Name: Akshay Hareshbhai Sitlani
Program: MSDS - Statistics
Subject: Data Wrangling and Husbandry, Spring 2020
Instructor: Jason M. Klusowski

Novel Coronavirus in India – A Brief Analysis

Objective Statement:
A brief analysis of the current situation of Coronavirus (COVID-19) in India.

Source of Data:
The datasets have been sourced from www.kaggle.com but the actual data has been obtained from:
The Ministry of Health and Family Welfare on India website – www.mohfw.nic.in and
The official COVID-19 India Tracker Website – www.covid19india.org

About the Data:
This project uses five datasets. They are described briefly as follows:

1. **covid_19_india.csv:** This dataset provides day-wise information of confirmed, recovered and deceased cases along with nationality (Indian, Italian, etc.). The information available ranges from January 30, 2020 to April 28, 2020.
2. **novel_covid_19_data.csv:** This dataset provides day-wise information of confirmed, recovered and deceased cases of all the countries in the world (along with province/state). The information available ranges from January 22, 2020 to May 05, 2020.
3. **ICMRTestingDetails.csv:** This dataset provide day-wise details of total samples tested, total individuals tested and total positive tests obtained between March 13, 2020 and April 27, 2020.
4. **ICMRTestingLabs.csv:** This dataset gives information about types of testing laboratories across various cities of India.
5. **HospitalBedsIndia.csv:** This dataset tells what type of and how many medical facilities are available in different states of India. It also gives information about the beds (and their types) available for active patients in various states of India.

A common problem with all these datasets is the **existence of missing values**. In this project, these datasets were manipulated appropriately to conduct the necessary analysis. Comments have been provided in the code wherever the missing values have been dealt with. Summary statistics have been generated in the .rmd files for reference.

Analyses and Visualizations:

1. **Number of Samples and Individuals between March 13 and April 28**

   This graph conveys the following insight:

   A. The increasing number of tests conducted in the latter half of the month period shows the proactiveness of the Government of India in arranging required medical facilities in short notice.

   B. Both graphs are almost identical, with minor change in the axis values. It can be noted in that period post 8th April, the gap between samples and individuals tested has increased, showing higher chances of an individual getting tested positive in first trial.

Figure 1 Number of Samples / Individuals Tested

2. **Number of Positive Test Results and Trend in Rise/Drop of Percentage Positive Cases between March 13 and April 28**

We see that:

A. More or less, we get a similar trend as the previous graph. So, the number of positive cases has been increasing almost exponentially.

