**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]](#footnote-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 page[[2]](#footnote-2) using a custom search. The format is also CSV and contains many fields that were not relevant to this project.

Finally the campaign finance data was obtained from the New York Times Campaign Finance API, specifically the Presidential State/Zip Totals.[[3]](#footnote-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.

**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.

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

1. Zip Code/Latitude & Longitude data: free-zipcode-database-Primary.csv
2. Per Capita data: census-percapita.csv
3. Campaign Finance: ${zipcode}.xml – one per town in New Hampshire

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 clean 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 NH 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:
   1. zipcode, city, state, latitude, longitude, republican total, democratic total, per capita
   2. 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 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 Lebannon is part of Lebannon. 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.

**Visualization ideas**

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 data less clear. Figure 1 represents the first idea around this visualization.

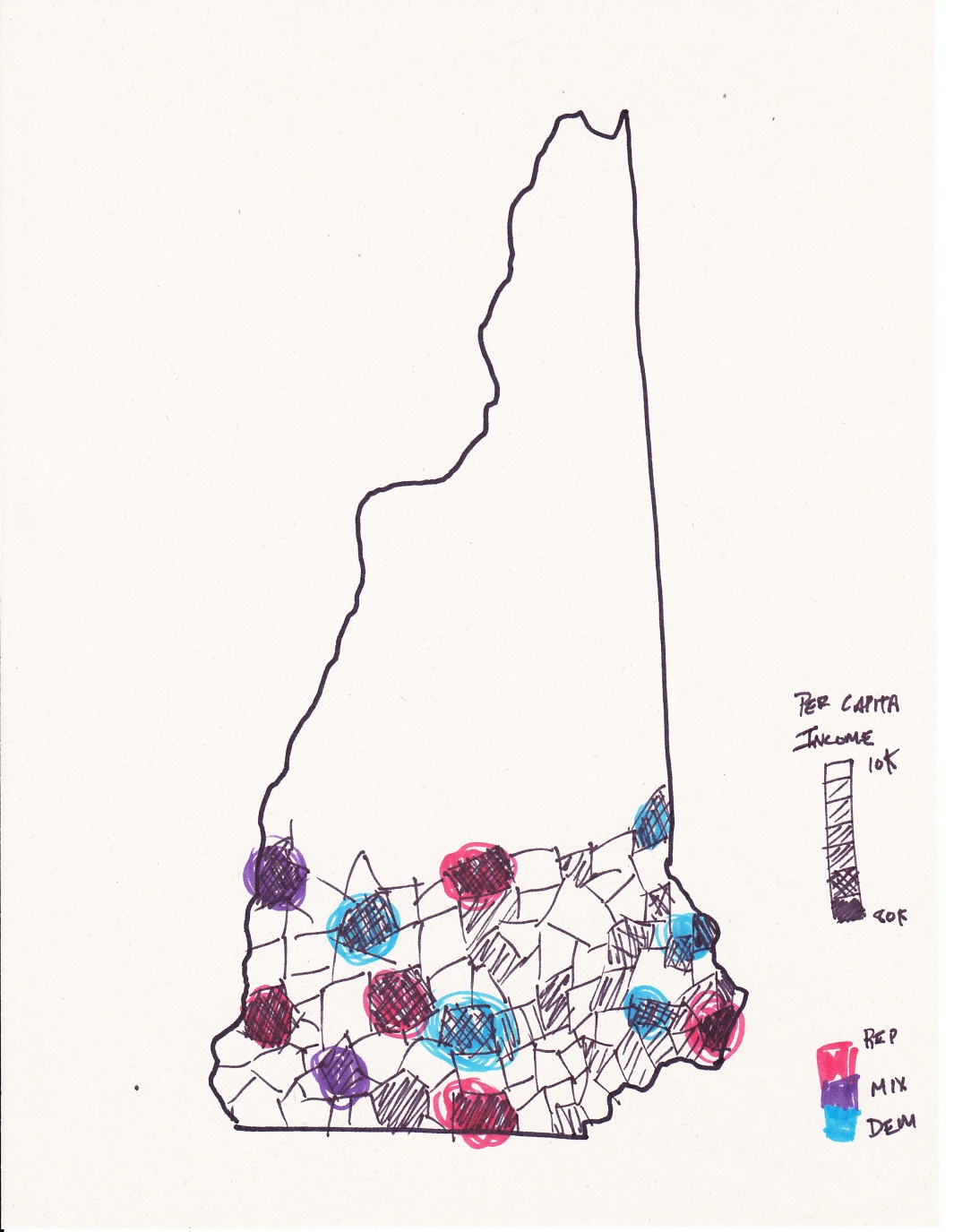


Figure 1: Map of New Hampshire showing per capita data and campaign contribution amounts

The second visualization that occurred to me was a scatter plot that might 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 process around this visualization.

