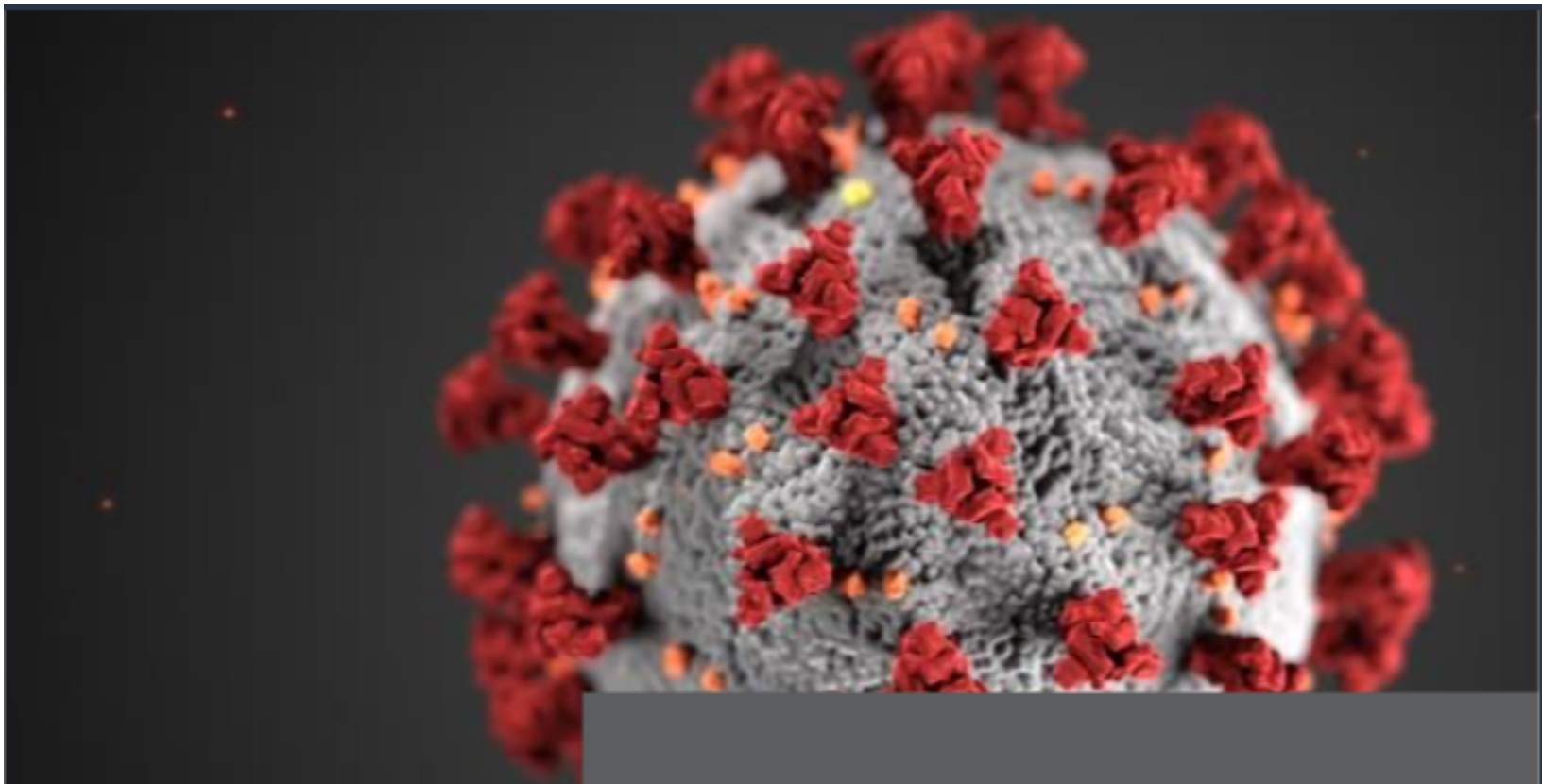


Process Book

Zhinuo Li
Wenjing Dai
Yinghui Jiang

3 June 2022

Intro-duction

Data mining usually provides unexpected insights by sorting through large datasets to identify certain patterns and relationships. While in the meantime, finding a suitable way to display the analysis result is significant. Thus, data visualization is introduced. In this project, we strived to present a visualized data story that allows viewers to interact with large amounts of data in an intuitive way.

