CSE 564 VISUALIZATION
PROJECT REPORT

Pavan Gangakhed(112689083)
Abhishek Kumar(112680754)
https://youtu.be/WWu8Udi6RiE

Background

COVID-19 is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which was first identified in Wuhan and is now spread across the globe. The number of new cases is increasing everyday around the world. Through the tough times we are going currently, we realize the significance of health care and research. The robustness with which a health care system performs and responds to ever changing needs is critical to saving lives during such a pandemic.

Well-designed Dashboards would help COVID-19 response teams to chart out strategic plans by understanding how different measures and policies implemented in other countries were effective in slowing the spread of the disease. It also helps them to better understand the data and derive insights from them and at the same time informing the public about the ongoing situation in their areas. So our aim here is to design a real-time dashboard with data visualization tools to better grasp the current status of COVID-19 as the situation unfolds.

Objective

We will develop a dashboard that would be helpful in tracking the cases that can be used by public or health officials to better understand the ongoing developments in real-time. We aim to visualize the data in a way that brings forth the most important details like the region which is leading in growth of new cases/death, high risk zones, most common symptoms people have experienced etc. Effects of Lockdown in certain regions, social distancing visually observed from statistics and insights drawn from them.

1. To visualize the spread of disease across different states in USA using various data visualization techniques.
2. Determine low risk zones where the disease is controlled to gain more insights into the underlying causes for the same.
3. The amount of testing for the virus being done in different states.
4. Filter the number of confirmed cases and deaths in different states.
5. Chart to visualize the number of cases and deaths and changes over time in a specific state or territory which helps to understand when the curve is flattening (no new cases are identified)
6. Different unrelated trends observed with strong correlation to the evolving cases.

Dataset

CDC, University of Virginia website and Kaggle are our sources of data (Appendix). It is a time series dataset which contains information about the number of new cases, deaths and number of recovered people on a daily basis.

One of the datasets shows fatality statistics for both male and female in different states. This data is displayed with a radar chart on the dashboard. Number of confirmed cases, deaths and recoveries across all the states in the US is sourced from another dataset mentioned in the Appendix.

Data in other forms which track the response of a public health system against the disease such as number of beds available in hospitals, number of samples tested on a daily basis, the resources available at disposal for proper management, etc. help in capturing other kinds of information from different perspectives to better assess the situation at hand.

