DV Final Write-Up

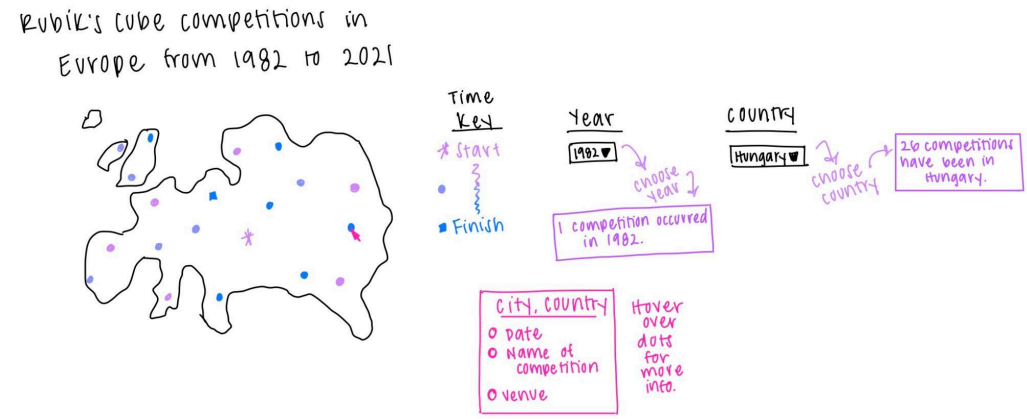
Data Set:

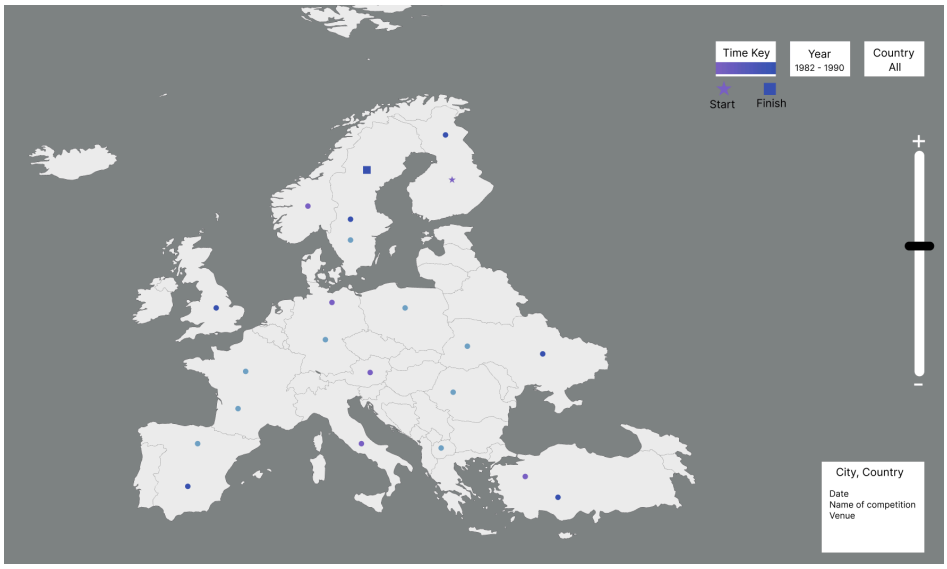
Upon searching for data to use for this project, we wanted to find something surrounding geographical data. At first, it was locations of hippos, but we settled on Rubik's Cube competition locations. We found a compiled dataset (.CSV) from Kaggle (found here), which contains data from the World Cube Association (WCA). The WCA has a wide array of data, spanning from competitors with the best medal collection, or the average event count by competition by country. It was also necessary to find a "geojson" of Europe, which I found here.

This dataset contained upwards of seven thousand entries, data we thought would be manageable to display, but quickly realized it was a task we were not quite suited for. We decided to shrink our scope down to just European countries because Europe has had the most competitions, the Rubik's cube was created in Europe, and the first competition was held in Europe. In order to filter the data down to just European countries, we required the help of ChatGPT. This aided in the production of a Python script that utilized the Pandas library to create a new .CSV file that contained just European countries. After we had this filtered dataset, we found it necessary to generate coordinates for the given country and city for each competition. This was done so using a Python script that utilized the Google Maps API, specifically the "gmaps.geocode" function to retrieve both the latitude and longitude for each location. Pandas was used to create a new .CSV which added these coordinates to each competition within the dataset. The Google Maps API required an API key, which was obtained without much trouble (and without money). After the data was filtered and coordinates were generated, we found that there was still data wrangling that had to be done.

