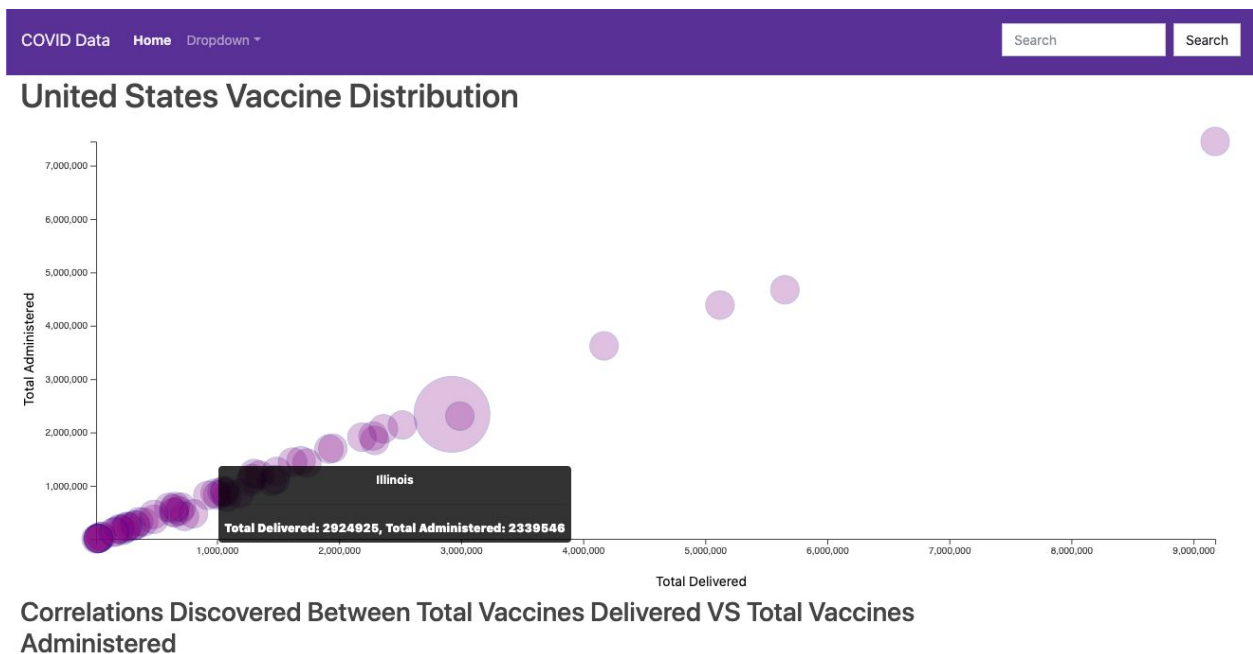


Group 5: Covid-19 Cases and Vaccination Rates

Our group was tasked with visualizing a story with data. We felt there was a story to tell with live Covid-19 data. We received our vaccine distribution and infection rate data from the CDC and Our World in Data. Our requirements were to make each visualization interactive for the user. It would have to include a Python Flask-powered API, HTML/CSS, JavaScript, and at least one database (SQL, MongoDB, SQLite, etc.) We were also required to have a combination of web scraping, Leaflet or Plotly. Lastly, everything we did needed to be contained into one dashboard page with links to the necessary visuals.

