DATA VISUALIZATION

CSCI55200 – SPRING 2023

TOPIC – EFFECTS OF COVID-19 ON LIFE

Team Members –

TEJAS HELWATKAR

SIVA KALYAN GANGALA

PRADNEYA BHOSALE

Term Project Report

I. BACKGROUND

We all know that the Covid-19 pandemic has had a serious impact on almost every sector of life, and still, we are witnessing some of those effects in our present life. The pandemic hit the world around the end of the year 2019, and infected millions of people (all age groups) across the world, which resulted in the huge number of deaths. COVID-19 has resulted in significant sickness, hospitalizations, and fatalities, particularly in older people and those with existing medical issues. Significant economic disruptions brought on by the epidemic include numerous businesses closing or operating at lower levels. With the loss of jobs and means of subsistence, the global economy has been badly hit. An important development in the fight against the pandemic has been the creation of potent COVID-19 vaccinations. However, there are still disparities in vaccine access around the world. In this data visualization project, our goal was to showcase the effects of Covid-19 on mainly four areas, which are the spread of the infection around the world, symptoms experienced, vaccinations, and deaths.

II. DATASETS

We collected the data to implement our visualizations using the websites Kaggle, Owid (Our World in Data).

1. Dataset from Our World in Data

Confirmed cases and deaths data: The World Health Organization Coronavirus Dashboard is where this information is gathered. The dataset for cases and deaths is updated every day. The data's time/date stamps show when the WHO last updated it. There is a lag between reporting to WHO and the dashboard update since it takes time to review and validate the incoming data. After these times, counts and corrections will be carried over to that region's subsequent reporting cycle. Delayed reporting for any nation, territory, or region may lead to the presentation of pooled counts for several days with a retrospective update to counts from earlier days to appropriately reflect trends. The WHO may fix significant data mistakes found or reported more frequently.

Our data on hospitalizations and admissions to intensive care units (ICUs) is compiled by Our World in Data from official sources. The Our World in Data team gathered this data from official reports; you can find more information in our post on COVID-19 testing, which also includes a checklist of questions to help you understand the testing data, details on the testing's geographic and temporal coverage, and a thorough breakdown of the sources used in each individual nation. The Our World in Data team has gathered this information about COVID-19 vaccinations from official reports. The information was gathered from a number of organizations, including the World Bank, the Global Burden of Disease Study, the Blavatnik School of Government, and the United Nations.

2. Datasets from Kaggle:

    i. COVID – 19 Symptom Checker

https://www.kaggle.com/datasets/iamhungundji/covid19-symptoms-checker

Based on some pre-defined standard symptoms, these data will assist in determining whether or not a person is suffering from a coronavirus condition. These signs are based on recommendations made by the Indian Ministry of Health and Family Welfare and the World Health Organization (WHO).

Seven key variables that will affect whether a person has coronavirus disease or not are included in the dataset, with descriptions of each one as follows, country, age, symptoms, experience of any other symptoms, severity, and contact. With all these categories. Data: Two CSV files have been submitted, one is raw data and the other is Cleaned-Data. Cleaned data is being used for the project as it is refined and better to get efficient results.

    ii. COVID – 19 World Vaccination Progress.

https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress?select=country_vaccinations.csv

Each day, data from the covid-19 GitHub repository of Our World in Data is gathered, merged, and uploaded. Data on vaccinations at the country level is compiled into a single file. This data file is then combined with a data file about locations to include information about immunization sources. There is also a second file providing information on the manufacturers.

Key variables in this dataset are as follows, Country, Country ISO Code, Data, total number of vaccinations, Total number of people vaccinated, the total number of people fully vaccinated, daily vaccinations, total vaccinations per hundred, the total number of people vaccinated per hundred, the total number of people fully vaccinated per hundred, number of vaccinations per day, daily vaccinations per million, vaccines used in the country.

