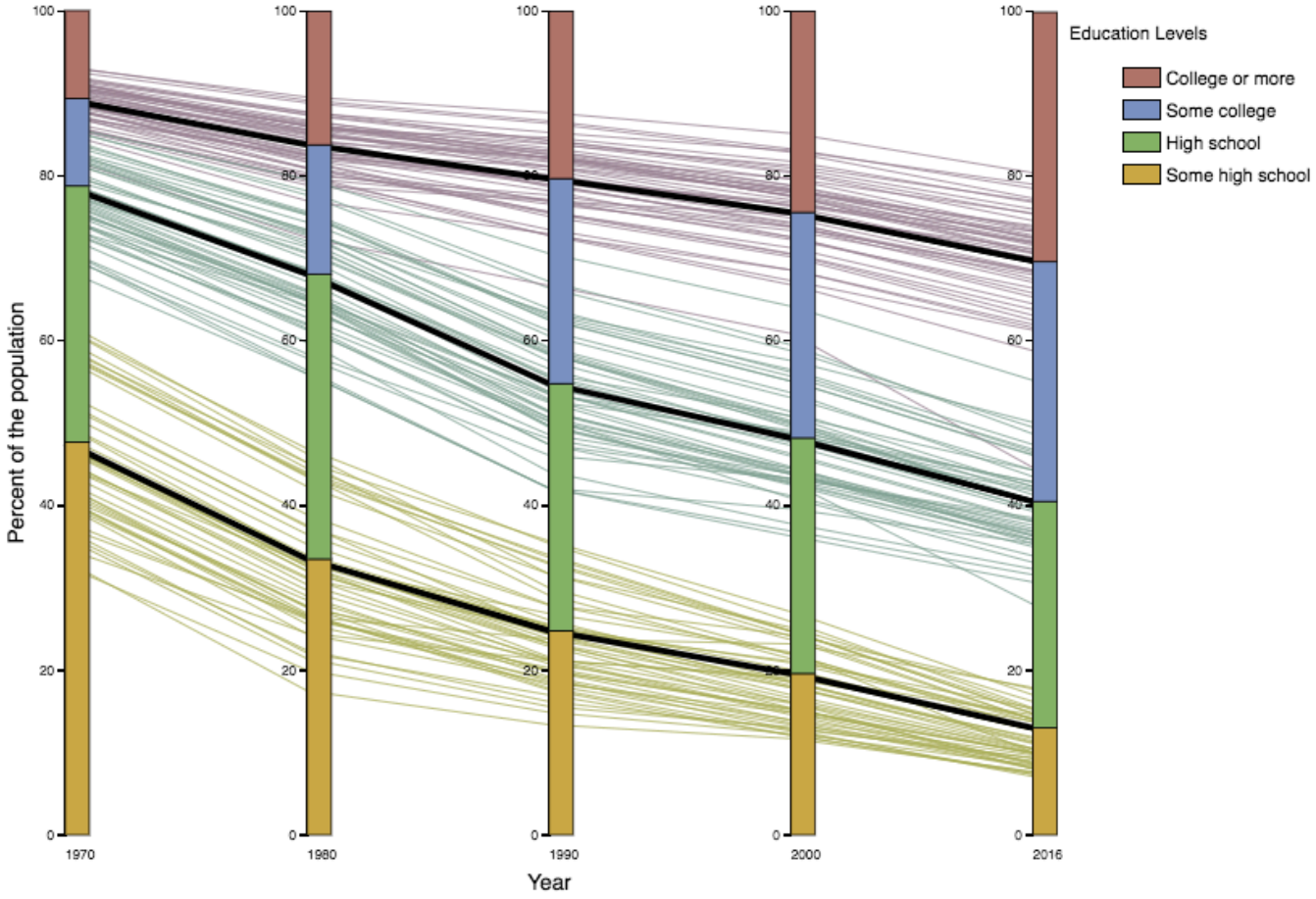
How the percentage of people in each state with different levels of education has changed from 1970 to 2016



This visualization brings together a stacked bar chart and a line graph to make a variation on a parallel coordinates visualization. The main bars show the proportion of people with each level of education in the United States as notes in the legend. Each of the lines represents where those divisions between education levels would be for each state. In this way, each state gets three lines to show how these proportions have changed for them over time. When one line is selected, all the corresponding lines for that state are highlighted and a label is shown.

