1

**Department of Information Technology**

**IT 7113 Data Visualization Development Project**

**Project Title**

United States COVID-19 Cases and Deaths by State over Time  
  
**Team Members**  
Chandana Avadhani | Sudarshan Badireddi | Durga Narayana Varma Addepalli



**Executive Summary:**

Before the 17th century, data visualization was mostly limited to maps, which showcased land markers, cities, roads, and resources. As the demand grew for more accurate mapping and physical measurement, better visualizations were needed. Theme visualization first appeared in the 18th century. Near the end of the 21st century, attempts were made at the thematic mapping of geologic, economic, and medical data. At this time, abstract graphs of functions, measurement error, and empirical data collection were introduced.

The latter half of the 20th century is what friendly calls the ‘rebirth of data visualization’ with the emergence of computer processing. Computers made it all possible to process large amounts of data at lightning-fast speeds. Over the last three decades, the field of data visualization has exploded into dozens, if not hundreds, of focus areas. Dashboards, data discovery tools, and a variety of other software tools allow businesses, researchers, and individuals to explore their data in new and increasingly creative ways.

In this project, we will showcase a variety of data visualizations in multiple dashboards to get meaningful insights from Georgia Schools’ data.

**Background:**

The graphical display of information and data is known as data visualization. Data visualization tools, which employ visual components such as charts, graphs, and maps, make it simple to identify and comprehend trends, outliers, and patterns in data. In the Big Data environment, data visualization tools and technologies are crucial for analyzing enormous volumes of data and making data-driven decisions. The primary purpose of data visualization is to clearly and effectively present information using graphical means. If this is accomplished, the user will be able to get the primary insights from the separate visualizations. In the corporate setting, it will assist stakeholders and business owners in making sound decisions that benefit the company.

Similarly, we intend to display the Covid-19 Cases and Deaths by State across the United States over time. The goal is to display Covid-19 data with many factors such as total deaths/cases, confirmed cases/deaths, new cases/deaths, and many more using various visualizations and Dashboards. The other goal is to provide a comprehensive picture of Covid deaths and cases data to the general public so that they can derive valuable information from it.

**Data Description:**

Regarding this advancement, the data source that we have selected is associated with the Covid-19 database. The following is a link to the original source of the data: https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36

The data sources that are shown in the links above are in a format known as CSV. After opening Excel to begin working on the file, we saved it as a workbook in Excel. The daily numbers of confirmed and probable cases, as well as deaths, that were retrieved by the CDC from states and territories over the course of time are included in this archived aggregate dataset. Adjustments may be necessary because these provisional counts are susceptible to change, which may include revisions to data that was previously provided. Because these modifications can lead to a lower total number of cases and deaths when compared with the data from before, the newly calculated numbers of cases or deaths may include negative values to account for the effects of these adjustments.

The formatting of the data was the most significant step that we took in the completion of this project. When creating the dashboard, only the columns that are necessary for its function will be examined. According to the requirements of the dashboard, a few of the news columns have also been constructed. Excel is the format of the data that is sent into the Tableau program for the purposes of data visualization.

**Dashboard Design:**

A dashboard is a visual representation of the most relevant information required to achieve one or more objectives; it is condensed and presented on a single screen so that the information can be monitored at one time.

We designed four dashboards specifically for this project. All of the dashboards provide useful information. Each Dashboard has different visuals, and the end-user or audience can obtain relevant and important information from each visualization. The primary goal of dashboard design is to present data in a meaningful way in order to bring more clarity and reduce complexity.

The steps for creating a dashboard are as follows:

* In the case of dashboard creation, the first and most important phase is requirements collection. We must ensure that all requirements are appropriately gathered and without errors.
* The problem statement must be expressed clearly after gathering and assessing the requirements. It is impossible to start with dashboard construction without proper clarification on this.
* Once the preceding procedures have been performed, the main and critical phases are data analysis and data purification. All of this is done to ensure that data is cleaned by deleting or eliminating unnecessary rows and columns.
* Align and separate data that is solely needed for dashboard creation. Create appropriate visualizations or select charts for the dashboard.
* Finally, design a dashboard using best practices.
* When constructing a dashboard, the bare minimum of filters, colors, and labels must be followed and considered.

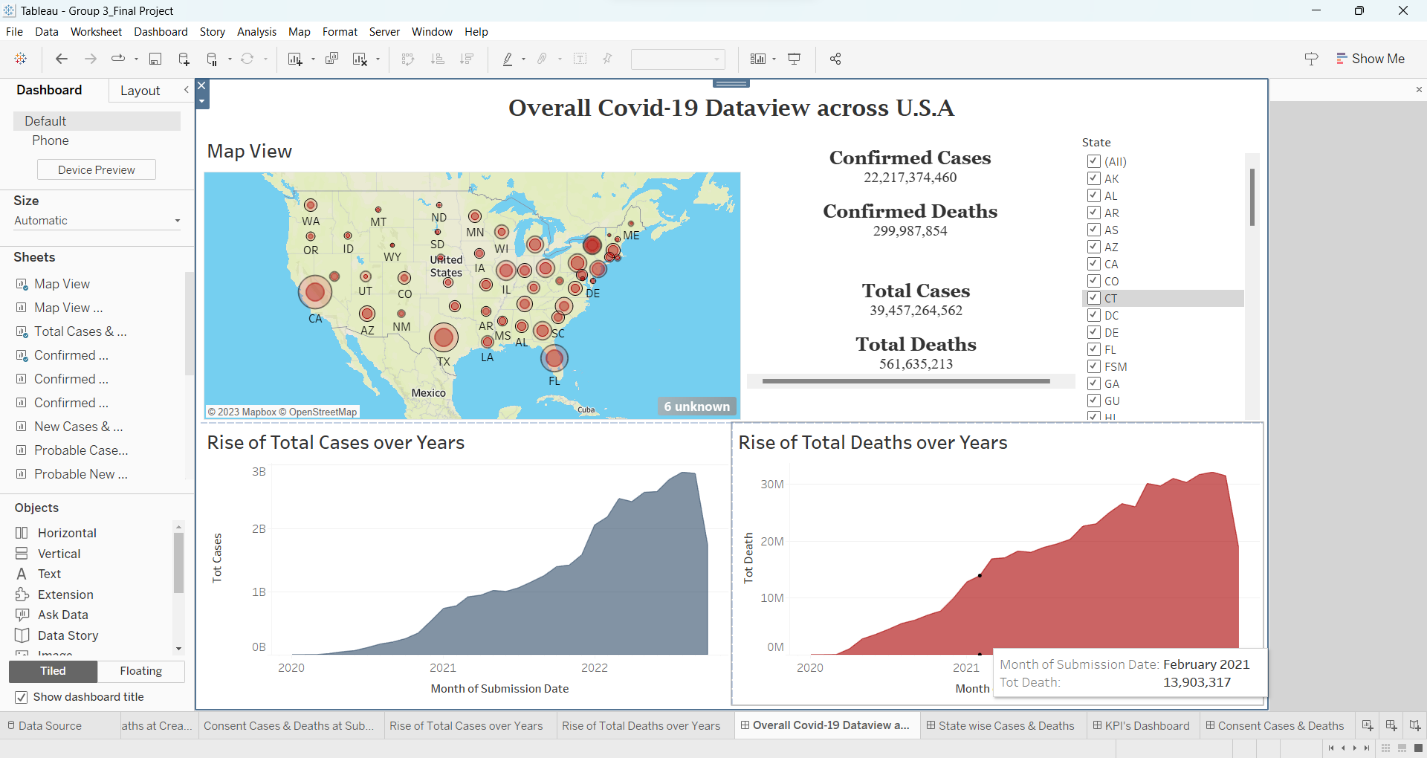
**Best Practices:**

A dashboard is successful when users can quickly find answers using it. Even if you have a beautiful dashboard with an interesting data source, it will be meaningless if your audience cannot use it to discover insights. This can be improved by adhering to best practices. In designing our dashboard, our team adhered to the practices listed below.

* The first and most important step is predicting the user's wants. To be more specific, what are his objectives?
* The next step is to select appropriate visualizations that will make a difference if they are included in the dashboard. Displaying too much data with too many visualizations makes it difficult to glean insights or data.
* We must select the appropriate metrics, and only those should be displayed in the dashboard or individual visualizations. So, in terms of serving user needs, the fewer is better.
* If correct interaction characteristics were employed when developing a dashboard, the major interactivity for the user with the dashboard would occur. As a result, when building a dashboard, we must ensure that the majority of the interactive elements are utilized.
* Finally, the entire dashboard must be displayed on a single screen with a good and meaningful dashboard name.

We can ensure that consumers can easily find vital information by applying excellent dashboard design practices.

