SI 649 Communicative Visualization (Interactive) Blog Entry

Michelle Cheng (michengz@umich.edu)

Learning Objectives

1. The viewer will be able to explain how the home value/household income distribution can be classified and recall Red Lake County's home value/household income relative to the US average.
2. The viewers will be able to explain where Red Lake County falls within both the unemployment and poverty rate distribution.
3. The viewers will be able to recall that Red Lake County's unemployment and poverty rate are lower than many of the counties with higher natural amenities ranking.
4. The viewers will be able to recall the importance of agriculture to Red Lake County's economy despite the county being ranked last in natural amenities.
5. The viewers will be able to recall that Red Lake County's farm earnings percentage and GDP growth over the years are higher than many of the counties with higher natural amenities ranking.
6. The viewer will be able to recall that the average commute time in Red Lake County is shorter than other states/country mentioned in the original article.

Project Goal

For the interactive version, I held on to the same goals as the static version I created previously -- to create visualizations for points mentioned in the original article proving that Red Lake County isn't really "America's worst place to live." It is expected that, after looking at the visualizations covering these two aspects, viewers will be able to get a better sense of the "good" sides of Red Lake County. This time, I was able to cover more information in each graph with interactivity, therefore, I focused more on comparing Red Lake County with other counties in order to show that Red Lake County isn't always the worst when looking at other measures, hence new learning objectives were added in the original list of objectives.

Again, I categorized the content into two domain questions that I want to answer through my visualizations:

1. Is living in Red Lake County affordable?
2. What is it like to work in Red Lake County?

Inspirations

To get familiar with all sorts of possible visualization interactions, I went on the Altair Gallery website to look for some inspirations. The screenshot below was the main inspiration for two of my interactive visualizations. By selecting one of the data points from the scatter plot, it filters the time

series, with only the selected item remaining on the graph. This solves my initial concerns of having too many lines for line graphs given that there are a large number of counties to compare.
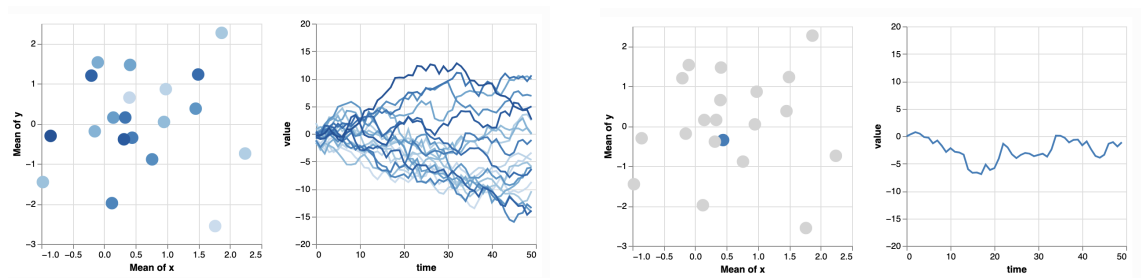
The other screenshot shown below also inspired me to add interactivity for my scatter/histogram plot for household income and home value. By selecting one of the data points, the histogram filters down to showing only the amount for the selected item. With the shift of bars when switching across different data, it would allow me to show the drastic distribution differences between states.
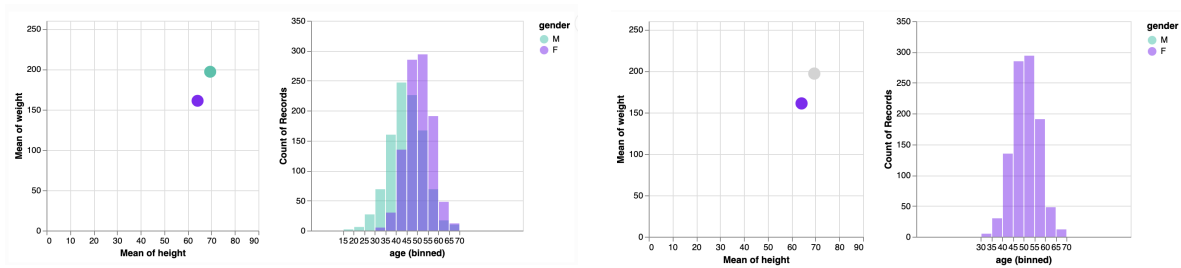
## Design Process & Design Rationale

