COVID-19 Case Analysis

Caitlin Chao, Michelle Lim, Katerina Thano, Karen Xu

Northeastern University

Abstract

Since its first discovery in late 2019, COVID-19 has swept through much of the world. It changed the way we interacted with each other. Thus, for our project, we want to analyze the impact it had on the human population through cases, deaths, and vaccinations. We want to analyze specific locations and timeframes that are of interest to our potential users. Our users will be able to quickly see the USA's trends with COVID cases and handling, along with a state-by-state analysis of such information. This enables users to have a more specific, reliable, and detailed version of COVID data that they can easily comprehend and trust.

Our GitHub link is included below:
https://github.com/DS4200-S22/final-project-cm2k

– Introduction

The severe acute respiratory syndrome 2 (SARS-CoV-19) was first reported in Wuhan, China. It quickly spread worldwide and, at the time of this paper, claimed almost six million lives. In the US, each state handled their own cases, vaccinations, and deaths differently. We want to analyze how different ways of handling cases affected the spread of the pandemic and its toll.

We believe that this state-by-state analysis is critical for future pandemic prevention. We want to investigate if strict mandates, such as vaccine, mask, or quarantine mandates actually slowed the spread of COVID, resulting in less deaths and/or cases. At the same time, we want to know if mandates that were less strict contributed to or caused more deaths and/or cases.

Our data and analysis are meant to summarize the different effects the varying mandates had on each state's cases/death as a proportion of the state's population. We believe that this information is critical to future generations. If we find a correlation between the strictness of a pandemic mandate and the cases/deaths count, we could use this information to minimize, or even prevent, the next widespread disease.

We hope that our visualizations and conclusions will lead to widespread social changes. Especially since the pandemic (at the time of this paper) is still an ongoing problem, we hope that our data analysis not only helps prevent future pandemics, but also change the way we are currently handling ours. With the rise of protests regarding masks and legislation regulating vaccine passports, we hope that our data can either support or protest such policies.

Finally, we wanted to study this data for our own interests. As a group, we have never had a "normal" year of college. Our college experience centered around remote work, recorded lectures, and asynchronous classes. Unlike previous years, we were thrown into a chaotic pandemic during a pivotal point in our lives. We wanted to analyze this data as a tangentially related search for an explanation and closure.

– Related Work

A study by Mora-Araus, Velastegui-Montoya, Jaramillo-Lindao, and Apolo explores the audio landscape of COVID-19 and how it changed over time as a result of lockdown [1]. This visualization stands out in particular because of its use of the subject matter. Not many subject matters can utilize large-scale study of sounds; the result visualization stood out compared to a lot of the others that appeared while researching. This study provided a look at a concentrated area in an Ecuadorian city. The resulting dashboard organized the data neatly and used various colors to set parts of the data apart from one another. While we noted that we hope our final product will look like this, we also wonder if a similar study has been conducted on an area in the United States. The unique study gives us a chance to make our project stand apart in how in-depth it goes into the Coronavirus and its impact on society.

Wang R., Guo, Li, Jiao, and Wang L. conducted about physical risk factors for patients who could contract COVID-19, namely with regards to genetic data [2]. It used a variety of visualizations to communicate the information presented. While there were line graphs and histograms, we really enjoyed the use of matrices and the graph of importance scores. In particular, the latter had the same bars split into different colors, whereas usually in bar graphs there are two separate bars side by side with one another. Both of these diagrams were extremely effective in communicating the different symptoms that those infected with COVID face as well as comorbidity with other chronic diseases. There was a balance between simpler and complex graphs, and nothing was ever overdone to make it extremely fancy or flashy. The visualizations in this study give us a good idea of all the visualizations we can use in our final presentation.

The unique nature of the study conducted by Zhao, Tu, Fang, Wang, Huang, Xiong, and Zheng immediately made it stand out from other studies. While this paper still dealt with COVID-19, it took a unique approach and examined the optimization of material delivery during the pandemic [3]. In doing so, it employed many unique visualizations that demonstrated the unique nature of the subject matter. We loved the spatial diagrams visualizing the COVID infection rates and estimated material delivery demands. In addition, smaller diagrams such as the one visualizing infection risk along a delivery route helped zero in on the risk of COVID in everyday life. While we plan to look at COVID on a larger scale, it was an important reminder that COVID has impacted people's day to day lives. This could be an additional visualization to include in our report to help users have a more personal stake in the data.

A study by Afzal, Ghani, Jenkins-Smith, Hadwiger, Ebert, and Hoteit was particularly interesting because it explores how exactly COVID-19 might have been spread throughout different types of demographics and groups of people [4]. It even took into account how the spread of COVID-19 affected available hospital beds, media alerts, school closures, and more. A specific visualization points out the individual cities within Oklahoma using color and through a contrast channel. Since the human eye will perceive differences more easily, Tulsa stands out since it saw the largest

rise in cases and other COVID-19 related fields. For our project, we may need to utilize color and contrast to emphasize specific points of our visualization, just like they did. We will also need to consider a number of related data points that are not just deaths and vaccinations. The broad scope of this study and the use of contrast in this reference will help to enhance our project as a whole as we continue to dive into our research of COVID-19 and its effects on people.

A research journal by Zhang, Sun, Barua, Bertini, Padilla, and Parker contained numerous helpful visualizations, but one that really stood out was a visualization comparing COVID-19 to the seasonal flu [5]. To most people, the seasonal flu is not considered life threatening, so this visualization brings a familiar illness to light and compares it to COVID-19 which is not very well known. Other included visualizations have interesting marks to convey the data. For example, using the outline of a human body to represent the number of people affected and using squares in a grid, like a calendar, to represent incubation time helps users to make automatic connections to the information at hand. The marks chosen make the most sense in their contexts, and this can be integrated into our project as well. When choosing certain marks to represent certain points of data, we can make sure that the marks have meaning within themselves and convey a message instead of just being a simple shape.

– USE CASE

To demonstrate the value of gathering COVID-19 data, a possible scenario could involve the federal government or those in charge of enforcing federal government regulations. For example, if they are tasked with informing their citizens about the rise in COVID-19 cases as a means to raise awareness about their new COVID-19 protocols, they will utilize available datasets to create some form of information to present to the public. In order for some sceptical citizens to follow the new rules, they may want to reference sound evidence to back up the new regulations. They may also want to provide the citizens with comprehensive and easy to understand visualizations that are much easier to interpret than the rows and rows of data that are available online. It would also enable the public to discern the truth from all of the varying news about COVID-19.

When searching for COVID-19 data to use in their visualizations that will be available to the general public, they may find many large, general datasets related to the issue, and they may come across the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [7]. Although this dataset is very large, it can also be extremely specific since it is categorized into many folders and subfolders. The data is also updated daily, so it is always reliable and relevant.

Once the data scientists who work for the federal government are able to locate and clean this data, they will be able to produce relevant visualizations and graphics to be available to the public. These visualizations will help people understand the urgency of the pandemic and back up the new regulations that the government set in place. They hope they will be able to convey the data in an objective way that will help the public have a better general understanding of the current situation of the pandemic.

– DATA

We are using the New York Times COVID-19 Case dataset. This dataset includes several links. For our project, we will be using the U.S. State-Level Data CSV file. The data begins January 20th, 2020 [the first reported case of COVID-19 in the United States was on January 21st, 2020], up until February 17th, 2022, and provides the number of cases per day in each of the 50 listed US states. This dataset includes 6 columns, and they are follows: date, state, fips (which is a code that corresponds to each state), number of cases, and total number COVID-19 related deaths. This dataset, of course does cannot show a completely accurate number of COVID-19 cases there were in the United States because often throughout the pandemic, testing has been limited, and not every case is detected in general.

Updates to the dataset: Our data set had to be updated. The following changes were made: The “fips” column was deleted as it was not useable for us and was extra data. A new column was added which included the abbreviation of each state name.

New York Times Covid-19 Data - https://github.com/nytimes/covid-19-data

– TASK ANALYSIS

See Figure 5.1 on the next page(s) for table.

Users of our visualization mentioned the possible effects of COVID on the general population. One of the most important domain tasks mentioned by our end users is the truthful information which is shared to combat the spread of false information, especially with COVID and vaccinations, and the notion of false news. Another critical domain task would be the up-to-date data that our visualization uses. With each day, the data is changing, so having updated information about new cases, deaths, and infections is vital for policymakers to keep their mandates and requirements current. Lastly, a possible domain task that our visualization can serve is reducing the preparation time required for spokespersons. With an easy-to-use visualization, a spokesperson can significantly reduce the time spent researching, understanding, and preparing their address to the public. It allows such a great amount of information to be displayed in a clear, concise way that still displays everything needed.

The primary user of our visualization are people who would need this data to make medical or policy decisions. Since our data is recent and provides a summary of cases and deaths, we want hospitals and doctors to be able to make informed decisions from the summarized data as well.

Another primary user we also believe that our visualizations could be useful for is online news articles, analyses, and people who are simply curious. For the average person, COVID data is lengthy and overwhelming. Summarized into color-coded charts with interactive displays would simplify the data into smaller, comprehendible pieces. Additionally, we believe that understandable visualizations help eliminate confusion surrounding COVID, which we hope will help clear misunderstandings about COVID and clarify misinformation that is a result of communication failures.

