

CS 171/CSCI E-64: Visualization

Project 1, Process Book

James Goodspeed – jgoodsp@fas.harvard.edu

Introduction

For project #1 I've chosen to look at how Presidential campaign donations from 2012 relate to per capita income in the state of New Hampshire. A secondary goal to this is to see if there are geographic locations within New Hampshire that are more liberal or conservative based on the cumulative amount the individuals in that town gave to Republican or Democratic candidates in the last Presidential election.

Data Sources

The New Hampshire town information relating to zip codes and latitude and longitude comes from federalgovernmentzipcodes.us.¹ This CSV file contains zip codes for all towns in the United States and territories as well as military zip codes. The file also contained data that was not required for this project such as estimated population and tax returns filed.

The per capita information for the New Hampshire towns comes from the US Census bureau's American Fact Finder page² using a custom search. The format is also CSV and contains many fields that were not relevant to this project, but did have the town name and per capita income for that town.

The campaign finance data was obtained from the New York Times Campaign Finance API, specifically the Presidential State/Zip Totals.³ This data had to be pulled one zip code at a time so doing this for each town in New Hampshire necessitated an automated solution. The format for this data was in XML.

Further on in the process it became clear that having population data would be beneficial. This data was obtained from the New Hampshire Office of Energy and Planning.⁴

Data Cleanup

Once I had the raw data I had to perform cleanup on that data to get the information required in a simplified format from the three different sources. The majority of this cleanup was done using a Python script – generateCSV.py. Please note that this script requires the lxml library.

The raw data files that had to be cleaned up were:

1. **Zip Code/Latitude & Longitude data:** free-zipcode-database-Primary.csv

¹ <http://federalgovernmentzipcodes.us/>

² <http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml>

³ http://developer.nytimes.com/docs/read/campaign_finance_api#h3-pres-state-zip

⁴ <http://www.nh.gov/oep/programs/DataCenter/Population/PopulationEstimates.htm>

CS 171/CSCI E-64: Visualization

Project 1, Process Book

James Goodspeed – jgoodsp@fas.harvard.edu

2. **Per Capita data:** census-percapita.csv
3. **Campaign Donations:** \${zipcode}.xml – one per town in New Hampshire
4. **Population Data:** This was primarily added to the existing data files using Google Refine

The generateCSV.py script performs the following operations:

1. Extracts the town name and per capita information from the census-percapita.csv file
2. At the same time also cleans up the town name so that it is in the same format as the town name in the free-zipcode-database-Primary.csv file (all uppercase).
3. Write the town and per capita information to a new CSV file called perCapita.csv
4. Loop through the free-zipcode-database-Primary.csv file and extract the New Hampshire records.
5. For each of the New Hampshire zip codes contact the NYTimes API and pull the campaign contribution data for that zip code and save the XML file to a xmldata directory.
6. Use xpath (using the lxml library*) to extract and sum the campaign contribution amount for each candidate by party for that zip code/town.
7. Write the following fields to a new file called NHZips.csv:
 - a. zipcode, city, state, latitude, longitude, republican total, democratic total, per capita
 - b. Note: the per capita field is blank and is a place holder at this point

Once I had the NHZips.csv and the perCapita.csv files I had to combine, or merge, them together. To do this I used Google Fusion tables to merge the files on the town name, which was the same in both files.

Once the data was merged there was some additional manual cleanup that had to be done as well:

1. There were a fair amount of zip codes that were not incorporated towns, but were part of a larger town. For example the community of Etna is located in the city of Hanover, West Lebanon is part of Lebanon. These records were combined into one record.
2. Some of the larger cities had more than one zip code (i.e. Concord). These records were also combined into one record for that city or town.
3. There were a handful of towns that did not have per capita data from the US Census Bureau, those records were removed.

Further on in the process I realized that having population data for the towns would be beneficial. This data was merged into my data set using Google Refine.