B. In general, the percentage of positive cases has also increased over time.

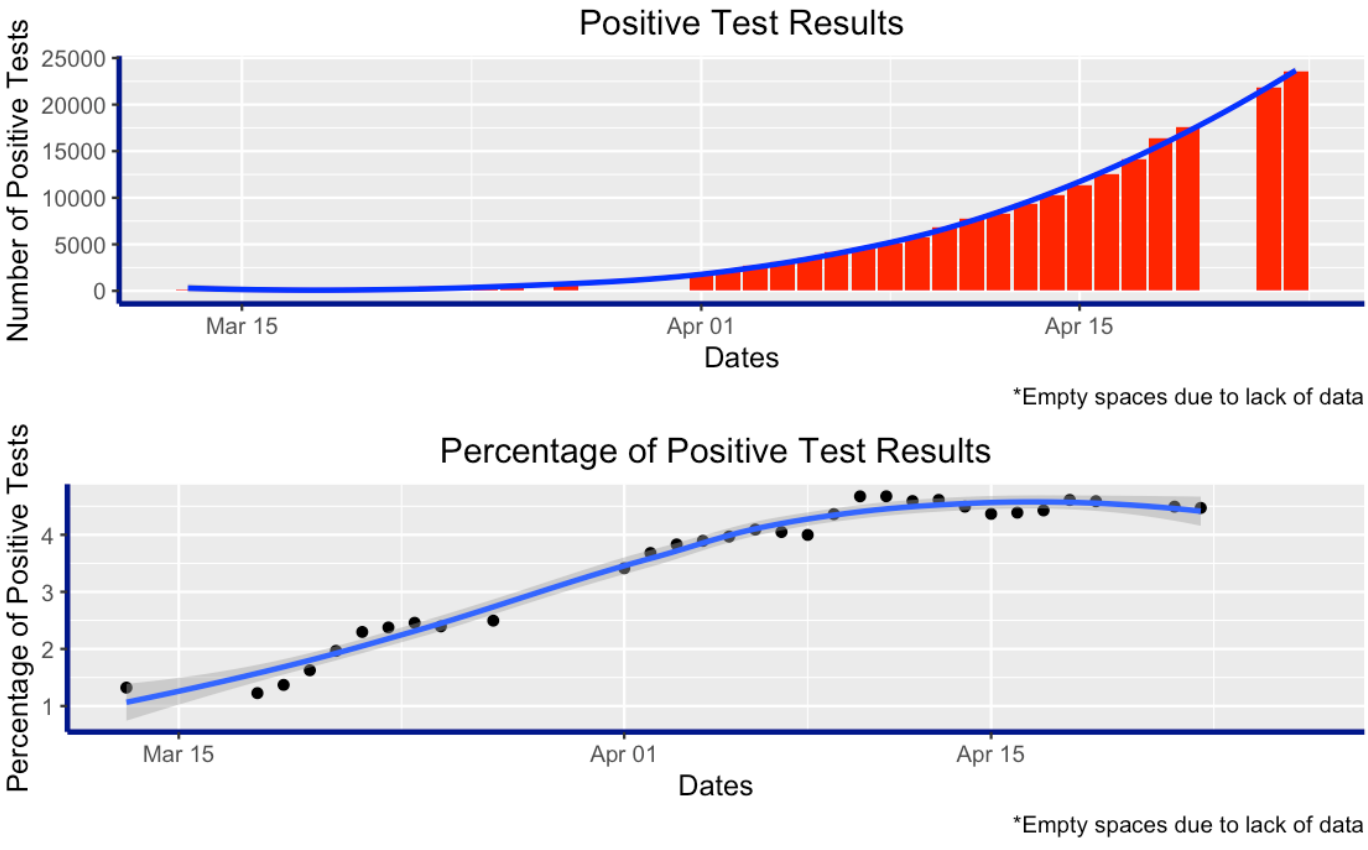

Figure 2 Number and Percentage of Positive Test Results

3. Comparison of negative and positive cases each day between March 13 and April 28

We can observe a positive correlation between total individuals tested and total positive cases.

