**Line Chart**

**Marks:** Points, Connections **Channels:** Positions, Color

For the line graph, I attached the year as the x-position and the count as the y-position for each point on the graph. The points are then connected by a line. The combinations of these two marks and position helps show a trend of the searched name. Since the line graph allows up to 5 names to be searched, each name searched get its own unique color to help the viewer distinguish which lines is for which name that was searched.

The y-axis of the line chart is scaled linearly based on the count of the searched name.

The x-axis is also scaled linearly based on the year for the searched name.

**Bar Chart**

**Marks:** Lines **Channels:** Vertical Length, Y-Position, Color

For the bar graph, the mark used are lines. The length or height of these lines represents the count for the searched name in each US state which is represented on the Y-axis or y-position of each bar. This creates a categorical graph that makes it easy for the viewer to see the count of the searched name and where it was most used within the US states. The color for the bars were used based on the color of the line for the name searched on the line chart. This helps the viewer to see that that bar chart corresponds to the selected name to see further details about it.

The y-axis of the line chart is scaled linearly based on the count of the searched name.

The x-axis is scaled band wise with a list of US state initials ordered alphabetically. The initials of each US state covers the whole width of the bottom axis in an equally spaced manner.

My first problem was literally trying to open my dataset in google’s open refine. The csv file was so large that the software could not open it up. Without open refine, I couldn’t clean the dataset. I often had to restart my computer and I feared that my computer was explode. I then tried to open the csv file in excel which seemed to work alright, but I wasn’t sure how to clean the data in excel. I did eventually figure it out. I removed only the Id column in the csv file since it seemed useless to me. The other columns that were left was the name, year, count, and gender. When I tried pushing this csv file to github, github threw it back at me saying that it couldn’t exceed the 1.5 MB file limit. This meant that I had to move more data from the csv file to make the file size small. I opened the csv file back in excel. This time I removed all rows that were within the years of 1910-1989. This saddened me as now I only would have a ‘part’ of the original data file, but something had to be done.

My next step was just creating the line graph. Creating the scale and labels were easy enough.

The next challenge I ran into was restructuring the dataset into subsets; I had to group all the years together and add the occurrences of that name during that year as a total count for that year. This created a lot of for loops within a for loops. Sometimes I had to create three for loops embedded in each other which caused my brain to malfunction trying to keep track of where I was in each iteration and what value(s) I was getting out from it.

With the subsets of data, I was able to create multiple lines on the line graph; each corresponding to a searched name.

Another challenge was connecting the colors between the name tags, the lines and dots on the line chart. I had to create an array of set colors. Then for each name searched, I would assign a color from the array of colors to each name. Sounds easy enough, but it got messy when a name tag would be removed and added again. A long story short, a lot of times, I ran into the same color popping up twice for two different names searched. I got around this using a secondary color array that would keep track of what colors where being used (I think…).

Next was css-ing the webpage a bit and then creating the bar chart which would show more info about the searched name during a certain year. Once again, creating the scales and labels were easy enough.

The next problem I ran into was figuring out how to connect the line and bar graphs together which was by color, name, and year.

The bar graph would look into the dataset for a match of the name and find the corresponding year with it. Each US state corresponding to these two variables would have its count variable correspond to the bar lengths. (I’m sorry if this makes no since, it’s 12am in the morning and my brain is mush).

Once everything was working and I saw the big picture, I’ve noticed that a lot of names searched have been declining; at least between 1990 to 2014. I was surprised as I was expecting some lines to be increasing overall, showing that the name was increasing in popularity. I then realized that the occurrences of names are declining because the US total population is also decreasing each year. I confirmed this with this [link](https://www.u-s-history.com/pages/h2061.html). The bar graph that shows the occurrence of the name in each state was especially high only in CA. I figured this was because CA, out of the other states, has a high population of people. With this all-in mind, I think the popularity of baby names can be determined by the slope of each line between points. The steeper the slope, the faster the popularity of that name is increasing or decreasing. When looking at the corresponding bar graph of the name, you need to keep in mind that the length of each bar does not tell you the population for each state which does affect the actual popularity of the name when comparing its occurrence in other state.

Ever carious about the popularity of a baby name over time? Look no further because the graphs on this website will show you just that! To start, enter a name in the text field and either press the enter button on your broad or press the search button. You will see a line and points pop up on the first graph. This line shows you the popularity of that name from 1990 to 2014. If you click on one of the points on the line graph, the graph below it will show a bar graph. These bars tell you further information about that name during that year. Each bar and its height shows you the occurrence of that name used in each US state. Remember, you can search up five names at a time!

If it this wasn’t assigned during the finals weeks and I had more time, I would’ve enjoyed making this project a bit more.

There’s minor tweaks like showing less ticks on the x-axis on the bar chart and fixing the dashed lines on the line chart to be drawn behind the actual lines and points.

There’s also a few things to add to this project, for example, adding the Voronoi grid to improve the line charts interactivity experience. This way, the user doesn’t have to exactly hover over a point to see the radius change in size and the mouse courser changing to the pointer icon.