DataSet

The main dataset we selected is from the New York Times(https://github.com/nytimes/covid-19-data). It released a series of data files with cumulative counts of coronavirus cases in the United States, at the state and county level, over time. The dataset includes cases from the first reported coronavirus case in the US on Jan. 21, 2020 till today and is ongoing. Another dataset on COVID-19 vaccinations in the United States is from Kaggle(https://www.kaggle.com/datasets/paultimothymooney/usa-covid19-vaccinations). It's state level data with detailed information about vaccination status and the distribution. The additional dataset we chose is from the US CDC (https://data.cdc.gov/Public-Health-Surveillance/Rates-of-COVID-19-Cases-or-Deaths-by-Age-Group-and/3rge-nu2a). It tracks rates of COVID-19 cases and deaths by vaccination status. It reflects cases among persons with a positive specimen collection and deaths among persons with a positive specimen collection.

Pre-processing

To provide a general overview of the data, we presented national level and state level information by filtering out county level data. Also, the dataset includes overseas territories. Considering the visualization of the interactive map, we only focused on 50 states of continental United States. To merge the first two dataset, we selected the overlapped time period. Besides, the dataset contains some duplicate and unreliable data. We removed them by checking the ascending trend. Although some are due to statistical revision by the organization, the final dataset is believed to provide solid insights. To better dig in, we performed temporal and categorical analysis to give out a general view of US COVID-19 outbreak.