Approach

Design

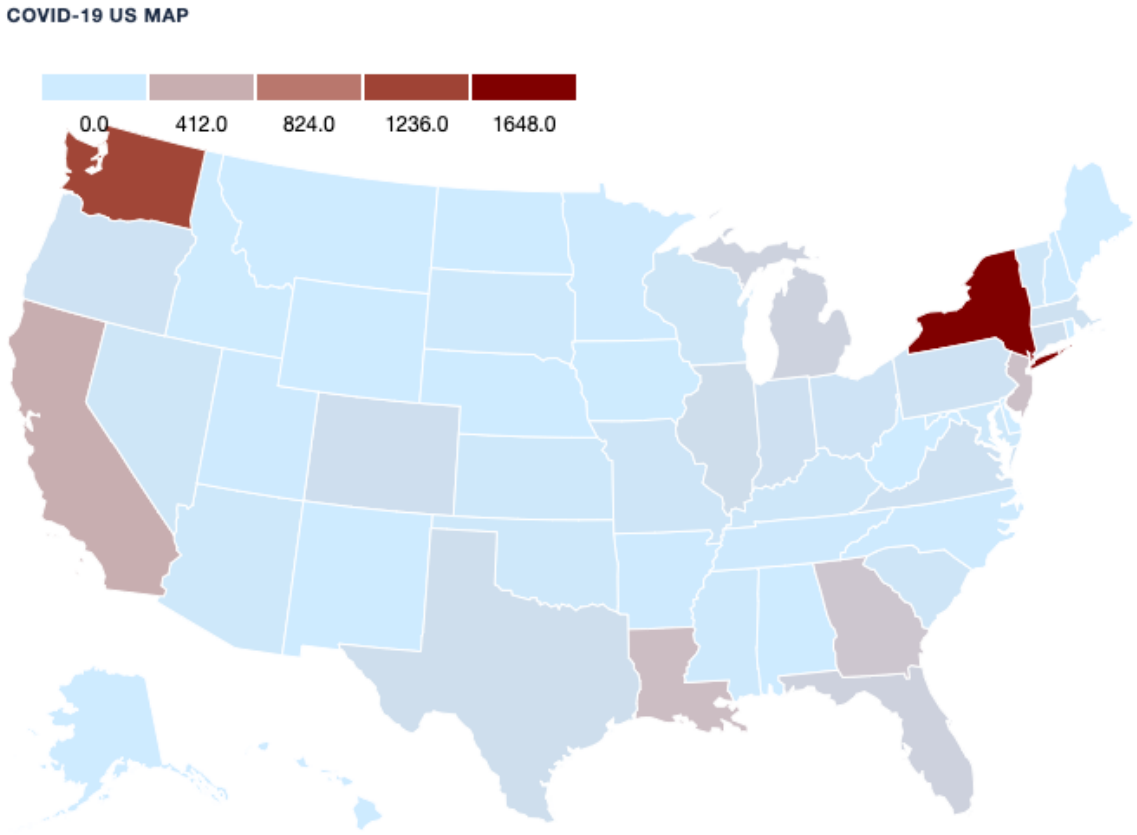
1. Choropleth map showing number of confirmed cases across different states in the US.
2. Time Series Line Chart of the confirmed cases across different states in the US.
3. Stacked Area chart showing the number of active cases over a period of 2 months.
4. Half Donut Chart representing total cases across different states in the US.
5. Sparklines to show succinct and short representation of time series along with the aggregated numbers on deaths, active cases and recovered cases.
6. Radar Chart showing the deaths distribution of across Age and Sex.
7. Parallel Coordinates with the medical resources obtained on the India dataset.
8. Interactivity between the graph elements and user input.

Methodology

1. Python - for data cleaning, munging and pre-processing for optimized performance and low latency.
2. Flask - for data exchange between back-end exposed via various endpoints and front-end to consume to draw various charts on the UI.
3. D3 - for data visualization.
4. HTML, CSS and Javascript.

