Sarita Yardi <yardi@sims>
Steve Chan <sychan@sims>

Money Can't Buy You Love…

…But Can It Buy You a Public Office???

The Data

From the description of the original assignment:

> *The data set being analyzed is a financial summary of U.S. campaign finance contributions and spending for each 2-year Congressional election cycle between 1993 and 2002. These data sets are published by the United States Federal Election Commission, and include summary information for every candidate for Congressional office (both the Senate and House of Representatives). The data was downloaded from: http://www.fec.gov/finance/disclosure/ftpsum.shtml*
>
> *Each row of data represents a single candidate running for office in the given year (years are listed using the last year of the election cycle), and contains information about contributions, expenditures, party affiliation, state, congressional district (or none for the Senate), and outcome (win/loss/runoff). The data set is fairly large, containing over 9000 rows. The only difference between the data set being given to you and the ones on the FEC website is that we have concatenated data for multiple election cycles and added a year column, enabling any number of trend analyses. As you'll undoubtedly notice, the data is highly multidimensional, with a large number of columns. The FEC website includes a detailed description of each column.*

But – you already knew that! A few observations that we have are:

- The data seems to be somewhat "dirty"
  - The Total (contributions) $200-$499 field ranges from -$249457 to $2480345
  - The Total Indiv Contrib field ranges from -$72816 to $35885552
  - It isn't clear how people can make negative contributions (unless they contribute a bill or something)
- An alternative explanation is that the data is correct, but we misunderstand the semantics of the different fields. But the description the FEC site seems to be very straightforward for the Total $200-$499 field:
  - *This amount represents the sum of all individual contributions received by the campaign in amounts between $200 and $499.*
- When we tried to sum the values for the different subcategories of individual contributions, and compare them to the Total Indiv Contrib to look at proportions of large vs medium vs small contributions, the values would not reconcile into sensible values (often resulting some categories of contributions with negative values)

Sarita Yardi <yardi@sims>
Steve Chan <sychan@sims>

- Because we either do not understand the semantics of the fields, or the data is inconsistent/dirty there were some hypothesis that we originally considered testing, but found to require too much massaging of the original data. As an example, the large vs. medium vs. small individual contributions mentioned above had to be abandoned.

Tools Used

Spotfire
Tableau 1.5

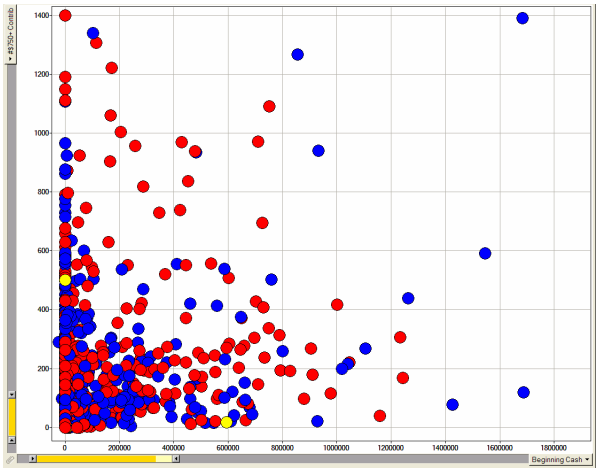
The Hypotheses

Hypothesis #1
Candidates with the highest beginning cash amounts will also have higher greater numbers of contributions over $750

"Are candidates actually able to "buy" money?"

Furthermore, we would speculate that the Republican Party sees a higher occurrence of this correlation than the Democratic Party, given the differences in trends and wealth patterns between the two parties.

We first did a Scatter Plot of #$750+ Contributions versus Beginning Cash with red mapping to the Republic Party and blue mapping to the Democratic (see Figure 1). While it appeared that there were more red circles at the upper end of both axes (more beginning money and more numbers of donations), the difference was not profound and there was not an obvious correlation between Beginning Cash and #$750+ Contributions. A large part of this could be because many candidates had $0 beginning cash, so testing the hypothesis became more challenging.



However, plotting Beginning Cash as a log scale made the visualization much more interesting and informative (see Figure 2). We scaled it to ignore the candidates with $0 beginning cash and zoomed in on the curve of interest. As Beginning Cash values increased (on a log scale) the # of 750+ also increased *proportionately*.

Figure 1: #$750+ versus Beginning Cash on a Normal Scale in Spotfire

Sarita Yardi <yardi@sims>
Steve Chan <sychan@sims>

This suggests that those candidates with more starting cash do, in fact, attract more donations whether it is because they are in correlating demographic environments or because they are more resources to "buy" their donations.

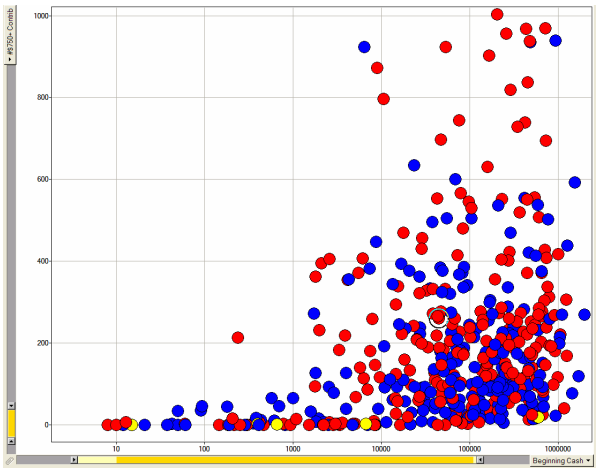


Figure 2: #$750+ versus Beginning Cash on a Log Scale in Spotfire

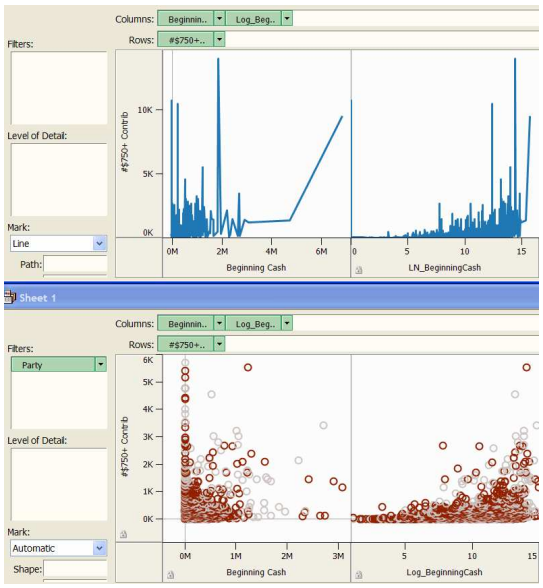


Figure 3: A visual comparison of #$750+ versus Beginning Cash on Normal and Log scales using circles and lines in Tableau

