COVID-19

Impact of corona virus in United states.

Team

Introduction of COVID-19 Project

- The COVID-19 is the pandemic that is affecting the whole world, especially USA.

- Since USA is the most affected country in the whole world, and hence we have chosen to perform case study on it by analyzing the data finding interesting insights in the dataset.

- The data will be gathered online and it will be imported in R to perform exploratory analysis and Tableau will be used to find some insights in the data by visualization.

- We have started working on this project since the first week of march and we have used the data from the outbreak till the 25th April 2020 for exploratory analysis.

The Problem

Covid-19

COVID-19 commonly known as coronavirus was first identified in Wuhan, China in 2019, since then it has spread worldwide and caused a pandemic. It is an infectious disease caused by severe acute respiratory syndrome with a fatality rate of nearly 1 percent.

Possible Entry In US

COVID-19 entered US through following possibilities

- Imported cases in explorers
- Cases among close contacts of a known case
- Community-procured situations where the wellspring of the disease is obscure.

Problem statement

- There have been more than **1.2 million** confirmed Covid-19 infections in the United States.
- Infections have remained stubbornly elevated.

Agencies responsible for covid-19

White House Coronavirus Task Force

The **White House Coronavirus Task Force** is a **United States Department of State** task force that "coordinates and oversees the Administration's efforts to monitor, prevent, contain, and mitigate the spread" of the **coronavirus disease** (COVID-19).

Centers for Disease Control and Prevention

The **Centers for Disease Control and Prevention** (**CDC**) is the leading **national public health institute** of the United States. Its main goal is to protect **public health** and safety through the control and prevention of disease, injury, and **disability** in the US and internationally.

COVID-19 DATA

Description of the data

COVID-19 dataset comprises the total number of tests conducted, breaking out positive, negative, and hospitalized patients. The data is collected from "The COVID Tracking Project" collects its data from state/district/territory public health authorities—or, occasionally, from trusted news reporting, official press conferences, or tweets or Facebook updates from state public health authorities or governors.

Description of Data

- State - State or territory postal code abbreviation.
- Positive - Total cumulative positive test results.
- Negative - Total cumulative negative test results
- Death - Total cumulative number of people that have died.
- Total test - Total no of tests conducted.
- Hospitalized - Total cumulative number of people hospitalized.

Limitations:

Not all the states consistently report their test results and regularly. In such cases, they use other reporting tools like directly asking state officials, watching news conferences, gleaning information from trusted news sources. Moreover, since the symptoms are not visible until 14 days, the number of actual positive cases may be more than reported to the state/district/territory public health authorities, etc.

Data Analysis Tools
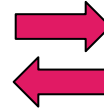
❖ R-Studio

RStudio is an integrated development environment for R, a programming language for statistical computing and graphics. It provides free and open-source tools for R and enterprise-ready professional software for data science teams to develop and share their work at scale. RStudio makes it easy to set your working directory and access files on your computer.

❖ tableau

Tableau is a powerful and fastest-growing data visualization tool used in the Business Intelligence Industry. It helps in simplifying raw data into a very easily understandable format.

