

Introduction

For this assignment I used the Spotfire and Tableau visualization tools to examine data on financial contributions to U.S. Congressional campaigns between 1996 and 2002. After surveying the data's components in Excel I made the following hypotheses:

1. In general, the candidates that raise the most money will win the race.
2. In races that involve an incumbent candidate, I imagine the incumbent will win in the majority of instances and that s/he will raise much more money than the challengers.
3. In Senate races between major party candidates, I imagine the individual who spends more of his/his own money will tend to win.

In my attempt to confirm or refute these hypotheses I was often struck by unexpected or curious tangential patterns within the visualizations; in several cases I chose to redirect my exploration in an effort to better illuminate these phenomena. What follows are highlights from my investigation.

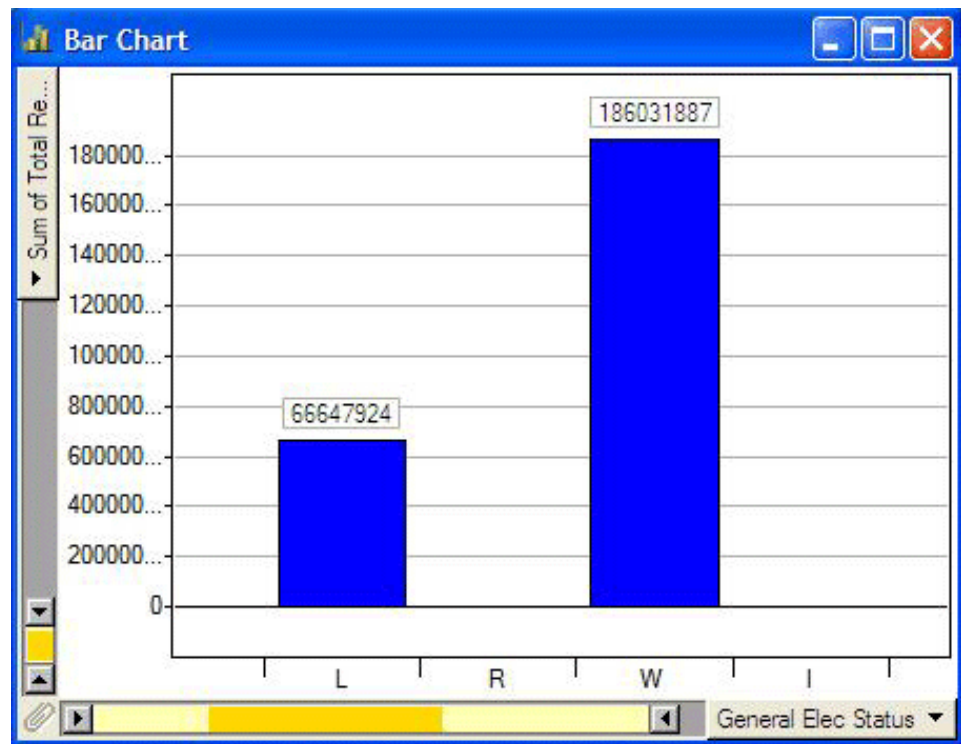


Figure 0. Win vs. Loss on x-axis, Total Receipts on y-axis.

Examination

I started my analysis using Spotfire. After familiarizing myself with the tool I decided to conduct analysis on a subset of the data: major party candidates for California's congressional seats. To investigate my first hypothesis – that candidates with the most money tend to win – I first constructed a bar-chart comparing General Election Results (limited to win or loss) on the x-axis and Total Receipts on the y-axis. Indeed, winners raised more money than losers (fig. 0). What this shows, though, is that the sum of receipts for all winners was greater than the sum or receipts for all losers. I didn't

find this visualization very informative. More interesting would be data showing how much the winner over- or under-spent the next closest contender in each race. Unfortunately, it would be quite difficult to organize candidates by race on aggregate (doing so race by race is easier given the district codes, but I wanted to make aggregate comparisons). Given this limitation I decided on a different tactic: I would compare Total Receipts (y-axis) to the Percentage of General Election Vote Received (x-axis). Doing so in a scatter-plot format allowed me to more clearly see individual instances (candidates) and hence look for clusters and trends. The resulting