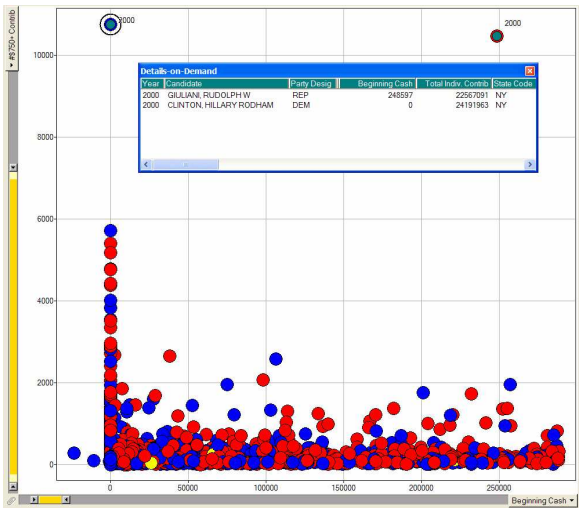


Figure 4: Hillary VERSUS Guiliani! This is a gratuitous but entertaining image demonstrating the importance of some statistical outliers. It has nothing to do with our hypothesis but it pretty solidly supports our overall theme of money and power.

Sarita Yardi <yardi@sims>
Steve Chan <sychan@sims>

Hypothesis #2
Greater expenditures will strongly correlate with winning an election.

"Can the election be bought?"

Some of the data is a little bit quirky! There are big "W" values and little "w" values to represent wins and they are separated into unique columns on the bar chart which is clearly a mistake because they should all be representing a single "Win" column! The same problem occurs in few select cases for "L" versus "l".

One of the challenges with testing this hypothesis was that the dataset did not provide a single sum expenditure value for me to easily chart expenditures versus wins and losses. Instead, we looked at the various relationships between different expenditure breakdowns, including:

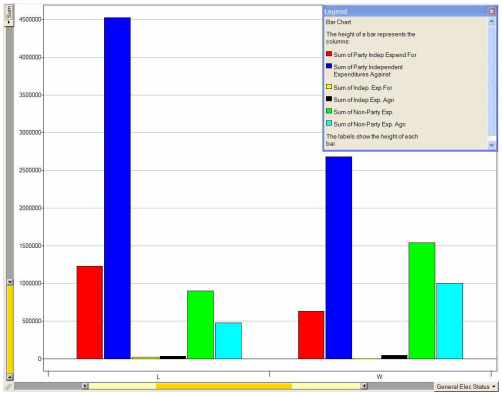


Figure 5

Non-Party Expenditures For
Non-Party Expenditures Against
Independent Expenditure For
Independent Expenditure Against
Party Independent Expenditures For
Party Independent Expenditures Against
Party Coordinated Expenditures
(see Figure 5)

This was particularly challenging because we didn't have a good grasp of the implications of each of these different categories so testing them involving a lot of trial and error testing. We also weren't sure if we might find more interesting relationship looking at the Count, Sum, or Average of these datasets.

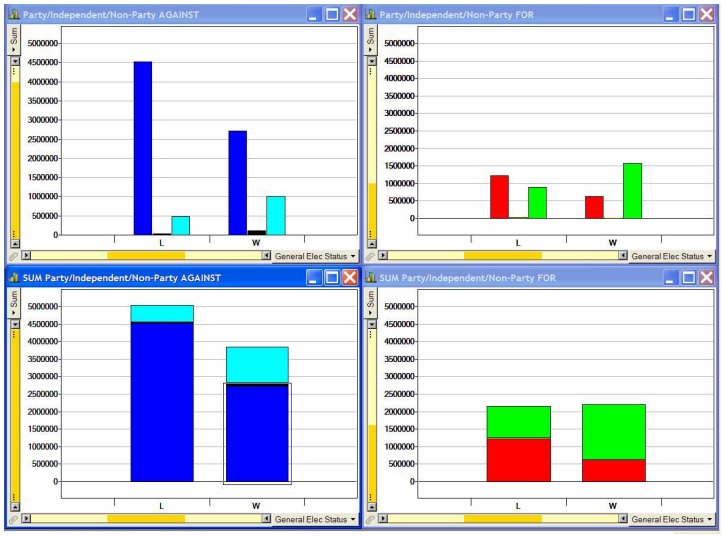


Figure 6

It turned out that the best way to visualize the information was to display it in four different graphs (see Figures 6 and 7).

The interesting relationships, to us, were not in the differences between party, independent, and non-party contributions, but rather the total of these combined and how that influenced overall results. In particular, the sum of the three expenditure types FOR the candidates were similar whether they won or lost. However, the sum of the three expenditure types AGAINST the candidates showed a stronger correlation. The more money spent AGAINST a candidate, the more likely they were to lose.