When condensed to only European countries, there were 495 missing coordinates along with about 100 incorrect coordinates. Therefore, we used a spreadsheet to find the cities where the coordinates were missing and manually inputted those coordinates. For the incorrect coordinates, we found them using our visualization. So, we would find dots that were in the ocean, find the competition in our spreadsheet, and input the correct coordinates for that location. Overall, we inputted or corrected over 500 coordinates to ensure all of our dots were in the correct places on our map.

Prototypes:


Rubik's cube competitions in Europe from 1982 to 2021
Time Key
Start
Finish
Year
1982
choose year
1 competition occurred in 1982.
Country
Hungary
choose country
26 competitions have been in Hungary.
City, Country
Date
Name of competition
Venue
Hover over dots for more info.


Time Key
Start
Finish
Year
1982 - 1990
Country
All
City, Country
Date
Name of competition
Venue

Task Analysis:

Using this map of Europe and different colored dots, we hope to represent the locations and timing of Rubik's Cubes competitions from 1982 to 2021. Users can analyze the data by identifying where most competitions were held based on the density of the dots on the map. The user can then hover over the dots in order to get the exact location of the competitions with city, country, and venue name listed. Furthermore, users are able to identify when most competitions were held based on the color of the dot representing the competitions. Utilizing the key on the side, users can tell which year had a larger number of competitions based on how many dots are that specific color. Along with this, users can compare the data using this map. We have added the filters so users can select different years and countries to display on the map. Therefore, users can compare how many competitions are held in those different years and countries. As a result, users can identify which country or year is especially popular for Rubik's Cubes competitions. For example, the popularity has increased in 1982. Therefore, a user can see the number of dots increasing over the years. The first competition was held in Hungary. So, the user can also see how competitions expanded geographically. Also, the user can identify ranges as they can see the countries and years with the highest or lowest number of competitions. Finally, users can search the data using our filtering system. They can select one year, all years, or any combination of years between 1982 and 2021 to display on the map. They can do the same filtering with countries. As a result, a user could search for a specific competition by selecting a specific year and country. Therefore, our visualization allows users to analyze, compare, search, and identify ranges for the data set of Rubik's Cubes competitions from 1982 to 2021.

Accessibility Analysis:

Our goal is for any user to be able to access our visualization and complete the same tasks. Therefore, a user can zoom into our map up to 400%. Using SVG's for our dots ensures that the quality of our visualization stays high as a user zooms in. So, a user with low vision who needs larger visualizations can zoom in and see the data points and country borders. Along with this, we need to make sure users with color blindness can properly utilize our map. So, we have highly contrasting colors in accordance. We also utilize luminosity to ensure colors are distinguishable in black and white. Furthermore, we avoid colors that affect many people with colorblindness such as red and green. For users with very low to no vision, we ensure our tasks can be done using a screen reader. We used WAVE to do this as it shows us what a screen reader could do with our visualization. We added alt text to describe our map, and the screen reader can access the information the tooltip displays.

The tasks we would like users to be able to complete are analyze, compare, find ranges, and search the data. Therefore, a user with visual impairments should be able to complete those tasks as well. For analyzing data, we have alt text to describe the map and our visualization. Also, a screen reader will be able to identify our filtering methods and information that the tooltips display. For comparing data, the user will be able to select different countries and years with the help of the alt text, zooming, and screen reader. These features and the screen reader will also ensure that the user can find ranges and search the data. Through these methods, the different colors of the dots and the ability to see the density of the dots are lost. However, rather than see the color of the dot to tell which year the competition is, the user can access that information with the tooltip and screen reader. Also, the user can also access the 'density' of the dots using the screen reader. Overall, our visualization is accessible to users with visual

impairments or blindness. Although they will need access to a screen reader and our alt text, they will be able to complete the tasks. Information that comes from seeing the map is lost such as density of dots, borders of countries, and color of dots. But, the user will still be able to find all of the data that is important to complete the tasks.