Tableau is an interactive, self-service reporting and analytics tool that enables faculty and staff to integrate and combine data from multiple sources into visualizations and be accessed in a single desktop environment using Tableau Desktop or through a shared dashboard.
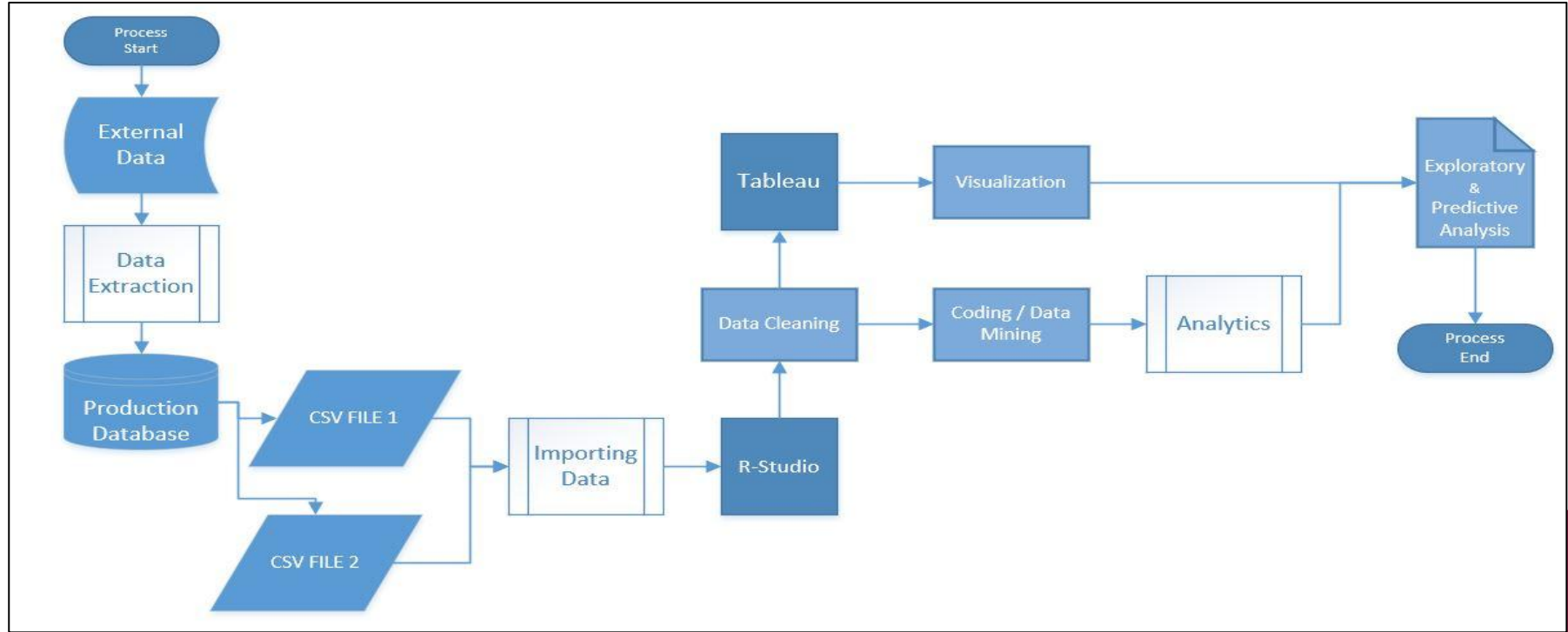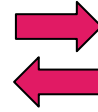
Data Cleaning and Manipulation
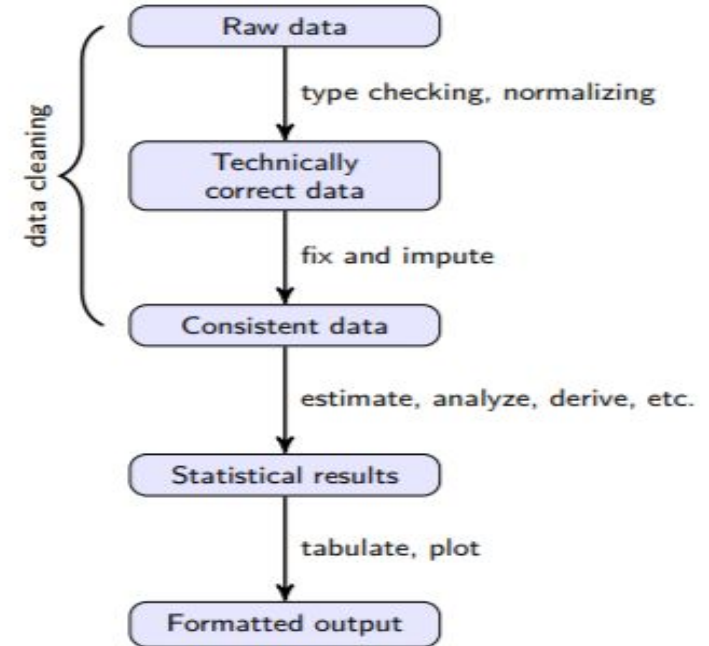
Exploratory Analysis of data

```r
# Top 5 states in US with highest positive cases and lowest positive cases
```{r}
cases1=cases%>%filter(date=="20200423")
cases1%>%arrange(desc(positive))%>%head(5)  # top 5
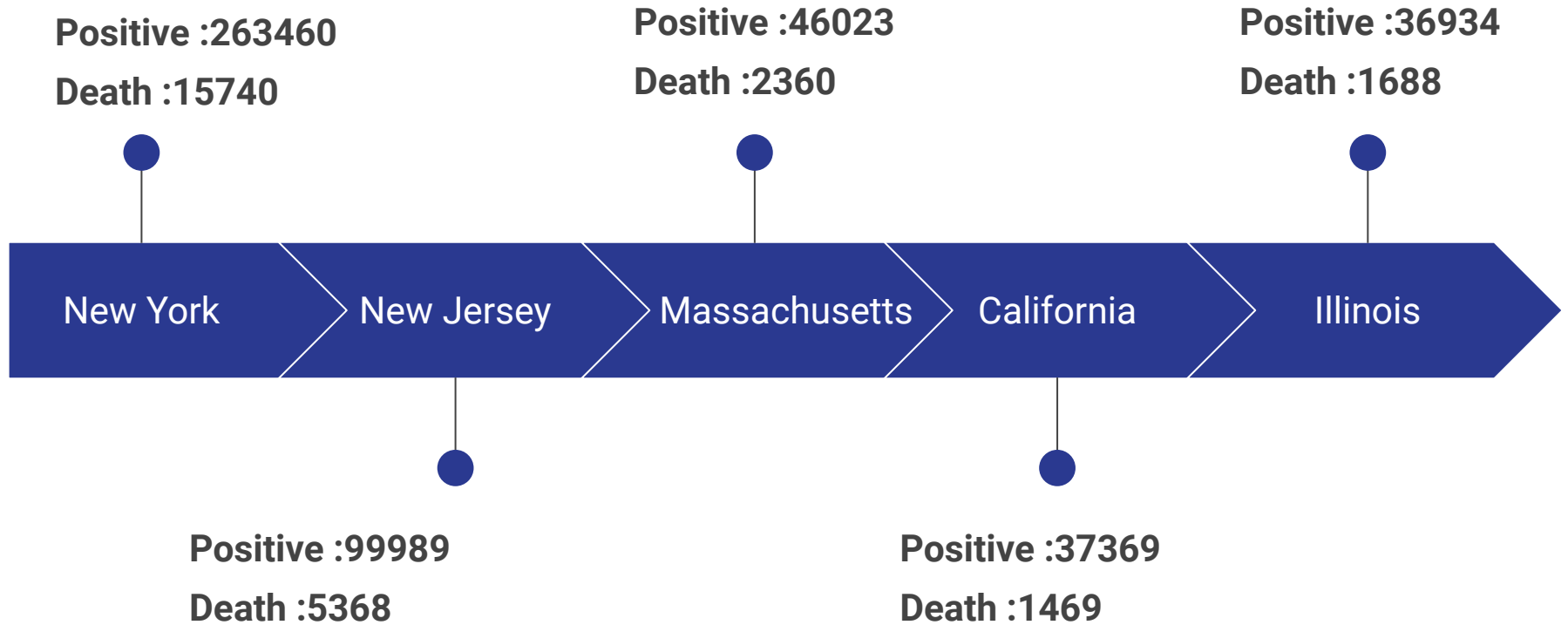cases1%>%arrange(desc(positive))%>%tail(5)  # bottom 5