Sarita Yardi <yardi@sims>
Steve Chan <sychan@sims>

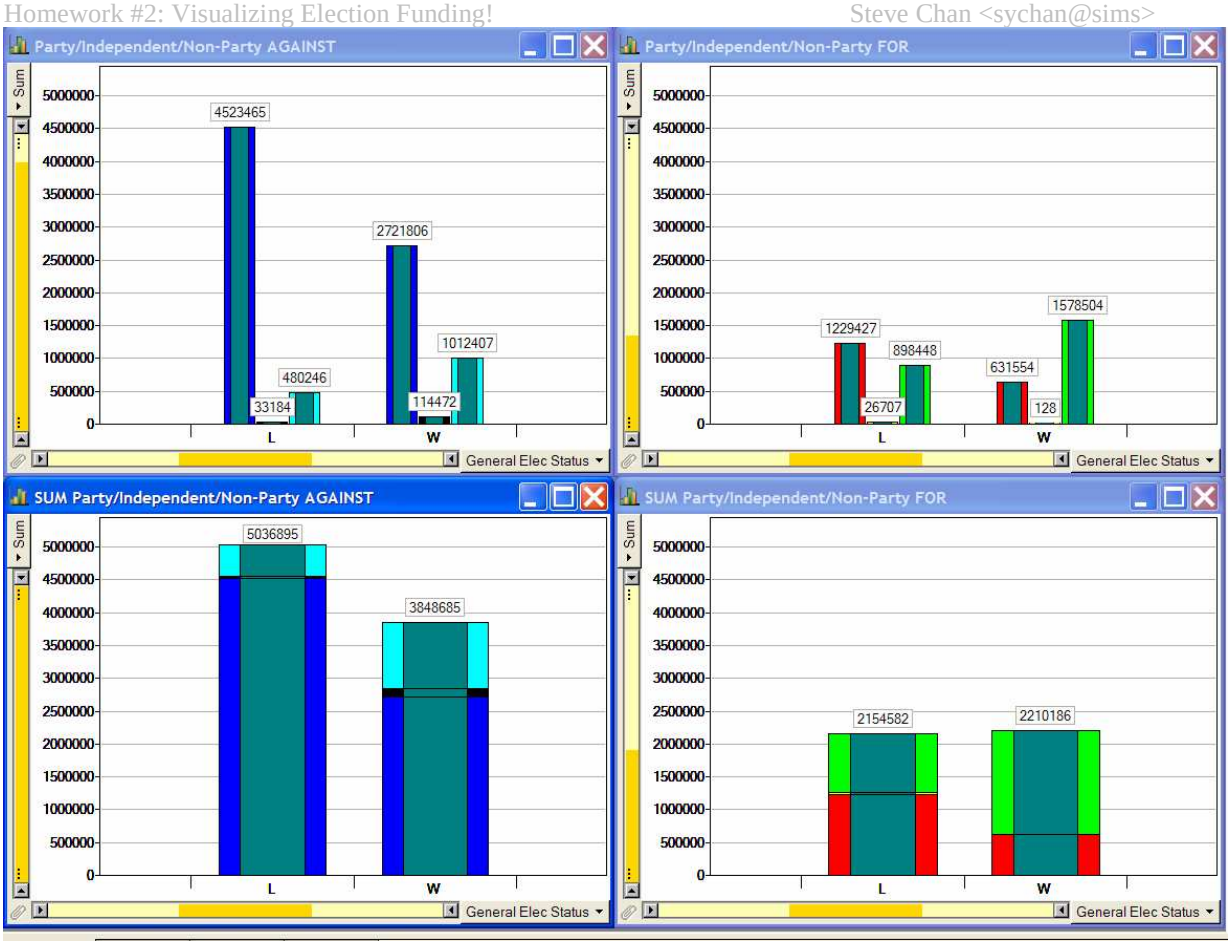


Figure 7: Using Sums to compare correlation between Wins and Losses and Expenditures For and Against Candidates

Since the total values spent FOR and AGAINST were very different, we also compared the amounts of money spent to campaign ***against*** winning and losing candidates, respectively. Losing candidates had a total of $5,036,895 spent AGAINST them, while winning candidates only had a total of $3,848,685 spent to against them. Winning candidates faced only 76.4% of the amount expended against them as those who lost. On the other hand, a total of $2,154,582 was spent FOR candidates who lost and $2,210,186 was spent on candidates who won. The losing candidates faced only 2.5% less money spent on them than the winning candidates. This difference could easily be small enough to be negligible, although we don't know that for sure. However, given that the order of difference between expenditures FOR and AGAINST is 10, which is very large for a data set like this, we would conclude that expenditures against candidates are likely an indicator for winning.

One of the most frustrating challenges we faced when using Tableau was that we wanted to put Win and Loss on the y axis and display the different graphs vertically but we couldn't figure out how to do that! We searched through the menu for ways to change the layout, which was a feature we had found in Spotfire, but we couldn't find it in Tableau. We tried swapping rows and columns, which aligned them vertically but then put Win and Loss on the x axis which defeated the purpose of what we was trying to accomplish (see Figure 8). Similarly, we tried manually switching the categories by dragging the row

Sarita Yardi <yardi@sims>
Steve Chan <sychan@sims>

categories into the column categories and vice versa but that had the same result as just swapping them. We didn't find a solution to this problem. It's likely one exists and we just never found it, but we spent more time than we felt we should have looking for the solution.

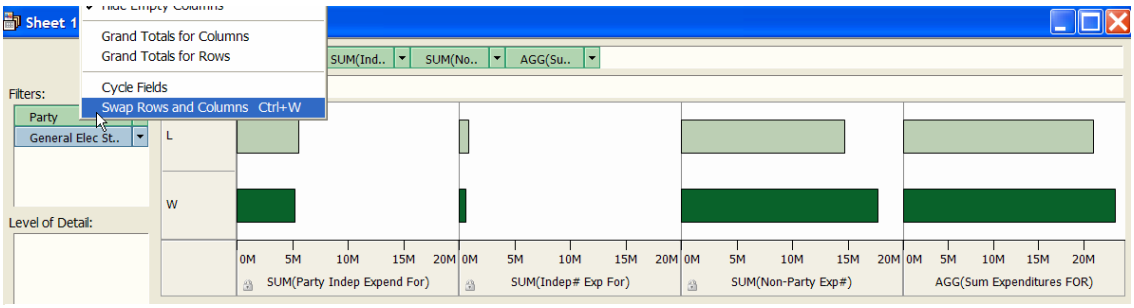


Figure 8: Using Tableau to swap rows and columns (unsuccessfully)

Because we couldn't figure out how to display the information in the way we wanted, we ended up displaying it in a different way, which turned out to be equally interesting because it caused me to develop an interesting comparison between the total amounts that candidates chose to spend FOR their own campaign versus AGAINST others (see Figure 9). Overall, the total expenditures FOR more than doubled that of expenditures AGAINST. However, this was in large part because the Non-Party expenditures also reflected this same trend and there were far more Non-Party expenditures than Party Independent or Independent expenditures.