First Pilot:

To conduct our test, we edited the given script by customizing the words to our visualization. Then, we met our participant at Union. Our first pilot was Maddie, a junior biology major and public health minor. David was the note taker and used his computer while Ally was the speaker who showed Maddie the visualization on her computer. Our script was as follows:
"We are evaluating our visualization and are asking you, the participant, to complete some tasks using the visualization and then provide feedback about the visualization and experience. As a reminder, we are evaluating the visualization, not you as a participant, so you don't need to worry about being "right" as you complete these tasks. There are three tasks, followed by a brief feedback session. The whole pilot session should take under 5 minutes. Do you consent to participate?"
[Wait for yes]
Speaker: "Thank you for agreeing to participate. We will start with the three tasks. Please 'think aloud' as you complete the task, meaning voice what you are thinking as you work through the task. Your first task is to analyze the data. Are you able to identify when or where many competitions were held?"
* You may answer clarifying questions, such as "what do you mean by x?" or "are you looking for an exact number or an approximation?"

[Pause to allow the participant to complete the task]

“Your second task is to compare the data. How do different countries or different years compare to one another?”

[Pause to allow the participant to complete the task]

“Your third task is to search the data. Are you able to seek out a specific competition, a specific venue, a specific country, or a specific year?”

[Pause to allow the participant to complete the task]

Speaker: “That is the end of the third task. For this last bit, we welcome any feedback you may have about the visualization or about your process for completing the tasks.”

[Allow participant to speak first, then informal discussion]

To analyze the data, Maddie zoomed in, filtered the data based on year and country, as well as hovered the dots for the tooltip to come up. Then, to compare the data, she focused on comparing different countries to each other. Therefore, she selected France, Germany, and Poland to see how competitions were dispersed in each country. Finally, while searching the data, she sought out the most recent competition in Reykjavik, Iceland. She found that the most recent competition was in 2020. While exploring our map, Maddie was able to identify areas that held many competitions based on the density of the dots. She specifically pointed out Gutersloh, Janusz Korczak in Germany having a cluster of dots. Using the tooltip, she could see that each competition was held in the same venue, but at different times. She could also tell that the competitions occurred during different years because each dot was a different color. Because the same venue was used multiple times, many dots are layered on top of each other. So, Maddie said that this hinders the ability to see or hover over some dots. We asked if she would prefer larger dots to represent venues used multiple times. Maddie liked the idea but said large dots

might cover up smaller dots that represent a venue that is nearby. At this point, she did ask what the colors represented. Therefore, we will change the visualization by adding a more prominent and clear key to show what the colors relate to. During the feedback session, her first question was if the winner of the competition was displayed. David and I thought this was a very good question. However, our dataset did not include the names of the winners, so we could not add that information to our visualization. Furthermore, Maddie commented on the good zoom-in quality and said she could easily navigate the map. Along with this, she said that the colors were very pretty, however someone with colorblindness might struggle as some of the colors are red and green tones. From this first pilot, we will change our colors. We must ensure that the contrast is high, the luminosity is good in black and white, and the colors ensure people with colorblindness can distinguish between the dots. We will also consider making dots larger if there are many competitions at the same venue. However, we need to make sure the larger dots do not cover other competition. Finally, we will add a very clear and understandable key for the colors used.

3 Other Pilots:

Our second pilot was Henry Chesley-Vogels, a chemistry major and music minor. We read the script and took notes while Henry attempted the given tasks. The first task was to analyze the data. Specifically, analyze where many competitions are held based on the density of the dots as well as when many competitions are held based on the color of the dots. He performed this task by zooming in and navigating the map closely. He hovered over dots to see exact locations and dates. He noticed that there were a lot of different cities in the Netherlands. Other countries held competitions in the same cities and even the same venues over and over.

However, the Netherlands held competitions in a variety of cities and venues. The next task was to compare the data by comparing different countries and years. Henry used our filtering to achieve this by choosing specific countries. He compared the amount of dots in one country to another as well as selected multiple countries at the same time. He also filtered by year to see how the number of competitions changed over time. He was interested in Covid times to see how the pandemic affected the competitions. He found that competitions did decrease due to Covid, but did not go away completely. The final task we asked him to complete was to search the data. He decided to search for a competition in Norway in 2010. He found that there was a competition in Oslo and Trondheim in 2010. He did this using our filtering again as we selected Norway and then 2010 to display 2 competitions. During the feedback session, Henry asked where the years 1983 to 2002 went. We informed him that there were surprisingly no competitions during those years. Overall, Henry was impressed by our visualization as he felt he could complete the given tasks well. He also liked the navigation, zoom, and filtering.

