

## Dataset

We used the dataset complete.csv for the group course work has 4692 observations and 10 features. The dataset is about the corona virus cases and deaths over different states of India over eight months from 30<sup>th</sup> of January to 6<sup>th</sup> of August. I am trying to look for the average number of New cases and deaths over different months and different states. As I went through the structure of the dataset, the death column is in character type, which is needed to be numeric. And the overall dataset has one null value in the Death column, which has been filled by 0.

## Questions

We will answer the following questions in this report.

Q1: The average number of new cases and deaths over each month?

Q2: Which top 10 states has the average number of new cases and deaths?

### Q1: The average number of new cases and deaths in each month?

To answer the question we will first convert the Date column into three sub columns (Day/Month/Year) to check the average numbers of New Cases and Death over in eight months.

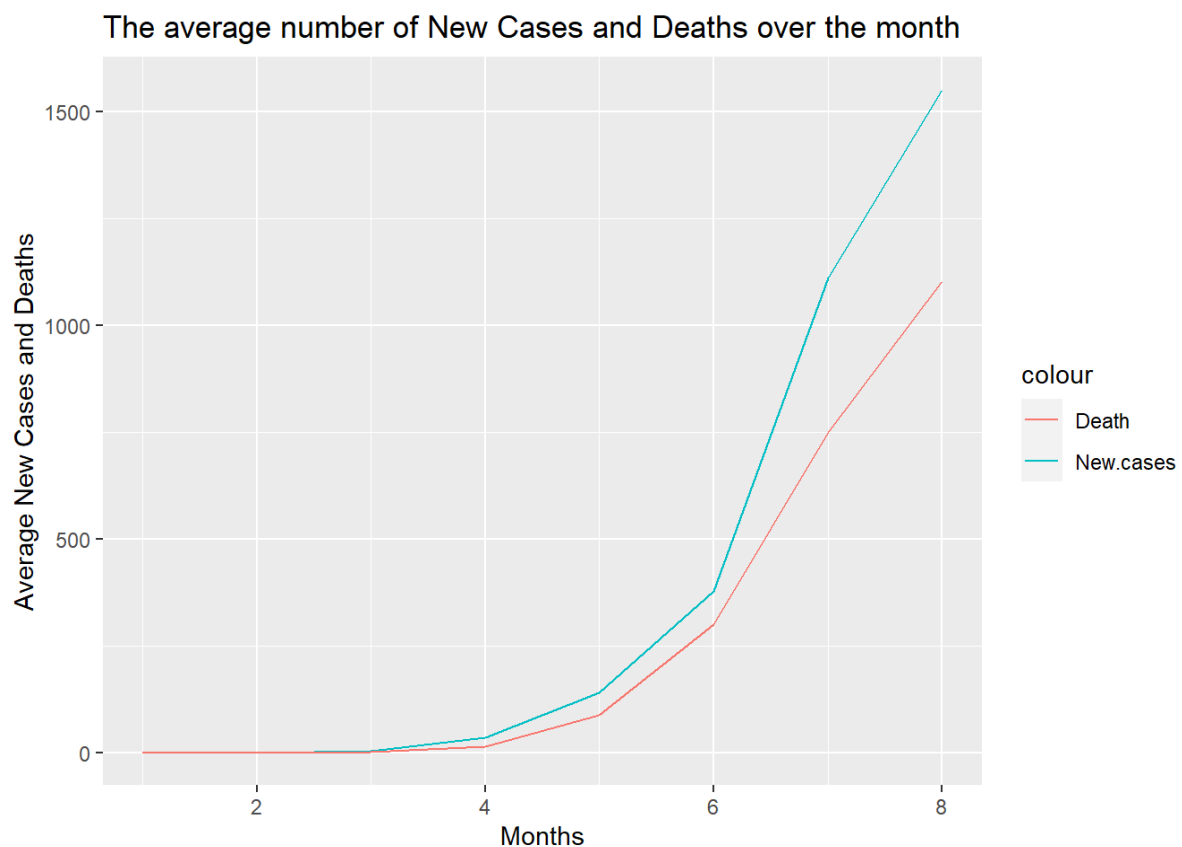


Figure 1

The Figure 1 show the average number of New cases and Deaths of each month. As we see from 30 January till April the new cases and death rate is considerable low. Majority of the cases and deaths happened after April. From April onward till June the new cases rose almost to 450 and deaths almost to 400. The average number of new cases and deaths constantly increases from June till August. Overall, we can say that the number of new cases are 3 times higher than previous months and deaths almost double.

