

Joshua Tlatelpa-Agustin - u1399218

u1399218@utah.edu

Kian Akhiary - u1201865

u1201865@umail.utah.edu

Visualizing Covid-19

Repo:

<https://github.com/dataviscourse2023/final-project-sigmaseniors.git>

Background and Motivation:

Our motivation for choosing a visualization project on COVID-19 is rooted in our concern about COVID-19 and how it is still present in the world and people are still being affected every day as a result of the virus. Everyone has been affected by COVID-19 in some way, whether it is a result of getting the virus, or how for months everyone's daily lives were changed when the pandemic began because of remote work, online learning, etc. The pandemic highlighted the importance of data and information in our lives. We believe data visualization has the power to transform complex data into something that is easily readable which can save lives by informing critical decisions.

Our primary motivation is to help contribute to the public's understanding of the pandemic through clear, accessible, and easy-to-understand visualizations. We aim to help individuals make informed decisions based on our visualizations. Visualizations have the potential to shorten the gap between raw data and meaningful insights.

Project Objectives:

Our objective is to identify COVID-19 trends, mainly health status differences between regions across America, and to use data visualization to increase public awareness of the current levels of COVID-19. Regarding what we want to learn and accomplish, we would like to

provide the public with a clear and data-driven understanding of COVID-19 to support informed decision-making. Variables that we would like to visualize include race, ethnicity, gender, and pre-existing conditions and how they have impacted the health status of individuals across the US. We would also like to use our visualization to identify COVID hotspots.

Data:

All of our data is from public datasets made available by the CDC. The Covid hesitancy dataset was last updated in 2021. The Case Surveillance data is an ongoing dataset that gets updated monthly.

COVID-19 Case Surveillance Public Use Data with Geography (2023)

<https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data-with-Ge/n8mc-b4w4>

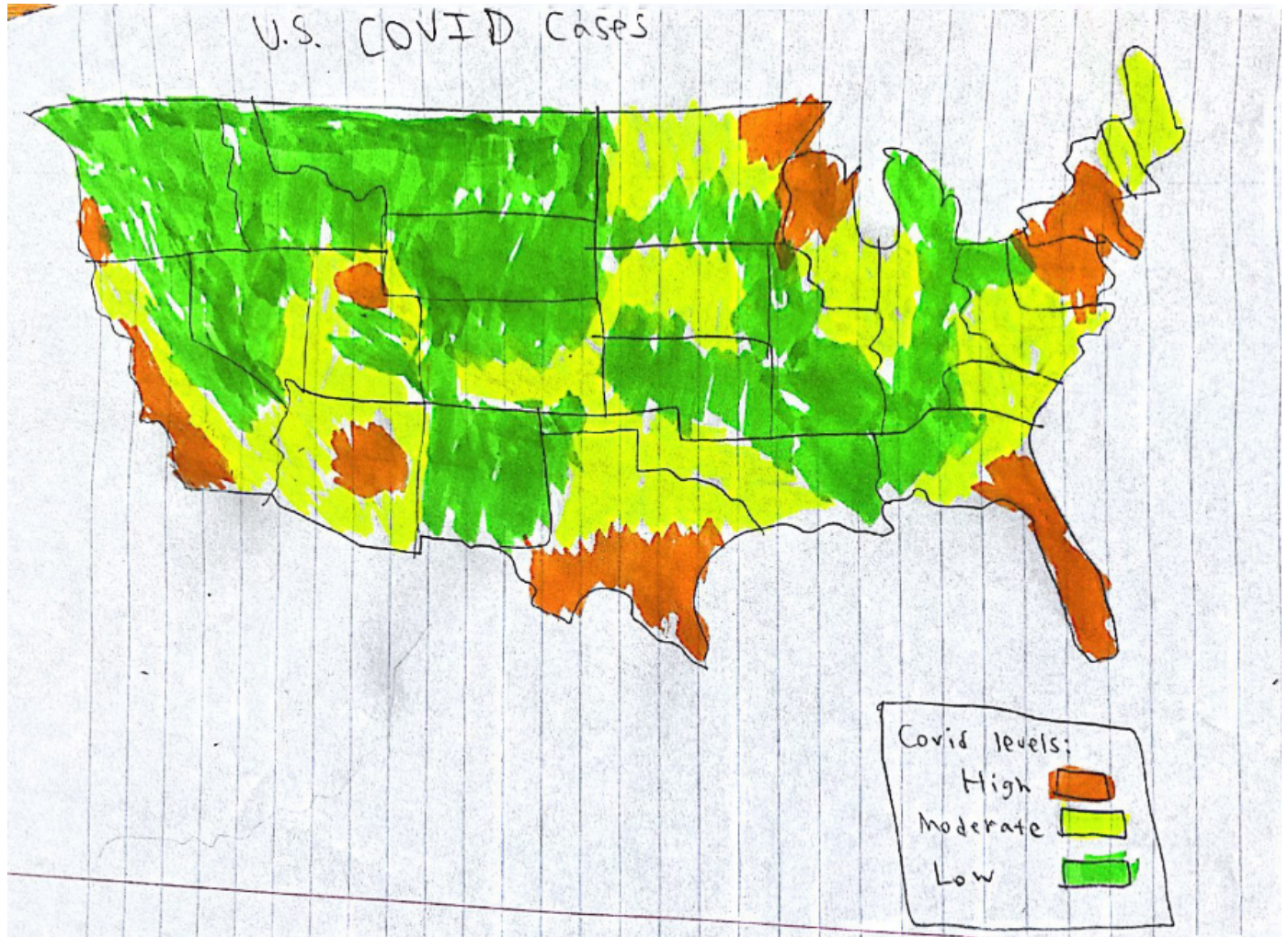
Estimates of Vaccine Hesitancy (2021)

