2020

COMPARISON OF GLOBAL PANDEMICS:

COVID 19 Vs SARS

Team Members :

ARAVIND KUMAR SANKARA NARAYANAN

VIPUL KOTA

VISHNU PRASANTH REDDY

COMPARISON OF GLOBAL PANDEMICS : COVID 19 Vs SARS

1. Abstract :

All proven facts are based on Data. Data is necessary to prove any future occurrences based on the statistical point of view of the past occurrences and particularly for the emergence of phenomena such as the COVID-19 outbreak. A viral pandemic is not a scenario in which intuition can provide a sense of how the spread is advancing. Collecting, analyzing, sharing, and ultimately making use of data is what is needed.This collected data needs to be efficiently conveyed to all sorts of individuals and groups, from lay people to experts and everything in between. Visualizing collected data can make its dissemination easy, and can help others understand quickly what has taken others so long to collect and analyze. After all, a picture is worth a thousand words. This project dealt the same by comparing the spread of two of the major pandemics the world has ever faced such as the COVID 19 and the SARS.

2. Primary Objective :

The primary objective of this project is to visualize the effect of COVID 19 in comparison with SARS. this comparison is done to visualize the effect both the viruses brings into the world and how much this pandemic situation could affect the future . This analysis is achieved by designing an interactive dashboard that demonstrated the effect of both viruses. in addition to that the growth factor and death rate caused by the pandemics has been visualized along with identifying the patterns of the spread of viruses using timeseries.to sum up all the effects across the globe an interactive choropleth map chart of the countries affected by the pandemics using geographical location.

3. Purpose :

The purpose of this project is to provide a statistical overview of the pandemics that occurred in the past and occuring in the present to gain insights.the observed insights can help the affected countries to take appropriate measure to stop the spread.the observed insights can help the unaffected or less affected countries to take appropriate measures such as imposing lockdown, improving medical facilities, creating awareness and imposing travel ban to affected countries

4. Project Lifecycle :

The following are the different phases followed while working in this project ,

- Data Collection
- Data Preprocessing
- Exploratory Data Analysis
- Visualization
- Building Dashboard

4.1 Data Collection :

Since the data is based on a global pandemic which was happened before and currently, it is important to have up to date information regarding the data and from trusted sources. The data used in the course of this project is fetched from John Hopkins University one of the renowned universities having the largest Bio- medical research center who keep tracking of the current COVID 19 happenings and all other previously occurred pandemics. Normalization of total cases was done to convert the varying scale of both SARS and COVID to [0,1].

4.2 Data Preprocessing :

The preprocessing stage includes converting the timestamp to Date format that is vital in order to plot any time series data. Feature creation technique was used to create a new attribute from the existing attributes that helped to visualize new inference.

4.3. Exploratory Data Analysis:

To map features to accurate aesthetics exploring the data is a fundamental step. It is essential to explore for outliers, null values which are a drawback for accurate visualization. Based on the requirement of visualization new grouped tables and features are created. All the attributes are grouped in all possible ways that could help visualize the data from different angles. We

explored the data by grouping all the countries based on date, and grouping all the regions based on each country.

4.4 Visualization:

4.4.1. - Plot 1: Geo Visualizing the spread of virus across the Globe:

Considering the spread of viruses being a global pandemic across the world, it is vital to visualize the effect globally to identify and segregate the countries that are highly affected. A Geographical visualization has been designed using Choropleth Map that helped us to identify the affected countries based on the frequency of cases in each country and color the respective area/region based on the range of values the number of cases falls into.