#Increase in cases in New York
caseny=cases%>%filter(state=="NY")

ggplot(caseny)+geom_line(aes(x=date, y=positiveIncrease))
```
```

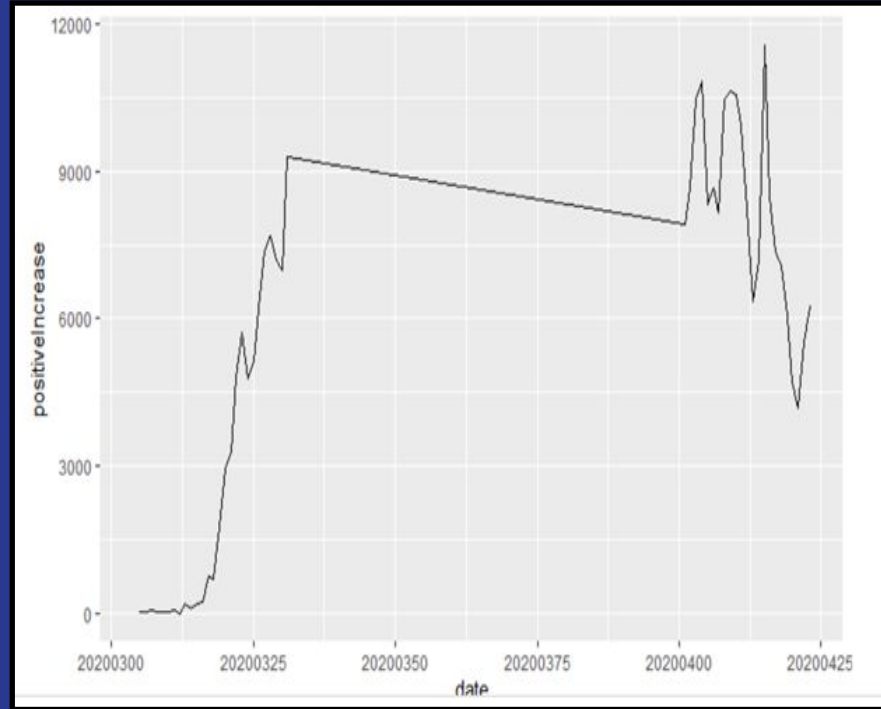|	date <int>	state <fctr>	positive <dbl>	negative <dbl>	pending <dbl>	hospitalizedCurrently <dbl>
1	20200423	NY	263460	432460	0	15021
2	20200423	NJ	99989	100159	0	7240
3	20200423	MA	46023	149053	0	3890
4	20200423	CA	37369	444728	0	4929
5	20200423	IL	36934	136382	0	4877

	date <int>	state <fctr>	positive <dbl>	negative <dbl>	pending <dbl>	hospitalizedCurrently <dbl>
52	20200423	WY	326	7241	0	0
53	20200423	GU	135	1180	0	2
54	20200423	VI	54	583	31	0
55	20200423	MP	14	51	0	0
56	20200423	AS	0	3	17	0

