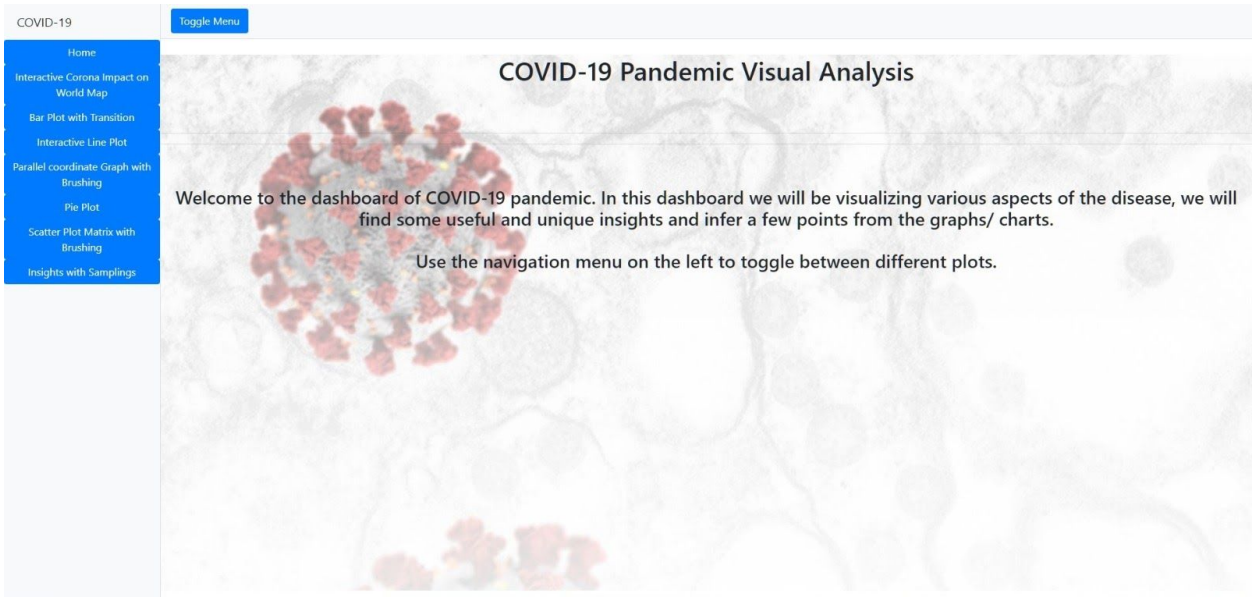
COVID19 Pandemic Visualisation and Analysis

Visualisation Spring 2020 Final Project Report



Sahil Mahesh Deshpande:112683210

Siddhesh Yadav:112670298

05.19.2020

MS in Computer Science, Fall 2019 Batch.

INTRODUCTION

In this project we will perform visualisation of public health surveillance data with respect to the COVID-19 virus, public health surveillance is the ongoing, systematic collection, analysis, and interpretation of health-related data essential to planning, implementation, and evaluation of public health practice, and at the time of this proposal we feel like visualisation of this data will help us better understand and accept the current pandemic. COVID-19 data visualisation can be used to help public health professionals and health care providers monitor the spread of COVID-19 in the United States and help develop a better understanding of U.S. illness, disease severity, and social disruptions associated with COVID-19. These visualisations help understand the U.S. national public health response to COVID-19.

APPROACH

Technologies we have used include HTML5, JavaScript, D3.js and Flask microweb framework.

We have a dashboard which contains links to navigate between graphs with analysis and statistics. This uses a client-server system: python for processing (server), D3 for VIS (client).

We present a menu to allow users to select a variable and update the charts. The standard graphs drawn include a bar chart that is displayed when a categorical variable is selected, an interactive pie chart and line chart, along with a scatter plot and scatter matrix with brushing. The charts are interactive such that they will have mouse-over display that shows the value of the bar on top of it, the bar will become wider and higher to focus on it on mouse-over and the mouse cursor on moving left will decrease bin width/size (for numerical variables only) and vice-versa.

In the non-standard graphs we included MDS, PCA and parallel coordinate. We implemented random sampling and stratified sampling(using k-means) on the datasets, then performed dimension reduction on both original and the two types of reduced data and provided visualization of both original and the two types of reduced data.