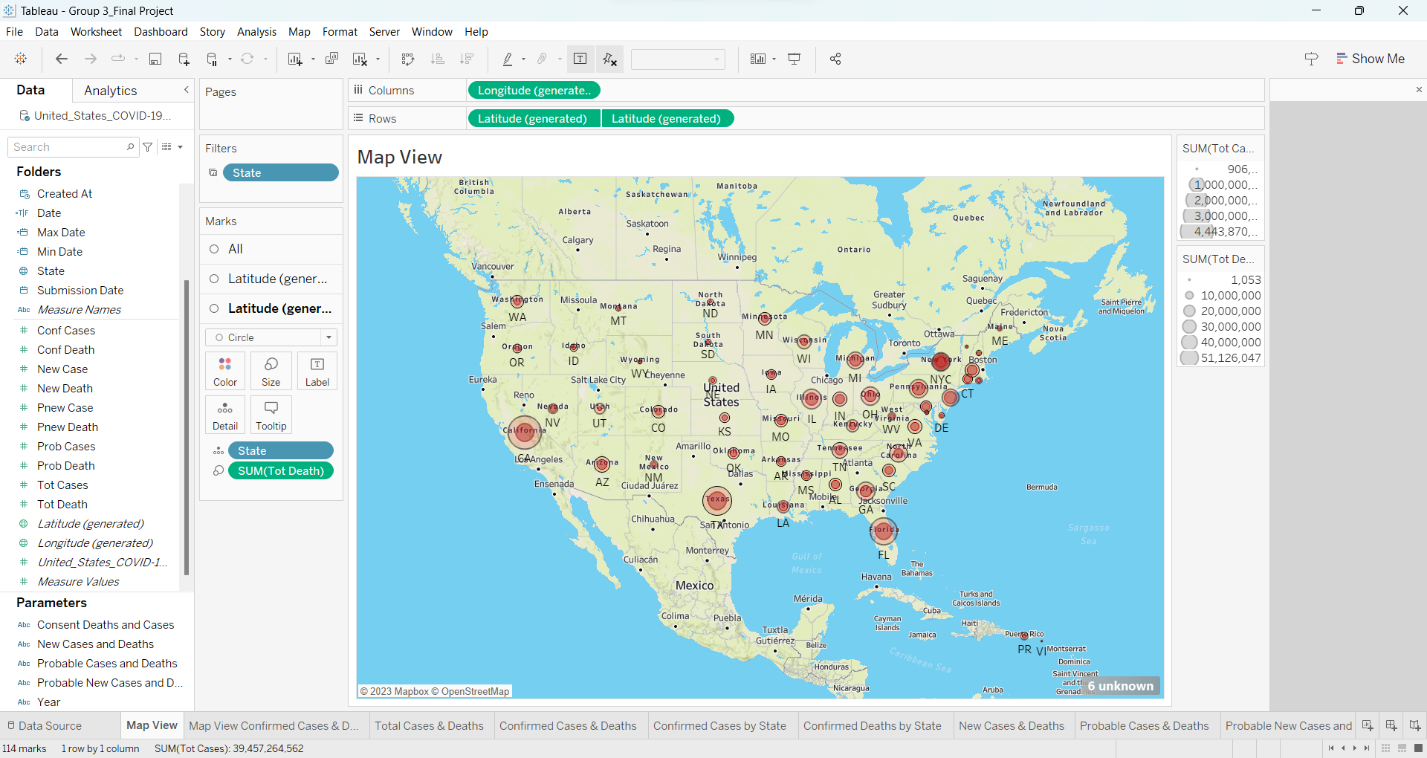
**Dashboard-1:**

 **Overall Covid-19 Dataview across U.S.A**

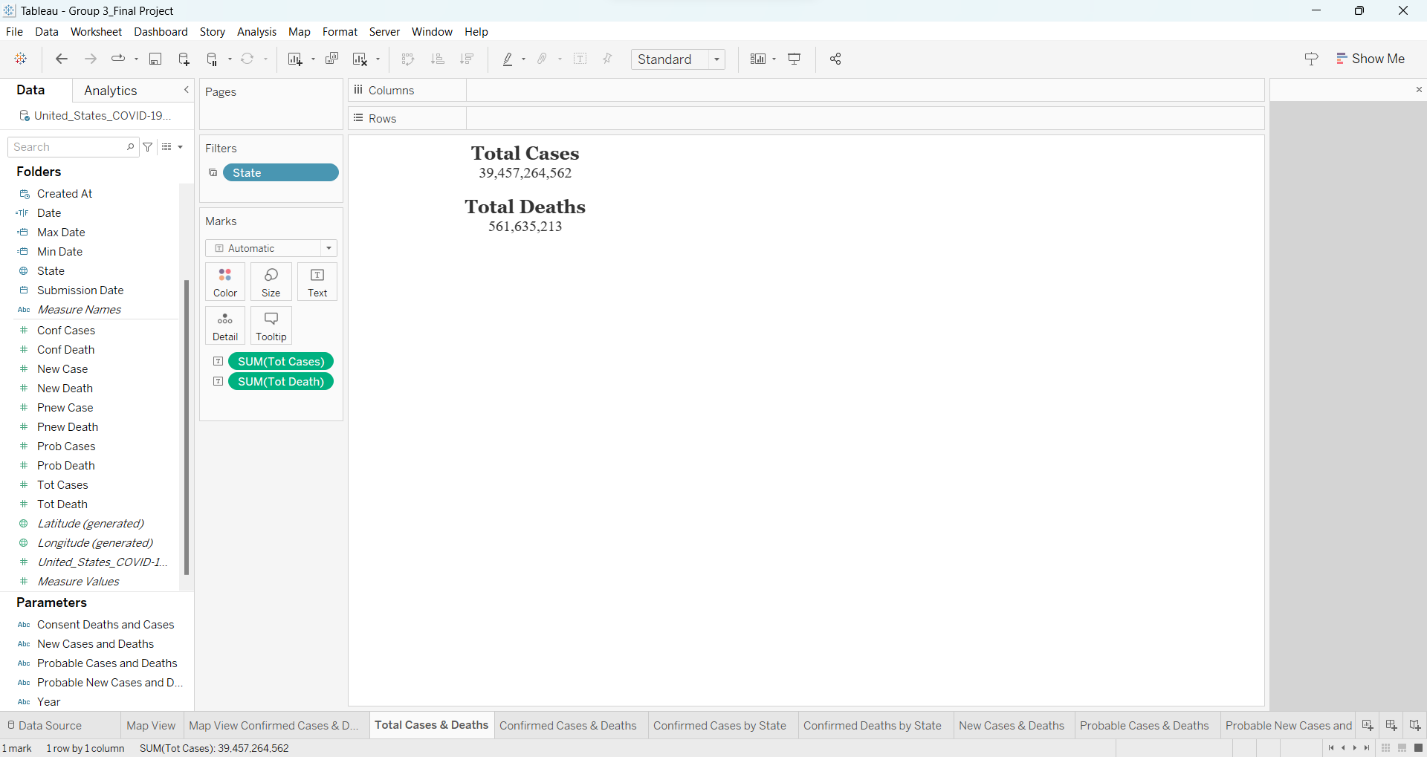
**Purpose of the Dashboard-1:**

The primary objective of the dashboard is to provide information on cases of Covid-19 and deaths that have occurred throughout the United States throughout time. It is made up of a variety of visualizations, each of which focuses on important parts of our project, and in addition, it includes a significant portion of the information that we intend to cover throughout the scope of this project. The geographical representation of the cases and deaths as a circle located at that state is shown in the map view as a circle. This dashboard gives a numerical representation of cases and fatalities, as well as a graph visualization that illustrates the rise in the number of cases and deaths over time. This dashboard includes a filter that allows users to zero in on states of interest.

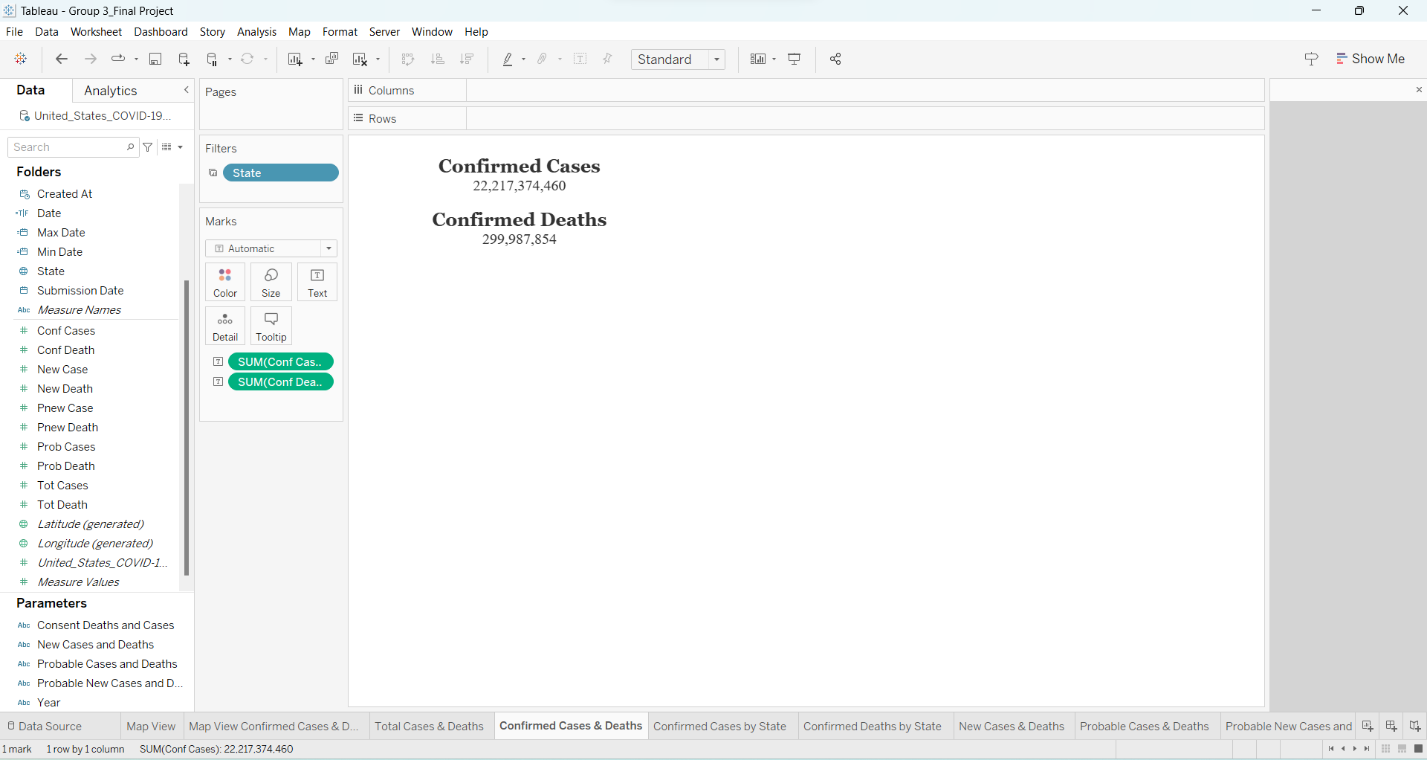
**Contents of Dashboard-1:**

 **Map View: Total Cases and Deaths in U.S.A**

* The above is the Map view for total cases and deaths
* Cases and deaths are represented by a shape(circle)
* The size of the shape varies based on the count.

