CS416: Data Visualization

Narrative Visualization

Madhumitha Sivaraj, sivaraj4@illinois.edu

University of Illinois, Urbana-Champaign

I. Narrative Visualization
My narrative visualization seeks to explore vaccine hesitancy and its impact on the spread of COVID-19 cases. To start viewing the interactive slideshow, please click on "1" or the "right arrow" on the navigation bar. Use the "right arrow" to move to the next scene or use one of the numeric buttons (1, 2, 3, 4) to skip directly to one of the four scenes.

Users can access the narrative visualization at: https://madhusivaraj.com/nv/.
My codebase is located at: https://github.com/madhusivaraj/nv.

Data Sources used (public datasets):
- https://aspe.hhs.gov/reports/vaccine-hesitancy-covid-19-state-county-local-estimates (made available by U.S. Department of Health and Human Services)
- https://github.com/nytimes/covid-19-data (made available by The New York Times)
- https://github.com/BloombergGraphics/covid-vaccine-tracker-data (made available by Bloomberg)

II. Messaging
COVID-19 is a dangerous disease caused by a virus discovered in December 2019 in Wuhan, China. COVID-19 most often causes respiratory symptoms that can feel much like a cold, a flu, or pneumonia, but COVID-19 can also harm other parts of the body. It is very contagious and has quickly spread around the world, resulting in over 198 million cases and 4.22 million deaths globally. According to the World Health Organization (WHO), equitable access to safe and effective vaccines is critical to ending the COVID-19 pandemic, so it is hugely encouraging to see so many vaccines proving and going into development. But it's not vaccines that will stop the pandemic, it's vaccination. Various governmental agencies are working to ensure fair and equitable access to vaccines, and ensure every country receives them and can roll them out to protect their people, starting with the most vulnerable. However, a significant portion of the American population has not yet been vaccinated, citing vaccine hesitancy as one of the primary reasons.

Through my narrative visualization, the message I am trying to communicate is that states with higher vaccination rates and lower vaccine hesitancy are more likely to achieve herd immunity from COVID-19, thereby reducing COVID-19 cases on a greater scale than states with low vaccination rates and high vaccine hesitancy. Hence, greater vaccine hesitancy has a negative impact on the spread of COVID-19 cases.

I use various data sources that including information including COVID-19 rolling data, vaccine trackers, vaccine hesitancy estimates to develop four scenes, each scene building off one another. The first two scenes are a bar chart that displays estimates of COVID-19 vaccine hesitancy rates across U.S. states and another bar that displays the percentage of populations that are fully vaccinated. Analyzing the first two scenes helps users identify the top six most vaccine-receptive states with the highest vaccination rates and the top six least vaccine-receptive states with the lowest vaccination rates. The third and fourth scenes focus on that insight and present line graphs of the rolling averages of cases per 100K individuals for the top six most vaccine-receptive states with the highest vaccination rates and the top six least vaccine-receptive states with the lowest vaccination rates. The last two scenes enable users to drill down on the six states in each set, enabling users to make comparisons between the two sets of six states and reach the conclusion that greater vaccine hesitancy has a negative impact on the spread of COVID-19 cases.

III. Narrative Structure

The narrative visualization was designed to follow the interactive slide show structure. This hybrid structure follows an author-directed path through a slideshow. Users can navigate a slideshow using a navigation bar or "slide advance" on the top left portion of the visualization. User exploration is allowed at some of the steps of the story.

My project starts off with an index page that provides context about COVID-19 and a summary of the purpose of the visualization and how to get started. Users can use the right arrow on the navigation bar to proceed to the Scene 1, and the following three scenes after, of the narrative visualization. In Scenes 1 and 2, users can investigate some tangent they are interested in by exploring details with a mouse-over via the tooltip or "details-on-demand" I included in the bar graph. Users can also choose just to continue on if they are not interested in the details of those two slides. Scenes 3 and 4 offer users the opportunity to "drill-down" and explore the various progression of COVID-19 cases over a 17-month timeframe for various states. Using the drop-down button provided under the slide advance on scenes 3 and 4, the graphic changes to reflect the progression of COVID-19 cases for the various state. Users can choose to utilize this feature to explore the rolling averages for various states on scenes 3 and 4, or they can advance to the next scene or reach the conclusion.

My narrative visualization enables moving from slide to slide, with options to investigate further in-between. The narrative structure I follow gives users the opportunity to drill down in a particular slide and then it takes them to the next slide after users have "drilled down" into a detail.

