Understanding Data: Variability & Context

Michael Surdek

Southern New Hampshire University

Graphical user interface, application

Description automatically generated

Graphical user interface

Description automatically generated

For the Module One assignment, I chose to analyze the Florida School and District Grades dataset from Google Public Data. I created visualizations for parents of Eighth graders in the Sarasota school district who are looking for information that will help them determine where to send their children for high school. The two distinct audiences were all parents in general and parents of English as a Second Language students. I made one visualization showing Graduation Rate over time (2015-2018) at 8 high schools in the Sarasota district, and another showing English Language Arts Learning Gains of the Lowest 25% over time (2016-2018) at the same schools.

For this assignment, I chose to modify the visualizations by turning them into bar charts, sorted from highest to lowest, and only used the data from the most recent year, 2018. I also modified the x-axis to a minimal range that included all of the points, instead of ranging from 0% to 100%. These modifications impacted the elements of variability and context that are displayed. When I showed how the rates changed over time in the initial visualizations, it was possible to get a better idea of where each school stood relative to the others. When I only included data from the year 2018, it is still possible to see how the schools compare in each category, but there is increased uncertainty for the audience. This is because the data for one year is more likely to impacted by factors such as individual graduating class proficiency, or particular events that occurred which might have affected the data for that year but would not be a factor if a parent were to enroll their child in that school going forward.

The modifications I made also impacted the context of the visualizations when it comes to the “what” and the “when”. The “what” of the visualization is impacted in a similar way to the variability. These rates over time are more likely to be related to the curriculum, faculty, and systems in place at the schools, but in a certain year the rate could be related to the student classes or isolated events. This means that the bar chart visualization might not be showing the same “what” as the line chart was showing. The difference in the “when” is self-explanatory because the parents are concerned with how the rates will affect their children over the course of the next four years. Having a “when” of one year is helpful information, but a “when” of multiple years can show consistency or trending directions for the schools, which are more likely to play a role over their children’s high school period.