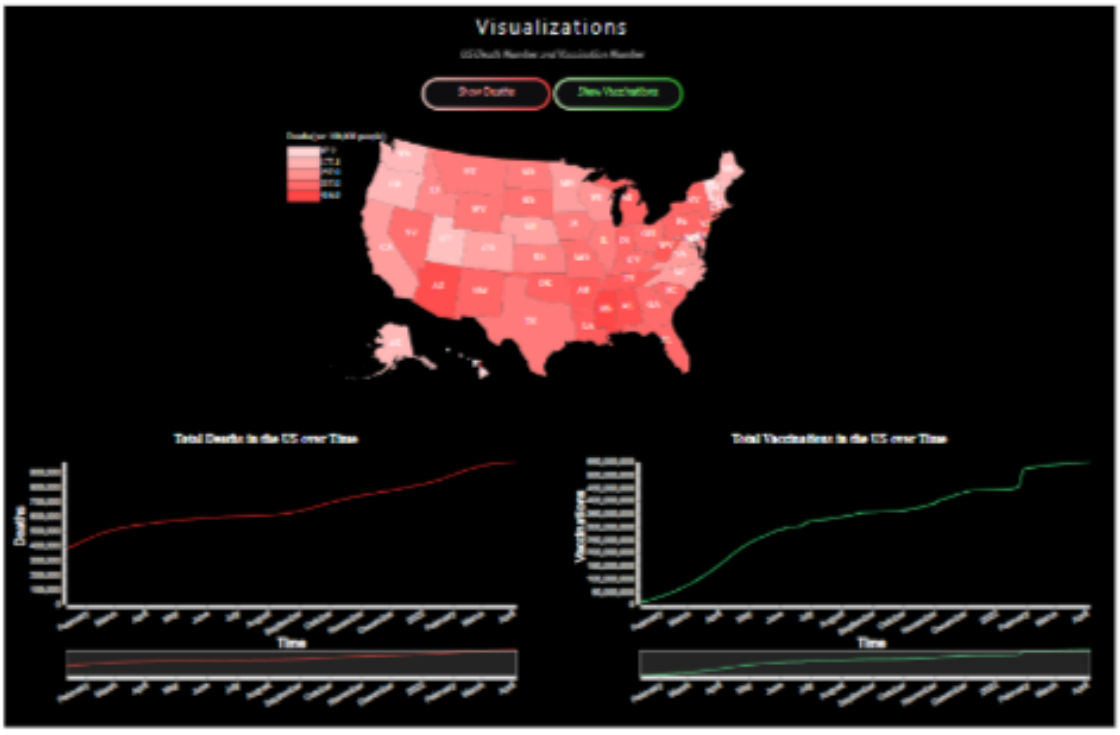
Implementation and Final Product

Core Visualization

The core visualization has a map graph and two line graphs. All three charts are first implemented using D3 separately and later lined together.

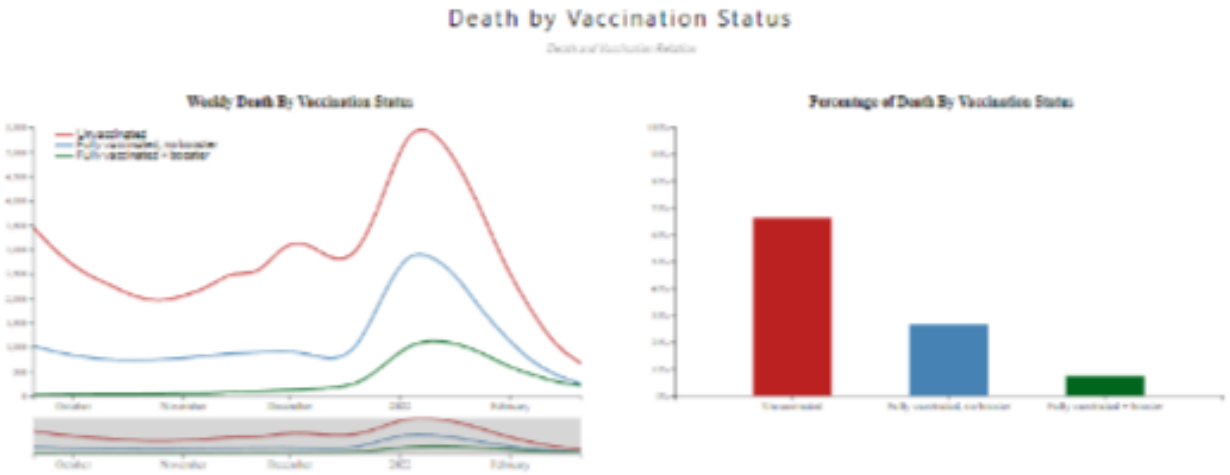
Ideas in lecture 08 helped us to implement the map graph. We used a GeoJson file to create the map and a choropleth map to color each state according to the data. Additional buttons were implemented so that the map could be colored by either the number of deaths or vaccination. Initially, all three graphs show the total data of all states in the United States. When a user selects a state on the map, all three charts change their representations using event listeners. The map changes the view box to only focus on the selected state, and line graphs only show the data for the state chosen using the filtering function. In addition, we implemented animations to make the transition smooth.

The axes of the line graphs use scaling and the d3.max feature, which also updates according to different data for different states. In addition, the line graphs have the brush and zoom function that interacts with mouse scrolling and selection.



