

DATA VISUALIZATION USING TABLEAU

ASSIGNMENT 5

November 6, 2022.

INDIANA UNIVERSITY
(BLOOMINGTON)

Authored by: VARSHA R
IU ID: 2000751388
Sem: Fall 2022



Introduction

Data visualization is achieved using Tableau, a software package focusing on business intelligence (BI). The word tableau refers to a graphic representation or description. As a novice with Tableau, I have experimented with the software and built a few visualizations. I have tried to implement some of my learnings from this tableau course. The purpose of this report is to discuss these visualizations. For each graph, the following information will be discussed:

- Plots built using the dataset.
- Plot analysis and inferences.

Dataset

For the visualizations, we will use the following given dataset that contains around a year's worth of medical related data. The dataset explores state-level data on a variety of COVID-19 metrics, including the latest hotspots, tests, hospitalizations, and deaths.

Name: The State Dataset on the COVID-19 pandemic for the US

Let's discuss a few attributes from the dataset that we will be using for the visualizations.

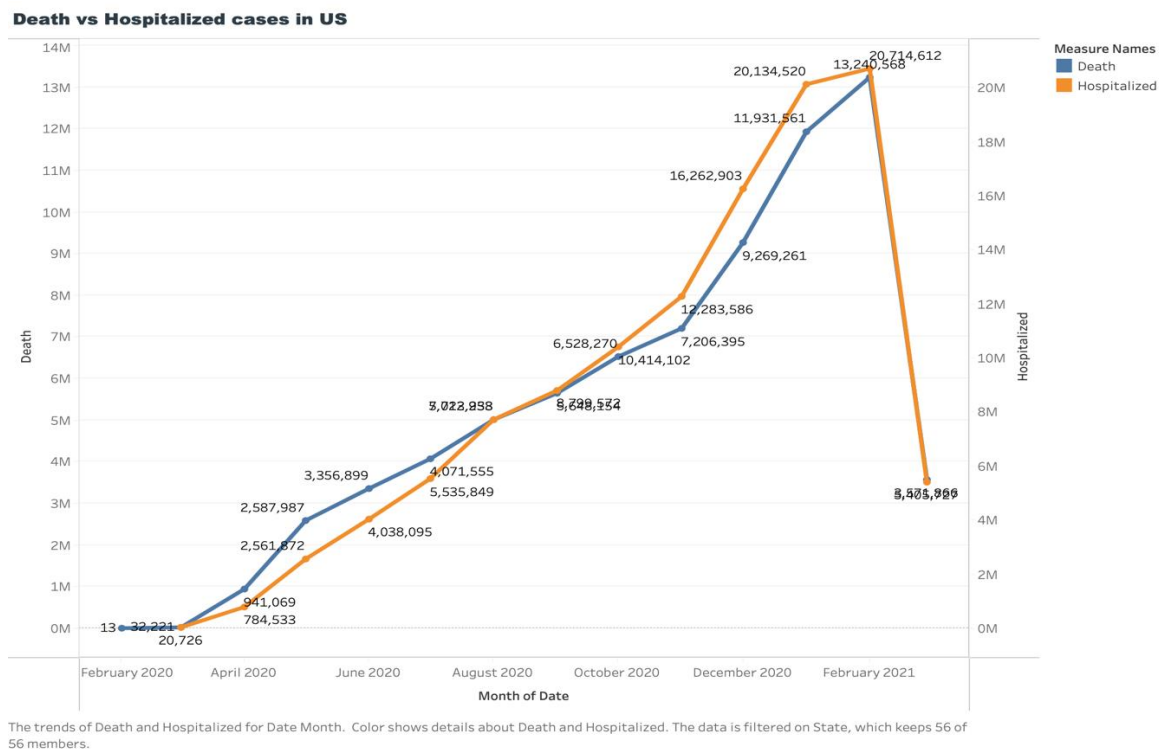
Date	– Holds the dates of when the data was collected
State	– Location of the data collected
Death	– Number of deaths
Hospitalized	– Number of individuals hospitalized
Negative	– Number of test results that turned out to be negative
Positive	– Number of test results that turned out to be positive
Total test results	– Total number of test results recorded

Visualizations

Sheets:

1) Do the time series analysis with Tableau's time functions for death and hospitalized cases in US.

