INLS641 Fall'22 Visual Analytics

An Online Tool for Visualizing the Impact of COVID-19 on Hospitals

Team Members: Viola Goodacre, Vibhor Gupta

Introduction:

The COVID-19 pandemic has had a devastating effect worldwide, impacting the lives of everyone. There have been almost 96 million cases and over 1 million deaths in the USA alone. It has placed unprecedented strain on the US healthcare infrastructure. Throughout the pandemic, it has been important for health systems, public health experts, researchers, and the government to track the impact of the pandemic on hospitals. Although the pandemic is not yet over, in recent months, with the strains having weakened over time and more people being vaccinated, the case and death rates have declined. COVID-19 is leaving in its wake healthcare systems with reduced staffing and increased workloads that impact the quality of patient care. Many hospitals continue in a crisis, which could worsen if cases increase in the winter of 2022. It is important to continue to monitor healthcare systems to identify emergent problems. In addition, looking at past data enables researchers to study the pandemic's effect on the US health systems over the past three years. Many measures can be used to identify if a hospital is in crisis or failing, including bed utilization, staffing levels, admission rates, readmission rates, and more.

This project gave us an opportunity to propose a dashboard that can quickly identify states that may have healthcare systems in crisis. Given the timeframe available for the completion of the project, the scope was kept limited to only a selected set of possible indicators. A future extension to the project could include the addition of more parameters. In addition, the selected dataset only contains data at the state level. Further drill-down to the county or hospital level would require cross-referencing other datasets. This could be an idea for another extension to this project.

We came up with three personas and highlighted the questions that might be interesting to them, as shown in the figure below:

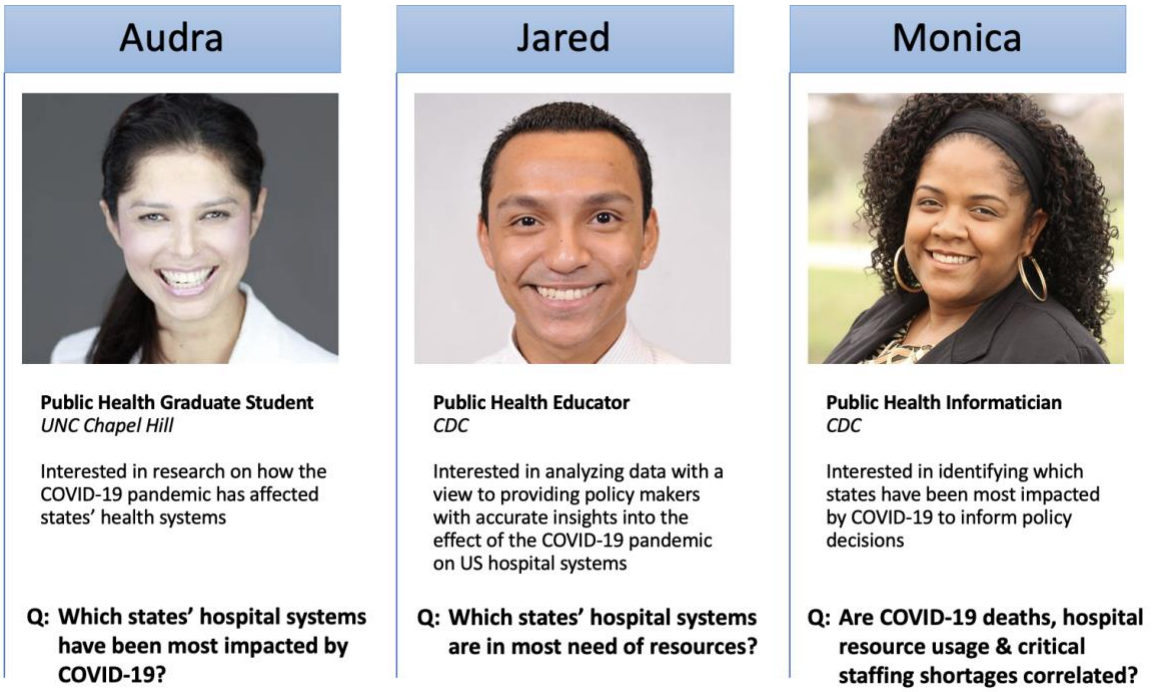

Figure1: Three personas and their question

The personas are from different fields and may be interested to ask the questions highlighted in bold.