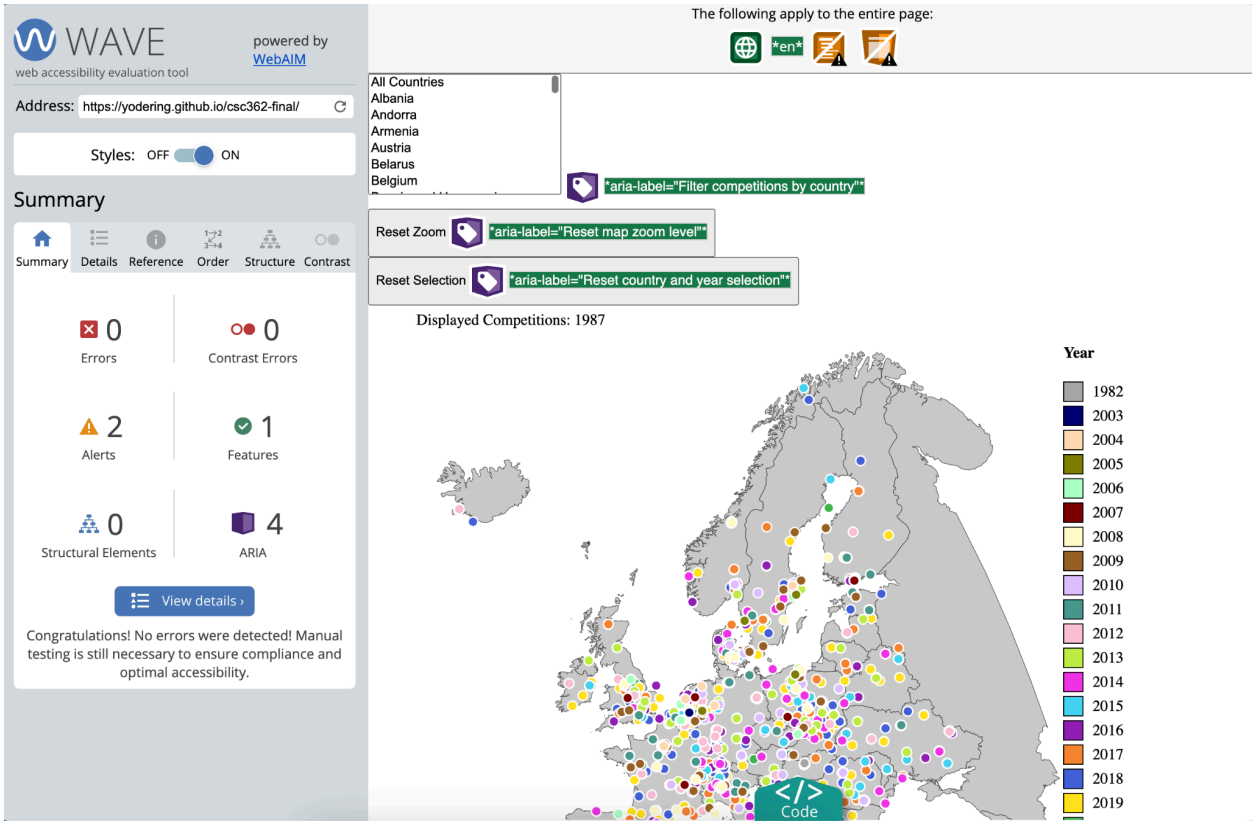
Our next pilot was Cooper Strauss, a political science major and classic minor. Once again, we read the script to Cooper, gave him different tasks to complete, and took notes as he executed the tasks. As we gave Cooper the task of analyzing the data, he noticed that there were a lot of competitions in southern Europe. He also noticed that there was a significant jump in the number of competitions between 2010 and 2014. He completed this task mainly by filtering by year and zooming in to explore the different dots. Then, we gave him the task of comparing the data. He selected one year at a time to see how competitions spread over time. Cooper observed that the first competition may have been in Hungary, but the overall trend is that competitions started in western Europe and moved east over time. Then, he used our filtering to see how Covid affected the competitions. Similar to Henry, he found that competitions decreased, but did