<https://data.cdc.gov/stories/s/cnd2-a6zw>

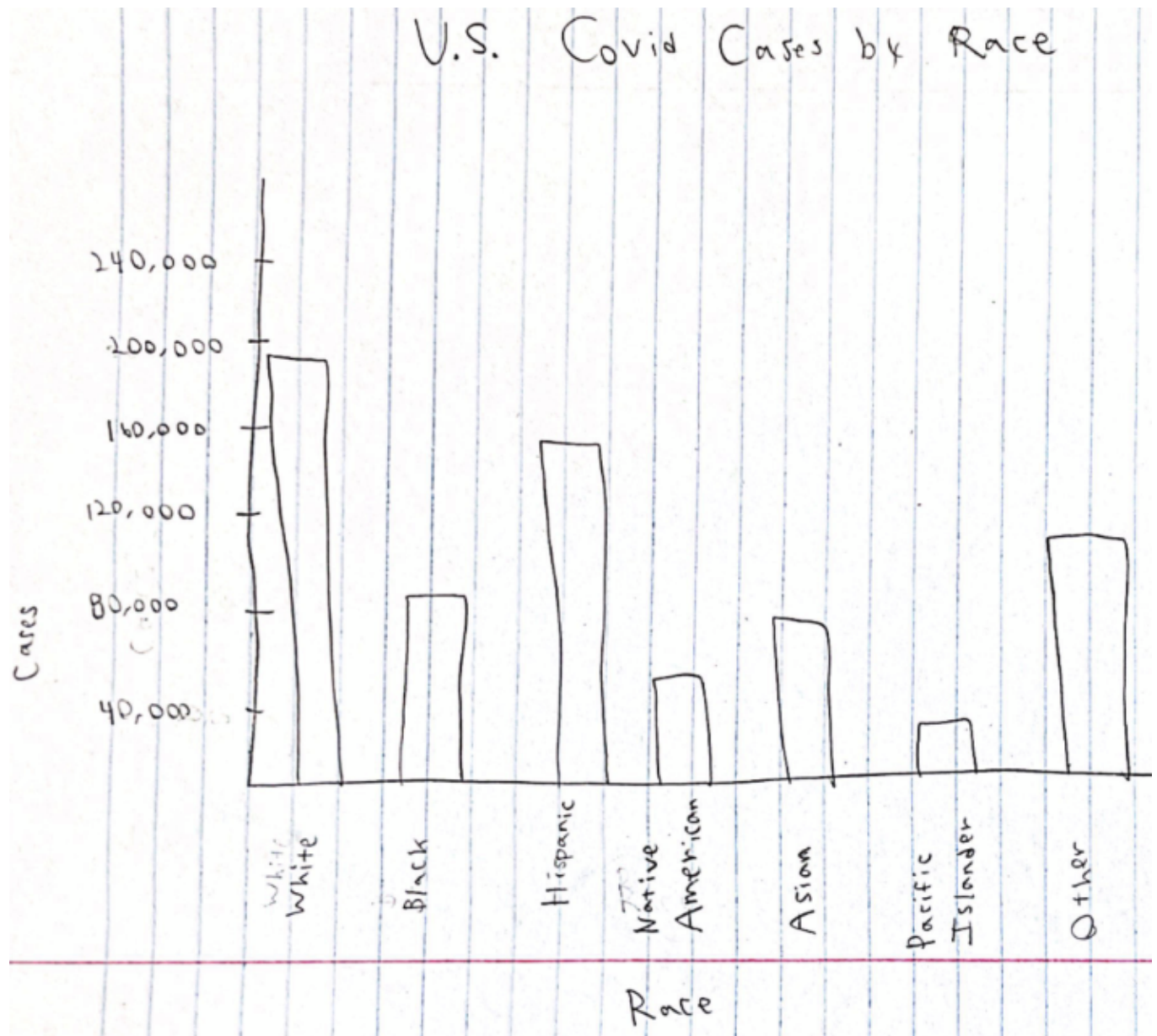
Data Processing:

