Visualization Summary
Visualization #1 looks at the correlation between GPA of females and the number of females enrolled in institutions. We expected to see an increase in GPA as more females enrolled, however the data rejected this. It seems there is a correlation between lower enrollment and higher GPAs.

Visualization #2 looks at the proportion of females enrolled in STEM programs to males enrolled in STEM programs. The visualization is a multigraph where all seven declared races are looked at individually, along with US citizens. All eight graphs are normalized on the y-axis from 0 to 1 to avoid spatial skew, and a context bar is added at y = 0.2 so that all graphs can be compared to one another quickly.

Visualization #3 compares the actual distribution of ACT scores to a normal distribution of ACT scores.

Visualization #4 shows "How enrollment by class(Freshman, Sophomore,...) changes over time for schools where students declare a major upon enrollment

Visualization Interaction
Visualization #1, #2, and #4 all use hover over interaction i.e., when the mouse is hovered over a data point a summary statistic will appear that provides more information, a summary, or specific data at that point.

Visualization #3 has a drop down interaction that allows the user to change the view year.

Design Process
Our design process started with cleaning the data set. This process was done in excel. Then we imported the data into Pandas and ran some quick tests with pd.describe(). This gave us an overall layout of which attributes were usable and which were too sparse to use. Once we had a basic understanding of the NCWIT data, we developed four research questions we wanted to address with the visualizations. The process of generated RQs was iterated a few times until we had questions we felt were complex but could still be answered with the data.

For visualization #3 John explored a similar visualization in another class and it looked like an interesting problem. Creating it in Bokeh was where most of the design challenges came in. The visualization could have been more effective if the bar chart had been scaled differently so it was easier to directly compare it to the normal distribution of ACT that year.

Data Representation
We chose to visualize the data using line graphs, bar charts, and scatter plots. The bar charts are effective when visualizing discrete data points such as GPA and ACT. Scatter plots are effective for pattern recognition in a single visualization. Line graphs are effective for multi-graphs where one of the axes is time because it is easy to quickly compare the shape of the lines, and see the differences.

Work Done
Tom, Taylor, John and Ben each created one visualization and created research questions to answer. John also cleaned the data set.

Running Project 2
There are four visualizations in the Team14 GitHub repo. They are all jupyter notebooks and can be run by ‘running all cells’. Visualization #1, #2, and #4 all produce HTML files, which is where the visualization will appear, and Visualization #3 runs the visualization in the notebook.

Visualization #1, #2, and #4 have hover over interactivity, and Visualization #3 has drop down menu interactivity.