Our visualization will be primarily developed for present case. We want to present the data in a way that allows our main domain users, the medical professionals, researchers, and policymakers, to quickly identify areas of interest to them and allow them to reach an informed conclusion. Our main goal is to summarize and

present the updated information, therefore allowing users to digest and then provide the general public with educated knowledge as well. Not only is the visualization for these important figures, it will also influence a much wider range of people through them.

While our focus will be presentation, we also aim to make the visualization as user-friendly as possible, allowing the discover aspect to be strong as well. We hope that users can use our visualization to learn more about COVID and the cases and deaths in our dataset.

Task ID#	*Domain Task*	*Analyze Task*	*Search Task*	*Analytic Task*
01	Researchers/Medical Professionals: I want to be able to see how the cases and the death rates compare to other previous infectious diseases, such as SARS.	Consume → Present	Explore	Compare
02	Policymakers/Researchers: I want to be able to allow users to quickly pinpoint which locations in the USA had high rates of infection.	Consume → Discover	Locate	Compare Identify
03	News Articles, Blogs, Spokespersons: I want a tool to easily display this COVID-19 data and the trends of this data for my viewers	Consume → Present	Explore	Compare
04	Historians/Researchers: I want to be able to see a two-year summary of the progression and spread of COVID-19.	Consume → Enjoy	Explore	Summarize
05	Anyone: I want to be able to look at COVID-19 case trends easily and quickly for selected timeframes	Consume → Discover	Browse	Compare
06	Anyone: I want to be able to see the effects of COVID-19 geographically	Consume → Discover	Locate	Compare

Figure 5.1 (Task Analysis Table)

– Implementation Plan & Preliminary Work

Given the dynamic nature of COVID-19 and its evolution in the last two years, it only makes sense for our final product to have a high degree of interactivity in order to help users visualize the timeline of COVID and its growth. Our main dataset focuses on COVID data within the United States, so we plan for our final product to visualize data on a similar scope. However, given the role that the U.S. plays on the global stage, it is more than likely that any data we visualize regarding any form of industry/business in America will touch on the impact of COVID in foreign countries. While we are focusing on COVID rates per state, we also want to look at vaccination rates per state/county and examine any factors that may have led to these results. We also want to visualize any potential datasets that examine the effects of COVID-19, similar to the aforementioned study in *Related Works* that examined the impact of COVID on living material deliveries.

We hope for our final product to be akin to a dashboard reminiscent of those seen in Tableau. Naturally, as per the rules enforced on this project, we will be using D3, but with our skills in front-end development we believe that a dashboard layout may be possible. We also plan to add interactive elements such as a slider timeline so we can display data on COVID from the last 2 years as well as isolate COVID data by state/county. We expect to clean our data via Python through Jupyter notebook, which will be stored in the form of CSVs. Our generated GitHub page will be structured as an HTML, CSS, and JavaScript file. If necessary, external interaction via Python and Flask will be integrated in order to create a seamless interactive experience for the user.

After careful consideration and deciding between various designs created by our team, we came to a conclusion about the actual components of our visualization dashboard. For our final project, we will need to incorporate various types of visual encodings. We will be sure to include a choropleth map since it is the easiest way to digest information geographically as our data spans multiple states. Our users will be familiar with a map of the United States, which makes it much easier to understand that the data will be divided by state and that each state itself represents a mark that more information will be relayed on. We will also make use of a scrollbar for the user to focus on certain time frames, helping the data to stay consolidated and easy to comprehend. This is important to understand trends in the data, especially since the scrollbar would change in real time with the rest of the visualization. We will include a stacked bar graph as well as a line graph that delegates the trends occurring in the time frame specified. The bar graph would show the rates of positive cases in the certain region chosen by the user.

Our interactions and interactive components will include brushing across multiple states. This will then be linked from our map of the United States and the charts and graphs across our

larger coordinated view visualization tool, enabling the user to clearly see which parts of the visualization are working cohesively. We will be using filtering when the user uses the scrollbar since the scrollbar will represent a certain time frame. When the user hovers over any component, we will use tooltips to display in depth details and information to the user. We will also make use of highlighting to emphasize certain components. For example, the specific state that a user moves their mouse on will be highlighted to aid in their visualizing and comprehension of the data.

The required parts of our project include the heatmap of Covid-19 cases/deaths, the tooltip that displays the number of cases/deaths and the date for the state it is hovering, and a line chart on the side that displays the difference in cases. We also need a scrollable bar so that the user can see the data over time. We would like to make a more intricate side chart, similarly to Sketch 6, where one can see the trends of the data and the change in cases/deaths.

Our project will mostly be done using regular D3 with JavaScript, but we will be using the Plotly D3 guide to create our heatmap U.S. map (https://plotly.com/javascript/choropleth-maps/).

– Visualization Design

In our final visualization tool, we decided to implement a map that shows the states and their locations. When users hover over a state on the map, our visualization shows a concise summary of the particular state's statistics. When users click on a state, rather than just hovering, we want the view to display the state's detailed cases and deaths.

The initial view that the users see is the map. The US map is meant to show which states have higher rates of COVID cases and deaths.

To create this map, we originally wanted to use the Plotly library and their map methods. We felt that using Plotly, rather than the other libraries we were taught in lecture was easier, since we did not go into depth about maps in other libraries. However, after multiple attempts at implementing the map, we are currently reconsidering our design choice. We want to continue to pursue the visualization that we had originally discussed, but we are flexible to adapting our goals. At this time, we are focused on other aspects of our visualization, having determined that the map is going to be difficult and time-consuming. Putting the map off also entails that the most important user interaction, hovering and clicking states, is currently in the works.

Some ways of conveying information that our visualization implements are through colour and area. Each state is a mark, an area that is coloured according to the data. Through using hues, we can show users information about that state; if the colour is a lighter hue, the state has less COVID cases/deaths and if the colour is a darker shade, the state has a higher COVID cases/deaths. Using colour allows users to quickly understand the cases and deaths nationally and users can quickly see which states may be most interesting.

Our group chose to focus more on the bar chart from the visualization. The bar chart displays detailed description of COVID cases and deaths for each state. Our group intended this chart to be displayed when the user clicks on a state; the bar chart would then show the bar chart with data specific to the selected state. However, since we do not have our map implementation finalized, we are working on just displaying the bar chart. We used D3 as our main visualization library for the bar chart. For now, some elements of the bar chart are hard-coded, but we aim to change that in the upcoming weeks.

Additionally, our group is also working on the line chart. The line chart, much like the bar chart, is also meant to be displayed when the user interacts with the US map. The line chart is also built with D3, since our group agreed that using information taught in lecture was the easiest way. The line chart, much like the bar chart, as some aspects that are hard coded for now.

Our bar chart uses areas (bars) to show the rates of cases and deaths. The heights of the bars allows users to quickly compare the number of cases and deaths for the interested state. Similarly, the line chart uses lines to demonstrate the changes of cases and deaths overtime. The trend of the line allows users to efficiently compare the rate of change from one time period to another.

Our group's main focus this week was to have a basic idea of how we were going to implement the visualization we want. We made sure that our files were in order and our data was uploaded correctly. Our group dedicated a significant amount of time to Plotly before deciding to focus on the other aspects of our planned visualization. Although our group feels that we did not accomplish what we wanted to this week, we still gained experience with Plotly, which should help us when we actually implement the map next week.

Even though our group was unable to implement the map this week, we still made good progress on our bar and line charts. We lost time due to the Plotly situation, but our group still put our best efforts forward into the bar and line charts. Next week, we hope to finalize our visualization and begin incorporating the interactivity.

This section will be updated accordingly when that happens.

REFERENCES

[1] M. Mora-Araus, A. Velastegui-Montoya, Y. Jaramillo-Lindao and H. Apolo, "Mapping the Sound Landscape During Social Isolation Due to COVID-19," *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 8340-8343. doi: 10.1109/IGARSS47720.2021.9554530.

[2] R. Y. Wang, T. Q. Guo, L. G. Li, J. Y. Jiao and L. Y. Wang, "Predictions of COVID-19 Infection Severity Based on Co-associations between the SNPs of Co-morbid Diseases and COVID-19 through Machine Learning of Genetic Data," *2020 IEEE 8th International Conference on Computer Science and Network Technology (ICCSNT)*, 2020, pp. 92-96. doi: 10.1109/ICCSNT50940.2020.9304990.

[3] T. Zhao et al., "Optimizing Living Material Delivery During the COVID-19 Outbreak," in *IEEE Transactions on Intelligent Transportation Systems.* doi: 10.1109/TITS.2021.3061076.

[4] S. Afzal, S. Ghani, H. C. Jenkins-Smith, D. S. Ebert, M. Hadwiger and I. Hoteit, "A Visual Analytics Based Decision Making Environment for COVID-19 Modeling and Visualization," *2020 IEEE Visualization Conference (VIS)*, 2020, pp. 86-90. doi: 10.1109/VIS47514.2020.00024.

