Alvaro Reynoso
COMP_SCI 333
Professor Jessica Hullman
Assignment 3

Interactive Visualization

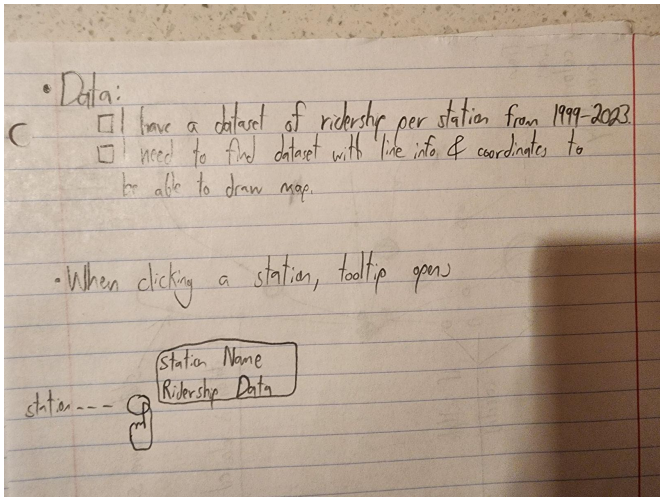
Link to Visualization: https://alvaro-regu.github.io/

Data Domain

For my visualization, I decided to look at the ridership data of Chicago's L ridership. I found a dataset covering the per station ridership numbers. It covered every month from 1999 to 2023 with total monthly ridership as well as the average daily ridership for that month. This dataset was found in https://data.cityofchicago.org/ which is the website the city has to access public data. Two limitations of this dataset was the lack of mention of which line each station belonged to. Additionally, since I planned to draw the map, it would have been useful to have the coordinates for each station. I obtained my other 2 datasets from a public repository which showed me how to draw the train map. This data set included the line location and coordinates as well as other extraneous variables I did not use.

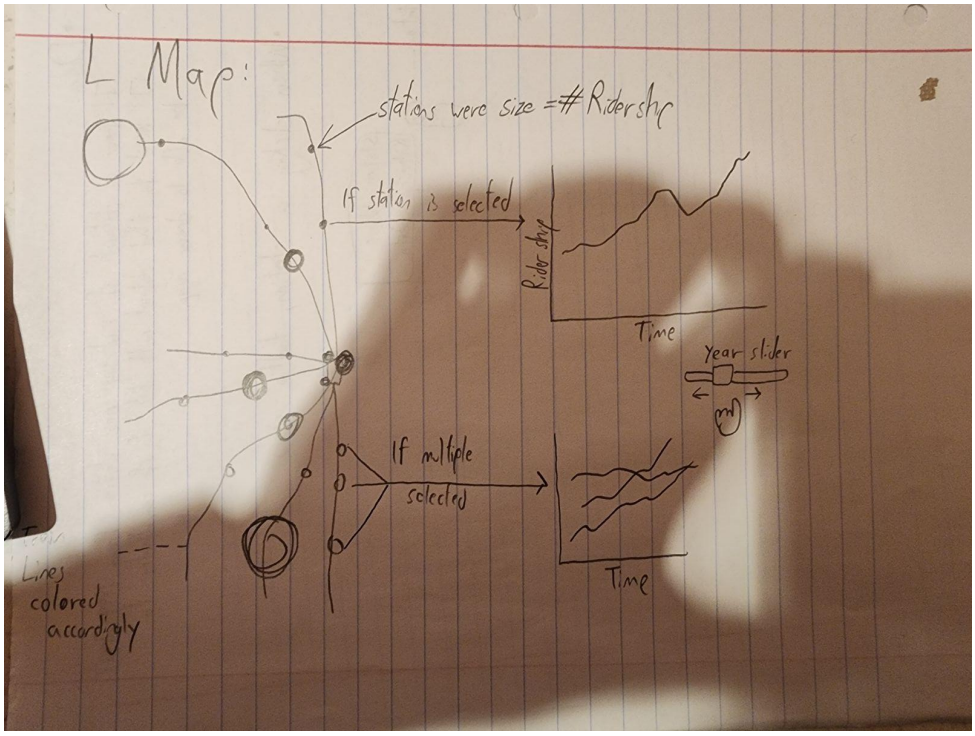
Storyboard

I began the storyboard section of this project by describing the data I had and the data I needed. This is essentially the same as what I mentioned in the previous paragraph. Below the data, I sketched my idea for a tooltip. I wanted the user to be able to click on a station and have a tooltip displaying the station name and ridership data pop up next to it