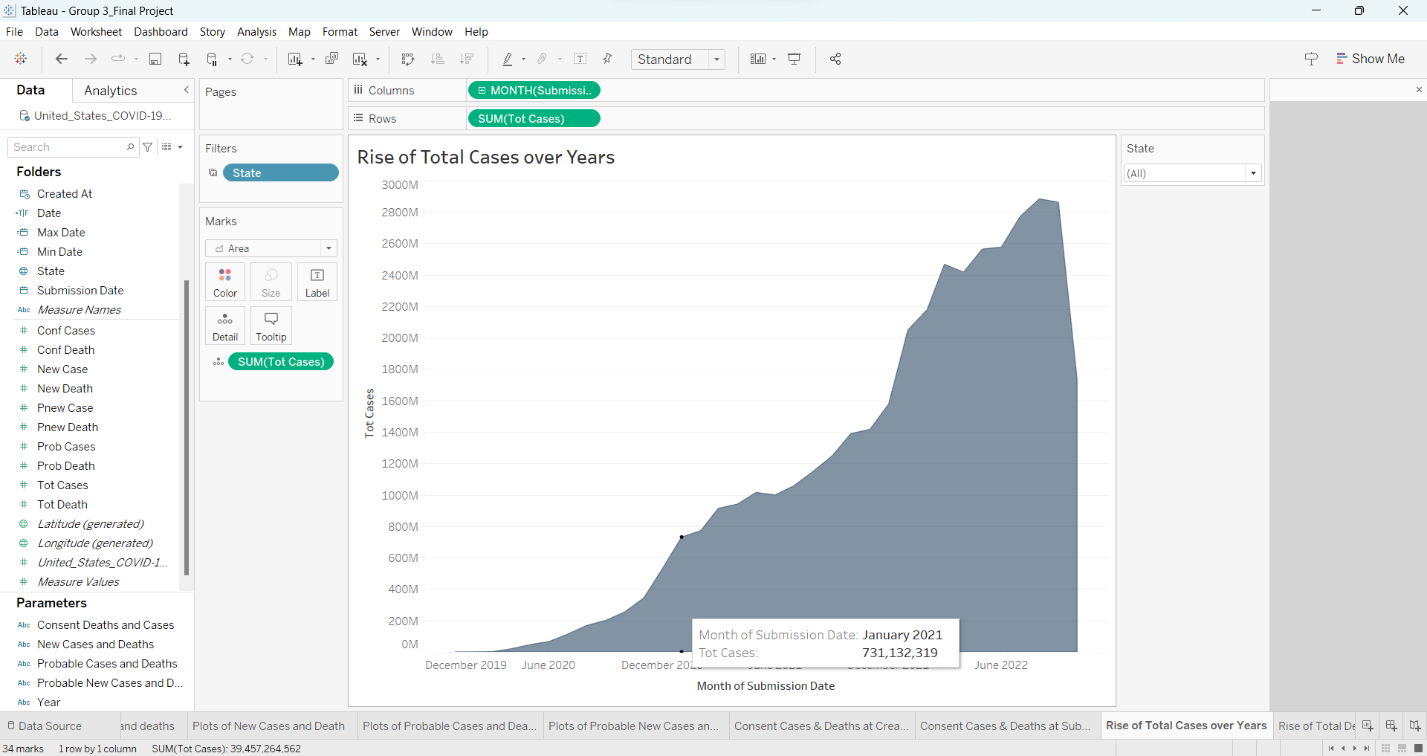
**Label: Total Cases and Deaths**

* The above sheet is the Numerical visualization of total deaths and total cases count.



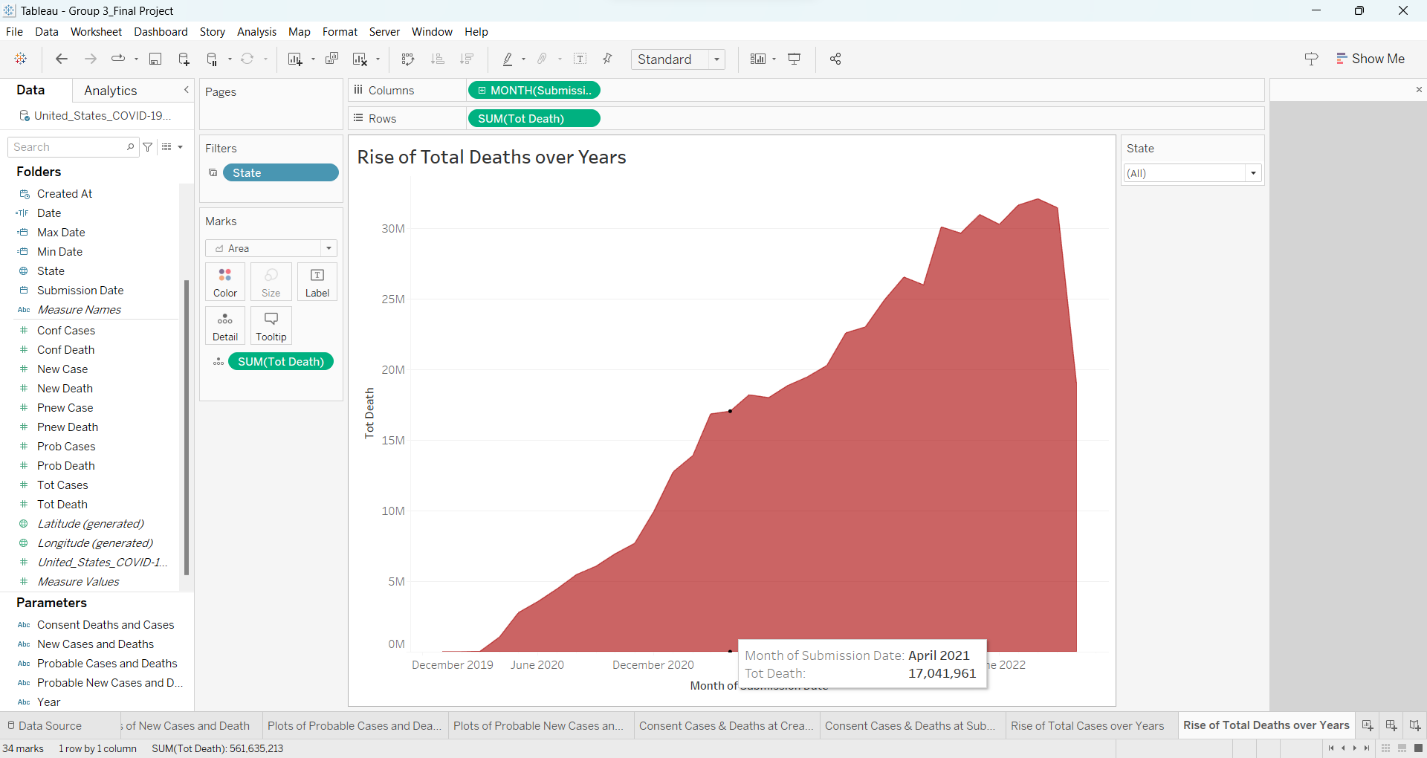
**Label: Confirmed Cases & Deaths**

* The above sheet is the Numerical visualization of confirmed deaths and confirmed cases count.



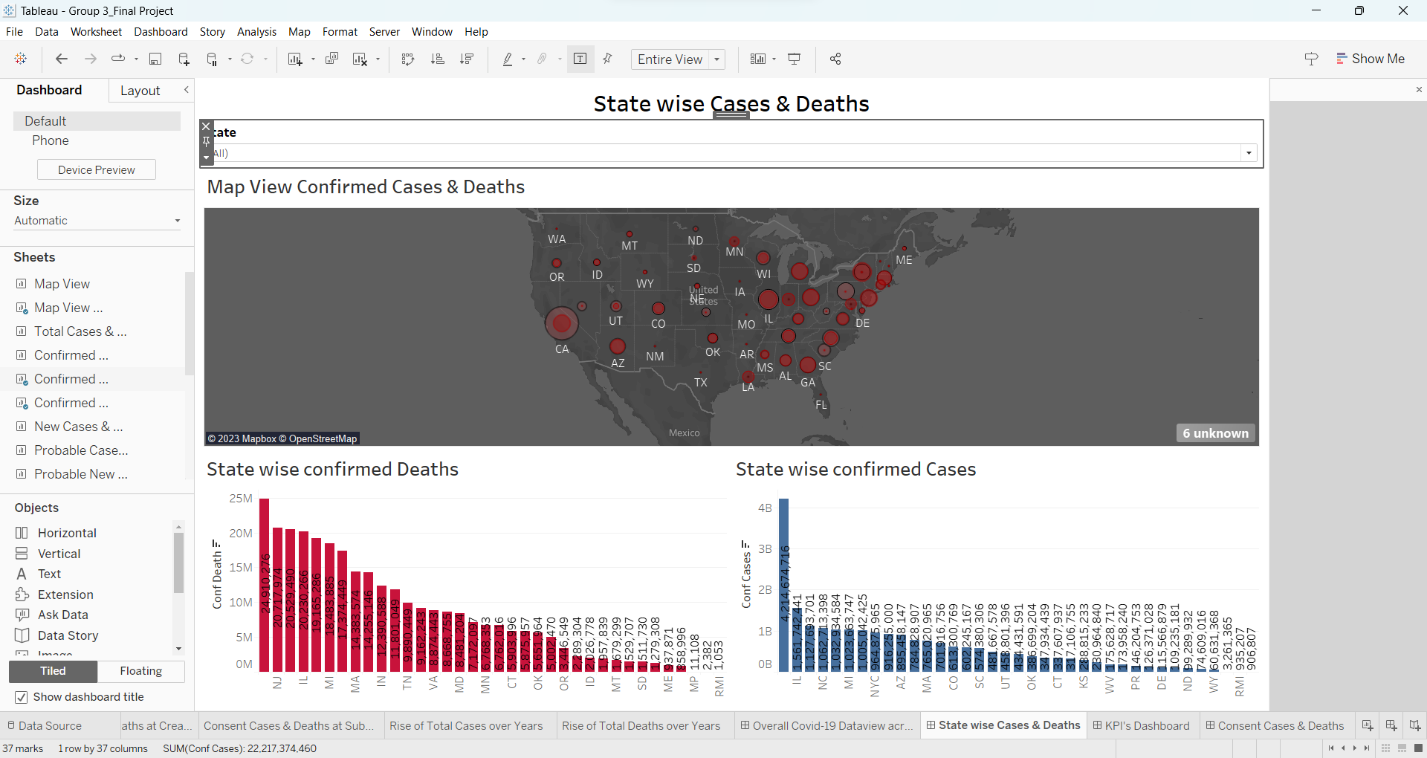
**Area chart: Rise of Total Cases over the years**

* The above is the chart that represents the growth of total cases over years.
* This constitutes of datapoints for every month provided in the dataset
* This sheet also has filter to get the total cases for a particular state(s) as a state(s) of Interest

