# COMPSCI X465.1 – Data Visualization Final Project – Project Report

## Discuss the objective, data, methodology, and conclusion of your project.

## What did you wish to achieve with this data presented?

The novel coronavirus, also known as SARS-CoV-2, is a contagious respiratory virus that first reported in Wuhan, China. On 2/11/2020, the World Health Organization designated the name COVID-19 for the disease caused by the novel coronavirus. This notebook aims at exploring COVID-19 data for US

Methodology:

1. Why did you choose this layout, color, and metrics?

The layout has been chosen taken into consideration the target audience of the report. A combination of pie-charts, bar-graphs, choropleths, and line graphs have been chosen for conveying the information. The colors have been selected in a way to make the visualization appealing to the users. For choosing the labels and axes, care has been taken to avoid visual clutter while at the same time highlight important information like “Total Number of Cases”, “Number of Deaths” etc.

1. What metrics, quantitative or qualitative models, or analysis was chosen?

The following metrices have been chosen:

* The cumulative distribution function for the total number of confirmed cases
* The cumulative distribution function for the total number of deaths
* The ratio of confirmed cases to deaths which signifies the fatality index
* Top 10 states with the maximum number of cases
* Top 10 states with maximum number of deaths
* Categorize the states of US according to the total number of confirmed cases
* Time series analysis of the monthly average of the cases to identify presence of waves
* More detailed view of the second wave of the pandemic
* Analysis of the state of Oregon data
* Number of people per 100 vaccinated in US as a whole and Oregon in particular.

1. Who was your audience; why did you choose this audience?

With more and more people getting vaccinated, the number of Covid positive cases has come down in the recent months. As people are getting ready to get back to normalcy, more and more states are considering lifting off precautionary measures taken to slow down the progress of the disease, like limiting social gathering and wearing masks in public places etc. The analysis report has been aimed at raising a general awareness among the people on the state and spread of the disease.

Data:

1. How was the data prepared and cleansed?

There has been lots of Covid 19 related information available for analysis. The dataset has been chosen keeping in mind, the authenticity and completeness of information. The following data sources has been used:

<https://github.com/nytimes/covid-19-data/blob/master/us-states.csv>

<https://github.com/nytimes/covid-19-data/blob/master/us-counties.csv>

<https://ourworldindata.org/us-states-vaccinations>

and <https://covid.ourworldindata.org/data/vaccinations/us_state_vaccinations.csv>

Conclusion:

1. Compare your feedback on the data you analyzed versus having to do it yourself.

A huge volume of structured and unstructured data has been available on Covid19. But it becomes really hard for the users to use the information in effective way without using any scientific tool for data analysis and visualization. The current project aims at analyzing the data and presenting relevant information in a user-friendly way by using appropriate data visualization tools and methods.

1. What did you notice?

Some of the graphs could have been overlayed to gain more powerful insights into the data. Providing a one-page summary with the observation would have made the analysis more effective. Apart from vaccinations, other factors like usage of masks, social distancing etc. could be analyzed to get a more comprehensive understanding of the factors impacting the spread of the disease. Analysis of the demographics of the confirmed cases would create awareness about who is more susceptible to the disease.

1. What would you have done differently?

Use more overlay graphs. Make the graphs more interactive. Create a dashboard with all the information. Create a pipeline for regularly updating the data so as to keep the information up-to date.

1. Does this change your opinion of step 1? Why or why not?

Step 1 helped me to identify the basic information which could be used for analysis given the timeframe of the project. Where as the questions 7,8 helped me take a deeper look at the work and identify the areas of improvement so as to make the visualizations more effective.