Afterwards, I designed a more complete vision for what the visualization and its interactions would look like. First, the entire train system would be drawn with each individual line being colored the corresponding color. The stations would be placed on the correct locations all across the lines. This is represented on the left side of my sketch. On the right side, in between the graphs, you can see a slider which would allow the user to select the year. If you look at the stations, they are different sizes. This would represent the ridership for that year. As

the user slides the timer slider, the stations on the map would grow and shrink accordingly. When the user clicked on a station, a line chart would show up displaying the change of ridership over time. Additionally, the user would be able to choose multiple stations to compare their data.



Changes Between Storyboard and Final Design

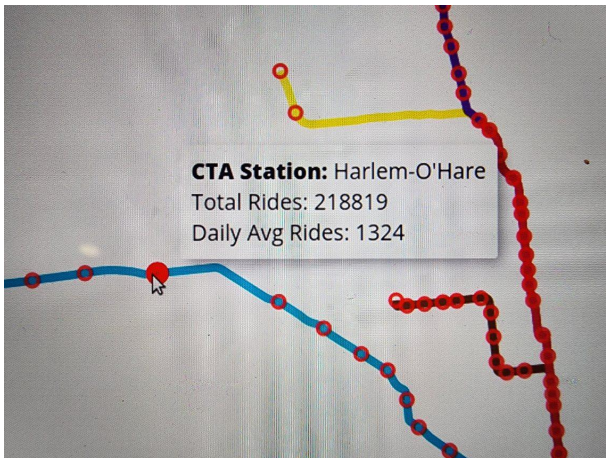
The biggest change I made was removing the entire station sizing functionality. Although I thought it was a good idea, the way the train system is designed itself presented issues. Because of the density of the station across the line and of the loop, there was a lot of overlap between the stations. This caused overlapping that made visualization seem very overplotted. I considered removing the stations that had really small ridership but this made it so the user could not select that station for comparison if they wanted in the line chart. Additionally, because of the train system design, the loop got by far the most ridership to the point that it dwarfed all other stations making for giant circles in the center that blocked each other and surrounding smaller stations. I was not able to come up with a satisfactory solution to implement this idea so I decided to simply get rid of it after spending too much time on it.

Everything else remained generally the same. I changed the tooltip popping up with hover rather than with a click. Additionally, I made it so that there was a smaller zoomed in window in the corner for the loop stations because of how crowded they were in the full sized map.