Descriptive statistics for the New cases and deaths for Q1

month	New.cases	Death
1	0.0000000	0.0000000
2	0.0689655	0.0000000
3	2.6323232	0.4080808
4	33.8271078	13.9658485
5	139.1443798	87.0087209
6	377.3152709	299.9546798
7	1110.1172840	748.3755144
8	1550.7904762	1102.1047619

Table 1

## Q2: Which top 10 states has the average number of new cases and deaths?

To see the mean New Cases and deaths of top 10 state. We group by state and take the mean the observations.

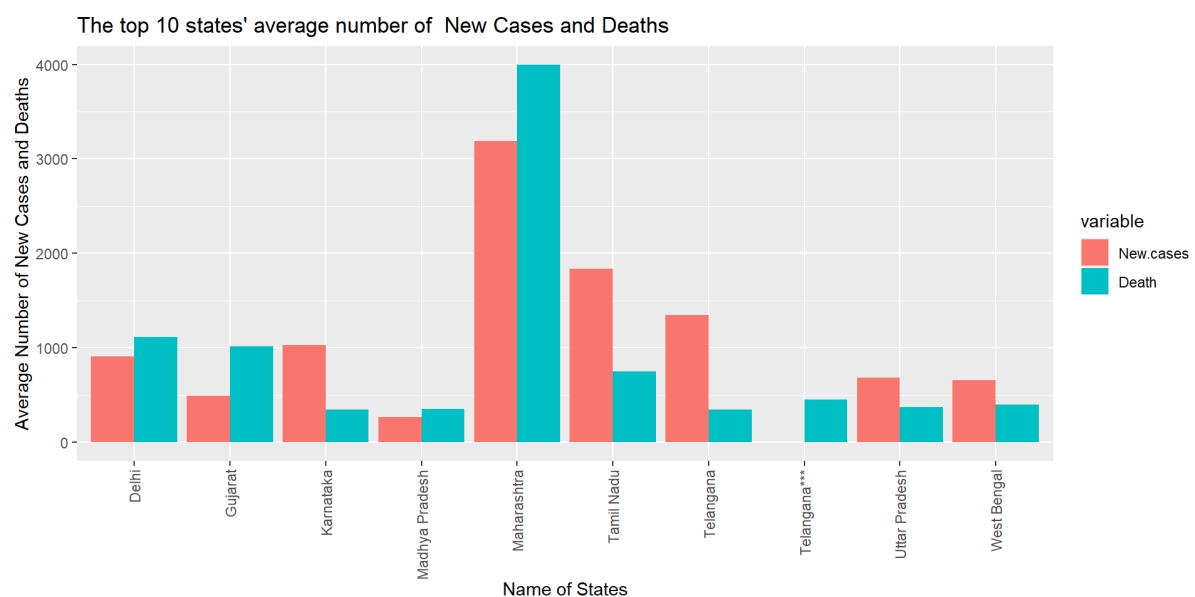


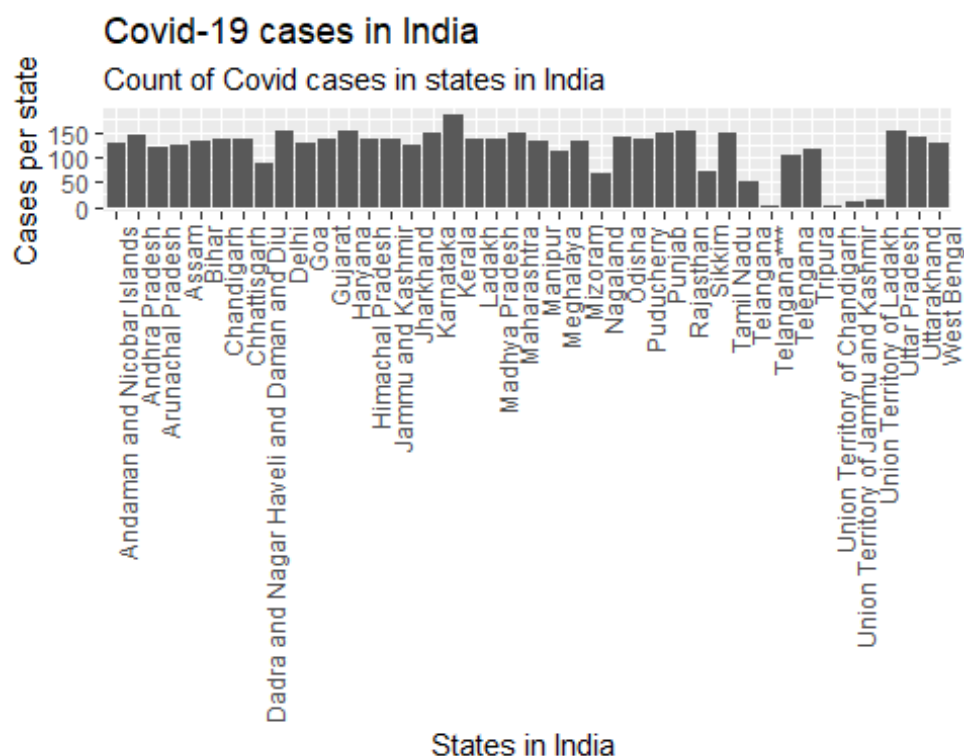
Figure 2

The figure 2 above indicates the top 10 highest average number of new cases and deaths over states of India. The Marhrasthra has consirderably maximum numbers of new cases and deaths with 700 new cases and nearly 1000 deaths . And the state which has the lowest number of new cases and deaths Madhya Pradesh with almost 150 new cases and over 200 deaths.

## Part 1

The dataset is the downloaded from Kaggle. It is the Covid dataset the states in India with variables that cover those tested positive to the virus, denoted as confirmed cases, deaths, Dates virus detected, the number recoveries, geographic locations denoted longitude and latitude and more. It gives a rich historical record the document to analyse and draw insight from.

The graph below summarises the figure of covid cases per state. The insight from the graph below is that the spread of Covid-19, across India, is widespread and the number case are not far between except for two or three states. The graph gives a good overview of the geographic spread and the distribution of the virus

