Communication Visualization - Static

Nate May 27 | Oct 2021

Introduction

This paper discusses my approach for supplementing Christopher Ingraham's 2015 article entitled "*I called this place 'America's worst place to live.' then I went there*". with information, visualizations to improve upon the effectiveness of his communication objectives. This first iteration is intended to be a static article with a subsequent push to make it interactive. Until then, you can access the static version [HERE](#) and the source code [HERE](#).

Data

I began this project by first looking for available data that could support it. Some links were provided but I wanted to end up with something unique from my peers. Having discovered a small dataset of [Population Census Data](#), it occurred to me that changes in population might be more informative than the USDA's natural amenities index referenced in this article. Ultimately, used the census data for an aside discussed later in this paper, [Unemployment](#) data from the same USDA source, and the [Natural Amenities Scale for U.S. Counties](#) provided in the assignment to supplement details in the content of the article. With the data in hand, I then laid out my objectives.

Learning Objectives

1. The user will be able to compare and contrast how counties across the US (with emphasis on Red Lake county) compare with respect to the metrics used in the article as well as other measures related to population growth.
2. The user will be able to understand the conceptual framework of the N8 Index.
3. The user will be able to recall that while Red Lake County was lowest on the Natural Amenity Scale it ranked moderate to lower on other measures mentioned in the article.

Inspiration

As a web developer, I set a personal goal to familiarize myself with the [D3](#) charting library. Knowing that this would be a steep learning curve, I decided to aim for visualizations with less complexity than those which we had created with Altair in previous assignments. I referenced D3's provided examples to understand the range of capability and as a source of visualization inspiration. While going through each example I considered how the data that I had gathered might be presented in such a visualization. Ultimately I decide on the [map](#), a [barplot](#), a [scatterplot](#), a [line chart with bands](#), and a [violin plot with jitter](#).

Design Process

Aside

My first set of visualizations were related to the census data and my argument for population growth as an alternative to the natural amenities index. Having written a short conceptual summary of the "N8 index" in support of <u>learning objective 2</u>, it was clear that my description would be well supported with descriptive visualizations.

Specifically, I wanted to communicate:
- Why using a single decade might be misleading and
- What were the rankings suggested by this new metric

My first iteration led me to have 4 bar plots showing the counties with the largest and smallest population growth in each decade, as well as 3 scatterplots comparing the rates of change between each of the subsequent decades. These were difficult to compare and contained redundant information, but I was able to consolidate them into the single scatterplot below.
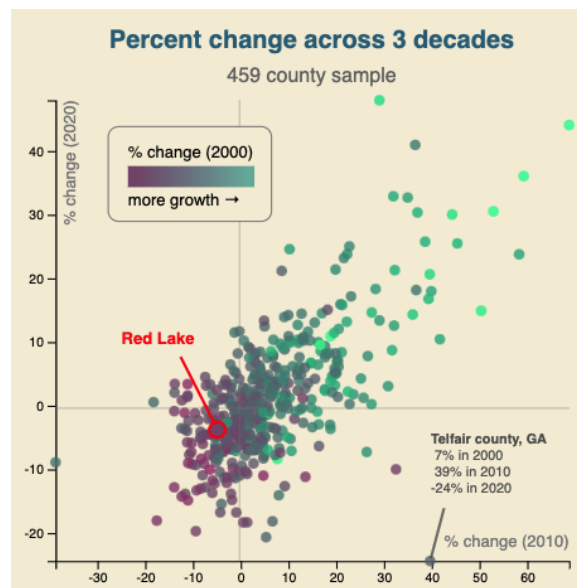
This visualization shows the percentage change in population in 2000, 2010, and 2020 against the color scale, x-axis, and y-axis respectively. The summary calls out Telfair Georgia as a county with significant population change variability as a means to justify my normalization process. I added an annotation to point out Telfair as a point of interest as well as Red Lake county. Throughout my visualizations, I point out the Red Lake data point to meet <u>learning objective 1</u>.

Having explained the concept of the N8 index, it was now time to visualize the rankings that it suggests. I began with 2, vertical bar plots for the counties scoring highest and lowest, but I felt that they were a bit difficult to compare given that they did not share an axis. After combining their axes, I found it difficult to label them properly. Finally, I stacked the bar horizontally and met all of my requirements.

Counties ranked by N8 Index™
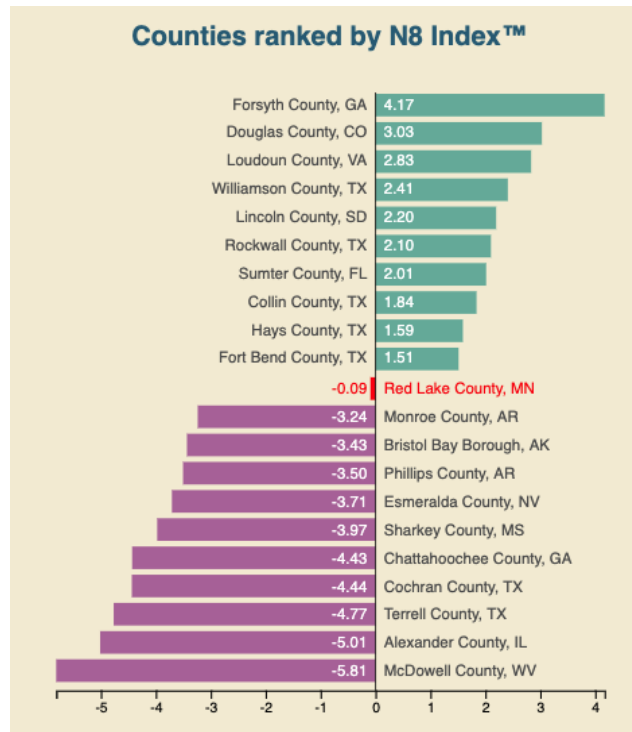
This visualization clearly describes the ranking while providing salient values, labels, and once again, points out where Red Lake falls within this ranking.