* The lxml library for Python must be installed to run this script

CS 171/CSCI E-64: Visualization

Project 1, Process Book

James Goodspeed – jgoodsp@fas.harvard.edu

Design Evolution

I started with two visualization ideas (one with the help of Shirley Zhou) for this data. The first was a map of New Hampshire showing the cities shaded according to per capita income. The Republican and Democratic donations would be represented by red and blue circles of varying size depending on the total campaign contribution amount for each party in that town. I also thought that it might make sense to have a color scale going from red to blue, with purple in the middle showing the varying contribution amounts, but ultimately decided against this as it would have made the underlying data less clear. Figure 1 represents the first idea around this visualization.

In the final version of this visualization I split the data out into three different maps as explained in the Visualizations section of this document.

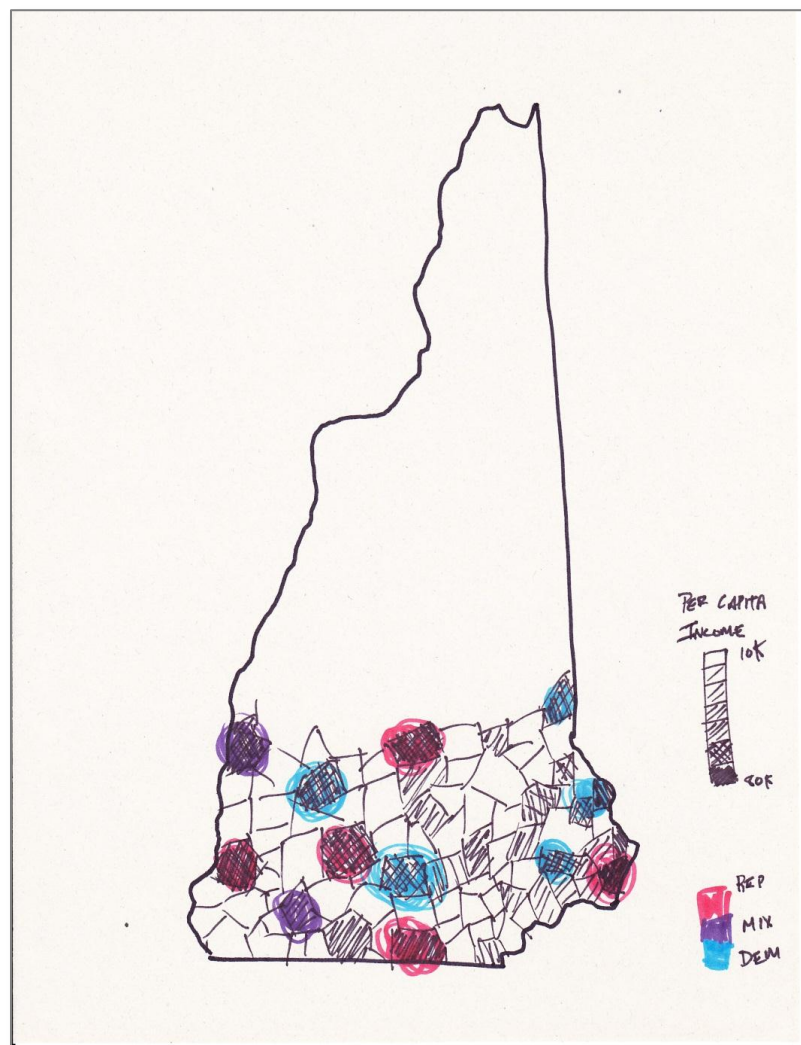


Figure 1: Map of New Hampshire showing per capita data and campaign contribution amounts

CS 171/CSCI E-64: Visualization

Project 1, Process Book

James Goodspeed – jgoodsp@fas.harvard.edu

The second visualization was a scatter plot that could show the correlation between per capita income and what party the people in that town primarily donated to. This could potentially answer the question of whether more wealthy towns gave to one party over another. Figure 2 represents the early ideas with this visualization.