The dataset used can be found at this link and then by clicking the "Educational attainment..." option. It is census data from 1970 to 2012 - 2016 with information for each county about its urban influence, number of people with a some high school, a high school diploma, some college, or a Bachelor's degree or higher, and the percentage of people in each of those categories. This information is also provided for the entire country and each state.

Eva Grench
Assignment 4
10/31/18

Dataset:

The dataset used can be found at https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/ and then by clicking the “Educational attainment…” option. It is census data from 1970 to 2012 - 2016 with information for each county about its urban influence, number of people with a some high school, a high school diploma, some college, or a Bachelor's degree or higher, and the percentage of people in each of those categories. This information is also provided for the entire country and each state. I used a few scripts to clean the data and made some changes manually, so the final version of the data in the .zip file are EducationFinal.csv and US.csv.

Caption:

This visualization brings together a stacked bar chart and a line graph to make a variation on a parallel coordinates visualization. The main bars show the proportion of people with each level of education in the United States as noted in the legend. Each of the lines represents where those divisions between education levels would be for each state. In this way, each state gets three lines to show how these proportions have changed for them over time. When one line is selected, all the corresponding lines for that state are highlighted and a label is shown.

Description:

When thinking about this dataset, the first question that came to mind was, How has the number of people attaining various levels of education shifted over time? This visualization does that for the U.S. in general with the stacked bars, but to answer a more complicated question, How has the number of people attaining various levels of education shifted over time in a given state as compared to others?, the lines or parallel coordinates aspect of the visualization is introduced.

As previously mentioned, each bar represents the percentage of the U.S. population attaining each level of education for the given year. Each line shows where those divisions on the stacked bar would be for each state. In other words, each line show the additional percentage of the population gaining a certain education. The last set of lines is assumed, but they would be straight across the top for each state (because 100% of the population falls into one of the four categories).

I used length when making the bars, encoding the percentage of the population, slope for the lines, showing how the percentage of that particular education level changed in relation to the others over time, and color to show categorical data about the education level.
Length is something that we are pretty good at interpreting, so I thought it was important to pick something quick to analyze for the main pieces of data: the bars. The slope of the line is pretty arbitrary because there's missing information in between, so getting the precise angle of the slope is not necessary, which is harder for viewer. Getting the general trend, on the other hand, is pretty straightforward and easy for the viewer to do which is the main goal for that aspect. The colors were meant to be distinct here because they represent categories. At the same time, I was careful to color the lines such that they blended the colors of the bars it divided. Using very distinct colors makes it very easy for the viewer to distinguish between the categories quickly, which is important to answer the main question.

During the critique, some feedback I got was that the lines would likely be difficult to distinguish. To remedy this I added in the additional colors for the lines. That way it is clear what line belongs where so it is not necessary to always hover or click to get a general idea of the data. On that same topic, I came up with the idea of having a drop down menu and allowing the user to select a region of the U.S. (or all of it) and only display the lines for states in that region. This would allow for easier interpretation and comparisons. Unfortunately I did not have time for this extension, but I think it would be interesting and worthwhile.

The interactions I included are similar to what we did for assignment 3. If a line is hovered over, all of the corresponding lines (including the 2 entirely separate lines) for that state change color. If a line is clicked, it becomes thicker. Again, so do all the related lines. Additionally, when a line is hovered over, a tooltip showing the name of the state is displayed. Because the dataset is rather big, allowing the user to see even just a little more information about each piece seemed important if they were going to gain much insight from the visualization.