IV. Visual Structure

The visual structure for the four scenes is as follows:

- <u>Scene 1:</u> I selected the vertical bar chart type because I wanted to display comparisons between categories of data. The charts display rectangle bars with lengths proportional to the values that they represent. In addition, I wanted to display independent ordinal data (state names) with dependent, discrete quantitative data (percentage of the population vaccinate-hesitant). The view is filtered by vaccine-hesitant population percentage, displaying states in order of percentage of the population most vaccinate-hesitant to least vaccine-hesitant. The bar chart ranges from 4.026% to 25.139% in descending order. Users can hover over each bar, assigned to a single state, and view information such as the percentage of the population that is vaccine-hesitant and its CVAC level of concern.

- Scene 2: I selected the vertical bar chart type because I wanted to display comparisons between categories of data. The charts display rectangle bars with lengths proportional to the values that they represent. In addition, I wanted to display independent ordinal data (state names) with dependent, discrete quantitative data (percentage of the population fully vaccinated). The view is filtered by vaccination rank, displaying states in order of percentage of the population most vaccinated to least vaccinated. The bar chart ranges from 28.501% to 62.938% in descending order. Users can hover over each bar, assigned to a single state, and view information such as the percentage of the population that is vaccinated and its vaccination rank.
- Scene 3-4: I selected the line chart type because I wanted to display the progression of continuous quantitative data over a timeframe. Line graphs are used to track changes over short and long periods of time. When smaller changes exist, line graphs are better to use than bar graphs. This is the best fit because it can also be used to compare changes over the same period of time across various states featured in the dropdown menu.

I also use the same scene template (header, slide advance, scene with annotations, summary at the bottom of the page) to promote a consistent visual structure and keep users oriented through transitions. Each scene is a single chart or a coordinated set of single charts via a dropdown menu. My annotations also promote visual consistency, each following a similar template. The annotations have clean and straight edges, gray font color, text format that includes a bolded title and description that follows beneath the title.

Color was a huge element in my visualization. I even added a soft off-white background color in my map chart to bring visual interest to my dashboard. I decided to not use a lot of blue in my visualization because the human eye has the most difficulty focusing on that color as opposed to red, green, and yellow (all colors I use in my charts). This way, all the colors are broken up and each chart does not get washed up from a visual or aesthetic point of view. In each chart, I use the same color (red for the bar chart and grey/black for text in Scene 1, yellow for the bar chart and grey/black for text in Scene 2, and an assorted colors for the line charts and grey/black for text in Scenes 3-4) to bring some consistency. I also tried to use warmer colors where possible because humans tend to focus better on warmer colors and bring them to the forefront. I opted for warmer tones of yellow for the bar chart and colors like off-white and soft grey in the graphics within "details on demand".

All scenes in the narrative visualization contain tooltips (details on demand), annotations and brief summaries at the bottom which ensure the user can understand the data and navigate the scene. The annotations also urge the viewer to focus on the important parts of the data in each scene. The slide advance helps viewers transition to other scenes. The annotations and bottom summaries help users to understand how the data connects to the data in other scenes, as my written analyses presented in both the annotations and summaries for all four scenes reference other scenes explicitly.

V. Scenes

My narrative visualization includes a landing page and four scenes. My four scenes are ordered in such a way that information from the previous scenes is essential in understanding the information on the current scene and for drawing a conclusion from all the information in the visualization.

When users go to https://madhusivaraj.com/nv/, they are met with a landing page which provides context about COVID-19, vaccines, and vaccine hesitancy. These details are intended to provide users with no knowledge of the COVID-19 pandemic with background information such that users have necessary understanding to navigate and process the narrative visualization. The landing page also provides brief information on how to navigate the interactive slide show.