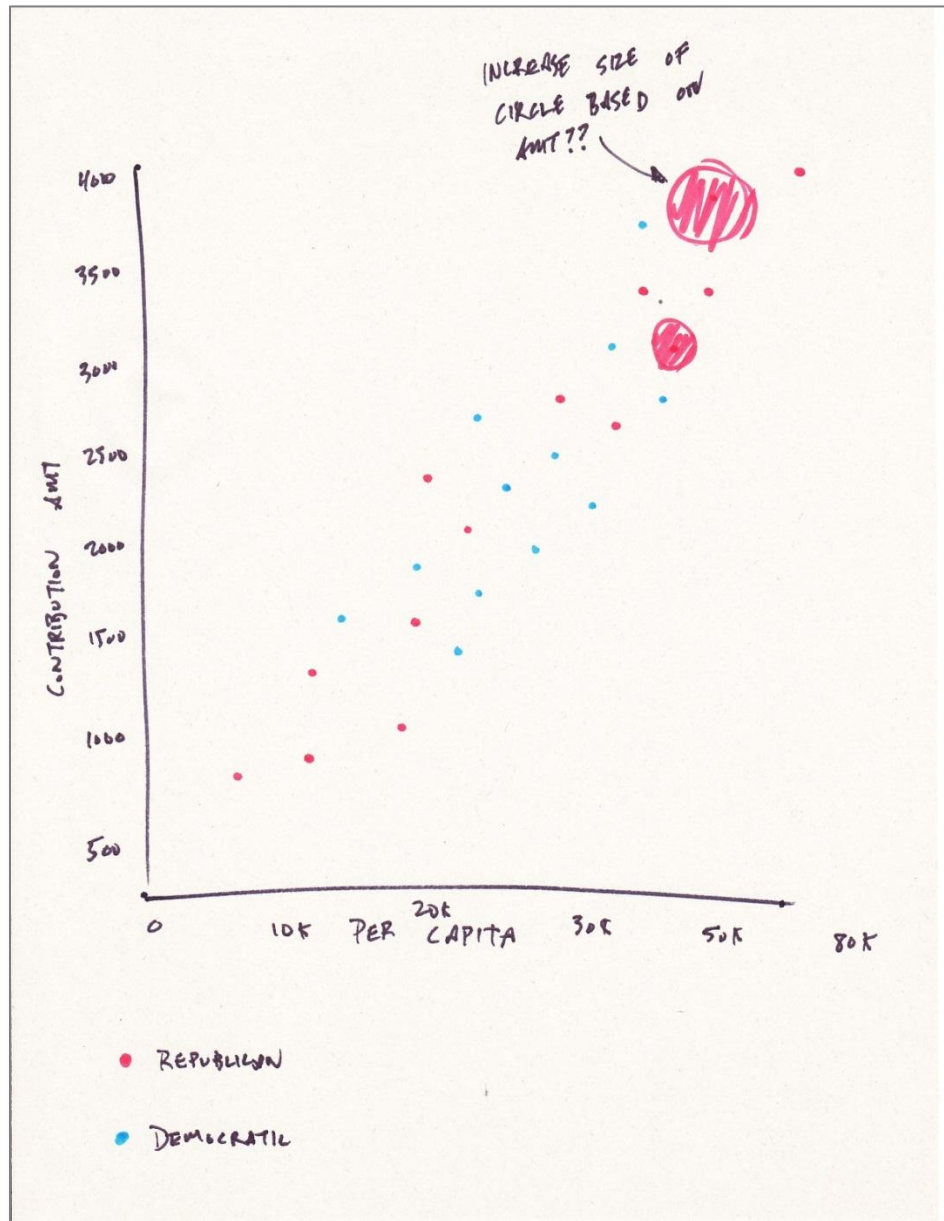


Figure 2: Scatterplot showing the potential correlation between per capita income and Presidential campaign contributions

CS 171/CSCI E-64: Visualization

Project 1, Process Book

James Goodspeed – jgoodsp@fas.harvard.edu

Later in the design process I thought of another visualization to try and gain some data insights. The grouped bar graph below was meant to potentially show a similar correlation to the scatter plot above, but in practice there were too many towns to represent on the real chart in Tableau.