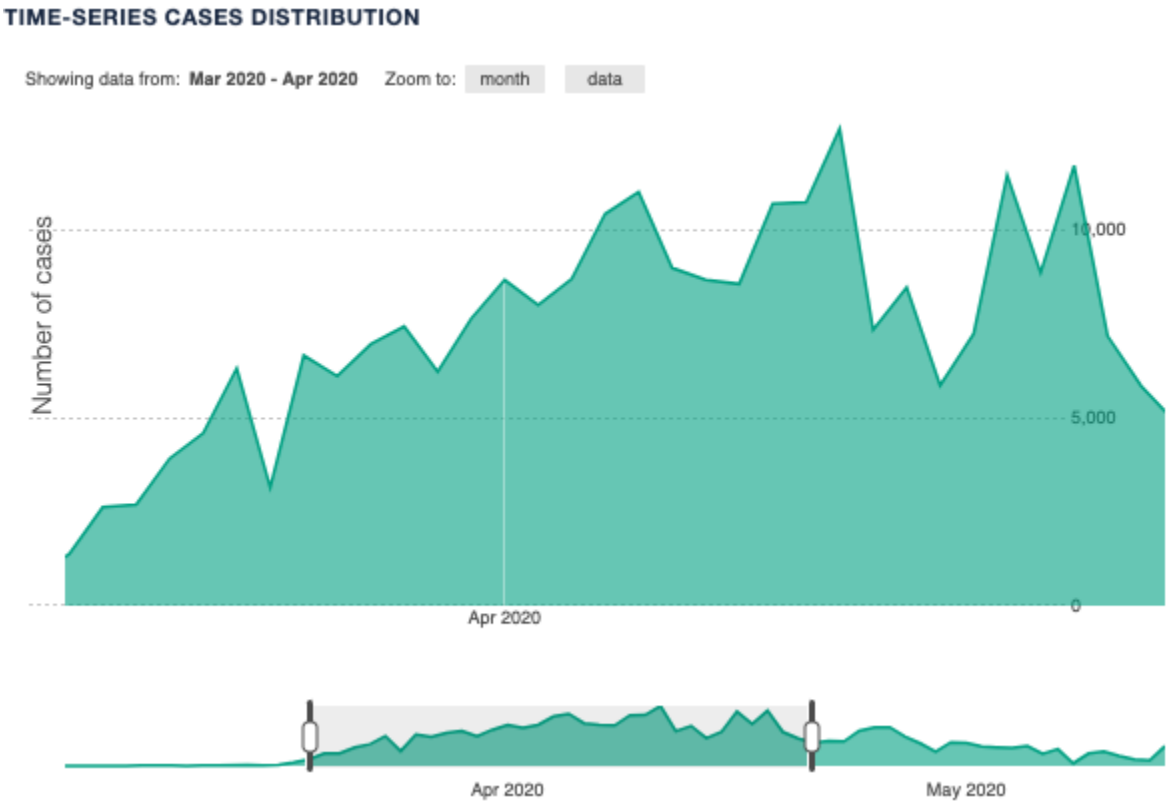
Components in the dashboard

Choropleth Chart



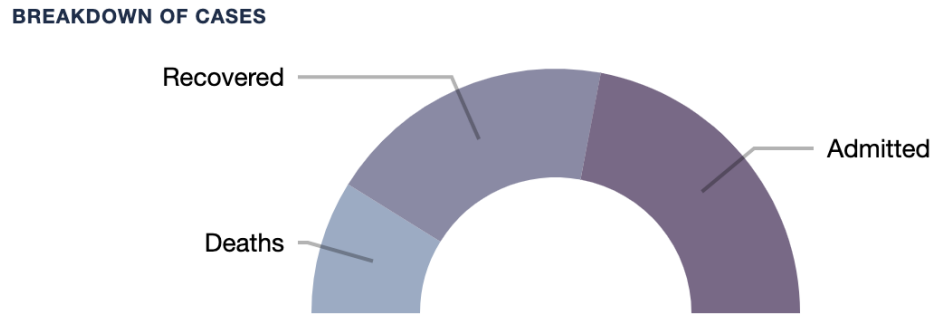
A choropleth map is used to represent distribution of the number of confirmed Covid-19 cases across all the states in the US. We can check how states stand against each other in the number of confirmed cases using the slider option on the time series chart shown on the dashboard to its right. The color scale assigns darker shades to states with a higher number of cases. This chart also interactive with the radar chart we used to show the deaths distribution across age and sex in the states.

Time Series Area Chart



Here we can see the day wise cases and custom date range can be selected via the buttons or the brush at the bottom. This chart is contextual - it displays the number of cases for the state selected on the map or United states.

Pie Chart for breakdown of Cases



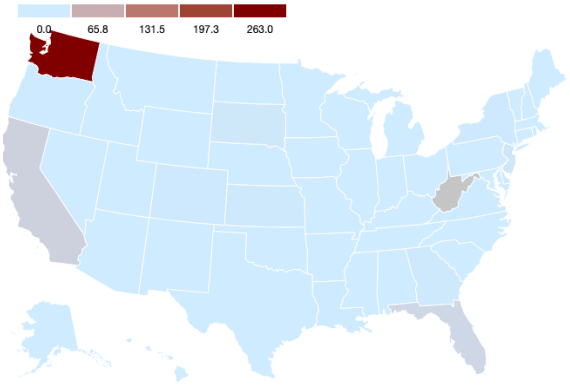
Breakdown of the cases for the state and time range selected is visualized with the help of half donut chart.

