

## Report

### 1. What kind of data pre-processing have you done on the server side?

Preprocessing includes taking the data and transforming it into 5 different short csv files; donations by state, donations by candidate, donations by date, donations by size (under and over \$75), and number of contributors by state. This is done by running `preprocessor.py`, which creates the 5 csv files from an input data file (`data.txt`) through judicious use of the `groupby` function, a smaller version of `P00000001-ALL.txt`

### 2. What transformations on the data do the client-side d3 apply to generate the plots?

For `amtbystate.js`, `bubble.js`, and `stateDonsByIntensity.js`, we take our input csv of states and drop all abbreviations that are not in the 50 US states; `stateDonsByIntensity.js` also takes in an input CSV of population of each state. For the others, we just map `parseFloat` over the contribution amounts to get arrays of numbers.

### 3. What is the intent of each visualization? i.e., what information do you want to present to the viewer?

`amtbystate.js`: The purpose of this visualization is to quickly inform the viewer which states are most influential in contributing to federal campaigns, giving information about where candidates should focus their campaigning in order to raise the most amount of money. Hovering over the bar will show the viewer exactly how much.

`bubble.js`: The purpose of this visualization is to show the viewer the projective relative voting power of each state against each other, based on the data we have, which is federal campaign contribution data. It shows where the most number of people have donated, but is not necessarily correlated with the highest total donations. Hovering over the bubble will show the viewer exactly how many people have donated.

`stateDonsByIntensity.js`: This allows the viewer to see at a glance donation amounts per capita geographically, possibly allowing the viewer to draw conclusions about the fundraising potential of geographic regions, such as coastlines or groups of states. Hovering over the state will show exactly how much money was raised.

`donsByCand.js`: This visualization shows the viewer at a quick glance which candidates are leading and by how much relative to other candidates. Hovering over the pie chart will show exactly how much money was raised by a particular candidate.

`donsByDate.js`: This visualization shows the trend of donations over time, allowing the viewer to draw conclusions about when the best time to be fundraising would be or draw conclusions about when candidates most care about money.

donsBySize.js: This shows the viewer a rough proportion of how popular each party is with people who donate less money and people who donate more money, very roughly dividing the data into popularity by income. Mousing over the data will cause it to change to the other party.

4. What are the right visualization techniques to present this information effectively? (e.g., which type of chart or graph is used? why?)

amtbystate.js: A bar chart is used because outliers can be easily seen, so the states with greatest fundraising potential are highlighted. Also, it is easy to see exactly how much more money is raised by some states over others.

bubble.js: A bubble chart is used because all that we care about is the relative proportion of number of donors; if a viewer really cares about the exact number, they can hover over. A bar chart might have worked just as well; however, the downside to a bar chart is that it does not grab the viewer's attention as much as a more exotic form of data visualization.

stateDonsByIntensity.js: A geographic map was used because the intent of the visualization was to allow the viewer to see the fundraising potential of not just individual states, but regions as a whole. Thus, a map where color intensity is used to represent total funds per capita raised fits our needs very well.

donsByCand.js: A pie chart is used because it is very easy to see who is ahead and by how much, percentagewise. In addition, it effectively communicates the most important data, giving the less important data (candidates we don't care about) less space rather than allocating an equal amount of space for each value no matter how irrelevant.

donsByDate.js: A scatter plot chart was used to try to show the trend of the data over time without having distracting lines like in a line graph.

donsBySize.js: A bar chart is used because it effectively shows the proportion of both large to small donations and Republican to Democrat donations. The color is also used to indicate to the viewer which party is being seen currently.