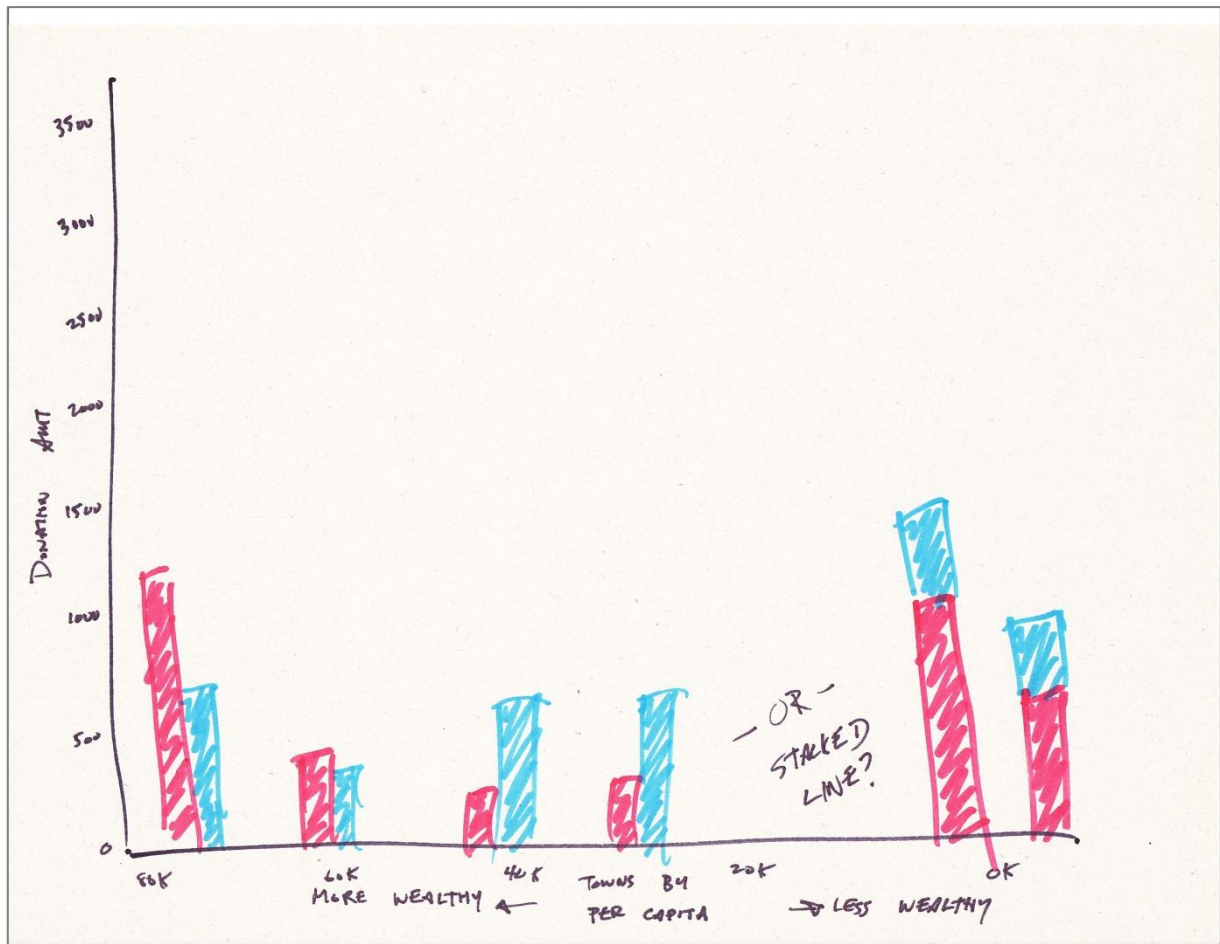


Figure 3: Grouped/Stacked bar chart

Ultimately I went with a combination and variation of the first two visualizations.

James Goodspeed – jgoodsp@fas.harvard.edu

Visualizations

For this project I chose to use Tableau. I briefly explored loading my data set into Many Eyes, however it became clear that while Many Eyes could easily create the scatter plot, Many Eyes could not support the map of New Hampshire I wanted for my first visualization.

The first set of visualizations comprises three different maps of New Hampshire. The first map shows the Democratic contributions represented by blue circles, centered over the towns their data represents with the size of the circle corresponding to the contribution amount. The second map shows the same data encodings for the Republican contributions colored in red.