We plan to clean up the data. Some of the rows have missing data which could mess up our visualizations. For example, if age is missing we might go through our data and replace the age with -1 so we can dynamically ignore it in our code. We will also most likely go through and delete rows that won't be relevant to our project. For example, there are hospital and county IDs in our dataset that won't be useful to us without further context. The dataset is close to 14GB so depending on performance we will have to delete some of these unnecessary rows of data.

Visualization Design:



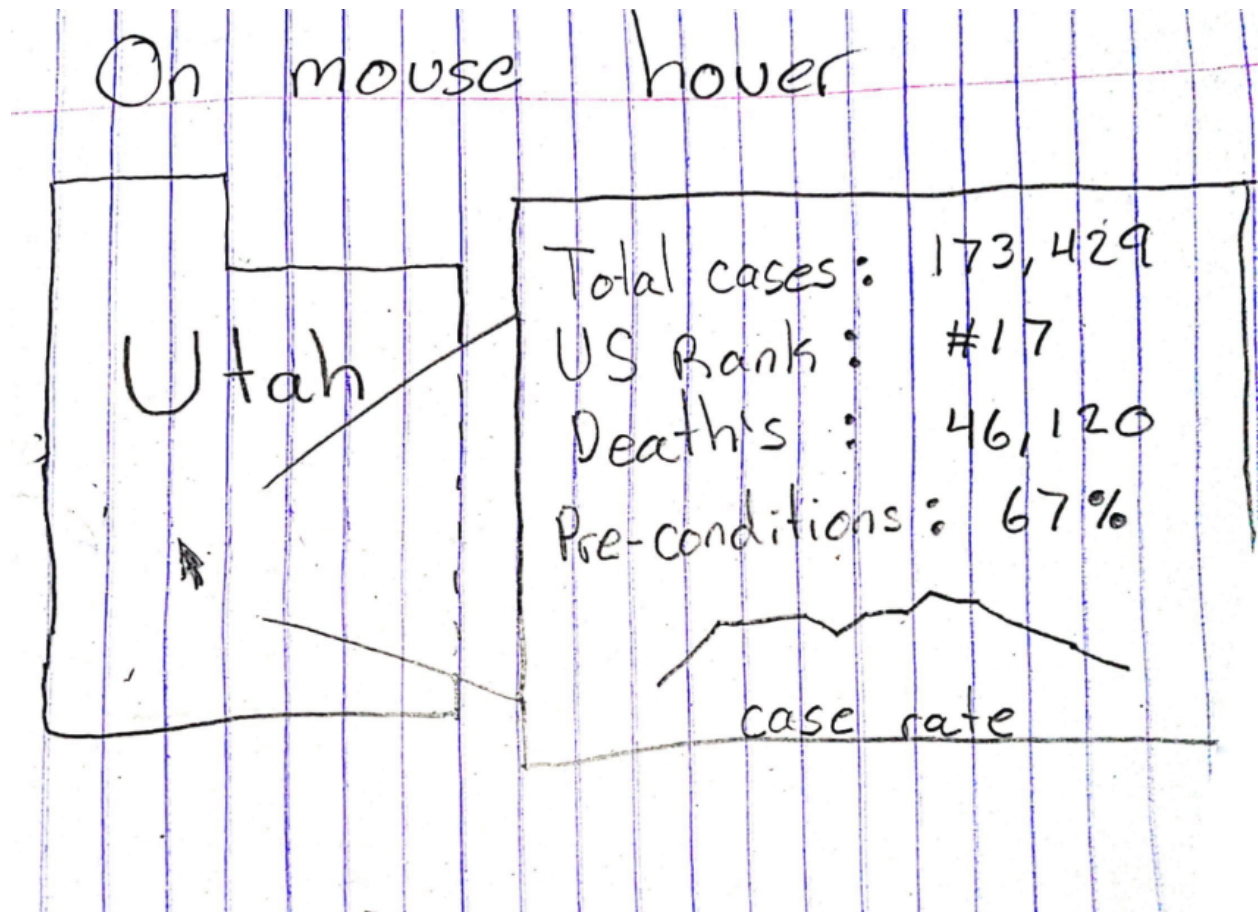
We chose a heatmap as one visualization because it allows the reader to see where there is a large amount of COVID cases. If a reader sees their state is in the red, they will begin to take precautions and it will help the person viewing the chart better understand where COVID is impacting more.