<u>Scene 1:</u> The first scene of the narrative visualization is a bar chart with estimates of COVID-19 vaccine hesitancy rates across 50 U.S. states using data from the U.S. Census Bureau’s Household Pulse Survey. I decided to put this scene first because my narrative visualization is about vaccine hesitancy and its impact on the spread of COVID-19 cases, and before exploring its impact, I felt that it was necessary to understand the current position of vaccine hesitancy across all states. The x-axis of the bar graph contains the 50 states which are independent variables. The y-axis is the percentage of population that is hesitant about vaccines. When users hover over each bar, a tooltip or “details on demand” pop up and display information like the exact percentage of the y-axis and the CVAC level of concern for that state. Scene 1 also includes two annotations, highlighting the top 10 vaccine-hesitant states and top 10 vaccine-resistant states. The annotations also provide geographical insights and information to support the findings that the top 10 states with a high level of concern about vaccine hesitancy are primarily in the South, Midwest, and Rocky Mountain regions, which are more conservative and rural. Meanwhile, the top 10 vaccine-receptive states all have low or moderate levels of concern regarding vaccine hesitancy. All these states, besides Hawaii, are in the Northeast. It is important to note that the COVID-19 vaccines that have received EUA in the US were developed by Pfizer in New York, Moderna in Massachusetts, and Johnson & Johnson in New Jersey, and this information may contribute to the skew in vaccine hesitancy based on geographical location. Below the graph is some text about my gathered insights and observations from the bar chart.

<u>Scene 2:</u> The second scene of the narrative visualization is a bar chart that displays the percentage of populations that have received the COVID-19 vaccines using data provided by Bloomberg's Covid-19 Vaccine Tracker. The x-axis of the bar graph contains the 50 states which are independent variables. The y-axis is the percentage of population that is fully vaccinated. When users hover over each bar, a tooltip or “details on demand” pop up and display information like the exact percentage of the population that is vaccinated and the Vaccination Rank for that state. This scene follows the previous scene about vaccine hesitancy, enabling users to analyze commonalities between the rankings of states based on vaccine hesitancy and vaccinated populations. Scene 2 includes two annotations, highlighting the top 10 states with the highest vaccinated populations and top 10 states with the lowest vaccinated populations. The annotations

also provide my findings regarding overlaps between 1) the top 10 vaccine-receptive states (scene 1) and the top 10 states with the highest vaccinated populations (scene 2); and 2) the top 10 vaccine-hesitant states (scene 1) and the top 10 states with the lowest vaccinated populations (scene 2). Thus, users can identify the top six most vaccine-receptive states with the highest vaccination rates and the top six least vaccine-receptive states with the lowest vaccination rates. Below the graph is some text about my gathered insights and observations from the bar chart.

<u>Scene 3:</u> The third scene of the narrative visualization is a line graph above displays the rolling averages of cases per 100K individuals across the 6 most vaccine-receptive states with the highest vaccination rates, using data provided by The New York Times. This scene follows the first two scenes because it builds off the insights gathered from Scenes 1 and 2. In Scenes 1 and 2, we identify the top six most vaccine-receptive states with the highest vaccination rates and the top six least vaccine-receptive states with the lowest vaccination rates. Scene 3 maps the progression of COVID-19 cases over a 17-month time frame for the first set of six states, the top six most vaccine-receptive states with the highest vaccination rates. The x-axis of the line graph is a timeseries from February 2020 to June 2021. The y-axis is the average cases per 100,000 individuals. I included a dropdown menu of all six states, enabling users to view the progression of COVID-19 cases for the selected state. This provides users with the option to either view one example, Connecticut, and the option to the progression for any of the other 6 most vaccine-receptive states with the highest vaccination rates. Scene 3 includes one annotation, providing information regarding the average number of cases per 100K, the total number of cases, the average number of deaths per 100K, and the total number of deaths for a state on the last day that data was provided. This annotation provides information to compare the latest developments of COVID across various states. Below the graph is some text about my gathered insights and observations from the bar chart.

<u>Scene 4:</u> The fourth scene of the narrative visualization is a line graph above displays the rolling averages of cases per 100K individuals across the 6 least vaccine-receptive states with the lowest vaccination rates, using data provided by The New York Times. This scene follows the third scene because it builds off the insights gathered from Scenes 1 and 2, and Scene 3. In Scenes 1 and 2, we identify the top six most vaccine-receptive states with the highest vaccination rates and the top six least vaccine-receptive states with the lowest vaccination rates. Scene 3 maps the progression of COVID-19 cases over a 17-month time frame for the first set of six states, the top six most vaccine-receptive states with the highest vaccination rates. Scene 4 enables users to see the progression of COVID-19 cases over a 17-month time frame for the second set of six states, the least vaccine-receptive states with the lowest vaccination rates. Scene 3 precedes Scene 4 because the latter relies on insights from comparisons between these last two scenes to draw a conclusion between vaccine hesitancy and the spread of COVID-19 cases. The x-axis of the line graph is a timeseries from February 2020 to June 2021. The y-axis is the average cases per 100,000 individuals. I included a dropdown menu of all six states, enabling users to view the progression of COVID-19 cases for the selected state. This provides users with the option to either view one example, Alabama, and the option to the progression for any of the other 6 least

vaccine-receptive states with the lowest vaccination rates. Scene 4 includes one annotation, providing information regarding the average number of cases per 100K, the total number of cases, the average number of deaths per 100K, and the total number of deaths for a state on the last day that data was provided. This annotation provides information to compare the latest developments of COVID across various states. Below the graph is some text about my gathered insights and observations from the bar chart.