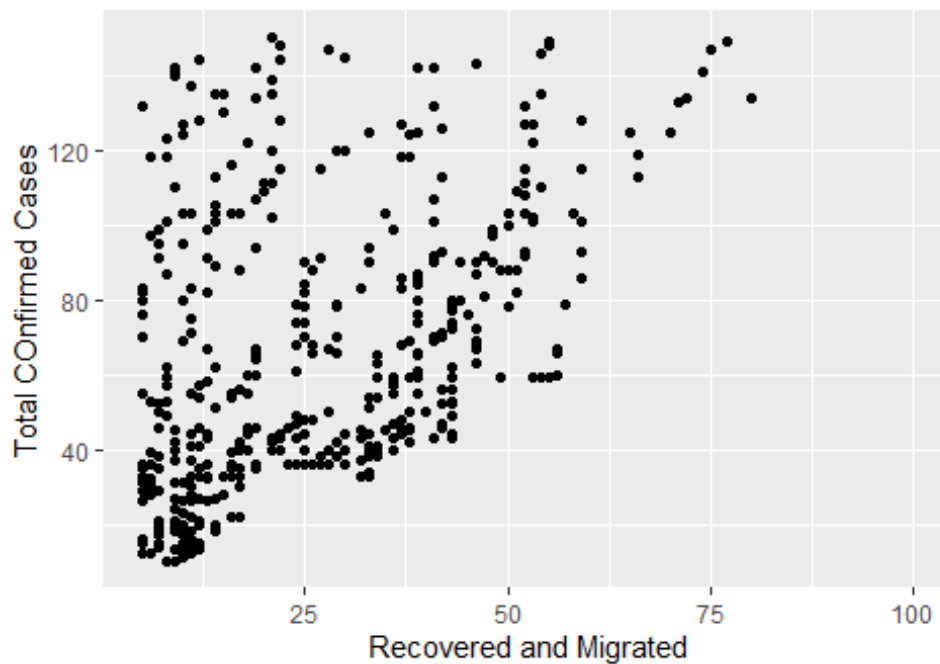


The graph below attempts to investigate the pattern in the implied distribution of Total Confirmed Cases as it relates to those that have survived and recovered from the virus. The implication of the relationship is that the recoveries seems to be suggests that as Covid cases rise, recoveries from the virus drop. The likely explanation is that new variants slow down the recoveries; the more time the virus has, the more new variants that are resistant to the vaccine being produced. Also, some people who may have developed natural immunity may not have immunity for the new variants

```
## Warning: Removed 4040 rows containing missing values (geom_point).
```