Data:

The dataset for the project is openly available at HealthData.gov. It contains state-aggregated time-series data for hospital utilization in the USA. Each record in the dataset contains the aggregated data from all hospitals in a specific state for one day (the day of reporting). The dataset contains data for all 50 states and for American Samoa, Puerto Rico, and the US Virgin Islands. It includes data for January 1, 2020, through to the present day and continues to grow by 53 records per day (one per state/territory). The dataset contains 135 fields, several of which were added periodically since the creation of the dataset. There are currently over 50,000 records. As this project's scope is small, only a small number of fields will be needed. Any extraneous fields will be removed from the dataset. The data can be found on the website or on the concept proposal.

Intended Visualizaion:

We intended to create a timeseries visualization that depicts the trend in a parameter with time. Along with the graph we plan to build a US map, with color code values of the parameter The initial proposal can be found in the screenshot below:

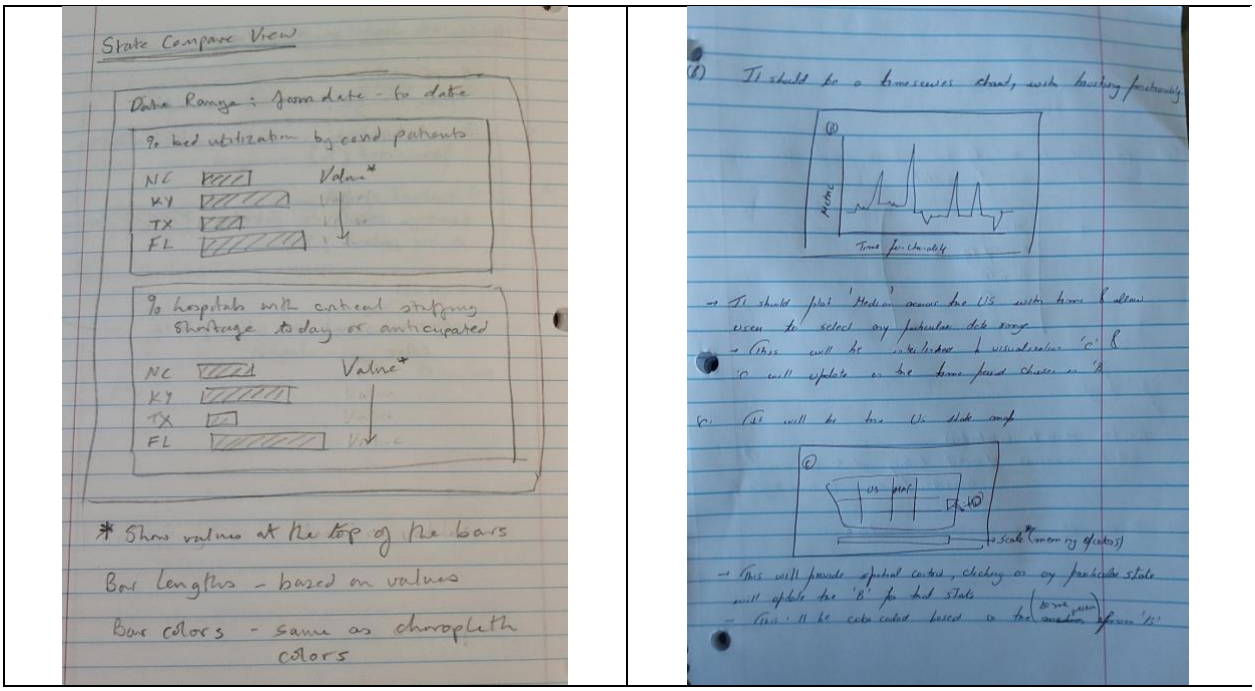

Figure2: Intented Visual Design

Data Transformation and Stastical Analysis

Since the data had more than 50 thousand records that were specific to each state on each day, some of which included null values, it was necessary to process the data. The steps used were to:

- Remove null values from the data.

- Find correlations between the selected attributes.

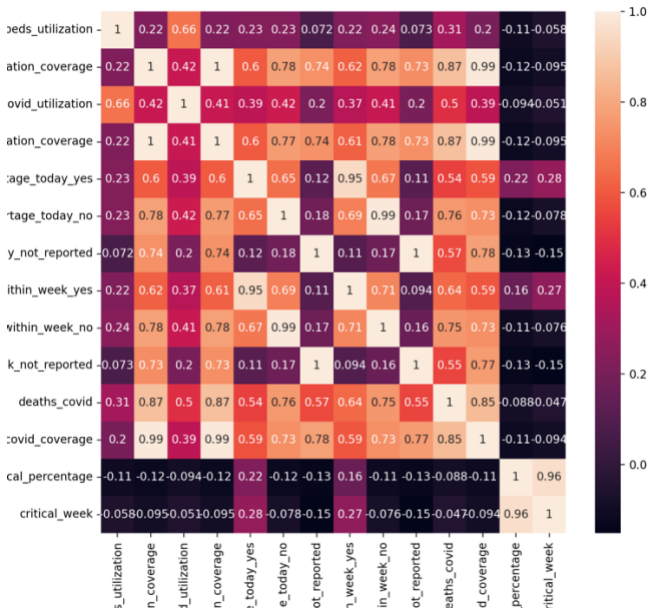

Figure3: Correlation plot for the data

The correlation between death from covid with other parameters were obtained. As it is clear from the plot in Figure 3 not many factors correlated with the deaths from covid. Based on our understanding we chose staff shortage today, staff shortage anticipated within a week, inpatient covid bed utilization as the parameters to plot for the dashboard.