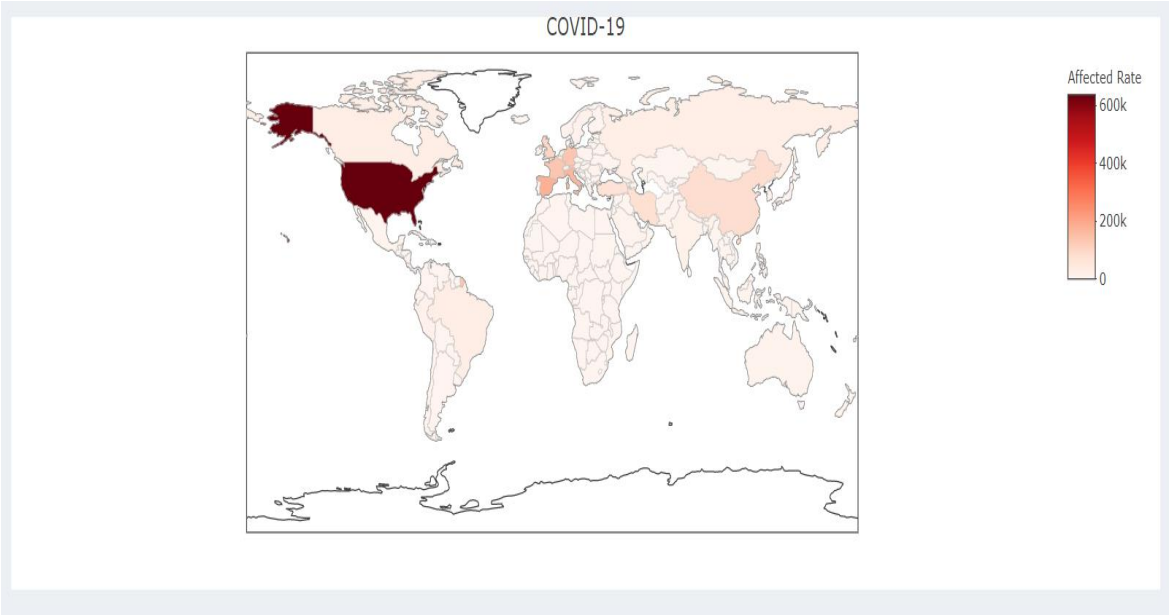

Figure 1.1 Choropleth Map on spread of COVID 19

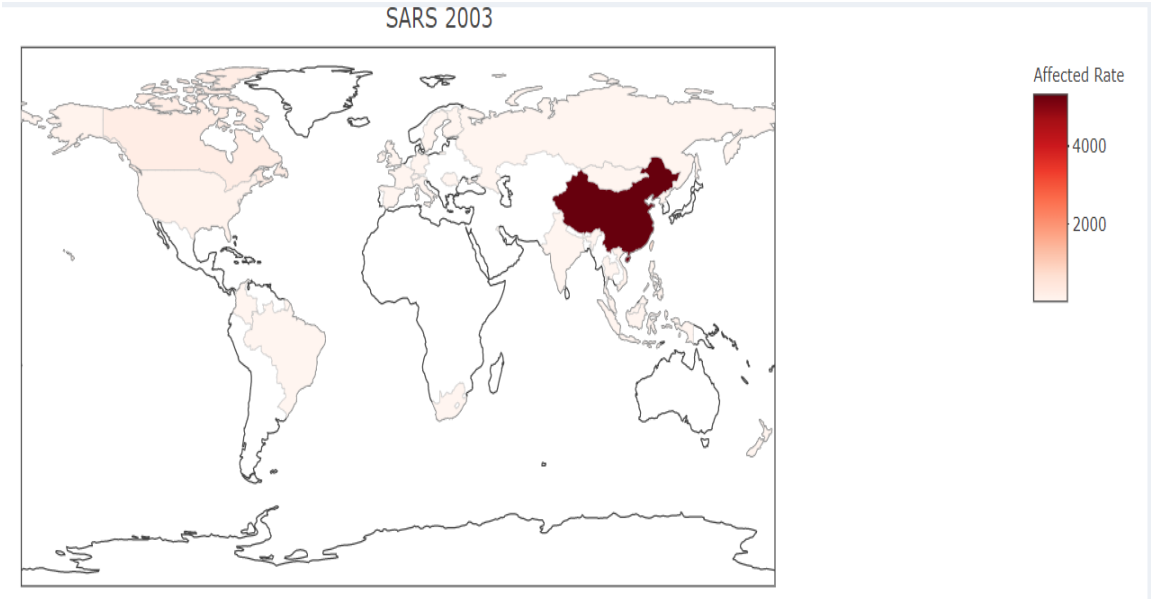

Figure 1.2 Choropleth Map on spread of SARS

The following figures (1.1 and 1.2) provides a globular view on the effect of viruses across the world. This choropleth map is achieved using plotly one of the powerful R package that uses the geographical location based on latitude and longitude to locate the respected countries and the color palette changes from light to dark based on the range of cases.

4.4.2 - Plot 2 : Visualizing the growth factor from the affected rate :

The spread of viruses can be determined using the growth factor. Growth factor helps to determine the change over time. Bar chart found to be the best fit for track the changes over time. In this case it helped our analysis in determining the growth factor. Growth factor is the factor by which the quantity multiplies over time.