Visualization 1:

For the first visualization, I thought of bringing the natural amenities ranking into the unemployment and poverty rate graph in order for users to directly compare Red Lake County with top-ranked natural amenities counties. This would make it easier for users to come up with more specific conclusions instead of having them interpret the distribution of Red Lake County among all counties across the U.S.


(Static Version)

At first, I created a geomap of the entire U.S. divided by counties and encoded each county's natural amenities score with color gradient. However, the counties turned out to be too small, making it challenging for users to locate Red Lake County. Moreover, this would use up way too much data ink, which would also mislead users that the natural amenities map is more important than the two unemployment and poverty rate boxplots.
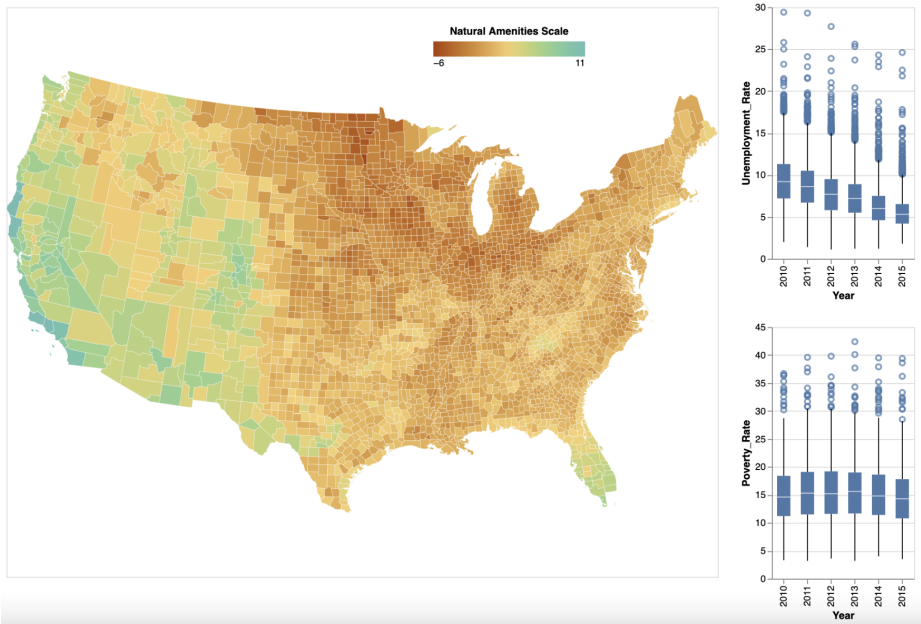
To solve this problem, I decided to only let users select the top 10 and last 10 counties from the natural amenities ranking, given that this still serves the purpose of comparing higher-ranking counties with Red Lake County while maintaining the visualization clear and easy to read/ interact.
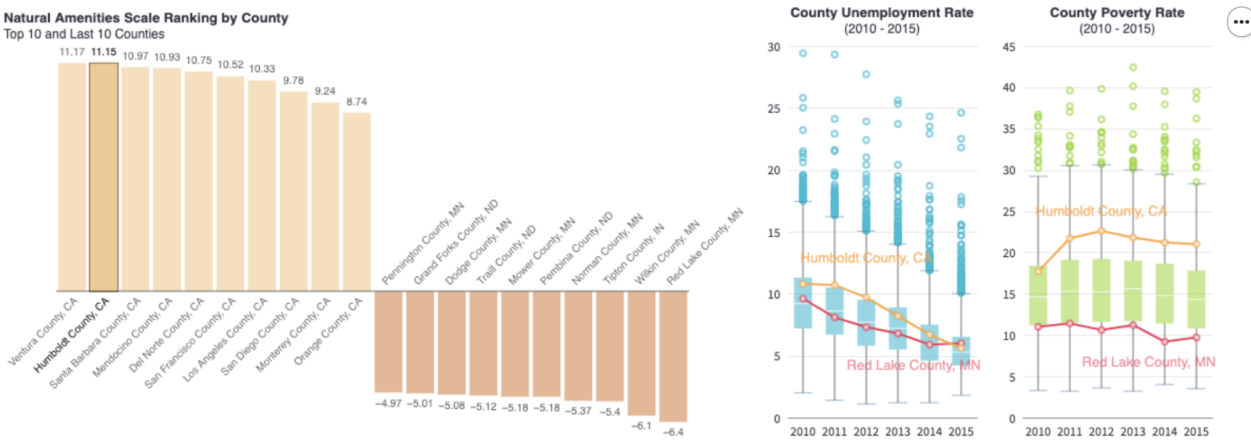