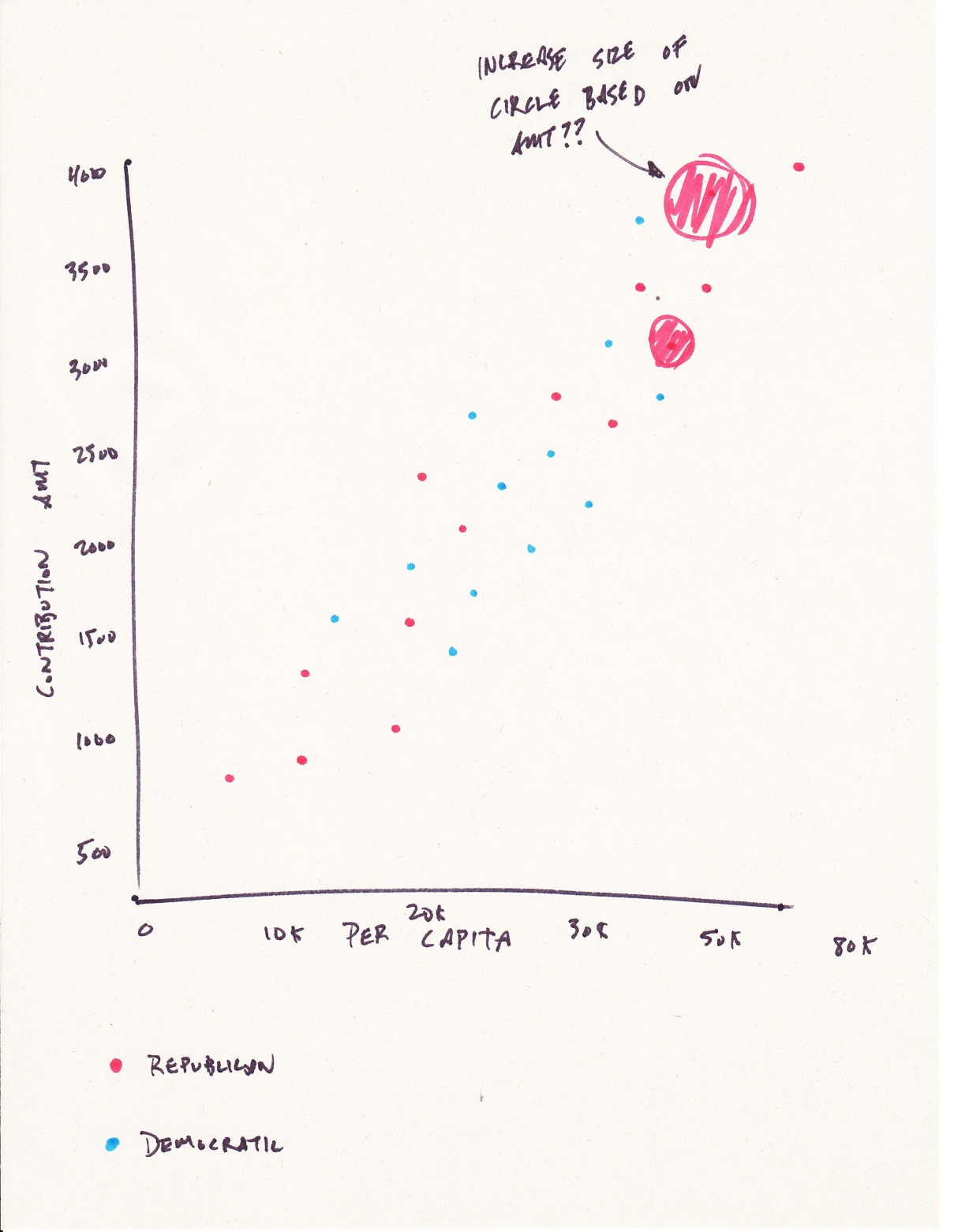


Figure 2: Scatterplot showing the potential correlation between per capita income and Presidential campaign contributions

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.

