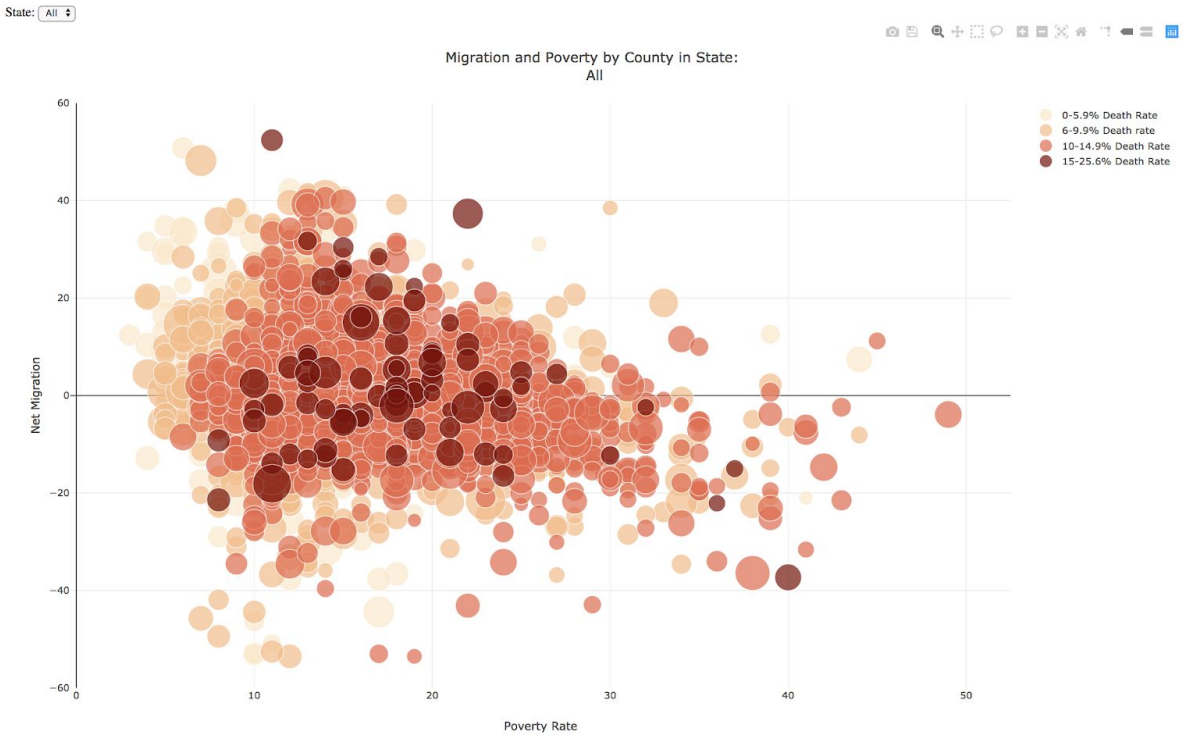


This plot displays demographic information for each county in the United States from 2016. It maps poverty against net migration (where negative migration indicates people leaving and positive migration indicates people entering), death rate indicated by the color scale shown, and percentage of people in the county with a bachelor's degree or higher, which is indicated by the area of the circles (smaller area means less with a higher degree, larger area means more with a higher degree). This story intends to highlight the differences in migration, education level, and death rate dependent on poverty level between US counties.

The data for this visualization is taken from the census bureau website. (https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/). It uses a combination of the poverty, education, and population datasets from 2016.

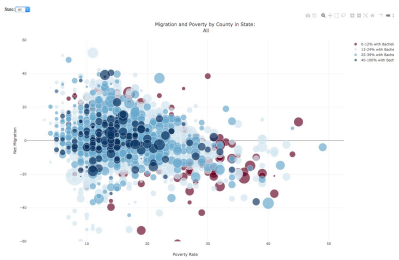
(Note: this is longer than 1 page but it includes pictures that help show the evolution of my visualization, as well as future work and sources)

The task that I was trying to accomplish was to answer the question: Is higher poverty associated with higher death rate, higher negative migration, and lower percentage of people with a bachelor's degree or higher?

In my initial draft, I was trying to compare death rate with poverty by only comparing the highest and lowest poverty counties against each other. However, after receiving feedback, I learned that it would be valuable to include information about every county, but this would be challenging to accomplish on a map since I would be using circles with varying sizes to represent death rate. Furthermore, I received feedback to dig deeper into the data, so I decided to include migration and education level, which ended up providing interesting insight into the data. In order to include all of this information, I used a bubble chart, which gave me an x and y axes in addition to color and size of the circles. Because of the geographical information lost with a bubble chart, I included a hover feature to obtain county information, as well as the dropdown allowing users to view data for each state.

