CS3300 Project 2 Write Up
Jonathan Wu (jzw8), Kyle Albany (kja56), Craig Andres (cra64)
April 28, 2016

For Project 2, our group decided it would be cool to see the correlation between voter demographics and popular vote for the 2012 Presidential Election. We found a data set online that provided the total votes each party received in every county in the United States except Alaska. We wanted to choose a demographic that we thought would influence political affiliations the most. We narrowed it down to income. We found a 2012 census online that provided statistical data on poverty percentages and median household income for every county in the United States. The only statistic we thought would be a consistent correlation was the median household income. After calculating the popular vote, we shaded in red for Republican states and blue for Democratic states and altered the shading depending on how many more votes a party got over the other.

Data

Extracting and Filtering Data

We were lucky to find two very complete data sets on the 2012 Presidential Elections and 2012 Income data. On top of that, they both included FIPS codes for the counties. They came as excel file downloads, so we saved them into 2 separate csv file. We wrote a python script to extract the data we needed from the election file and placed it into a new csv file. The election dataset contained vote numbers for every party out there as well as total votes; however, we decided to only extract the data relevant to the GOP and the Democratic party because those are the two main political parties. We grouped the rest of the parties into an "Other" category.We ended up extracting the FIPS codes, states and their county names, total votes, and all info relevant to the GOP and Democratic party.

We wrote another python script that extracted relevant data from the income dataset and placed it into another csv file. The income data had the state and county FIPS codes separated, so we had to concatenate them and account for the county FIPS codes that didn't have 3 digits. The fields we ended up extracting from this file were the FIPS codes, the states, the county names and the median household income. We wrote another python script that read in both summarized csv files and compared the FIPS codes. The script only extracted the counties that were matched. The fields on this complete summary were states, county names, FIPS codes, GOP voter data, Democratic voter data, total votes and median household income. We put this final summarized csv into another python script that parsed csv files into json files.

2012 Election Data:
http://www.theguardian.com/news/datablog/2012/nov/07/us-2012-election-county-results-download

2012 Income Data:
https://www.census.gov/did/www/saipe/data/statecounty/data/2012.html

Mapping Data to Visual Elements

To represent our data, we use a d3 topographical map of the US (including Hawaii and Alaska). The idea was to show the election results as a color (red for Republican, blue for Democrat), just as most election maps already do. The colors would be a fading gradient from 30% to 90% from white to red/blue depending on the majority percentage of votes for each county/state. Since our data only had county data, to label the states, we just aggregated the county data within each state.

Since color isn't the best at giving specific data, we used the "Zoom and Filter" ideology and had more detailed breakdown of state/county specific data displayed in an information box upon a click on the map. This allowed the user to have certainty in the data they were viewing.

To integrate the average income data, we thought awhile about how to do this. Using more colors or an overlay texture would have made the map appear noisy and difficult to read. Instead, we decided to use a range of income values that the user could specify themselves, and then gray out the counties that didn't have an income that fell in that range. This way, the user could have customized views, and data ranges.

The Story of the Data

Surprisingly, the data shows that there really isn't that big of a correlation between economic classes and political alignment. Part of this might be that the average household income is a large distribution itself, and had we done this on a voter-by-voter basis with their income, we might have seen a more noticeable trend. That said, however, from our chart, we can see that political alignment seems to be almost more based on Geographical placement. The middle of the country is much more aligned with the Republicans, while the East and West coast urban areas are more aligned with the Democrats. Had we had more time, a population density dataset integration might have also made this observation more clear.