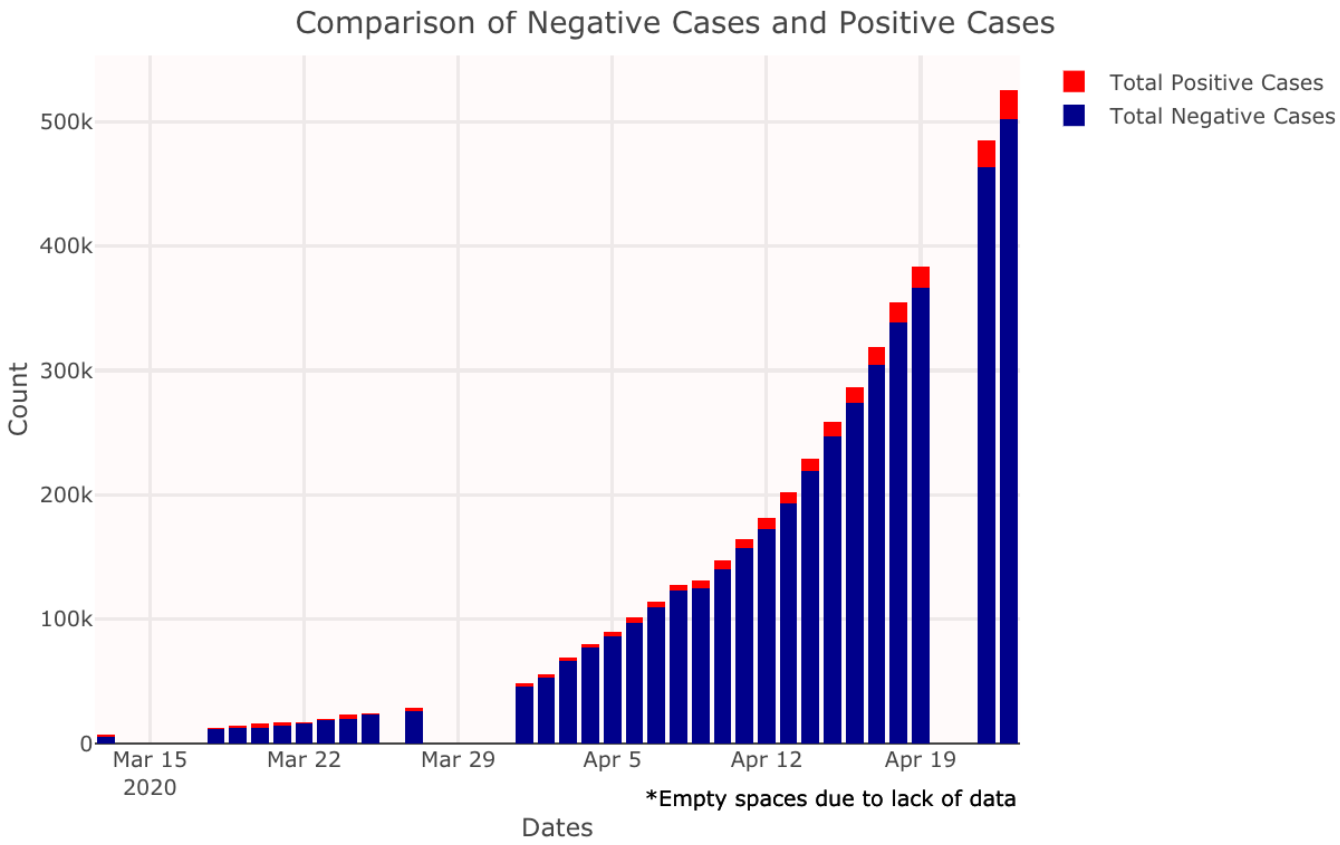

Figure 3 Comparison of Positive and Negative Cases

4. Types and Counts of Testing Centers in India

Important Facts about the above treemap:

A. 3 states with no labs yet have been omitted from the plot. These states are: Daman and Diu, Lakshadweep and Nagaland.

B. Some abbreviations have been used in the plot due to long state names and limited space on the plot. These abbreviations are: A and N (Andaman and Nicobar), Arunachal (Arunachal Pradesh) and D and NH (Dadra and Nagar Haveli).

Insights:

A. There are many government labs that are testing people for COVID-19. However, the private sector has not been given that discretion by the government. For fast detection and containment of this virus, private sector must be given equal if not more responsibility.

B. The states with the lighter color indicate that they have higher number of labs compared to the darker colored states.
This may have two aspects:
   a. The positive part of this is that these states are forerunners is detecting and curing COIVD-19 affected people. States which are on the darker side, need special attention from the government. If more testing labs are not made available to them, it might lead to a bigger problem.

b. The negative part of this is that the lighter colored states probably can have higher number of COVID-19 cases, due to which the government has set up more testing labs there in the first place. So, these states act like a hotspot and extra care need to be taken in such states so as to prevent further transmission esp. mass transmission.

