Project Report: STAT 430

Shantanu Solanki

Introduction

The first step to solve a problem is to understand it. And an efficient way to understand a problem is using visual representation of its aspects. This project deals with the visual representation of data regarding COVID-19 in the different states and counties. I have build an interactive web tool using Dash to view the different features. The following features have been presented:

- US map showing the percentage of vaccinated population (partially and fully) in different states.
- Map of different states with the percentage of vaccinated (partially and fully) in different counties.
- A time-series plot of 7-day moving average of percent COVID-19 positivity rate for different counties on different selected date ranges.
- A time-series plot 7-day moving average of new COVID-19 cases per 100K for different counties on different selected date ranges.

Dataset Description

The data required for this project has been downloaded from the official website of Centre for Disease Control and Prevention. There were three datasets required for the visualisation. these are mentioned as follows:

COVID-19 Vaccination in US - State

This dataset has the data regarding the number and proportion of population in different states with partial and full vaccination. The dataset has a total of 80 columns. However, here I have used the following 4 columns to plot the Choropleth map:

- `Date` : Date of updation of the dataset (Datatype: String)
- `Location` : State concerned entered as their codes, eg. IL, CA, etc. (Datatype: String)
- `Administered_Dose1_Pop_Pct` : The population percent of the state with first dose of one of the several vaccines (Datatype: Numeric)
- `Series_Complete_Pop_Pct` : The population percent of the state with full vaccination with one of the several vaccines (Datatype: Numeric)

COVID-19 Vaccination in US - County

This dataset has the information about vaccination of the population at the county level. It has similar columns as the dataset above. However, this data was visualized on the map using the column 'FIPS' which stands for Federal Information Processing Standards, which is used to indentify a particular geographical location. In all, the following features were worked upon:

- `Date` : Date of updation of the dataset (Datatype: String).
- `Recip_County` : Names of the counties (Datatype: String).
- `FIPS` : Geographical codes of the counties (Datatype: String).
- `Administered_Dose1_Pop_Pct` : The population percent of the county with first dose of one of the several vaccines (Datatype: Numeric).
- `Series_Complete_Pop_Pct` : The population percent of the county with full vaccination with one of the several vaccines (Datatype: Numeric)

COVID-19 Community Transmission - County

This dataset has columns about the transmission of COVID-19 among several counties. The dataset has a total of seven columns. This was used to plot 7-day moving average positivity rate and no. of new cases in different counties. The following columns were used to make the time-series plots:

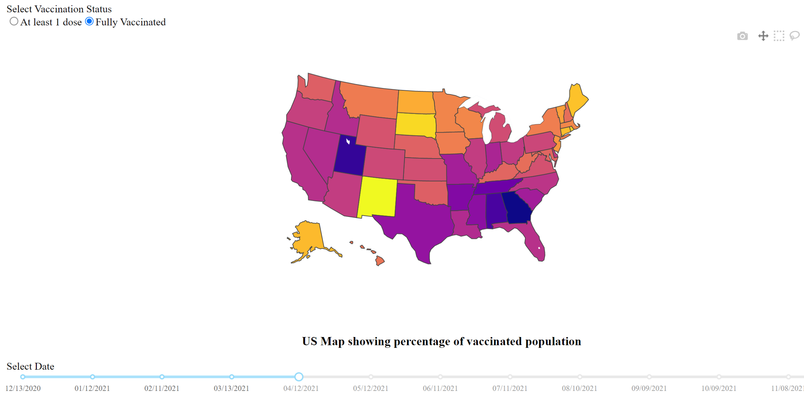
- `report_date` : Date of updation of the record (Datatype: String).
- `state_name` : Name of state (Datatype: String).
- `county_name` : Name of county(Datatype: String).
- `cases_per_100K_7_day_count_change` : 7-day moving avergae of the no. of new cases per 100k people (Datatype: Numeric)
- `percent_test_results_reported_positive_last_7_days` : 7-day moving avergae of the percent positivity rate on the COVID-19 tests (Datatype: String later converted to numeric for plotting)

