**CS 3300 Project 1 Description**

Leah Greenstein, Michelle Nelson, Emily Gonzales

For our first project, we decided to research the relation between spending on education and test results. Specifically, we wanted to see the relationship, 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 into five variables for every state: Name, Region, Spending on education, Average SAT score, and Total percentage of students who took the SATs.

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 average SAT score variable. We also used this PDF to obtain data for the variable indicating the percentage of students 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. 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, since these two variables are reasonably related.

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. We didn’t want to place the legend above the graph since the viewer’s eyes would have gone there first. Below the graph was also an undesirable location since we wanted to utilize the maximum amount of screen space for the graph itself while still having the legend appear evident. Color was a very critical aspect of our visualization design. We first decided against a monochromatic color scheme because it made the graph appear bland and uninteresting. Rather than adding color simply to enhance the graph visually, we used color to introduce a new variable: state region within the U.S. This way not only do we have a fairly even representation of five different colors, but it is now easy for the view to immediately group certain data points together into meaningful subsections, adding to the story the data tells. We used circles to graphically represent the data points because we found it most visually pleasing when compared to 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. Finally, we used transformation on the circles to change the relative size to demonstrate the percent of eligible students taking the SATs. 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.

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, 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. 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.