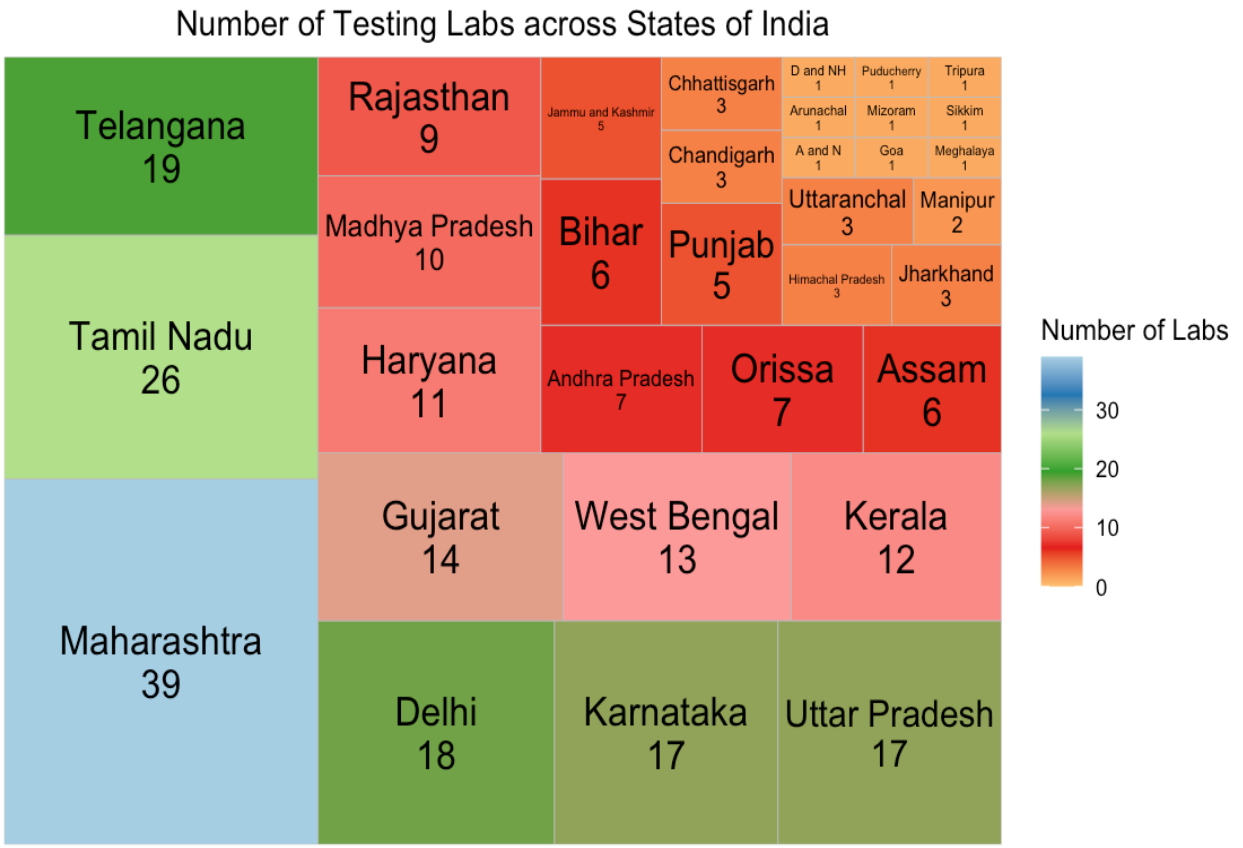

Figure 5 Number of Testing Labs across States of India

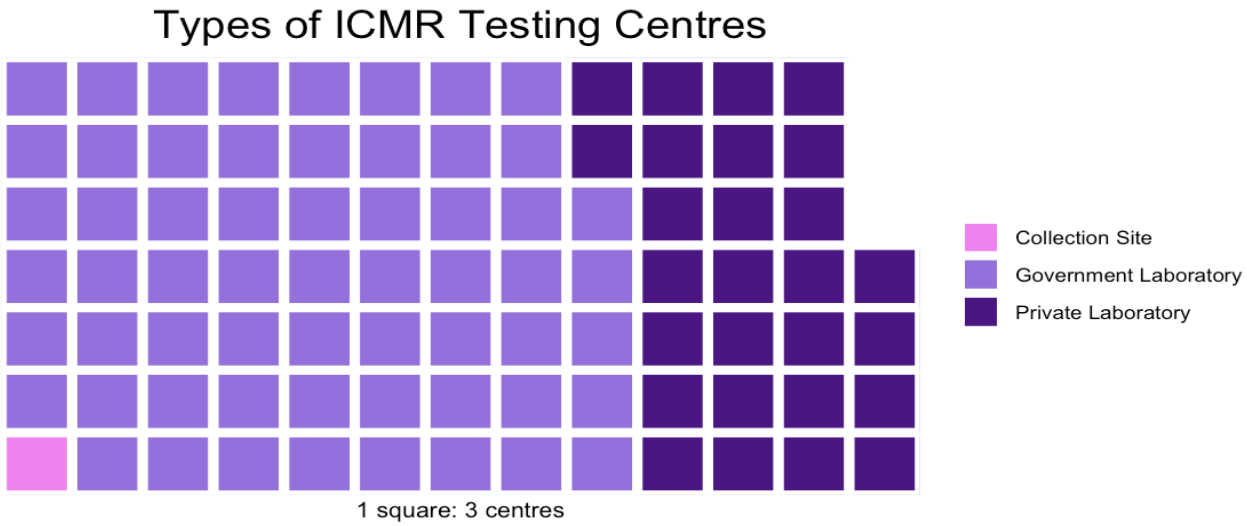

Figure 4 Types of ICMR Testing Centers

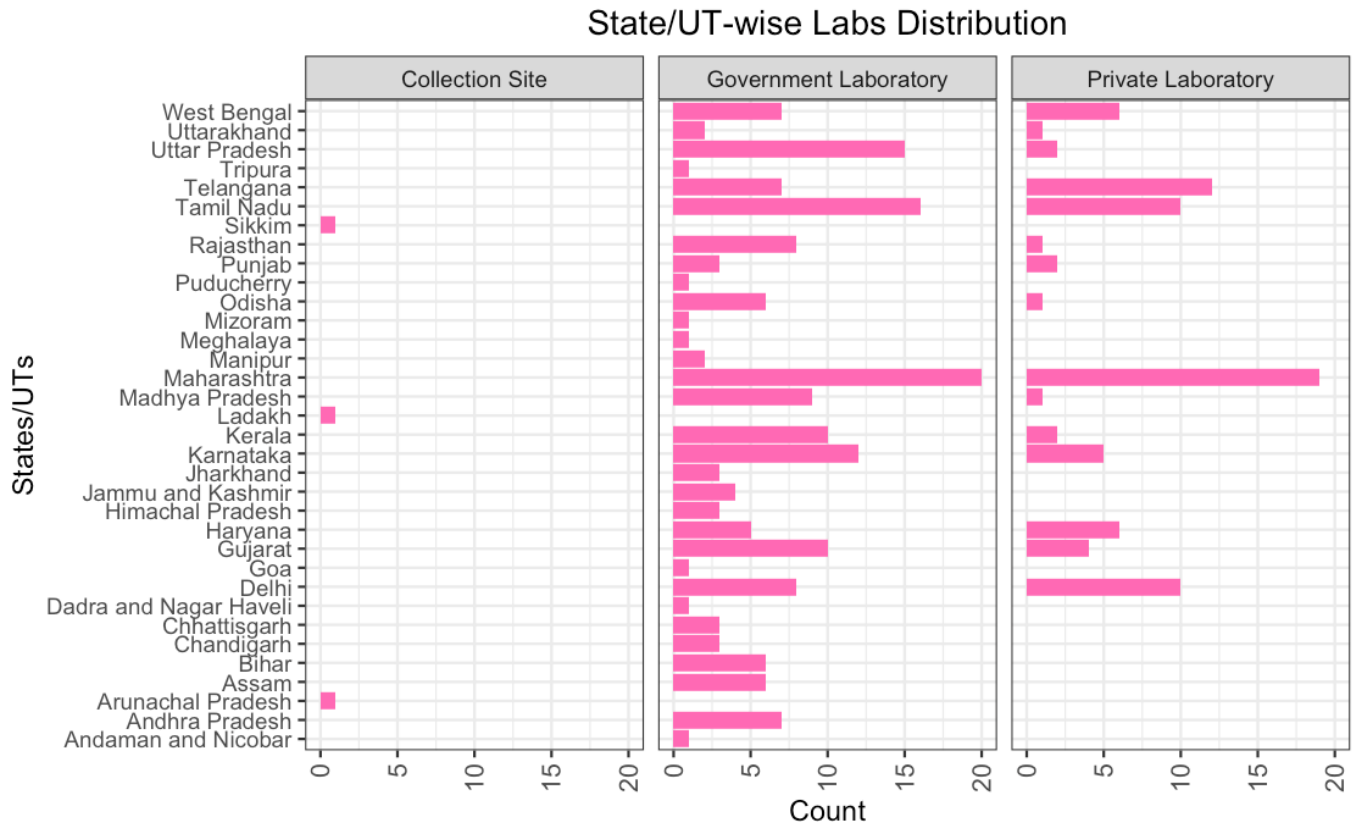

Figure 6 State/UT wise Labs Distribution

5. Type and Count of Health-Care Facilities across various states of India

There are some states and/or Union Territories like Andaman and Nicobar Islands, Bihar, etc. that critically lack in hospitals and other healthcare facilities. More focus should be given on increasing resources in such regions in order to contain the spread of COVID-19.