[5] Y. Zhang, Y. Sun, L. Padilla, S. Barua, E. Bertini, and A. Parker, "Mapping the Landscape of Covid-19 Crisis Visualizations." *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021. doi:10.1145/3411764.3445381.

[6] "Covid-19 Hospitalizations." Centers for Disease Control and Prevention. Centers for Disease Control and Prevention. Accessed February 24, 2022. https://gis.cdc.gov/grasp/COVIDNet/COVID19_5.html.

[7] CSSEGISandData. "CSSEGISANDDATA/Covid-19: Novel Coronavirus (COVID-19) Cases, Provided by JHU CSSE." GitHub. Accessed February 24, 2022. https://github.com/CSSEGISandData/COVID-19.

[8] CSSEGISandData. "Covid-19/who_covid_19_situation_reports at Master · CSSEGISANDDATA/Covid-19." GitHub. Accessed February 24, 2022. https://github.com/CSSEGISandData/COVID-19/tree/master/who_covid_19_situation_reports.

Appendix A: Group Charter

– Group Purpose

Our group had met each other in our previous courses but had not had the chance to work together yet as a team. We decided to take the opportunity to work together for this class's project and see how our individual skills come together to form a great project. COVID has defined our college experience so far, and now that we're almost halfway complete, it's safe to say we've never known a higher education experience without COVID. As such, it is crucial to look back on what has happened and what will happen in the world, so we know what knowledge we need for the future.

– Group Goals

Each member of our group intends on putting our best effort forward. We would like to create a meaningful project and a visualization that follows all of the design rules of thumb to go along with it, as well as integrate all of the skills we have learned in this class so far. We are all committed to the highest level of performance and are collectively aiming for the highest course grade possibly. At the end of this project, we hope to see significant improvement in our skills with managing large datasets as well as front-end development to visualize such data.

– Group Member Roles/Responsibilities

Our group leader will be Karen, the meeting facilitator will be Katerina, the document coordinator will be Caitlin, and Michelle will be the information manager.

– Ground Rules

Our group will meet either virtually or in-person. We have created multiple channels of communication (text group chat, Teams group chat, Discord group chat, and email). We are all comfortable speaking with each other and are comfortable voicing our concerns with one another. If an issue arises, we will simply talk through it and move forward. We will keep specific deadlines and expect to commit an equal amount of time.

– Potential Barriers and Coping Strategies

As mentioned above, we are comfortable speaking with one another and being honest with both ourselves and each other. If a problem arises with a certain group member, we will bring it up as a group and work through it together. We do not foresee any big problems arising.

– Mid-Project Reflection

All members of our group have been abiding by our agreed-upon guidelines. We continue to feel comfortable with our group roles and dynamic. The only problem we have encountered is time because we are all very busy this semester, so it's often tough to find a time where everyone can work, however, this is something that we have continued to be able to work through.

- Katerina -> Karen does a great job at formatting our document, and always contributes positively to our conversations while we work.
- Karen -> I love that our group caters to what each member does best. We let the artist draw and the coders code. We are flexible with our assignment of tasks and trust each other fully to handle a task. As a group, I believe that this division of labor allows for higher-quality work, especially since tasks are handed to the person who is both best suited and the most interested in a particular task. The quality of work by each group member
- Caitlin -> I love Michelle's attitude towards our work. Though it isn't always easy, Michelle is very positive and has a good sense of humor to cheer us up when we are all feeling down. She helps boost morale a lot since she reminds us that we aren't just working for ourselves; we're working for each other. We know when to focus, but also know when to take a step back and take a break.
- Michelle -> Caitlin is doing a wonderful job at digital drawing since she is a designer. She is really doing her best and we are all grateful that she has these design skills to assist in a lot of the visualizations. As a group, I'm so glad to be able to work together with my partners for this project since we have all had classes with each other before and are finally able to come together as a team and put together our individual skills. We are all also able to connect with each other since we are all on co-op search this semester, and it is easy to work through any problems with this group because we are all comfortable enough to talk things out and make sacrifices if needed.

– Appendix B: Data Exploration

– Data Review

The date column is ordered, ordinal, sequential; The state column is categorical, the levels are the states of the 50 United States of America; The fips column is categorical, the levels are the codes that represent each U.S. State; The cases and deaths columns are ordered, quantitative, and sequential.

Our cases and death counts had a very wide distribution of numbers because COVID cases and death rates varied widely through the 2 years of the pandemic. The death values range from 0 to 85,682, with a mean of 8035.98 deaths, and a median of 2849 deaths. The case values vary from 1 to 8,978,975 cases a day, with a mean of 477943.71 cases, and a median of 170789 cases. The dates range from 01-21-2020 to 03-01-2022.

We did not encounter any data quality issues and had to perform no data cleaning for our dataset. We believe that this is due to the fact that our data set is very simple and has also been used to create a dashboard for a credible news source (The New York Times), which meant that the data had to be credible to begin with.

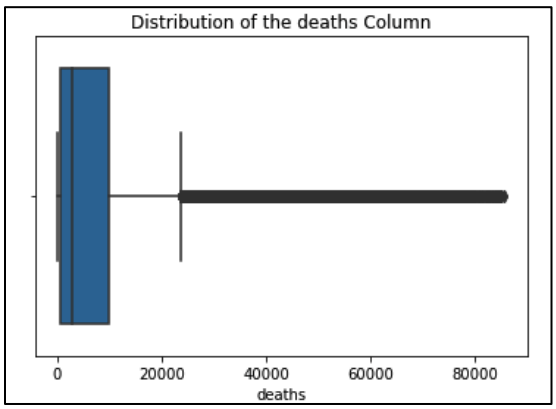


Figure 1

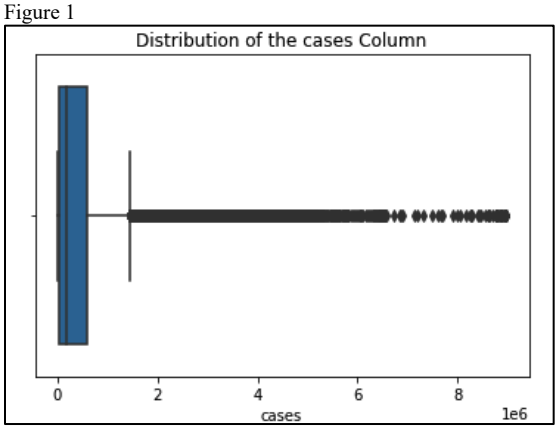


Figure 2

– INSIGHTS

As expected from the COVID dataset, we began to see larger and larger amounts of cases and deaths as time went on. We had visualizations that split the case/death data into both cumulative and monthly numbers. Unsurprisingly, the states with the largest populations – in particular, those with large cities containing clusters of immigrant populations, such as New York, California, Texas, and Florida – had the most cases and thus the most deaths. However, while on paper they may have the most cases and deaths, this does not consider the percentage of the state population that these numbers represented. When we look at the visualizations that break cases and deaths down month by month, the largest spike in cases and deaths was at the start of the pandemic. These numbers finally started to decrease by Fall 2020, and though COVID cases/deaths haven't been constantly decreasing, even the small spikes in numbers have not surpassed the initial wave of cases and data.

One interesting observation we made during our data exploration was the gap between the number of cases and number of deaths computed over a monthly timeline. The largest gap was, naturally, during the first wave of COVID. However, this began to significant shrink during April 2021. This lines up with increased vaccination rates as the COVID vaccine was made available to the general public.

Another interesting observation we made was about the rates of COVID per state. Though the monthly breakdown of cases/deaths does not specify by state, by dragging the slider on the map, we saw an increased rate of cases and deaths in states that eased COVID restrictions earlier on in 2021. Many of these states were in the south-eastern United States, which lines up with the states that had more of a controversy towards wearing masks and vaccinations.

In future observations, we would like to do a monthly breakdown of cases/deaths per state to see if there are any further correlations in cases and rates over time. Additionally, as mentioned before, we could add an additional dataset containing the population of each state as well as the demographics to get a rough breakdown of the different groups who got COVID. This could potentially help us identify any possible connections to external causes of higher COVID rates within certain groups, such as racial discrimination or natural disasters (displacing populations).

It is worth noting that the data for March 2022 is incomplete as data is updated daily through Johns Hopkins University, and this project was written in March 2022.

– SCREENSHOTS

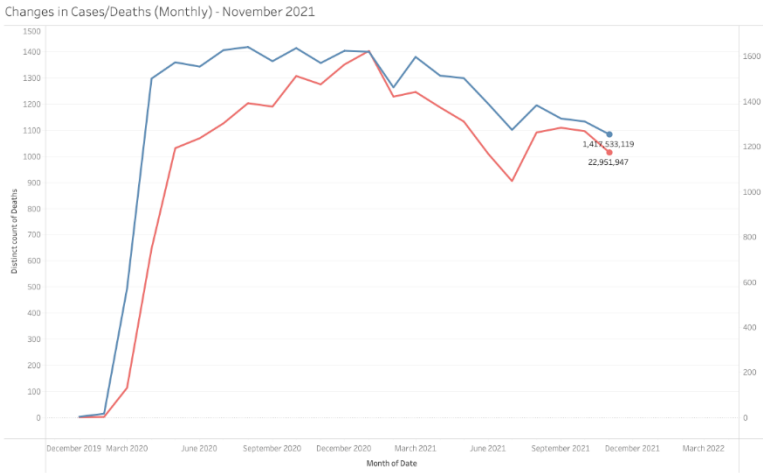


Figure 3

This visualization explores the case and date rates across the recorded months from our data set and is not cumulative. This data is presented as a double line graph where the different colors represent case counts (represented by the blue line) and death counts (represented by the red line). This visualization helps us see that death rates and cases rates have a direct relationship, whereas one increases, the other one also increases and vice-versa. We can also get a clear visualization of the gap between the amount of COVID cases vs the number of deaths, which has shrunken over time now that the vaccine and booster shots have been widely distributed. However, there are still small peaks, which could likely be due to newer waves of COVID variants, or increasing cases because of eased restrictions.