Infocards with Sparkline

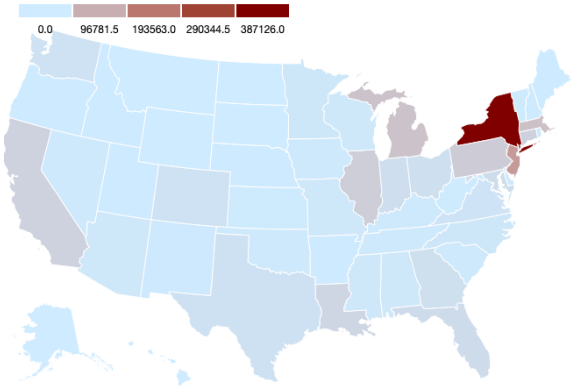
This section displays the aggregated number of confirmed cases, total deaths and recovered cases for the state/country and date range. It also shows the time series distribution per day of the same metric as a sparkline as can be seen to their right.

Results and Analysis

Trend Analysis



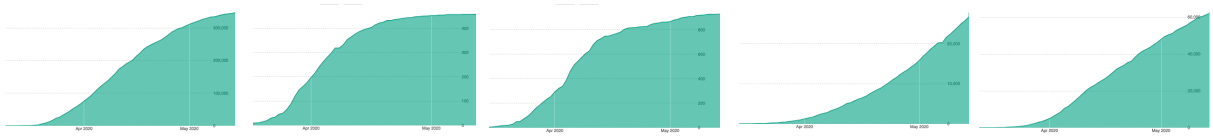
Hotspot - Washington (early March)



Hotspot - New York (mid May)

The onset of the covid-19 outbreak in the US started with Washington being the hotspot and then new cases flattened in a month. The trend shifted gradually to New York until it became the new leading hotspot by the mid May as shown in the charts used for comparison above. Different trends across different states can also be observed via the same choropleth by shifting the slider in the time series chart.