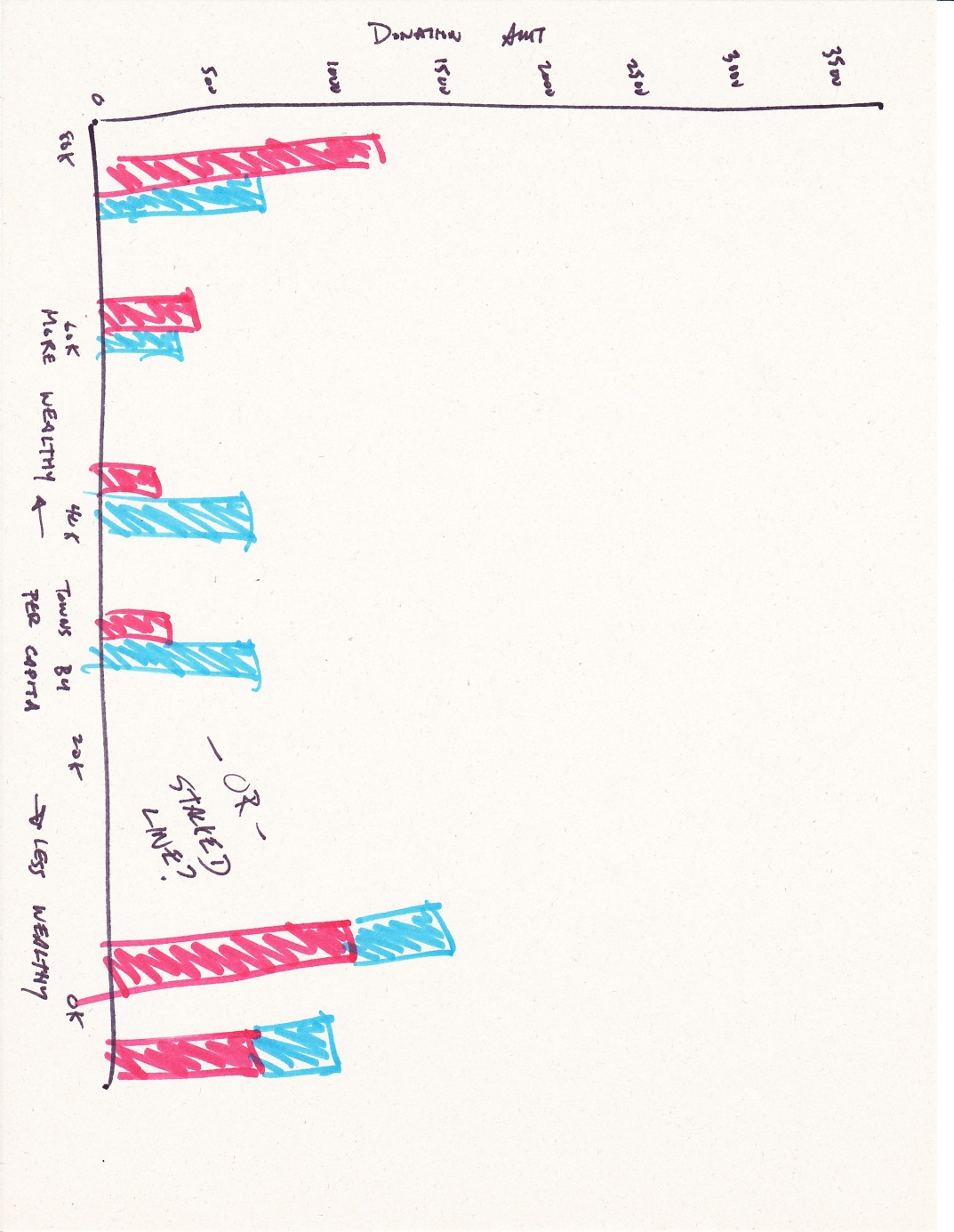


Figure 3: Grouped/Stacked bar chart

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

1. http://federalgovernmentzipcodes.us/ [↑](#footnote-ref-1)
2. http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml [↑](#footnote-ref-2)
3. http://developer.nytimes.com/docs/read/campaign\_finance\_api#h3-pres-state-zip [↑](#footnote-ref-3)