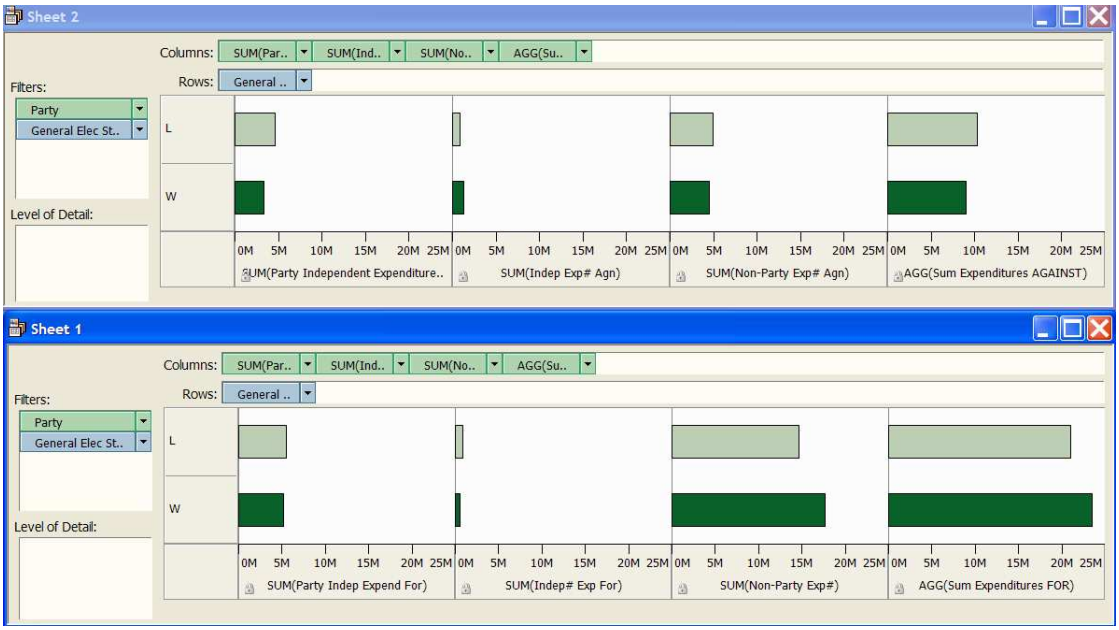


Figure 9: Aligning graphs in Tableau to compare total expenditures For and Against candidates

By lining the graphs up visually the way we did, we were able to quickly make this comparison. It isn't directly related to our original hypothesis, but it did provide a compelling example of how important a role the positioning and layout of the graphs plays in how the data will be interpreted.

Sarita Yardi <yardi@sims>
Steve Chan <sychan@sims>

Hypothesis #3

Our third hypothesis is that incumbents have an advantage in raising money. There are many reasons why we may believe this to be true:

1) Incumbents are demonstrably more effective at campaigning, having already won their office
2) Incumbents are a known quantity to campaign donors, having established a track record already by their time in office, so people "know what they are buying" (perhaps because they've already bought some before!)
3) There may be some other biases in the system that allow incumbents to raise money more effectively.

Our first analysis only goes so far as to see if there is a difference in the amounts of money that incumbents and challengers are able to raise. We use a simple bar chart to compare the average total receipts of challengers versus incumbents versus open seats.

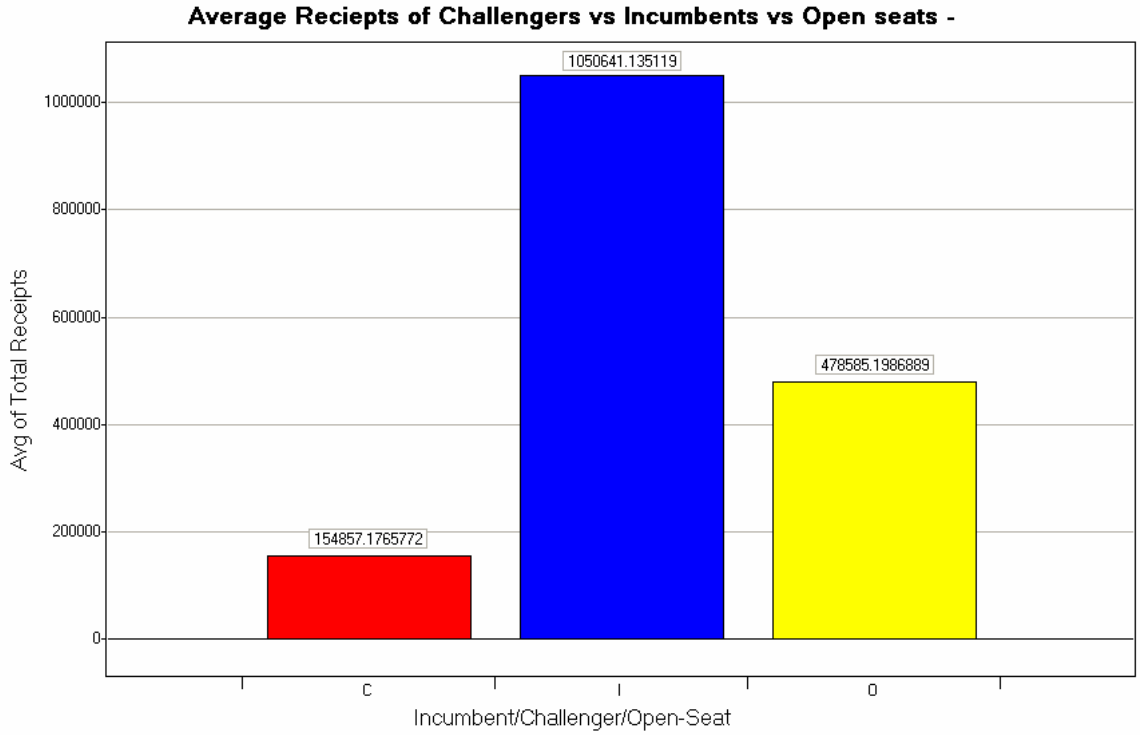


Figure 10: Money raised compared to position as challenger, incumbent or for an open seat

What we see is an almost 7:1 difference between the average amount of money an incumbent raises versus what a challenger raises. We also see that for open seats, there is significantly more money raised (roughly 3:1) compared to challengers. Could this be related somehow to the number of challengers versus incumbents in the dataset? A very quick chart shows the counts for each type of entry.