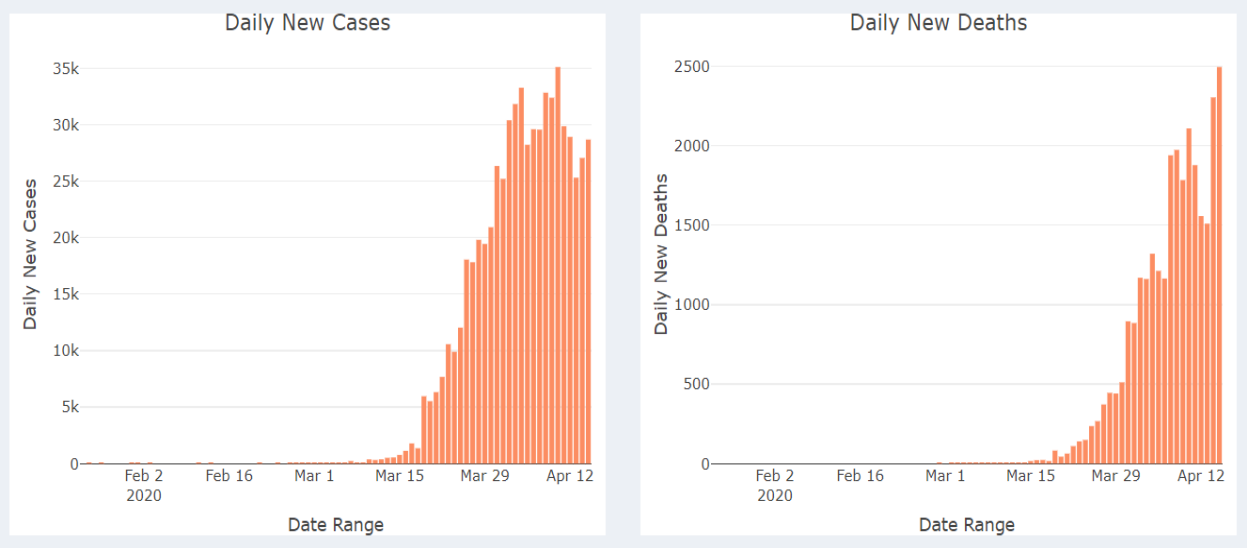

Fig 2.1 Visualizing the growth factor of COVID 19

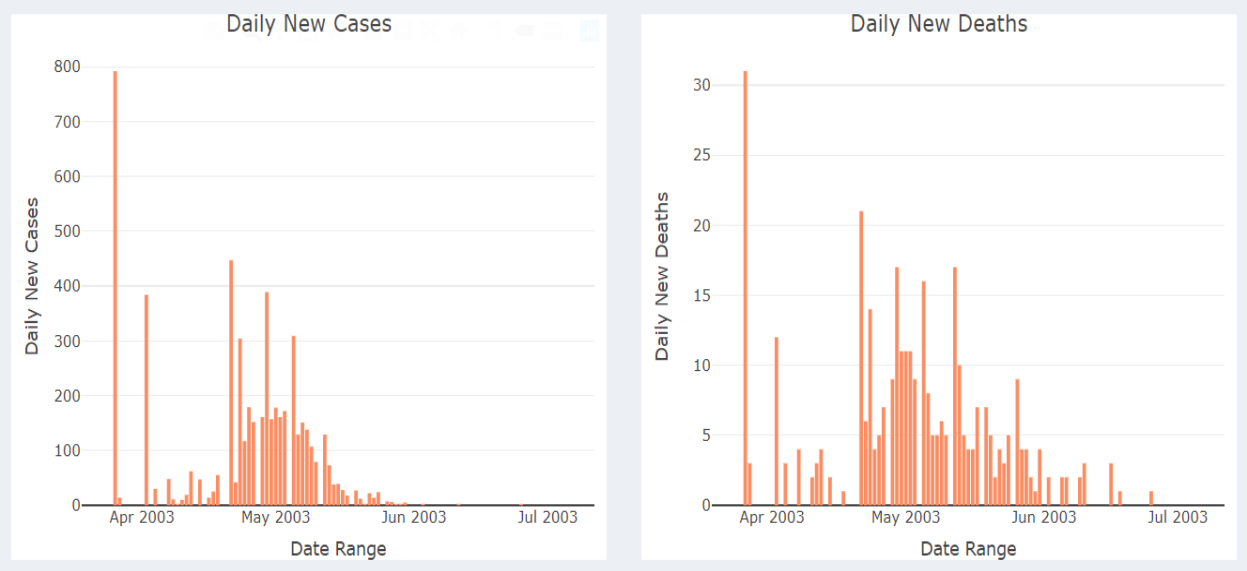

Figure 2.2 Visualizing the growth factor of SARS

The two attributes which could determine the growth factor are Daily New cases and Daily New Deaths occurred over time. The figures 2.1 and 2.2 shows the growth factor on the spread of viruses by tracking the number of cases and death recorded each day.

4.4.3. - Plot 3: Visualizing the part to whole relationship within the data:

The next objective of our analysis is to visualize the part to whole relationship within our data.

The data consisting of Confirmed Cases, Death and Recovered are the three attributes of our dataset considered in order to establish the part to whole relationship. Using feature engineering technique done in the data preprocessing stage, feature creation is implemented by creating new attribute from the above-mentioned attributes.

The new feature created is Active Cases, which is nothing but the people who are currently being affected by the viruses and are in treatment. This has been achieved by finding the difference between confirmed cases and the summation of death and recovered cases.