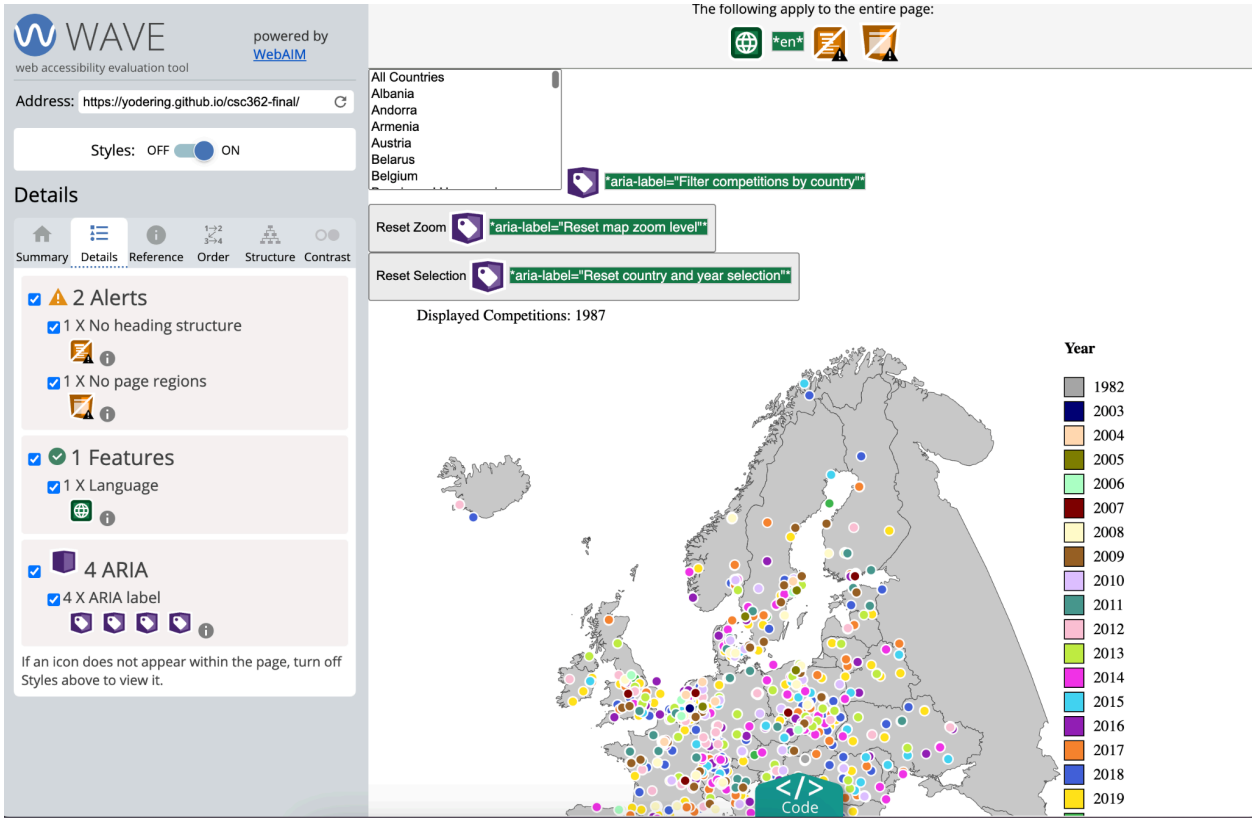
not end completely. Finally, the last task was to search the data. Cooper wanted to find the competition that is the most east on the map. So, he selected the eastern countries and zoomed in to find the dot that was the farthest right. Then, he hovered over the dot to see the city, venue, and date of the competition. During the feedback, Cooper expressed that he would have liked to click on the actual country on the map and have the number of competitions displayed. However, we have a filtering system where users can select the country on the side and then the number of displayed competitions is shown below. From our observations, we think Cooper was able to complete all of the tasks efficiently and was able to come to good conclusions based on his exploration of the visualization.

