Assignment 2

Allen Chen (allen.sl.chen@gmail.com)

March 8, 2012

Using 2 late days; 2 late days left.

1   Data Preprocessing

To preprocess my data, I had to perform a number of tasks for each visualization to coerce it into the nice JSON format that each visualization required.

1.1   County per-capita Donation Amount Map

The map required the most preprocessing. The map actually combines three different datasets: the donation datasets, a ZIP code-to-FIPS code (county ID) dataset[1], and FIPS-to-population dataset[2]. First, I had to process all ZIP code data from the donation dataset and translate them to FIPS codes. Then, I took the donation dataset and bucketized each total donation count by FIPS code, and divided each bucket by the FIPS population. The result was a JSON object which was keyed by FIPS code and had the value of average per-capita donation amount.

Some preprocessing was done for the diff view as well (though it could have been implemented on the client side). I took each per-capita donation amount for Republican and Democrat, and subtracted them. I normalized the data to [0,5], which allowed me to scale the data between red and blue (2.5 was a indifferent color, whereas 0 was dark red and 5 was a dark blue).

1.2   Donator Occupations

For the donator occupations, I keyed a hash on occupation and populated the hash with occupation data from the donation dataset. Then, I took the top 50 occupations from each of these datasets and output them in a JSON format for the Bubble visualization. The radius is a function of the **square root** of the actual occupation count, since without the square root, the graph would be unreadable (Retired would dominate the graph).

1.3   Donations over Time

I decided that, for donations over time, each bar should represent the cumulative donations over a 7-day period to give decent granularity while maintaining some sort of readibility on the graph. I bucketized each donation into a 7-day period and added them up.

1.4   Individual Donation Amounts

The problem initially with individual donation amounts was that the dataset was huge - it basically required passing a significant portion of the full donation dataset (all of the donation amounts). To keep the amount of data passed relatively low, I instead passed a subsample (a randomly-selected subset that was 10% the size of the original dataset).

1.5   Overall Donations Per Party & Candidate Donations

These two visualizations essentially required the same dataset. I preprocessed the data by bucketizing each total candidate donations by party, and passed it to the client. With the bar graph, there were a few extra data points I needed to pass (mean value of Republican, Democrat and Libertarian donations).

---

[1] http://mcdc2.missouri.edu/pub/data/georef/zcta_master.Metadata.html
[2] http://www.census.gov/geo/www/2010census/centerpop2010/county/countycenters.html

2   Client-Side Data Transformations

The client performs very little data transformation for most charts.

For all charts, the client needs to normalize the data in order to fit the data within the appropriate bounds. For the **Donations over Time histogram** (stacked view), the client needs to calculate the heights for the combined histogram. For the **Individual Donation Amounts boxplot**, the client needs to perform analysis to find the information for box plot. For the **Overall Donations per Party pie chart**, the client needs to transform the individual (annotated) candidate donation amounts to calculate the total party donation amount.

3   Visualization Intent

3.1   County per-capita Donation Amount Map

This map portrays the average donation bias per county. I want to show what areas of the country are most enthusiastic/politically-active and which way they lean. Another way of displaying this information is a geographically-positioned bubble chart (size of bubble corresponding to per-capita donation amount). To effictively display this geographic data, a map is required.

3.2   Donator Occupations

I wanted to show what kind of people (occupation-wise) are most influential in the donations to different parties of candidates. A bubble chart was used because I wanted to display occupation frequencies and compare them to other occupations very quickly. A histogram was considered, but since there are so many data points, it would require too much space to display this data.

3.3   Donations over Time

The purpose of this visualization is to show what times of year the political fundraising campaigns are most active. With this method, we can easily compare the total amount of donations to all campaigns (compared to a line graph, where we can't easily have a stacked view to see the total amount per party).

3.4   Individual Donation Amounts Box Plot

The box plot shows the typical (median) donation from a donator, and emphasizes the vast difference between a typical donator and a major sponsor (Republicans: $250 small donators vs. $1,000,000+ from major sponsors). A box plot because it conveys the most important details (median, 1st/3rd quartile and outliers) for the question: How much money is each person donating?

3.5   Overall Donations per Party Pie Chart

From the second level, we can see the distribution of the donations amongst Republican candidates, and we can see that the donations is extremely skewed. I chose this method to display the overall donations because I wanted a direct comparison that showed what percentage of donations go to Republicans, Democrats and Libertarians (and corresponding candidates). It very quickly answers the question: Which party has more donations? About what percentage of total donations does a particular party have?

3.6   Candidate Donations Bullet Chart

The candidate donations bullet chart shows the exact donation amounts per candidate and the average donation amount per party (the black line that is common amongst all candidates is the mean value of the Republican donations). We can see that there are three Republican candidates that are considerably ahead of that mean. I chose this method to portray the disparity in donation amounts for Republican candidates. It answers the questions: Who has the highest Republican donation amount? Who is "winning" in donations amongst Republicans?