- Plot built using this dataset



- Plot analysis and inferences

Attributes used for this plot: Date, Death, Death confirmed, Hospitalized, State

Columns: MONTH(Date)

Rows: SUM(Death), SUM(Hospitalized)

Marks: Color – Measure names, Details – Death confirmed

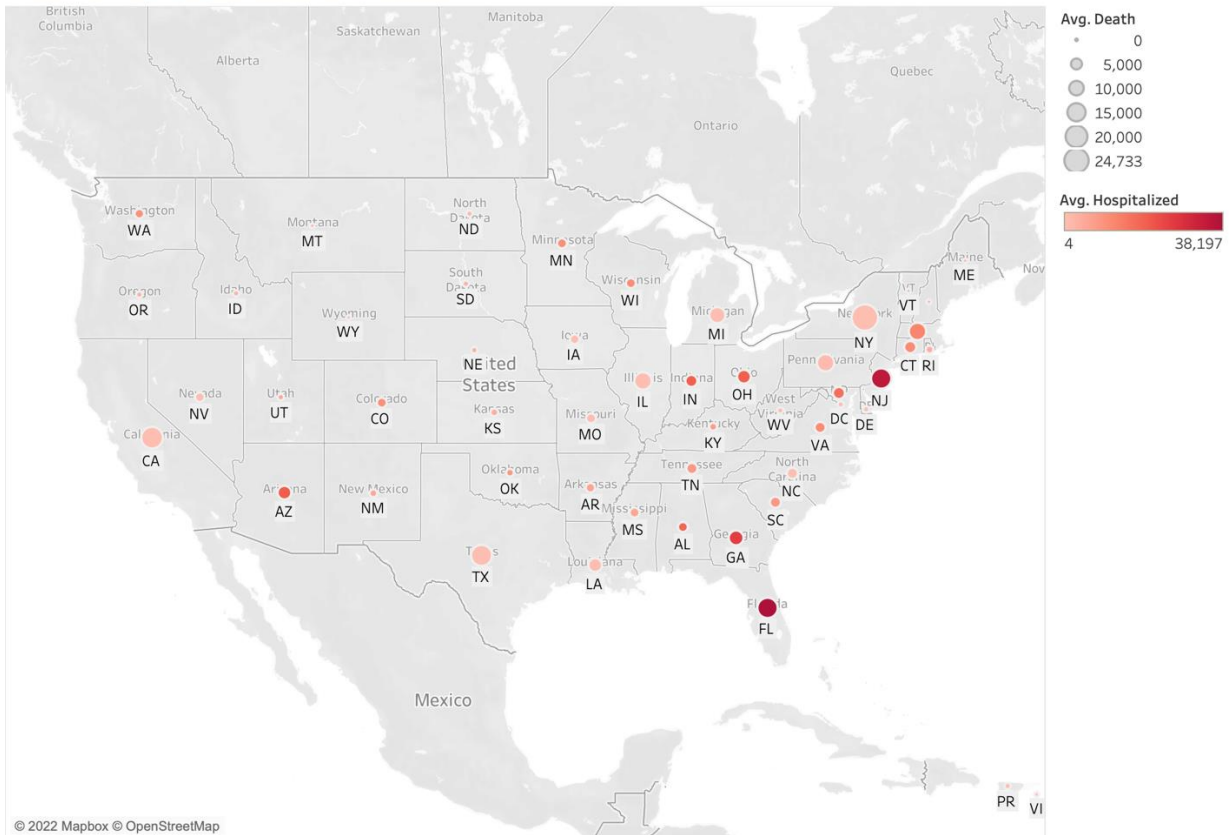
Filter: State

- Here we are trying to compare and see the number of people hospitalized vs number of deaths over the whole year in a time series format.
- As you can see, the hospitalized cases and death cases both started steeply increasing during November 2020 and peaked during February 2021.

