CS 3300 Project 1 Description

Leah Greenstein, Michelle Nelson, Emily Gonzales

For this project, we decided to research the relation between spending on education and test results within each U.S. state. Specifically, we wanted to visualize the correlation, if any, between spending per student on education and the average SAT scores of students for every state in the U.S. We integrated three separate data sets together and mapped five variables for every state: 1) name, 2) region of the U.S., 3) education spending per student, 4) average SAT score, and 5) total percentage of eligible students who took the SATs. In addition to plotting this data for each state, we also included a point that represents the U.S. national average SAT score and education spending per student.

All relevant SAT data came from the following online PDF: http://www.ipsr.ku.edu/ksdata/ksah/education/6ed16.pdf. This data is sourced from the National Center for Education Statistics and lists SAT scores for high school seniors who are continuing their education into college for 2013-2014. Since the data was presented as three separate subgroups (Critical Reading, Math, Writing), we manually added the three average sub scores together to obtain our total average SAT score variable for each state. We decided to graph the cumulative 2400-scale SAT score rather than each sub-score group because one score is easier to understand than three separate scores. We also used this PDF to obtain data for the variable indicating the percentage of high school Seniors who took the SATS in 2013-2014 for each state. No manipulation of this data was required. For the data concerning the funding of education per student for each state, we obtained data from an article reported in The Washington Post, accessed through the following link: https://www.washingtonpost.com/news/local/wp/2015/06/02/the-states-that-spend-the-most-and-the-least-on-education-in-one-map/?utm_term=.d9406c322508. The author of the article was able to gather the data directly from the U.S. Census Bureau. This article also indicated the U.S. average funding in education per student. Finally, in order to group the states by region, we used a categorization given from aacc.nche.edu. We ended up discovering a large amount of data concerning education and spending for states. In choosing which data to use, we wanted data that was simple to explain to the average viewer, and data that could be easily visualized in a graphical format without adding too much clutter. Ultimately, this is how we decided to relate education spending and SAT scores on our graph, as these are the two variables we were most interested in studying.

The main component of our project is the graph with amount spent per student on the x-axis and average sat score on the y-axis. In order to make this evident, we placed the graph directly in the middle of the page so that the viewer's eyes reach the graph first. The legend is placed slightly within the graph and we were careful not to allow it to interfere with the data points. The position of the legend still allows the main graph containing the data points to be front-and-center without the legend making the page appear lopsided. For the x-axis, we mapped the amount spent per student using a d3 linear scale. We did not start the x-axis at zero because this would make the data harder to interpret. For the y-axis, we also used a linear scale to map the average SAT score for each state. We manually inputted the point on the graph for the U.S. national average SAT score and spending per student.

Color was a very critical aspect of our visualization design. When we originally created the graph of education spending vs. SAT scores, we realized that it would be interesting to see if there is any pattern within certain regions of the U.S. Thus, we introduced another variable: state region within the U.S. We assigned unique, different colors to each of the five main regions of the U.S. and displayed the respective colors in the legend. Using color to visualize the region of each state allows the user to immediately group certain data points together into meaningful subsections, adding to the story the data tells. For example, upon looking at the graph, the user notices that on average, the Northeastern states (New York, New Jersey, Vermont, etc.) tend to spend more on education per student.

We used circles to graphically represent the data points because circles are easier to interpret than other shapes such as squares or triangles. The circles gave a softer rather than more harsh effect, allowing the viewer to concentrate on the data instead of the shape itself. We used transformation on the circles to change the relative size to demonstrate the percent of high school Seniors taking the SATs. We did this by mapping the percentage (0-100) to a radius value of the circle using a power scale. We chose this specific type of scale to account for the fact that the radius is squared to produce the circle on the graph. For instance, a circle with a small radius indicates a small percentage of eligible students while a circle with a larger radius indicates a high percentage. It is important to include these percentages in our data to show how SAT scores cannot always be a reliable method to measure student output or success. Without this scaling of the circles, one would assume that states like Illinois and Minnesota are getting "more bang for their buck" than New York or New Jersey because they have a higher SAT average compared to the amount they invest per student. However, with the scaling of the circles, the user can infer that the states like Illinois score higher because a smaller proportion of students are taking the SATs.

Ultimately, all of the data selection and design choices were made with clarity in mind. We wanted the viewer to easily digest all of the data in order to reach the conclusions that the data supports. In our case, the visualization of the data demonstrates many different interesting conclusions. At first glance it is easy to see how the states are grouped by region: the North Eastern states tend to spend more on education and have a much larger participation in the SATs, while all the other regions tend to spend less and are more spread out and less uniform. In addition, it was interesting to see how states that spend the most on education do not necessarily have the highest SAT scores. Without the use of transformations on the size of the data points indicating the percent of eligible students taking the SATs, this could be very misleading. With this variable added, we can see that while states that spend more on education may not have the highest average scores, they tend to have much more students taking the SATs, explaining a higher variability in the scores. We found a negative correlation between a state's participation rate and how high the average SAT score was—the higher the participation, the lower the score. Perhaps this data suggests even larger, more important insights that are not evident by simply looking at the raw numbers. For example, it could be that families on the east coast culturally place a higher importance on higher education, leading to the higher participation rates, where it is less expected across states in the Midwest. Insights like this would be extremely difficult to conclude without the proper visualizations. Before graphing the data, we expected a relatively linear positive relation between amount spent on education and the average SAT scores. This was surprisingly not the case, as we came to learn that relationships between data is not always so simple. In fact, by choosing to add variables that we found relevant to the overall story, the data became much more clear and easy to understand.