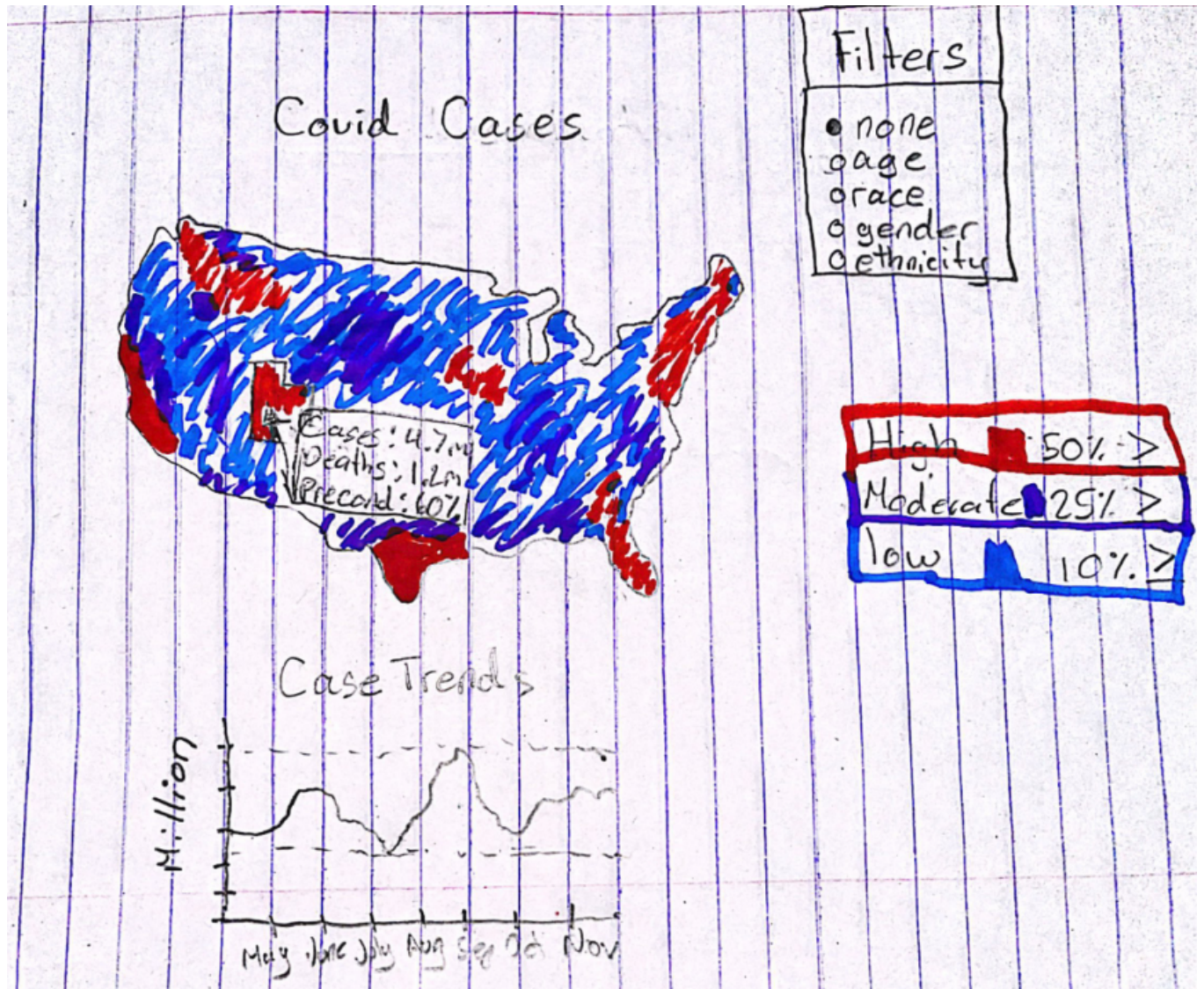
For the bar graph, we decided to show the COVID cases on race to see if certain races are being impacted more by COVID and have higher positive cases. This design however might be changed to show the percentage of each race that has gotten COVID rather than just total numbers since the total number will most likely result in ethnicities that don't make up much of the population showing low cases when in reality the transmission rate might be higher in that ethnicity versus other ethnicities. This could also be used to visualize other variables in the data.



