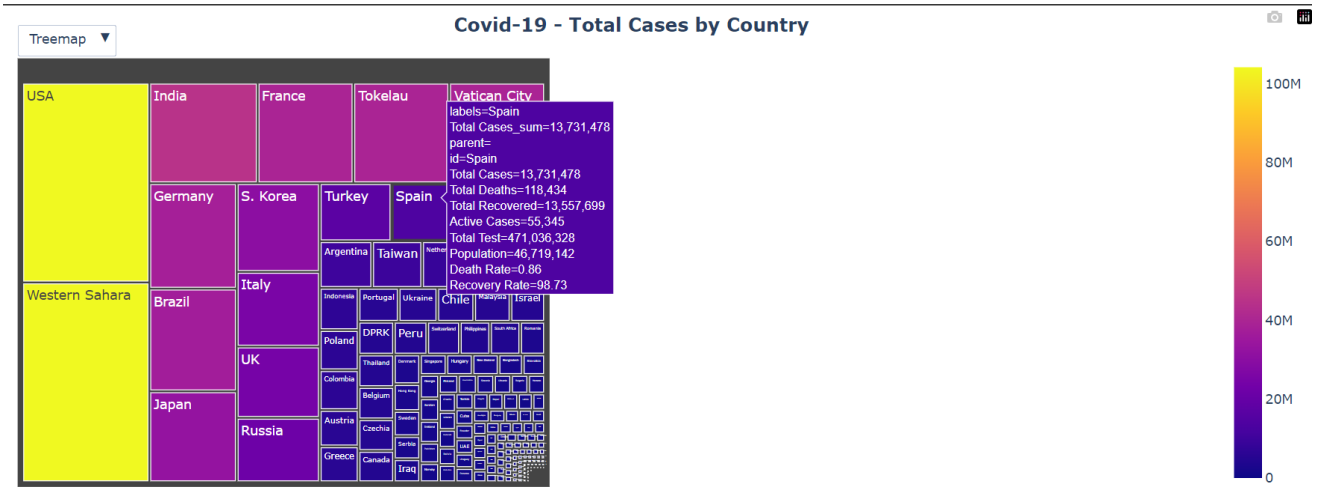


   From the above Choropleth map of the death rates of the countries presented accordingly, it can be clearly visualised that the regions with a comparatively darker shades of the colour have a greater death rate.

2. Data Encoding:

   The decision on how to encode the data can also affect the visualisation's effectiveness. The bar charts encoded the data using bar lengths, whereas choropleth maps encoded the data using colour shades. Treemaps used different encodings for the parent and child nodes, such as size and colour, making the visualisation more understandable. Likewise, sunburst diagrams also used colour and arc size to encode the data, making it easy to identify various trends too. For this design decision, choropleth map as well as Treemap are being considered to be the best choice since it gets easier and more understandable to visualise a system with the colour intensities and the area covered rather than a simple bar chart.

From the Treemap shown below, it could be clearly visualised that there is a colour-scale which differentiates the countries based on their pandemic statistics and rates while on the other hand, the bar chart depictions use a uniform shade of colour and the only understanding one could attain through the visualisation of data is the bar lengths.



3. Interactivity:

The level of interactivity provided by the visualisation had a significant impact on the user experience. Hovered bar charts provided tooltips when the user hovers over the bars, thereby, displaying the statistics of the country and to resolve the data granularity issue, it also had an interactive feature to zoom-in and out of the bar chart to have an elaborative view of the bars whereas choropleth maps enabled the user to filter data by selecting a range of values. Treemaps enabled the user to zoom in on individual nodes, and even switch views with a sunburst diagram that enabled the user to interactively highlight specific arcs. According to the dataset and this design decision, a treemap with a sunburst diagram view is the most interactive design decision because it highlights each and every minute detail while the user hovers over the arcs.