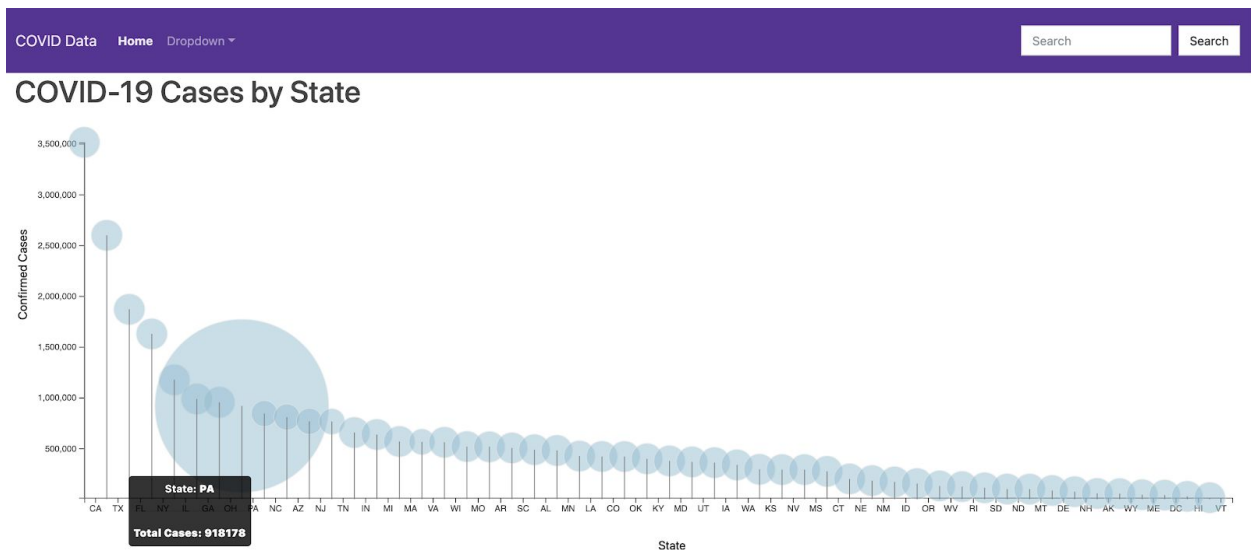
The first visualization we did was a scatter plot of vaccines allocated and administered by each state in the country. What we learned on a basic level with this data was the states with the higher populations had more positive cases and therefore would have more vaccine distribution. This proved to be correct once you independently search the states with the higher population count. The plan to distribute vaccines to the harder affected areas was shown to be working. There was some discrepancy in our data such as numbers registered by Veterans Health and the Republic of Palau was included in this dataset. We were tasked with not cleaning the data so this limited us in having our data completely clean.