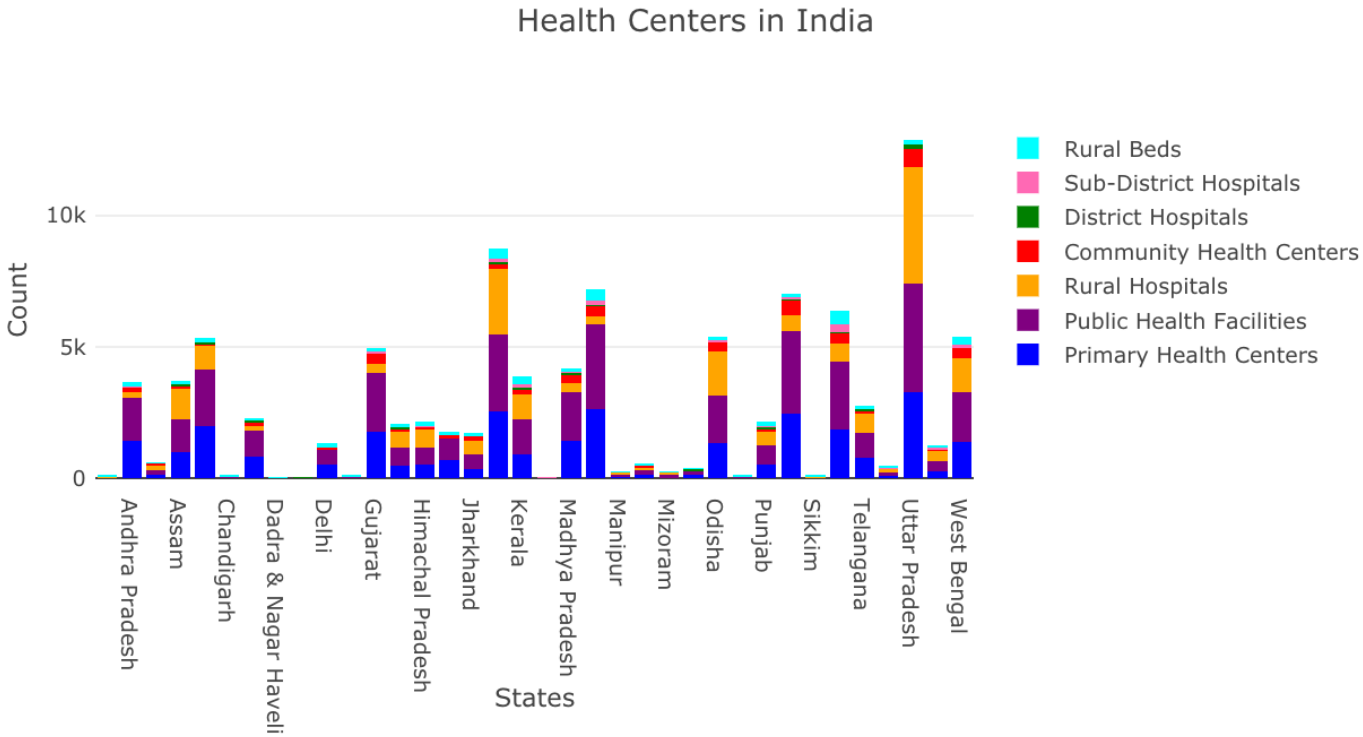

Figure 7 Health Care Facilities in India

6. Types and Count of Beds available in various states of India

There are some states and/or Union Territories like Andaman and Nicobar Islands, Bihar, etc. that critically lack in beds for active patients. More focus should be given on increasing resources in such regions in order to contain the spread of COVID-19.