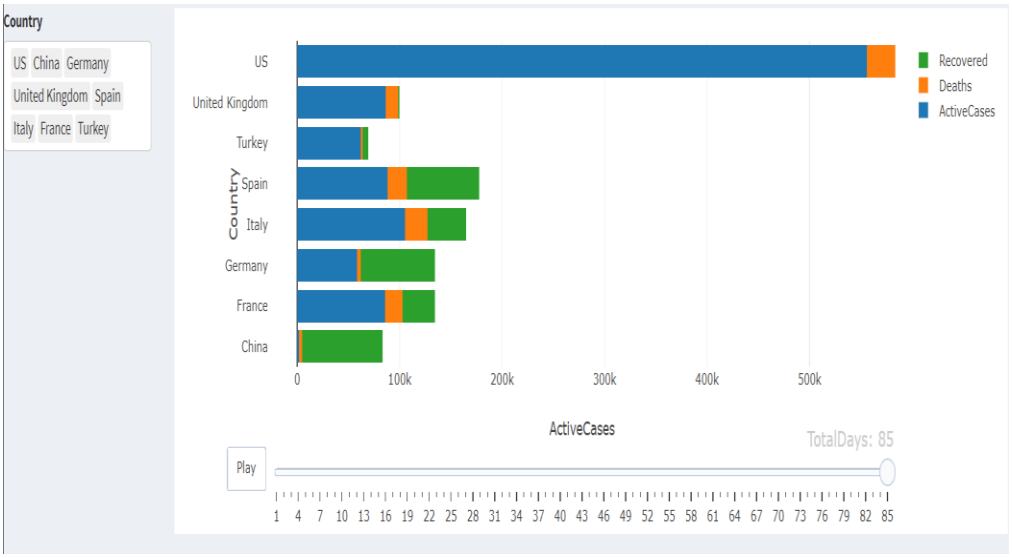

Figure 3.1 Stacked Bar Chart of COVID 19

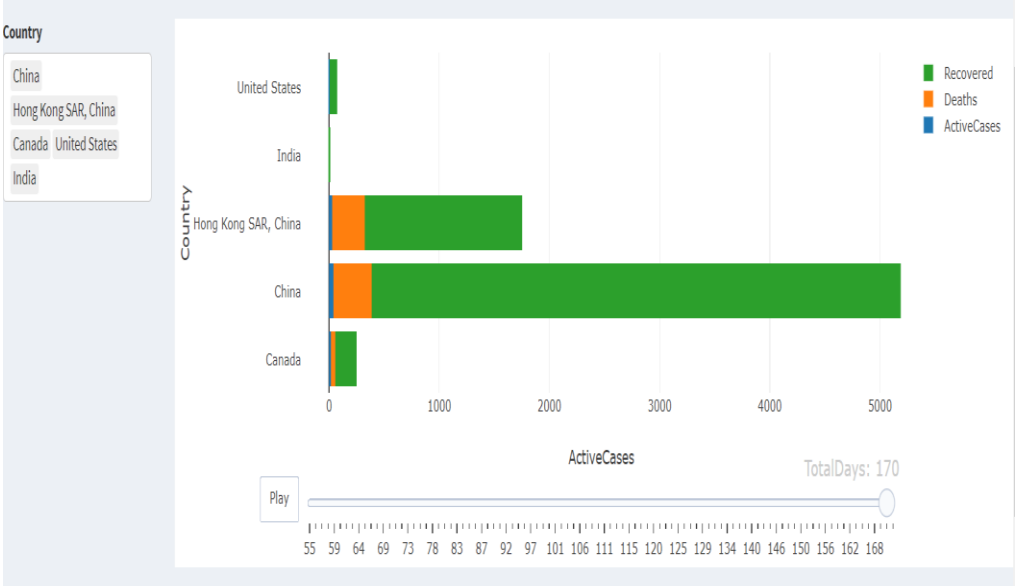

Figure 3.2 Stacked Bar Chart of SARS 2003

The above figures 3.1 and 3.2 visualizes the relative decomposition of each primary bar based on the levels of categorical variable. Here each bar comprised of number of sub bars which determines the number of Deaths, number of Recovered and number of active cases in relative with the total number of Confirmed cases over time.

4.4.4. - Plot 4: Spread of Covid and SARS:

The seriousness of the disease is understood by the acceleration of disease. The acceleration is analyzed by the rate of change in the number of cases and deaths. The Speed is effectively analyzed with an animation scatter chart. The seriousness of the disease is inferred by the drift of the bubble. The correlation of deaths and cases also can be visualized. The main approach of the visualization is to have emotional connect with the audience who travelled in similar time frame.

In this visualization, by data pre-processing a new feature called Total days is created where start date is initialized to 0 and incrementing it with passing days. The cases and deaths are mapped to x and y respectively, color is mapped to country and total days is mapped to frame.

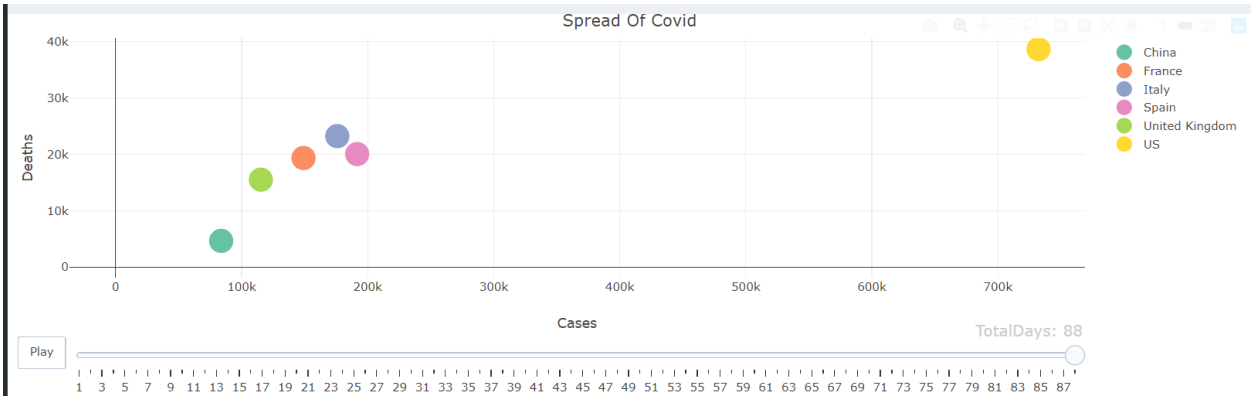

Figure 4.1 Animation Chart of COVID 19

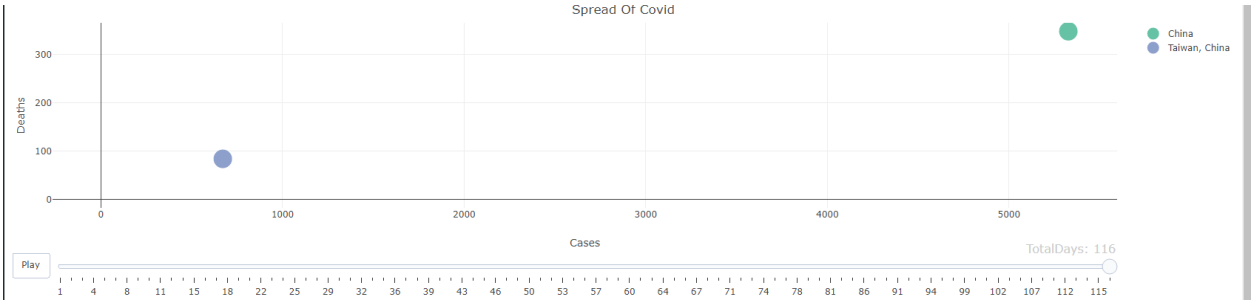

Figure 4.2 Animation Chart of SARS

The above figure 4.1 and 4.2 analyzes the acceleration of disease by observing drift of the bubble. Here each bubble is represented by different country. Drifting of bubble relative to time is studied.

4.4.5. - Plot 5: Fatal trend of Covid and SARS:

To study the fatality of diseases it is important to know the percentage of deaths to the total number of cases. The fatal cases can be predicted by visualizing the trend of fatal percent over time. The Data is prepared by grouping the values by date and country and taking cumulative sum of deaths and cases by date and country.