 **Area chart: Rise of Total Deaths over the years**

* The above is the chart that represents the growth of total deaths over years.
* This constitutes of datapoints for every month provided in the dataset
* This sheet also has filter to get the total deaths for a particular state(s) as a state(s) of Interest

**Dashboard-2:**

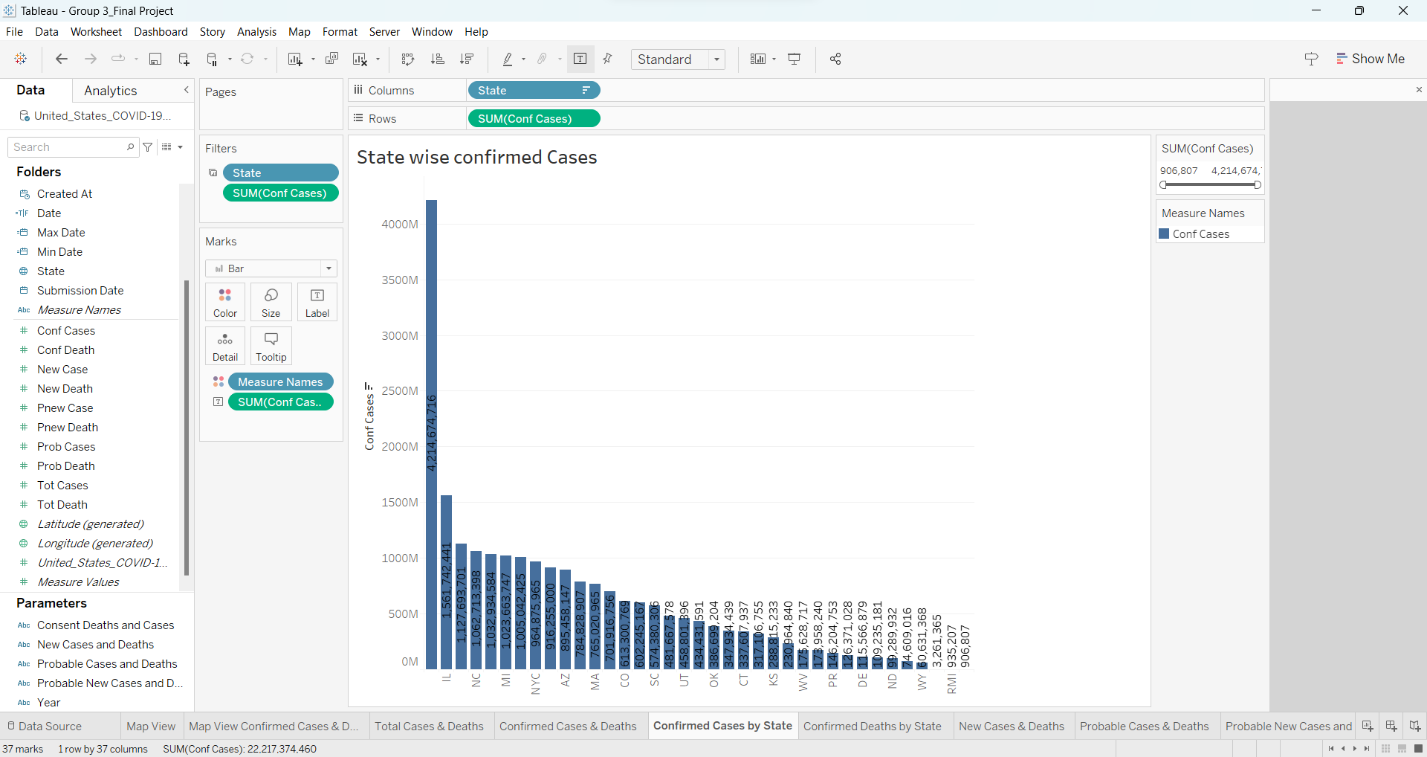
**State wise Cases and Deaths from Highest to Lowest**

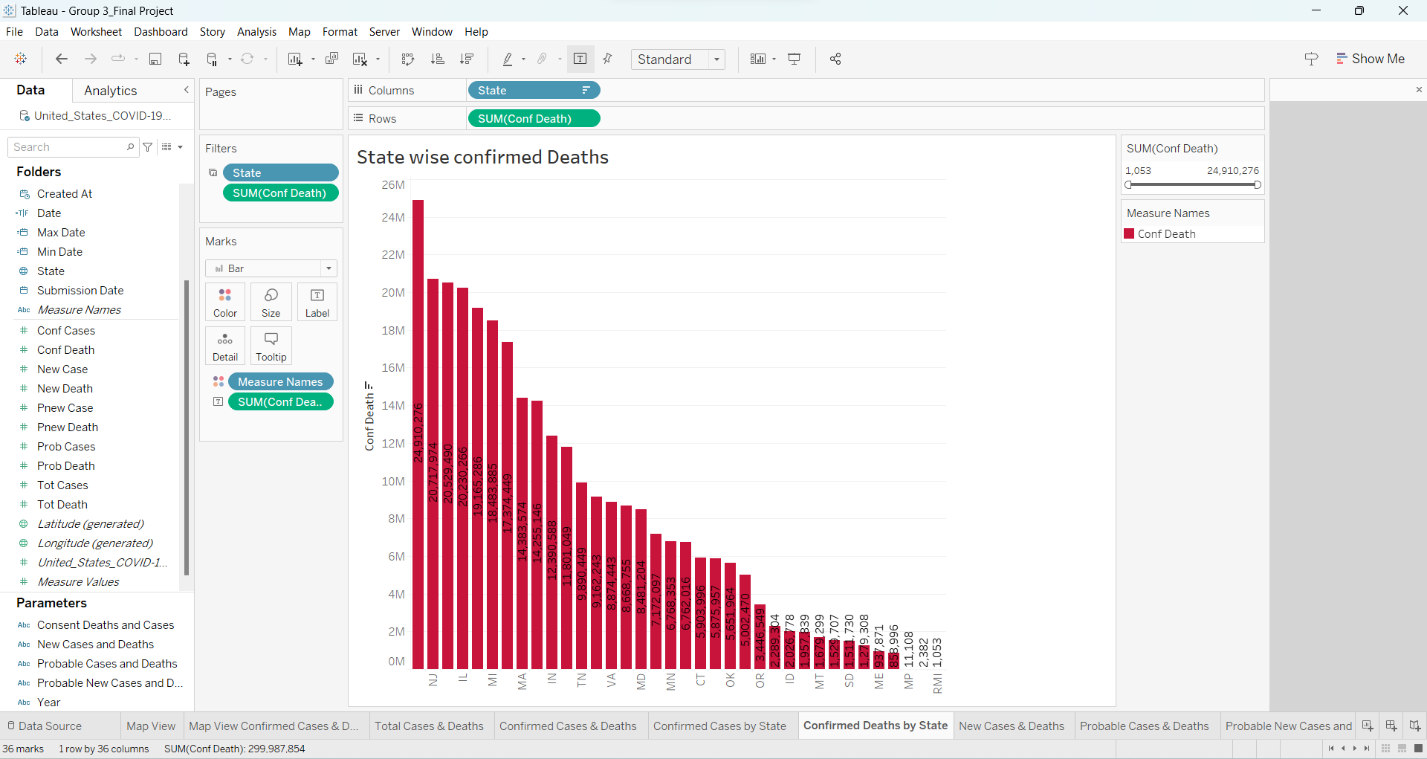
**Purpose of Dashboard-2:**

The primary objective of the dashboard that was just displayed is to provide information about the states that have the highest/lowest number of cases and deaths. This dashboard, like dashboard 1, features a map view that illustrates the overall number of cases and deaths in a certain state. In this example, the state in question. This dashboard also contains bar graphs, with the deaths shown in red and the cases shown in blue, with the counts descending from highest to lowest. This provides a concise knowledge of which state has the greatest recorded case numbers as well as deaths. In addition to that, it possesses a filter that allows one to zero in on a specific condition.

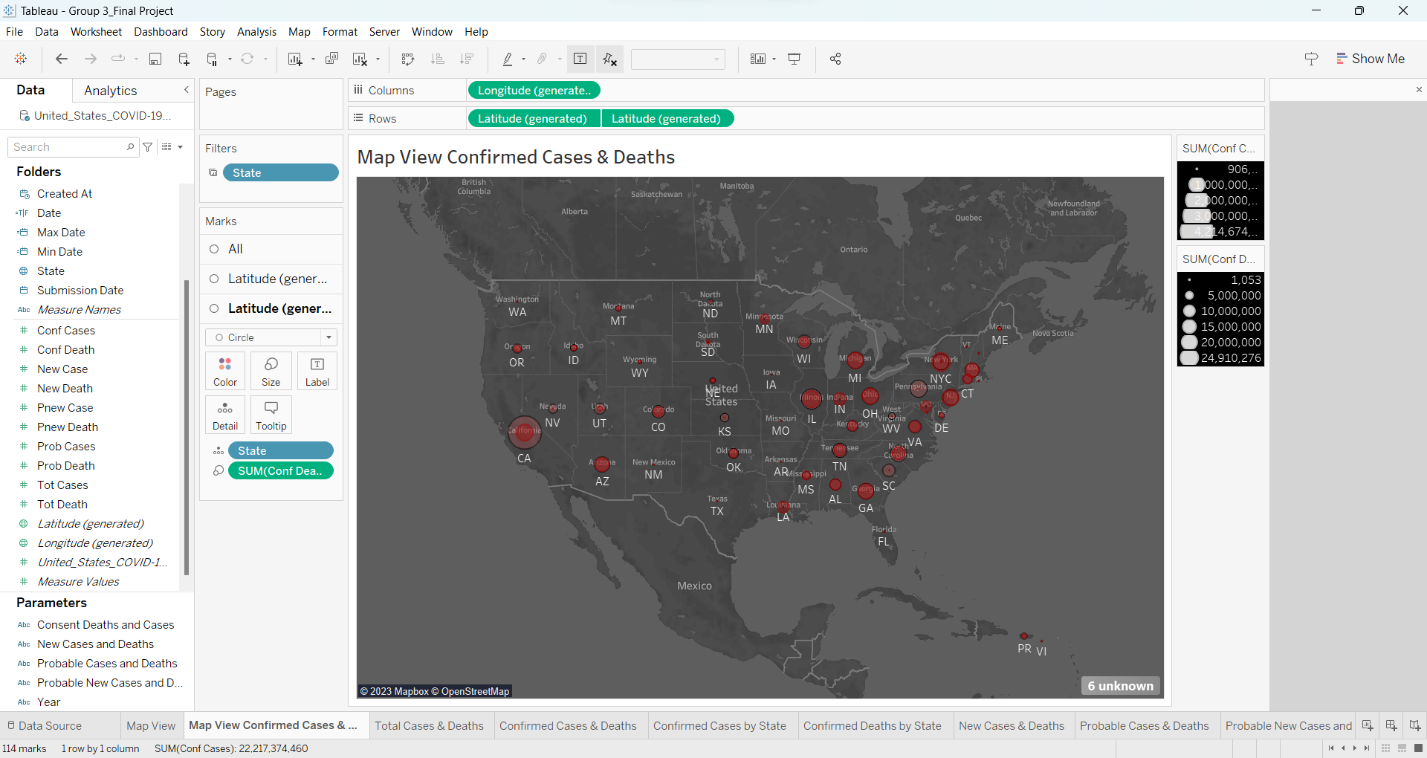
**Contents of Dashboard-2:**

* The below is the bar chart represents the total cases recorded by a state which is arranged from highest number to the lowest.
* Each bar in the representation has a label with total count of cases.

