

CS765: Design Challenge 1 Description

Due on October 2, 2019 at 11:59pm

Katie Mummah

Question 1

What tools you used to create the visualizations. If you did any programming, explain what the program does (e.g., “I wrote a Python script to transform the data by aggregating by state. This made it easier to make the graphs I wanted in Excel.”). (this can be a 1-2 sentences, but you might have separate descriptions for each visualization). Also tell us what libraries you used.

Four out of my five visualisations were created using R. I relied mainly on the libraries contained in the [tidyverse](#), including the plotting library [ggplot2](#), the data manipulation package [dplyr](#), and the read package [readr](#). Beyond these three useful libraries, I used the [viridis](#) color palette(s) and a package called [ggrepel](#) to help prevent my text labels from overlapping.

I used an R notebook with small individual scripts to generate the plots. I used dplyr functions to filter the data, such as only looking at counties with a median household income of more than \$100,000 (Images 1 & 1a). I also used dplyr’s group and summarize functions to find the average unemployment rate and natural population change for each urban influence code (Image 3).

My final plot (Image 2) was generated using Tableau. I divided the 2015 population change by the 2015 population estimate for each state to get the variable I displayed on the map.

Question 2

For each visualization (including the alternate), explain what the story you were trying to convey is, and why you think the design that you chose is appropriate. See the grading criteria. (this doesn’t need to be too extensive 1-3 sentences per visualization is usually sufficient)

- The wealthiest counties in the US typically exist in the wealthiest states, with only two exceptions. Visualizations 1 and 1a both highlight this discrepancy. The bright yellow and orange immediately grabs the attention as an outlier demonstrating that even the median household in Los Alamos or Williamson counties makes more than 2.2 times the average income in the state.
- Image 1a (Figure 2) is the same information as Figure 1, but this time the story is to convey the relative linearity of increase in household income and increase in income relative to the state norm. In this case, the colors are not the eye-catching part of the outlier, but rather the position of the dots. I believe it is effective because the viewer has multiple visual clues that lead them to immediately recognize the importance of the two outlier counties as the main story.
- I chose a map to look at the states that are gaining and losing population (relative to their total number of residents) for Image 2 (Figure 3). This design is appropriate because we all visualize the US when trying to make sense of patterns across the country, and this design makes it easy to see which regions are staying steady or losing population (northeast, midwest) and which populations are growing (west, southeast).
- Image 3 (Figure 4) tells an unfortunate story of struggling rural America, right here in Wisconsin. Counties that are more rural (higher urban influence number) have higher unemployment rates (making them less attractive), which is conveyed using the familiar length encoding of a bar chart. More rural areas are also not growing naturally, even having a negative natural growth rate that represents a faster rate of citizens dying than being born. I believe a bar chart was appropriate because I was displaying one categorical variable (urban influence) alongside a quantitative one (unemployment rate) and a tertiary contrived quantitative value (natural population change).

Census areas with the highest median household income are almost all located in states with high average incomes.

Two major exceptions are Williamson County, TN, a suburb of Nashville that boasts many of Tennessee's richest, and Los Alamos County, NM, home to a well-paying national laboratory... and not much else.

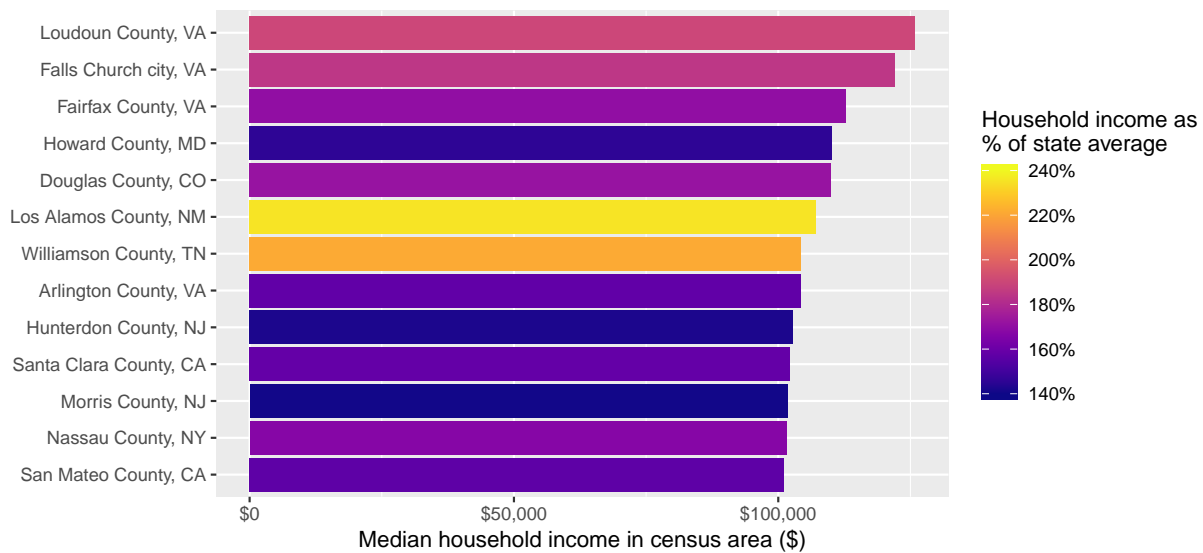


Figure 1: Image 1

- Image 4 (Figure 5) is a fairly straightforward and logical story— counties with a majority of their population holding at least a bachelors degree are more likely to have a higher median income and a lower (5%) unemployment rate. I chose a log-log scatterplot because the data is almost linearized in this transformation, allowing the viewer to observe the gentle but notable shift from lower degree-holding counties (purple) in the upper left corner to higher degree-holding counties (orange, yellow) in the lower right hand corner.

