Process Book

COVID-19 Data visualization with D3

CS480X Final project

Xiaowei Chen

Haowen Wei

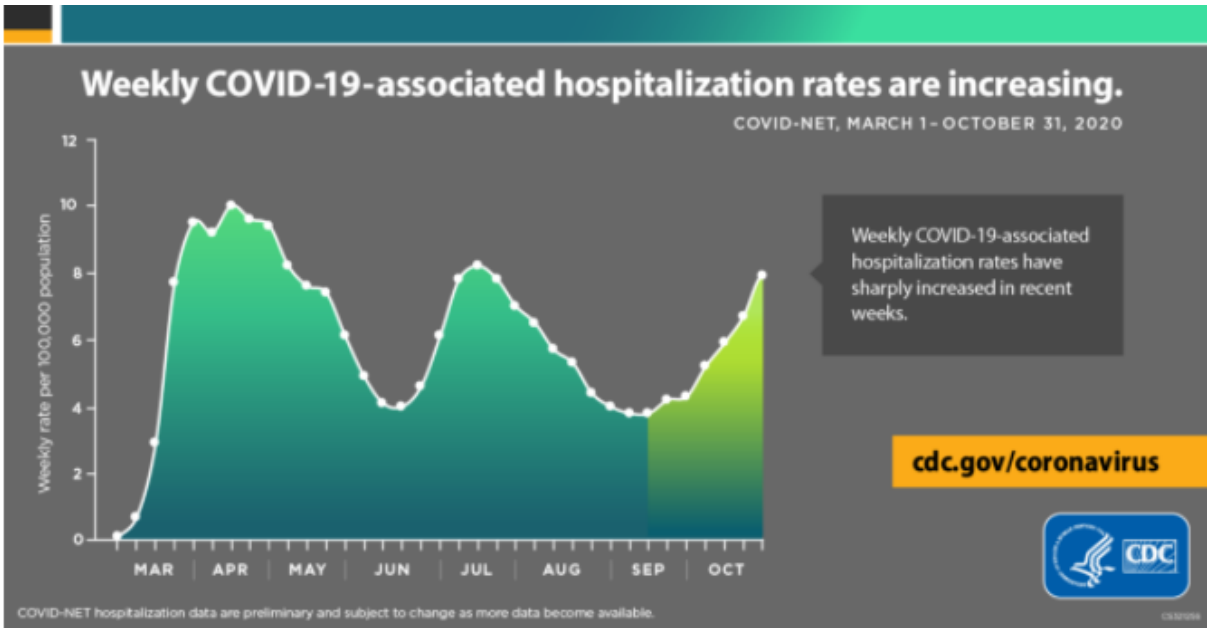
Matthew Hurlbut-Coke

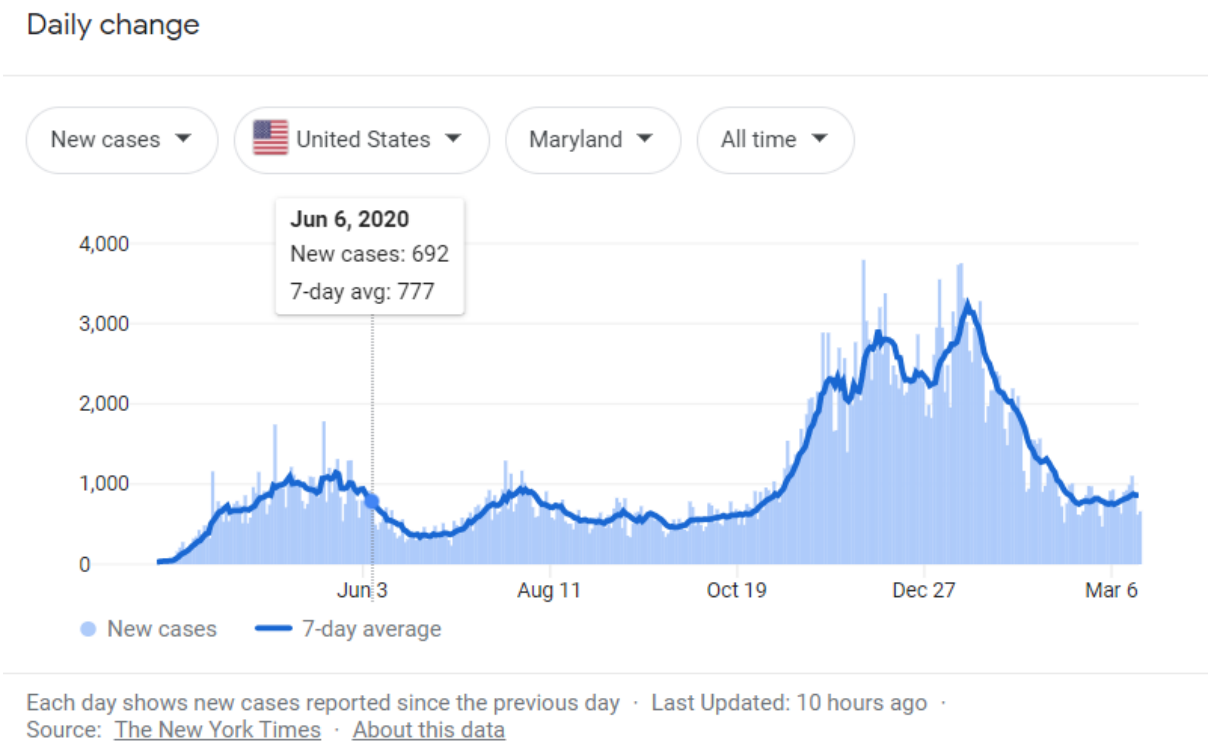
Overview and Motivation:

Since the first time Covid-19 appeared, it has lasted for more than one year. Millions of people died because of it, many countries' social orders were hit hard and their economy suffered great losses. Therefore, this is a topic that relates very close to our life and is worth paying attention to. By using the data visualization we created, users can quickly learn about the pandemic situation not only for the whole country but also for the states. We filtered the data and created several visualizations that show users several aspects about the pandemic to help them understand what is happening at this point.

Related Work:

The following are two related works for Covid-19. One is from CDC, showing the weekly rate per 100, 000 population versus each week; the other one is from Johns Hopkins University, showing both the new positive cases and week average calculated based on the statistics of each day.





Massachusetts government is continuously updating the pandemic-related information on its official website: https://www.mass.gov/info-details/covid-19-response-reporting
Here it has detailed descriptions on many aspects, like data reporting and vaccine information. From the weekly public health report section, we can find some data that generally summarize for cities in Massachusetts:

City/Town	Total Case Count	Case Count (Last 14 Days)	Average Daily Incidence Rate per 100,000 (Last 14 days)[1]	Relative Change in Case Counts[2]	Total Tests	Total Tests (Last 14 days)	Total Positive Tests (Last 14 days)	Percent Positivity (Last 14 days)	Change in Percent Positivity[3]
Williamstown	224	<5	4.0	Lower	66130	7846	4	0.05%	No Change
Wilmington	1909	76	22.2	Higher	39034	2665	79	2.96%	Higher
Winchendon	603	28	18.6	Lower	13961	856	30	3.50%	Lower
Winchester	1111	31	9.9	Lower	59430	4475	41	0.92%	No Change
Windsor	19	0	0	No Change	884	45	0	0%	No Change
Winthrop	2137	59	22.3	Lower	62366	3304	79	2.39%	Lower
Woburn	3749	119	20.5	Higher	82837	5406	136	2.52%	Higher
Worcester	21150	569	21.2	Lower	649416	49656	645	1.30%	Lower
Worthington	30	0	0	No Change	1655	104	0	0%	No Change
Wrentham	923	40	25.3	Higher	25395	1512	41	2.71%	Higher
Yarmouth	1164	60	17.8	Higher	33512	1670	67	4.01%	Higher
Unknown[4]	1463	11	*	*	267048	13200	11	*	*
State	562394	19791	20.3	Lower	17019491	1245674	22843	1.83%	Lower

When looking at this piece of data, we have a stronger sense that these are not only numbers, but there are real people and real lives behind the data. In Worcester, a total of more than 20,000 people are tested positive for Covid-19. Think about their family, the pandemic has brought a great pain to the city that

we are currently living in. This strengthened our thought to make the final project topic to be the Covid-19.

Questions:

How many deaths were occurring in each state? If we are evaluating the data based on different time schemes (each month? each day?), what does it look like and how is it changing based on the time scheme? What are the differences in the total number of deaths among different states? What if we are not looking at the total number of deaths but the deaths per 1000 people in each state? Airport is an important hub of transportation communication, and how is it related to our topic? If we turn to look at the increase of the number of people tested positive, what will we find out?

Data:

To draw the U.S. map on the first page once we open the app, we need a json file which includes details about the positions and boundaries for all the states in U.S. Here is it:
https://drive.google.com/drive/folders/1c9ZcQUV1w0rBR-NbbyplIJPL9pVAl2MG
To get the complete data which summarizes Covid-19, we went to a website called The Covid Tracking Project. Its data collected is used by many federal organizations like CDC, HHS and FDA, so it is trustworthy. Here is the link for data source:
https://covidtracking.com/data/download

To calculate the deaths per 1000 people in different states, we need data about population. Here is the link for data source:
https://github.com/jakevdp/data-USstates/blob/master/state-population.csv
To find the number of airports in each state, we used a site that provided information and statuses of all the airports in the US.
https://www.globalair.com/airport/state.aspx

Exploratory Data Analysis:

For the airport view, we initially looked at the distribution of airports in each state using simple bar charts to get an idea of which states had more or fewer, and which states if any were outliers.
For the positive increase view and the death brush view, we initially opened Power BI and have the tool to display a line chart for us. We found that there are very clear and obvious changes in the number of deaths and positive increase for each day and each month. Therefore, we are encouraged by that to use D3.js to draw bar / pie charts to show the data based on time schemes.

Design Evolution:

While working on the positive increase view, at first we only had a pie chart. Even if it can show users about which month has the most positive increase cases, it is hard to see how the pandemic is developing through a timeline. Therefore, we decided to add a bar chart. From the bar chart, we can clearly see that in 2020, the new positive cases are rapidly increasing, but in 2021, the new positive cases drop a lot.