I ended up breaking these two maps out because I thought the map would become too crowded with data if both parties were represented on a single map. With the data side by side it is easy to make a quick visual comparison. The red and blue colors were chosen due to their traditional representation of the Democratic and Republican parties.

Another important note about the campaign maps is that there is a filter on the maps to only display a circle if there was at least one dollar donated to either party. There were many towns, especially in Northern New Hampshire, that did not donate to either party and adding this filter ensured that a circle did not appear for that town and potentially portray false data to the user.

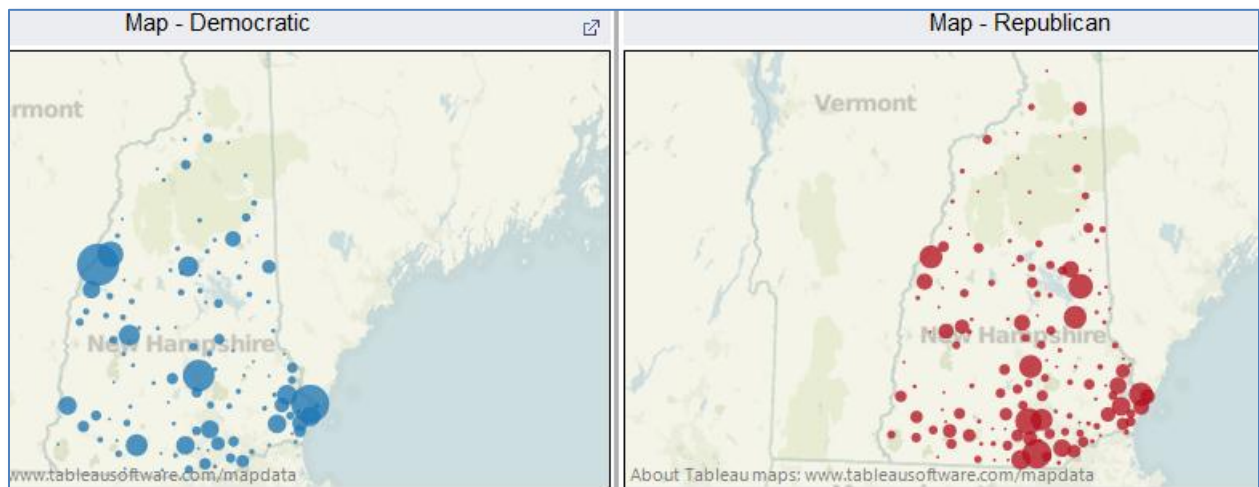


Figure 4: Democratic and Republican Campaign Contribution Amounts by Town

The third map is a gray scale map of New Hampshire representing the per capita income for each town. The darker colors represent a higher per capita income for that particular town. I did experiment with using the same encoding as with the campaign contributions (the bubbles), but with each town having a specific per capita income it was very difficult to see the differences between the towns as shown in Figure 5. This was not a problem on the campaign contribution maps because not all towns had Presidential donations so there were far fewer data points.