Positive :263460
Death :15740

Positive :46023
Death :2360

Positive :36934
Death :1688

New York

New Jersey

Massachusetts

California

Illinois

Positive :99989
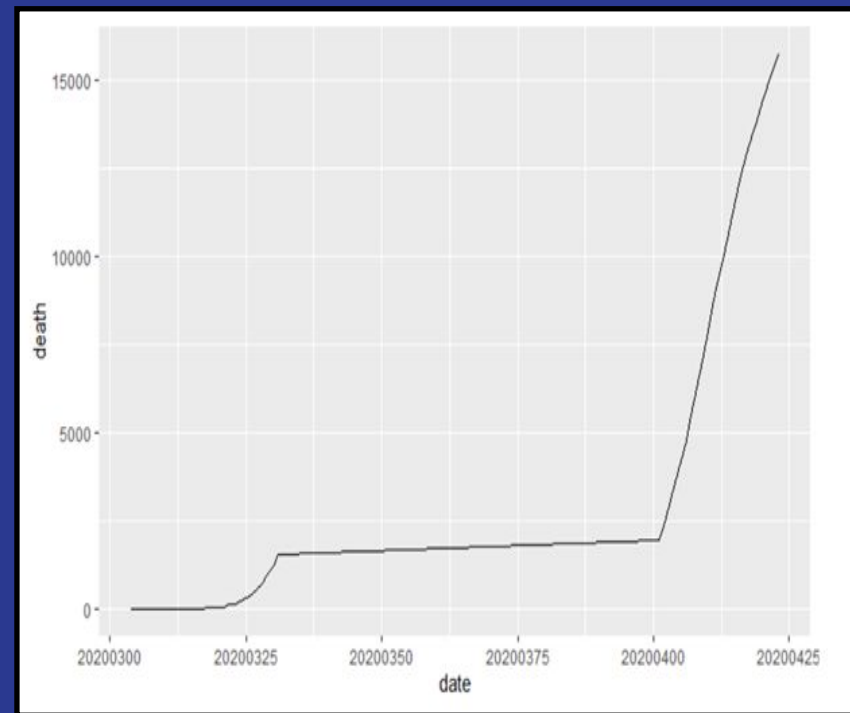Death :5368

Positive :37369
Death :1469

Observations

Trendline to get more insights about the positive cases in NY
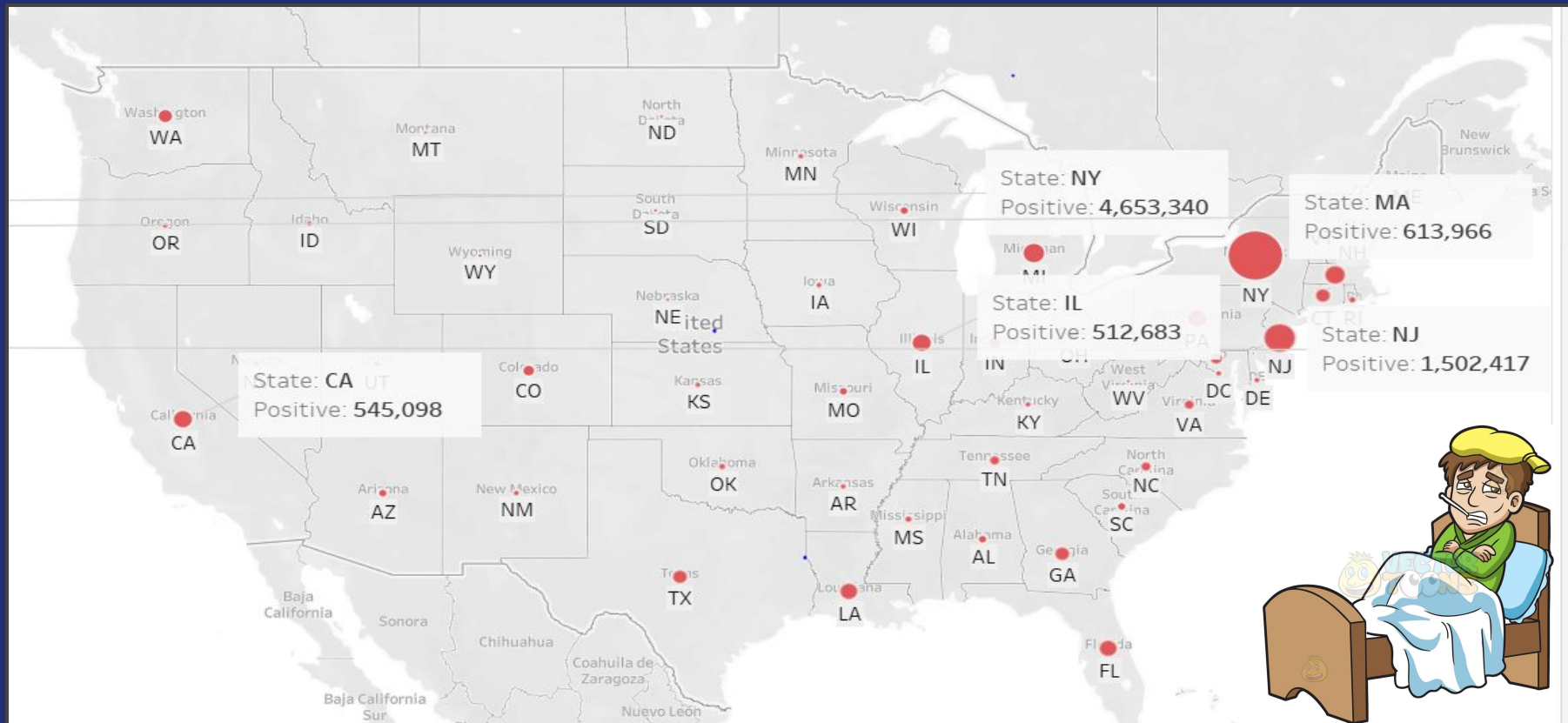
Observations