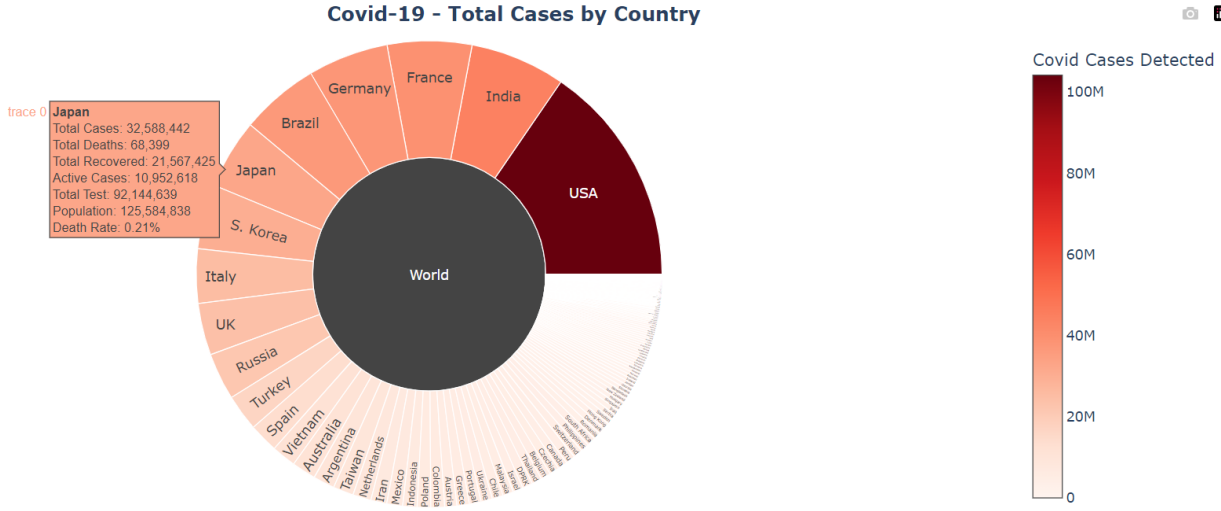


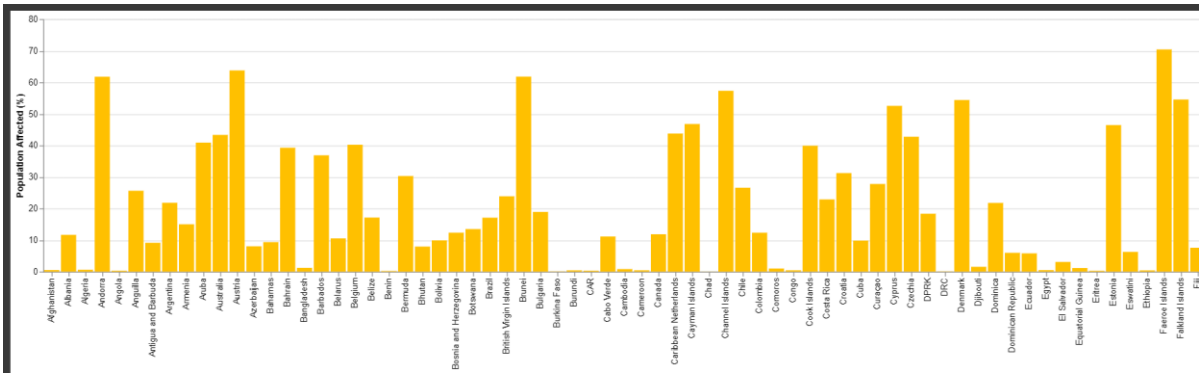
Figure: Interactive Sunburst Diagram with added hover-template showing the statistics of the hovered Country

4. Scale:

   The scale of the visualisation impacted the levels of details being displayed to a great extent. A larger scale would undoubtedly provide more granularity but it also depends on the type of visualisation system being used. For the hovered bar chart, an overall scale of 0-100% was used while the choropleth map implemented a different range of percentages for each of the views generated i.e.,

   - Death rate view : 0-10%
   - Recovery rate representation : 90-100%
   - Rate of active cases view : 0-5%
   - Rate of population affected: 0-20%

   The Treemap used an entire range of 0-100M in order to showcase a hierarchy of the countries based on the calculated percentages. In this perspective, Bar chart is the best decision choice since it uses a simple and easily understandable scale for the visualisation of data while choropleth has a different range for scaling each of the multi-view compositions and in terms of treemap, it gets very difficult to visualise the almost negligible data due to sizing.