 **Bar chart: State wise Confirmed cases**

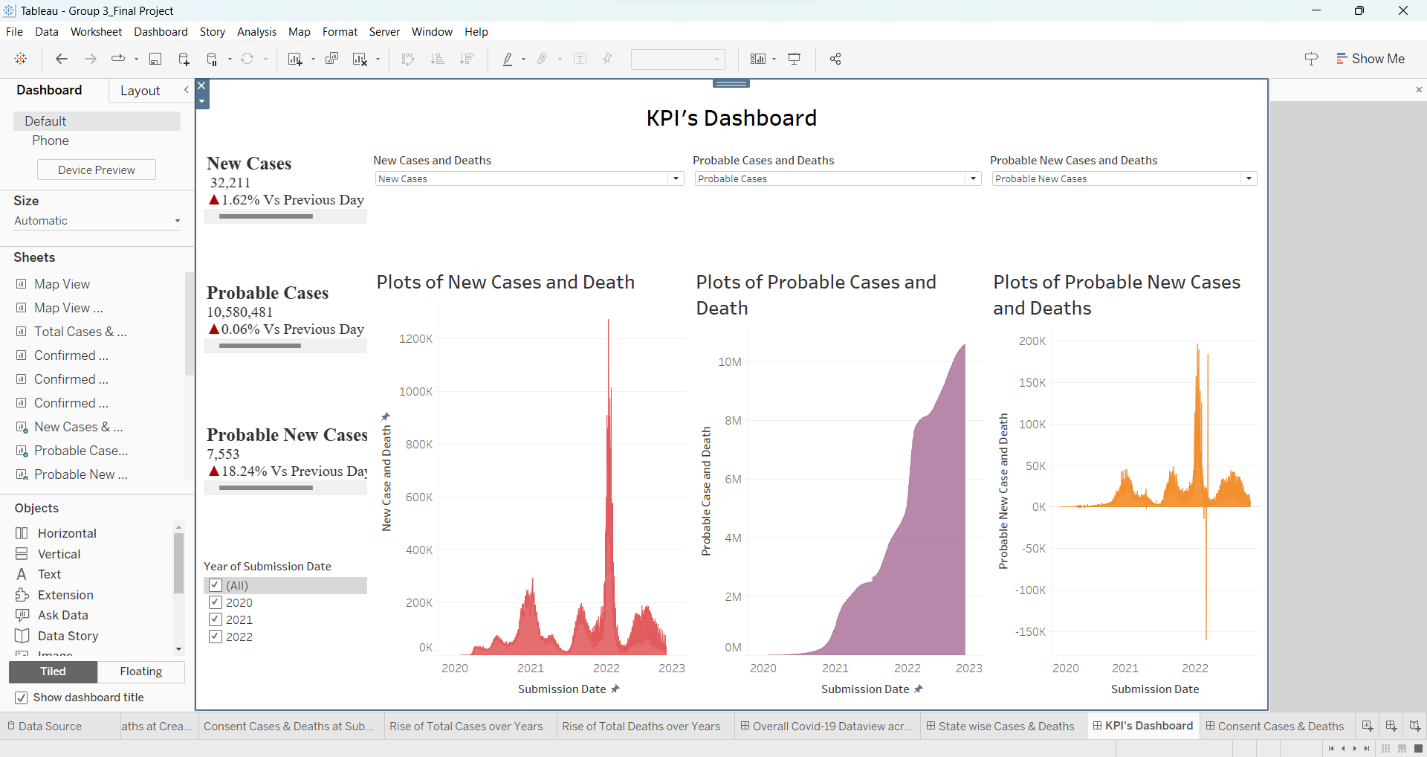
 **Bar chart: State wise Confirmed Deaths**

* The above is the bar chart represents the total deaths recorded by a state which is arranged from highest number to the lowest.
* Each bar in the representation has a label with total count of deaths.

**Map: Confirmed Cases and Deaths**

* The above is the Map view for total cases and deaths
* Cases and deaths are represented by a shape(circle)
* The size of the shape varies based on the count.

**Dashboard-3:**

 **KPI’s Dashboard**

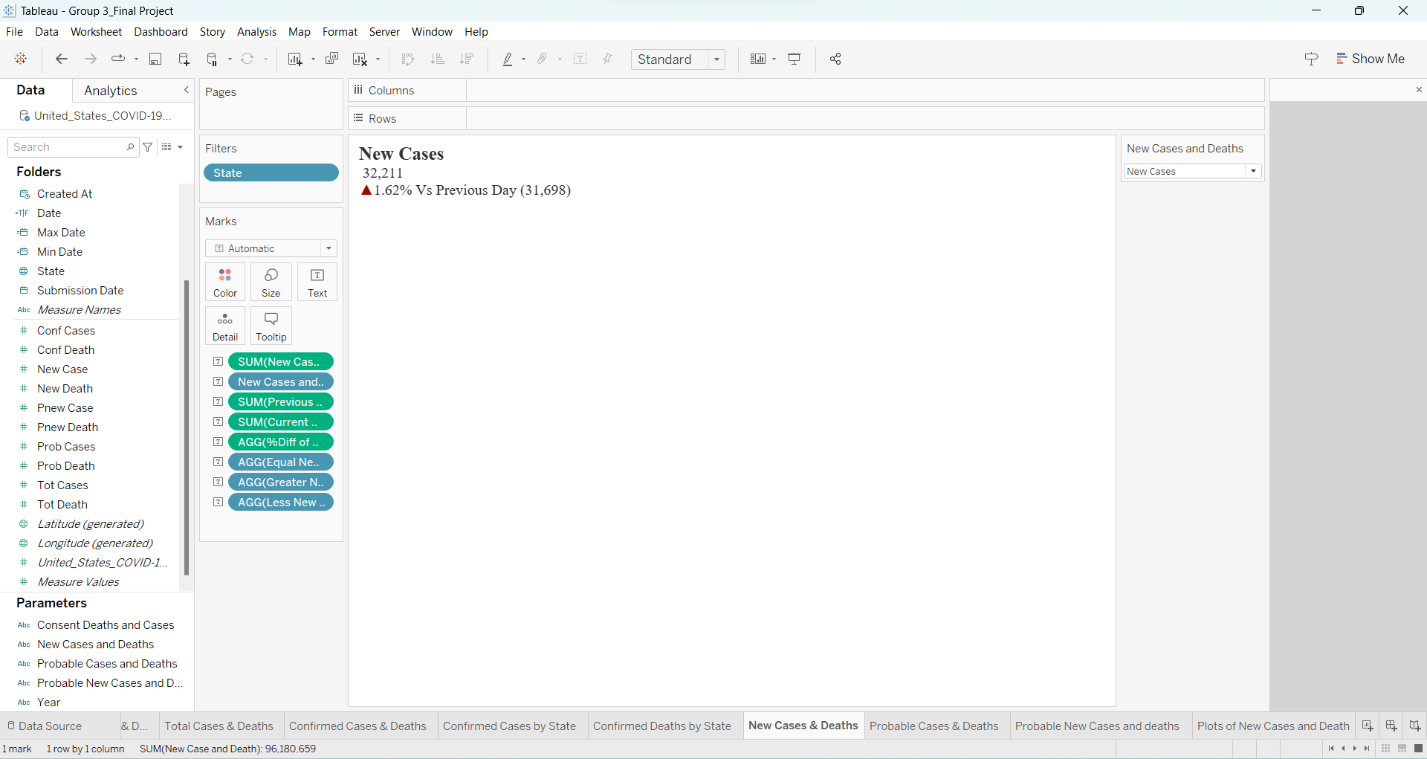
**Purpose of the Dashboard-3:**

The main purpose of the above dashboard is to understand the Key Performance Indicators (KPIs) for this project. This comprises the numerical and symbol representation for New Cases, Probable Cases and Probable New Cases. And these three representations include the following symbols

* Triangle: Growth in total cases/deaths from previous day
* Inverted Triangle: Decrease in total cases/deaths from previous day
* Dot: No change of total cases/deaths of that day when compared to the previous day