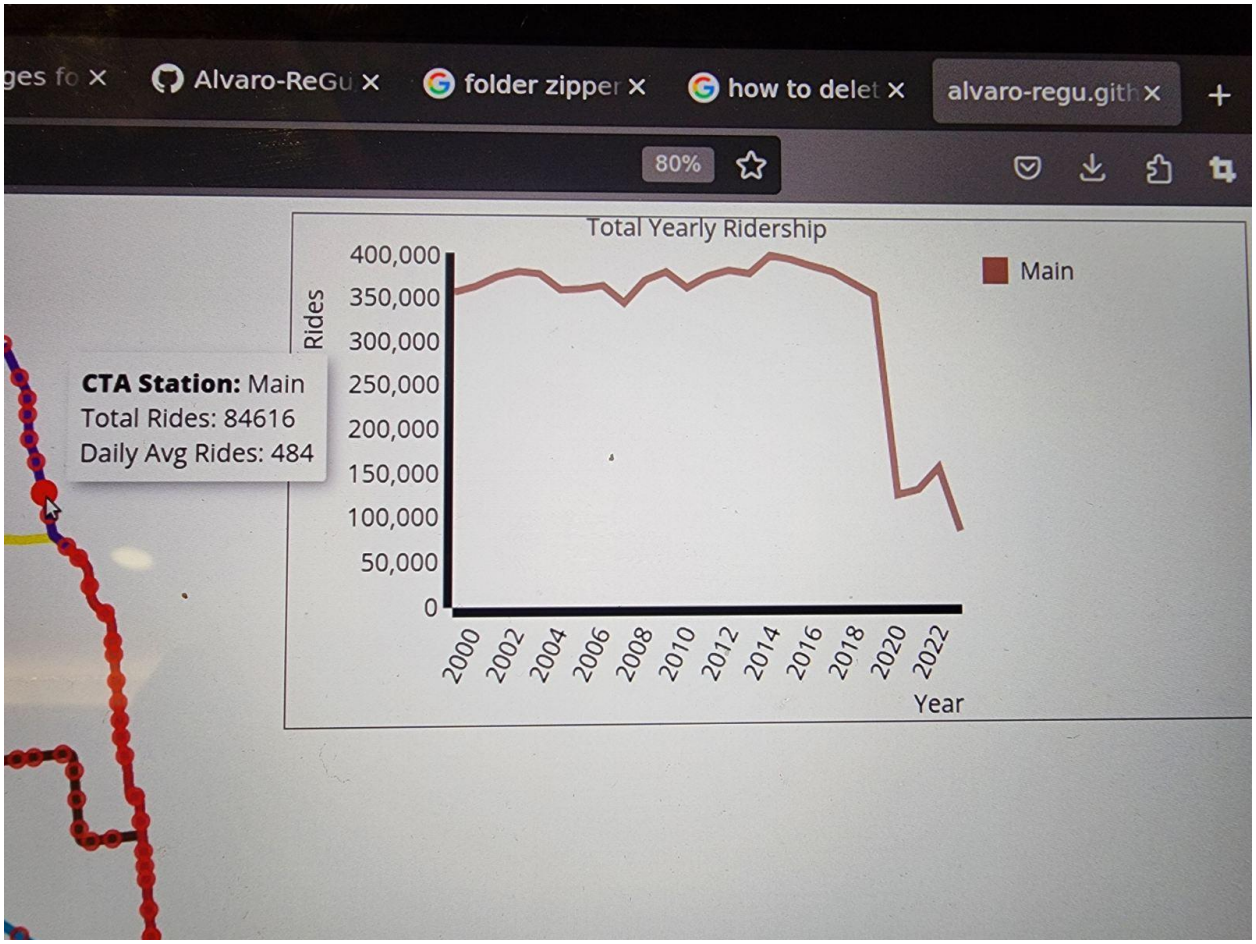
Final Application

This is how the final visualization looked:

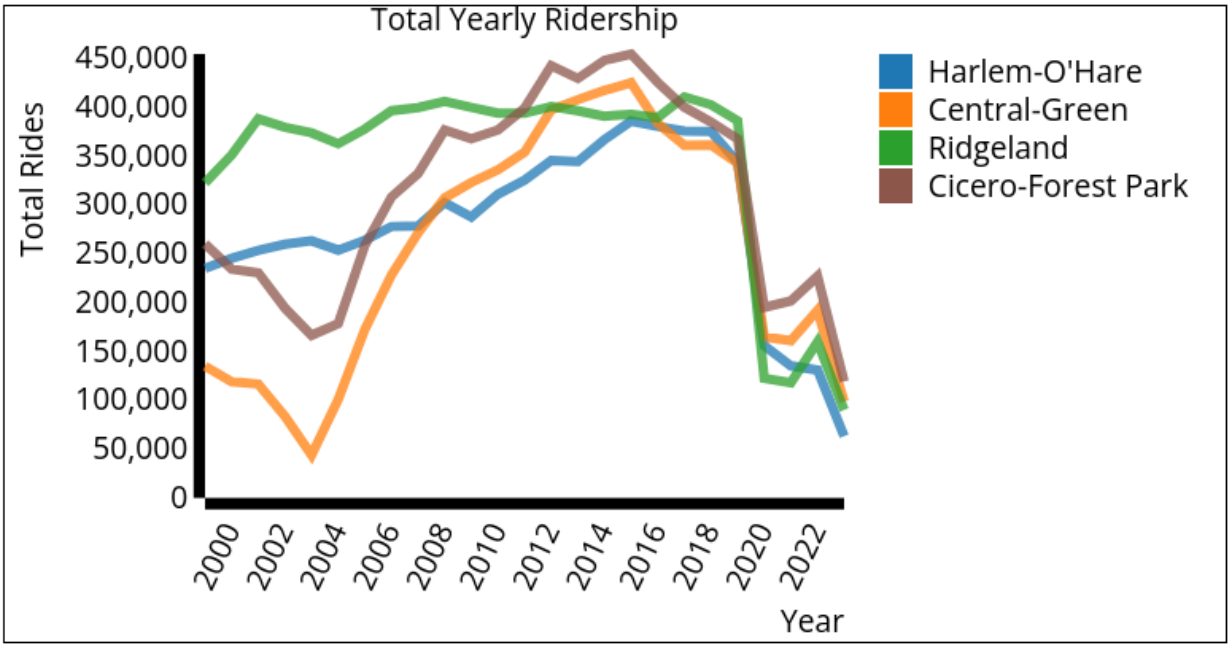
We have the map of the train in the center of the page with each line colored accordingly. Each station is represented by a red dot and is placed where they would actually be located. We have a simple title describing what the data is and a few short sentences explaining a few things. If, when a station is clicked, the graph overlaps with a map, you can zoom out. This is to account for differences in screen size. I also explain that only up to 5 stations can be chosen at a time and that you click a station again to remove it from the chart. Lastly, I explain what the zoomed in window for the loop is for those who don't know what the Loop is. I also have a toggle below where you can select which year to look at for the tooltips.