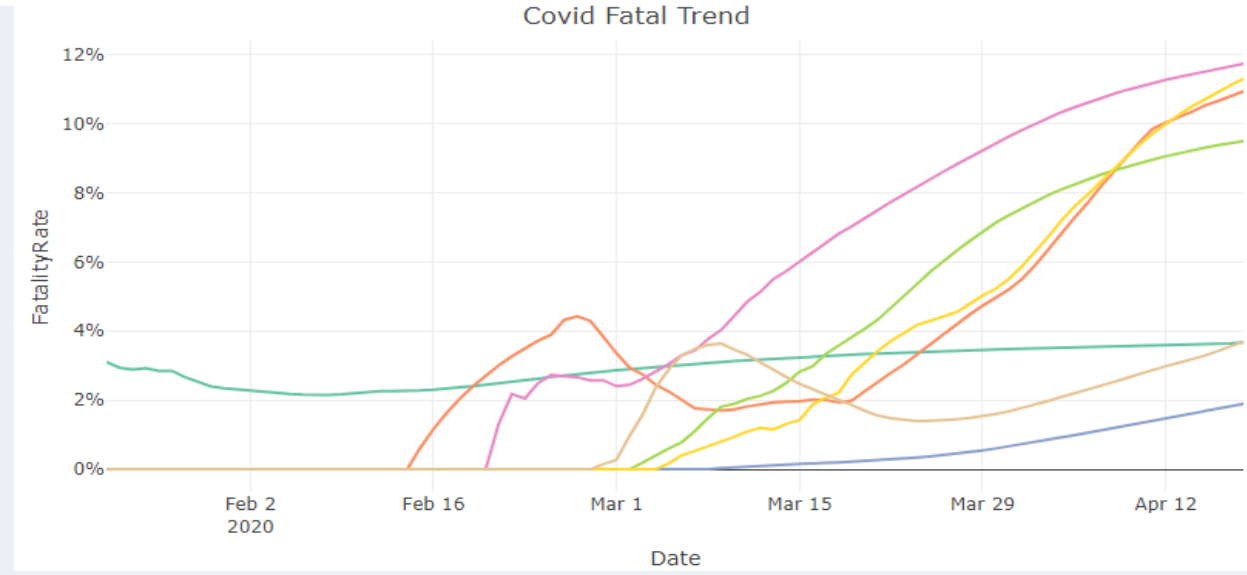

Figure 5.1: Fatal trend of Covid

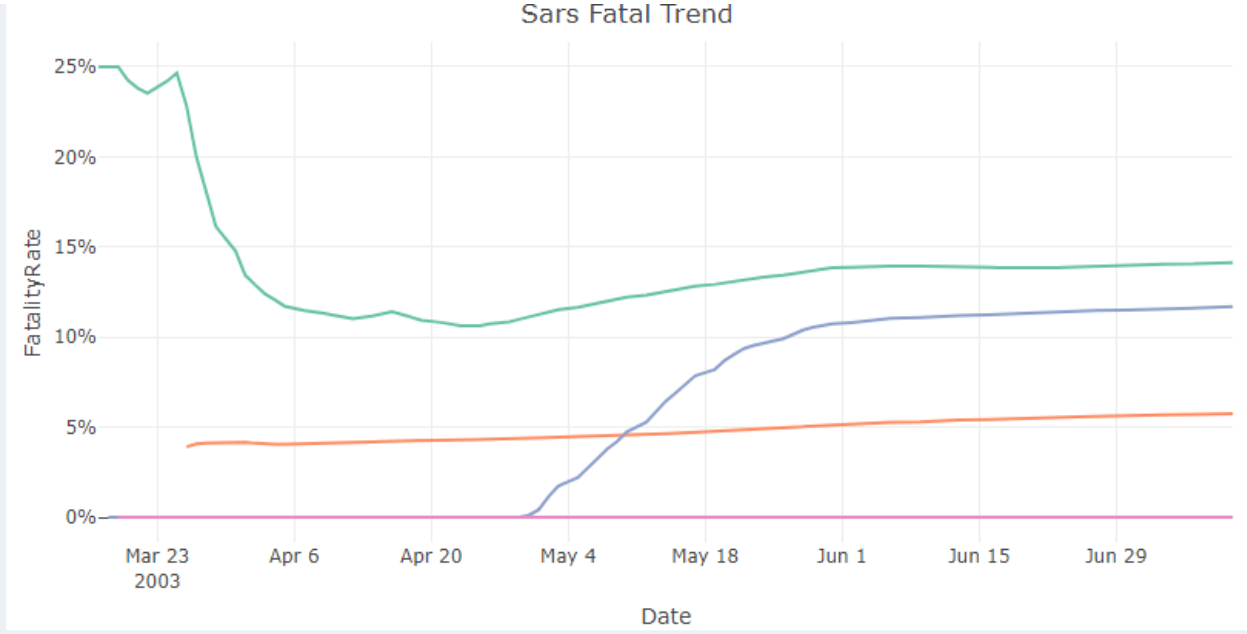

Figure 5.2: Fatal trend of SARS

In above figure 5.1 and 5.2 the fatal trend for different countries over time is analyzed.

4.4.6. - Plot 6: Recovery trend of Covid and SARS:

To study the recovery trend of diseases it is important to know the percentage of recovered cases to the total number of cases. The recovered cases can be predicted by visualizing the trend of recovery percent over time. The Data is prepared by grouping the values by date and country and taking cumulative sum of recovered and cases by date and country.

