Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Module:CS7DS4 Data Visualization 2019-20

Student Name : Snehal Dey

Student No: 19303333

Declaration:

"I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar.

I have also completed the Online Tutorial on avoiding plagiarism ‘Ready Steady Write’, located at http://tcd-ie.libguides.com/plagiarism/ready-steady-write ."

1. Introduction

The project aims to visualize the quantitative data and information of the spread of Covid-19 in India from the time when the first case was reported i.e on 30th of january 2020 till 22th of April 2020 into a graphical representation. This visualization helps us to perceive the data in a comprehensive manner and helps us in getting more insightful information from it.

2. Choice of Visualization tools.

There are many software tools and programming languages like Processing, D3 available that helps us to visualize the data. One such tool is Tableau which has become one of the most used tool in the industry for data visualization purposes.
Even though initially I started the project using python by using the plotly library, I have found the limitation that Plotly does not have the map for India till now and hence it cannot be used for visualizing the spread of coronavirus over the indian map. Hence I have used the most sought after tool in the industry for the visualization purpose for this project.

3. Data Source

The data used for this visualization is obtained from the internet. The links to each of these resources will be provided in the reference section.

1. I have used a dataset from Kaggle called Covid-19 in India to get the dataset of all the corona cases in India till 16th of April.
2. To get the data from 17th of April to 22nd of April I have used the daily updated data from the site of the Ministry of Health and Family Welfare.

4. Description

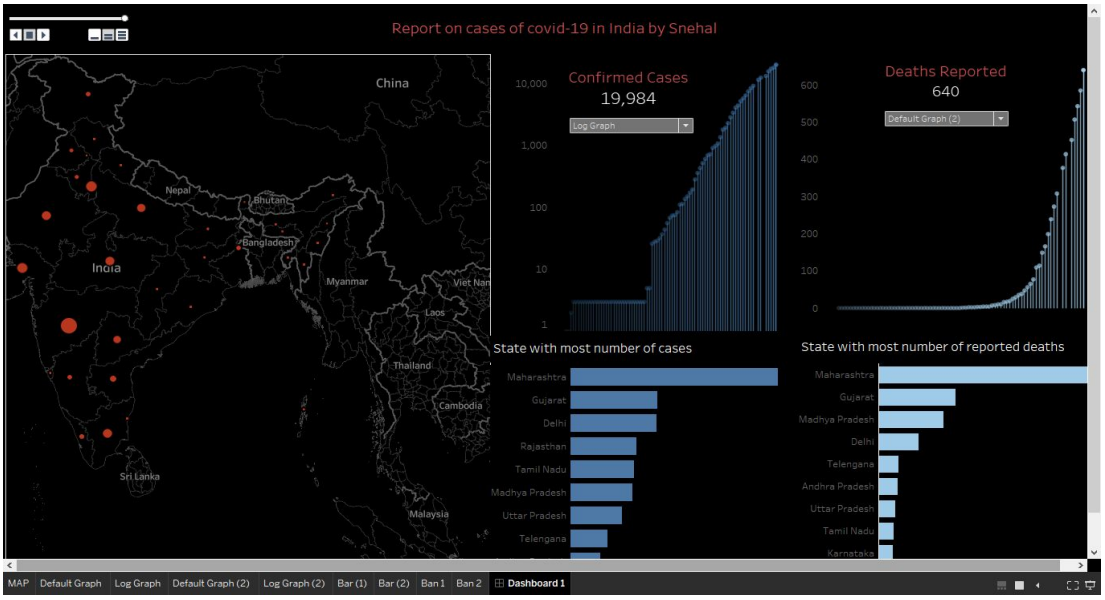
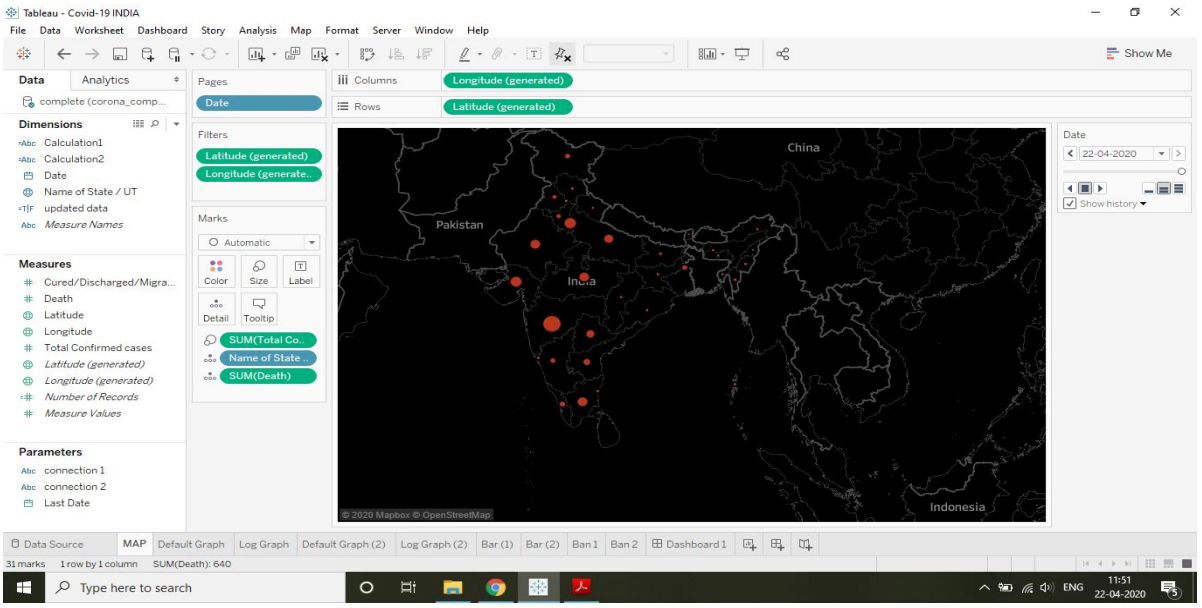