The visualizations with the respective datasets used in them are listed below:

Choropleth Graph – Our world in data

Line Graph – Our world in data

Spider Chart – Symptoms Checker (Kaggle)

Lollipop Graph – Symptoms Checker (Kaggle)

Stacked Bar Graph – World Vaccination Progress (Kaggle)

Circular Bar Graph – Our world in data

We have cleaned the datasets according to our requirements. The cleaned datasets are also attached along with the original datasets in the zip file. We have also attached the data cleaning file.

III. IMPLEMENTATION OF VISUALIZATION TECHNIQUES

We implemented all of our visualization techniques using the D3 technology. The techniques implemented and their purpose of implementation are listed in the table below.

Table 1

Name of the technique	Purpose
Choropleth Map	To visualize Covid-19 cases worldwide.

Line Plot	For visualizing the trends of Covid cases in ten countries of various population sizes.
Spider Chart	Used for visualizing people experiencing symptoms per million affected people.
Lollipop Chart	To visualize the various Covid-19 symptoms experienced by people in numbers in Korea.
Stacked Barplot	For visualizing the progress of top-10 vaccinated countries so far.
Circular Barplot	To visualize total number of deaths in ten countries of various population sizes.

IV. METHODOLOGY

Choropleth Map: A typical choropleth map has a particular theme and it uses different shades or colour patterns to represent different values of a specific attribute within defined geographic boundaries such as countries, states or provinces. The colour or shading pattern in a choropleth map reflects the magnitude or density of the attribute which is being mapped. The darker or more brighter values highlight the higher values whereas the lighter values indicate lower values. Using Choropleth map we have visualized the number of Covid-19 cases notified in different countries through different colours. These are the steps we followed to generate our first visualization in the form of Choropleth map.

- Since the Choropleth map is a geographic visualization, we first loaded the geographic data and the data to be visualized in the script.
- Then we transformed the geographic data into the screen coordinates by creating a projection.
- On the basis of the range of values in our dataset, we set the colour scale of our choice.
- Then we bound our data with the geographic attributes by using the data() and join() methods.
- Finally, we added the geographic attributes to the SVG container using the append() method.

Line plot: It is commonly used visualization technique to show trends or changes in data over time. It consists of series of data points connected by straight line segments. The x-axis represents a continuous variable such as time whereas the y-axis represents the value of the variable being measured. Each point on the line represents the value of the variable at a specific point in time. Line plots have various applications such as displaying data of stock prices, temperature readings, and website traffic. In addition to this, they can be used to reveal patterns in the data such as seasonal trends or periodic fluctuations.

Steps:

- We started off by converting our dataset into an array of objects, such that each object will represent a data point.
- Then we created a SVG element and also set its dimensions, and margins.
- Then we created a line generator using D3's line function.
- Similar, to our previous visualization, we appended a path element to the SVG element and to set its features.
- The final step was creation of the axis, and adding axis labels.

Spider graph chart: A spider graph chart is a 2-D chart type designed to plot one or more series of values over multiple quantitative variables. Each variable has its own axis and all axis are connected at the centre of the figure. This chart consists of a series of spokes, each representing one variable, and set of rings, each representing a data point. Spider charts are commonly used in doing data analysis, and comparison. Other applications also include performance evaluation of different options. In our project, we have used this technique to visualize the number of people experiencing different symptoms out of 1 million people. For e.g Korea. The creation of this visualization is similar to the Line plot, and the Lollipop chart.

Lollipop chart: This chart combines elements of the bar chart and scatter plot visualization techniques. In this chart, each category is represented by a vertical line also called as stem that extends from a horizontal axis to a point indicating that category value. The line length represents the value of the category, while the marker size represents the magnitude of the difference between the categories. Specific applications of the lollipop chart involve displaying data with large number of categories as they allow easy comparison between individual categories. We used this visualization for visualizing Covid-19 symptoms experienced by patients.

Stacked Barplot**:** A stacked bar plot in D3 displays the composition of a whole dataset as well as the proportion of each component. It consists of rectangular bars stacked on top each other where the height of each bar represents the total size of the group it represents. Each segment of a bar represents the proportion of a group, which belongs to a specific category. Commonly stacked bar plots are used to represent data which is divided into multiple categories. Additionally, they show how different categories contribute to the overall dataset and is also used to compare the composition of multiple datasets at once. We have used this technique to visualize top 10 countries with their vaccination rates. For creating this visualization, we have used the same process as that of the previous visualizations.