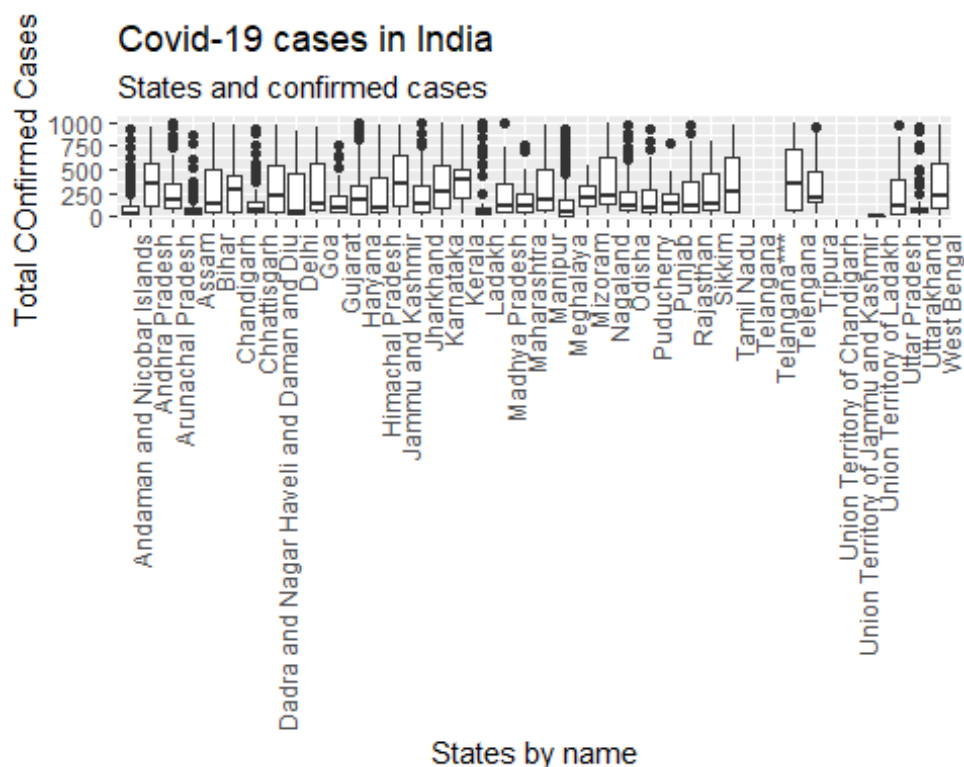
## Covid-19 cases in India

Total confirmed cases as it relates to recoveries



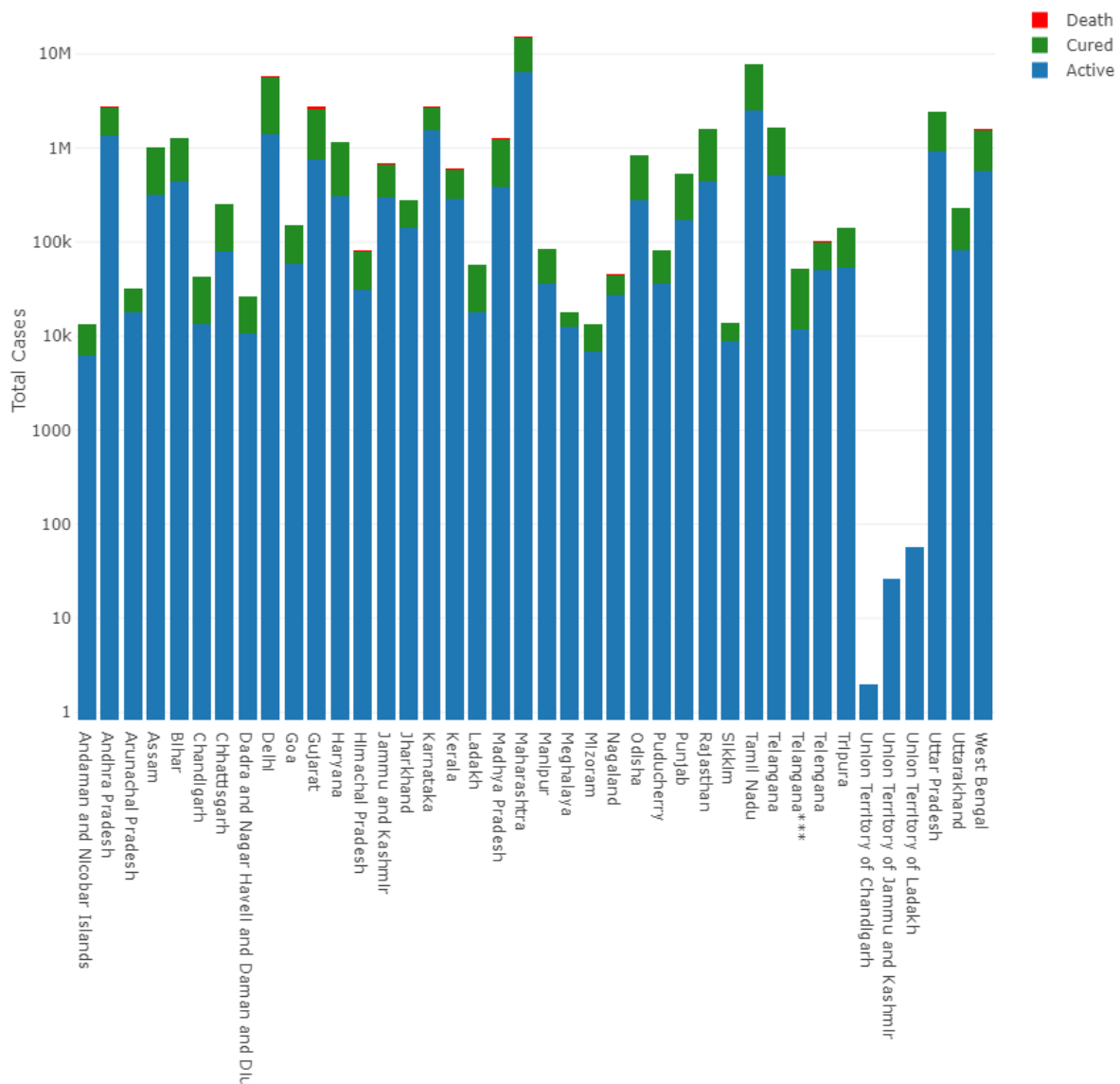
The graph below shows covid cases and the outliers. Some states regular or arithmetic increases while other showing geometric in denoted by the number of outliers they posted above the box plots. The sudden jump in the number of total confirmed cases maybe explained by population density, the geographic relationship of the states. For exampmple, states that borders may have similar rate of spread