Figure: Dashboard on Report of COVID-19 in India from 30/01/2020 till 22/04/2020

The dashboard provided above is a combined representation of nine different worksheets representing the same data but with different perspectives. **The visual encoding channels used in this are Position [Demographic], Size, Colour, Orientation and Shape.**

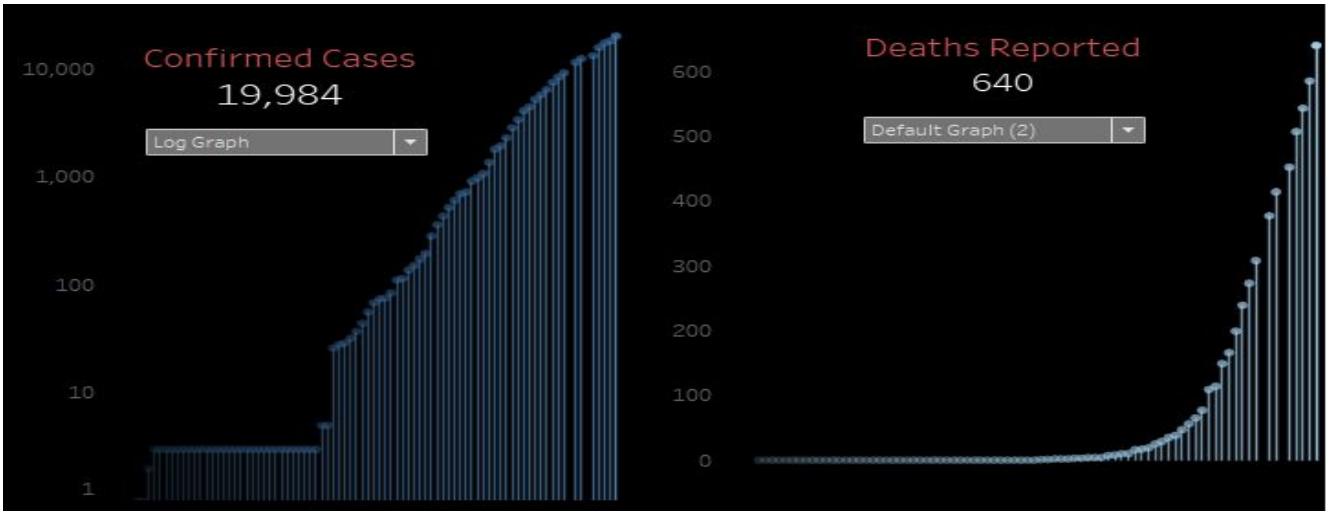
Map Chart

The worksheet on the right of the dashboard contains the Map object representing the political map of India. There is a slider present on the top of the map that shows how the virus has spread across different states of India based on the timeline. Also, the user can click or hover on the map to get the details of the total number of cases reported as well as deaths caused due to coronavirus as tooltips.