This data was visualized with a slider on Tableau, and this screenshot was from the slider displaying November 2021 data.

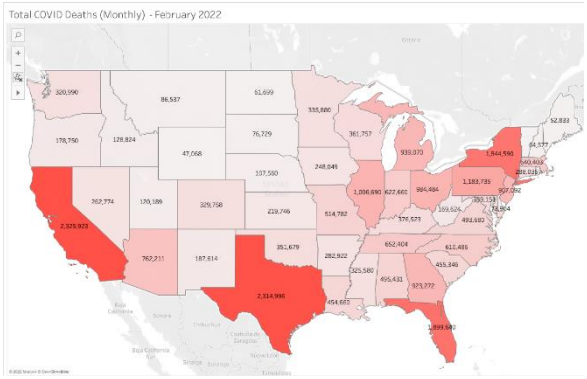


Figure 4

This visualization shows the cumulative amount of COVID deaths by month by state from March 2020 to March 2022. This data is presented as a map, where varying shades of red are being used to indicate the number of deaths from COVID. Lighter shades of red are used for less deaths, while the states with the most deaths are a more vivid, dark shade of red. Since the data is cumulative and not broken down by month, at some point all the states will be dark shades of red. However, since the number of cases (and therefore deaths) is not constantly increasing, it will likely be a long time before this map reaches that point. Additionally, this map does not account for the percentages of people per state that have died from COVID. Thus, since the states with larger populations are darker red, they'll have larger amounts of deaths, meaning many of the rural states will always be light red in comparison.

This data was visualized with a slider on Tableau, and this screenshot was from the slider displaying May 2021 data.

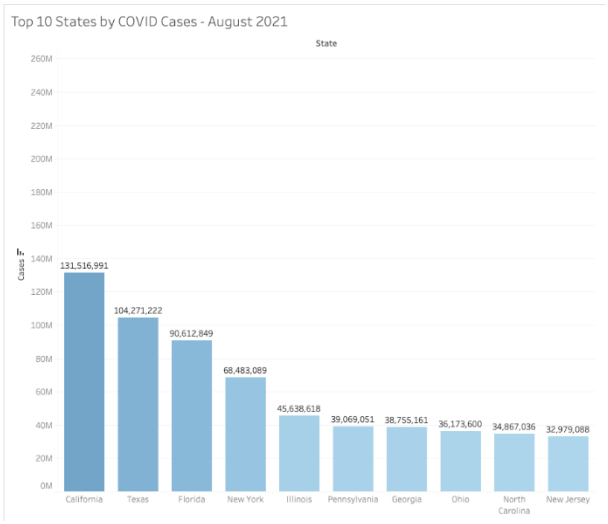


Figure 5

This visualization shows a ranking of the top 10 states with the largest (cumulative) amount of COVID cases, with data spanning from March 2020 to March 2022. This data is presented as a bar graph, where varying shades of blue are being used to indicate the number of COVID cases. Lighter shades of blue are used for less cases, while the state with the most cases is the darkest shade of blue. As we have observed before, the states with the largest populations also have the most cases. Many of these states also have cities that serve as safe havens for immigrants. California has San Francisco and Los Angeles, for example. As a result, these cities tend to be the most densely populated (as evidenced by New York, the state with the highest population density). This means that there could be additional factors, especially racial and socioeconomic ones, that could contribute to increased COVID rates within certain minority groups. It is also worth noting that while these states do have larger amounts of cases, this is in proportion with larger populations. We think that having data displaying cases over the total population would give a better picture of the COVID data, as these are purely just numbers.

This data was visualized with a slider on Tableau, and this screenshot was from the slider displaying August 2021 data.

– **DATA SNIPPET**

date	state	fips	cases	deaths
2020-01-21	Washington	53	1	0
2020-01-22	Washington	53	1	0
2020-01-23	Washington	53	1	0
2020-01-24	Illinois	17	1	0
2020-01-24	Washington	53	1	0

APPENDIX C: INTERVIEW

– **END USER PERSONAS**

Nora Green is the current White House Press Secretary. Her job is to collect information about important current events and speak about them to reporters and viewers in America. It is crucial for her to have up-to-date and accurate information about current events and pressing matters such as the COVID-19 pandemic. Nora has used visualizations before, but they have not been super scientific, as she is a communications professional and not a scientist.

Nora is looking for a way to quickly look at the COVID-19 cases of the day, week, or month and see how it is trending without having to manually create excel graphs every day. She also needs to know if any states are COVID hotspots.

Moreover, Nora is looking for a tool she can present when she is holding press conferences and COVID-19 related questions arise. Because of this, the visualization tool she is requesting needs to be extremely easy to use and comprehensible for people of different backgrounds and abilities.

Nora believes that this visualization tool would reduce her pre-conference prep significantly and would improve the quality of her conferences.

Our second end user is Doctor Anthony Fauci who is an immunologist and has served as the Director of the National Institute of Allergy and Infectious Diseases since 1984, meaning he has clearly seen a lot of changes in the way data is presented to the public over a wide range of years. He wants a good data visualization that will present facts to the public that are easily understandable and easy to comprehend without additional information.

Since he is in charge of releasing many statements related to COVID regulations and updates, Dr. Fauci places a great emphasis on how this visualization tool will affect its users and how it can be used to explain ideas to many kinds of people in the general public. This includes people that need a visualization to understand the new regulations and the reasoning behind them, and people that need to be informed about COVID in general so they can ground their knowledge in factual evidence rather than other information that may be posted online.

– Interview Script

Some questions that we would like to ask are:

1. What do you need from the visualization?
2. What do you intend to do with the visualization?
3. How do you think the visualization will help you accomplish your tasks?
4. What information do you think is most important for the public to know about COVID-19?

As the interview progresses, we intend to ask more specific questions regarding their responses to the preliminary questions. One example would be (regarding the intended use of the visualization): on a scale of one to ten, how much interactivity would you prefer? We would also like to know information such as how the visualization is being presented. If the visualization is being shown on a large projector, we may want to choose colours that are high contrast, since larger projectors tend to have colour washouts. We would also want to use high contrast to be inclusive to those who have any type of sight disabilities. On the other hand, if the visualization is being printed on paper, some of our features, such as the slider bar for time progression, would not be available. If the visualization is presented on paper, we would want all of the information to be successfully translated and while this may result in more paper being used, it will still convey all of the information necessary.

– Interview Notes

1. What do you need from the visualization?
    a. The visualization should be simple, informative, updated, and accurate.
2. What do you intend to do with the visualization?
    a. Simplify my preparation before I meet with the public, so it is easily understood
    b. Analyse the current data, plan for next steps since the data continues to update and grow as time goes on
3. How do you think the visualization will help you accomplish your tasks?
    a. Makes analysis more efficient and accurate.
    b. Visualizations are good summaries for presentations
    c. With a time-lapse, could help show the cause and effects of mandates, such as the mask mandate in various areas and the vaccine requirements that different regions and states of the US implemented.
4. What information do you think is most important for the public to know about COVID-19? (This question is intended to address the use of our visualities by the general public. Since our visualization will likely be made publicly accessible to all, we would like to know, in addition to helping the targeted end users, how to help the public better understand the spread and death caused by COVID-19.)
    a. The death tolls. There are people out there who believe that COVID is a hoax. They believe that the vaccine is fake or designed for the government to track its citizens. I hope that this visualization helps citizens realize that COVID is very real, and there are very real consequences. I also hope that it can help increase the amount of factual, sound knowledge about the illness.
    b. The spread. Despite recent news, COVID is still a very real issue. I believe that we can expect COVID cases to rise in the currently war-torn areas, due to the lack of masks, close physical contact with others, lack of sanitation, and other preventive measures.

– Interview Results

Q: Given the visualization, what are some things you need from the visualization?
A: Something that I can see briefly that will give me the information I need before speaking to the reporters. Given that I have limited time to prepare for each public speaking session—and that I typically have to prepare multiple topics and ready answers to potential questions that the reporters might ask, I must be extremely efficient. This visualization, I hope, will make my preparations much faster.
--
I need the visualization to be concise, accurate, and informative. Given that I will be making recommendations based on the information presented, it should be as accurate as possible and as up to date as possible. Given that there are actual lives on the line, I need the visualization to show me what I need to know as accurately as possible.

