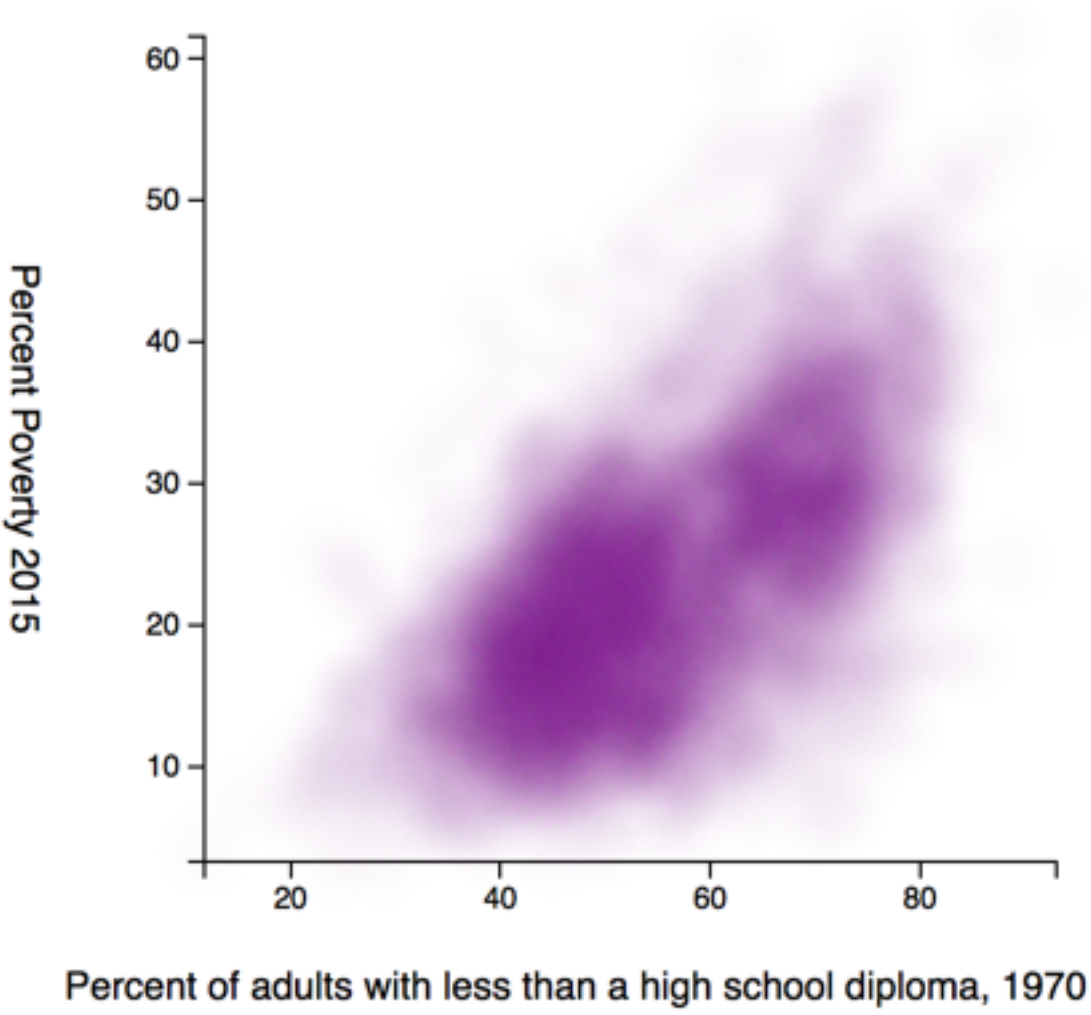
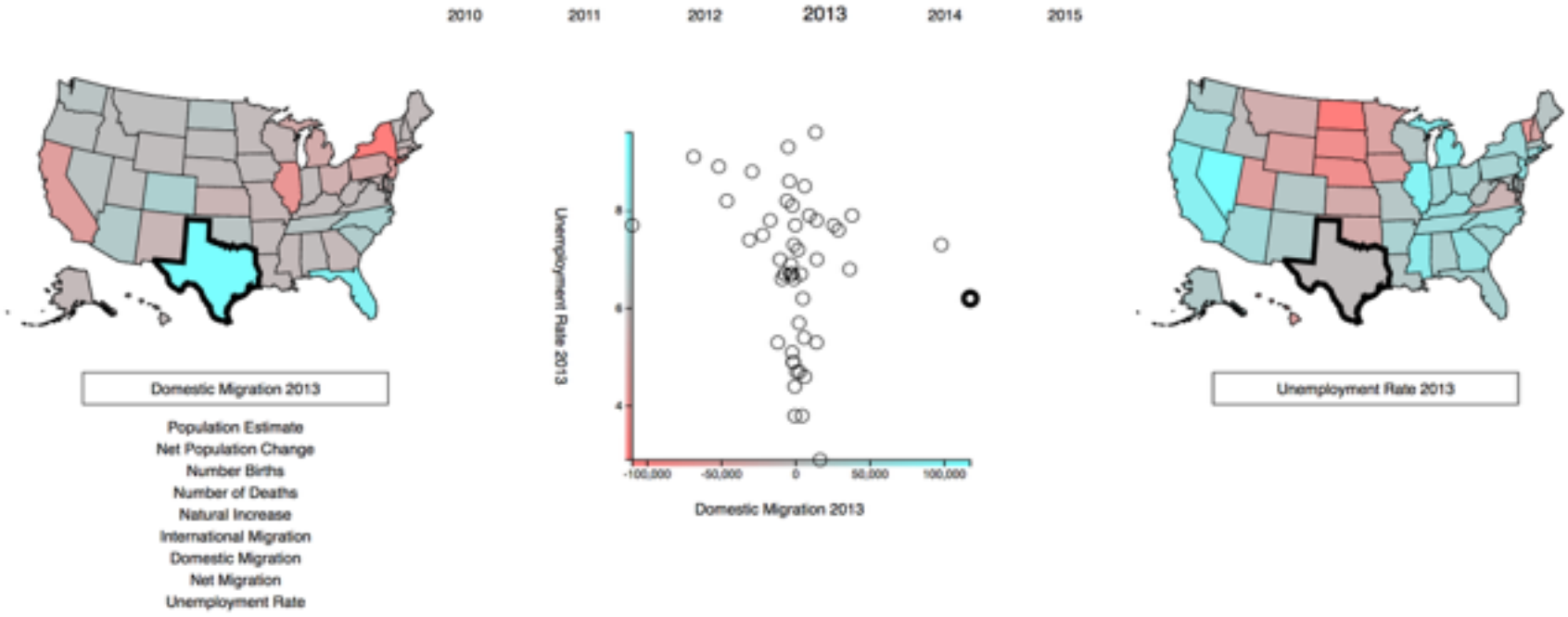
We chose to us the US census data, specifically the education, unemployment and poverty data. This dataset includes a variety of statistics including the percent of adults with a variety of levels of education, and population data with births, death, migration, and total population.