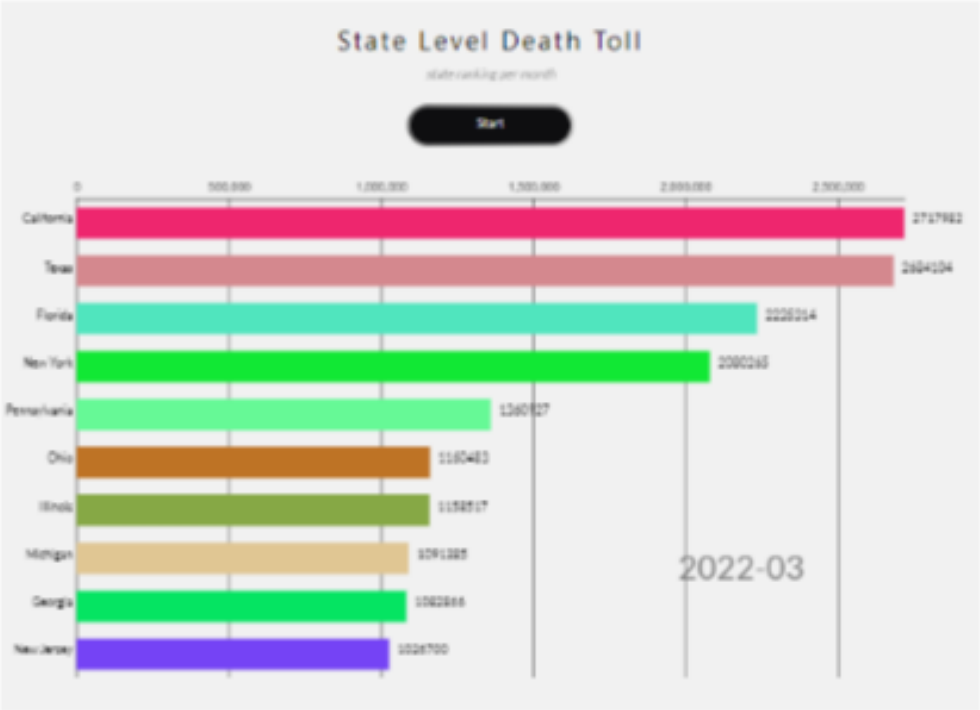
Vaccination Status vs. Deaths Graph

Using D3.js, we implemented a linked line and bar graph that showcases the effectiveness of vaccination against covid death. The line graph shows the general trend of weekly death over time, categorized by the three different vaccination status. The bar graph further shows the percentages of deaths that each category accounts for, which makes the difference more obvious and visible. We first implemented the two graphs separately using D3. We used ideas from Exercise 05 to implement zoom and brush of a basic line chart, and then added another event listener for them that animate the changes in percentage of deaths related to different vaccination status. We considered layering to contain three lines on the line chart.



Bar Chart Race Animation

Using D3.js, we implemented a linked line and bar graph that showcases the effectiveness of vaccination against covid death. The line graph shows the general trend of weekly death over time, categorized by the three different vaccination status. The bar graph further shows the percentages of deaths that each category accounts for, which makes the difference more obvious and visible. We first implemented the two graphs separately using D3. We used ideas from Exercise 05 to implement zoom and brush of a basic line chart, and then added another event listener for them that animate the changes in percentage of deaths related to different vaccination status. We considered layering to contain three lines on the line chart.



Summary

The final product is a website consisting of three main visualizations. The core part is an interactive map linked with line graphs. It could both display the US level and the state level data of vaccination and deaths by clicking on different parts. The user could zoom in or zoom out to check the US Covid information of a specific period by manipulating the time slider below it.

The second visualization shows deaths of different vaccination statuses, the line graph and bar chart are linked. The user could also explore it by manipulating the time slider.
The final one is a bar chart race animation. It provides a state level ranking of US Covid deaths for each month. The temporal evolution gives hints on how fast the pandemic spread and which state lost most in different stages.