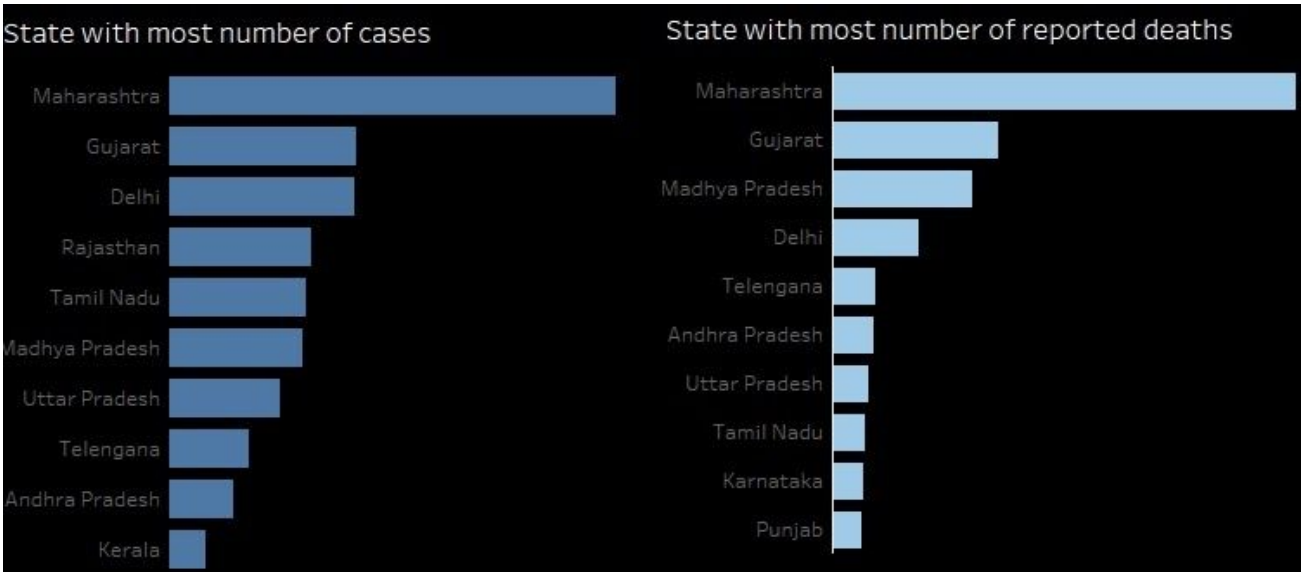
Line Graph

On the top right corner of the dashboard we find the line graphs which shows the rise in the number of cases of Covid-19 as well as the number of deaths reported due to the virus. The graphs are represented in both normal and log scale and one may switch between the normal view and log view by just selecting the view from the drop down menu. The dashboard also counts the total number of cases and deaths till 22nd of April and displays them above the respective graphs. All these actions are controlled by the slider present on the top left corner of the Dashboard. On hovering over the graph the number of Covid-19 cases reported and the date is represented as tooltip and for the graph which shows the rise in death the number of deaths reported is shown as a tooltip with the date.



Bar Graphs

The bar graphs are present on the down right corner of the dashboard. The bar graph shows the top 10 states that are most affected by Covid-19. The first graph represents the number of cases and the second the number of deaths. The results are sorted based on the most number of cases. On hovering over the bar chart we get to see the number of cases and deaths for that state.



5. Conclusions

In this visualization one can easily estimate values and trends of the increase in Covid cases throughout India from 30th of January 2020 till 22nd of April 2020 . The dashboard is interactive, informative and visually distinct with different options for analysing the spread of CoronaVirus.
From the visualizations we can infer that Maharashtra is the most affected state in India having the largest number of cases and deaths. Even though the first case was reported in Kerala, they were able to contain the disease by not letting it spread. Maharashtra being one of the most developed states in India had a lot of foreigner tourists as well as indian tourists coming influx and that might be a reason for the spread in this extent along with its neighbor Gujrat having second highest state. But the not so developed north east part of India has the least number of cases where states like Sikkim have not reported a single case and the cases in Nagaland and Mizoram being 1.
Doing this assignment I have also learnt how to use Tableau for the purpose of data Visualization.

References

[1] https://www.kaggle.com/sudalairajkumar/covid19-in-india

[2] https://www.mohfw.gov.in/#state-data