CS 3300 Project 2 Report

Antara Palit ap843, David Marchena dpm247, Justin Peterson jwp262

Data:

We utilized two datasets to create our visualization. The first was a dataset of the average life expectancy of each state in America, which we got from www.kff.org. The second was a more complex data set detailing the top causes of deaths per state per year from 1999 to 2015, which came from the Center for Disease Control website.

In the cause of death dataset, there are four main variables we used: year, cause of death, state, and number of deaths. There is additionally an age-adjusted death rate variable, which we didn't use due to lack of context in the dataset and no necessity for it in our visualization. To make the visualization easier to interpret, we averaged causes of death over the time frame of the data for our bar plot.

The life expectancy dataset is somewhat simpler; it contains just state names and life expectancies for each state, which made processing it to display the data easy.

Life expectancy data source link:
https://catalog.data.gov/dataset/age-adjusted-death-rates-for-the-top-10-leading-causes-of-death-united-states-2013/resource/0e603f1d-31bf-4809-8f10-a994b305b379

Cause-of-death data source link:
https://www.kff.org/other/state-indicator/life-expectancy/?currentTimeframe=0&sortModel=%7B"colId"%3A"Location"%2C"sort"%3A"asc"%7D

Mapping from data to visual elements:

The life expectancy dataset was very easy to handle due to its simplicity, but the cause-of-death dataset required some substantial manipulation to find overall percentages of different causes of death by state. To handle this, we used a series of iterative procedures to find sums for each cause of death for each state, and used these sums to calculate probabilities. The code for this data manipulation is found within the activateEvent() method in the visualization file.

We integrated both of these datasets into our final visualization, using information from both to color the map and generate a dynamic bar graph. Since both datasets were about mortality, we decided that having information from both datasets told a more in-depth story through our visual. We prominently showed the data from the life expectancy dataset and then depicted data from the causes of death dataset underneath it.

The first dataset was primarily used to color each state using a red gradient to represent the difference between life expectancies of different states, as well as for estimation of the year of death of people who input their info into the form at the top of the visualization. The second

dataset was used for the bar chart on the right side of the visualization, allowing us to change the bar chart dynamically based on user selections.

Besides these two datasets, we also used an SVG file found online for our map of the United States.

United States SVG link:
https://simplemaps.com/resources/svg-us

The Story:
The visualization shows the different life expectancies and causes of death for each state in the country. It also allows a user to input their personal information through a form, so they can find out what year they will die (based on the their age and average life expectancy of the state they are from). The coloring of the map shows clearly which states have higher life expectancies. Clicking on each state results in a bar graphs showing the break-down for cause of death - for certain states, suicide accounted for a surprising percentage of deaths.

A surprising realization from our visualization is that the states with the lowest life expectancy in the US are grouped around Mississippi and Alabama. By clicking on Mississippi as well as the states bordering it (which also have lower life expectancies) it is clear that this drop in life expectancy is not due to a different cause of death; these states still have heart disease as their most common cause of death just like most of the other states. This leads us to our second realization, which was the cancer and heart disease are overwhelmingly the most common cause of death. Almost every state had both of these causes in their top three..