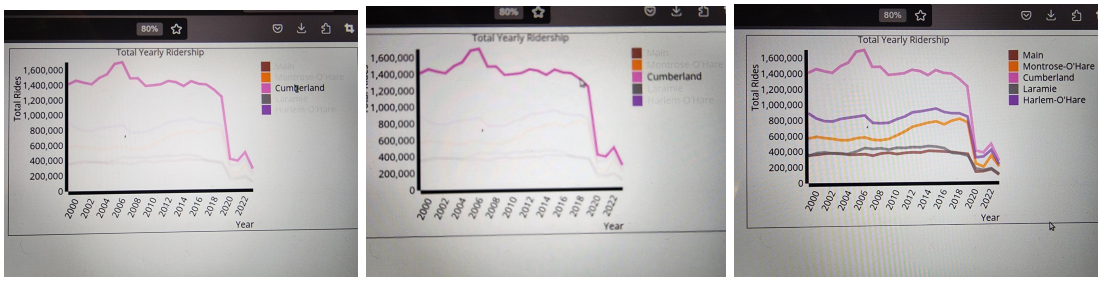
The above screenshot shows a zoom in on what happens when you hover over a station. The station changes size and stroke thickness. The tooltip displays the station's name, its total ridership for the year selected in the slider, and the daily average riders that year. If data is missing, it tells you so.



This screenshot shows what selecting one station looks like. We have the line chart showing up representing that station's data on the top right. The x-axis are the years while the y-axis is the ridership count. You can completely remove the chart and all of the selected stations by clicking on the clear button or deselecting the stations. If data is missing, it shows a blank chart. Also, all selected stations, are filled black until unselected or cleared.



Here is what the chart looks like when multiple stations are selected. They are represented by different colors with a legend on the side. The colors do not correspond to the line so that you can compare stations on the same line without confusing them. You can select up to 5 stations.



If you hover over either the line or the name on the legend, the other chosen stations become more transparent to focus on that line. This is in case the ridership line is very similar across stations so the user can distinguish them more easily.

Development Process

After finding my data, I followed instructions I found online here to draw the map. It took me a few days to fully understand how it worked but once I did, it was not too difficult. I connected the map to my ridership dataset I got from the city of Chicago but that was an issue. The station names were not the same across all stations so I had to go through and manually change all of them which took a whole day. This is also when I realized a few station's data was missing here and there but also almost the entirety of the pink line.

Afterwards, I worked on making the tooltip with the hover. This was pretty easy once the datasets were standardized. Took me about a day as well because I was just getting used to d3. This also included the making of the slider to choose the year being looked at.

I spent multiple days following this working on making the stations' circle sizes correspond to the ridership amount. While making this technically work wasn't too hard, it was tweaking the design and asking people for input on how to make it more legible that took a while. I ended up wasting a lot of time on this as I found no reasonable solution to it as mentioned previously in this report and I ended up scratching the idea entirely.

I moved on to making the station's charts show up. This took a surprising look as I had a very odd bug for a weekend. The issue was that the ridership data was being stored as a string rather than an int. After that the rest was pretty smooth including selecting multiple stations for the same graph. I chose to limit the number of stations selected to 5. The main reason for this was to avoid overplotting. If you chose stations that are very close to each other, odds are they will have very similar ridership numbers so the lines almost overlap. By having only 5, it is more manageable. Additionally, I implemented the hover over the line or legend to focus on a line to also combat this issue.