CS 171/CSCI E-64: Visualization

Project 1, Process Book

James Goodspeed – jgoodsp@fas.harvard.edu

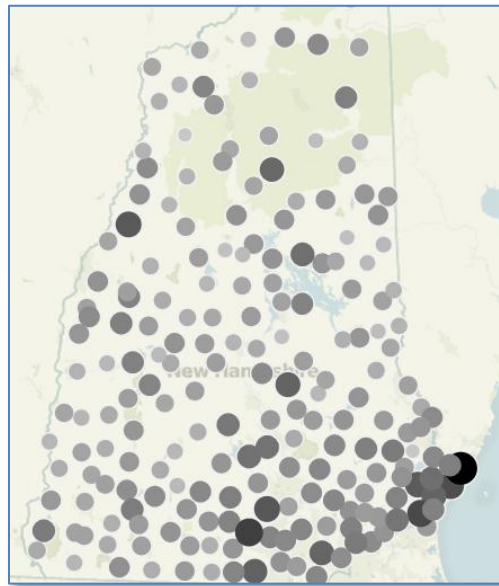


Figure 5: Abandoned Per Capita Map with circles

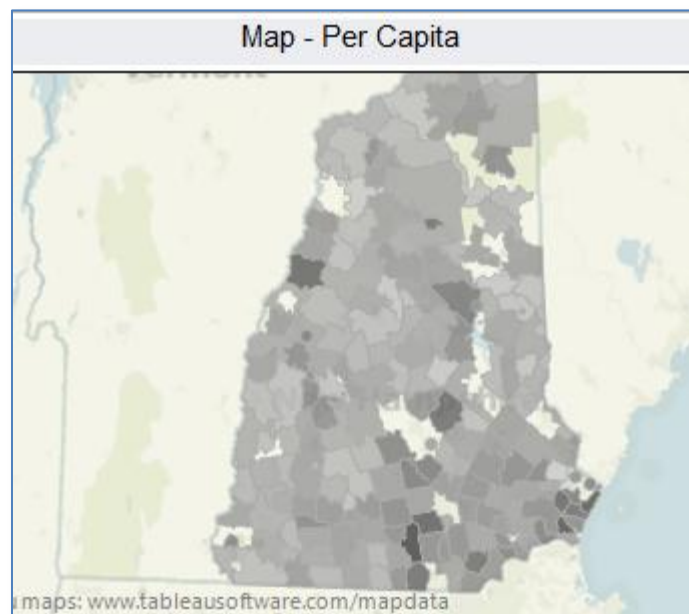


Figure 6: Final version of the per capita map of New Hampshire

The final visualization is a scatter plot that I had hoped would show a correlation between per capita income and party donations for that town. As can be seen below from the chart there isn't a

CS 171/CSCI E-64: Visualization

Project 1, Process Book

James Goodspeed – jgoodsp@fas.harvard.edu

strong correlation based on party. Surprisingly there isn't even a correlation between per capita income and overall amount given. I had expected to see the higher per capita towns have higher overall campaign contributions, but that was not the case for the data I gathered.



Figure 7: Scatter plot of Democratic and Republican Presidential donations by per capita income

As can be seen, the highest overall donation amounts come from cities and towns with a per capita income in the lower half of the state. What this data does not include is population of the cities and towns. The overall contribution amount from the New York Times API is the total amount for that town and is not adjusted for the population of that town. I was able to find population data for each town and add that to my data set and then create an adjusted scatter plot with a calculated field so that the contribution amount was adjusted for the population in each town. Even with this adjustment and the addition of trend lines there does not appear to be a strong correlation to me.

