Final Data Visualization

Prepared by:

Jason Lewris

**Background:**

For the final data visualization assignment, I wanted to explore the relationship between income and election results from the 2012 election with Mitt Romney and Barack Obama. I was interested in gaining a granular view into the country and how various counties voted and what the income levels were for those particular counties. My original theory was that wealthier counties tended to vote Republican, and I wanted to explore this theory through a visual interface.

**Introduction:**

In order to gain a better understanding into the relationship between election results and income data per county, I had to pull data from a multitude of sources, organize it, and format it in a way that was easily digestible by the visualization engine. Table 1.1 below shows the data sources pulled for this particular visualization:

|  |  |
| --- | --- |
| **Data Source** | **Data Location** |
| 2012 County Election Data | The Guardian [1] |
| 2012 County Election Data | NY Times [2] |
| County Income Bracket Data | IRS [3] |
| County Median Income Data | Census [4] |

I wanted to provide a more granular view into income, more so than just median income per county. The IRS provides data regarding the distribution of tax returns per income bracket for individual counties. The following income bracket schedule was utilized within the dataset retrieved from the IRS:

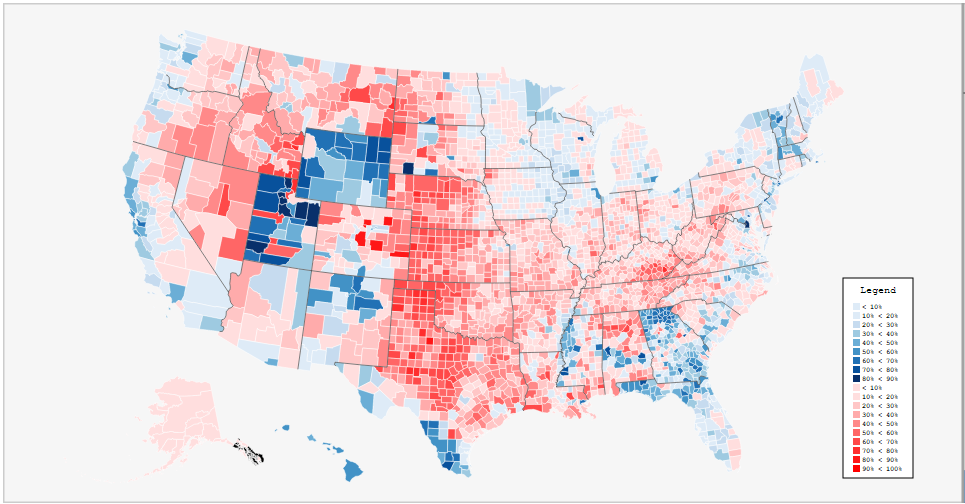
|  |
| --- |
| Under $1 |
| Under $10k |
| Under $25k |
| Under $50k |
| Under $75k |
| Under $200k |
| Over $200k |

This data needed to be manipulated and formatted in many ways. The original form in which the data was delivered was separate folders per state, and then within each state folder contained county information. Using python, a master dataset was created for IRS income bracket data to develop a single file containing all counties, states, income bracket distribution information, and county FIPS codes. This was then combined with the county median income data retrieved from census.gov. From here, two separate datasets were maintained, one containing election data retrieved by the Guardian, and the second containing income data from both the IRS and Census. Some election data was missing and had to be retrieved manually from the New York Times presidential election map in order to fill in the missing values in the master election spreadsheet.

