

## FEC Data - Further Analysis

In the previous [post](#) we showed how Federal Election Commission data could be loaded into [Neo4J](#) and manipulated using [Gremlin](#). In this follow-up posting, we'll modify the data structure and do some further analysis of the data.

### The FEC Data Graph

The FEC data is represented in the following graph. Each committee supports a candidate. Some candidates may be independent from a committee. Individuals contribute 1 or more times to a committee. For this demonstration, we've haven't separated out city/state/zip and created a common location.

A couple of notes on the data. Some of the committees did not have a treasurer so I added in a value of "No Treasurer". Some of the candidates were referenced to non-existent committees. In this case, I've created entries for those committees in order to load the data and create the links. Additionally, the individual contribution file has [overpunch characters](#) to different amounts or negative amounts. Those values were adjusted in the database so the data could be loaded as an integer value.

In this design, we see that a contributor (individual making a contribution) can make several contributions over time. These contributions are given to a committee in support of a candidate. Additionally, we've added in an additional data set that is a summary of all contributions where detailed donors reporting is not required because they have not given more than \$200.

Just to give people an idea of the volume of contributions, in September when I downloaded the data, there were 437,726 contributions. When I downloaded the latest file on November 3, there were now 598,306 contributions. That's about a 36 percent increase.

### Let's look at a candidate

We'll use the gremlin language to perform some graph manipulation to analyze giving to some candidates. For this case, I've decided to look at Mr. Herman Cain.

Using the Gremlin pipes (for a great description, see this [post](#)). To check on the committee supporting Herman Cain, we run this command:

In this instance, there is only a single committee, FRIENDS OF HERMAN CAIN INC, that is supporting Herman Cain.

Our pipe looks like this:

### Campaign Contributions

Let's take a look and see how many campaign contributions have been made to Herman Cain. The data is as of mid-September. I haven't downloaded an updated data set.

We see that the FRIENDS OF HERMAN CAIN INC have received 1201 individual contributions. The average contribution is determined using the following Gremlin command:

In the next analysis, we'll use some filter steps to remove objects from the flow of computation. In this example, we see that there are three contributions above \$4500.

To see who the people are, we'll use a more complicated pipe that starts with the committee, filters out the contributions that are greater than \$4500 and then passes those results as input to find out who made those contributions.

Our next analysis will be to see who is contributing multiple times to the Cain campaign. We use the following code to see who has contributed. In order to remove the reflexive path, we add in the filter. Without the filter, we would see double the amount of contributions.

Who is Julie Waddle? Using additional commands, we find out she is a homemaker from Herman Cain's hometown.

To show the contributions, we will run this command:

Let's do a little more digging and see what the top reported occupations were that contributed to the Cain campaign. We'll use the following command and get the following results which show retirees, homemakers, physicians and lawyers are the top contributors:

**Next Steps**

The next steps will be to reload the graph with updated data and look at different groupings of data (occupation, location, time series, etc).