CS 171/CSCI E-64: Visualization

Project 1, Process Book

James Goodspeed – jgoodsp@fas.harvard.edu

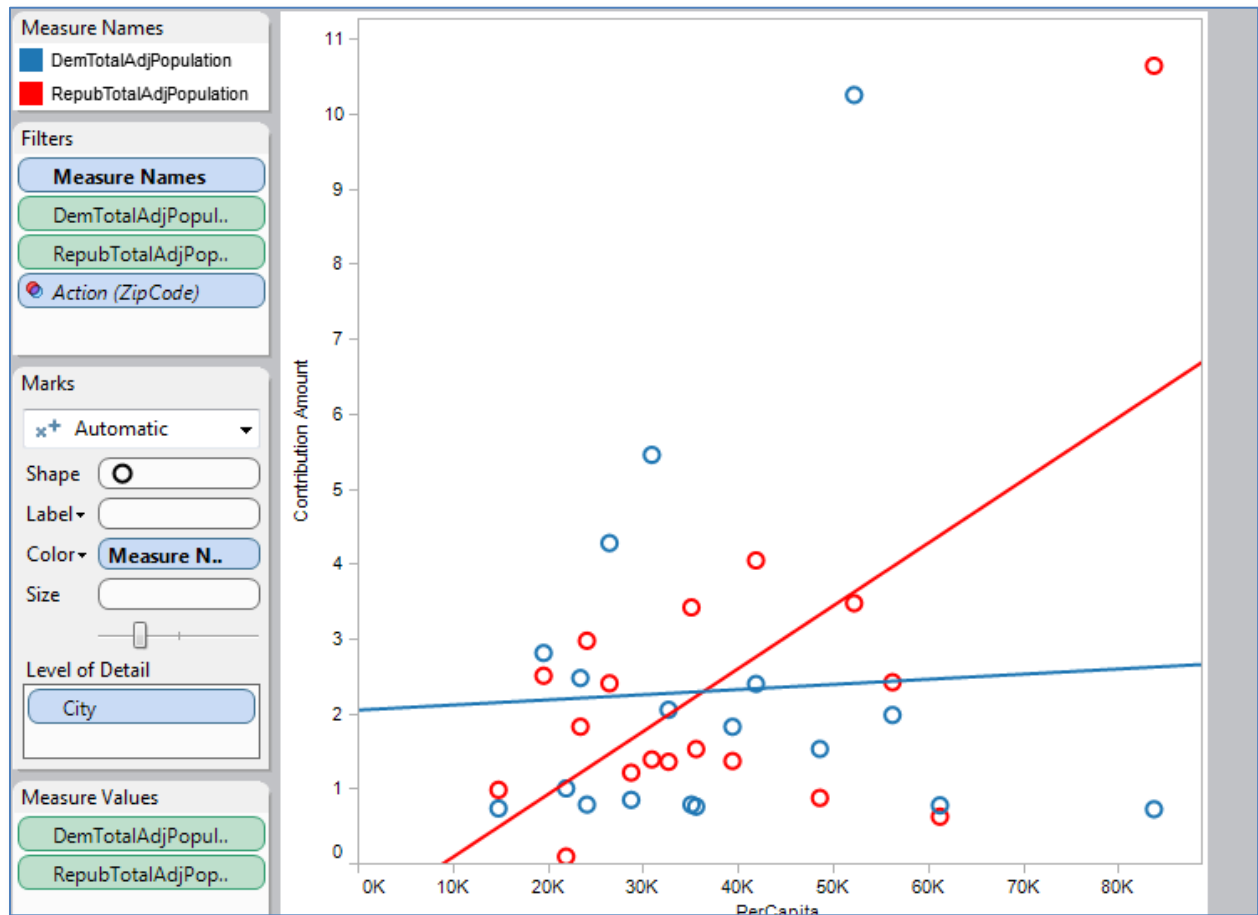


Figure 8: Scatterplot with data points adjusted for population

The final visualization is a dashboard of the three maps and the original scatter plot linked by filtering. Clicking on a data point for in one of the visualizations will filter the results in the other visualization allowing the user to gain insights into a specific town.