In the Death vs Time view, we used the linking view with brushing interaction. One of the functionality for interaction is providing users more detailed information in a general visualization and relating two different dataset. By applying the brushing view between different graphs, users can view the detail

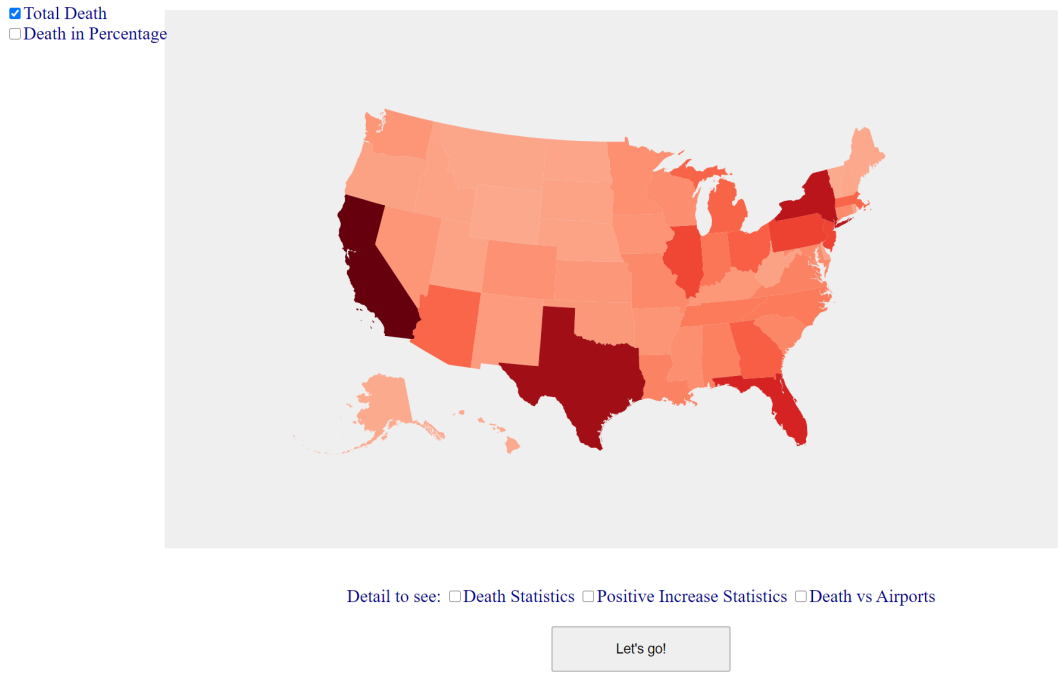
in one dataset while interacting with another graph. This method can help users to understand the relationship between those attributes.

While working on the airport view, some changes were made to better show the data. We hypothesized that states with a higher number of airports per capita might have had outbreaks of COVID-19 sooner than other states, as it might be easier for infected individuals to enter said state. The initial graph plotted airports per thousand people against the total number of deaths per state, but the total number of deaths may not accurately mirror the speed of the outbreaks, and it did not appear to have any significant correlation with the number of airports. Instead, we changed the graph to plot airports against the date that state reached one thousand deaths.

Implementation:

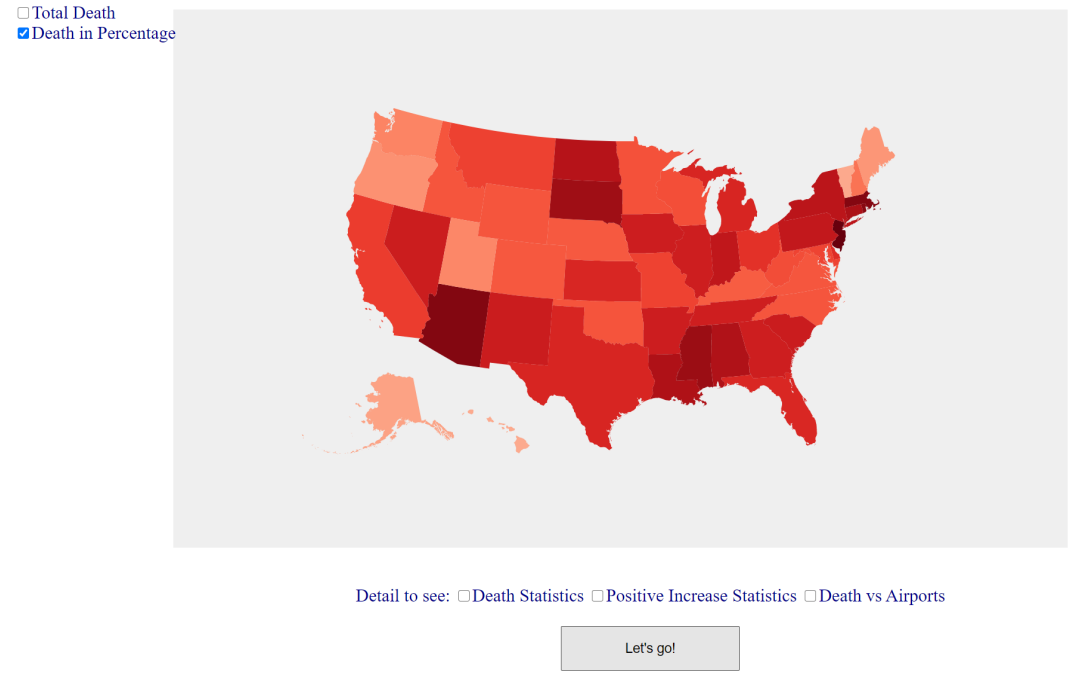
In this final project we have four different data visualization views with interactions.