Map

While I was proud of my aside, I felt that I had condensed it so much that I had fallen below my intended scope. There were 3 sections of the article that I felt I could support with visualizations. The first of which was the natural amenities score. Having seen a mapping of the natural amenity category (1 to 7) on a US county map, I thought it might be better to map on the actual amenity score (-6 to 12).
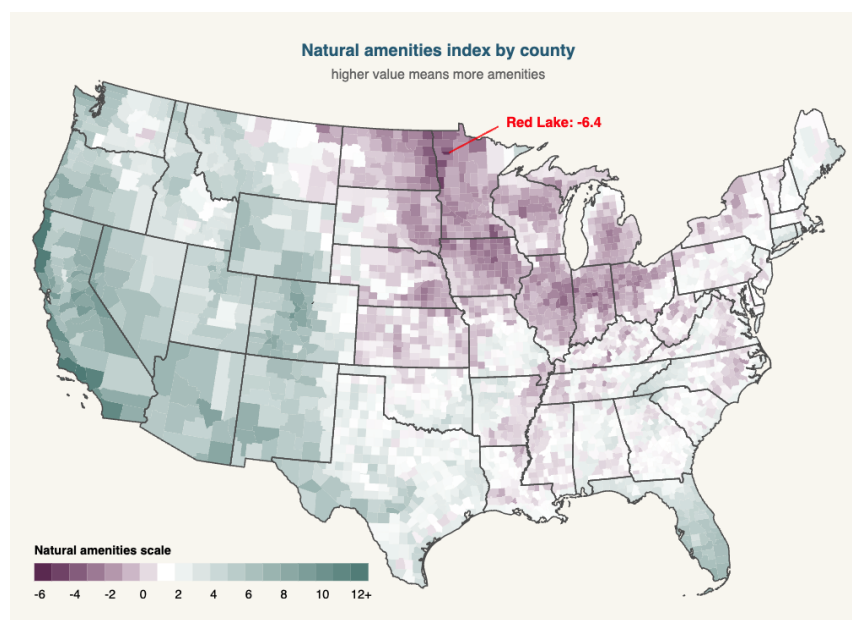
While this gave some context to the index and how it is distributed geographically, it does not directly address any of my learning objectives. However, I do annotate Red Lake county as usual.

Line Chart with Bands

Christopher mentions the employment rate [sic] of 4.6% as well as some other economic metrics. By itself that number may not mean much, so I've visualized unemployment rates for the last 30 years showing the higher and lower bounds among all counties as well as the national average.
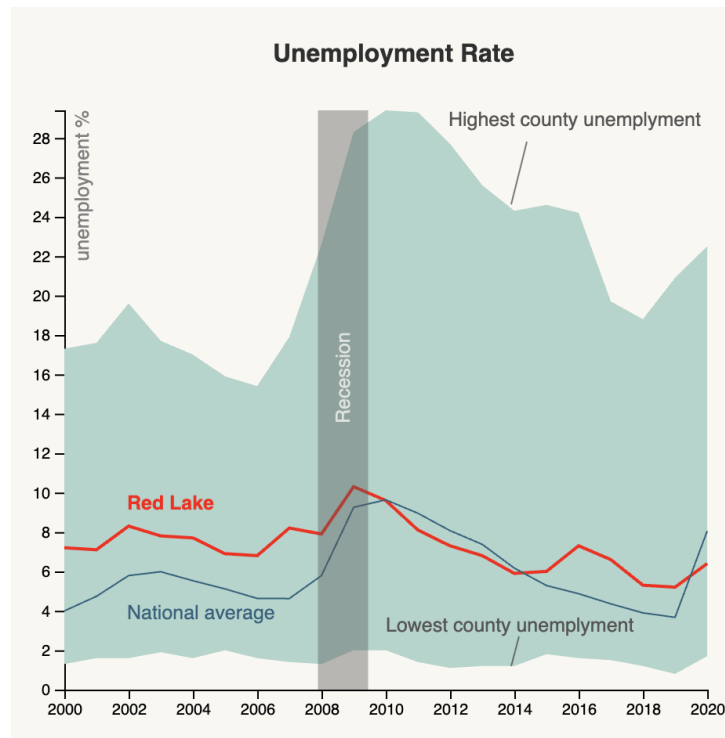
This graph support learning objectives 1 & 3. My hope is that viewers understand that Red Lake county has not struggled economically relative to many other counties and that it's pretty close to the national average. I've even displayed the 2008 recession to provide an explanation for the uptick in unemployment during that period.
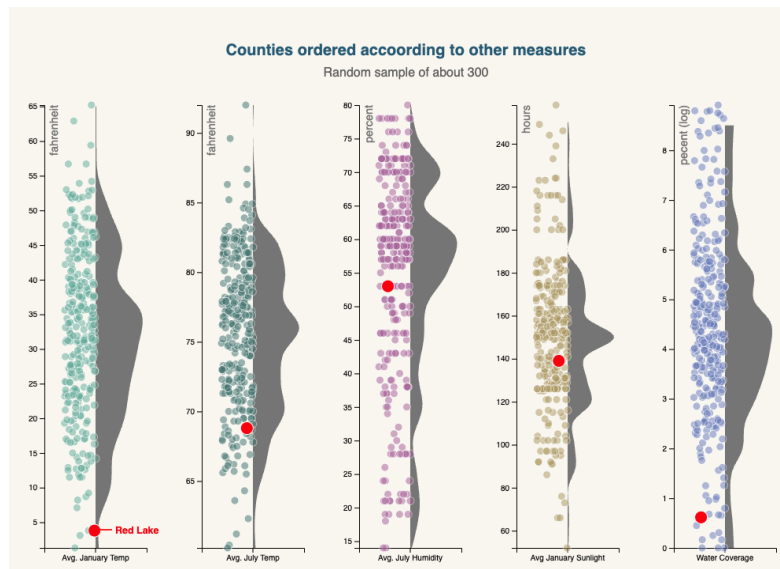
Violin Plots