2) Create a geographical map and analyze death and hospitalized cases for all states in US.

○ Plot built using this dataset

Death vs Hospitalized cases for all states in US



○ Plot analysis and inferences

Attributes used for this plot: Death, Hospitalized, State

Columns: Longitude

Rows: Latitude

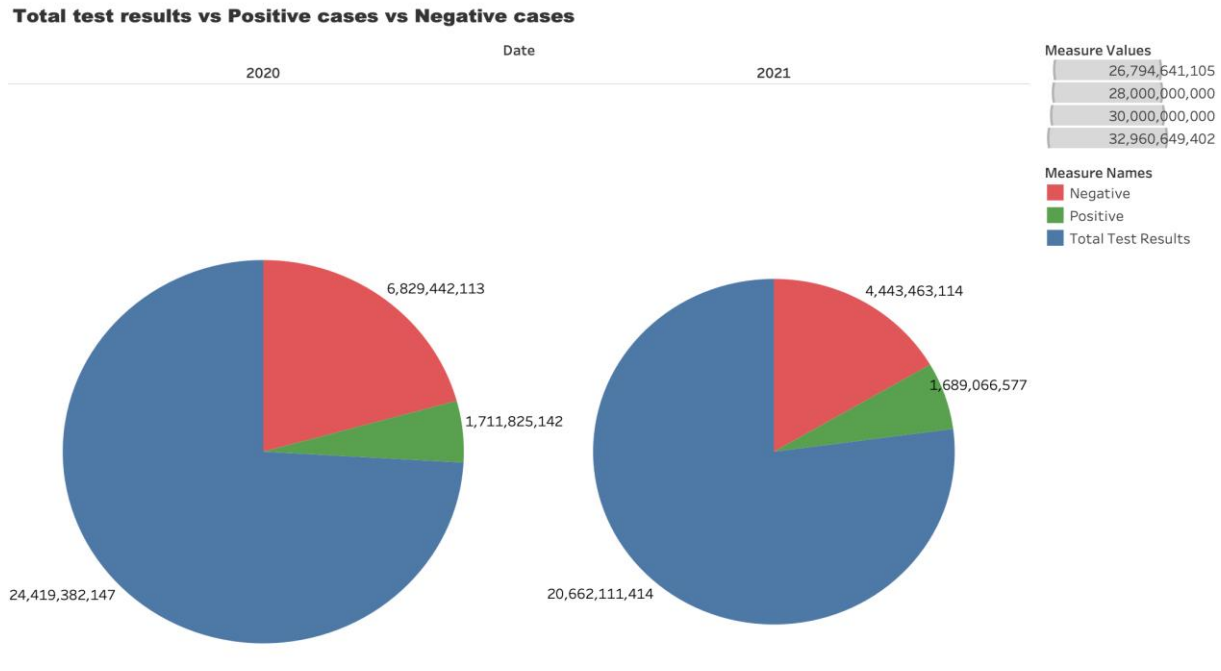
Marks: Color – AVG(Hospitalized), Details – State, Label – State, Size – AVG(Death)

Filter: State

- Here we are trying to compare and see the number of people hospitalized vs number of deaths over different states.
- We notice from the visualization that the size of the circle is largest for New York state indicating that the number of deaths was highest in that state even though not a lot of people were hospitalized.
- We also notice that Florida is colorized the most indicating the highest number of hospitalized individuals.

3) Create doughnut chart (pie charts) to analyze the 'Total Test Results', 'positive' and 'negative' cases in US.

○ Plot built using this dataset



Negative, Positive and Total Test Results broken down by Date Year. Color shows details about Negative, Positive and Total Test Results. Size shows Negative, Positive and Total Test Results. The marks are labeled by Negative, Positive and Total Test Results. Details are shown for Negative, Positive and Total Test Results. The data is filtered on State, which keeps 56 of 56 members.

