CMSC 23900

Project 1

<p style="text-align:center">Project 1 Writeup</p>

<p style="text-align:center">By: Andrew Vallaster, Omar Salemohamed, and Yasoob Rasheed</p>

Purpose

The purpose of this visualization is to examine the relationship between different measures of poverty within a geographic context. More specifically, we have taken both the participation rates of the federally funded Supplemental Nutrition Assistance Program (SNAP) as well as the number of housing units available to extremely low-income (ELI) renters as indicators of poverty and found data with regards to both variables at the county-level. We map both the county-level SNAP participation rate, as well as the number of housing units available to ELI renters per 100 units within that county. As such, our visualization lends itself to questions such as: in which regions of the U.S. do people have access to affordable housing with little need of nutrition assistance? Additionally, we are trying to investigate if there is an underlying relationship between a county's participation in the SNAP program, and the number of housing units it has available to ELI renters. Moreover, we want to see if such a relationship can be understood within a larger geographic context. For instance, are there notable trends between these two data points and the geographic identity (urban vs. rural) of a given county?

Pre-Processing

Our goal for pre-processing was to create a JSON object where counties were keys and each county was associated with a SNAP data point and an ELI housing figure. We were able to access CSV files for both data sources (listed in our data sources). For the SNAP data, counties were identified by their names, and the participation rates were all listed in one column, so it was easy to create a dictionary with counties as keys and SNAP data points as values. For the housing data, we found that the data point we wanted could be in one of two columns, so we had to account for that when processing the CSV. We created a new dictionary with county IDs as keys, and stored the housing data points. We then migrated the data in the SNAP dictionary into this new dictionary. Counties were only identified by county name/state in the SNAP data, whereas counties also had an ID in the housing data, so we first needed to create our SNAP dictionary before we could migrate the data. We decided to use county IDs as keys

instead of county names, as the TopoJSON file we used to render the map identifies counties by their IDs. Thus, when binding the data to the svg, the only way to associate the TopoJSON data with our housing/snap data was via a county ID. Once our dictionary was populated, we wrote it as a JSON object.

How It Works

We use a map of the U.S. with county-level granularity to layout our SNAP and housing data. We felt that the map was particularly useful for identifying the kind of regional trends that our questions about the data were getting at. Moreover, given that our data showed that a county's SNAP and ELI housing values are often similar to those of its neighbors, a map is better suited for blurring the county-level (categorical) borders and identifying broader clusters with similar data points.

We encoded SNAP and housing data values through color. The reason behind the decision was that we believed that color could best be used to encode ordinal information. Because our intention was to map the relationship between SNAP participation and the number of housing units available for ELI renters, it was important for us to compare how higher or lower SNAP participation rates corresponded to a greater or fewer number of units available. Thus we were less concerned with the actual data values and more interested in ordinal differences. As such one of the trade-offs we made was the reduction of both our SNAP and housing data from ratio and interval values, respectively, to ordinal values. Another decision we made for the purpose of making out visualization easier to understand was to have higher intensity, as conveyed by decreasing lightness in the color scheme we chose, correspond to both higher SNAP participation rates and fewer ELI housing units available. One could argue that it would have made more sense for increasing intensity to correspond to higher values of both data points, but given the questions we are trying to ask, we'd argue our approach is more appropriate. Given that in a broader sense we are examining measures of poverty, it is more natural for our viewers to conclude that increasing severity (as denoted by decreasing lightness in the white-black color opponent color channel) corresponds to greater degree of poverty as indicated by both data points.

Alternatives & Arbitrariness

Other possible representations of our data include:

- Different choices of color to represent both variables
- State-level granularity as opposed to county-level granularity within the map

- A scatter plot representation of our data with SNAP rates and # of housing units on the x-y axes and text labels or color encodings for each county.

One of the major design driven decisions we made in regards to our visualization was to represent our data with a map. The choice of a map as opposed to some kind of chart is that the former, as aforementioned, is more conducive to answering questions related to the relationship between our county-level data, and the geographic location of the county. Moreover, we thought it was important to show the data at the county-level as opposed to the state-level, because data points for both values often varied extensively from county to county within a given state, and we thought it was important to represent these variations to determine if there was an underlying geographic reason that could be attributed to these variations.