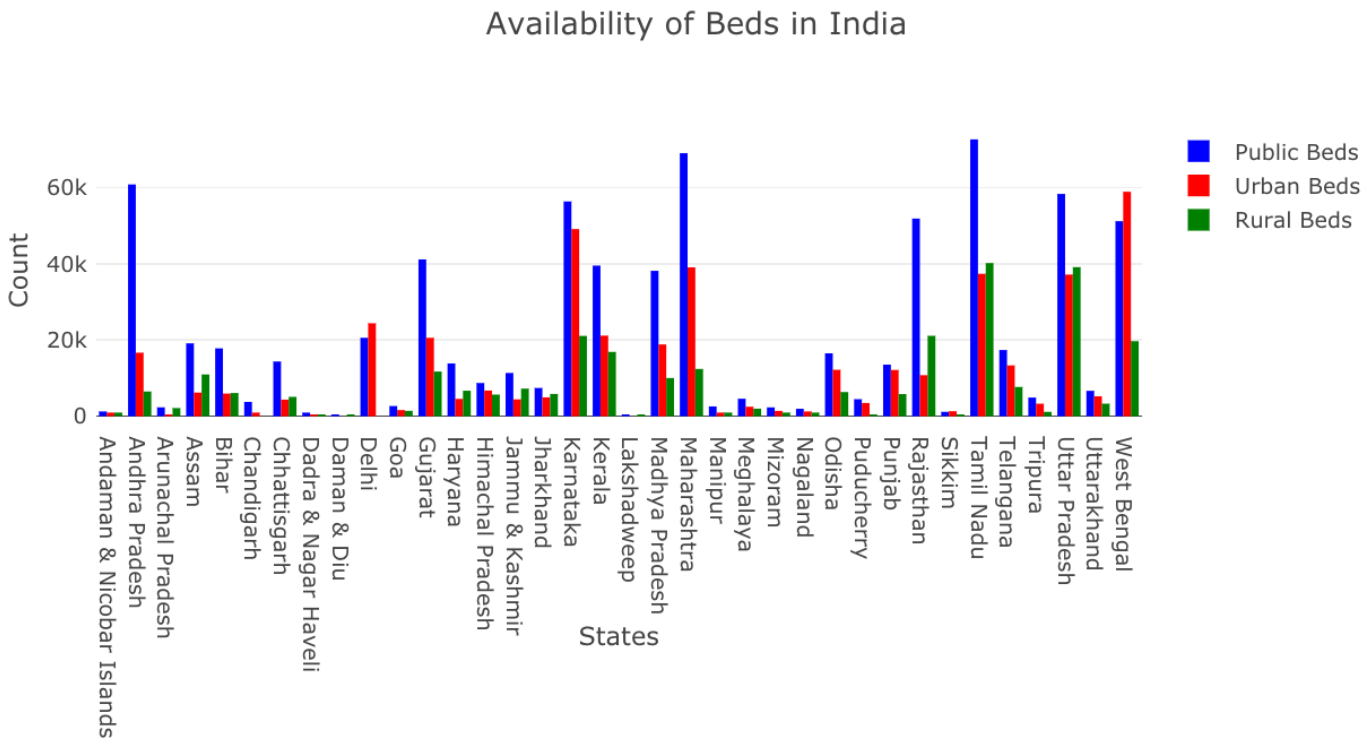

Figure 8 Availability of Beds in India

7. Daily Cases Trend for Top 9 States in India

We observe that **Maharashtra** has shown sharp rise in daily number of cases. So it can be classified as a **COVID-19 HOTSPOT**. **Delhi, Gujarat and Tamil Nadu** are next in line. So, steps must be taken to prevent these states to turn into a hotspot.

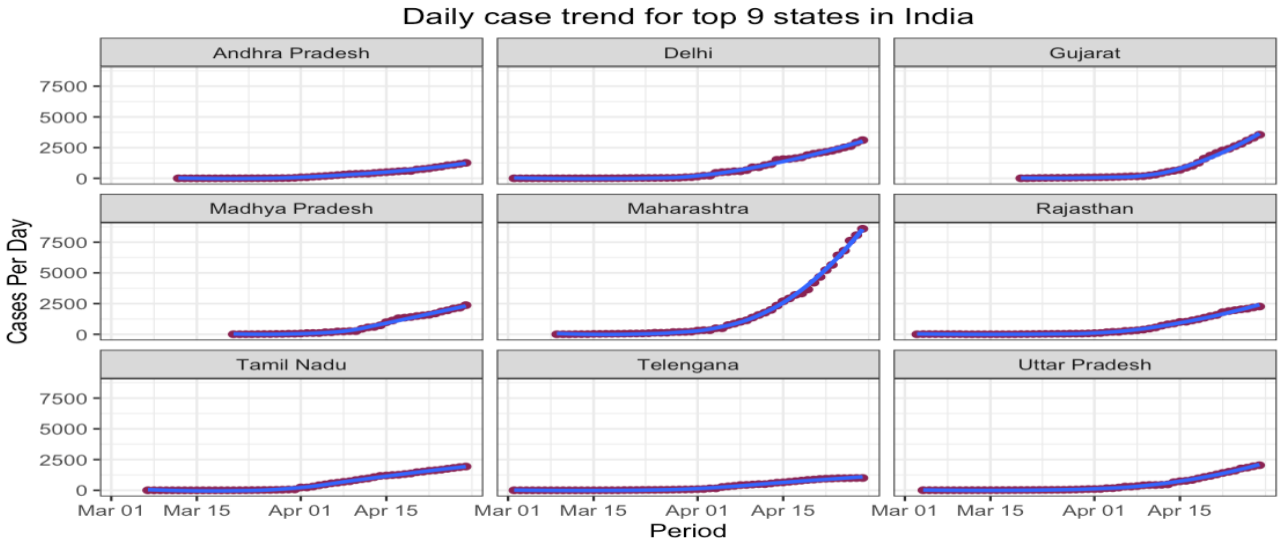

Figure 9 Trend of Daily Cases in Top 9 States

8. Number of Beds per Active Cases in India

We see that **Maharashtra** has very small ratio of beds to patients (17.36). And it has the highest number of cases in India. That means, if more beds are not made available soon, the **Maharashtra Govt.** may have to set up temporary hospitals like the one made by **Iran** and **China**.