Question 3

For the alternate design, discuss which of the 2 different designs you think is most effective at conveying the story. (1-3 sentences)

I think that design 1a is more effective at conveying the main point quickly— that Los Alamos County and Williamson County are outliers among this particular subset of data. While the colors stand out in design 1, the points in 1a not only grab the viewer's attention equally quickly, but also provide a quicker sense of whats going on, since one only has to glance at the nearby Y axis as opposed to having to decipher the legend and color scheme.

Census areas with the highest household income are almost all located in states with high average incomes.

Two major exceptions are Williamson County, TN, a suburb of Nashville that boasts many of Tennessee's richest, and Los Alamos County, NM, home to a well-paying national laboratory... and not much else.

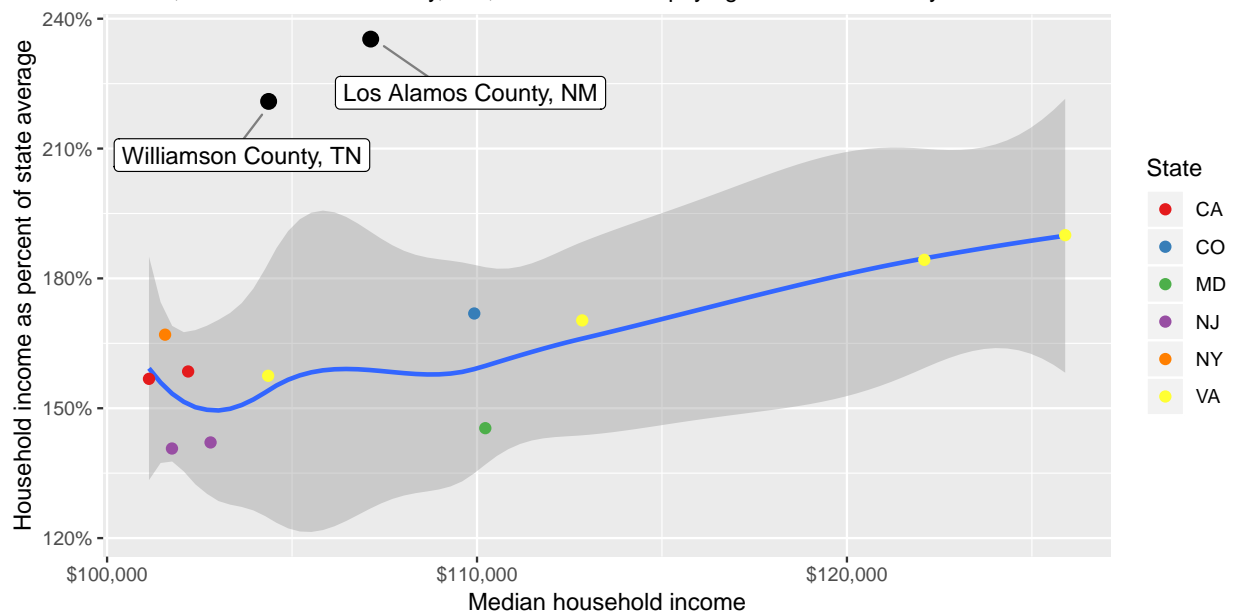


Figure 2: Image 1a

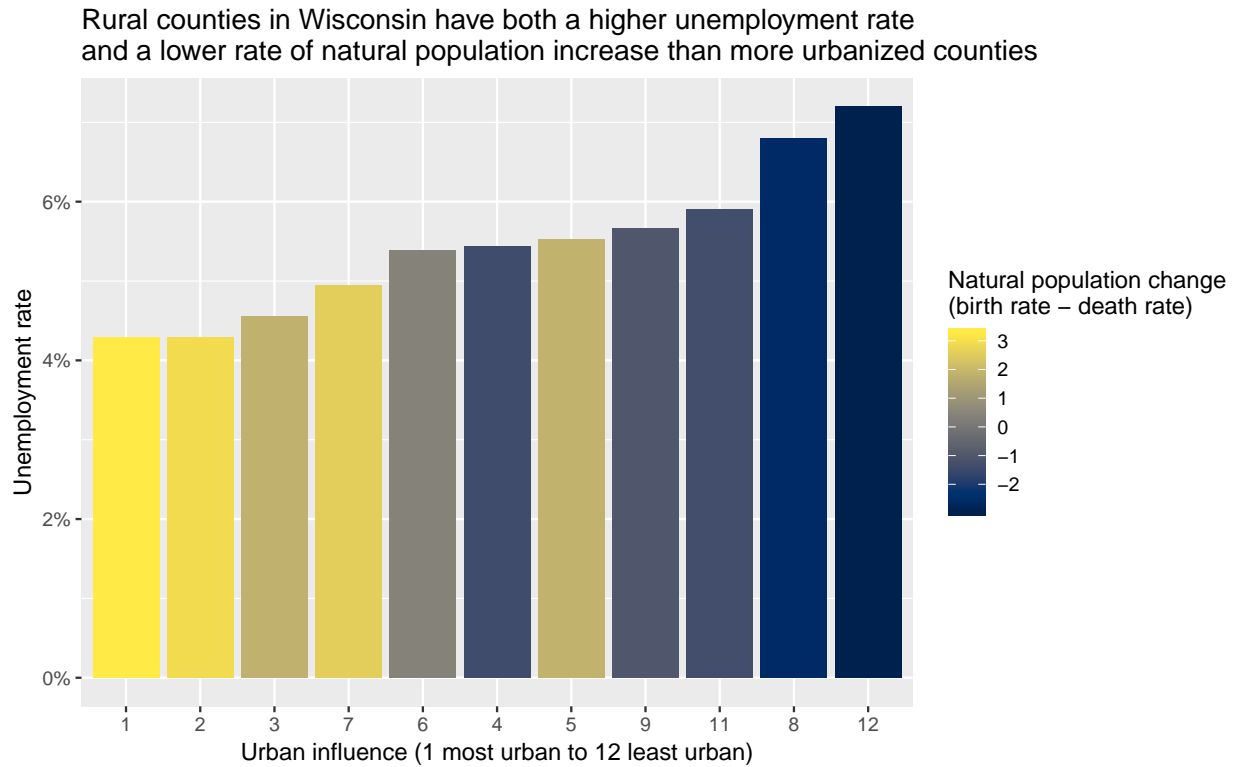


Figure 4: Image 3

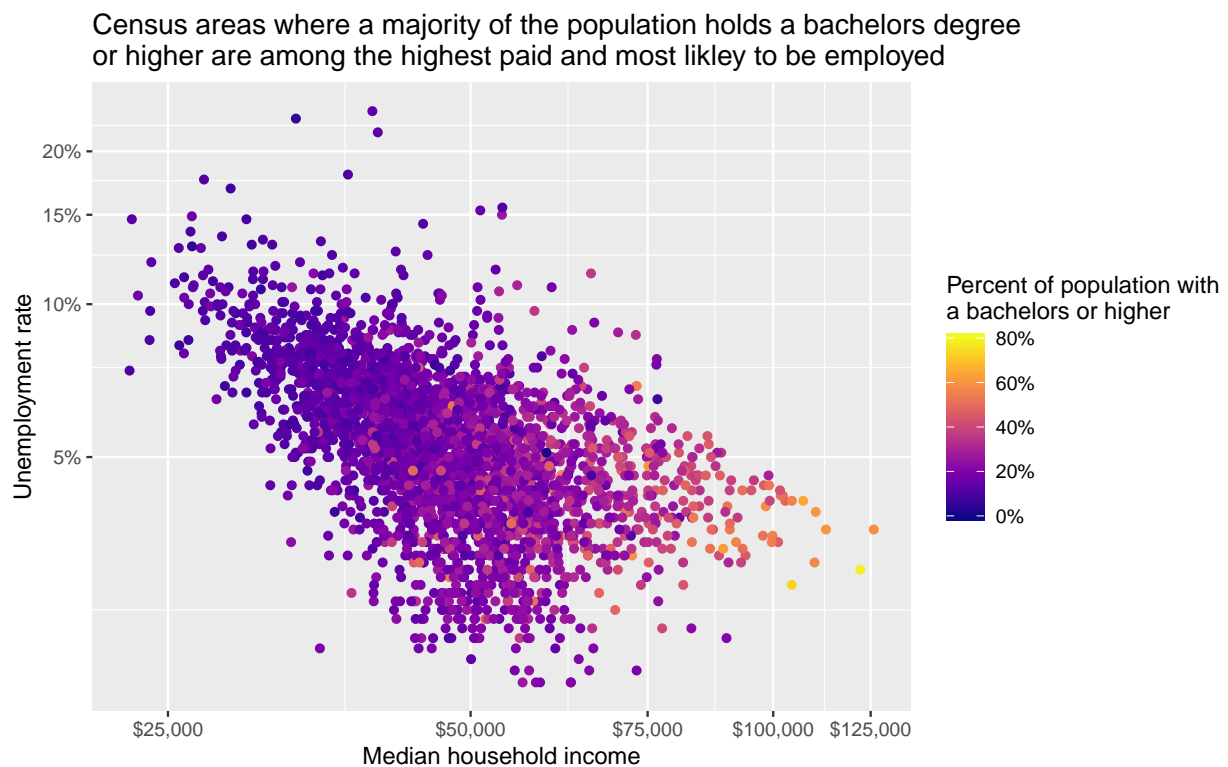


Figure 5: Image 4