For the final design (as shown in the image above), I used a positive/negative bar chart to represent the natural amenities ranking of the top 10 and last 10 counties so that users will be able to easily tell which bars represent the tops and which represent the bottoms. To let users realize that the bars (counties) could be selected and are clickable, I added a mouseover interaction for the bars so

that when users hover over, the bars would be highlighted. For the lines that pop up in the boxplots after selection, I chose a relatively eye-catching color so that users could easily tell that the popped up lines are the counties they selected for comparison. To allow users to quickly compare Red Lake County with the rates across different counties, the line for Red Lake County constantly stays on the boxplots and are also encoded with a salient color (red).

Visualization 2:

For the second visualization, I was also debating on whether I should include the selection for counties with higher natural amenities ranking. However, from the original static version, it was already obvious that Red Lake County has a relatively lower median home value and higher household income than other counties over the U.S. Therefore, I believe that it wouldn't be that worth it to have an additional graph for this visualization as compared to the first visualization.
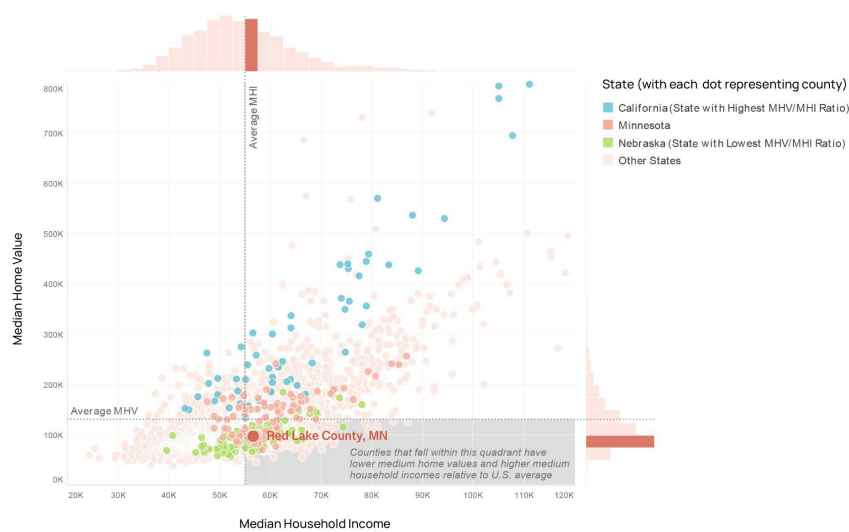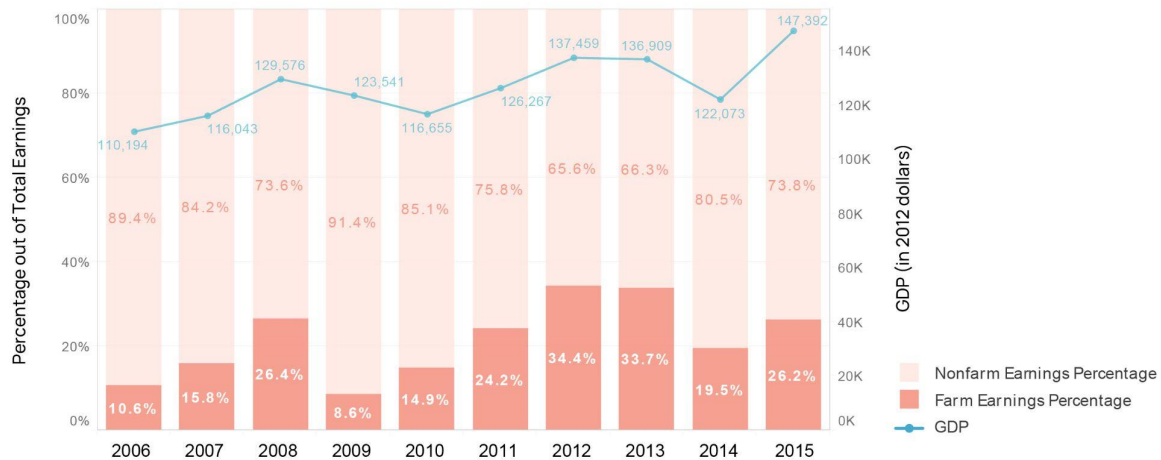