Homepage:

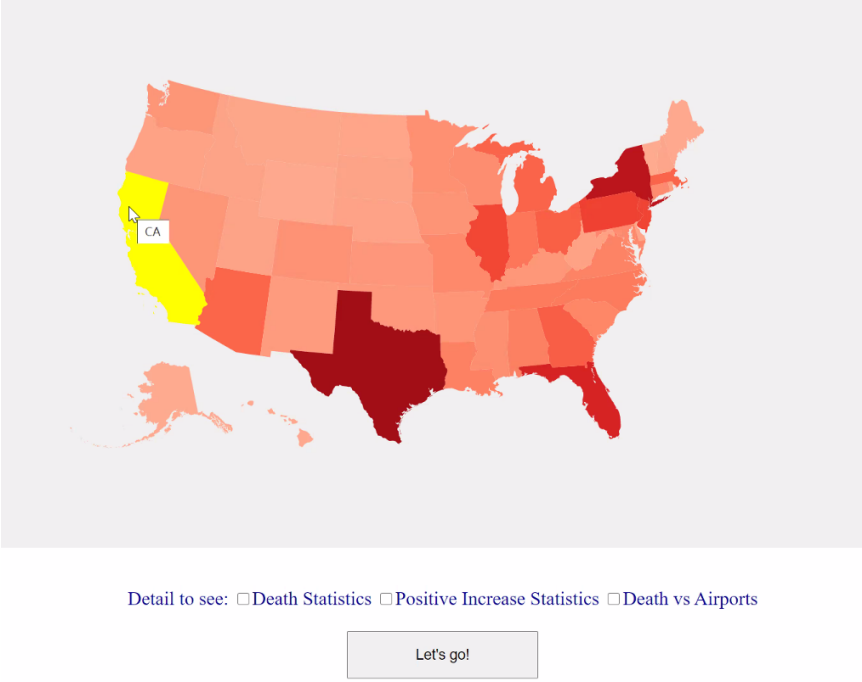


This is the first page users will see when they open the web application. In the center there is a U.S. map. The colors for the states are scaled using d3.interpolateRed.

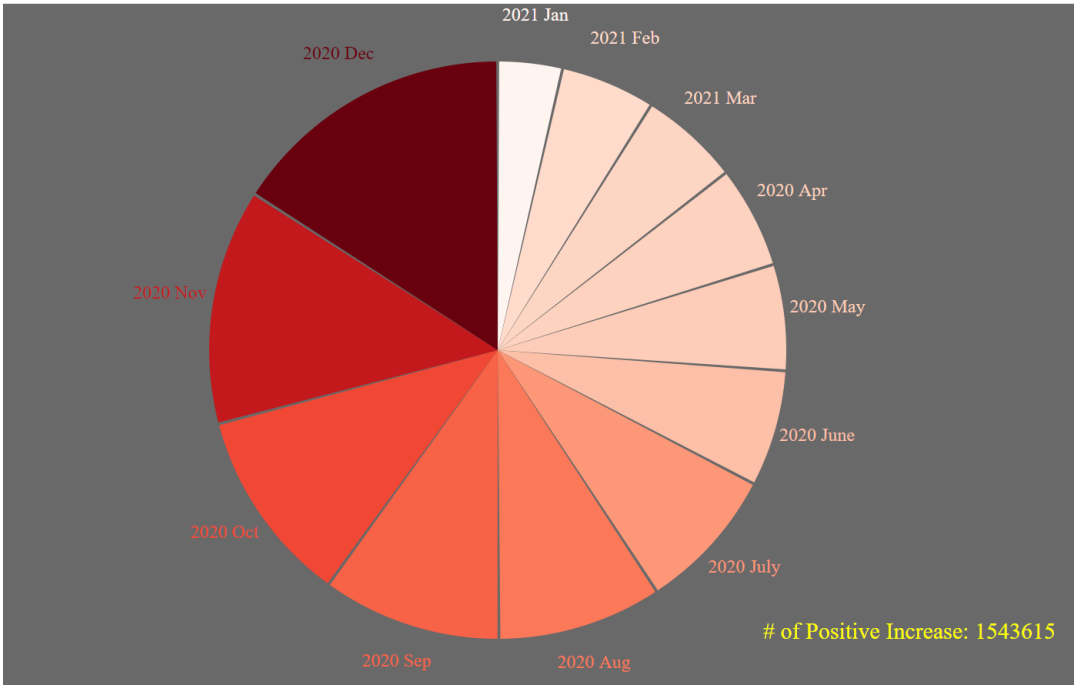
The default colors are showing the number of total deaths for all months in different states, but users can also switch to the mode that shows the colors of total deaths in percentage (total deaths divided by total population)