We would argue that the color scheme we chose was both a consequence of data-driven encoding as well design considerations. We opted for a blue-yellow color scheme, as they are a color opponent channel. As such, they allow for a clearer delineation between the two variables we are showing. Although one could argue that a different color scheme could have produced a similar representation, the blue-yellow parallelepiped out of the LAB color space is one of the few uniquely suited options for making ordinal comparisons. An alternative could have been a red-green parallelepiped out of the same space, but we opted for blue-yellow, as people with color blindness are typically more receptive to the latter. Furthermore, we used the white-black opponent color channel for encoding trends across the horizontal portion of our legend - namely where housing and SNAP data values for a given county had the same level of intensity such as high SNAP participation rate - few housing units available for ELI renters. Another arbitrary choice was the decision to encode the states with missing data for either data point as green. We did not want a visually striking color because we didn't want the viewer to be focused on absent data.

SNAP Participation vs. Units Available to ELI Renter Households
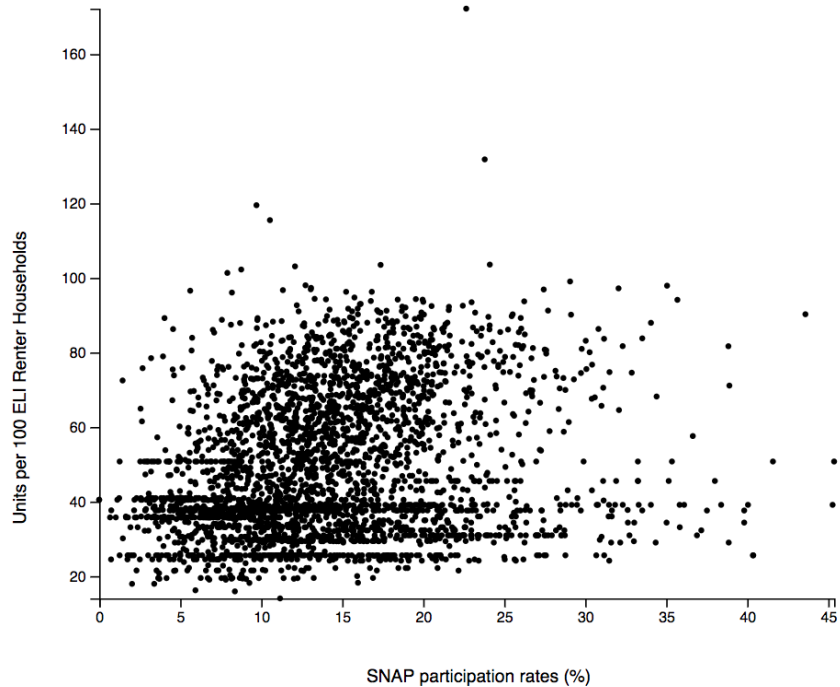
From the scatter plot, we did not see an obvious way to divide the data so we experimented with 3x3, 4x4, and 5x5 grids for both the snap and housing data. We ultimately settled on a 4x4 grid because the data showed more variation than a 3x3 grid accounted for but was not as cluttered as a 5x5 one. We wanted to be able to show enough variation between the counties; thus, choosing a 4x4 grid allowed us to make use of enough distinct data points from the scatterplot above.

Distribution of work:

Andrew:

   Preprocessing housing data

   Identifying and implementing color scheme for bivariate choropleth

Omar:

   Preprocessing SNAP data and writing util functions

   Rendering empty county-level map of US

Yasoob:

Preprocessing housing data and writing util functions

Creating scatter plot to determiine distribution of values

Group:

Identifying data sources and choosing topic

Writeup

Data Sources

For both of these data sources, we downloaded the csv files linked at the sites.

Housing Data:

http://apps.urban.org/features/rental-housing-crisis-map/

SNAP Data:

http://www.frac.org/snap-county-map/tables/snap-county-tab-2016.html