DATA

We collected multiple datasets for the purpose of these visualisations, namely stateCurrent.csv, StateHistorical.csv, USCurrent.csv, USHistorical.csv
The data was sourced from https://covidtracking.com/api. The data collected can be used to monitor spread and intensity of COVID-19 disease in the United States, to understand disease severity and the spectrum of illness, to understand risk factors for severe disease and transmission, to monitor for changes in the virus that causes COVID-19 and to estimate disease burden.

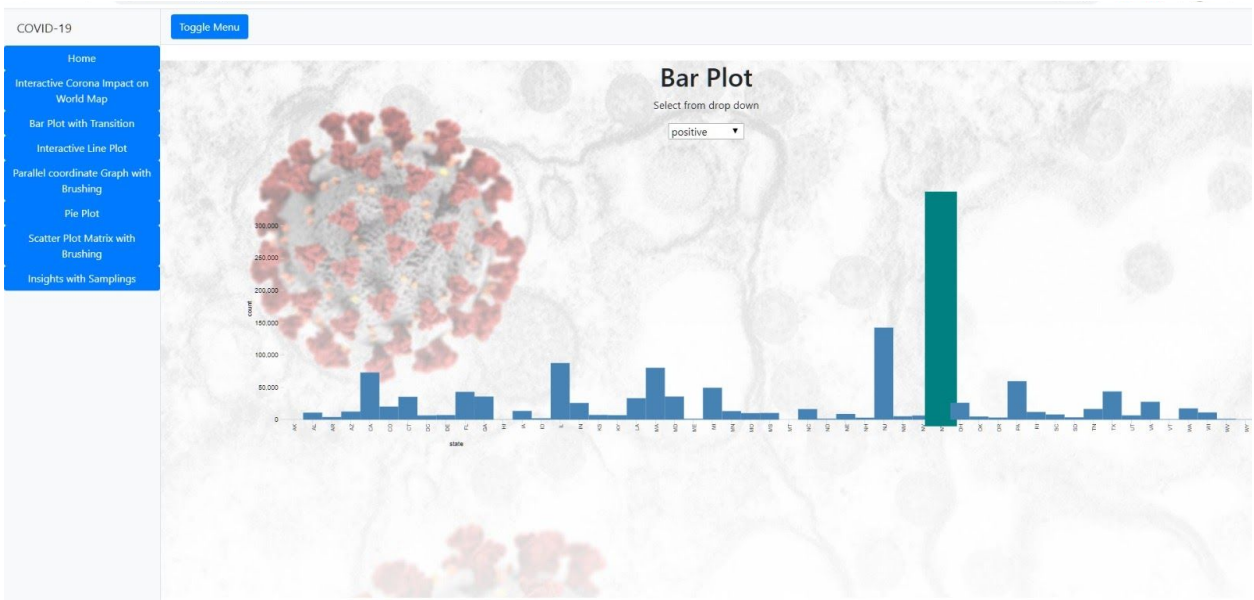
Dataset	Data
stateCurrent.csv	Shows latest data collected and the situation at that point for each state.
StateHistorical.csv	Shows data collected over time for each state.
USCurrent.csv	Shows latest data collected and the situation at that point for the U.S.A
USHistorical.csv	Shows data collected over time for the U.S.A.
John Hopkins University Data Repository.	Contains additional Coordinates for map plot.