**Visual:**

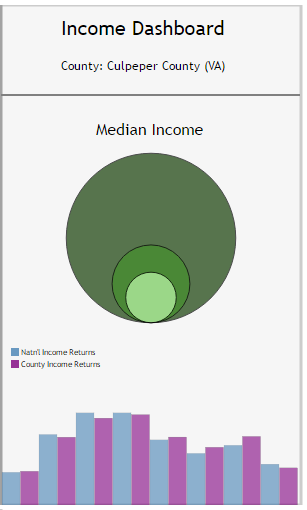
As mentioned previously, the goal was to develop an interactive visualization to enable users to explore how election results and income were disbursed across the nation. In order to do this, several prototypes were developed, many of which included the use of an interactive d3 map. Learning lessons from a tutorial provided by Mike Bostick titled “Choropleth”, I was able to generate a choropleth map of the United States, where each county was colored based on how it voted [5]. This was a slight pivot from the original map which incorporated income data, followed by election data that was presented upon clicking. Figure 1.1 displays the results of this choropleth transformation:

Figure 1.1



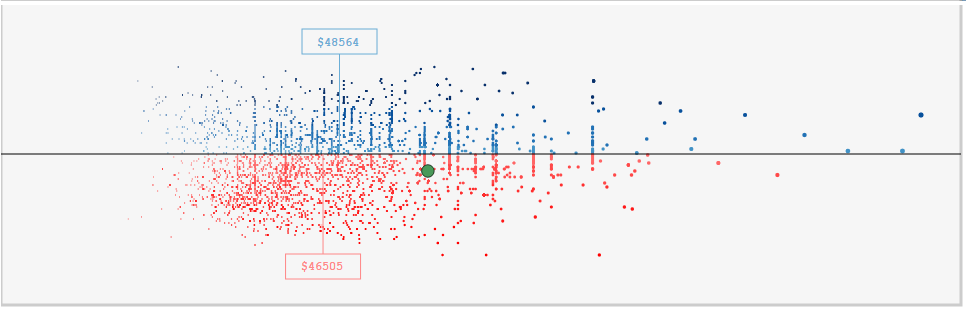
In order to give the user a deeper sense of interactivity and exploration, the dashboard approach was expanded to include many other elements that react based on how the user interacts with the map. The right quadrant of the dashboard presents the user with income level data. A histogram bar chart was developed to map out the national averages for each income bracket as a percentage of the total, and then upon a user clicking on a county, county level income distribution is drawn alongside the national averages. Additionally, to give the user a sense for the overall economic health of a particular county, a circle packing type visual was developed where the smallest and largest circles represent the smallest national median income county level, and the largest national median income level county. The middle circle initially represents the average median income, but upon clicking a particular county, the middle circle resizes based on that counties median income. Figure 1.2 shows an example of the right quadrant of the dashboard for Culpeper County (VA):

Figure 1.2



Lastly, in order to give the user a sense for how the distribution of income changes based on party support, a separate quadrant was developed below the interactive map to display all county level median income data, in conjunction with separating the democratic and republican voting tendencies. The x-axis represents county median income level, while the push upward for the blue area representing democrat represents how strongly that county voted democrat, and the push downward represented by the red represents how strongly that county voted republican. The average median income for each party is displayed, and when a user clicks on a particular county, where that county is on the graphic is highlighted in green. Figure 1.3 below shows the bottom quadrant of the dashboard:

Figure 1.3



**Sources:**

[1] Rodgers, S. (2012, November 14). Full US 2012 election county-level results to download. Retrieved

from The Guardian website: http://www.theguardian.com/news/datablog/2012/nov/07/

us-2012-election-county-results-download

[2] President Map. (2012, November 29). Retrieved from New York Times website:

http://elections.nytimes.com/2012/results/president

[3] SOI Tax Stats - County Data. (n.d.). Retrieved from IRS website: https://www.irs.gov/uac/

SOI-Tax-Stats-County-Data

[4] 2014 Median Household Income. (n.d.). Retrieved from Census.gov website: http://www.census.gov/did/

www/saipe/data/interactive/saipe.html?s\_appName=saipe&map\_yearSelector=2014&map\_geoSelector=aa\_c

[5] Bostick, M. (2016, April 30). Choropleth. Retrieved from https://bl.ocks.org/mbostock/4060606

[6] Bostick, M. (2016, February 8). click-to-zoom via transform. Retrieved from blocks.org website:

https://bl.ocks.org/mbostock/2206590