From the above bar chart, one can easily understand that the countries have been plotted against the percentage of the population affected over a range of 0-100%.

5. Data Labelling:

   Data labelling refers to the process of adding text labels to the visualisation to provide additional context or information. Here, we are considering various attributes or columns for the representation of data in the visualisation systems. Hovered bar charts and choropleth maps typically had limited options for data labelling, while treemaps with a sunburst view provided detailed labelling options for individual nodes or arcs. The tree map/Sunburst Diagram would be the best choice for this design decision due to its varied range of data display, however, the choice of data labelling depends on the user's requirements.

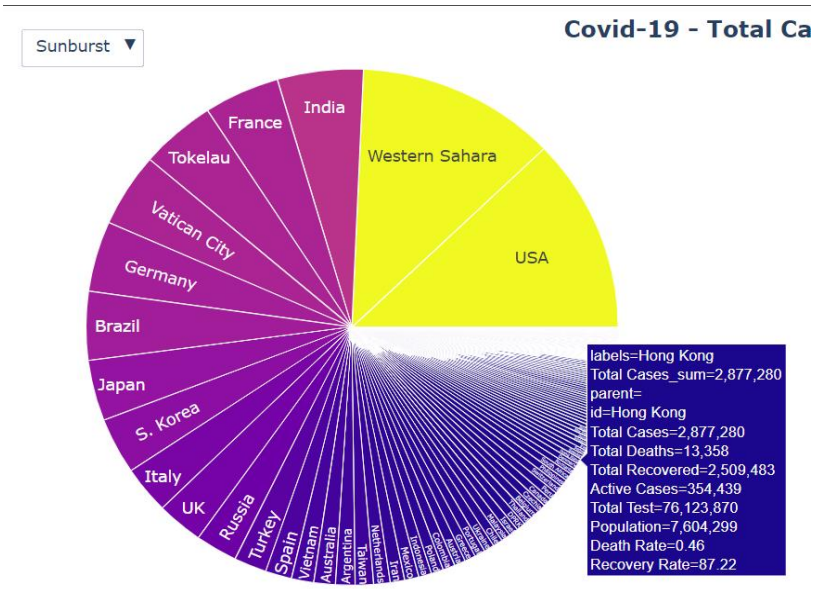


Figure: Interactive Sunburst Diagram with added hover-template showing the statistics of the hovered Country

6. Layout:

   The layout of the visualisation affected how the information is presented to the user. Hovered bar charts and choropleth maps typically have a linear layout, with the data displayed in a single row or column. Treemaps with a sunburst view offer a more hierarchical layout, with data displayed in nested nodes. Hence, the treemap/Sunburst diagram offered the most interactive design.