Flattening

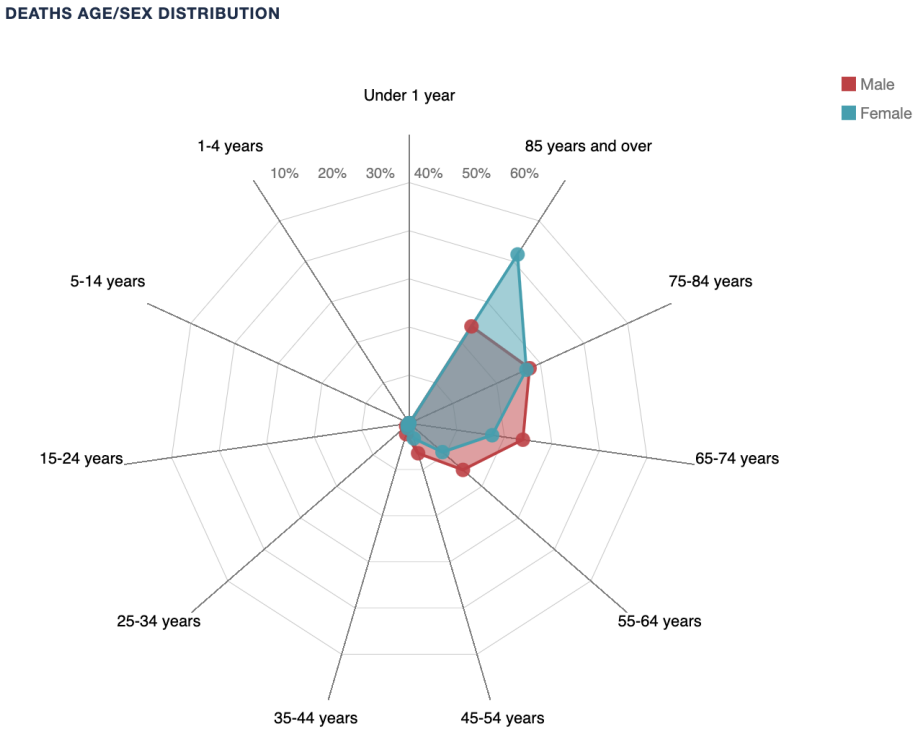


New York | Montana | Vermont | Virginia | Pennsylvania

Effects of Lockdown

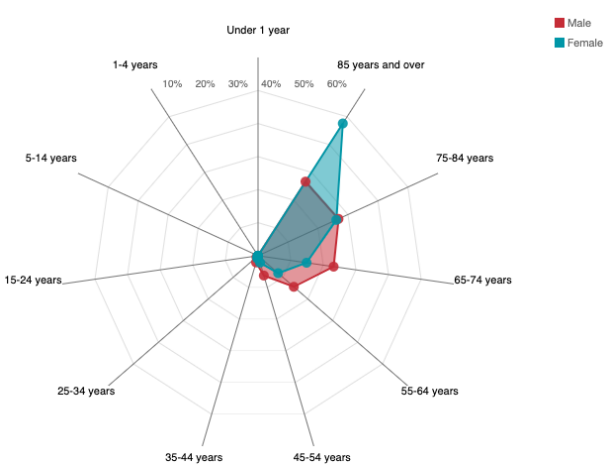
States such as Montana, Vermont and New York are flattening the curve as can be noticed on the cumulative distribution of the confirmed cases whereas states like Virginia and Pennsylvania are still rising and haven't flattened it out yet. Since the outbreak in these states happened much later than the New York counterpart it is expected to follow the curve for New York.

Fatality Rate for male and female across different Age groups

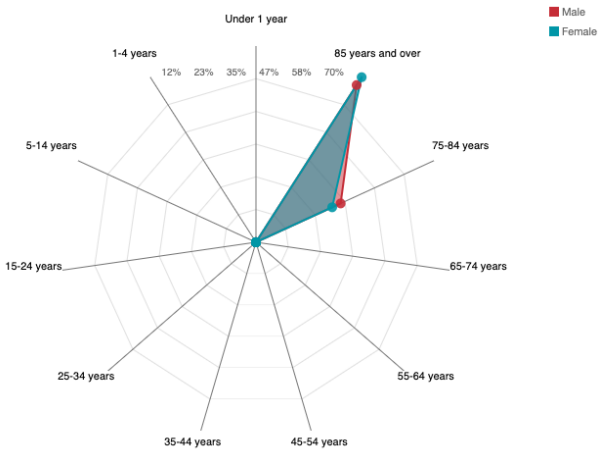


United States

State wise comparison of cumulative deaths across Age and Sex



New York



New Mexico

Some inferences from the data:

- Younger females are less susceptible than males of same age group
- Very less fatality for younger population across all the states in US
    - < 0.5 % for both males and females < 35 years
    - > 65 years – 75.3% for males, 85.9% for females
- In states like Vermont and New Mexico (0% for age < 75 years) deaths in age group > 65 years is far more skewed
- Females have less fatality rate for age group < 65 years than male counterparts