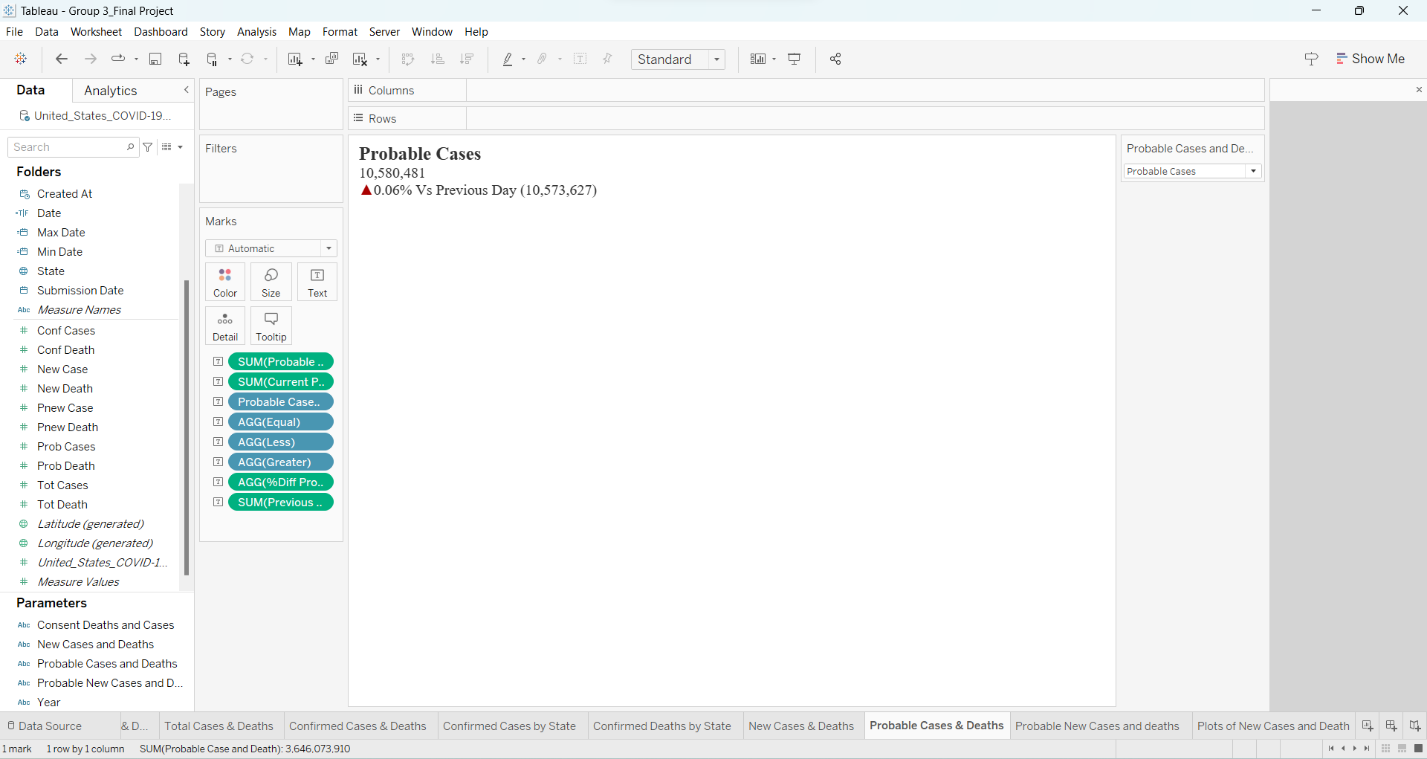
This dashboard also has 3 visual representations which depicts new cases and deaths, Probable cases and deaths, Probable new cases, and deaths. With all the visualizations in the dashboard, it gives better understanding on how the cases are increasing/decreasing for the day

**Contents of Dashboard-3:**



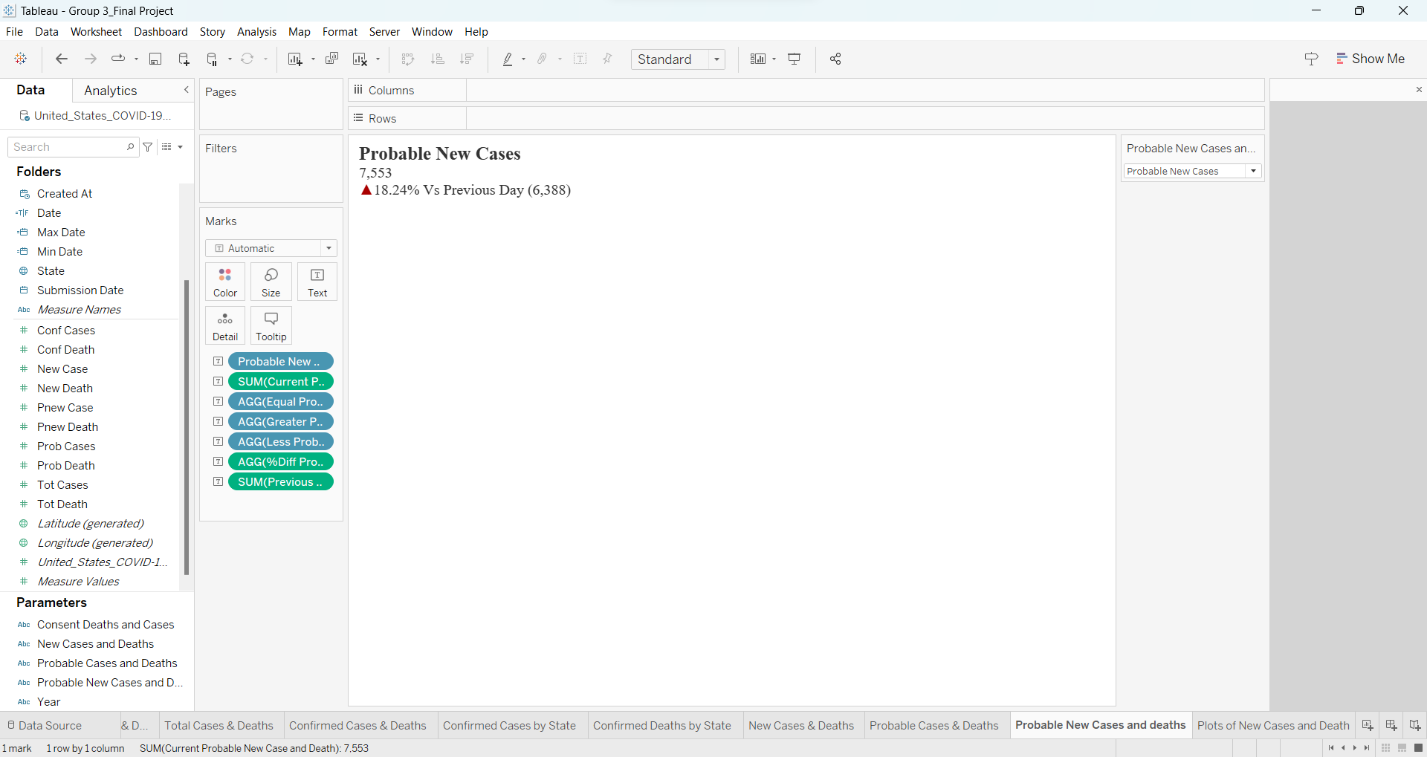
**Label: New Cases**

* This sheet depicts the numerical representation of new cases.
* It has a symbol (Triangle/Inverted Triangle/Dot) which represents the percent change in total count of that day in comparison to previous day.



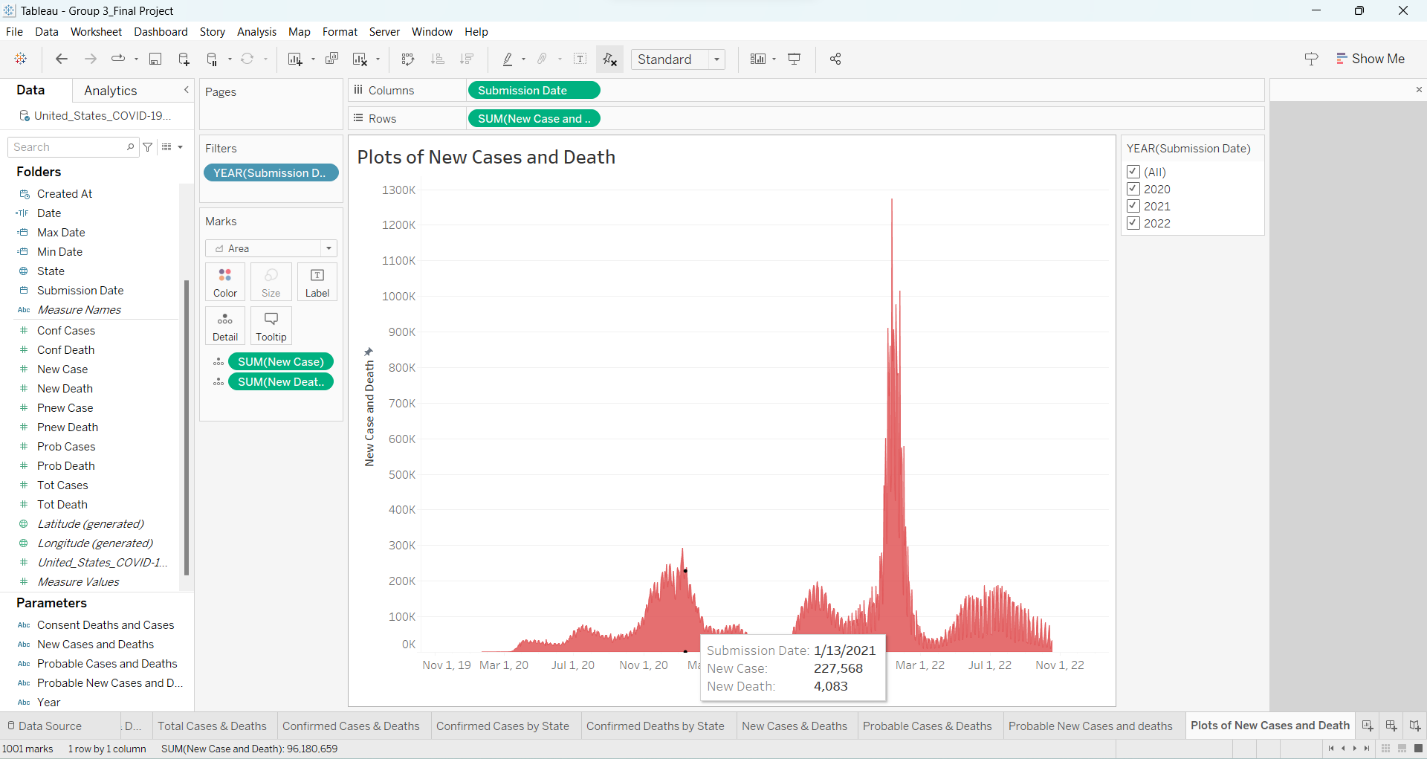
**Label: Probable Cases**

* This sheet depicts the numerical representation of probable cases.
* It has a symbol (Triangle/Inverted Triangle/Dot) which represents the percent change in total count of that day in comparison to previous day.