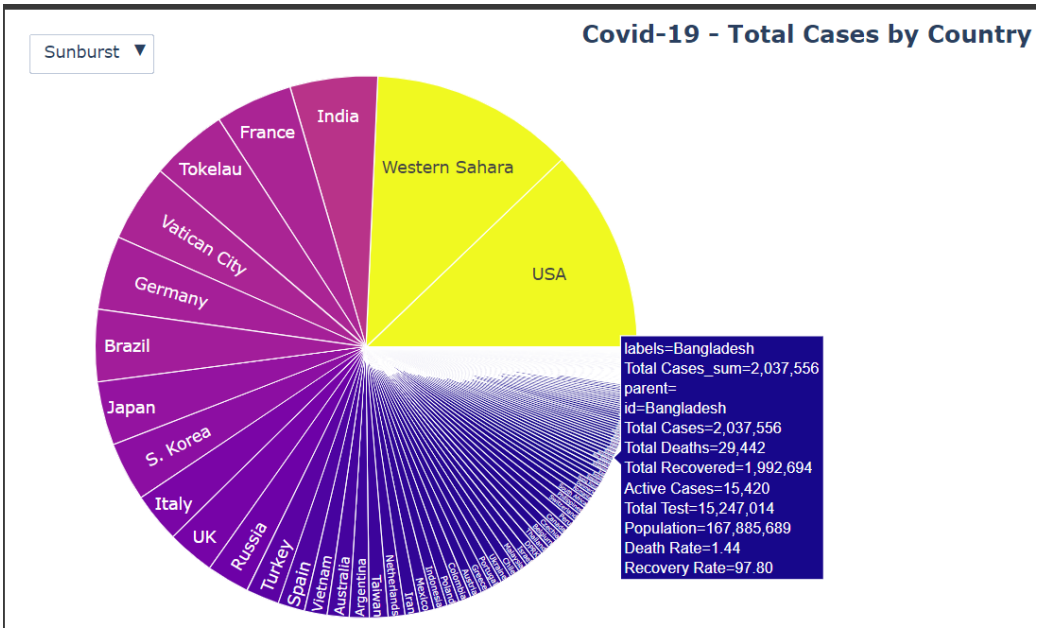


Figure: Interactive Sunburst Diagram with added hover-template showing the statistics of the hovered Country

Task 7: COVID Visualisation Comparison: A Comparative Evaluation of Three Visualization Systems

Introduction

The COVID-19 pandemic has led to a significant increase in the amount of data available on cases, deaths, recoveries, and active cases across different countries. To help users make sense of this data, several visualisation systems have been developed that allow users to explore and analyse COVID-19 data in various ways. In this report, we will evaluate and compare three different visualisation systems that aim to provide insights into COVID-19 data: a bar graph, a choropleth map, and a tree map.

Evaluation Methodology

To evaluate the three visualisation systems, we conducted a user evaluation where each system was evaluated by at least five people. We used a standardised evaluation methodology that involved the following steps:

Preparation: We prepared a list of questions that aimed to assess the effectiveness, usability, and usefulness of each visualisation system. These questions were designed to be answerable on a scale of 1-5, with 1 indicating "strongly disagree" and 5 indicating "strongly agree."

Participants: We recruited participants who had some knowledge of COVID-19 and data visualisation. The participants were diverse in terms of age, gender, and educational background.

Evaluation: We provided participants with access to each of the three visualisation systems and asked them to explore the data and answer the questions. Each participant was asked to evaluate each system in terms of effectiveness, usability, and usefulness. We created a google form and asked the participants to evaluate the models for themselves. We gave them some small tasks to perform on each model such as finding the country with the highest number of COVID-19 cases. The participants rated the model based on how easily they were able to interpret the information provided in the models.

Post-Task Evaluation: After completing the tasks for each system, participants were asked to rate the effectiveness, ease of use, and overall user experience of each system using a 5-point Likert scale. Following are the responses collected from the users via Feedback Forms:

Data Collection: We collected the answers to the questions from each participant and recorded their responses on a spreadsheet. We then aggregated the responses and analysed the data to determine which visualisation system performed the best.

Data Analysis:

The data collected from the survey was analysed to identify trends and patterns in the responses. Mean scores and standard deviations were calculated for each evaluation criterion (effectiveness, ease of use, and overall user experience) for each visualisation system.

Results and Analysis

Based on our evaluation methodology, we found that the three visualisation systems performed differently in terms of effectiveness, usability, and usefulness.

Effectiveness

In terms of effectiveness, the tree map was the most effective visualisation system according to our participants. The tree map provided a clear and intuitive representation of the COVID-19 data, allowing users to quickly identify hotspots and trends across different regions. The choropleth map was also rated highly effective, but the bar graph was the least effective in terms of providing clear insights into COVID-19 data.

Usability

In terms of usability, the tree map was rated the most usable visualisation system. The participants found the sunburst variation of the tree map to be particularly helpful in identifying the most affected regions. The choropleth map was also rated highly usable, while the bar graph was the least usable of the three systems.

Usefulness

In terms of usefulness, the tree map was rated the most useful visualisation system. The sunburst view of the tree map made it easier for users to identify the most affected regions and the severity of the situation. The choropleth map was also rated highly useful, while the bar graph was the least useful of the three systems.