Circular Barplot: This technique displays categorical data utilizing a circular layout. It is also known as circular histogram or a polar bar chart. In a circular bar plot, the data is divided into groups and each group is represented by a bar. The bar are arranged in a circular fashion, where each bar occupies a portion of circle proportional to the group frequency it is representing. So circular bar plot can be used to display time series data or data with periodic patterns. We can also use this to visualize large groups of data in a compact representation by maintaining the relative proportions. We have used this to visualize number of deaths in various countries. The process for creating Circular Bar plot is same as Stacked bar plot.

V. RESULTS

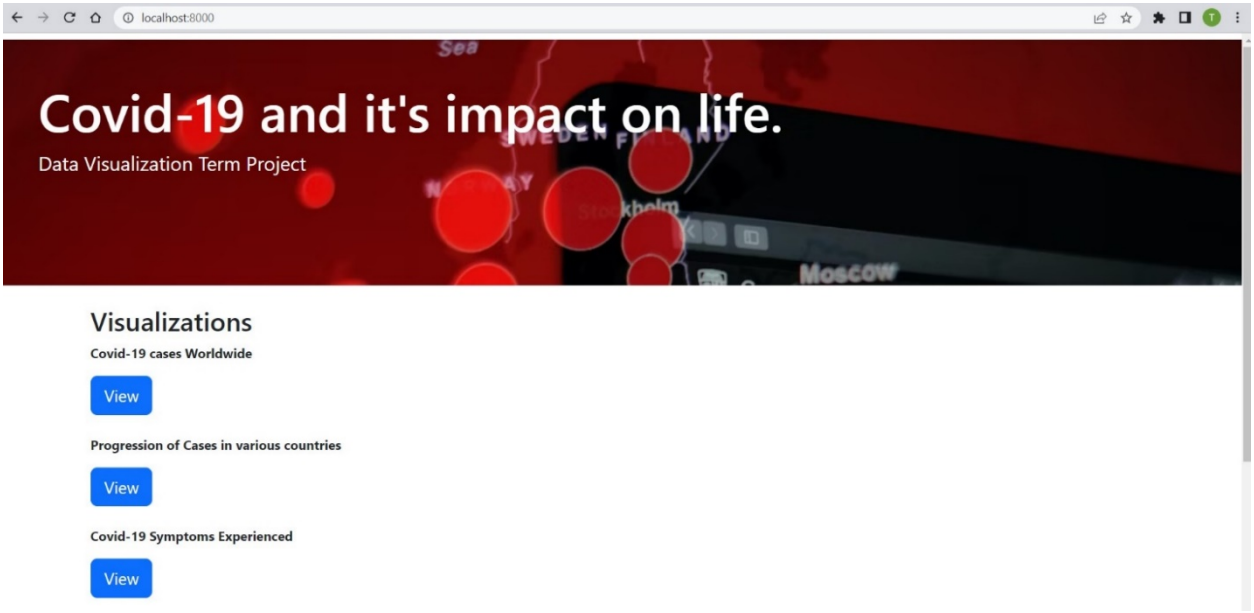

Figure 1: Main page with hyperlinks to view our visualizations.