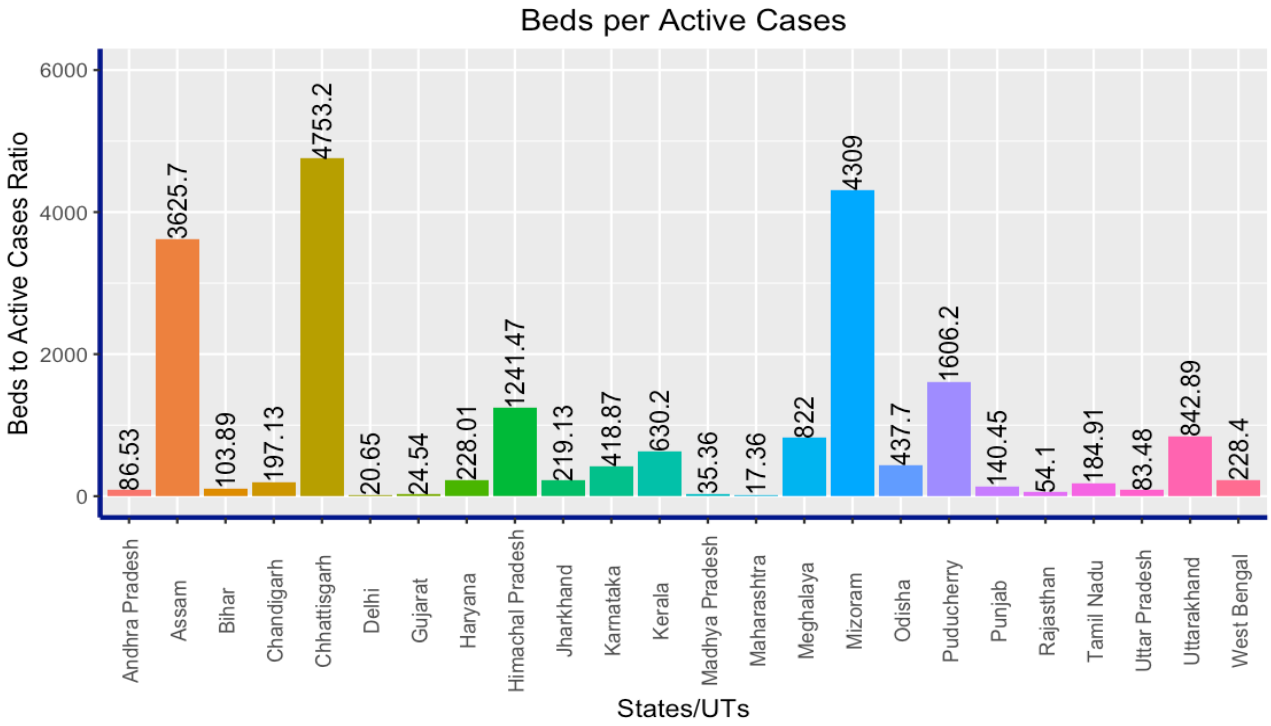

Figure 10 Beds per Active Cases in Different States

9. Animation of Rise in Confirmed Cases across Top 10 states between January 30 and April 28

As you can see in the animation, the first ever case of COVID-19 was found in Kerala. But, as time progresses, **Maharashtra** remains at the top for quite some time, whereas **Gujarat** rises to the second spot very quickly.

Kindly refer to the .gif file provided in the project resources. It is also visible in the .html file generated for the project.

10. Novel CoronaVirus - India Dashboard

This Dashboard presents the following:

A. Total COVID-19 cases in the World and India
B. Confirmed, Recovered and Deceased COVID-19 cases in the World and India
C. Trend of rise/fall of confirmed, recovered and deceased cases in the World and India
D. Top 9 most affected Indian states
E. India specific data table with filters for every category

I have made this using the "flexdashboard" package in R. Kindly refer to the .html file provided in the project resources to view the dashboard.

References:

General R programming related doubts: Official R documentation https://www.r-project.org/other-docs.html, https://community.rstudio.com/ and https://www.stackoverflow.com/

Dealing with Plotly graphs: Official Plotly website https://plotly.com/r/

CSS color codes and schemes: https://www.rapidtables.com/web/css/css-color.html

Ggplot issues: Official Ggplot website https://ggplot2.tidyverse.org/reference/

Information about types of graphs: https://www.r-graph-gallery.com/

Everything else: Random searches on www.google.com

THANK YOU!