Our final pilot was Mickey, a psychology major and public health minor. For the final time, we met with Mickey, gave her our speech, gave her tasks, and took notes while she completed the tasks. For the first task of analyzing the data, Mickey zoomed in and noticed that there was a high density of dots in Belgium and Netherlands. She said that because these countries are smaller than the ones surrounding them, maybe that is why the dots are so dense as competitions are held closer together. She also noticed that there are a decent amount of competitions in the United Kingdom. Then, she started to select different years and found that starting in 2009, the number of competitions significantly increased. To compare the data, Mickey wanted to compare Denmark and Sweden because she studied abroad in both of those countries. She determined that they have relatively similar amounts of competitions. However, she noticed that Denmark's first competition was in 2008 while Sweden was having competitions before then. Finally, to search the data, Mickey focused on Belarus in 2019. She used our filtering and zooming to hone in on Belarus. She found that there were 10 competitions in Belarus in 2019. She also found that there was a competition in Novopolotsk on December 21st.

During our feedback session, Mickey suggested possibly adding labels to the countries so that users can see what the countries are without having to hover over the dots. We think that this might make the visualization more cluttered as the dots and labels would often overlap. However, Mickey was able to complete the tasks very well and navigated the visualization efficiently.

Accessibility Checks:




WAVE
web accessibility evaluation tool
powered by WebAIM
Address: https://yodering.github.io/csc362-final/
Styles: OFF ON
Details
Summary
Details
Reference
Order
Structure
Contrast
2 Alerts
1 X No heading structure
1 X No page regions
1 Features
1 X Language
4 ARIA
4 X ARIA label
If an icon does not appear within the page, turn off Styles above to view it.
The following apply to the entire page:
en
All Countries
Albania
Andorra
Armenia
Austria
Belarus
Belgium
aria-label="Filter competitions by country"
Reset Zoom
aria-label="Reset map zoom level"
Reset Selection
aria-label="Reset country and year selection"
Displayed Competitions: 1987
Year
1982
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Code

Personal Reflections:

Ally: I have really enjoyed this project because I think accessibility in computer science is often overlooked. Last year, I had a Verna Miller Case poster about how to make STEM classrooms more accessible to students with blindness or visual impairments. Therefore, I have looked into screen readers, contrast, and ways in which to make coding more accessible. Also, my dad works in blind rehabilitation, so he teaches children with little to no vision how to navigate the world. As a result, this project really piqued my interest as I think math, computer science, and data visualization all need to do more for people with visual impairments. During the design process, David and I wanted to showcase the entire globe. However, the dataset would have been too large and difficult to clean. So, we cut it down to Europe because the first competition was in Hungary. I think this was a good idea because it is easier to compare countries and regions in a single continent rather than seven. Also during the design process, we wanted to make sure users could filter the data and display the years and countries they want to see. Cleaning the data was a grueling process as we had to manually input many coordinate. Because we used a map, a screen reader and alt text are necessary for users with visual impairments to access the data. Therefore, we made sure to use WAVE and add alt text to describe the map. The testing process was interesting as we asked 4 people who are not familiar with computer science. This gave us very unbiased feedback because they did not know to look for certain features or attributes. The testing gave us insight into how to make our visualization more understandable for a user that is not us, the creators. Since we made it, we know what the colors mean, where we can select years or countries, and that we can zoom in. However, a random user does not have that knowledge and we were able to add clearer features based on

those testers. Overall, we created a very cool visualization that many different kinds of users can access.

David: This project put a lot of things into perspective for me, especially considering how I've seen the popularity of the Rubik's cube throughout time, and more recently, the explosion in this popularity. The Rubik's cube definitely became a "quarantine hobby" for most people, so in present times there is a staggering amount of competitions in comparison to the past. However, being able to see where competitions were held and the frequency of them makes it fascinating to observe the trends over the years. I find that data visualizations become the most compelling to make and analyze when they're about a topic you *actually* enjoy. I could not find any Rubik's cube data visualizations, so from the start of the semester, I wanted to create something using data from the World Cube Association. I had known of the WCA's statistics page, but was not sure if I would ever be able to make something with what was found on there. I feel very accomplished having created something I wanted to from the beginning.

Aside from my likeness for the dataset, I thoroughly enjoyed seeing the piloting of the visualization and the other testing. I had some experience with this last semester when I took Critical Web Design, so being able to put into practice the use of constructive criticism was something I found enjoyable. I have always been fond of viewing differing perspectives, so being able to see different perspectives on my own work was fulfilling.

I very much enjoyed the focus on accessibility for this project, an aspect of STEM I've felt to be underlooked, but slowly garnering more attention overtime. The truth is that, in my other projects, I have not considered accessibility as something to implement, but will now take the time to do so.

Altogether, this project brought together my favorite topics and taught me new things; I feel quite accomplished.