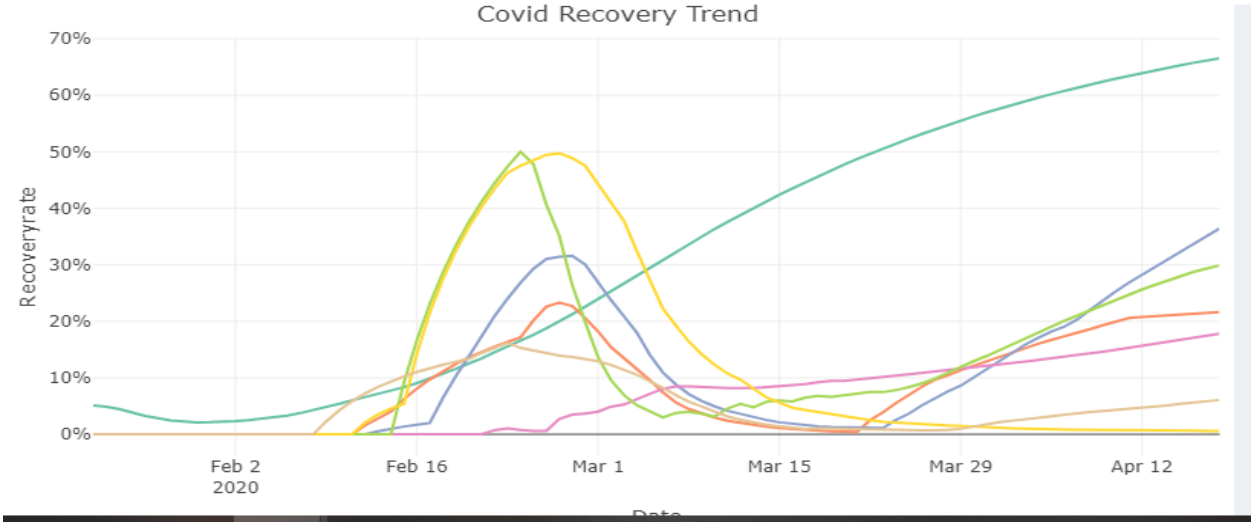

Figure 6.1: Recovery trend of Covid

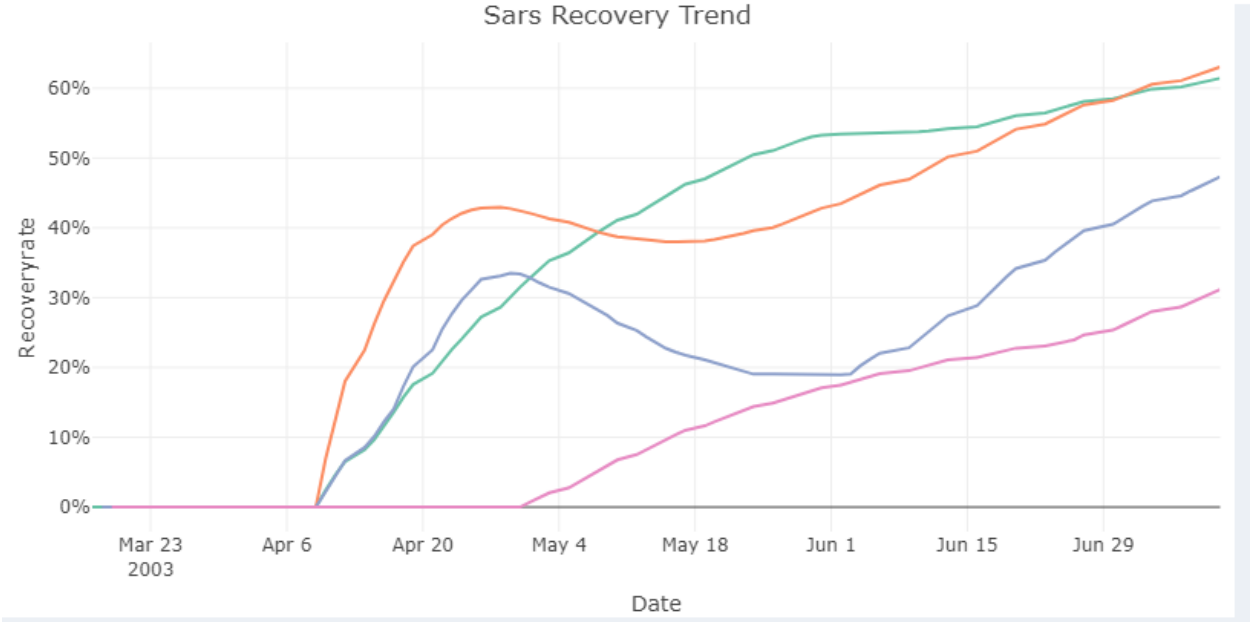

Figure 6.2: Recovery trend of SARS

4.4.7 – Plot 7: Visualizing the total cases using Timeseries Line Chart

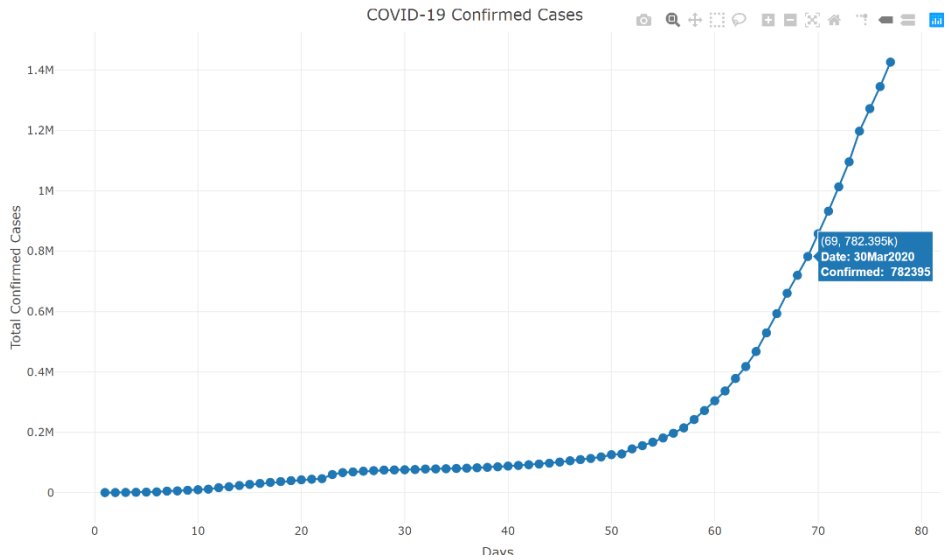

Figure 7.1 Line chart of COVID-19 total cases

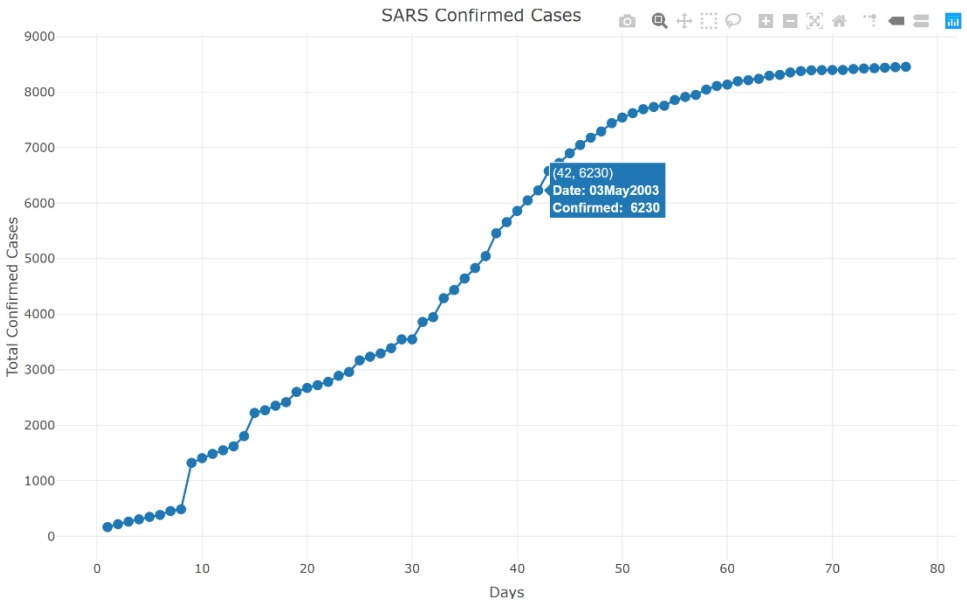

Figure 7.2 Line chart of SARS total cases

A time series plot is a graph where some measure of time is the unit on the x-axis. In fact, we label the x-axis the time-axis. The y-axis is for the variable that is being measured. Data points are plotted and generally connected with straight lines, which allows for the analysis of the graph generated. In the figures 4.1 and 4.2 total cases of COVID-19 and SARS were measured on Y-axis and the time stamp is converted into number of days and represented on X-axis.