Overall, our evaluation showed that each visualisation system had its strengths and weaknesses, and the choice of the best system would depend on the specific needs of the user. For instance, if the user is interested in comparing COVID-19 data across different countries, the bar graph may be the best choice. If the user is interested in exploring the distribution of COVID-19 cases across different regions, the choropleth map may be the best choice. Finally, if the user is interested in an interactive and engaging visualisation, the tree map may be the best choice.

In conclusion, our evaluation highlights the importance of using effective and user-friendly visualisation systems for COVID-19 data. Such systems can help users make sense of the data and provide valuable insights that can aid in the fight against the pandemic.

Task 8. Future work

Based on the results of our evaluations, we have identified several areas for improvement in the three visualisation systems created for the COVID-19 dataset. In this section, we will describe in detail the changes we would make to improve each system.

For the bar graph visualisation, one area for improvement is the use of more contrasting colours. While the current colours used in the graph are visually pleasing, they do not provide enough contrast, making it difficult for some users to distinguish between the different categories.To address this, we would modify the colour scheme to use more contrasting colours, such as dark blue for active cases and bright green for recovered cases.

Another area for improvement is the labelling of the y-axis. Currently, the y- axis label simply reads "Percentage," which does not provide enough context for users to understand what the percentages represent. We would modify the label to read "Percentage of Population Affected," providing users with more information on what the percentages represent.

For the choropleth map visualisation, we identified several areas for improvement. One issue is that the map does not provide any context or information on the countries being displayed, which can make it difficult for users to understand the data being presented. To address this, we would add labels to the map displaying the names of each country and include a legend that explains the colour coding of the map.

Another area for improvement is the use of colour in the map. While the current colour scheme is visually appealing, it is not effective at conveying the differences in the data being displayed. We would modify the colour scheme to use a darker colour for countries with a higher percentage of cases, and a lighter colour for countries with a lower percentage of cases.

Overall, the changes we would make to these visualisation systems are aimed at improving their usability and effectiveness in conveying the data being presented. By addressing these issues, we can create more user- friendly visualisations that are better suited for a wider range of users.

References:

https://en.wikipedia.org/wiki/Treemapping

https://xdgov.github.io/data-design-standards/visualizations/bar-chart#:~:text=Bar%20charts%20can%20be%20visualized,demonstrate%20differences%20between%20each%20bar.