Q: What do you intend to do with the visualization?
A: I intend to use it as a source of information. A summary of the recent developments. I may even have to present the visualization on a posterboard of sorts to show the reporters to emphasize my points. Since COVID is almost always a guaranteed question that reports will ask these days, I intend to use the visualization to provide me a quick, memorable summary regarding the current state of COVID. I may even have to consult other policymakers with the visualization and ask regarding the response and other efforts being made to combat COVID and prevent the spread.
--
Draw conclusions. I intend to use the visualization to make assumptions about the current state of COVID in the US. I need a quick, accurate summary of the current state so that I may plan for the future. With the time-lapse that the visualization provides, I can quickly determine which mandates, such as the mask mandates, quarantine mandates, and vaccine mandates, worked, and which mandates did not. I hope to discover trends that were, perhaps, overlooked by the general public and other policymakers to improve future ones.

Q: How do you think the visualization will help you accomplish your tasks?
A: Summary. Summarizing the information to make it more comprehendible and easier to read. Saving time by compiling the data into easily understandable images reduces the stack of paperwork I must parse through. This saves me time and I can dedicate more time to other things that need to be done and covered. The visualization would also allow me to remember details better and with more clarity, so it would not only save me time from parsing through paperwork, but it would also save me time in the memorization department. These two factors combined save me a significant amount of time.

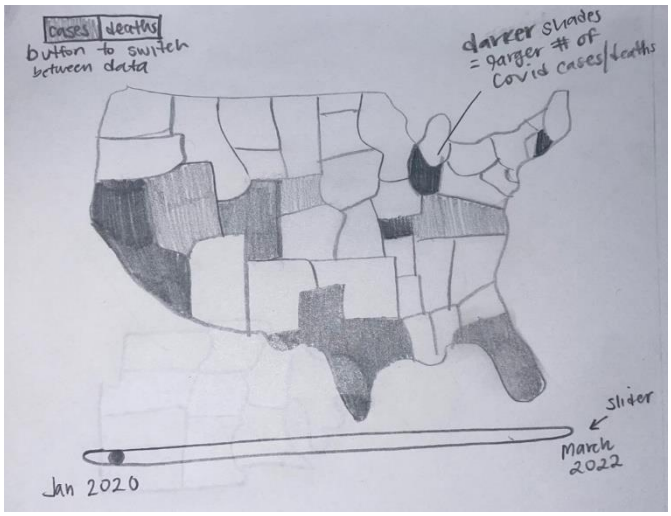
--

The time-lapse that the visualization shows are very important to my planning. Since the enforcement of mask mandates, we can see that there are significant dips in the total cases and the total deaths. We can also see that since the vaccine rollout efforts began, the curves have also begun to trend downwards. This is a result of the efforts to limit the spread and susceptibility of citizens to the disease. With the time trend, I will be able to say with conviction that certain policies and mandates have effects on the cases and deaths that the disease is causing.

Q: What information do you think is most important for the public to know about COVID-19?
A: I think that the most important thing for the public to remember about COVID-19 is that it is not yet over. There are still recommendations about vaccines and masks that some citizens are ignoring or choosing not to follow. Thus, I would like to remind citizens that while COVID may appear to be phasing out, we cannot let our guard down and simply return to what life was before. There are significant changes, not just in our healthcare, that have been impacted. A significant number of jobs are still being added back into the marketplace; students are re-adapting to being in in-person classes, and workplaces have begun to return to the office, obviously with remote work still in mind. I believe that COVID will have, and already has, a lasting impact on society and the way we view things such as healthcare. So, I believe the most important thing for citizens to remember is that it is not over yet, and although life may never return to what it once was, I hope that we can use this experience to prevent the next one.
--
I believe the most important thing for the public to know is the death toll. Some may believe that COVID is a hoax or a joke, but the death toll does not lie. There are some who refuse to wear masks or refuse to receive the free vaccine, simply because they do not trust the government and believe that the government is tracking them with so-called “chips”. The reality is that COVID is very real and has impacted the lives of millions of people in the United States. Citizens need to remember that their actions and choices have consequences for themselves and others around them. Too often, we hear numbers being thrown around on the news, with no context to truly how large the number is. I want these visualizations to put into context how many people have suffered and died from this very real disease.
Also, with the recent developments of the war in Ukraine, I can say that we expect COVID cases to rise, due to the lack of masks, sanitation, and the destruction of infrastructure. We hope that citizens, especially in the US, do not take what we have as granted.

– APPENDIX D: DESIGN SKETCHES

– SKETCH 1 - KATERINA



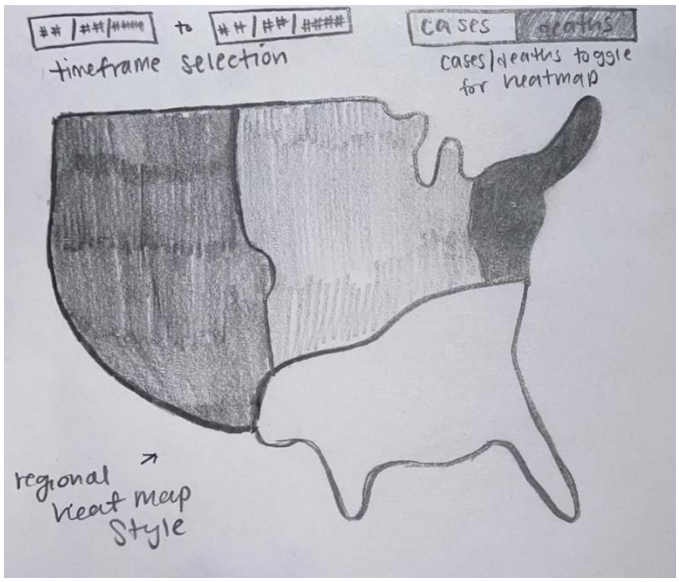
This sketch addresses tasks 2, 3, and 4. This visualization allows users to see the geographic Covid hotspots over time. Hue is used to represent the number of cases or deaths. Specific statistics show up when each state is hovered.

There are buttons to choose between seeing the data for cases or deaths, and there is also a scroll bar on the bottom where data can be seen over a different amount of time. Each state “shape” represents each U.S. state (Hawaii and Alaska not shown).

– SKETCH 2 - KATERINA

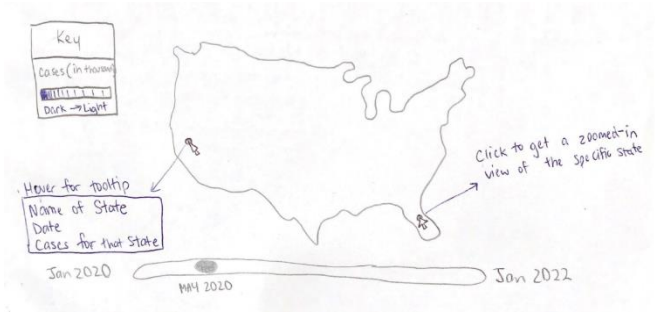
This sketch addresses tasks 6 and 3 and displays covid trends in each U.S. region or state. Each state shape is selectable, and the user can choose to see the trends for either the entire region or up to 3 states at a time. The line graphs on the side use lines differentiated by colors that represents cases vs. deaths. These lines display the number of cases/deaths over the selected time frame. This visualization has 4 sections for the 4 U.S. regions (drawing excludes Northeast + the South for brevity). There are also small side bar charts where the area of the bar represents the total number of cases and deaths for the selected timeframe.

– **SKETCH 3 - KATERINA**



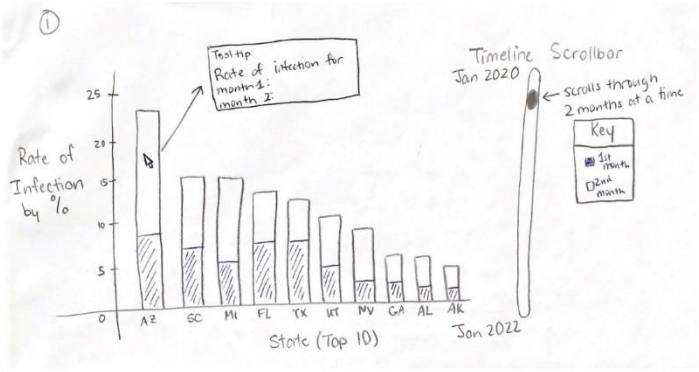
This visualization addresses tasks 3 and 5. This visualization is a U.S. map split by the 4 regions, where each area has a saturation that represents the number of cases or deaths based on what the user chooses, and the time frame they selected. Each region is also a different color (where again, the darker the color, the higher the number of cases), and each color represents the different region (for example, the west is represented by green, the Northeast by red, Midwest by blue, and South by yellow). On the side, there is a chart that indicates whether each region's cases/death rates are trending upwards or downwards.

