Joshua Tlatelpa-Agustin

Kian Akhiary

Process Book

Week 4: Finish Project Proposal

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

For this week we finished creating our project proposal. We decided to change up what we are doing for our project and have decided to go with a proposal dealing with COVID cases, deaths, etc. by different variables like ethnicity, race, gender, etc.

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

We met together and discussed what we will talk to the course staff about and what we might want to bring up during the meeting. After the meeting we talked to each other about what the TA brought up during the meeting and how we might address what the TA told us during the meeting about our project.

Week 6: Update our project book

We began looking into how to view the data we gathered from the CSV. Currently our CSV contains every single individual COVID case in the US up until the point of when we downloaded the CSV. So this is more than 100 million cases, accessing this will most likely take a long time so we will need to figure out how to just total all this data. Currently we are thinking of either using R or pandas to clean this data up.

Week 7: Start cleaning data from the dataset

This week I started cleaning the data using R Studio. I used this in my stats class so I was quite familiar and I figured it'd be the easiest way to clean so much data. Considering that

our csv file is around 14gb it took a long time to import the data into R. On multiple occasions, it took around one hour to import the dataset each time we started up our project. We spent a good amount of time optimizing so that we would not have to deal with the loading times. Josh ended up upgrading the RAM on his computer since we noticed it was crashing because of insufficient RAM. We also dug into the implementations of various libraries to find the most efficient one. We discovered the native library to read csv files was the most stable with low ram systems despite others being more highly rated online. We then explored the data by printing unique tables, the min/max values, and identifying missing variables. Each of these queries took around 15 minutes each. We cleaned up data and removed unnecessary info such as NA values when finding mean. Took quite a while to figure out what data we wanted to use for our visualizations and how to filter out the data we wanted. We decided on using case_month, res_state, age_group, sex, race, ethnicity, death_yn, underlying_conditions_yn and ignoring state_fips_code, res_county, county_fips_code, case_postive_specimen_interval, Case_onset_interval, process, exposure_yn, current_status, symptom_status, hosp_yn, Icu_yn.

Week 8: Fall Break

This week was fall break and we decided we will take the week off and begin working on the project again after break with fresh minds!

Week 9: Begin coding and forming visualizations

For this week, we began creating some of the visualizations of the project. We began by creating a heatmap of the United States which will be colored based off of total covid cases for now. We also created a line chart that for now will be using mock data, for now we want to just make sure the structure works correctly and will be easy to work once the data is ready for it. We also began to create a filter that will be implemented in the future that will filter the maps based on what the user selects.

Week 10: Finish our prototype

For this week, we finished setting up the heat map and line chart. Currently they are using mock data that was set up just to make sure that it correctly works and has no problems with the heat map and line chart. The structure of the filter the user can use if they want has also been set up as well, this however will remain unfunctional for the time being. A legend was also created for the heat map.

Week 11: Milestone

We compiled the dataset and cleaned the data and put the total cases into a JSON file. In this JSON file, each state's total COVID cases are listed. By doing this we no longer need to access a very large CSV of COVID cases that takes several minutes to access and we can now access the total COVID cases quickly. Using this data, we combined it with the visualizations we had in place and we have replaced the mock data of the heat map with the actual data. We also styled the visualizations to look nicer, they now have titles and are designed better.

Week 12: Peer Feedback

For this week, we gave our peers feedback on their projects. We looked at their projects and saw how they were coming along and gave honest feedback to help them get their projects finished up. We also began implementing the final parts of our project, starting with setting up the COVID chart to use actual data rather than the mock data we had in place.

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

We reviewed the feedback our peers gave us on our milestone. Looking at this feedback helped us better understand where we could use room for improvement on our project. Some of

the feedback given also suggested we add more stuff to the project, many of these were already planned on being implemented! We also began to finish cleaning and getting the rest of the data we would need to get the visualizations all working. As stated earlier, since it is a large amount of data it would take awhile to go through the data and get it all.

Week 14: Finalizing Project (Thanksgiving Break)

For this week we began to finish implementing the final features of our project. All the data that we will need has been cleaned, and stored in our repo, since we totalled the data, accessing these totals is now much easier and faster. Filter box has been set up correctly so it now updates the map to display values of whatever filter is selected. If the user selects deaths for example, the heatmap will be updated to show deaths instead. Changing to cases makes it show the cases of each state, etc. The chart for vaccine hesitancy has also been changed to a bar chart instead of a line chart and it now actually uses the actual data rather than dummy data. The chart now displays every state's vaccine hesitancy rate.

Week 15: Final Submission

For this final week, we got the last parts of our visualizations working. We added a pie chart that will show up when you hover over each state that compares that state's stats to the entire country. This helps the user better understand how big or small one state's stats are in relation to the entire country. For example, for COVID deaths, when you hover over California, you can see that almost a quarter of COVID deaths were in California, while states like North Dakota had a very small number of deaths compared to the rest of the country. We also created a video showcasing our visualizations and deployed