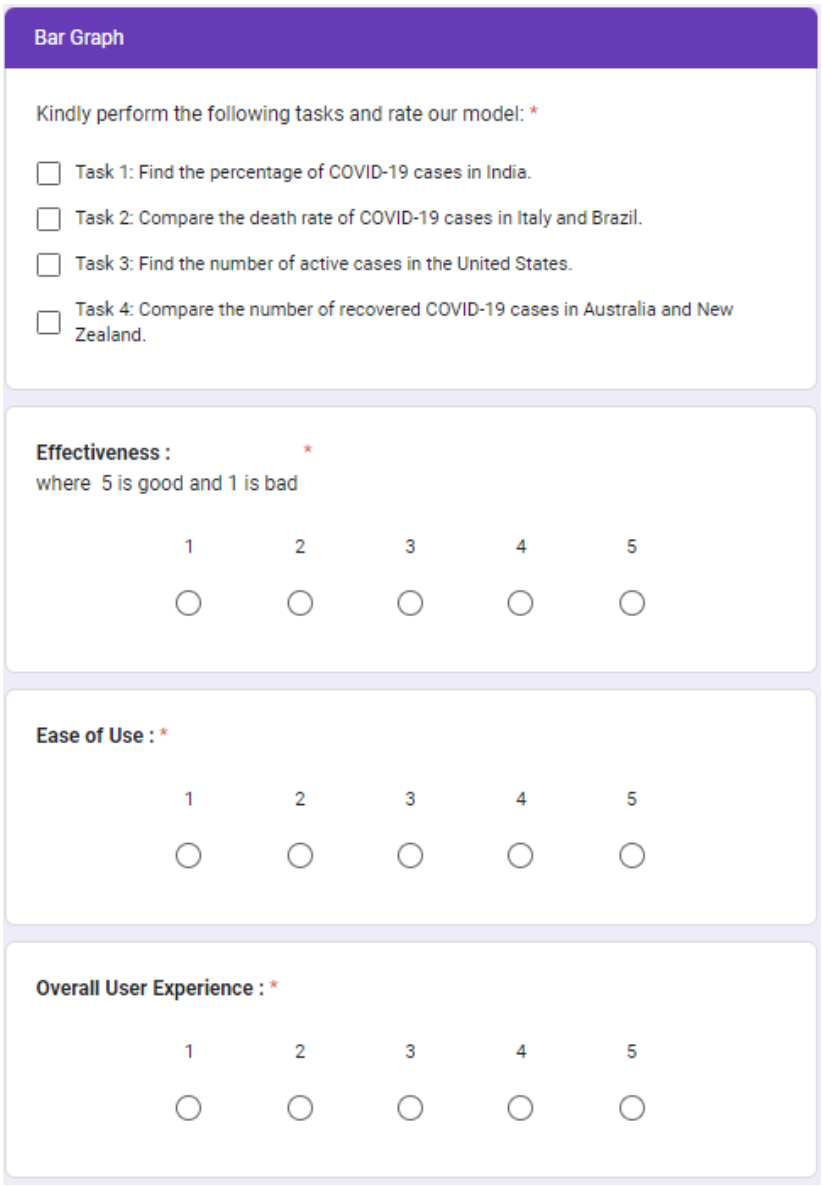
https://www.googleadservices.com/pagead/aclk?sa=L&ai=DChcSEwiz6JvKuPz9AhULjGgJHcL8CRgYABABGgJ3Zg&ohost=www.google.com&cid=CAESaeD2E1pQGgJsmIz-zf-W-plBlDHzBDob87Ffbz0xTj5xK_vb5Cuvqvx8MEkTSCEoVm0seAqDLV6Lfd1ZOtEXjwMgxa6UPrxB0sGZDAtH6xnEQ-CslFGPYgrfrfF8-xeq99H7Vj1tGwam7w&sig=AOD64_1BVYVjBzgFTqSya0V0OXYgEeq-WA&q&adurl&ved=2ahUKEwjShZPKuPz9AhVkgv0HHTtAARYQ0Qx6BAgJEAE

https://www.data-to-viz.com/graph/choropleth.html

https://www.qlik.com/us/data-visualization/interactive-data-visualization#:~:text=Create%20Interactive%20Visualizations-,What%20is%20Interactive%20Data%20Visualization%3F,better%2C%20data%2Ddriven%20decisions.

Appendix:

Google Forms



Bar Graph

Kindly perform the following tasks and rate our model: *

- [ ] Task 1: Find the percentage of COVID-19 cases in India.
- [ ] Task 2: Compare the death rate of COVID-19 cases in Italy and Brazil.
- [ ] Task 3: Find the number of active cases in the United States.
- [ ] Task 4: Compare the number of recovered COVID-19 cases in Australia and New Zealand.