○ Plot analysis and inferences

Attributes used for this plot: Date, State, Negative, Positive, Total test results

Columns: YEAR(Date)

Marks: Color – Measure names, Details – Measure names, Label – Measure values, Size – Measure values

Filter: State, Measure names

Measure values: SUM(Negative), SUM(Positive), SUM(Total test results)

- Here we are trying to compare and see the ratios between total test results, positive test results and negative test results.
- We notice that the number of negative test cases are way higher than positive cases.
- We also can see that the total number of test cases tremendously surpass the combined value of negative and positive cases which is obviously weird. This might be because a lot of results must be of other type of cases too.

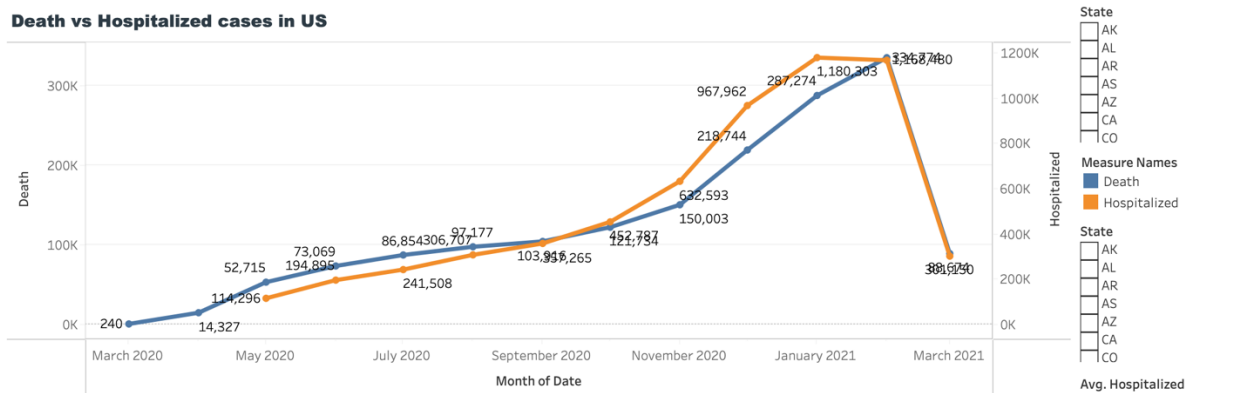
Dashboard:

Dashboards are a combination on one or more worksheets. Multiple worksheets can be merged into a single dashboard. Tableau worksheets are the primary building blocks for the dashboard. In this assignment, there is 1 dashboard created after grouping the three sheet based on their similarity of focus attribute which in this case will be deriving the covid cases stats for Indiana state.

○ Dashboard view

Medical history of all the states (USA)

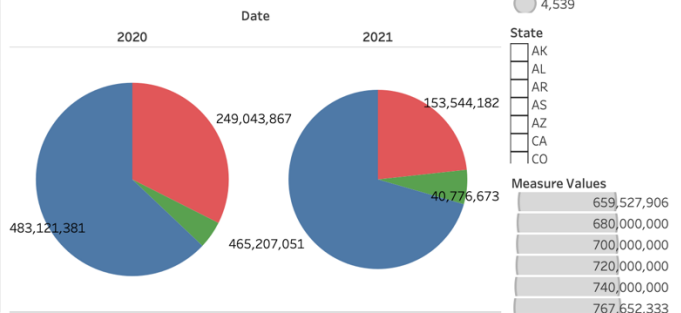
Death vs Hospitalized cases in US



Death vs Hospitalized cases for all states in US



Total test results vs Positive cases vs Negative cases



○ Analysis and inferences

Filtering attribute: Indiana state

- The average number of hospitalized - 19,467
- The average number of deaths – 4539
- Peak season with maximum number of cases – February 2021
- Total number of tests recorded in 2020 is higher than that of 2021
- Negative test cases decreased during 2021 which is a concerning factor
- Indiana had comparatively a considerable amount of cases to deal with during the pandemic.