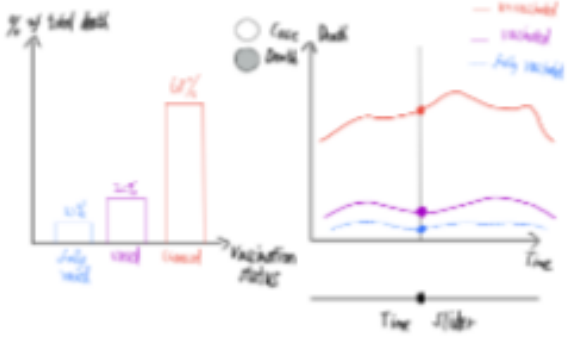
Challenges

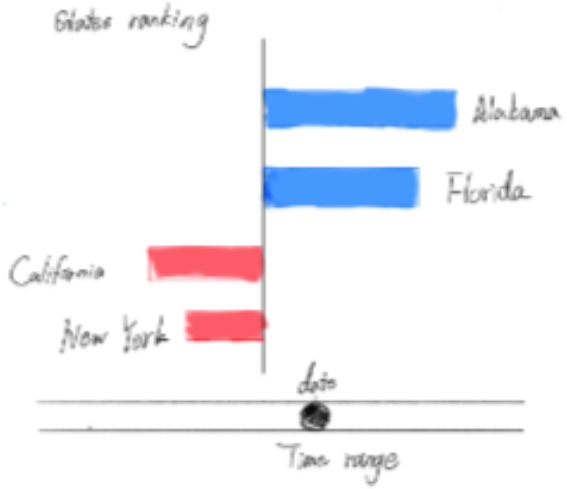
We've encountered both design and technical problems during implementation.
Firstly, during the design stage, by intuition, we assumed vaccination may probably have correlation to covid deaths. Thus, we decided to display the causal relationship between them through multiple visualizations. We tried to prove that getting vaccinated could effectively lower the death rate. However, after preliminary data processing and analysis, we surprisingly found they may not have significant connection. This striking fact determined the shift of our project goal and implementation. Although we still included deaths of different vaccination status by adding additional dataset, we finally decided to present a dashboard that provides detailed US covid information in general.
Secondly, each of us worked independently on different visualizations. Although each visualization runs fluently, in the end, we found that our styles are not aligned. It's what we ignored when we did the designing. So we decided to align the global style and finalize it.
Thirdly, for the technical part, it's quite different to test the website locally and in different browsers. We tried hard to fix this problem. This is what we didn't realize at the very beginning.
Fourthly, we planned to add a navigation bar for the interactive map in the sketches. However, we found this feature may be similar to clicking on different regions of the map. Because they would show the same graphs. So we decided to quit it in the end.
Finally, compared to the original design, we changed the last bar chart with a time slider into a bar chart race animation. Because we realized the first two visualizations are already the same type. To enrich our design, we decided to add a different type, so that we have interactive maps, linked graphs and animation in the end.

Changes

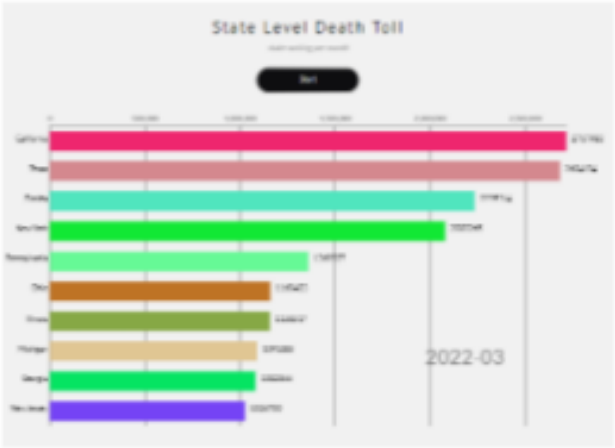
Here are the comparisons between our original design and the final product.

Initial Drafts





Finial Views



Peer Assessment

The work is reasonably distributed among us three. All members actively participated in discussion and made decisions together during the design, implementation and testing stage. Each of us mainly works on different visualizations. The final result is contributed by all.

Wenjing Dai

Website design, Style, Interactive Bar and line graph, Interactive Map and linked graph, Process book implementation section, Screencast voice, Data cleaning

Yinghui Jiang

Website design, Data cleaning, Preliminary analysis, Bar chart race animation, Process book content, Global style

Zhinuo Li

Website design, Website skeleton, Interactive Map and linked line graphs, Screencast edit, Process book implementation section, Website Style, Process book design, Data cleaning