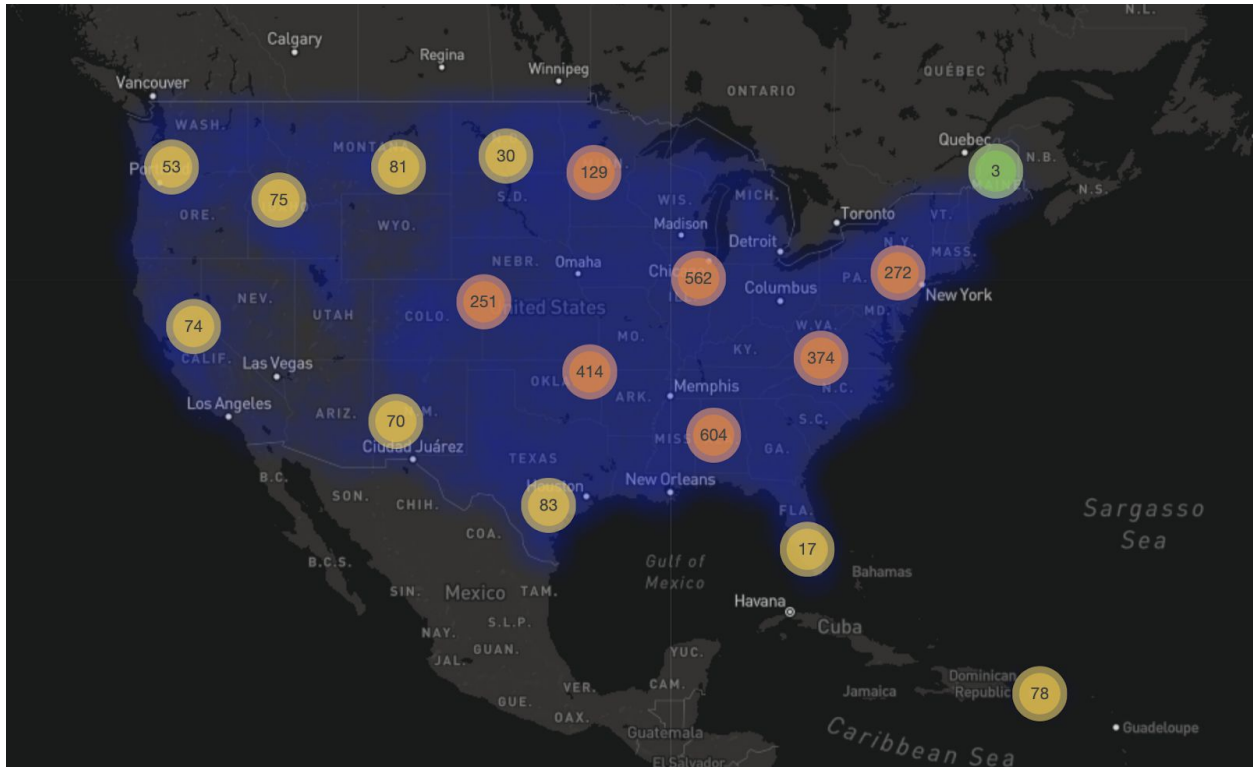
Created by Group 5 circa © 2021

Next, we created a lollipop chart using d3 to showcase how each state was doing in terms of Covid-19 cases. The purpose of doing so was to compare this data to the scatter plot and see if states that were getting more vaccines actually had more cases. While some states did match up, there were others that didn't. For example, The scatter plot ranked California, Texas, Florida,

New York, and Pennsylvania as the top 5 states that were getting the most vaccines in the country thus far. However, by looking at the lollipop chart, one can see that while the first 4 states match up, Pennsylvania actually ranked number 8 instead of number 5 for Covid cases. This may be because vaccine rollouts have been slow in some states and every week efforts are amped up. Over time, the scatter plot and lollipop chart should match up in state rankings as states with more cases should naturally have more vaccines available.

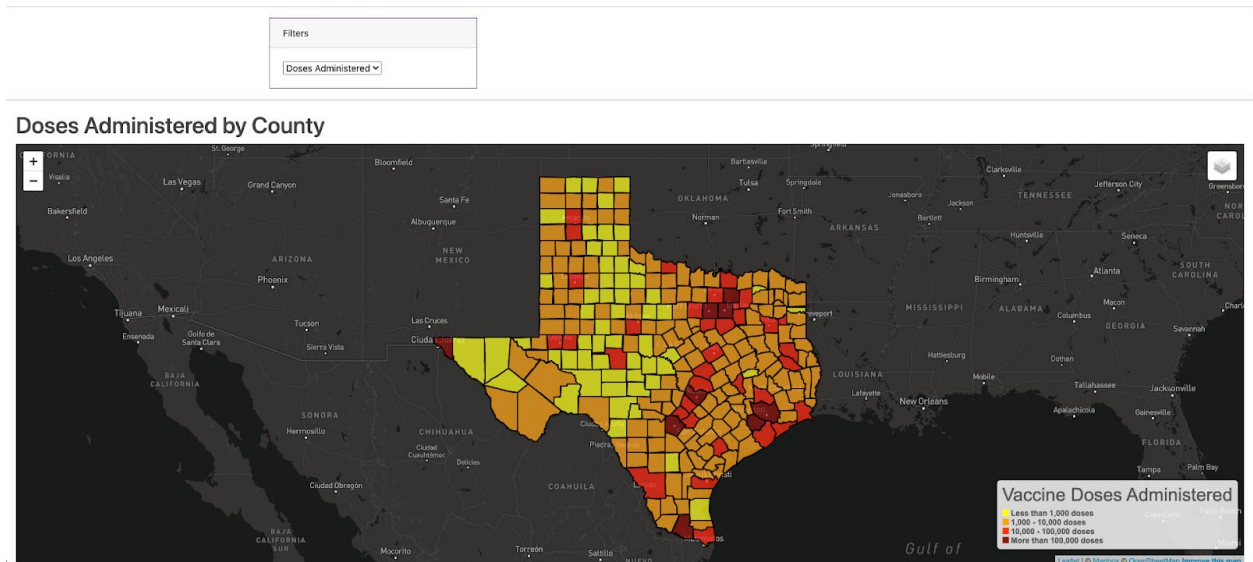


Our third visualization was a heatmap made using Leaflet. The heatmap was for the United States and showcased Covid cases. We added markers and then marker clusters to make the map look cleaner. When analyzing the heatmap, it's important to notice that the numbers and color scheme on the clusters represent how many individual markers are in that region rather than the number of Covid cases. Clicking on the clusters shows the individual markers, where each marker then has a number associated with it. These numbers are what actually represents the number of cases.