## Warning: Removed 2778 rows containing non-finite values (stat\_boxplot).



## Analyze State wise Total Cases

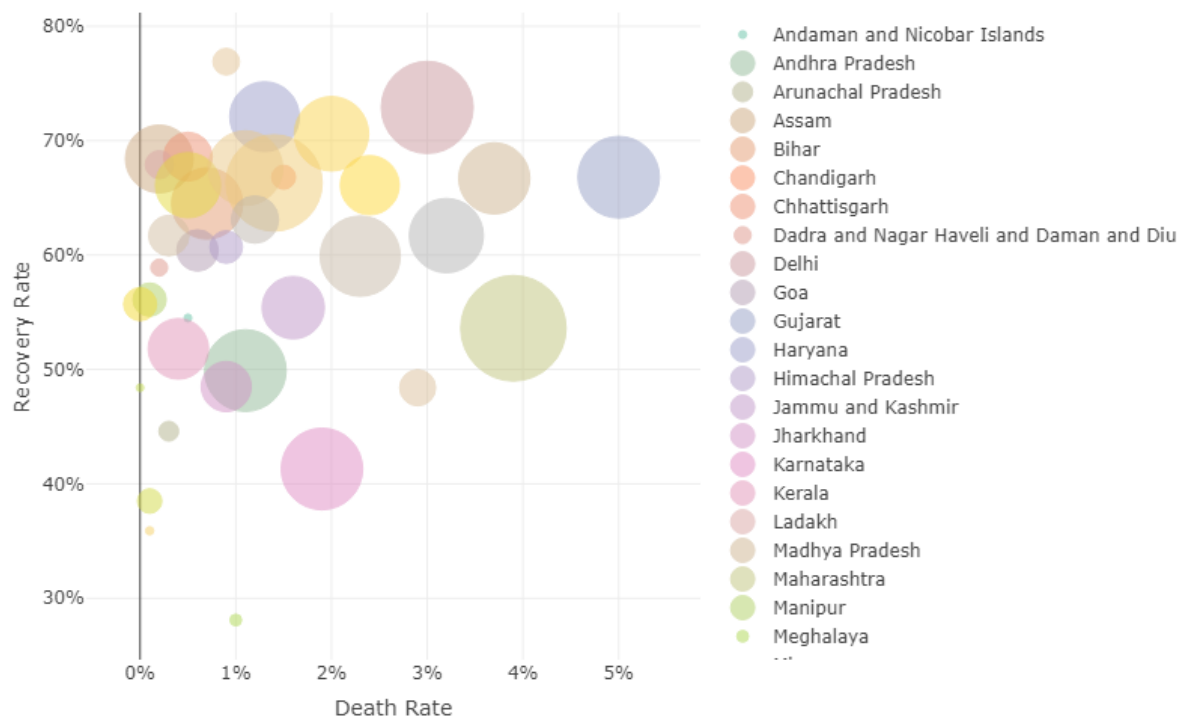
The bar graph of total cases in Indian states and according number of cases how many death and cured number of cases shown below graph. As shown in graph Maharashtra had over 10million number of cases between 30-01-2020 to 06-08-2020 which is highest number of cases overall to compare to other states and also active case, cured cases and death cases is around 6.458M, 8.146M and 587.648k number of cases which is more than to compare different stats. A second highest is Tamilnadu for cases of active, death and cured. Union Territory of Chandigarh has only 2 active cases and 0 deaths and cured cases. As following data also find 0 death and 0 cured cases in Union Territory of ladakh and Union territory of Jammu and Kashmir.