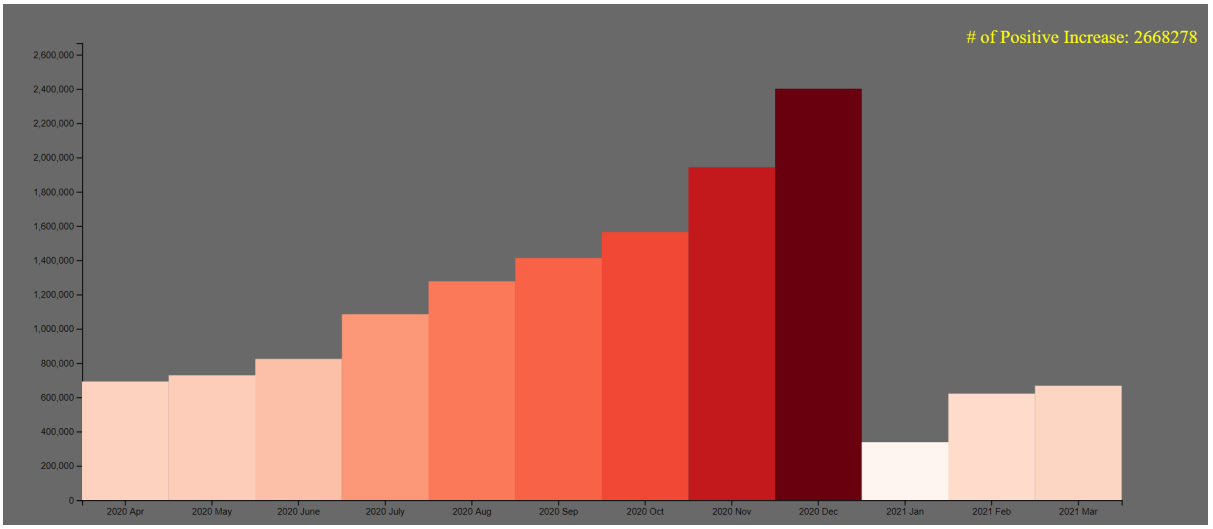


At the bottom of the page, there are three options, and users can only check one at a time. To get redirected to the correct page that users want to see, 1) first click on the state in the map, 2) then check a box, for example, the Death Statistics, 3) finally, click on the Let's go button, 4) page redirected!
Besides, by hovering over each state, users can see the name of the state so that they can know which state they are going to look at.



Positive Increase Statistics:

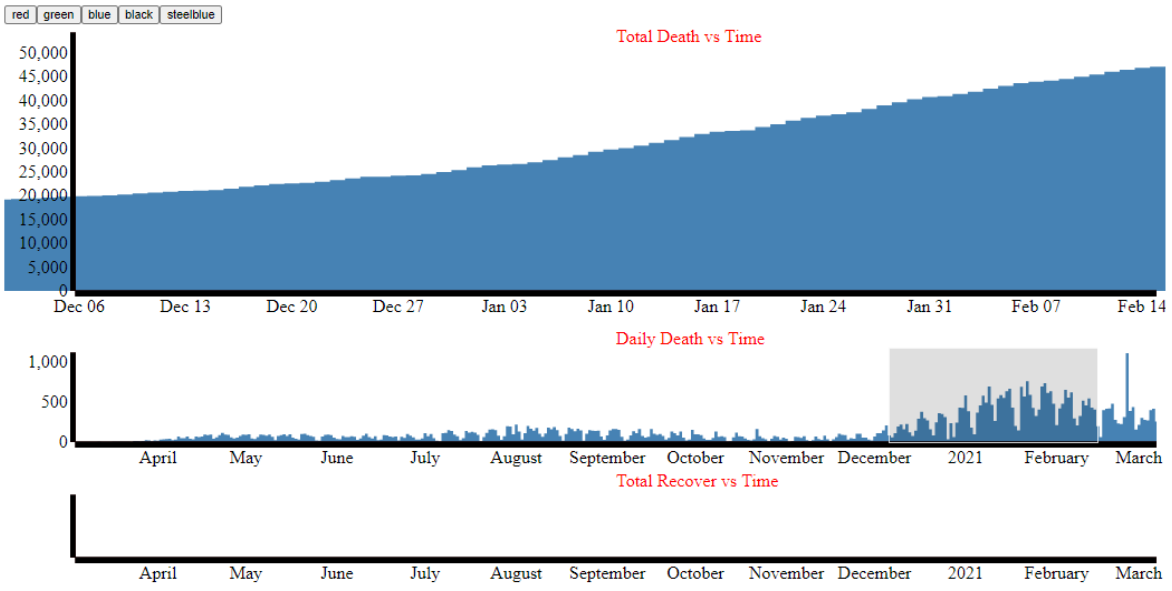




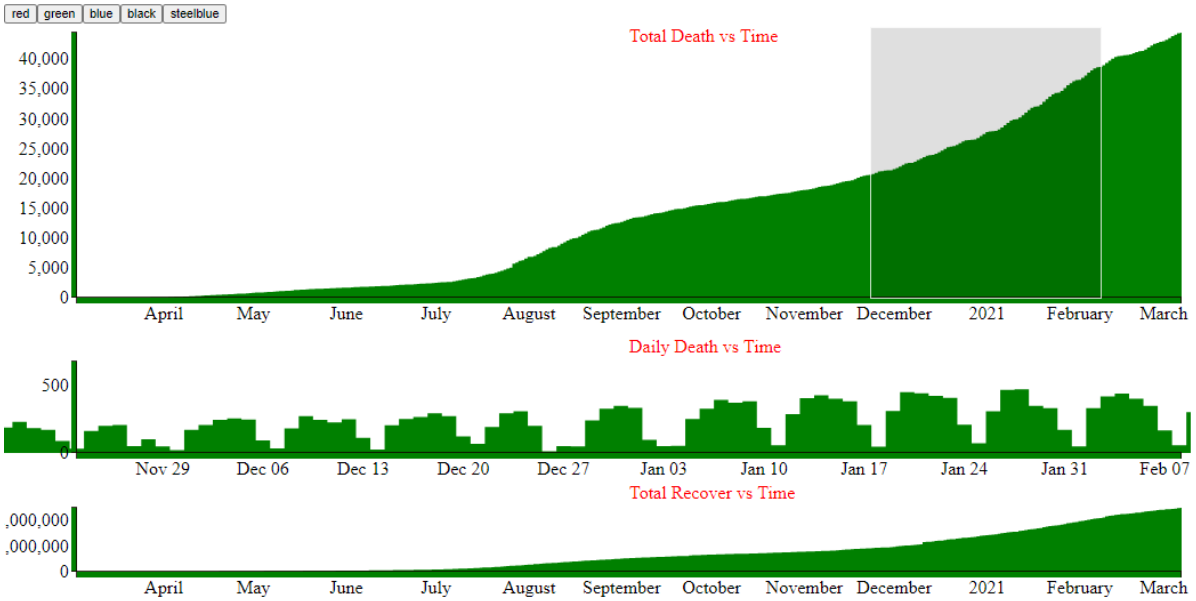
The pie chart shows the portion of the number of new positive cases each month a state has among all the available months, while the bar chart shows how the number of new positive cases is changing . By hovering over each slice, there will be a tooltip showing the total new positive cases at the bottom right corner. The colors for the slices are also scaled using d3.interpolateRed. The darker the red is, the highest the number it is representing.

Death Staticstics:

After the user enters this view, the user can see the 2-Way linked statistical graphs between “Total Death vs Time” and “Daily Death vs Time” for the state the user selected in the previous page. The data extends from the beginning of the COVID-19 pandemic to 2021 March. In this linked view the user can view the detail in one graph by brushing a section in the linked graph. (Some states do not have Total Recover vs Time, so it sometimes can be empty.)



Note: The Total Recover vs Time is not available for this state, so it is empty

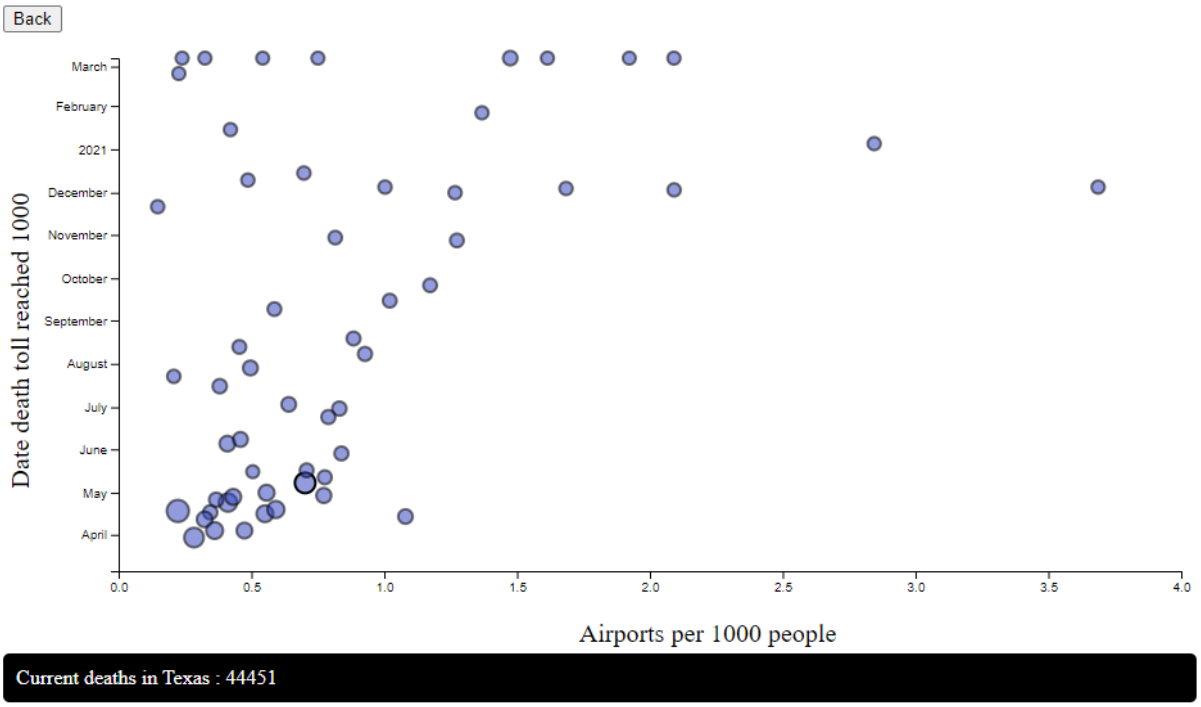


Before applying the brushing view to another graph, make sure you cleared the brush in the previous graph. The third Graph is a one way linked view between the total recover and Total death. However, some states do not have “Total Recover vs Time” information in the dataset we are using, so it might be empty. :)

Users also can change the color in this visualization by clicking the buttons on the top left.