A comparison of United States county education levels in 1970 and poverty levels in 2015



Comparison of US population data (migration, births, deaths, etc.) and unemployment from 2010 to 2015

Visualization 1: County Data by Historic Education and Modern Poverty

What is the relationship between historical education levels and modern poverty levels?

This is a fairly standard scatterplot, however we add color, a gaussian-blur, and opacity to the points to give a sense of color density for areas of the graph where many datapoint overlap. On the x-axis we encode the percent of adults in a given county who have less than a high-school diploma. On the y-axis we have the percentage of the population in 2015 who are in poverty. Because of the density effect, we are encoding the correlation between those two variables in saturation, where low saturation indicates few counties with both characteristics.

Distance is a fairly intuitive and accurate encoding that we thought would allow people to easily pick a point and find the corresponding values in a quick manner. We chose saturation because of the high pop-out value of color, which supports scanning for relationships and ease of finding zone boundaries. Upon finding these pop-out areas or boundaries, the distance provides a way to easily bring this to accurate values.

Our rough-draft for this visualization had hover-over functionality that brought up a line graph of poverty over time for a particular county. However, we received two pieces of feedback about this choice. First, people pointed out that since time is a fixed value on the x-axis, it would be confusing to see a “detail” view that has a timeline. It would just be a re-presentation of the data. Additionally, we realized quickly that due to the sheer number of counties, it would be nearly impossible to find any given county and looking at one random county over time was essentially a meaningless task. Given this feedback, we decided to scrap the line-graph but add in the gaussian-blur, decreased opacity, and color.

Visualization 2: A Comparison of US Census Data and Unemployment

What is the relationship between unemployment and population statistics, like migration or number of births? Are people more likely to move to places with lower unemployment?

This visualization is highly interactive and therefore allows for side-by-side comparison of a high number of variables. Specifically, our data includes Population Estimate, Net Population Change, Number of Births, Number of Deaths, Natural Increase, International Migration, Domestic Migration, Net Migration, and Unemployment Rate. Whichever two dimensions are chosen, we use distance to compare the two in a scatterplot. We also use hue (blue to red) to show comparison on the two maps within that particular metric.

The graph also supports user interaction. When the user hovers over a scatterplot point or a state, if draws a thick border around the associated states and point. In the above picture, the user is highlighting Texas. The user can click on either of the boxes to reveal all of the other statistics they can compare. Clicking those statistics will change the scatterplot and map for that data. Finally, the user can also click on the year on the top of the screen to show the data for any of the given years.

We wanted to users to be able to see if there was correlation between any two variables, but also see what that statistic looked like in the country at large. Specifically, we thought that the color on the US states, would provide a high pop-out effect. For example, in the above visualization, the left map pops out Texas, Florida, NYC, and California strongly. After a state is identified, it is easy for the user to hover over it and see the pop of the border on the scatterplot. This pop-out effect invites comparison of the two statistics, and because of the double encoding of each metric, also allows the user to get an accurate number for a given statistic. Importantly,

it also allows users to quickly see their state in comparison to others in a way that wouldn't be afforded by a simple scatterplot.

This groundwork allows an easy answer to our initial question. The user can simply select unemployment and net migration (the default settings), and quickly scan the relationship on the scatterplot. They can also look for patterns in one of the maps (dark red areas, grey areas) and check if that is also the case on the other map. Then, hovering over these points will allow for a more numeric analysis of those given states.

We had initially planned on doing a line graph with the unemployment data over time, that would also have circles with migration encoded in color along the lines. This would be complimented by a map; when folks hovered over a particular state, the line graph would switch to showing the data for every county in that state. We again realized that this was far too much data to show in one graph, as the overdraw would obscure most relationships. We also realized that color across the small circles on the line graph doesn't support efficient pattern scanning when paired with lines changing y-axis values. Finally, we realized that the largest trend in the data was a massive decrease in unemployment over the time period, which meant that most comparisons over time would simply highlight this trend.

Because of these many factors, we decided to switch to the current set-up of comparing two variables with two maps and a scatterplot. By taking the time variable out of the scatterplot, we were able to move our second variable onto the x-axis, which supported a much better comparison. It also had the effect of taking out the comparison on unemployment over time, which wasn't the point of this visualization. By moving the color to the two maps, we used the strengths of the color encoding (pop-out) while minimizing its weaknesses (comparison when not directly adjacent and lack of precision).