Sarita Yardi <yardi@sims>
Steve Chan <sychan@sims>

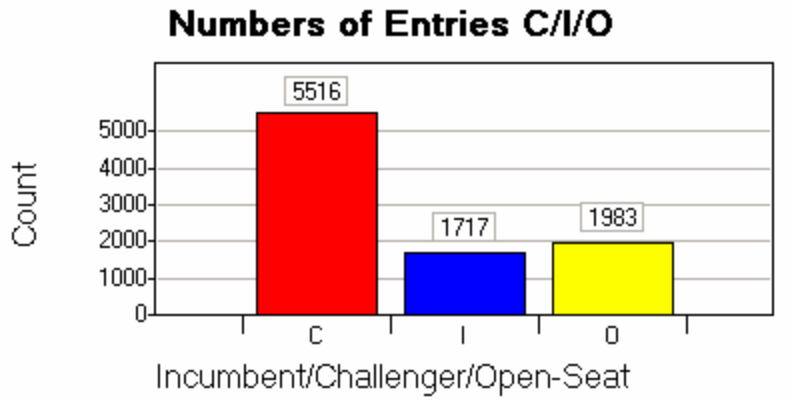


Figure 11: Numbers of each type of entry - challengers, incumbents, open seats

We see that there is a roughly 3:1 ratio between challengers and incumbents. It isn't clear what this tells us – we might speculate that multiple challengers are competing for the resources of those who are opposed to the incumbent. But with a 3:1 ratio between the number of challengers versus incumbents, and a 1:7 ratio between the average amounts raised by challengers versus incumbents, it seems that incumbents still have access to more money than all the challengers combined. An examination of the sources of contributions broken down between challengers, incumbents and open seats may prove instructive.

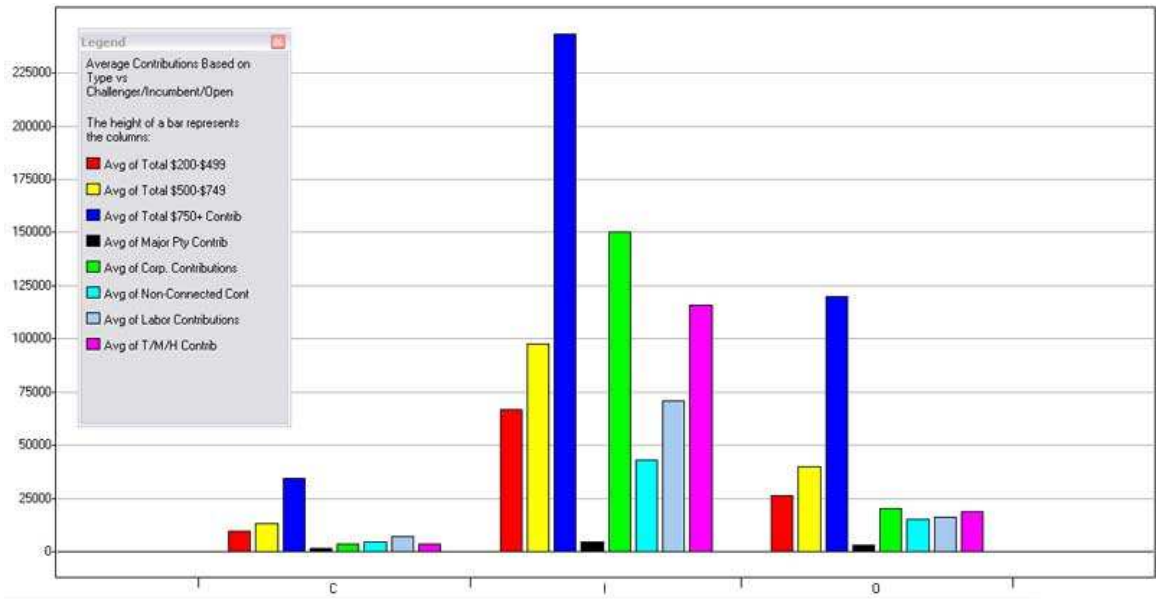
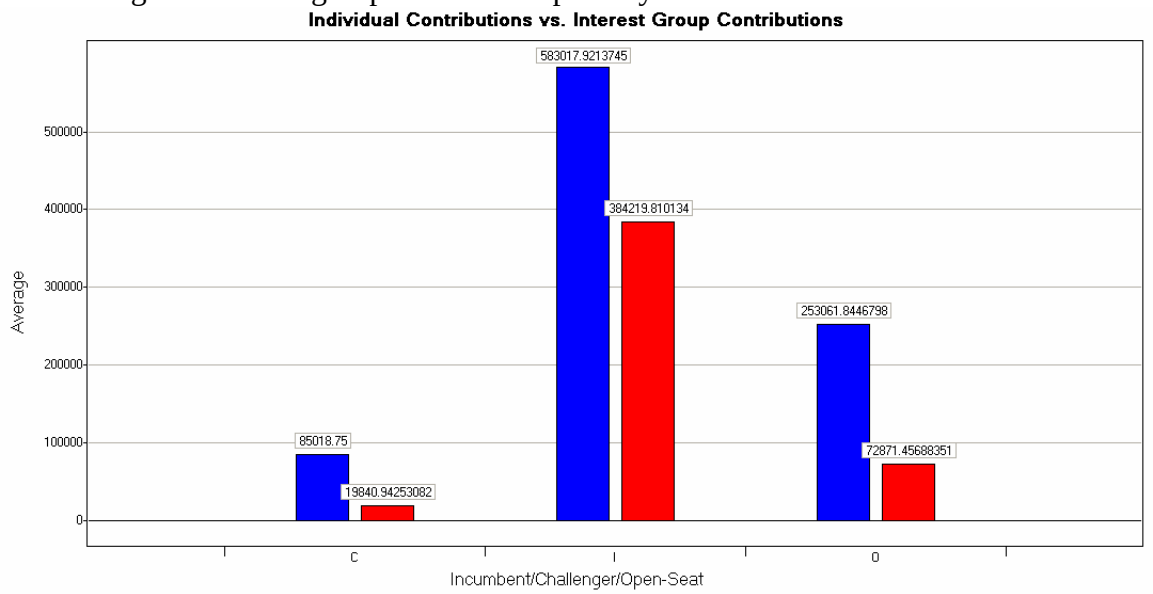