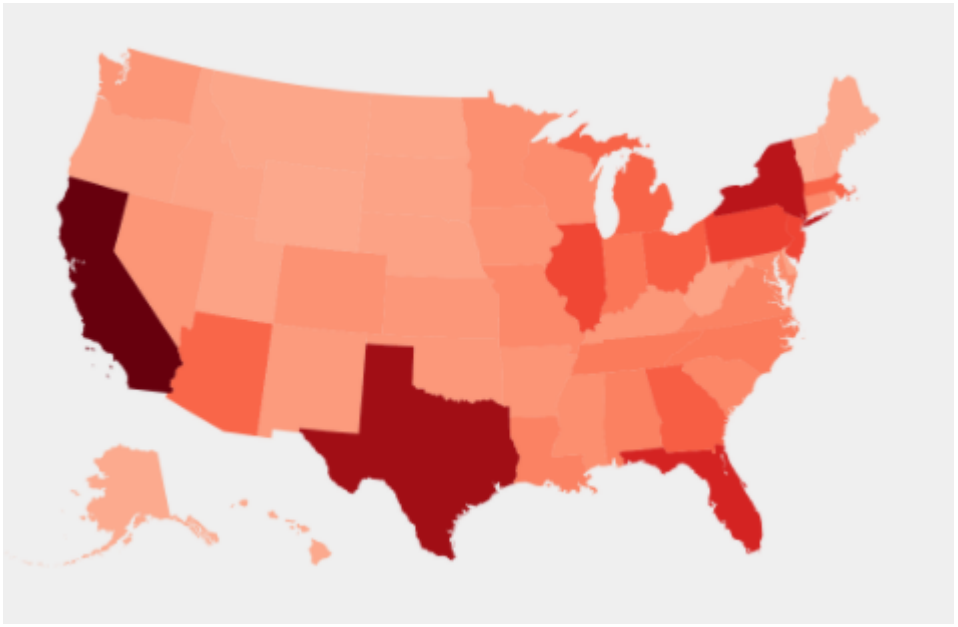
Death vs Airports:

This view shows a bubble chart which plots the number of airports per thousand people in each state against the date at which the death toll in that state reached one thousand. The bubbles vary in size slightly, with bigger bubbles corresponding to higher total death tolls. When the user moves their mouse over a data point, a popup will display the name of that state, along with the state’s death toll.

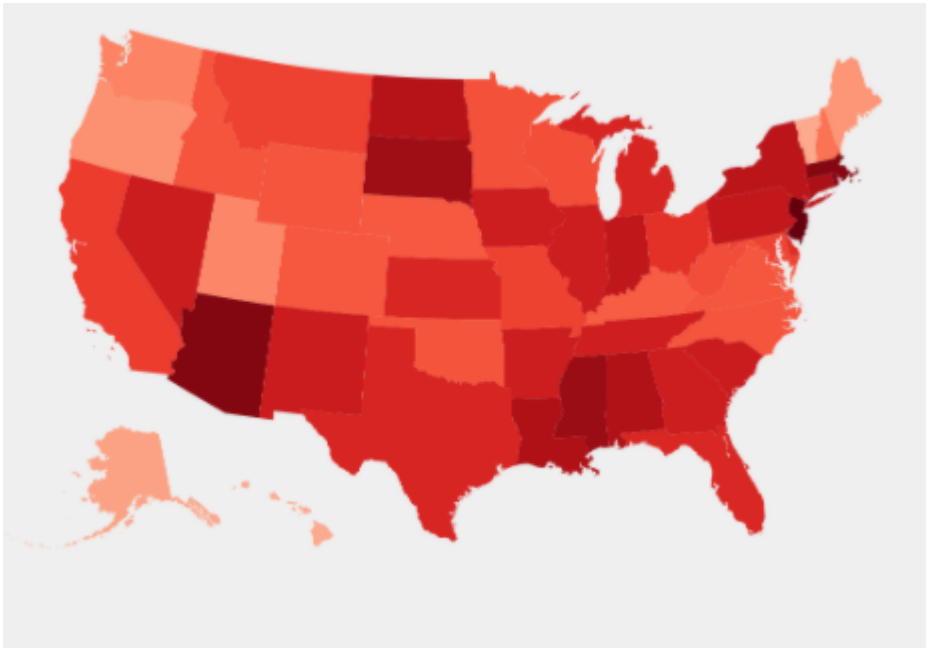


Evaluation:

In the home page, the default color on the map is showing the number of total deaths in that state.