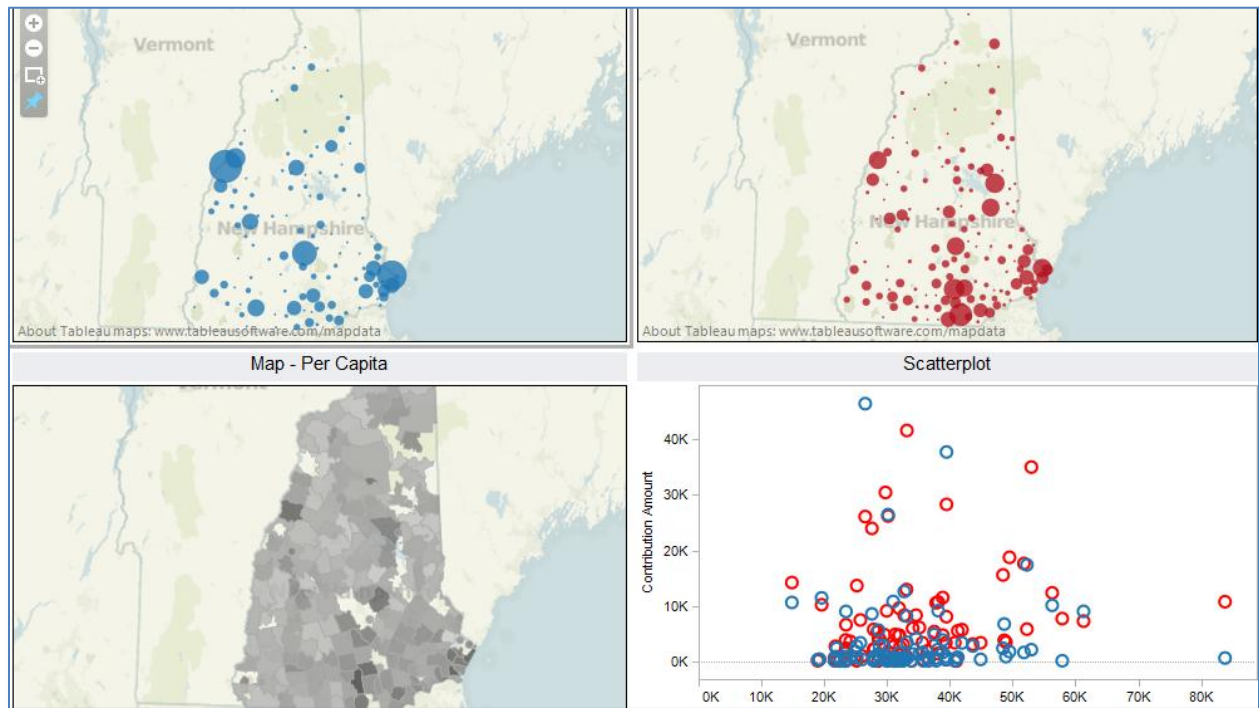


Figure 9: Dashboard view

Analysis of Data

As stated above I was not able to find a strong correlation between per capita income and campaign contribution amounts to a specific party. It is possible that with a larger data set analyzing the entire United States that additional information could be obtained.

The map visualizations did provide some interesting insights into the demographics of New Hampshire. Over the past 10 years New Hampshire has moved from a conservative state to a liberal state and in the election of 2012 was not only solidly for President Obama, but also elected a Democratic governor, two Democratic U.S. Representatives and took back the State House.

Looking at the maps of campaign contributions for New Hampshire there are strong liberal pockets in the seacoast region, near Hanover, home to Dartmouth College, and in the state's capitol in Concord. The Republican map on the other hand is much more evenly spread throughout Southern New Hampshire (the population of Northern New Hampshire is low and did not have high numbers for either party) and did not have strong bases in any particular region. One final interesting note, especially given New Hampshire's move towards the left is that New Hampshire citizens gave twice as much money overall to the Republican Party than they did to the Democratic Party.

CS 171/CSCI E-64: Visualization

Project 1, Process Book

James Goodspeed – jgoodsp@fas.harvard.edu

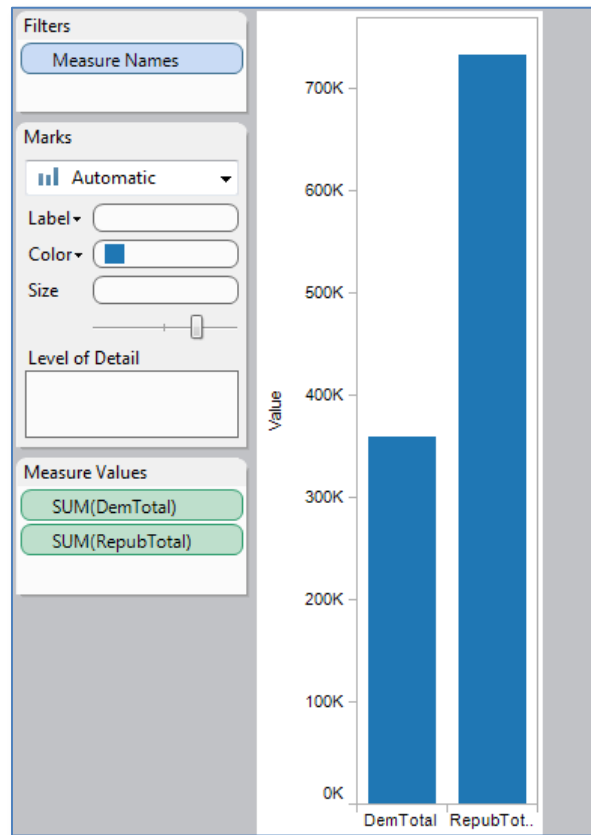


Figure 10: Democratic vs. Republican Donations