Figure 12: Average Contribution broken down by size and source for Challengers, Incumbents and Open Seats

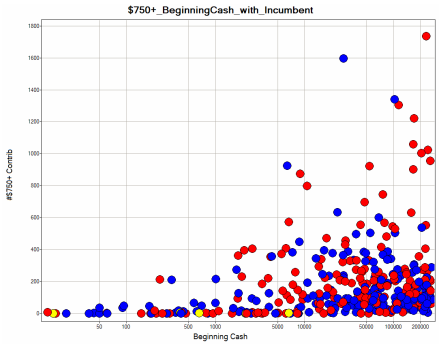
The relative proportions of many types (such as small, medium and large individual contributions) are similar. But what is very clear is that incumbents have relatively huge average contributions from corporations, non-connected, non-party committees and committees that are somehow (T/M/H Contrib) connected to the party. Because this money comes from various groups, it would not be inaccurate to refer to them as "interest groups". We see that corporations and connected committees are especially heavy contributors to incumbents.

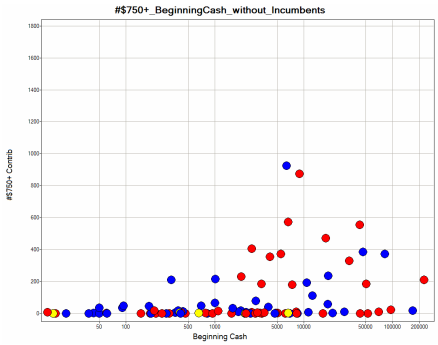
Sarita Yardi <yardi@sims>
Steve Chan <sychan@sims>

If we sum all the contributions that we've categorized as "interest group" contributions and compare them to total individual contributions, the relationship between incumbency, fundraising and interest groups becomes especially clear.



This graph shows that the proportion of money contributed to incumbents by interest groups is much higher, being roughly 65% as much as individual contributions, while challengers' special interest funding is only about 24% of what they receive from individuals. We also see that in contests for open seats, the proportional contribution by interest groups is somewhere in between what incumbents and challengers receive.





As we were exploring the dataset around different hypotheses, we tried a scatterplot that looked at large individual contributions versus starting cash. The plot on the left shows the number of large contributions versus starting money, while the sparser plot on the right is the same graph but without entries for incumbents. While this plot was for exploring another hypothesis, it also shows that incumbents typically start with a lot more money, and are also have a high number of large individual contributions.

Based on these graphs, it seems clear that there is a correlation between incumbency and the amount of money that can be raised (with a heavy advantage going to incumbents). Every category of contribution source is significantly higher on average for incumbents

Sarita Yardi <yardi@sims>
Steve Chan <sychan@sims>

than challengers. The other observation is that interest groups fund incumbents far more heavily than challengers (and more so than in elections for open seats).

Hypothesis #4

Given the results from hypothesis #2, where we see that negative campaigning seems to be effective at *denying* victory to an opponent, we speculate that spending on campaigning against candidates would trend upwards over time. We generated a bar chart that for a column summed the different categories of money spent against candidates, versus the year. What we found is that overall, there was a downward trend in the total money spent against candidates – with the notable except that during presidential election years (1996 and 2000) there were notable spikes in negative campaigning.

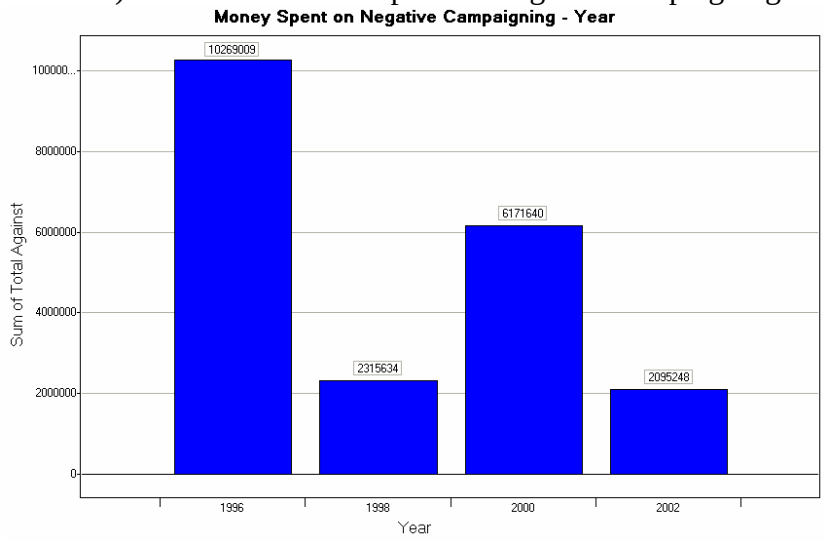


Figure 13: Total money spent on negative campaigning versus year

During presidential election years, the amount of money spent on negative campaigning more than doubles. Could this be due to simply having more money during election years? We look at the sum of Total Receipts over time to see what the trend is for the overall amount of money candidates receive during election years.