VISUALISATION AND ANALYSIS RESULTS

1. Map Plot

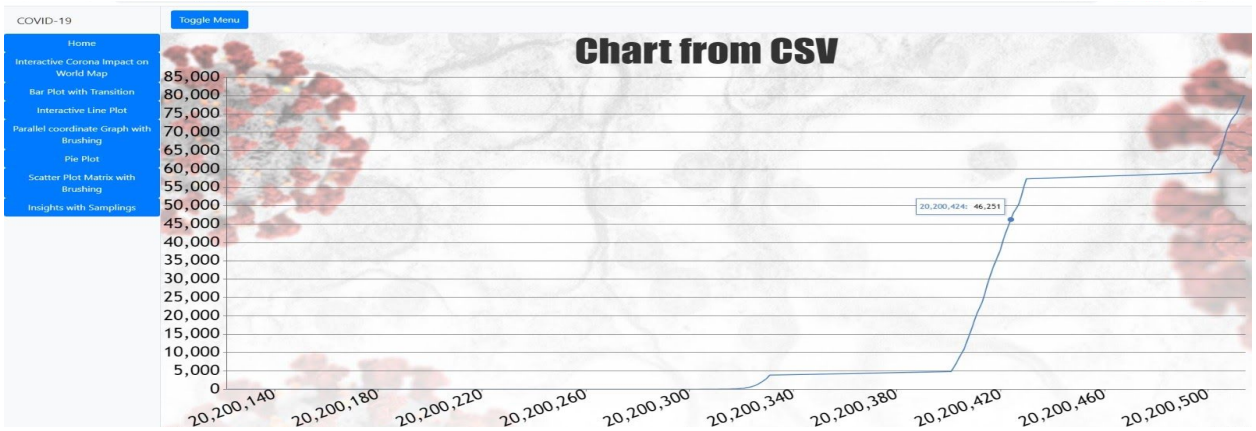
In the above map plot we see the spread of the pandemic throughout the world and how at the time of data collection the United States had the highest concentration of cases.

2. Bar Plot



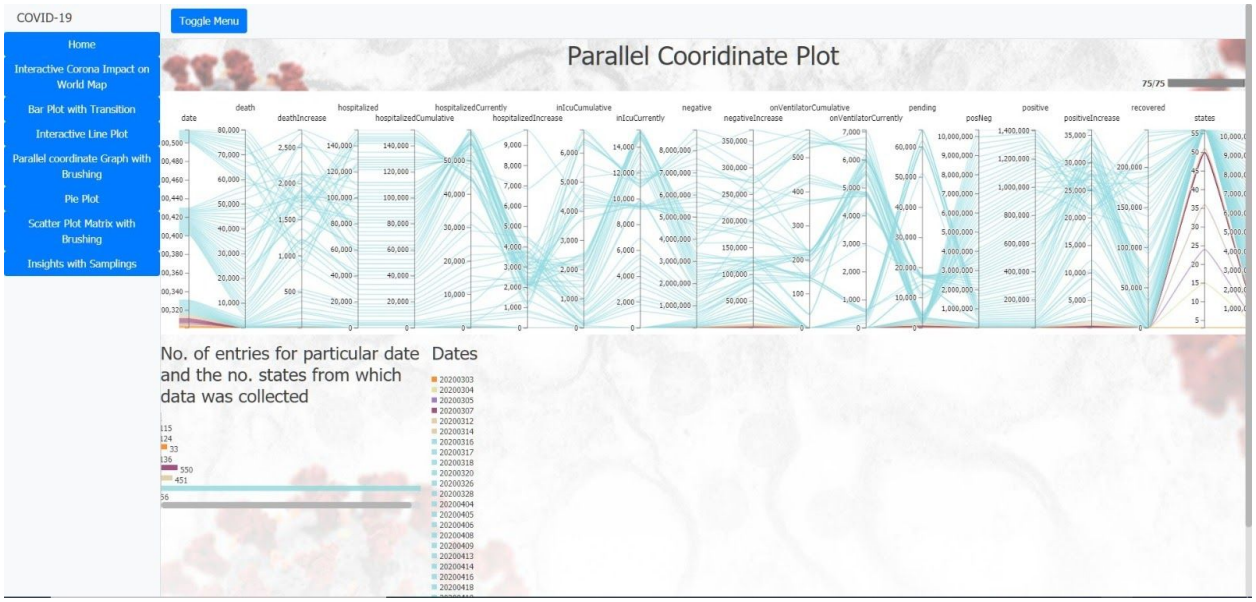
In the above bar graph we see the current situation of the pandemic in individual states of the United States, New York is hit the worst.

