Alex Levine

Endert
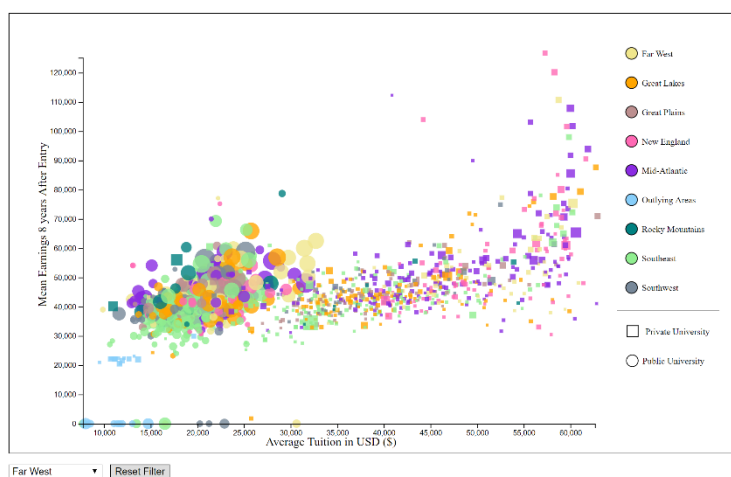
CS 4460

18 April 2019

<div align="center">P5 Description: Colleges Dataset</div>

My visualization supports retrieving value by hovering over a data point for more detail; filtering by selecting a region to show exclusively on the graph; finding extremum by displaying all points so that the user can see highest and lowest points; finding anomalies also by displaying all points so that the user can see outliers; clustering by showing points as separated by color and shape to allow the user to see groups of similarly located, colored, or shaped points; and correlating by showing all points so the user can identify trends in the graph's overall shape.

The visualization begins by showing the overview of all universities, laid out horizontally along the x-axis based on their average tuition, and vertically along the y-axis based on their mean earnings eight year after graduation. The data points are all color-coded according to their location, as explained in the legend along the right side. Their shapes are encoded based on whether they are publicly or privately run, as shown in the same legend. I felt these were the most important, basic information to shown about school as tuition tends to emphasize their prestige. People also would want to know which colleges would give them the best return on investment of earnings after paying for four years of college. In addition, more privately-run schools tended to more

expensive than public colleges. This answers questions about correlation between any of the dimensions, such as more public schools being cheaper and more private schools being expensive.

The visualization allows the user to filter the data points by a certain region, which can be especially helpful if the student has a particular area in mind for where they want to live for the next four+ years. The data points, after being filtered, retain their original color so that the user is still easily able to follow points from before the filter is set. Once the user is done with this setting, they can reset the filter and all the original data points reappear as before. This aids the user to answer questions regarding a school's location and its correlation with other dimensions, as the user can more easily see the location. It can also answer which locations have the greatest number of schools at different rates of price and success.

The visualization also allows for zooming and dragging on the graph itself, so the user can more easily select data points to compare or view more detail on. The user can zoom by scrolling on the graph box, and then they can click and drag to whichever area of the graph they wish to see. This can help the user more specifically and easily answer questions about specific schools by narrowing down the number of distractors on the screen.
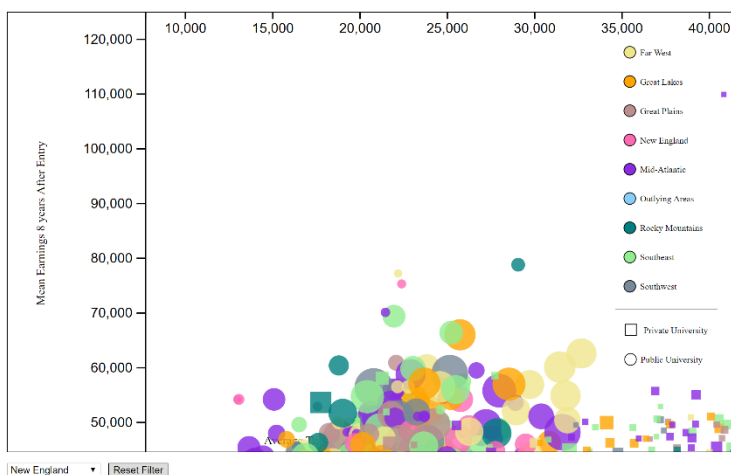
At any point during this process, whether the user has zoomed or not, dragged or not, filtered or not, they can mouse over the data points to see more important and specific details about each

university. The main point of the mouseover tool-tip is to see which university the user is selecting by displaying its full name. The tooltip continues by specifying the exact tuition and mean earnings after

eight years to which each point is mapped, in addition to the racial makeup of the university and a calculated return on investment based on the tuition and mean salary. This also helps the user more specifically and easily answer questions about specific

schools to figure out where they can possibly attend, where they might best fit in or feel comfortable, and which has the highest reward for attending.
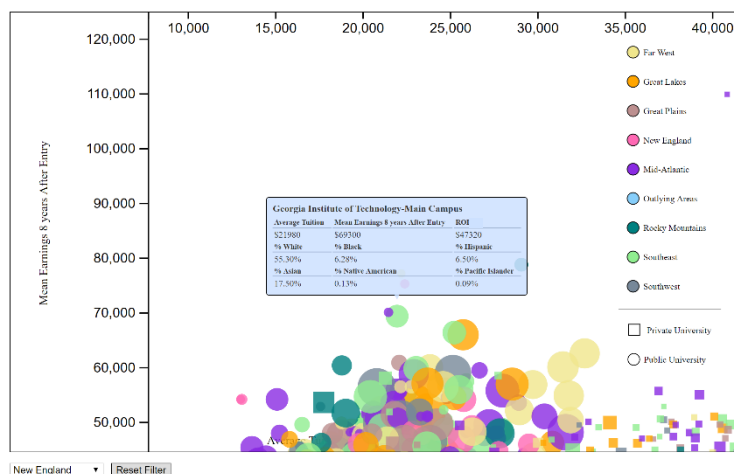
Something to note about this visualization that is more affected by the data than the code is that some universities most likely did not provide a mean earnings eight years after entry, so they have a zero entered for that value. This makes it appear that these universities all put out graduates who are unemployed, but this is highly unlikely. Therefore, the small cluster of

universities at the bottom of the graph at $0 mean earnings eight years after entry are most likely a result of survey error.