Appendix

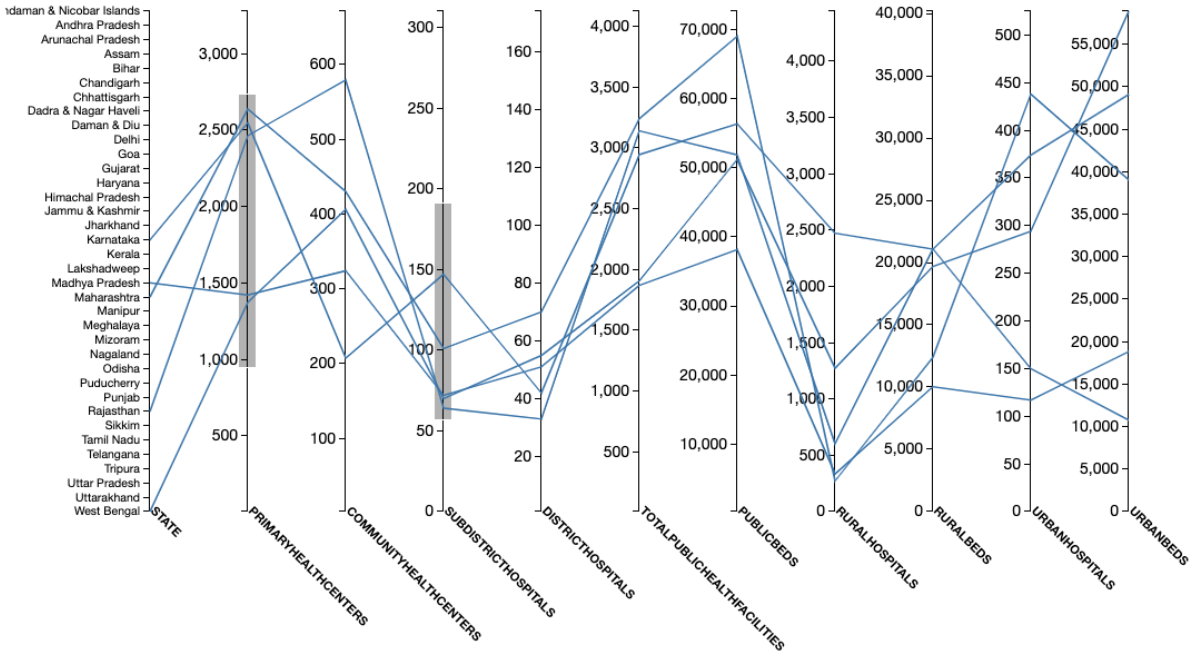
Dataset Links:

- https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Sex-Age-and-S/9bhg-hcku
- https://www.kaggle.com/sudalairajkumar/covid19-in-india
- https://nssac.bii.virginia.edu/covid-19/dashboard/data/nssac-ncov-data-country-state.zip

Analysis for Covid-19 in India

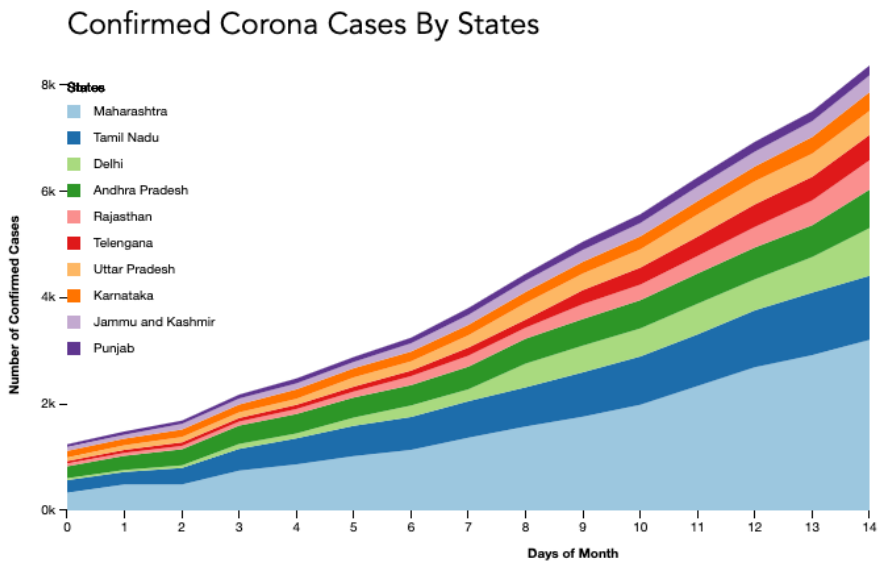
Parallel Coordinates

- To understand, plan and utilize hospital resources and medical need
- Some states need more resources in terms of the number of beds, number of hospitals to be better equipped with scaling to the intensity of the pandemic.



Stacked Area

This stacked area chart helps quickly identify relative performance of different states in India with respect to the emerging cases and how tightly measures such as lockdown are coming into effect in these areas.



Screenshot of the dashboard