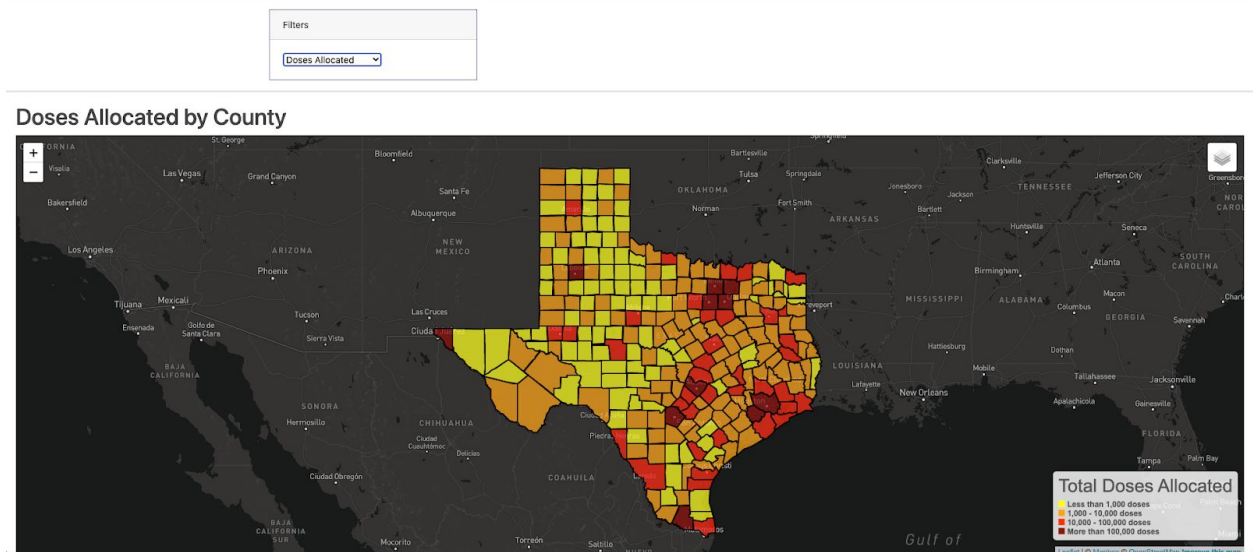


Our final visualization was a choropleth map of Texas counties which was made through Leaflet. The counties are colored by the amount of total vaccine doses administered, and when a county is clicked, the county name and dose information is displayed. This visualization showed that the highest concentration of vaccine usage is centered around the major cities, followed by the suburbs and smaller cities, with the rural areas at the bottom. We also added a filter to our map which allows the user to compare how the vaccine doses have been allocated. A similar trend was observed between vaccine doses administered and allocated.

Doses Administered



Doses Allocated



Limitations:

We encountered limitations through our work on this project. It was difficult to find APIs with geographic coordinates combined with COVID-19 data. We relied on a CSV file for the Texas counties' geographic data, for instance.

Secondly, not all areas of the country have received vaccines in significant quantities. The natural limitations of the supply chain are certainly present. It would be difficult to tell how much the supply issue has affected the overall numbers.

Lastly, the COVID-19 data included 274 countries. The data set was quite large and we struggled to filter the data at first.

Future Work:

There is an anomaly in the data which is skewing the overall picture by state. “Veterans Health”, which is the Veterans Administration, is treated like a separate state in the data. So, veterans receiving the vaccines through the VA are not counted as citizens of the state that they live in. Given more time and resources, it would be interesting to explore the degree to which state counts are under reported. States with high numbers of veterans would likely be affected more than state with smaller veteran populations.