Elsewhere in the article, Christopher mentions that the average January temperature is 4 below, so I thought I could help readers understand where Red Lake county falls within the distribution of all counties according to this and other measures provided in the Natural Amenities Scale dataset. These measures include
- the average sunlight hours in January
- the average January temperature
- the average July temperature
- the average July humidity, and

● the percent of water coverage in the county

I considered visualizing these with a histogram but I discovered that binning made it difficult to isolate Red Lake. Ultimately, I settled upon a violin plot with jittered offset datapoints.


Like a histogram, the violin plot gives a sense of the distribution, but in this case, the individual data points are still displayed to support the histogram with more granular detail. This also allowed me to highlight Red Lake county to give it more salience for easy comparison to other counties. My hope is that the reader observes that Red Lake county falls near the average on many measures despite having the lowest natural amenities score, thereby supporting learning objective 3.

Design Principles

Color

Red and green are a typical color scale for showing values considered to be better vs worse than one another. With that being said, I also need to highlight Red Lake county which I felt should share the semantic color of "red". This meant that I needed a highly salient red to highlight Red Lake county and a distinguishable and shaded offset of the red hue for my scale values. Having chosen a magenta color for this, I then used the site Muzli to generate a palette that could support red, green, and red-green scales. Later I also added a blue of a similar tone and chose other styles to complement the aesthetic of the palette.
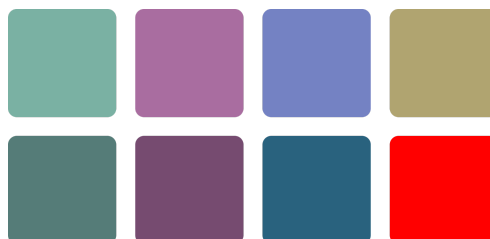
Sampling

The 3000 counties resulted in too many data points for visualization clarity in my scatter plot and violin plots. For each of these, I chose to reduce the number of data points by using a [stratification sampling](#) method that ensures that the sample distribution is representative of the population distribution. Giving the data points opacity also helped to communicate density more effectively.

Evaluation

In the context of my three learning objectives:

1. The user will be able to compare and contrast how counties across the US (with emphasis on Red Lake county) compare with respect to the metrics used in the article as well as other measures related to population growth.
2. The user will be able to understand the conceptual framework of the N8 Index.
3. The user will be able to recall that while Red Lake County was lowest on the Natural Amenity Scale it ranked moderate to lower on other measures mentioned in the article.

I would test each of these in different ways. For learning objective 1, I would ask a viewer to rank Red Lake county according to the specific measures mentioned in the article such as unemployment and daylight hours to test their recall in the context of Red Lake's ranking of a single measure. I could also ask them which regions of the US scored higher on the amenities scale. For visualizations that show density, I could also ask what was the most common or average value and where Red Lake county fell in relation to that average.

For learning objective 2, I would provide the viewer with a few countys' census data and a calculator and ask them to calculate the N8 index. I'm not sure at this point if it would be better to give them the percent change formula or simply give them the change rate for each decade directly. In either case, they should know to multiply the changes rates and omit counties that have decades of both growth and decline.

For learning objective 3 I would ask a viewer to rank Red Lake county. I would point out that the natural amenity index had Red Lake in dead-last, but that they should take into account the other factors presented in the article as well. My hope would be that viewers would Red Lake in the interquartile range and cite other measures mentioned in the article to support that ranking.