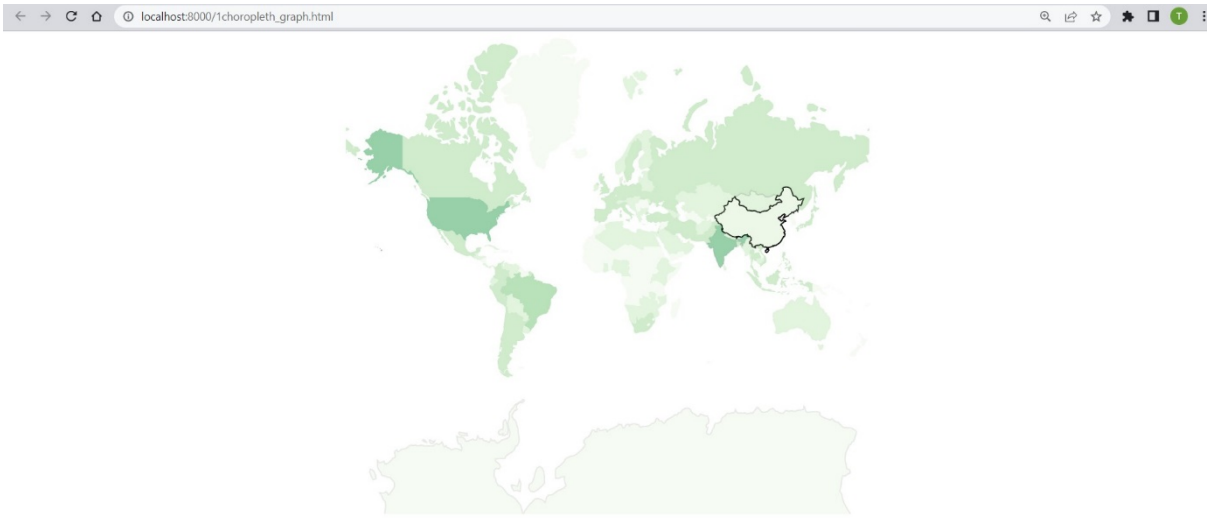

Figure 2: Choropleth Map

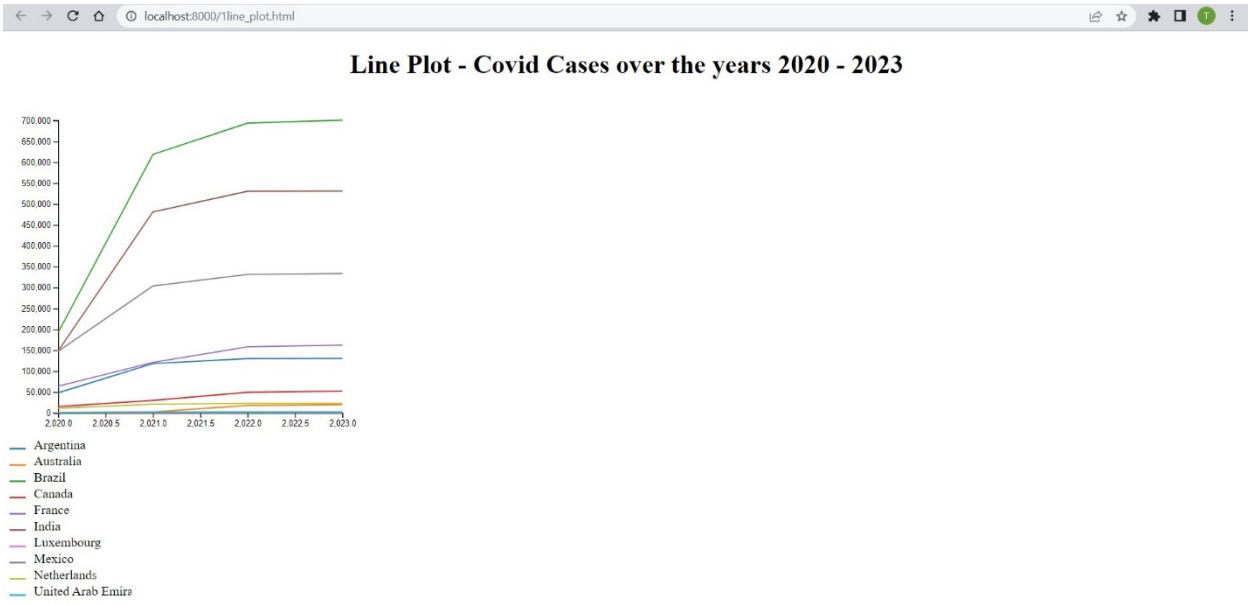

Figure 3: Line Plot

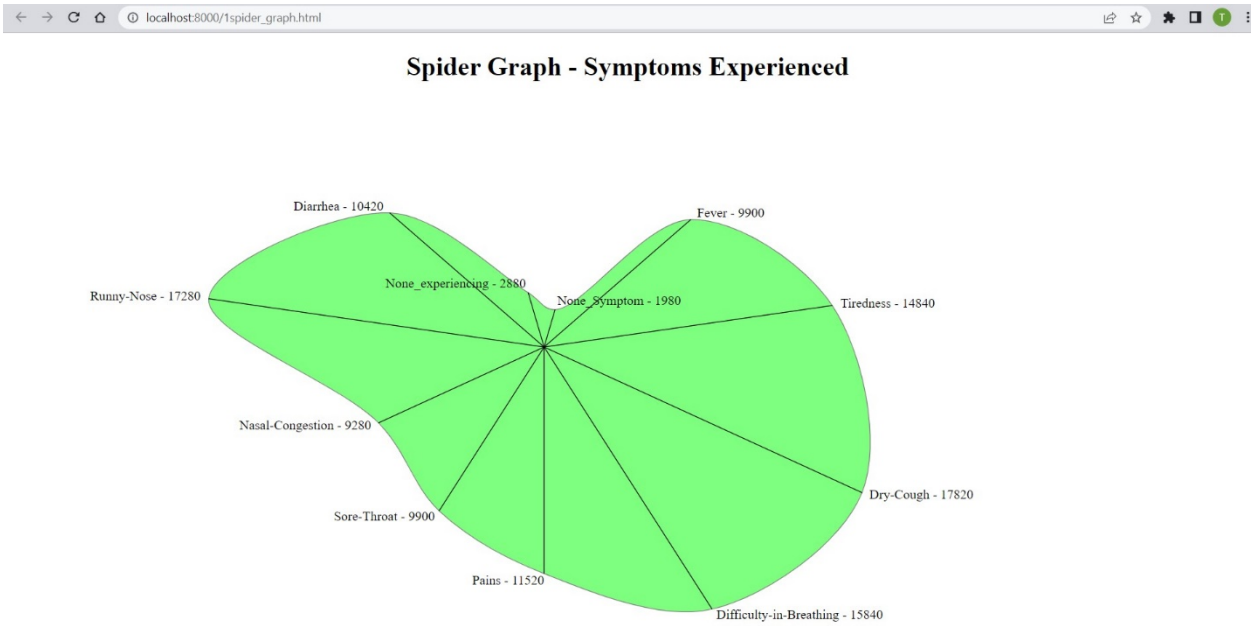

Figure 4: Spider Chart

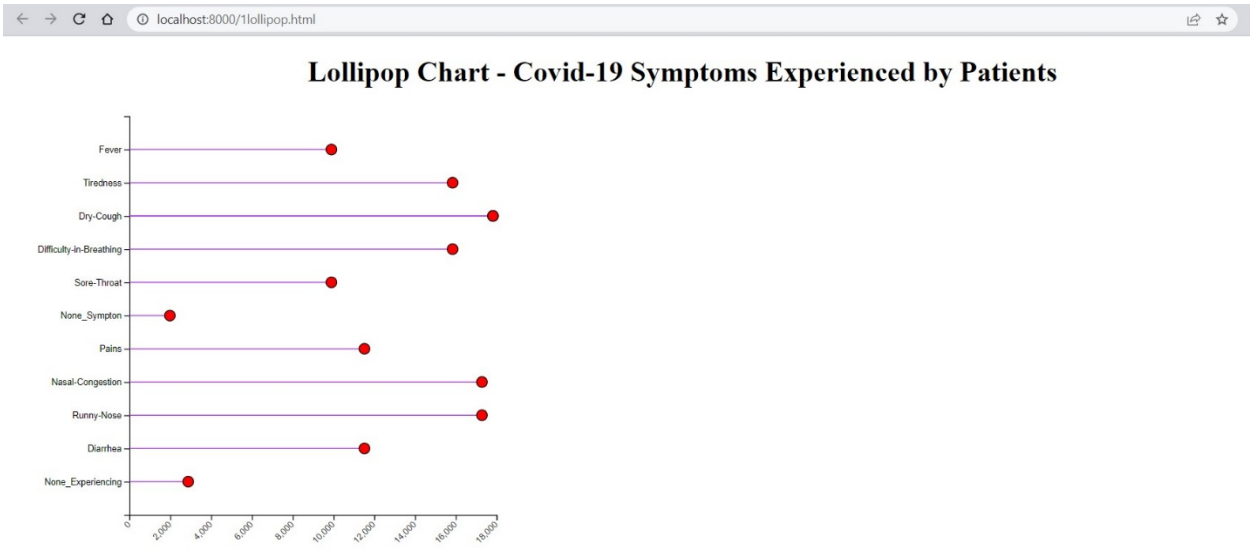

Figure 5: Lollipop Chart

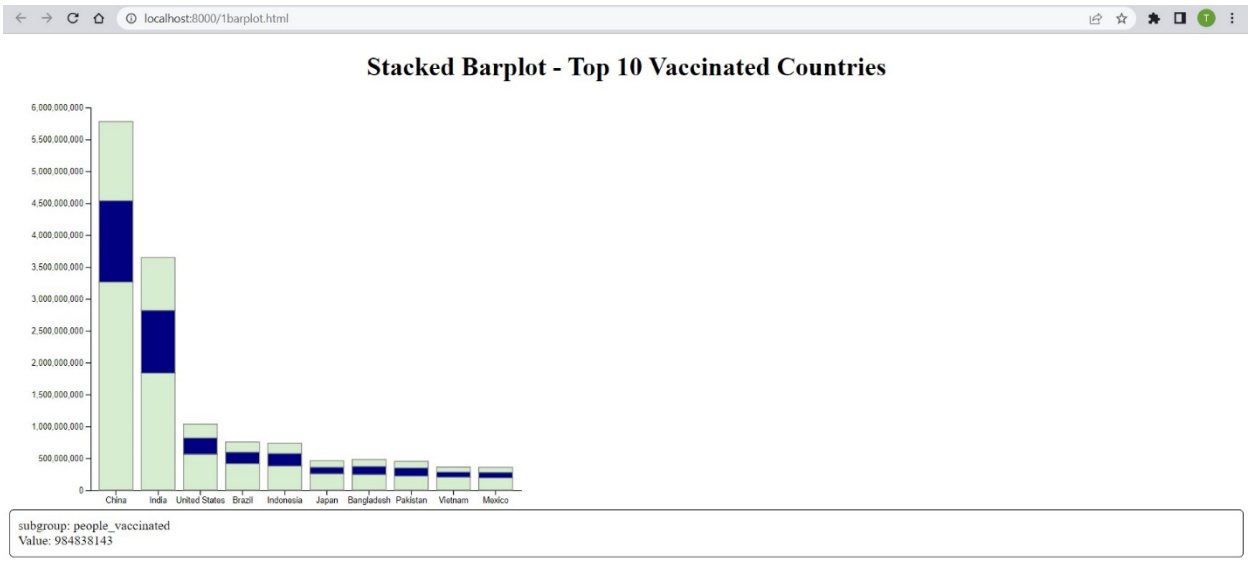

Figure 6: Stacked Barplot

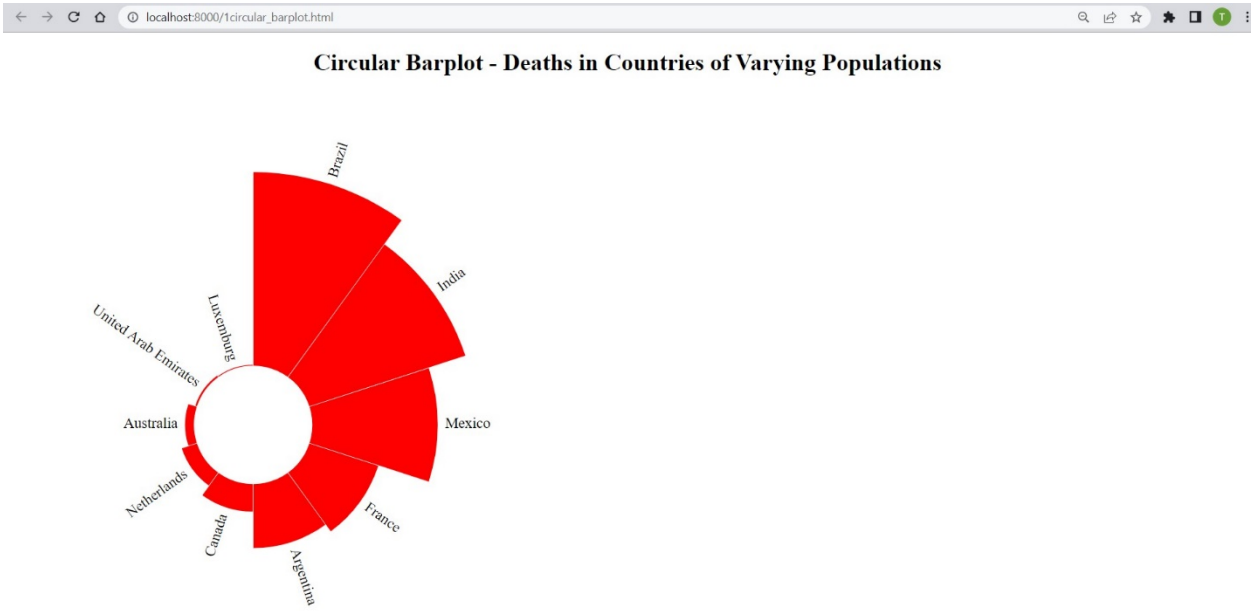

Figure 7: Circular Barplot

VI. INSIGHTS

Choropleth Map: Our first visualization was Choropleth map. Since it is a geographic visualization, we used it to show number of Covid cases worldwide. So, if we hover over a particular country, we can view the total number of Covid cases in that country. For e.g. China has 98315 Covid cases as shown in Fig.2.

Line plot: Our second visualization on the list was Line plot. We have utilized this visualization to view Covid Cases in different countries