Objective and outcomes

Using the above three data sets, I made four dynamic visual representations of the data (2 maps and 2 time-series plots). The description fo the maps are as follows:

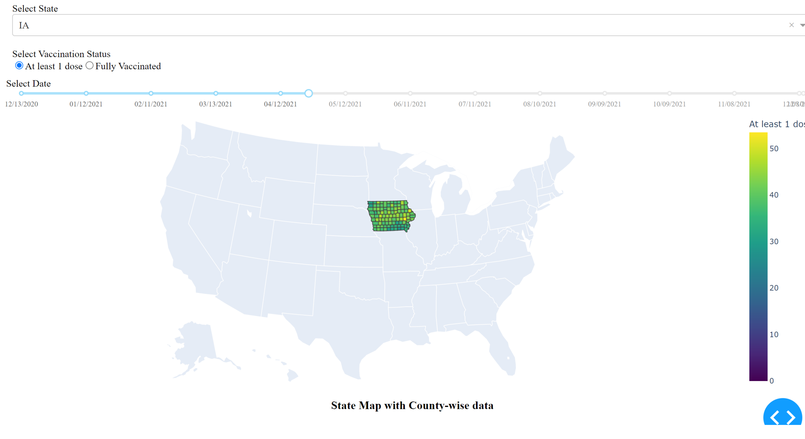
Map 1:

This is a choropleth map of the US with two inputs: Vaccination Status (Partial of Full) and Date. Here is a demo:



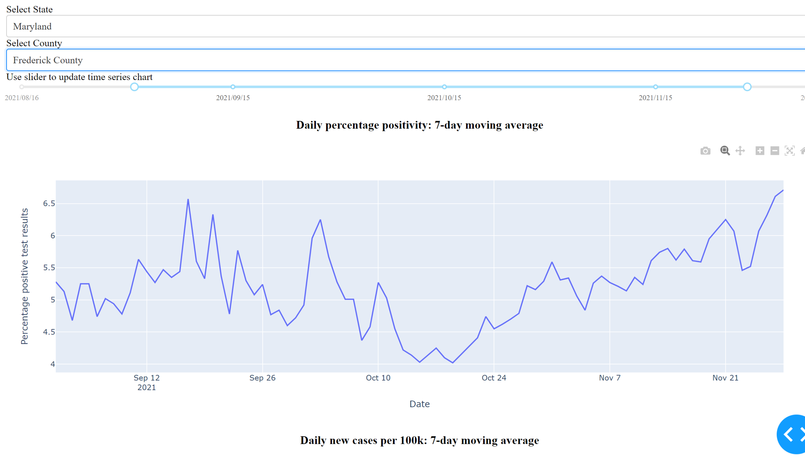
Map 2:

This is a choropleth map of the US states with 3 inputs: State Code, Vaccination Status (Partial of Full) and Date. The date input slider refreshes upon selecting the state (chained callback). Here is a demo:



Graphs 3 and 4:

These are two time-series charts with 3 common inputs: State name, County name and Date range. The county name select options depends on the state selected through the State Dropdown menu (chained callback).The graphs 3 and 4, respectively plot the 7-day moving average values of percent positivity rate and new cases per 100k people. Here is a demo:



Conclusion:

The above demonstrated figures provide a decent understanding of the COVID-19 spread and the vaccination drive. These can be used by anyone to understand the level of transmission of COVID-19 in different counties as well as can be used to educate common public by showing the correlation between the level of vaccination and positivity rate. Though the values are not direct representation of what is happening, the two values, total no. of cases and percentage positivity, together, can give us a broad understanding about which regions are more affected and thus, can help the officials in their course of action.

One of the aspects I could improve upon is the speed of updation. Since the datasets are pretty big, the Map 2 takes some time to update. I would process the data to make it quickly retrievable so that it reduces the updation time.