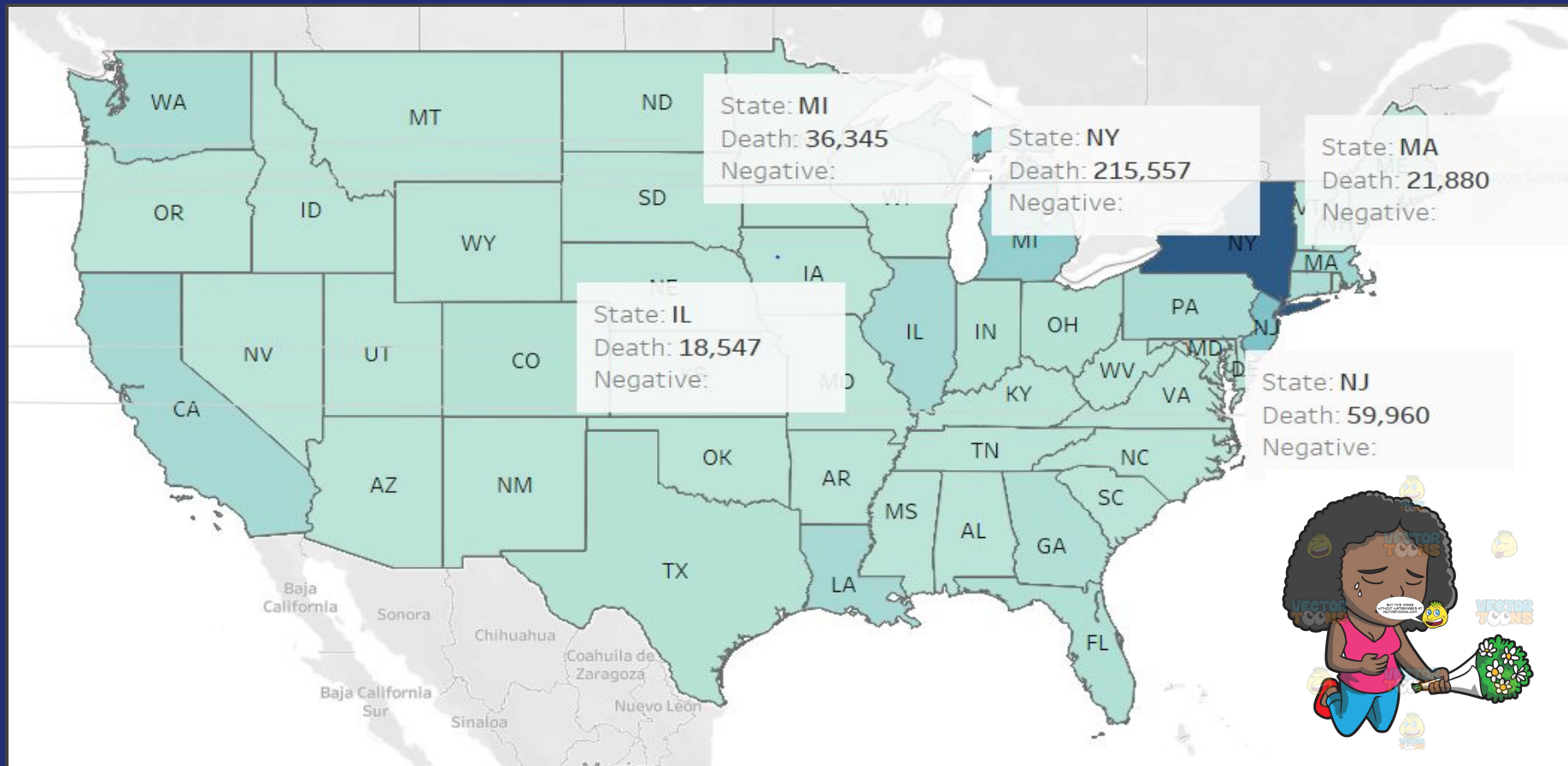Trendline to get more insights about the death rate in NY.