Effectiveness : *

where 5 is good and 1 is bad

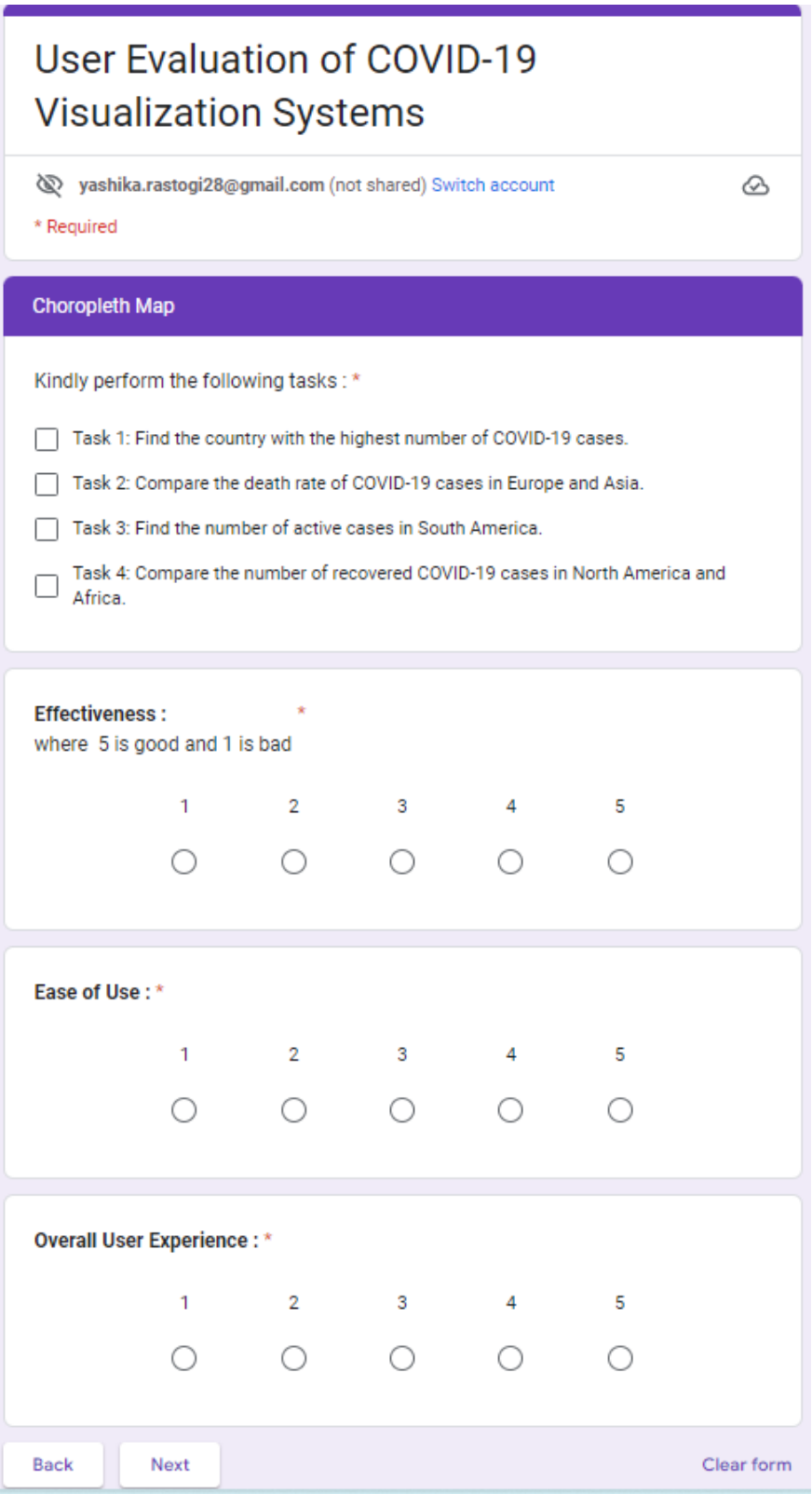
1 2 3 4 5

Ease of Use : *

1 2 3 4 5

Overall User Experience : *

1 2 3 4 5



User Evaluation of COVID-19 Visualization Systems

yashika.rastogi28@gmail.com (not shared) Switch account

* Required

Choropleth Map

Kindly perform the following tasks : *

- [ ] Task 1: Find the country with the highest number of COVID-19 cases.
- [ ] Task 2: Compare the death rate of COVID-19 cases in Europe and Asia.
- [ ] Task 3: Find the number of active cases in South America.
- [ ] Task 4: Compare the number of recovered COVID-19 cases in North America and Africa.

Effectiveness : *

where 5 is good and 1 is bad

1	2	3	4	5
○	○	○	○	○

Ease of Use : *

1	2	3	4	5
○	○	○	○	○

Overall User Experience : *

1	2	3	4	5
○	○	○	○	○

Back Next Clear form

Feedback from Google forms

	BAR CHART			CHOROPLETH MAP			TREEMAP		
Name	Effectiveness	Ease of Use	User Experience	Effectiveness	Ease of Use	User Experience	Effectiveness	Ease of Use	User Experience
Becca	3	3	4	5	4	5	3	4	3
Peter	3	4	4	5	4	5	3	4	4
Riya	4	4	4	3	5	4	4	5	5
Mahek	5	5	5	4	3	4	5	4	4
Emma	3	3	4	3	3	4	4	2	3