3. Line Chart



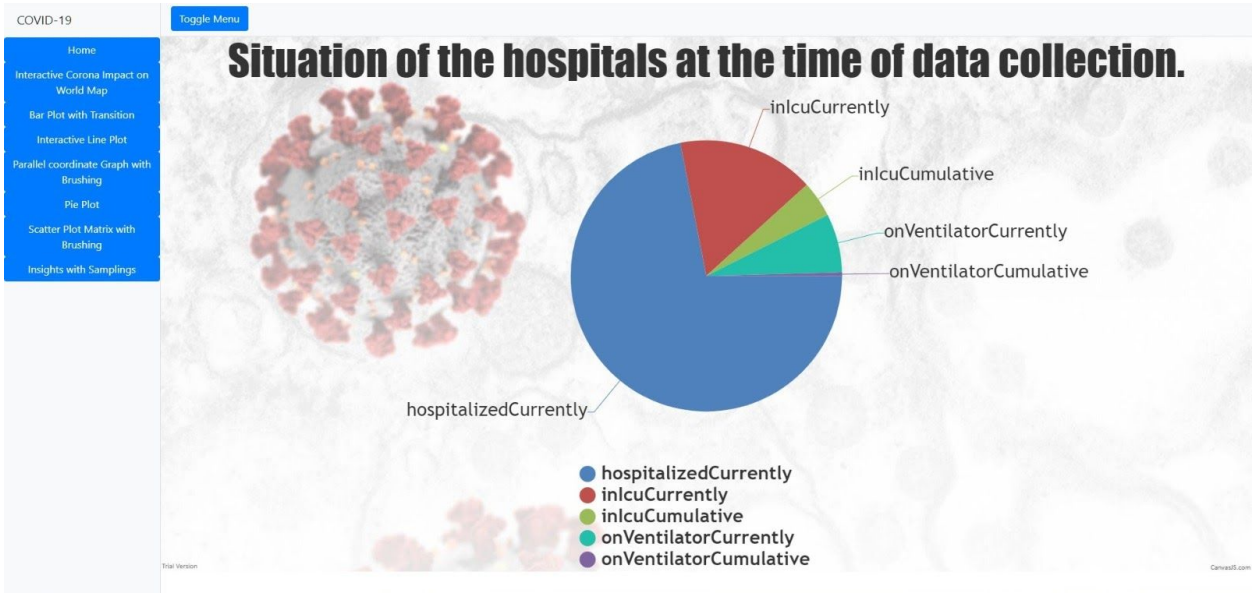
In the above line plot we see the increase in total number of deaths caused by COVID19 across the data collection dates.

4. Parallel Coordinate Plot



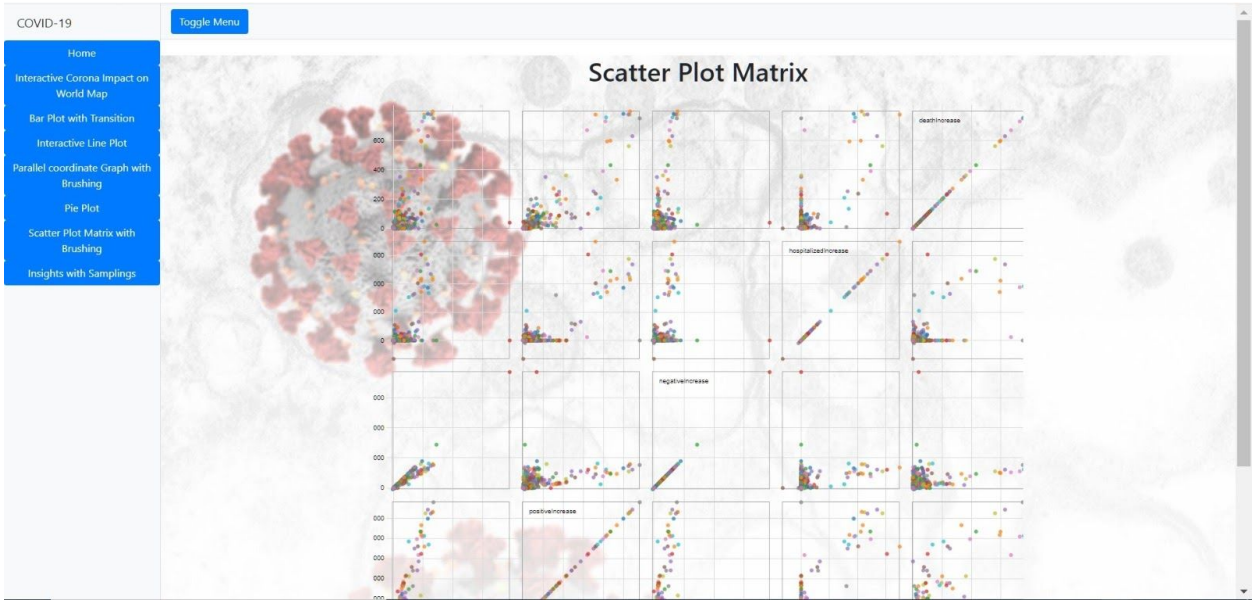
In the above parallel coordinate representation we can see the overall situation of the pandemic, the data can be analysed over various days and we can clearly see that the pandemic kept getting worse till the latest data collection date throughout the United States.