Method to Visualize data using Geo-mapping technique

1. Load the data file into Tableau and create a new project called COVID-19.

2. Drag the state dimensions and drop into the mark section. Select the geo-mapping view under "show me" drop down option.

3. Drag the required column into the marks section and aggregate using sum function.

4. Drag state into the label.

5. Select the states which are required to highlight and annotate.

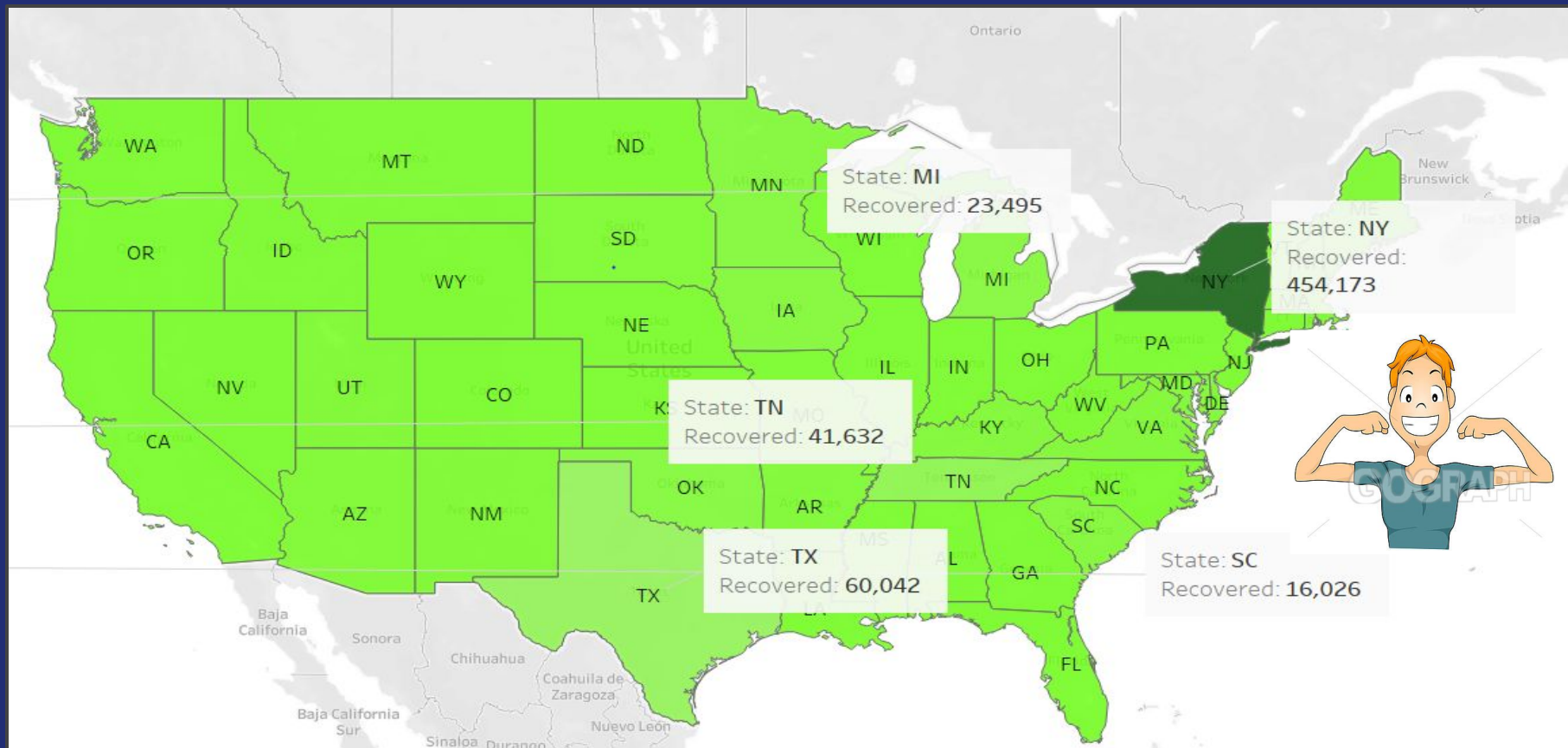6. Select the colors in the marks option and edit for the desired color.