(Static Version)

One concern for the original static version that I had was that only three states were able to be highlighted because if each state had their own color, the scatter plot would turn out too messy and difficult to read. With the help of interactivity, I was able to fix this problem by utilizing interactive selections to filter down different states.
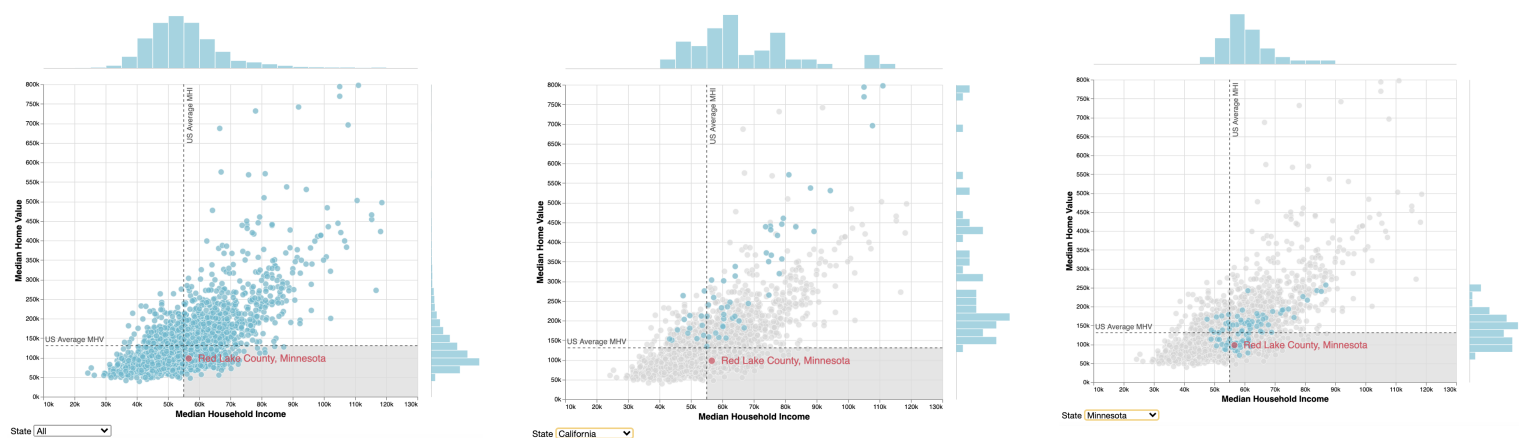
For the final design (as shown in the image above), I used a dropdown selection box for users to switch between states. When a certain state is selected, the data points of counties outside