Samuel Shapiro & Nick Patel

CS 4460

P5

4/21/19

For this project, we chose to create a visualization using the colleges dataset. This data set includes nominal fields such as a college’s name and region, as well as quantitative fields such as a college’s average cost, average family income of students, and median ACT score of students. With this dataset, we decided to create three visualizations that would allow users to explore relations between variables and discover correlations through their own exploratory analysis.

The first of our visualizations is the focus of the user interface and features a scatterplot with an alterable x-axis variable. The y-axis of this scatterplot shows a college’s average cost, while many of the other quantitative variables are options for the user to apply to the x-axis, such as average family income (shown by default), mean earnings 8 years after entry, and median debt after graduation. When the user chooses a new x-axis from the dropdown menu below the x-axis and presses the button to apply changes, the x-axis will rescale and all of the points on the scatterplot will animate and transition to their new appropriate spot. If a data point does not have a proper value for that variable, the point will shrink to a size of zero and animate into the y-axis. When the user hovers over a point on the scatterplot, the border of the point will darken, and the point will fill with a light blue color. When the user clicks on a point, the border will darken, the point will fill with a light green color, and statistics of the selected college will appear to the right of the scatterplot.

The second visualization uses a linking feature only appearing when the user selects a school. The chosen visual is a pie chart showing the user’s chosen school’s diversity. The chart considers all 8 diversity factors, and a new derived variable called other. Other is a representation of the percentage needed to have a 100% as many schools did not add up to 100. This was also done to ensure a well-balanced pie chart as the total percentages taken into consideration should add to be equivalent to 100%. Each slice is then represented by a different color and labeled. There is also a sorting feature to allow each variable to always be represented by its corresponding color.

When creating a data visualization, it is always important to keep in mind of what the objective of the visual is or trying to get across. In this implementation, the chosen data set was a CSV of colleges with different variables and categories. The first step in approaching the problem was to answer the question, “What information do we care about?” In doing so the approach was to first implement a scatter plot to visualize all the schools. The Y-axis for the scatter plot is always a representation of the average cost of attending a school. When comparing schools, one major factor that usually stands out and in some cases is a deciding factor of admission is the average cost. The next taken was organize the data across an x-axis. For this we chose a dropdown box with 8 key factors that can used to re-arrange the colleges to the new axis. This allows the user to sort through the colleges by one of these 8 ways to gain an entirely new scatterplot visual. This was created to allow a dynamic and interactive vis to gain a better understanding of the data as it allows the user to see how different a single college can be placed depending on the X-axis. This allows us to further our understanding of the data itself. After the implementation of the scatterplot, we wanted to attempt to explore a data point more in depth. When a user selects a school by clicking on it on the scatterplot, the node is then highlighted to easily keep track of the selected node. In addition, a 2nd svg is then generated and appears on the screen displaying a pie chart representing a selected school’s diversity. It helps analysis see even more than just schools; it could also allow users to see diversity trends within a certain area due to the dynamic scaling of the scatter plot to begin with.