Top 5 Positive cases

State: NY
Positive: 4,653,340

State: MA
Positive: 613,966

State: IL
Positive: 512,683

State: NJ
Positive: 1,502,417

State: CA
Positive: 545,098

State: MI
Death: 36,345
Negative:

State: NY
Death: 215,557
Negative:

State: MA
Death: 21,880
Negative:

State: IL
Death: 18,547
Negative:

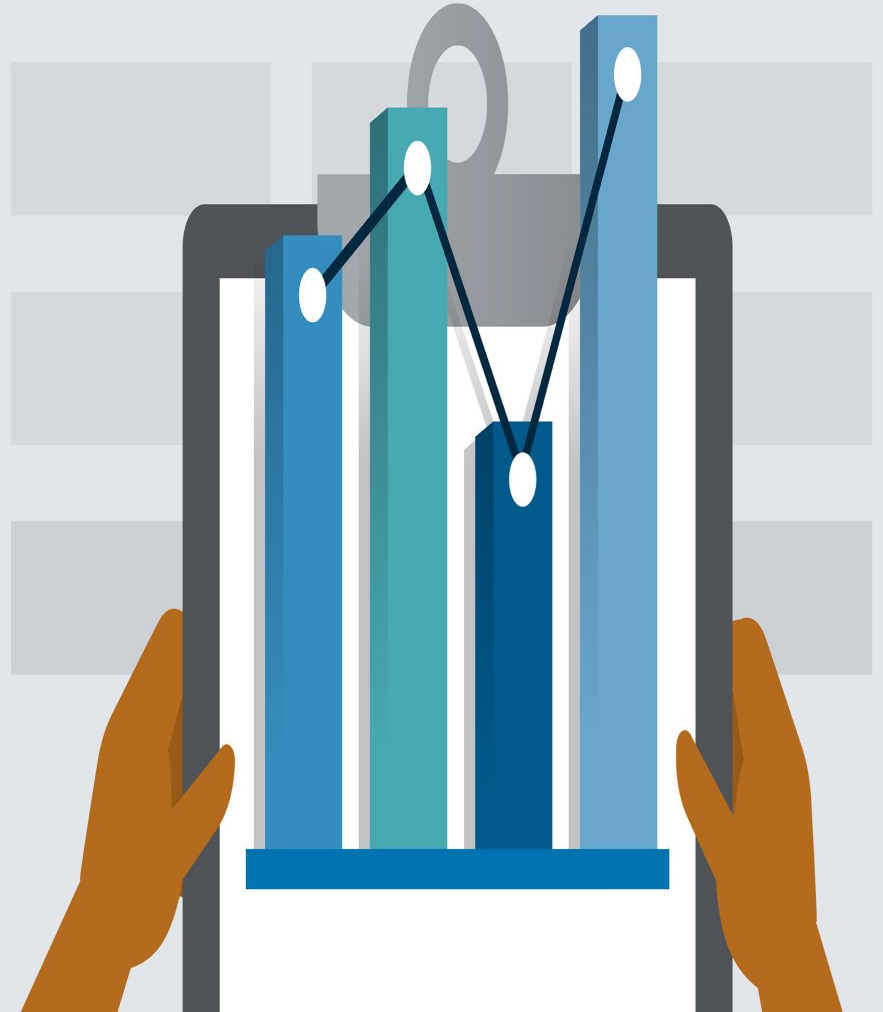State: NJ
Death: 59,960
Negative:

Top 5 Death States

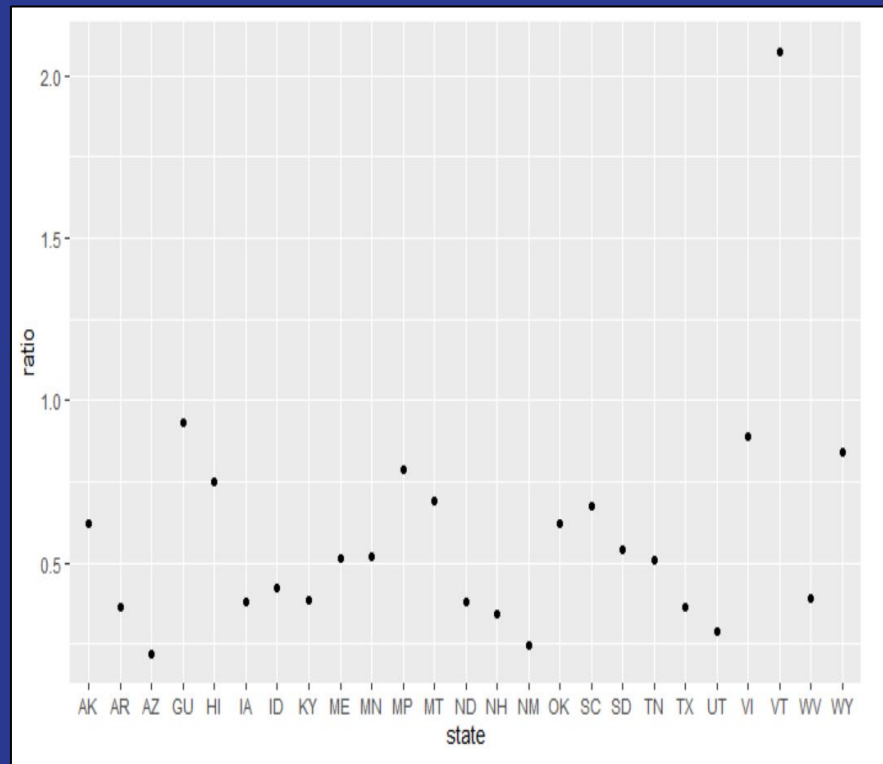Top 5 Recovery states

Analysis of the data

Analysis

State that will be the first to get out of this pandemic

This Graph Shows the representation of the score of the states that will get out of this pandemic. Highest score implies the fastest recovery and from the result we can
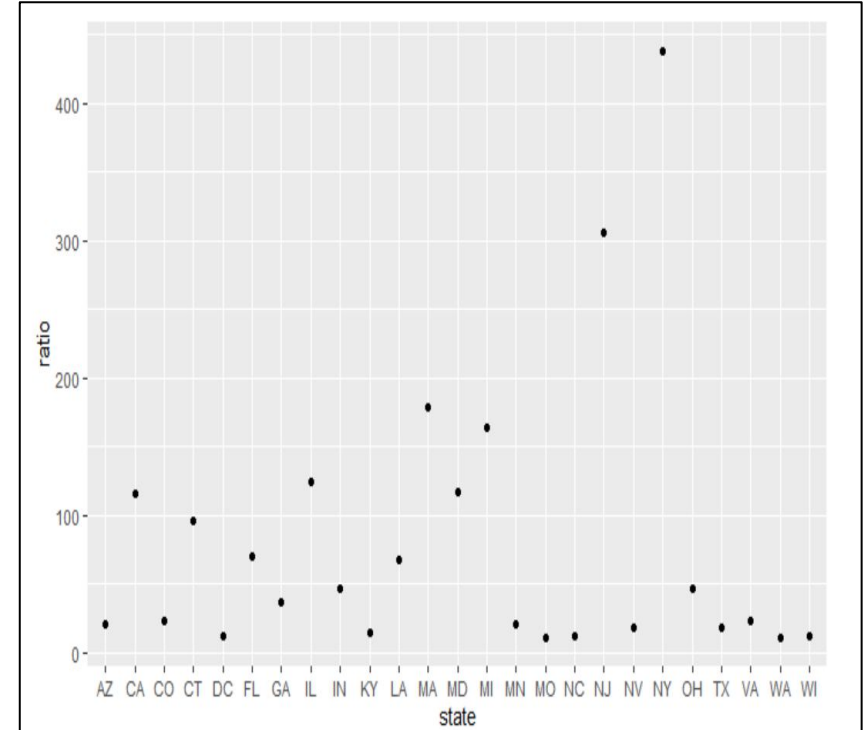see that Vermont(VT) has the highest possibility to get out of this pandemic and is safe compared to other states.

Analysis

State that will be the last to get out of this pandemic

This graph shows the graphical representation of the score of the states that has been hit hardest and will be the last to get out of this pandemic. The highest score implies the least recovery and And from the result, we can see that New York(NY) is in a critical situation and New Jersey(NJ) is in the second position.

Observations from Analysis

➔   New York is leading in terms of positive cases followed by New Jersey, Massachusetts, California, and Illinois.

➔   New York is in critical situations and will be the last to come out from this pandemic. We used the ratio of death increase to that of recovery cases.

➔   We have observed that New York data might show that it can recover fast but still New York is in a critical situation because of the death ratio.

➔   Vermont will recover fast and will be the first to come out of this pandemic. We have used a ratio of recovery cases to that of positive cases.

Summary

1. Hence we conclude the introduction of covid-19 its effects, data cleaning process and the dataset used for the exploratory analysis.
2. Additionally, we have also discussed the visualization using Tableau and we have encountered various observations.
3. As per the observations and the analysis carried out on the dataset, we have concluded with few recommendations given by two task forces namely White House Coronavirus Task Force and Centers for Disease Control and Preventions.
4. We have taken certain mathematical ratios to acquire desired result and make graphs of the results for better understanding using Rstudio

Recommendations

➔ Digital platforms or applications to keep citizens informed, enable public participation, and/or offer open data; Digital tools to enable public participation.

➔The government should take prior measures as the pandemic is with spurt increase every day otherwise the situation in socio-economic problems may lead to riots and chaos by poor and ill people.

➔ Controlling rumors to avoid panic among the public.

How to control it ?

- ➔ Keep at least 6 feet between yourself and others.

- ➔ Wash your hands with soap and water often.

- ➔ Cover your nose and mouth with a tissue or sleeve when sneezing or coughing.

- ➔ Do not touch your face with unwashed hands.

- ➔ Monitor your health more closely than usual for cold or flu symptoms.

THANK YOU
ANY QUESTIONS?

Presented by :
Gaurav Kondapuram
Deepen Patel
Akhil Podila
Ankit Singh
Sweni Thakkar
Smit Ajmera.