VI. Annotations

I followed the susielu/d3-annotation (https://d3-annotation.susielu.com/) framework because they were visually consistent. My annotations follow a similar format, with clean and straight edges, gray font color, text format that includes a bolded title and description that follows beneath the title.

The annotations in scene 1 are used to section off the top 10 vaccine-hesitant and top 10 vaccine-receptive states. The annotations in scene 2 are used to section off the top 10 states with the highest vaccinated populations and top 10 states with the lowest vaccinated populations. The details in the annotations on scene 1 provide geographical insights I realized after analyzing commonalities between the top and bottom 10 states on the bar chart. These generalizations I formed from scene 1 are supported geographical similarities identified between 1) the top 10 vaccine-receptive states (scene 1) and the top 10 states with the highest vaccinated populations (scene 2); and 2) the top 10 vaccine-hesitant states (scene 1) and the top 10 states with the lowest vaccinated populations (scene 2). The annotations on scene 1 and scene 2 help users can identify the top six most vaccine-receptive states with the highest vaccination rates and the top six least vaccine-receptive states with the lowest vaccination rates.

The annotations in scenes 3 and 4 provide the average number of cases per 100K, the total number of cases, the average number of deaths per 100K, and the total number of deaths for a state on the last day that data was provided for 6 most vaccine-receptive states with the highest vaccination rates and 6 least vaccine-receptive states with the lowest vaccination rates, respectively. These two annotations provide information to compare the latest developments of COVID across various states. These annotations coupled with the annotations from the first two scenes enable users to process all insights and reach a conclusion regarding the direct correlation between low vaccine receptiveness and a decline in COVID-19 cases. Thus, the annotations support the messaging that greater vaccine hesitancy has a negative impact on the spread of COVID-19 cases.

The annotations change within a single scene for only scenes 3 and 4. This is because both scenes offer different line graphs of the progression of COVID based on the state selected in the dropdown menu. To have annotations stay consistent would not make sense as case and death data varies across all states. My annotations are cleared between every scene.

VII. Parameters

The parameters are the variables that are used in the charts to control the scene and the elements in the chart, and everything else. In scene 1, my parameters are the 50 U.S. states and the percentage of population hesitant to get the vaccine, but when I transition to scene 2, the latter parameter changes to the percentage of population that is vaccinated. Another parameter in my visualization can be seen on both scenes 3 and 4. In both scenes, I display the rolling averages of cases for the current state in a state machine, based on the state selected in the dropdown menu. When I select a state in the dropdown menu, that U.S. state becomes “the current state” of the narrative visualization and this parameter triggers a chart unique to the state to be displayed. This chart is a timeseries progression of COVID-cases in that U.S. state over a 17-month timeframe, with an annotation providing information pertaining to that U.S. state.

VIII. Triggers

Triggers are the connections between the parameters. When one parameter changes, it can cause another parameter to change as a result and that first parameter is triggering a change in the second parameter. In scenes 3 and 4, I have a dropdown menu so depending on which state the user selects, it impacts the behavior of the data my D3 code uses to produce a line graph which displays the progression of COVID cases over time for that U.S. state. So, if we do not change the default option of the menu bar in Scene 3 (Connecticut), a line graph pertaining to Connecticut will display, including annotations pertaining to COVID-related information in Connecticut. However, if we changed the option to say New York, a line graph for New York will appear as will relevant annotations. The on-event handler uses the data value (the parameter, state name) to filter the data visualized using line charts.

My visualization also contains other window events in my interactive slideshow, including mouseover and mouse out to display tooltips (or details on demand). To create these, I pass X and Y axis parameters in my code to inform where to display the tooltips for scenes 1 and 2. Each button in my slide advance also has affordances using color and setting the current slide as “active” to help the user know which buttons are clickable and which slide they are currently on. Similarly, the dropdown menu provides a concrete list of states for which graphs can be generated in scenes 3 and 4 which is another affordance to let the user know which states they can explore for those scenes.