## A rate of total cases in India base on different type of cases (Death and Cured cases)

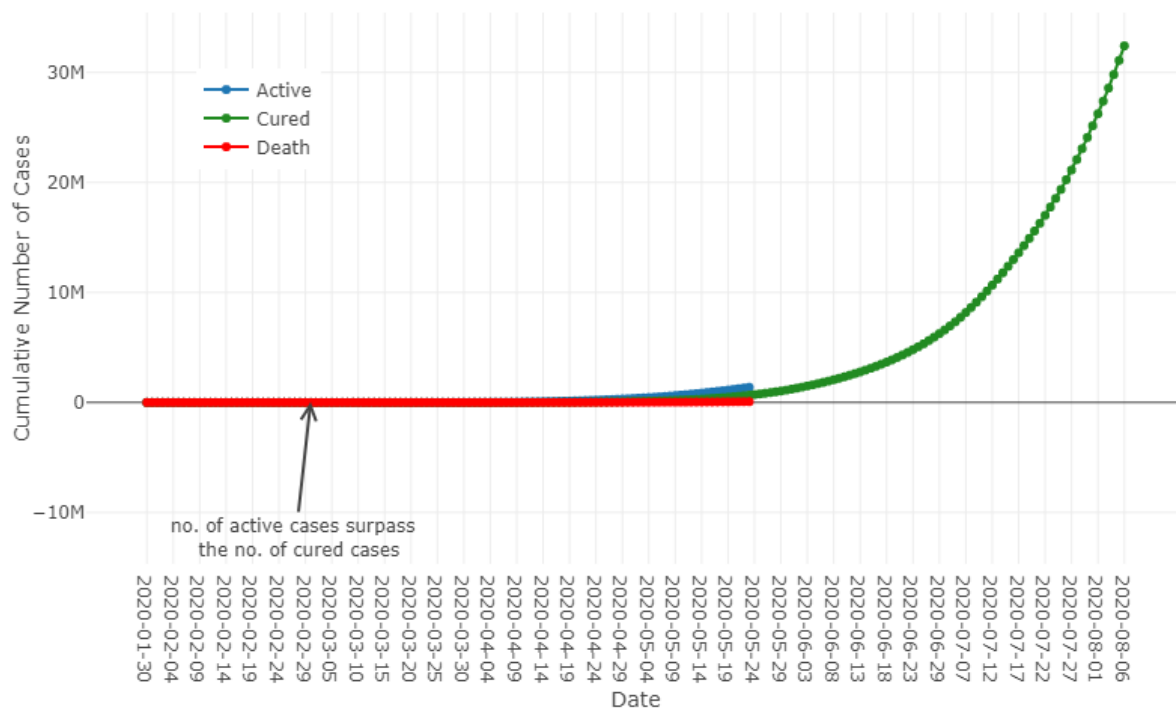
As below graph is showing death rate and recovery rate by states wise and comparison of each other for confirmed cases, cured rate and death rate. In this scenario Gujarat is most death rate compare to other states which is 5% death rate. Telangana\*\* has highest cured rate of cases which is 76.9%. Mizoram and Puducherry has 0% death rate total out of 13335 and 82967 confirmed cases. Meghalaya

is one of the states which has minimum cured rate it is 28.1%. As to comparison clearly show which states has how much rate of death, cured and active cases.



### Date wise cases differ and changes in numbers of cases

A scatter plot describe date wise how many number of cases active, cured and death. As per dataset 23-05-2020 date has 1.378M total number of cases and after that cured cases is surpass of active cases. Total number of death at 23-05-2020 is 67.317k. At the end of 06-08-2020 total cured cases are around 32.413M out of all confirmed cases. As clear plot describe the simply way date wise growing cured cases of corona virus in India.

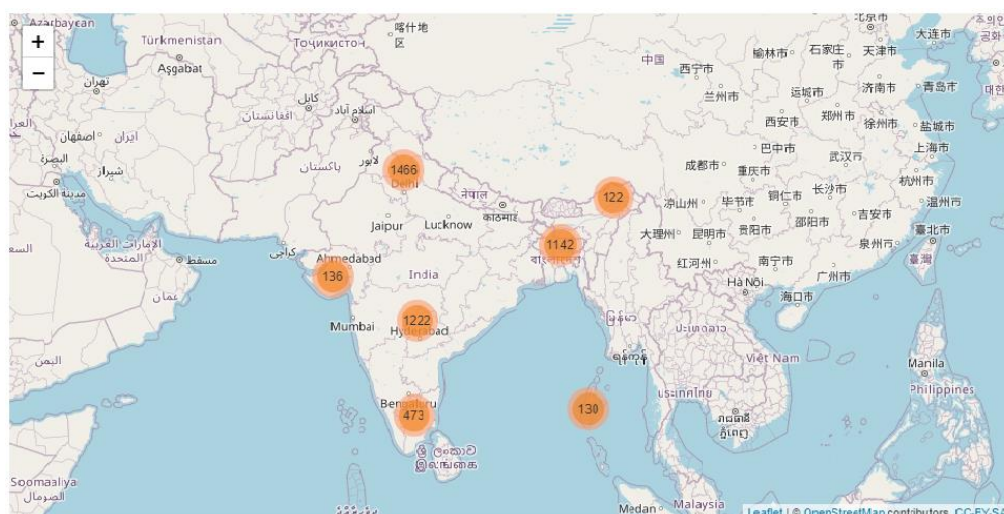


## Introduction:

In the history year 2020 will be written as one worst time in human era. According to Johns Hopkins University (JHU) more than 2,686,366 deaths were reported globally due to spread of virus out of which 159,216 deaths were reported in India. In order to understand how things went we will be exploring covid19 dataset which have records of India cases from date 1/30/2020 to 8/6/2020.