of the state will become less visible, making it easier for users to spot the distribution of counties from the selected state. The filter also applies to the histogram on the top and right side of the scatter plot as selecting different states, which helps emphasize the difference in distribution for each separate variable (median home value and median household income). Additionally, tooltips were added to provide more details for each data point so that users could actually tell the county it represents and the exact MHV/MHI values.
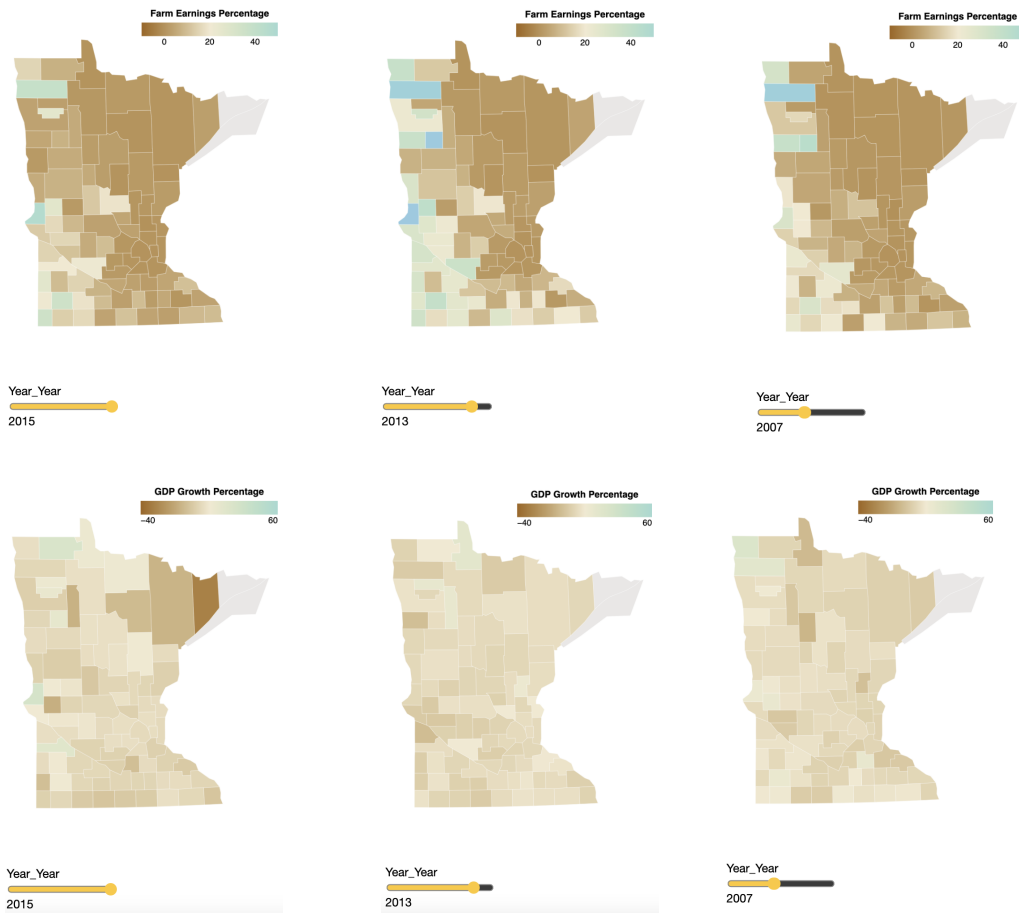
Visualization 3:

For the third visualization, I made bigger revisions from my previous static version as I received feedback on the static visualization that dual y-axis isn't really a good way of presenting data. Therefore, I went on brainstorming new methods of presenting the two variables.
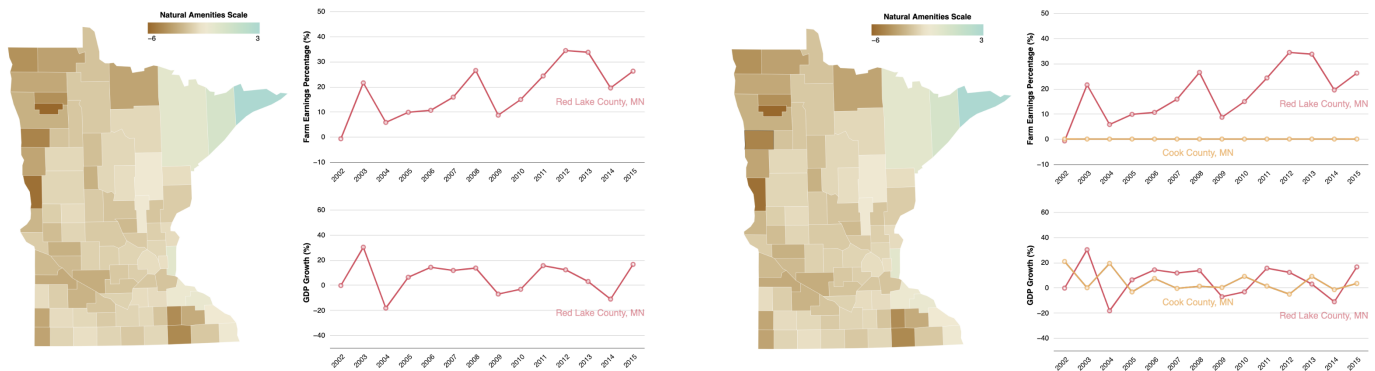

(Static Version)

Before that, I made the decision to expand the data and involve data from other counties so that there could be more possibilities for this visualization. One idea that I came up with was to show a Minnesota map divided by counties and encode the two variables with color gradients. In terms of interactivity, users could use the slider to switch between years to see the changes over time (as shown in the graphs down below). After coding out this graph in Altair, I found that it didn't turn out as effective as I expected. The differences across different counties were hardly noticeable and difficult for users to come up with quick conclusions.

Therefore, I went on brainstorming better ideas that would be suitable for the domain questions. As I was rethinking about the relationship of the visualization with the domain questions, I noticed that the reason for emphasizing the importance of farming to Red Lake County wasn't really explained from the static version, which may affect the overall flow of the narrative. Therefore, I decided to again include the natural amenities selections in this visualization so that users could directly compare counties that have higher natural amenity scores with Red Lake County.

For the final design (as shown in the images below), I created a geomap of Minnesota for county selection rather than selecting from overall ranking because, in terms of agriculture, it would be fairer to compare Red Lake County with counties that have similar environments and climates. This will also help narrow down the number of selections so that the graph could remain neat and clear. Similar to the first visualization, when a county is selected, a line representing the time series data of that county will pop up both in the Farm Earnings Percentage chart and the GDP Growth Chart. Again, this helps users to quickly compare information of specific counties with Red Lake County, making it easier for users to come up with explicit conclusions.
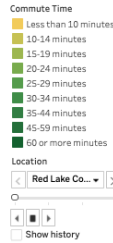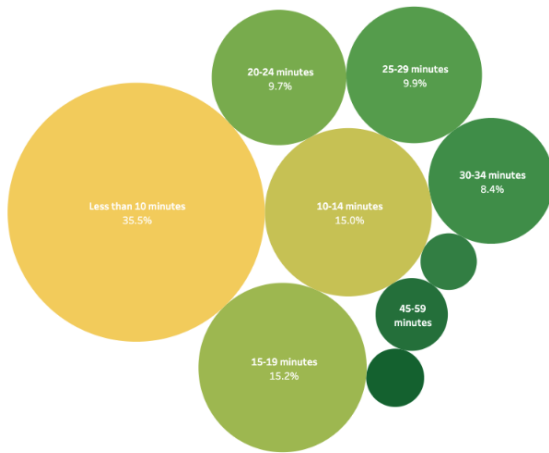
For the last visualization, I took the idea that I brainstormed for the static version -- using packed bubble charts to encode the percentage of each time interval with bubble size.
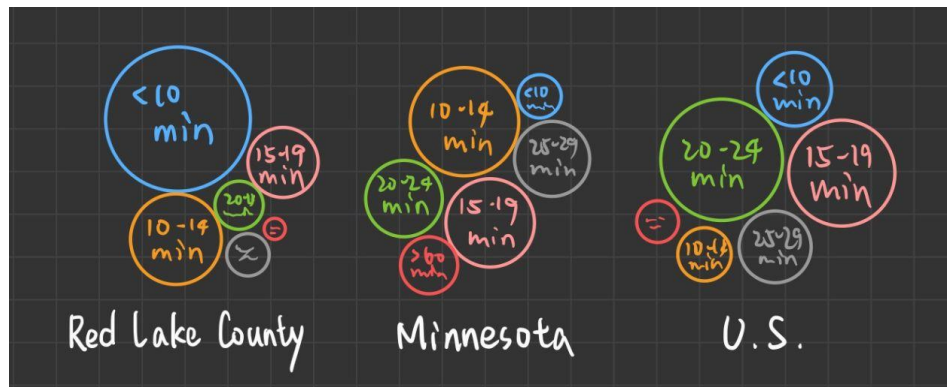