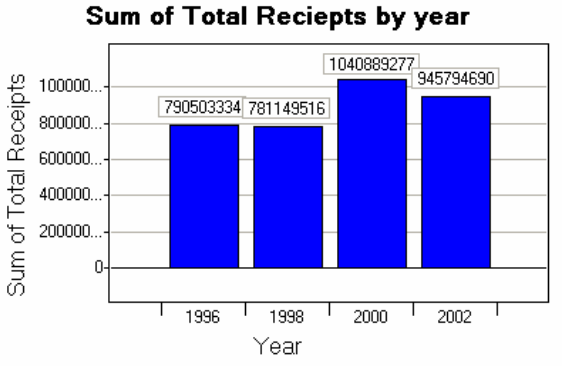


Figure 14: Total receipts by election year

What we see is that the amount of money spent against candidates doesn't seem to match up with the money that candidates receive. Where does this negative campaigning money

Sarita Yardi <yardi@sims>
Steve Chan <sychan@sims>

come from? Breaking down the components of the "Total Against" column in figure 13 tells us a more intriguing story.

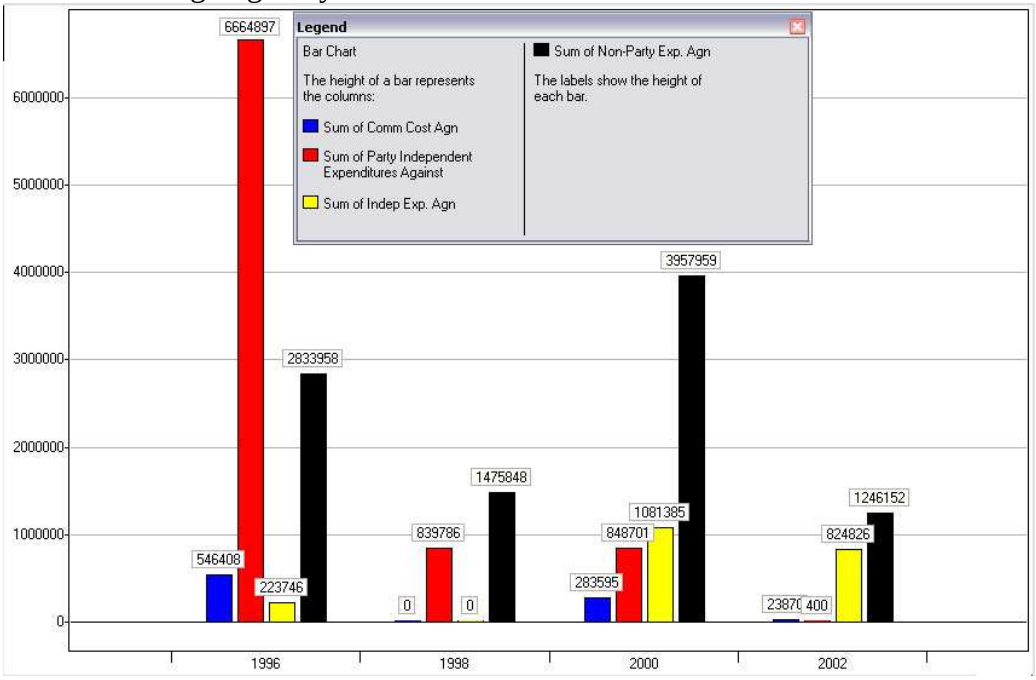


Figure 15: Sum of money spent against candidates, grouped by source of money over time

We see a dramatic drop in monies spent by the candidates themselves, and the independent monies spent by their party. However, independent and non-party monies stay relatively constant (taking into account the peaks during presidential election years). In the 2000 election, the non-party monies actually increased by over 1/3 compared to the previous presidential election year.

What seems to have happened is that candidates themselves have spent less money on negative campaigning, but interest groups have not backed off on the negative campaigning, and in the 2000 election they actually increased their negative campaigning efforts.

As a final tie in between hypotheses 2, 3 and 4, we offer the following simple chart that shows the relative amounts of money spent on negative campaigning between challengers, incumbents and for open seats.

Sarita Yardi <yardi@sims>
Steve Chan <sychan@sims>

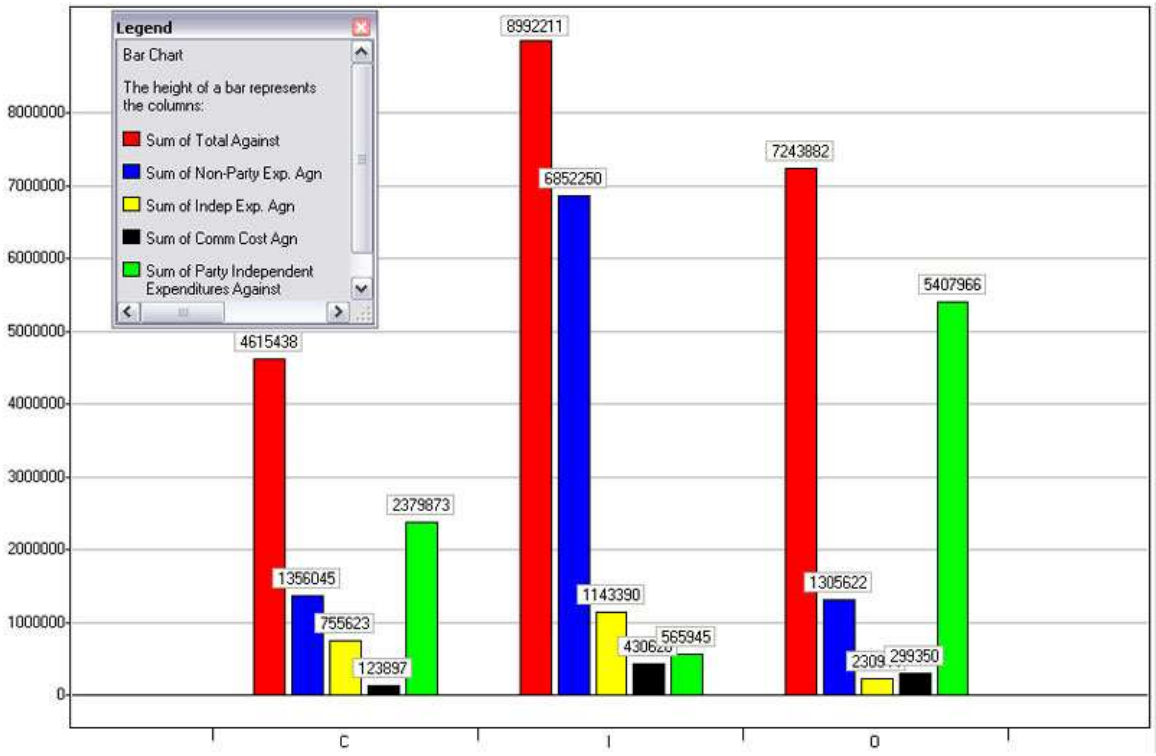


Figure 16: Money spent against a candidate broken down by source

The chart shows that there is roughly twice as much money spent **against** incumbents as there is spent against challengers. On the face of it, incumbents seem to be used as a *piñata* for negative campaigning – until we realize that our analysis shows incumbents have roughly seven times as much money as challengers.