– **SKETCH 4 – MICHELLE – FAVORITE**



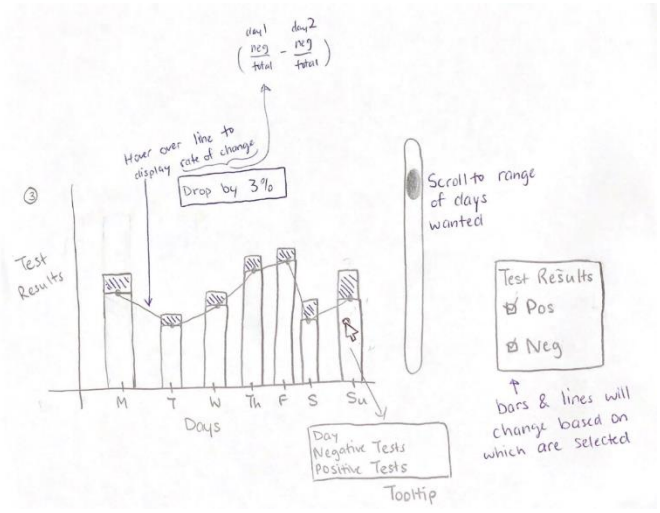
This visualization addresses the domain task for historians and researchers. This would allow them to quickly pinpoint which locations in the USA had high rates of infection and have additional information by hovering over the specific state through a tooltip. The visualization mainly uses color as its channel to convey information. This is shown through the key with the bar showing that the darker colors will represent more positive test cases and lighter colors indicate less positive test cases. It also uses shapes as a channel to differentiate the states from each other. Ideally, this map would have outlines for the states which would use the shape channel (Since it was a sketch, the outlines are only part of the concept of this drawing) as well as specific numerical values in its tooltip and key. The states themselves would be the marks for this visualization and they can be considered areas because they are a two-dimensional mark. This visualization utilizes a scroll bar to represent the month and year that the user wants to specify. Once this scrollbar changes, the map should change in real time. It will also include a tooltip depending on which state the user hovers their mouse over. To get even more information, the user would be able to click on a specific state and see a zoomed-in version of that state.

– **SKETCH 5 - MICHELLE**



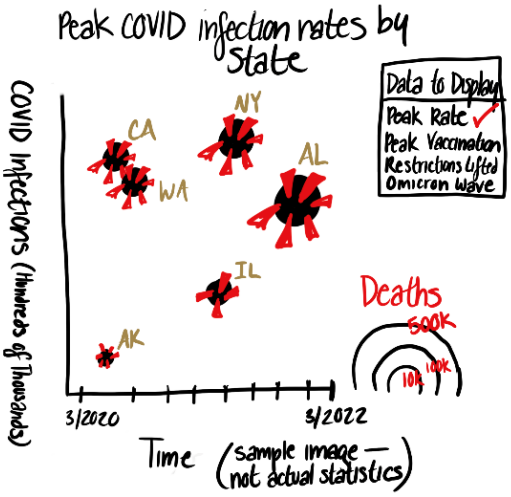
This visualization addresses the domain task for policymakers and researchers. It allows them to see a two-year summary of the progression and spread of COVID-19 by scrolling through, two months at a time. Since they would already be familiar with the topic based on their field of study and their prior knowledge, I decided to go for a visualization that would allow them to extract more information in one single view. This visualization simplifies the data down into rate of infection for the top ten states, which would allow people like policymakers to quickly find the top areas of interest and address the public's concerns in that specific area. I wanted to include both months in a stacked bar chart because it would allow the users to understand a trend in the data without having to explicitly communicate it. The marks would be the areas that each bar individually fills up, and the more area the bar covers, the higher value it represents. The different colors also help to differentiate between the 1st month results and the 2nd month results. This is detailed in the key to the right of the visualization. Overall, this is easy to digest and only contains the most important information for its users to be able to relay the information to others later on.

– **SKETCH 6 – MICHELLE – FAVORITE**



This visualization addresses the domain task for the general public. This would allow them to be able to look at COVID-19 case trends easily and quickly for selected timeframes due to the scrollbar, and it would be easily digestible for someone who is not as comfortable with medical terminology. This visualization combines a stacked bar chart with a line graph of the averages. Ideally it would include exact numbers along its y-axis to represent the actual number of results. The scrollbar would also contain exact dates for the user to explore. The box on the right labelled "Test Results" would enable the user to see different results at once. For example, if they only chose to see "Pos," the graph would change to just have plain blank bars representing the positive tests. Otherwise, if they chose to only see "Neg," the bars would only display the negative test results. Clicking on both options at once allow the user to see both the stacked bars and the line graph which represents the trend in the data. Users would be able to use this trend easily thanks to this visualization since its numerical value would show up as a tooltip when they hover their mouse over the line. Another tooltip I included would pop-up once a user hovers over the bars. It would simply include the numerical data of the bar they are viewing. It would have the date, the number of negative tests, and number of positive tests. This visualization uses color and area to represent different things and show contrast. The positive and negative tests are different colors while the size of the bars themselves represent their values. Marks such as the dots and lines will also be extremely helpful for the user to find where they can find detailed information.

– **SKETCH 7 - CAITLIN**

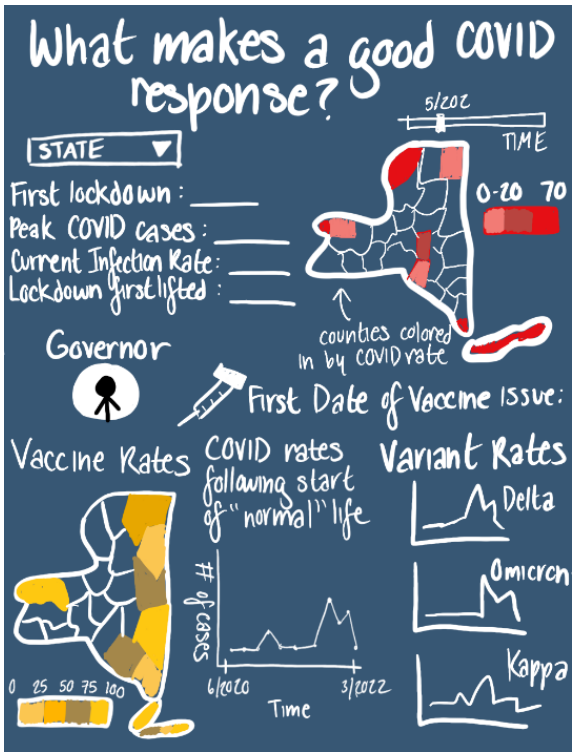


This sketch, drawn by Caitlin, provides a unique take on a bubble graph. Rather than using bubbles, it uses mini images of COVID-19 as bubbles (marks), with the size (channel) indicating the amount of deaths from COVID-19 for a particular state. While the size indicates deaths, these mini viruses would be plotted along a timeline, with the x-axis representing time and the y-axis representing the number of infections in thousands. This visualization would be useful for plotting the relationship between multiple factors during the pandemic, and in the case of COVID-19, plotting the time, infections, and death rates would be helpful to get an overview of the virus fatalities overall. While the purpose of the graph still serves as a bubble graph, I chose to visually design in this way to provide a fresh take to viewers and personalize the graph to the subject matter.

Additionally, since this graph utilizes multiple factors, we can take advantage of that and display other datasets to help users get a better idea of how COVID has impacted society. Other datasets to display besides infection rates are vaccination rates, COVID infections vs. the day restrictions lifted in states, and the infection rates of the Omicron wave.

This visualization would be useful for general users who want to see COVID-19 case trends, but media personnel can use this visualization to easily interpolate and present data about the virus. Since the graph plots 3 factors, users will be able to see a lot of data at first glance, especially in regards to breaking down COVID data per state. With multiple views, users can also compare overall trends in COVID infections and compare various factors that have impacted the data being shown.

– **SKETCH 8 - CAITLIN**



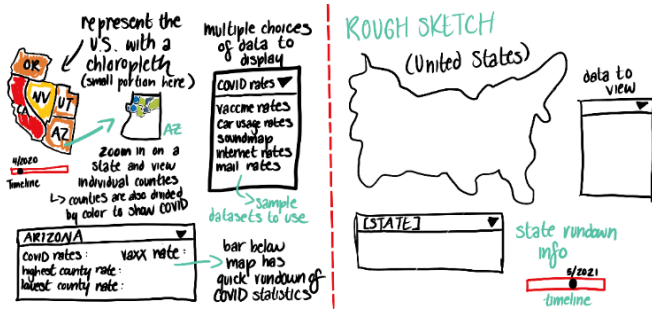
This sketch, drawn by Caitlin, aims to provide a comprehensive dashboard of COVID responses across each state. There is a dropdown menu for users to select a state, and then this visualization displays summary statistics of the state, including information such as the date of first lockdown, the current governor, and the current infection rate. We also see chloropleth graphs on vaccination rates, and line graphs on COVID variant infection rates as well as looking at infection rates once life for Americans went back to "normal."

One channel we would use in this sketch would be color. For the two maps of a given state displaying vaccination and effection rates, our lighter colors would indicate lower infection/vaccination rates, and darker colors would indicate higher infection/vaccination rates. This helps visually distinguish counties within states for users and provides a

geographical element that can easily see the overall cases of COVID-19 across different parts of a site. Additionally, the marks being lines for the graphs in the bottom half of the visualization help the user keep track of the infection rates over time.