5. Pie Chart



In the above pie chart we see the hospital traffic and resource utilisation throughout the United states.

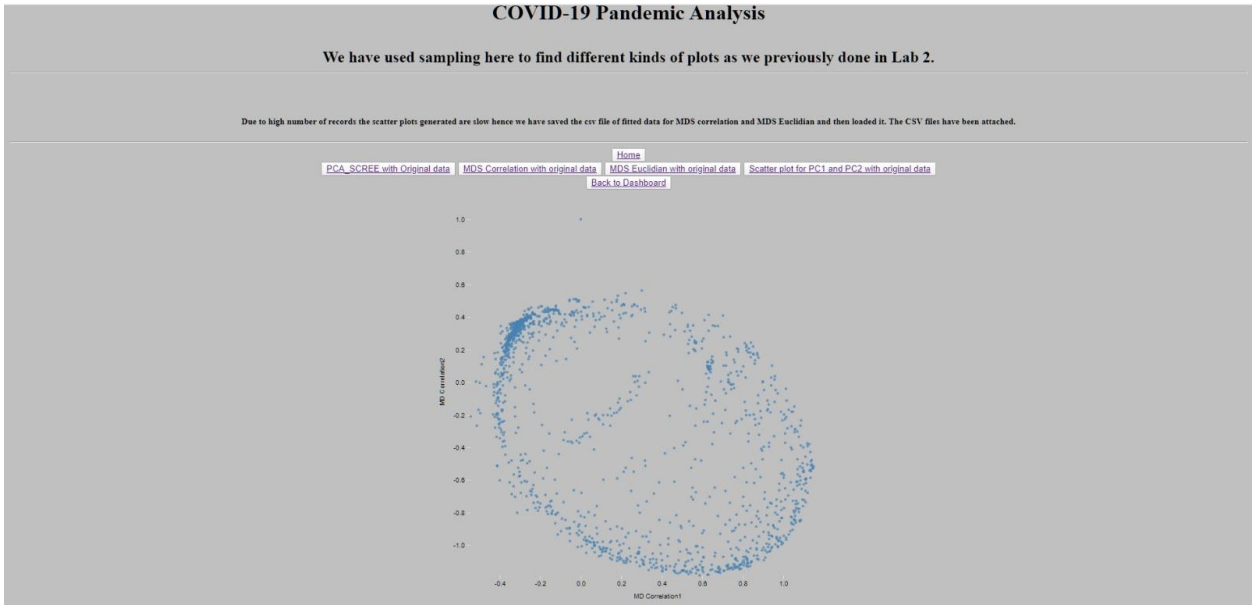
6. Scatter Plot



A scatter plot matrix is a grid (or matrix) of scatter plots used to visualize bivariate relationships between combinations of variables. Each scatter plot in the matrix visualizes the relationship between a pair of variables, allowing many relationships to be

explored in one chart. A scatter plot matrix is composed of a grid of mini-plots and one larger preview plot that shows a selected mini-plot in more detail. The variables in question here are deathsIncrease, hospitalizedIncrease, negativeIncrease, positiveIncrease, totalTestResultsIncrease.

7. Non-standard plots



From the Non-standard plots that we generated, we can find how closely the variables in concern are related to each other. The further away loading of variables known as vectors are away from PC origin, the more influential they are. Even the observation in the MDS graph reveals that further away two points are the more dissimilar they are, and conversely the closer two points are the more similar they are.

YOUTUBE LINK: https://youtu.be/OTNJ2I_rnJQ

CONCLUSION

In conclusion, after constructing interpretable visualisations of our dataset we were able to track the pandemic; the cases, deaths, hospital activity; and clearly understand the current situation which can be used by people in charge of allocation of resources and curbing of the pandemic to optimise their efforts.