We also considered using line graphs as a visual encoding to be able to show trends in the data. We plan on adding light lines on the x and y intercrosses for easier readability. Here we modeled what we expect an age line graph to look like from what we know about COVID-19. This will be helpful so that individuals are aware of the current trends in COVID-19. It might be the case that there are unforeseen spikes. For example, the fatality rate at 50 may be a lot closer to that of 80+. This could spread awareness to people who shrug off the disease as something that their age group doesn't need to worry about.



We also plan to add on mouse hover functionality. Here we have a prototype of the type of encodings we wish to display when a user hovers over a state. They will be shown relevant information as well as a quick-line graph of past trends in the data. We think seeing the raw numbers is a good way to showcase this information since if the user is hovering over the state it is most likely because they are interested in a deeper dive into the data. We save this for on-hover functionality so the user isn't overwhelmed by all the data at once.



This is one possible end prototype that incorporates aspects of all our ideas. We maintain the on-hover functionality but decrease the amount of information shown. We moved the bar graph to the bottom because we think the information is crucial enough that the user should have quick access to it. We also plan on adding buttons that will allow the user to filter the heat map based on that table in the dataset. This is to provide users with a better understanding of the data. We want to add this because it's not immediately obvious that hot spots will be the same amongst the varying variables.

Must-Have Features:

Our project must have an interactive heat map of the country that will allow the user to hover their cursor over the states and get information on that state's COVID-19 case statistics. We will also have a filter that allows the user to filter the visualization. For example, limit the visualization to people above 60. The visualization will also be normalized based on the population. There will also be vaccination data when hovering over a state based on census data in 2021.

Optional Features:

We will allow the user to change the color map of the heat map. For example, changing the default hue color map to grayscale. We also plan to add supportive line graphs below the colormap when hovering over a state. We are mainly planning on overlapping the state's normal distribution with the country's based on the statistics they're hovering over. Adding a zoom option would also be something we'd want to do if we have time.

Project Schedule:

Planning on all work being partner based.

Week 4: Finish Project Proposal

<https://www.dataviscourse.net/2023/project/#project-proposal>

Week 5: Project Review w/ Staff (scheduled w/ staff)

Week 6: Update our project book

Week 7: Start cleaning data from the dataset

Week 8: Fall Break

Week 9: Begin coding and forming visualizations

Week 10: Finish our prototype

Week 11: Milestone

Week 12: Peer Feedback

Week 13: Applying Peer Feedback, try to add optional features

Week 14: Finalizing Project (Thanksgiving Break)

Week 15: Final Submission