## Geo mapping of Data Point

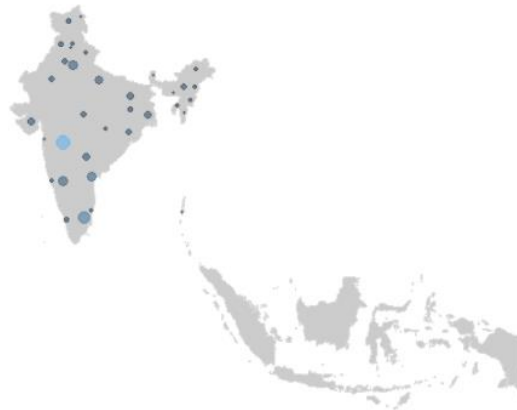
Below is the geographical mapping of data points for the covid19 cases. With help of this one explore exact locations of reported covid19 cases.



Below is geo mapping of confirmed covid19 cases using geolocation data points.

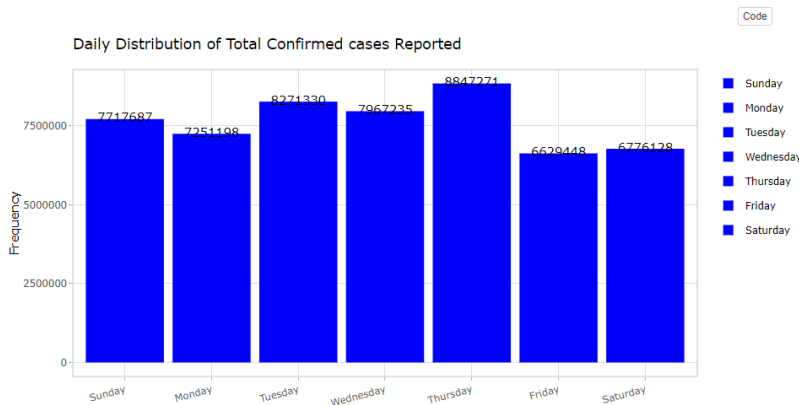
[Code](#)**2020-08-05**

## Corona Virus Total Confirmed cases

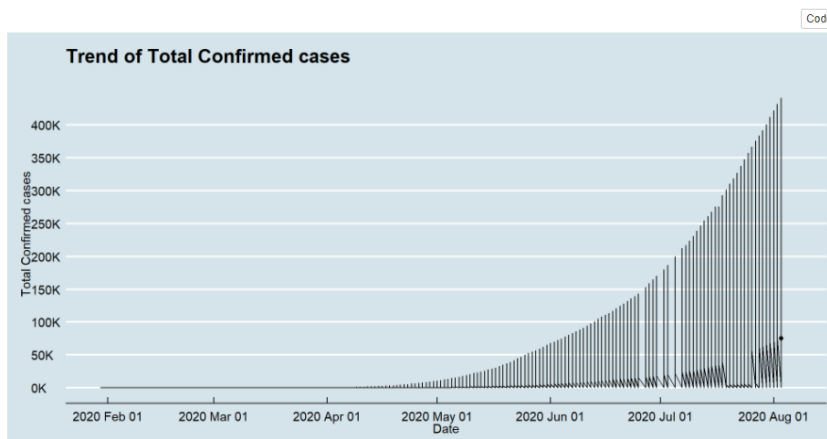
**COVID-19**

One of the key finding of below visualization was that during the **weekdays** total number of confirmed **covid19** cases was pretty high as people use to work and run their business whereas during weekends confirmed covid19 cases were not that high as people use to stay at home and spend time with their friends and family.

### Daily Distribution of Total Confirmed cases Reported



### Trend of Total Confirmed cases





## Top 10 states of India with the highest confirmed cases.

In the below graph we can visualize the time series bar plot for the top 10 states of India with the highest confirmed covid19 cases form date 1/30/2020 to 8/6/2020. It was found the **Maharashtra** had reported highest number of confirmed covid19 cases.