In compiling information together about infections and vaccinations, I wanted to create a one-stop visualization for the most vital part of COVID data. With fast facts about a state, users looking for information about COVID have a reference that easily provides them with numbers and facts surrounding how a state has been dealing with the pandemic. Users impacted by this would be policymakers, researches, media personnel, and the general public.

– **SKETCH 9 – CAITLIN – FAVORITE**

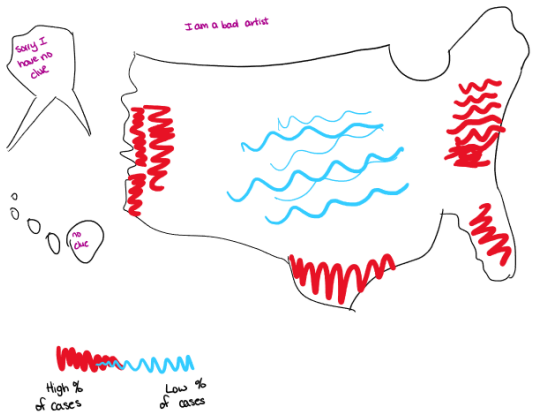


This sketch by Caitlin utilizes the U.S. map to create a chloropleth capable of displaying multiple datasets relating to COVID that are broken down by American state. Users would have a map of the United States mainland as well as Alaska and Hawaii, colored accordingly based on the data that the user is currently viewing. This data would be information on rates of infection, vaccination, sound maps, internet usage and mail. I chose this because I wanted to provide users with a better picture of how lifestyles have been impacted by COVID. Regardless of which dataset the user is looking at, the colors representing the rates will still be the same. Darker colors denote higher rates of infection/usage/vaccinations, while lighter colors in turn represent lower rates. In addition, zooming in on a state would break down the data further by county. This coloring would follow the same rules as the ones that states would follow – darker colors for higher rates, and lighter ones for lower rates.

In addition, selecting a state would pull up a small sidebar containing information about that state. This would be summary statistics of that state's fight against COVID-19, including total infection, vaccination, and death rates. While these summary statistics will not change between datasets, we also have a slider at the bottom of the graph to visualize data across time for the pandemic. Users are able to view data from the start of the pandemic to as recently as 2 weeks ago.

This visualization relies on color as the main channel. In providing a visual aid to users, users are able to distinguish the different amounts of infection/vaccine/death rates, which in turn help paint a better picture of how a state has been fighting COVID-19. Users looking for trends in data, especially geographically would benefit a lot from this graph. From the chloropleth, we hope to provide users with the numbers of COVID-19. However, the summary statistics at the bottom of the visualization also serve to give a more well-rounded picture of the state. With so many different datasets, I thought this visualization would be useful because it gives us the opportunity to visualize data that is more geographic than numerical. We are able to give locations to the data on a map rather than staring at values in a table.

– **SKETCH 10 - KAREN**

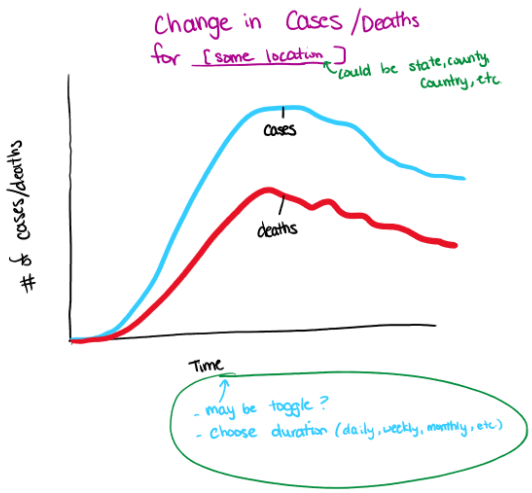


This sketch is drawn by Karen. She is not an artist.

The map of the US is meant to display the entire nation and the rates of cases and deaths. Although the image only displays cases, we were hoping to be able to toggle between viewing cases and deaths. This heatmap allows users to quickly identify which areas of the nation had the most cases and deaths, along with a map that contains the locations of each state for further analysis.

One channel we intend to use for this visualization is color saturation. We want a deeper, more saturated color, such as red, to represent a higher number of cases/deaths. On the other hand, we would represent a lower number of cases/deaths with a less saturated color. We want our map to be a heat map, where areas of high infection are shown in an eye-catching color. We want to user to be able to toggle between viewing cases and deaths, but we are still deciding if we want users to be able to view them at the same time, as a sort of combined-layering technique.

– **SKETCH 11 – KAREN**

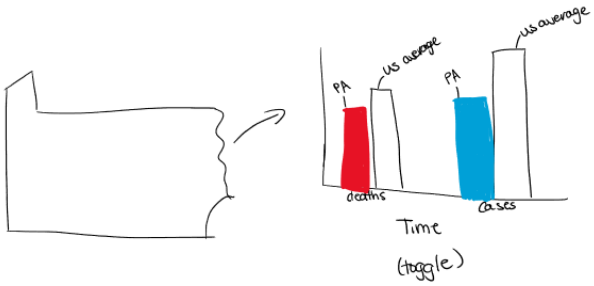


This line graph shows the comparison of the total daily changes for a particular state. We want to be able to compare the cases and deaths, along with user interaction for the time. We want the user to choose how they wish to view the time, whether

it be days, weeks, or months. We want the time to be adjustable, such as having a slider bar to show the progression of the cases over time.

One mark the visualization implements is the line. The line shows the previous progression of the cases/deaths up to and including the current time the user has selected. One channel the visualization implements is colour. We want to use colour to distinguish between the cases trendline and the deaths trendline. This way, it would be must easier for a user to see the correlation between the two and identify any potential relationships. Additionally, we want the colour difference to allow users for quick and easy access to the information they desire.

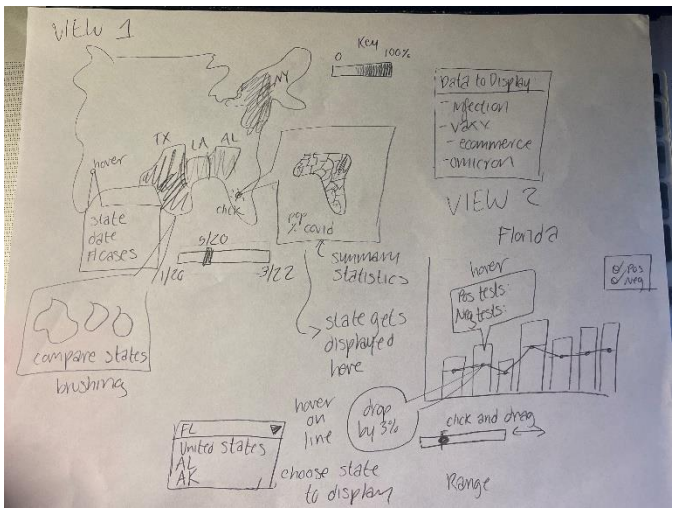
– Sketch 12 - Karen



For this last sketch, we wanted to show what it would look like if a particular state was selected. In this sketch, we chose Pennsylvania. Here, we have a bar chart showing the comparison of the selected state's cases compared to the national average. We thought this comparison would be a good visualization to include, since it shows users how individual states are faring compared to the rest of the nation. However, one thought we had regarding this visualization is the scale. An example would be if a state with a higher population would have more cases, even if the percentage of the population infected was comparatively lower. Thus, my group brainstormed various ways to combat this disparity: we could use the percentage of a state's population compared to the national percentage of the national population, we could use the percent change from one day to the next and compare that to the national change, or we could use some other method of making the statistics more comparable, especially for states with higher populations.

One mark that the visualization uses is areas. Each bar is an area, which represents the selected state or the national average. The bars' heights allow us to compare how high or low or similar state's infection and death rates are to the national average. One channel we are using us the colour. We want to use a brighter, more eye-catching hue for the state's bars and a duller, but of the same colour, bar for the national average. The difference in hues allows users to more clearly identify the state's infection and death rates compared to the national average.

– Final Sketch



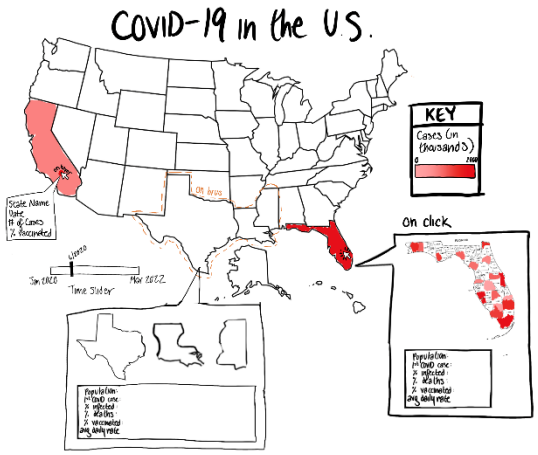
For our final draft sketch, we want to combine Sketches 4, 6, and 9 from the previous section. There are three main goals with our final draft sketch: display a map of the US so that users can identify states and their locations, allow users to select a state or region by clicking and able to view details about the state's cases and deaths, and we want to implement a slider to represent time so that users can see the changes in cases and deaths over time.