4.4.8 – Plot 8 : Recovered and death cases of COVID-19 and SARS

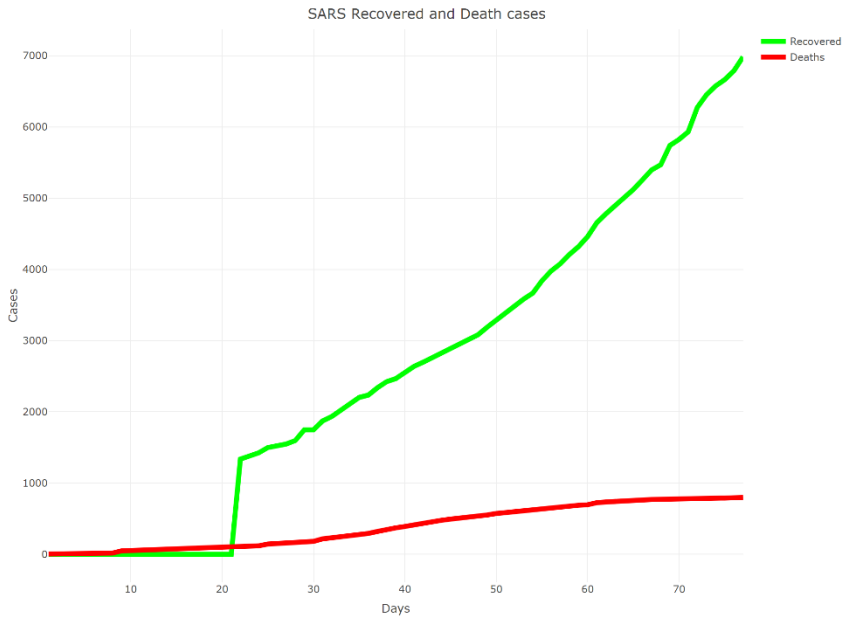

Figure 8.1 Recovered and Death cases of SARS

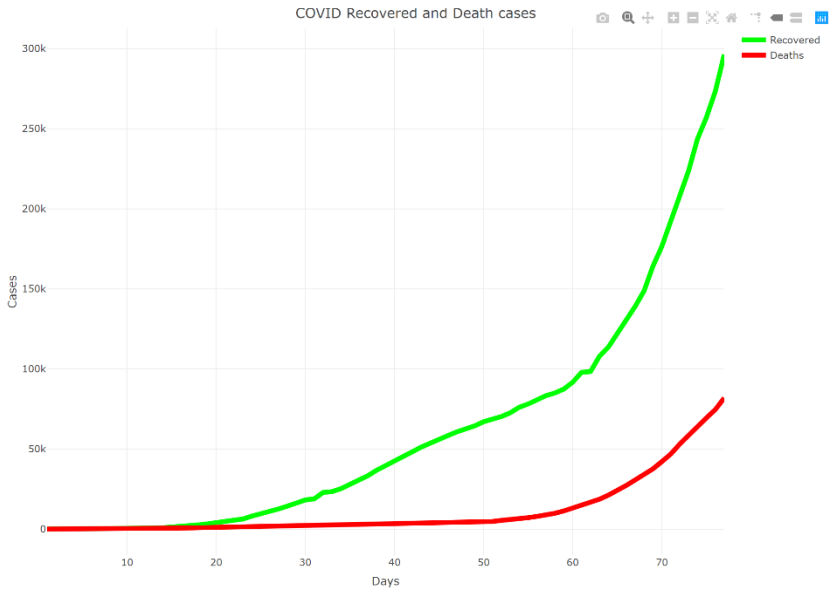

Figure 8.2 Recovered and Death Cases of COVID-19

The Above figures shows how Recovered and death cases are varying as the number of days increases. we can notice that there is a linear trend in Recovered cases of SARS after 20 days of its outbreak and exponential trend in recovered and death cases of COVID-19. We can also infer that the proportion of recovered cases are more compared to deaths in both the pandemics.

4.4.9 – Plot 9 : The Trend of COVID-19 and SARS

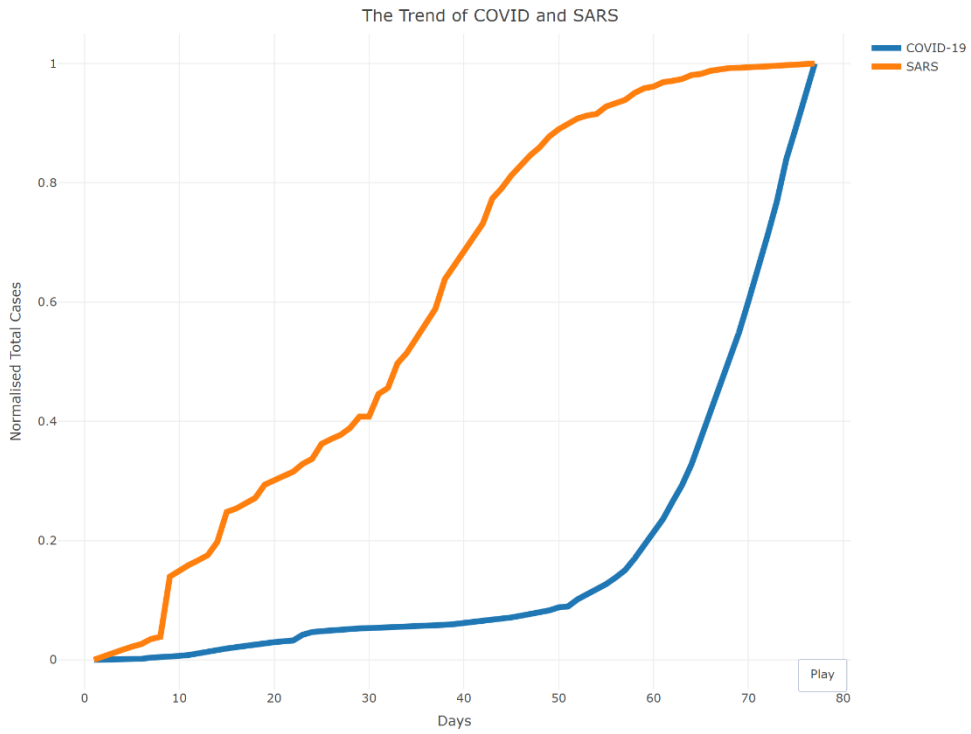

Figure 9.1 The Trend of Pandemics

Another objective of using line chart is to identify and display the trend of both COVID and SARS in comparison with each other. But the range of total cases of SARS and COVID are on different scale. So, we have normalized the total cases and plotted it against days. We can observe that the SARS cases increased Logarithmically and COVID cases increased Exponentially.

4.5. Dashboard:

To Make the above plots reachable to much wider audiences it is important to group the above plots in a graphical user interface called dashboard. The dashboard gives a glance overview of different key metrics used in the comparison of COVID and Sars. The dashboard has been made quite interactive with radio boxes to switch between different diseases and text boxes to switch between different countries additionally moving the pointer on the graph gives required Information.

4.5 Interactive Dashboard

4.6. Tools and Frameworks:

The following are the tools and frameworks used to build the entire project .

- R studio
- Shiny
- Plotly
- Excel

4.7 Project Management and Efforts:

Sl.no	TASK	Members	Duration
1.	Data Collection	Aravind, Vishnu and vipul	3days
2.	Data PreProcessing	Aravind and vishnu	4days
3.	Explolatory data analysis	Aravind and vipul	2 days
4.	Chart selection	Aravind, Vishnu and Vipul	1day
5.	plot1, plot2 and plot3	Aravind	3days
6.	plot4, plot5 and plot6	Vipul	3days
7.	plot7, plot8 and plot9	Vishnu	3days
8.	Code integration and Dashboard Creation	Aravind, Vipul, Vishnu	3 days

4.8. Conclusion :

From the interactive plots and animations, the comparison of 20th century pandemics and interesting insights were outlined. The choropleth map displays the intensity of the spread, the bar charts displays the count of cases in each country and line chart displays the trend of cases. Since both the outbreaks took place in different decades, conversion of date into to number of days played a key role in comparison.

In the future, the dashboard can be updated with live data of covid-19 and compelling insights can be visualized. This helps in taking appropriate measures to stop and contain the spread of virus in highly affected countries.

4.9. References :

John Hopkins University Corona Virus Resource Center

https://coronavirus.jhu.edu/

SARS 2003 complete data

https://www.kaggle.com/imdevskp/sars-outbreak-2003-complete-dataset