- Calculate median values for the data aggregated by time and by state. This was for plotting purposes.

Intended Visual Designs Developed.

We developed two prototypes for the project. Both the websites have a timeseries chart along with the map, as stated in our concept proposal. We were not able to add the state comparison on the dashboard, due to time constraints and a few challenges, as will be discussed later.

Final Design Implementations

We tried to stick to the original plan as much as possible but had narrowed the scope due to the aforementioned challenges. We worked separately on different parts of the visualization, then attempted to

integrate them towards the end of the project. The final design can be seen in the following screenshots:

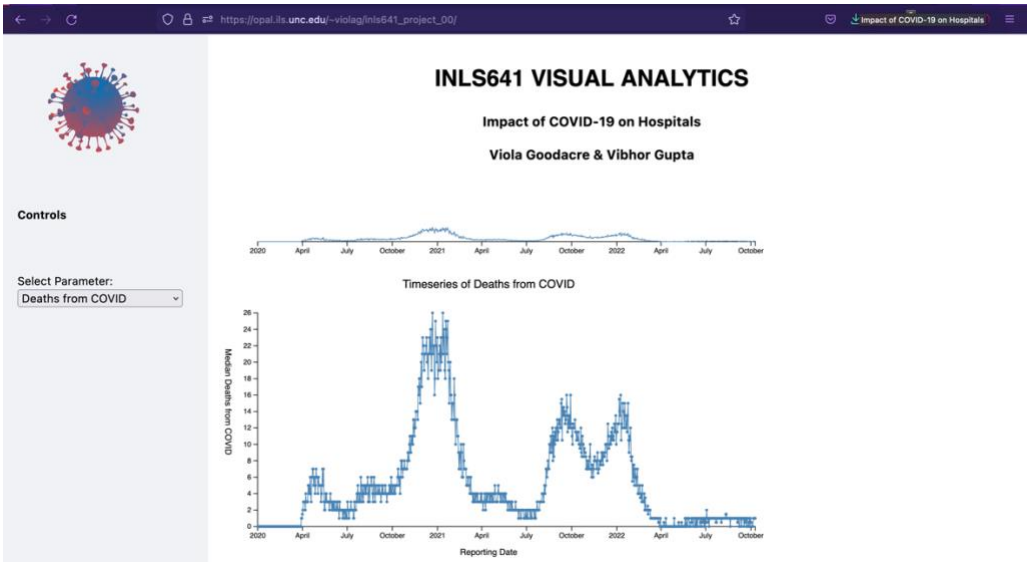

Figure4: Website developed using D3 and python

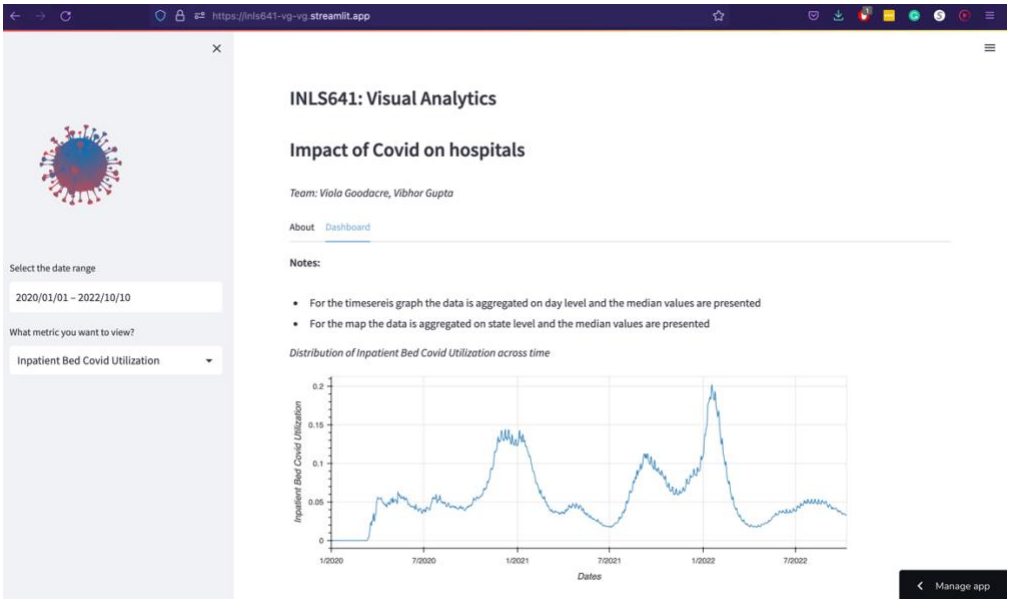

Figure5: Website developed in Python

The main challenge of this project was that one team member was coding in JavaScript and d3, while the other was coding in python. Integration was problematic and the problem was not fully resolved. So we came up

