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 which appears only when the user selects a school. The chosen visual is a pie chart showing the breakdown of the selected school’s diversity. The chart considers all 8 diversity factors, and a new derived variable called other, which is a representation of the percentage needed to total 100%, as many schools did not add quantities adding 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 in place to ensure each variable will always be represented by its corresponding color.

When creating a data visualization, it is always important to keep in mind what the objective of the visual is or the message it is trying to get across. The first step in approaching the problem was to answer the question, “What information do we care about?” In doing so, our approach was to first implement a scatter plot so that the user has a view of all data points. 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 choosing a college is the average cost of attending. The next objective was to organize the data across an x-axis. For this we chose a dropdown box with 8 key factors that can be used to create entirely new scatterplot visuals. This was created to allow a dynamic and interactive visual experience which, through exploratory analysis, allows the user to see how colleges relate to each other in multiple ways. After the implementation of the scatterplot, we wanted to attempt to explore a single data point more in depth. When a user selects a school by clicking on it on the scatterplot, a second SVG is generated and appears on the screen displaying a pie chart showing the selected school’s diversity breakdown. Adding this feature allows any user to have a greater interactivity with and insight into each node. As users interact with the data, not only will they be able see how diverse a single school is, but they will be able to draw new conclusions about how diversity factors into the correlations visible on the scatterplot.