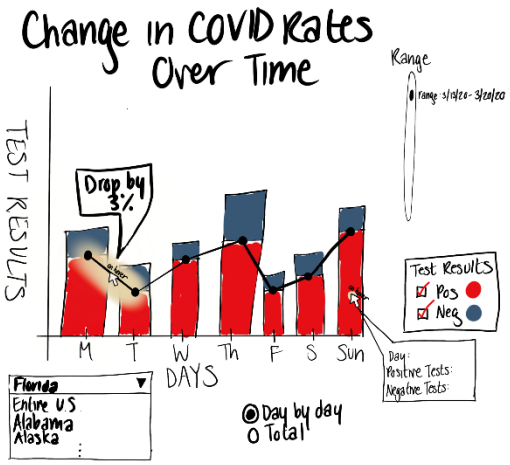
Using these three goals, we narrowed down our options to three sketches. Sketch 4 contains a national map, a time-toggle ability, and state-by-state COVID summary. This sketch caters to most of our goals and would need very little revision to implement.

Another sketch we selected was Sketch 6. We liked that the graph was multi-faceted. The graph consolidated a lot of the information and is still simple enough to understand, especially for our 'general population' domain user. We wanted to include something with more statistics, for those that are especially interested in one particular state.

Lastly, we selected Sketch 9 because of its information. Sketch 9 had a regional approach to the issue, which some users might find useful. We hope to add the regional divisions onto our final sketch, so Sketch 9, along with its multitude of information, was a good choice.

– Appendix E: Digital Sketch





In our final digital sketch, we are keeping the three goals in mind.

- Display a map of the US so that users can identify states and their locations
- Allow users to select a state or region by clicking and enable users to view details about the state's cases and deaths
- Implement a slider to represent time so that users can see the changes in cases and deaths over time.

The first goal of our visualization is to allow users to, at a quick glance, see hotspots of COVID cases and deaths. A map with a colour gradient would not only be the most effective, but incorporating a map also allows users to quickly identify states of interest. The national map is designed to give users a quick summary of COVID, which some end users are looking for. The concise and colour-coding is efficient at highlighting hotspots that may be of particular interest to the user. The map also allows other end users who are curious to explore the map, instead of directing them immediately to numbers and statistics. We also want to implement information text boxes, much like Sketch 9, with information about a particular region or state. This quick summary, along with the summary statistics, allows users to quickly gain an understanding about the current COVID situation of that state. Additionally, the map serves two main functions: allow users to see broader areas of the nation and to serve as a medium between user and the visualization.

The second goal of our visualization is connected to the first. We want to implement brushing and linking so that the user can select a state on the map from the first goal that they wish to view details beyond the summary and have the visualization display details about one state. We are considering allowing the user to select multiple states to compare, but we are not sure if that is something that is feasible with our current skillset and time limitation.

The final goal of our visualization is for users to see the COVID cases and deaths over time. We want to implement a time-toggle ability, so that users may observe the timeframe that they are interested in. We also want the users to be able to select dates to view. This time-lapse will allow users of all kinds to quickly learn about the cases and deaths over time, the effects of certain policies implemented during the timeframe, such as mask mandates or vaccine mandates, as well as observing the spread of COVID.

– Appendix F: Usability Testing

– Preparation

Our COVID-19 dashboard visualization tool allows users to compare COVID-19 cases and deaths across the United States. Users have the ability to switch between seeing COVID-19 cases and COVID-19 related deaths from the entire period of recorded COVID-19 data in the U.S. (starting January 2020) depending on which view they prefer. We can ensure that this data is accurate since it is from the Johns Hopkins Coronavirus Resource Center, and it is updated daily with current information and statistics. Our users are also able to see which U.S. states have the largest number of cases and deaths on a U.S. map (the darker the color of the state, the larger number of cases/deaths) through our choropleth visualization. Our use of contrast and color will allow the user to easily deduce information in a very efficient and effortless way. Despite the complexity of information, we will be able to convey the trends and compare the data from different states easily through our visualization. The tool's intended use is to allow users to quickly look at the data provided and be able to observe trends about the number of COVID cases and deaths, along with comparing between states.

- Task 1: Identify the state(s) with largest number of COVID-19 cases/or deaths.
    - Our visualization will make it very easy for the users to identify the state or states containing the largest number of COVID-19 cases and deaths since we will use channels such as color and area to define these amounts and compare the amounts versus other states. We will be testing this by making sure the users understand what the different shades/colors within our visualization mean and that they are able to find the highest trends which will be made obvious in our visualization. In addition to this, we will make sure our visualization is intuitive so that users do not have a difficult time using it. We hope that our visualization will provide the information in a simple and digestible way for the user, and we will be able to determine this through our usability testing. We also want to make sure that our dashboard is accessible for everyone, and we want to take into account universal design, so we will be making use of contrasting shades for this visualization.

- Task 2: See COVID-19 cases/deaths trends for a week's time frame.
    - We can fulfil this task by allowing our users to choose certain time frames to view based on which time frames they select. We will implement this by using the brushing skills we have acquired from past assignments. Instead of displaying too large of a time period at once, we will simplify the view for our users so that they will be able to specify exact weeks they want to see, and the data will be easily digestible this way. This task will be tested through the user's interaction with the visualization and their brushing using their mouse. We will test if the user is able to successfully scroll over a specific week's time frame and the visualization is able to specify the data in those specific points. Specific data will be relayed to the user through mouse hovering and the use of tooltips, which the user should easily access and understand.
- Task 3: Find exact statistics about the COVID-19 cases and deaths on a specific date.
    - We will be able to complete this task using tooltips and mouse hovering functions. Since it would be too chaotic to include exact data and statistics for each data point statically on the visualization itself, we will instead use the hover functions in d3.js to convey more specific information to our users. Not only does this make the visualization cleaner and more aesthetically pleasing, but we will be able to include more detailed information since the tooltip does not take up space. This can be tested by seeing if the users are able to understand the tooltip intuitively and without extra information. The tooltips should show up automatically when the users hover over the data points so that they do not have confusion when trying to comprehend the data. Hopefully, our visualizations are intuitive and helpful to our users.

– **RESULTS**

One major issue we ran into was that our map of the US is not completely implemented, so we were having trouble asking our user to test it. Otherwise, our other issues were minor. Specifically, our tooltip was difficult to find at first since we still need to improve its location relative to the user's mouse. Currently, the tooltip hovers above the user's mouse but it is hovering too far above it instead of nearby. Although the tooltip exists, we still need to work on its location. Overall, thanks to the labels on the axes of our graph, the participant in our usability testing was able to successfully and clearly understand the function of each mark and channel in our visualizations. They appreciated the clarity of our visualization and did not claim to dislike anything about our visualization tool.

Task 1: Identify the state(s) with largest number of COVID-19 cases/or deaths.

Our main issue, which is an aspect of our dashboard that we are currently working on, was using our data function with our map of the United States for the user to complete this task. Since it is still in progress, the user was not able to use the map for its function, but we will make sure to ask a user to perform usability testing as soon as this part of our project is completed and ready to be used. Therefore, our design definitely requires some changes before it is totally complete. We would potentially look for metrics such as the name of states and the specific data from those states through the use of tooltips.

Task 2: See COVID-19 cases/deaths trends for a week's time frame.

While we are still working on the brushing part of our visualization, we intend to allow the user to choose a specific date in which they will be shown the visualizations in a single week's time frame. This, along with the tooltip when the user hover's their mouse over the data points, should allow the user to successfully complete the task we are aiming for. Specific metrics include the exact number of cases and deaths, and the trend in this data will be shown through the steepness of the line in the line graph. Since the visualization will span a week's time, the trend will be fairly obvious for the week's time frame that this task entails. Our test results do indicate that a change to our design is necessary since we will have to complete the part of our visualization that allows the user to specify which exact week they would like displayed.

Task 3: Find exact statistics about the COVID-19 cases and deaths on a specific date.

Our user pointed out that having the exact date on the x-axis of our visualization gave them a lot of clarity on the data and we believe this to be a positive aspect of our tool. We made sure to make our visualizations simple and easy to comprehend so that the user would not need additional information to understand it. Specific outcomes and metrics we expected were for the user to identify these dates and make use of the tooltip that appears when they hover over the bar graph. This tooltip would allow the user to identify total number of COVID-19 cases and total number of COVID-19 deaths on the given day. Something we still need to work on is the total number of COVID-19 deaths within our tooltip. We are still working on fixing this tooltip and making it accurate for the user.

We will be modifying our visualization dashboard after receiving this feedback from our usability testing and other general improvements we want to make in the future. Specifically, the first change we will implement is hopefully getting our map to show accurate readings of the COVID-19 data for the user's specified time frame. This was our first task that we wanted to complete and it will give a lot more insight to our users since it is a huge visual aid. While this next improvement does not necessarily affect functionality of our dashboard, we will also look to improve the tooltip so that it hovers closer to where the user's mouse is on the visualization. This should help with the user's comprehension of the data and trends displayed. Overall, we will aim to have the best functioning visualization dashboard that continues to provide clear and easily digestible information to our users through the use of labels, tooltips, and d3 functions.