A more interesting story is who is paying for the negative campaigning. About 50% of the money spent against challengers is from the "party independent expenditures against". This is money spent by opposing political parties (presumably by the incumbent's party). Against the incumbent, we see that over 75% of the money spent comes from "Non-party" groups. For open seats, we see relatively large amounts of money spent on negative campaigning, with 75% of it coming from the opposing party ("Party Independent").

Incumbents are attacked by interest groups, challengers and open seats are attacked by the opposing party.

The final observation on these analysis comes from what is **not** represented in the data. Looking at the columns reported by the FEC, we have the following:

- *Non-Party Expenditures For*
- *Non-Party Expenditures Against*
- *Independent Expenditure For*
- *Independent Expenditure Against*
- *Party Independent Expenditures For*
- *Party Independent Expenditures Against*
- *Party Coordinated Expenditures*

The first 6 items have a breakdown of for/against by groups not directly controlled by the candidate. However the seventh item does not have a *for/against* breakdown. There is

no smoking gun that shows how much money a candidate directly coordinated or spent on negative campaigning – only "independent" or "non-party" groups ever spend money on negative campaigning. We can only know how much money was spent against a candidate by relatively anonymous groups.

Analysis of the Tools

Overall, we preferred Spotfire to Tableau, feeling that it was easier to use and offered more interactive functionality. It was also a more visually appealing software tool. In Tableau, we liked being able to see the row and column options on the left hand menu and dragging them to the top row and column bars was simple and intuitive, however, the color highlighting and scatterplot features and zooming features made Spotfire a lot more exciting to use!

Spotfire defaulted Party 1, the Democratic Party to the color red when we chose to display Parties by color. It defaulted Party 2, the Republic Party, to the color blue. Needless to say, my initial conclusion, before was that my hypothesis could not have been more wrong. (Why we should assume that Spotfire can recognize particular parties and color them accordingly is a mystery, but we guess that is what my eye is trained to assume!)

We found the size scale to be unhelpful as a method for distinguishing categories. We had already settled on red for the Republican Party and blue for the Democratic party so we tried to use size to distinguish between states. (we were playing with comparing Texas and California). However, since size is scaled for all the possible values, in alphabetical order, the California circles were very small and the Texas circles were very large and covered a number of the California circles. Spotfire would be more useful if it could recognize how many and what range of the categories are actually being displayed and adjust it accordingly.

Setting colors in Tableau is more challenging than in Spotfire. We want to make the Republican Party red and the Democratic Party blue but we cannot figure out how to do that we can get one of them to be the right color, but the other color always defaults to grey. Given that this is a lesson in visualization, it is ironic that we can't find a way to compare Spotfire and Tableau results using the same colors, one of the most fundamental aspects of visualizations!

We found it far easier to use Spotfire's range tools to quickly adjust scales and get an idea of what my data looked like. In Tableau, we usually ended up double-clicking on the axis and then entering a guesstimate for my range by looking at the graph. We also found it really challenging to figure out how to put it on a log scale to compare it in Spotfire. Although this may have been demoed in class, we had a hard time remembering what we saw then, so we pulled up the help menu and followed those directions. We created a New Calculation and that syntax was analogous to Visual Basic syntax so we had no problem creating the log calculation, but then it wasn't initially intuitive how to add it to

the graph. Now that we know they calculations are added to the Measures box, we think using Tableau for creating custom calculations will be super easy (and interesting!). So we like that Tableau has such a short learning curve.

It would have been really useful to calculate a regression for hypothesis 1 in Spotfire, because we could internally visualize a logarithmic correlation in my data (see Figure 2) but we believe that our license didn't support regression calculations so we had to forego that hypothesis (not that we expect a student license to have all the features!).

Final Observations

As mentioned in the first section, we were somewhat hampered by what are either bad values in the FEC dataset, or a lack of insight into the semantics of the data. If we assume that the negative values for donations were errors, it is actually possible to drop them from the data being examined in Spotfire, either by using the query panel, or by graphing them and then marking the "bad" data for deletion. However, because we didn't know if those entries were actually *bad*, or if they were *merely misunderstood*, we decided to avoid doing any hypothesis that would be obviously effected by this.

While these tools are fairly powerful in how they allow you to visualize the data, there seem to be many usability issues. These may be due to the difficulty of creating an "expert" interface that provides enough power for flexible visualizations while also being concise and easy to navigate. Despite the initial learning curve, it was possible to test many hypotheses fairly quickly. Spotfire's ability to use the query panel to quickly and interactively filter data was very useful: the scatterplot in hypothesis 3 showing distribution of large donations with and without incumbents was very helpful in seeing the relationship between different subsets of the dataset. Overall, the tools did provide valuable insights into the dataset.

An interesting question, after having seen Martin Wattenberg's visualizations, is whether it is possible to create a general visualization tool (or even toolkit/api) that can give the sorts of insights and interactions that Wattenberg's specialized visualizations provide. Wattenberg seems to spend a lot of time examining the specific dataset he is working with before coming up with a particular visualization. Modulo the aesthetics of Wattenberg's custom visualizations, could something like Spotfire come close? It might be argued that Spotfire could be a tool that allows someone to really scrutinize the dataset in different ways, looking for a general idea of the sort of visualization that would work best – and then using that as an input to develop a custom visualization. Another interesting direction for these visualization tools is the role of collaboration and social interaction in creating visualizations. This seems to be the current direction of Wattenberg's work, and it would be interesting to speculate about the ways that Spotfire or Tableau could be extended to support collaborative or social interactions.