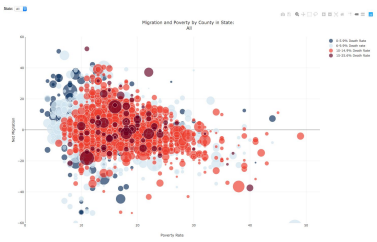
I also received feedback on how to address overlap between the circles which looked like a blob without any identifiable patterns (see below). I tried using a random sample of the data, but ended up losing a lot of the extreme data (very low and very high poverty) which I felt was important to the story I was trying to tell. Furthermore, I also wanted to display state-by-state data, so a random sample meant that we would lose a lot of data points that were essential to observing trends on the state level. I decided to keep all of the data and find another way to address the issue. Ultimately, I expanded the width and length and reduced the range on the y-axis to omit outliers (though it would be best to exclude outliers within the data itself). The "by state" dropdown allows users to hone in on a particular subset of data, which was a concept that we learned in class on how to handle a lot of data (which was inspired both by Munzner's idea of filtering and Heer and Schneiderman's ideas of view specification). Although there is still significant overdraw in the "All" view, it still manages to paint a broad overview and encourages exploration with the dropdown. Furthermore, plot.ly has a built-in feature that allows users to select which traces (death rates) to view, so this isn't as much of an issue on an interactive plot (though it could be improved on a static image by using a random sample on the "All" view).

Other feedback that I received was to not use a categorical color scale to represent sequential data. When I initially made this graph, I was encoding education with color and death rate with circle size. I interviewed people on what color they thought of when I said the word "education," from which I decided to use blue. However, the low versus high education was not popping out, making it difficult to make comparisons between high and low educated counties.. I addressed this by changing the low education trace to a dark red so that it would maintain the same luminance but would pop out because of its hue (see below). This was particularly helpful in making comparisons between states, with some states having a lot of red circles (indicating low education) and others having none.



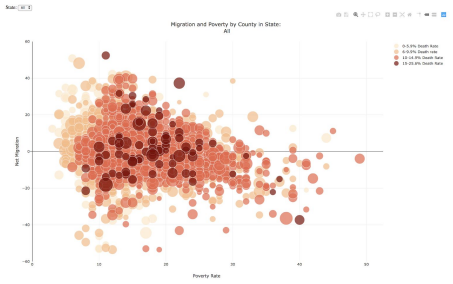
After conducting three (informal) usability tests, I realized that users were not observing death rates, and the stark difference in low education rate versus high education rate was being pronounced the most with the contrast between red and blue. The main task that I wanted to accomplish was to visualize poverty and death rate, and since it is difficult to compare area, this was a challenging task for users. Instead, I used color to portray this difference and used circle area to show education level (I would have liked to but was unable to include a scale and legend for circle size).

Initially, I used a sequential color map using the color red, which became darker with increased death rate, but it did not have the stark contrast between low and high death rate that I wanted to stand out. I tried a diverging color map



to accomplish this, since it is easier to compare hue than chroma (ideally I would use the same hue and change the luminance, as described by Stone as being the best way to show contrast, but this problematic when there are overlapping circles). Whether to use a sequential or diverging color ordering was something I wrestled with for a long time. After trying about 20 color mappings, I ultimately decided that a sequential ordering makes the most logical sense given the principles of color we have learned, especially given the type of visualization this is. This is meant to be an exploratory piece and allow the audience to arrive at their own interpretations rather than telling them what I want them to see by using two different hues, when really the data is representing the same thing. Furthermore, diverging color maps on sequential data implies that the data diverges on the exact point 10, which was a mostly arbitrary decision and not related to the data itself. Therefore, I ultimately came to the decision to use a sequential color map.

Current visualization:



Future work:

- Add legend to show that circle area is related to education and what each size represents.
- There are currently circles included that do not represent counties but instead State averages, which is an issue given by the structure of the data. This needs remedying, perhaps by making those points different than the others.
- Restructure where the dropdown is positioned so that it is more apparent and accessible.
- Get rid of the “US” and blank options on the dropdown. Instead, highlight the US average data point on the “All” view.
- Fix some more of the overdraw. Perhaps use a random sample in the overall view and the full sample in the state-by-state view.

Sources:

https://plot.ly/python/bubble-charts/
https://plot.ly/javascript/dropdowns/
https://plot.ly/javascript/plotlyjs-events/
https://www.w3schools.com/jsref/met_document_addeventlistener.asp
https://developer.mozilla.org/en-US/docs/Web/API/Document/querySelector
https://codepen.io/plotly/pen/QbZmZY
https://plot.ly/javascript/bubble-charts/
https://plot.ly/javascript/filter/