Though circle sizes are usually harder for users to compare, the differences in bubble size difference across locations would be obvious when it is done interactively. In order for users to see the gradual changes of bubble sizes when switching between regions, I chose to use Tableau for creating this visualization because the packed bubble charts in Tableau are much more dynamic and easy to animate. For the final design (as shown in the images down below) I used color gradients to encode different time intervals so that users could intuitively get a sense that the lighter the color, the shorter the commute time is. In terms of interactivity, I used the slider widget for switching between locations so that users could switch through the options quickly to compare the size differences for each time interval.

Commute Time to Work by Location
Red Lake County, Minnesota, New York, U.S.
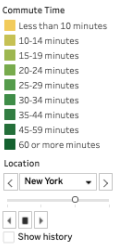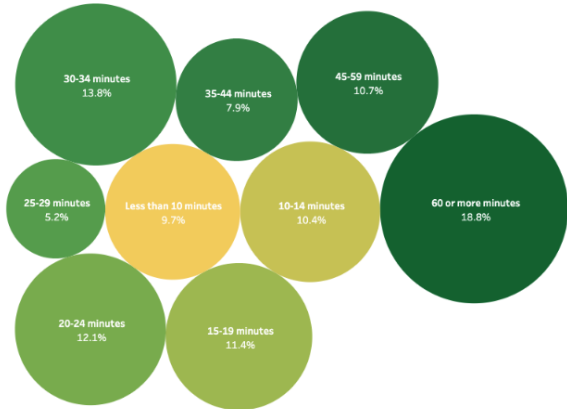
Evaluation

To evaluate the interactive visualizations that I created, I would again conduct an experiment to see if users could successfully meet the learning objectives. But this time, I would also be focusing on the "usability" of the graphs and see if users were able to properly interact with the interactive widgets/tools. Again, for this experiment, I won't be looking for a specific target audience/user, but users will be required to finish reading the original article written by Christopher Ingraham prior to starting the experiment. I would start the experiment by having the users look at each of the graphs and ask them to "think out loud" as they interpret the visualizations. By doing this, I would be able to not only test if they were able to understand the graphs, but also make sure that the users are focusing on the right information when viewing them. During the process, I would observe how they use the graphs and see if all of the intended interactive tools were used as intended. After having them view the graphs, I would ask them a few questions related to the learning objectives such as:

1. How can the home value/household income distribution be classified? Which quadrant was Red Lake County in and what does that mean?
2. Could you explain where Red Lake County falls within the unemployment rate and poverty rate distribution?
3. How is Red Lake County's unemployment and poverty rate compared to the counties with higher natural amenities ranking?
4. What industry plays a significant role in Red Lake County's local economy despite the county being ranked last in natural amenities? How is Red Lake County's farm earnings percentage

and GDP growth over the years compared to the counties with higher natural amenities ranking?

5. How would you compare commute time in Red Lake to MInnesota, New York, and the entire U.S. ?

Towards the end of the experiment, I would again have them talk about whether these data and visualizations successfully supported the concept of the original article and whether it made them feel that Red Lake County was overall "not that bad of a place to live".