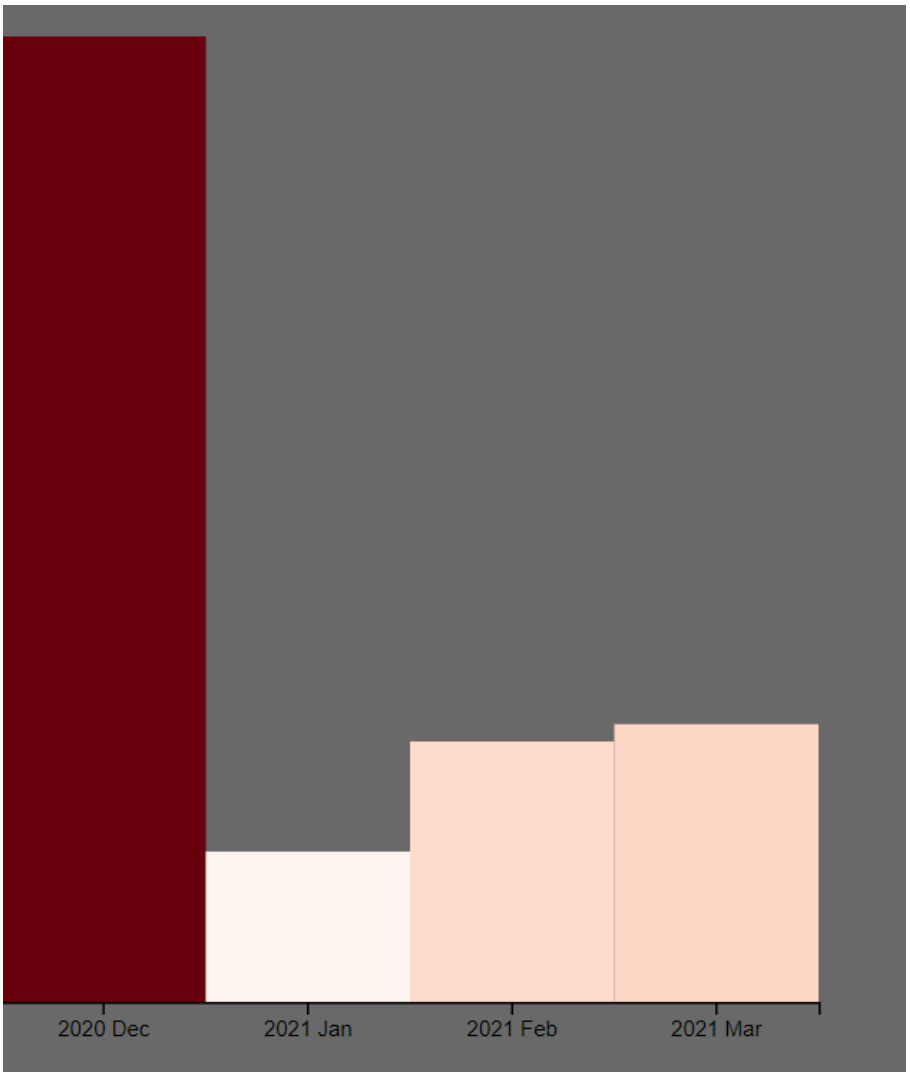


We can see that states like California, Texas and New York have the darkest colors. The darker the red is, the higher the number of total deaths (including all months) is for that state. However, it is not enough to only take a look at the total value. Different states have different populations. What will happen if we divide the total deaths by the population? By switching to the 'Death in Percentage' mode, we will see that the colors change a lot.



It turns out that states like Arizona, New Jersey and Massachusetts have darker colors than others, which implies that the death rate in these states are higher than others.

In the positive increase view, we can see that the number of new cases of people tested positive reaches a peak in December, 2020 and falls dramatically in 2021.

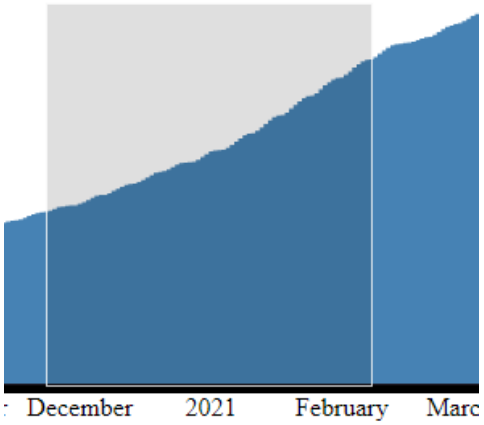


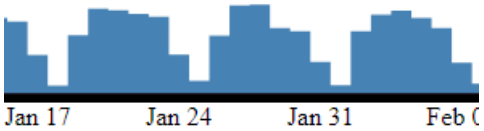
This is a correct trend, because if we refer to the official visualization posted by Johns Hopkins University (the full graph is put above in the ‘Related Work’ section), we can also see that the number of new positive cases falls in 2021.

This may be because that more and more people are taking vaccines at the beginning of 2021, but because the pandemic is still spreading, there is a slow increasing trend of new positive cases. However, in general speaking, things are getting much better in 2021.

By viewing the death recovery statistics, one interesting thing we find is that the general trend of the death number is the same in each state. One interesting thing we find is that the reported death number in a short period of time follows

a certain pattern. The period of this pattern is around 7 days.





The reported death number changes periodically, and the period is around 7 days. The death number increases after Sunday and starts decreasing after Wednesday.

In some states with a lot of people, for example California and Florida, the numbers of airports per 1000 people are low (because of too many people), but the death rates in these states are high. There is a negative relationship between the number of airports and the death rates. So there may be a stronger relationship between the death rate and the population instead of the death rate and the number of airports per 1000 people.