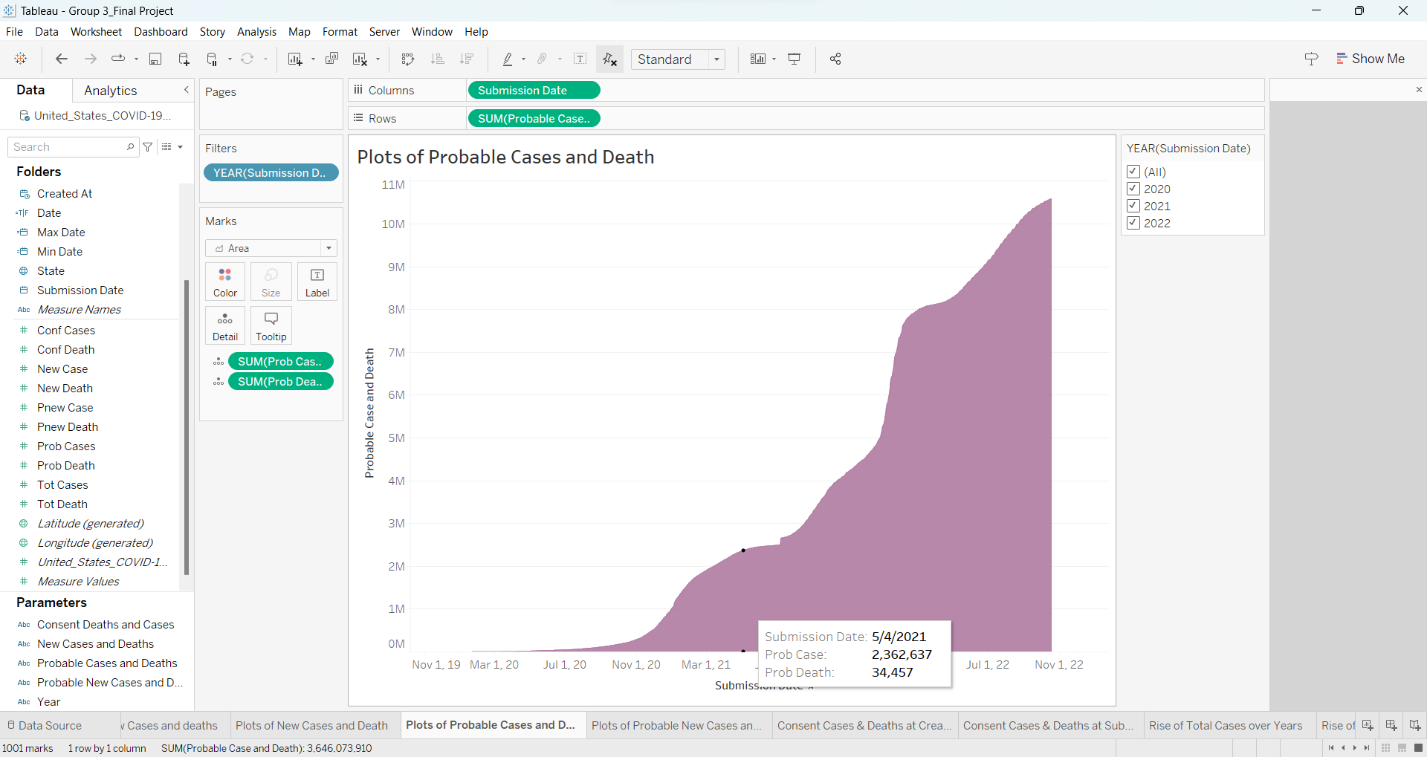
**Label: Probable New Cases**

* This sheet depicts the numerical representation of probable new cases.
* It has a symbol (Triangle/Inverted Triangle/Dot) which represents the percent change in total count of that day in comparison to previous day.

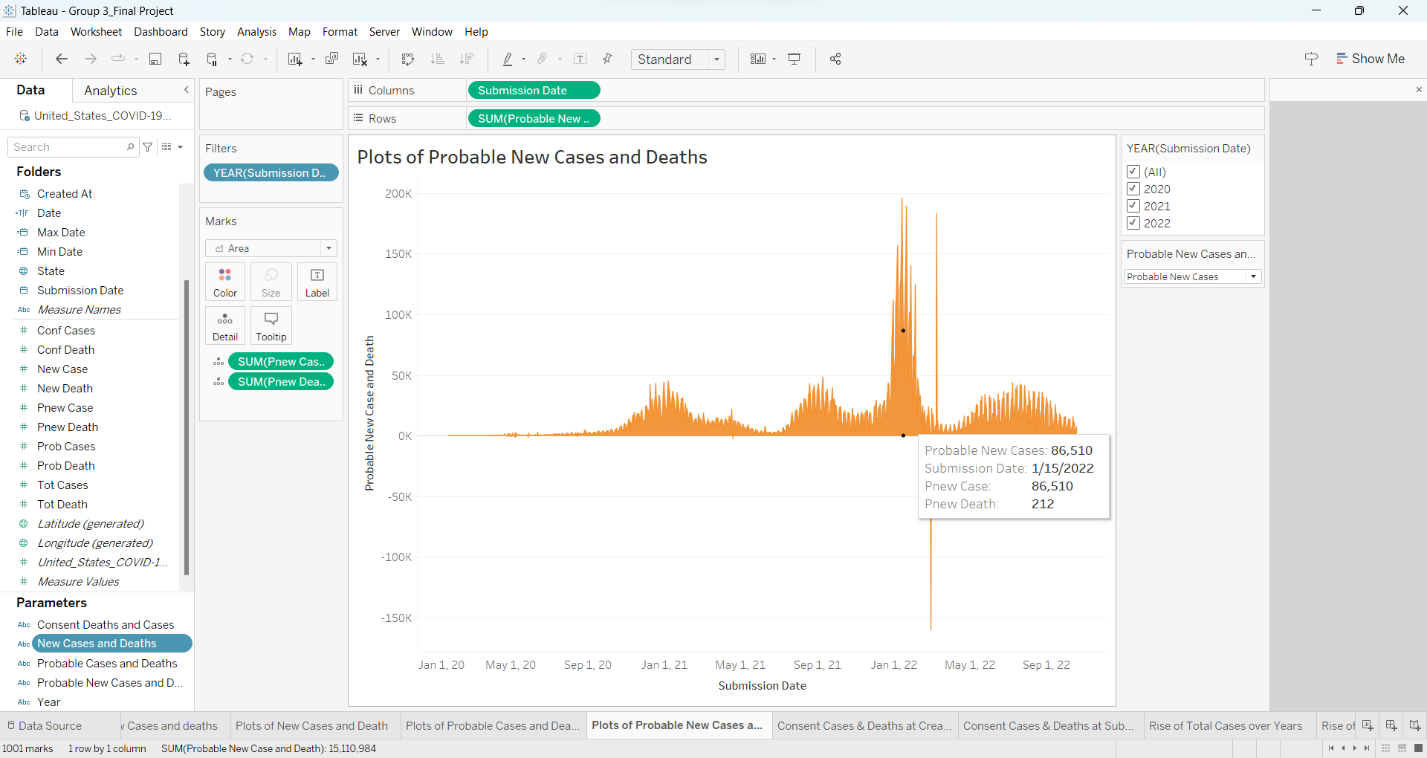


**Area Chart: new cases and deaths**

* The above sheet depicts the total cases and deaths on that submission date
* It has a tooltip which consists of information of the submission date, new cases and new deaths.
* This sheet helps to better understand the daily statistics of cases/deaths.

**Area Chart: Probable Cases and Deaths**

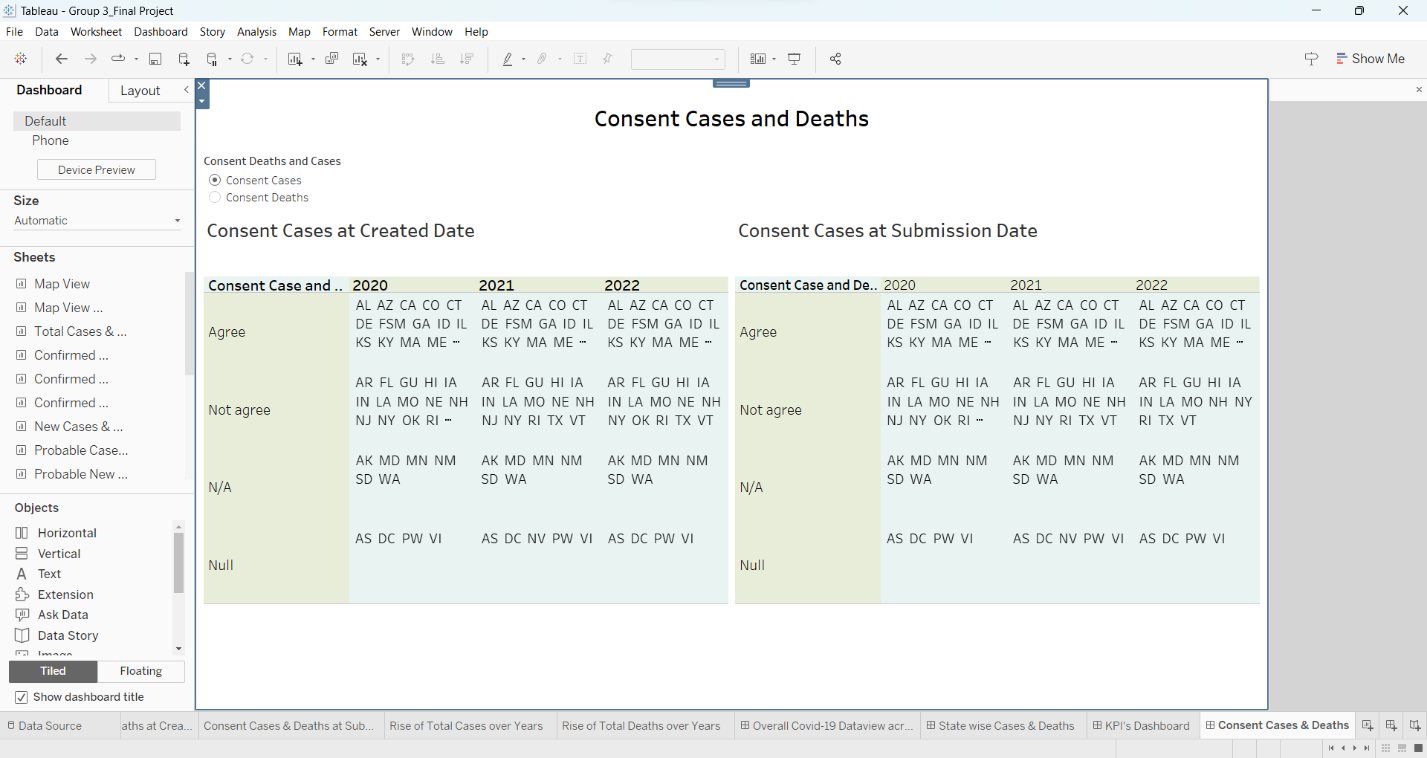
* The above sheet depicts the total probable cases and deaths on that submission date
* It has a tooltip which consist information of the submission date, probable cases, and probable deaths.
* This sheet helps to better understand the daily statistics of probable cases/deaths.