such as Argentina, Australia, Brazil, Canada, France, India, Luxembourg, Mexico, Netherlands and United Arab Emirates (UAE). In this visualization we showed the variation in the Covid cases for each country for 4 years starting from 2020 to 2023. We also used different colour scales, for each country so that it becomes easy to compare values of total covid cases with that of other countries.

Spider Chart: We used this chart to visualize the number of people experiencing different symptoms out of 1 million people. So based on our visualization in Fig.4 we can see that there are 10420 people who experienced Diarrhea, 9900 people who experienced Fever, 2880 people who didn't experience any of the symptoms, so on and so forth. So we have showed 11 different categories of people experiencing different types of symptoms.

Lollipop Chart: Through this we visualized the Covid-19 symptoms experienced by patients such as tiredness, fever, dry cough, difficulty in breathing, sore throat, none symptom, pains in joints, nasal congestion, runny nose, diarrhoea, and people experiencing nothing. So, we can get an insight that there are maximum people experiencing dry cough, nasal congestion, and runny nose, whereas people having no symptom are minimum.

Stacked Barplot: This is a very interesting visualization which we used to show the progress of top 10 vaccinated countries. Based on Fig.6 we can say that the total number of people which have been vaccinated are 984838143. Also, we can see that China tops the list with 5750000000 people successfully vaccinated. On the other hand, there are only 400000000 people in Mexico who have been vaccinated.

Circular Barplot: We used this visualization to view the deaths in countries with varying population. From fig. 7 we can note that Brazil has had maximum loss of deaths, then following it is India, Mexico and so on. Luxemburg has had minimum deaths. But this again varies depending on the population of the country.

VII. CONCLUSION

From these visualizations we are successfully able view the impact of the novel Corona virus on the lives, health, and deaths of millions of people across the world. During the pandemic times governments and health organizations around the world had implemented various measures such as lockdowns, social distancing, and vaccination drives to prevent the spread of novel Corona virus. The positive part of this is that many countries showed interest in technological innovation and also collaborated with each other to develop vaccines. This pandemic showed us how we can face difficult and challenging times together.