with two prototypes as shown if the figure above. Both the websites look same just has minor differences in them.

The data visualization reveals few findings, which are thoughts to ponder upon. Example is:

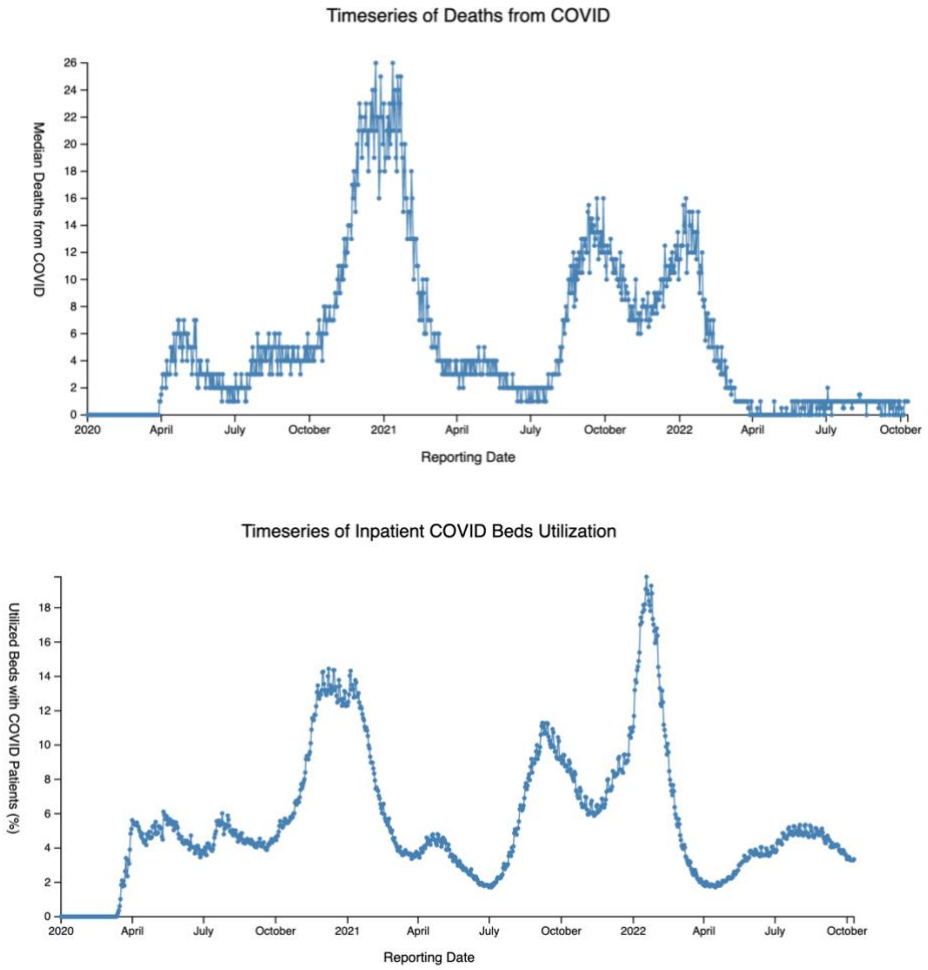

Figure6: Comparison between Covid Death and Inpatient Bed utilization

As you can see, the peaks occur on a similar timeframe, this hypothesis why are deaths from covid increases when people are admitted in hospital. There are different speculations that came to our mind, while analyzing the data. This highlights the power of data visualization.

Overall, the project was a good learning experience. However, it was difficult due to persistent health problems for one team member that lasted longer than a month. Even so, with better planning and collaboration between team members the project may have been better.

Ethical and Societal Considerations

The project aimed to provide a holistic view of the impact of COVID-19 on hospitals. Data used was already collected, depersonalized, aggregated, and made publicly available. We did not identify any potential negative ethical or societal impacts of this project.

Links:

- Link to the D3 website for the project
- Link to python based website