**Area Chart: Probable New Cases and Deaths**

* The above sheet depicts the Probable new cases and deaths on that submission date
* It has a tooltip which consist information of the submission date, probable new cases, and probable new deaths.
* This sheet helps to better understand the daily statistics of probable new cases/deaths.

**Dashboard-4:**



**Consent Cases and Deaths**

**Purpose of the Dashboard-4:**

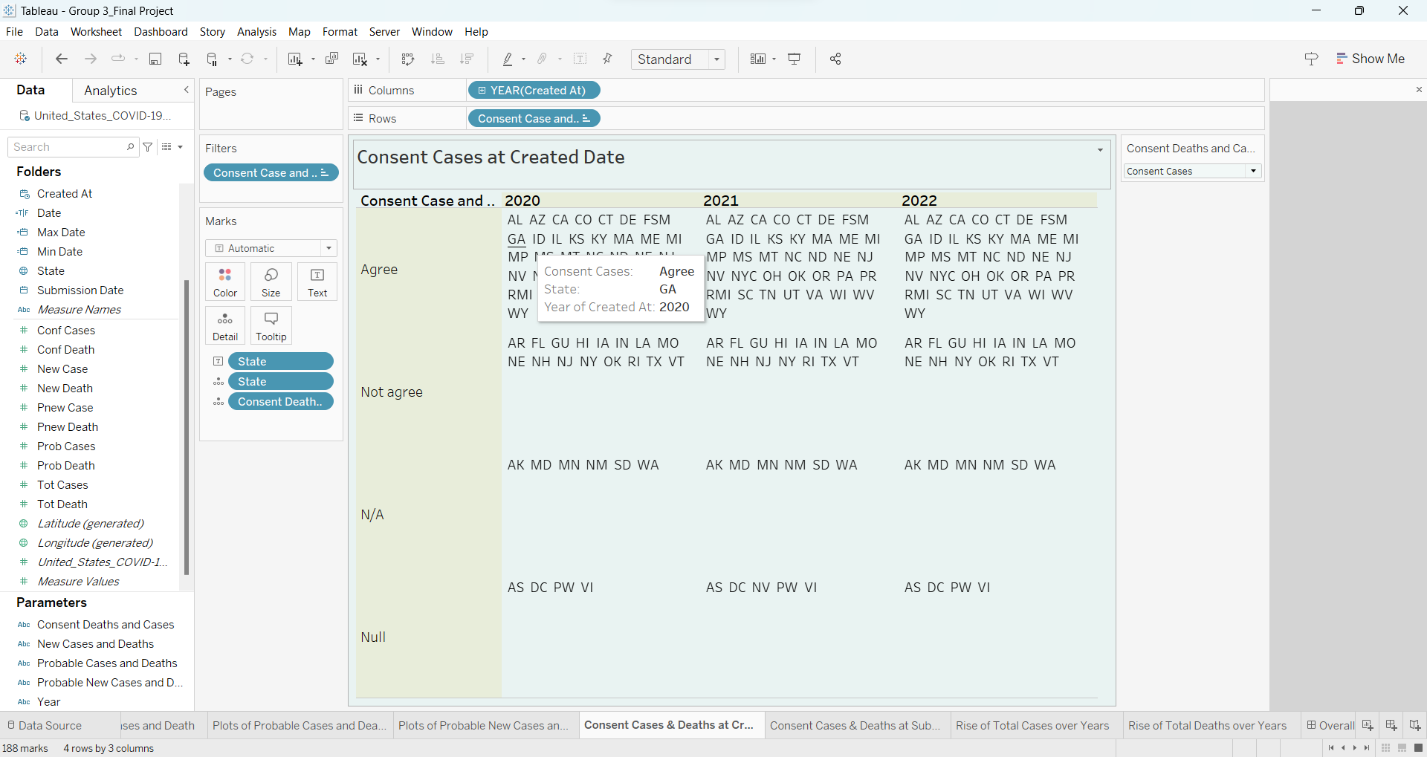
The main purpose of this dashboard is to separate the states based on their consent towards death and cases based on created date and submitted date over the years (2020,2021,2022).

Consent has 4 values:

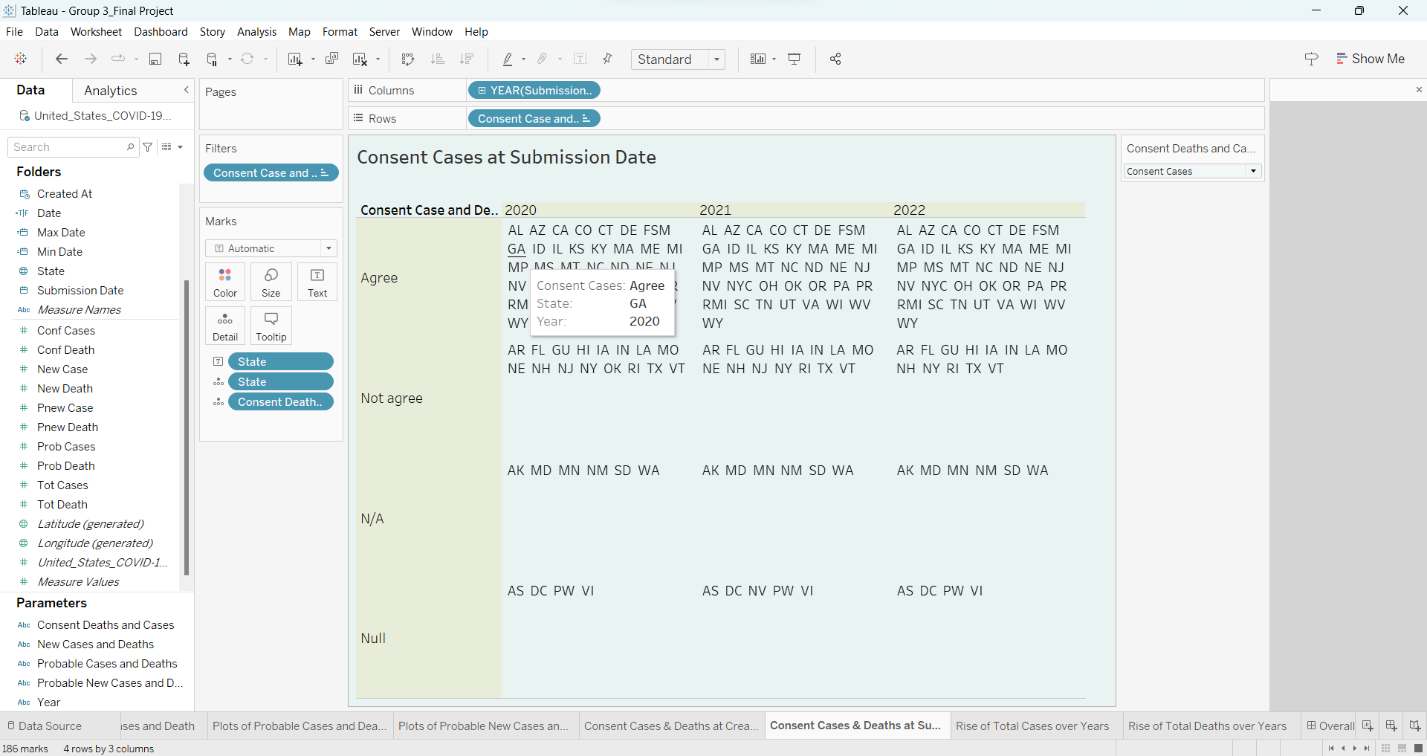
* Agree: If consent is agreed then data is the total of confirmed cases/deaths and probable cases/deaths
* Not Agree: If consent is no agree then data is the total cases/deaths
* N/A: Not applicable
* Null: Data is Absent

With this consent we can get an understanding of what kind of data that is being provided for that state. This dashboard also has filter to toggle between consent of cases and deaths.

**Contents of Dashboard-4:**

 **Consent Cases/Deaths at Created Date**

* This sheet divides the data into four consent values as described above based on the created date.
* It is over the three years (2020, 2021, 2022).
* Filter to toggle between consent cases and deaths.

 **Consent Cases/Deaths at Submission Date**

* This sheet divides the data into four consent values as described above based on the submitted date.
* It is over the three years (2020, 2021, 2022).
* Filter to toggle between consent cases and deaths.

**Conclusion:**

A user can gain useful insights from each of the dashboards described above. Each dashboard provides information that is both important and useful. With a manner analogous to this, we can generate an unlimited number of dashboards with tableau. The only catch is that we must make sure we follow all the correct design stages and industry standards.

**Learnings from this project:**

While working together to finish this course, our group picked up a huge number of new skills. To begin, we were able to acquire a distinct understanding of the way data is managed during the building of dashboards and the extraction of valuable insights from it. Understanding the data and limiting ourselves to the characteristics that were necessary for the project was the most significant and vital thing that we picked up from this experience. As soon as the data is ready, then conduct an analysis of the most effective visualizations for the data that is accessible. This became clear to us after putting in the necessary work. Finally, when building the dashboard, we made sure to follow the best practices. These are all things that we learnt through a process that took place day-by-day and week-by-week.

**Acknowledgment:**

Today, we can construct a dashboard using Tableau by adhering to all the recommended standards. We would like to express our gratitude to Professor Dr. Jack Zheng for teaching this class to each student and for supplying them with relevant references and essential documents. Our understanding of data visualization is much better because of all of this.

**References:**

* Data Visualization with Examples: https://www.tableau.com/learn/articles/data-visualization
* Intro to tableau: <https://help.tableau.com/current/guides/get-started-tutorial/en-us/get-started-tutorial-home.htm>
* Conditional Formatting: <https://community.tableau.com/s/question/0D54T00000F33w6SAB/conditional-formatting-and-symbols>
* KPI in tableau: <https://help.tableau.com/current/pro/desktop/en-us/kpi.htm>