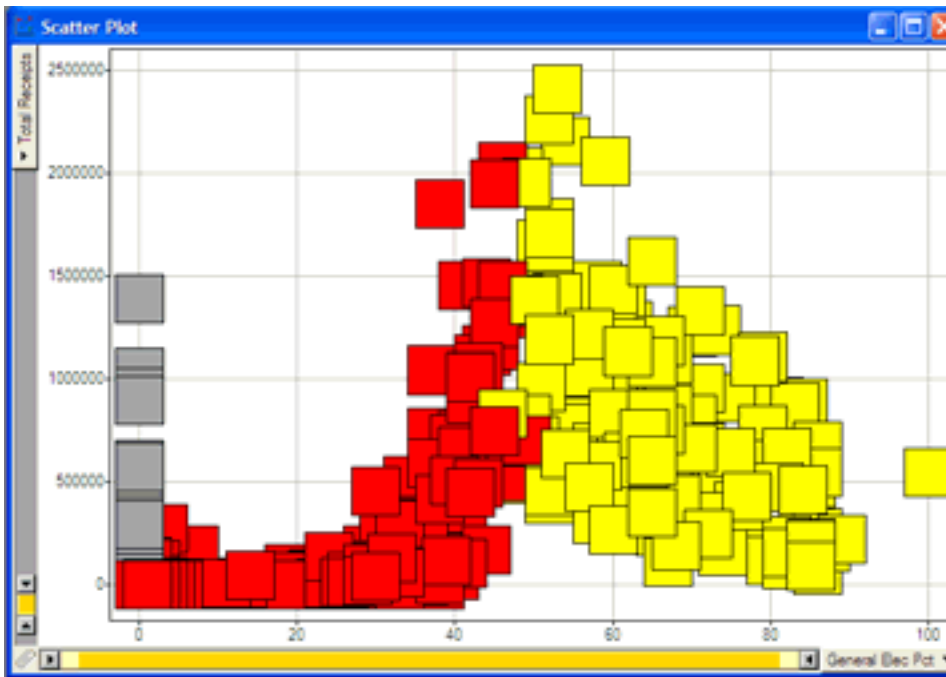


Figure 1. Percentage of general election vote (x-axis), Total Receipts (y-axis). Red is a loss, yellow is a victory.

visualization (fig. 1) initially surprised me but soon made clear sense. The scatter plot tended towards a bell curve, with a peak in Receipts occurring amongst candidates who received approximately 50% of the vote. To better show this trend I focused my scope by eliminating the Senate candidates: their fundraising efforts dwarf those made by House candidates, and their sample size is too small for useful trend analysis when the overall scope is limited to the races in a single state. I also used color to delineate winners and losers (yellow for a victory, red for a loss). The bell-shaped clustering suggests that candidates in close races outspent those in less close races [holding all other factors constant, a caveat that applies to all interpretation presented in this paper]. This makes perfect sense:

candidates that are expected to win by a comfortable margin will often raise less money than those in competitive races. But it also suggests that plenty of candidates lose despite raising much more money than victorious candidates, they just probably do so in a different race. Once again, what seems to matter is spending relative to your competitor. This visualization also illustrates the dominance of our two-party, and hence two-candidate, system. Very few candidates won the election with less than 50% of the vote. In a system with more candidates we would expect to see more candidates winning with a lower percentage of the vote.

Two additional features in this visualization deserve comment. The first is that candidates that win a large percentage of the vote tend

to raise more money than those that win a small percentage of the vote. We can see this by the rightward leaning bulge in the bell curve. Indeed, a good portion of losing candidates raise little-to-no money; we can hypothesize that these were essentially uncontested races due to the presence of a strong incumbent in that district – someone ran as a token candidate for the opposing party, but little resources were raised. The second feature worth noting are the gray boxes on the left-hand side of the visualization. A quick examination reveals that these are candidates that raised money for, but did not win, primary elections and hence received 0% of the vote in the general election.

After analyzing this visualization, I wanted to see if the pattern persisted for smaller subsets of Total Receipts. One of the subsets I viewed, Contributions by Major Political Parties, caught my eye because of an extreme outlier (fig. 2). At this point I had expanded the scope to once again include Senate candidates. Nevertheless one candidate, Matthew Fong, was way off the charts with nearly three times the party contributions than those for the next closest candidate (not surprisingly, the next closest candidate was his opponent Barbara Boxer). I decided to select both candidates and start building additional visualizations. I was curious to see why Fong

was such an extreme case, and, why he lost the race despite being supported by a political party to an unprecedented degree. The first comparison I constructed compared Major Party Contributions with Total Receipts (fig. 3). I did so to see how Boxer compared when all fundraising activity was included.

Not surprisingly, there is a positive correlation between Total Receipts and Major Party Contributions (a movement up and to the right). Yet the pattern also suggests that this correlation is relatively weak – there are many candidates who have higher total receipts and yet lower contributions from major parties. This visualization also shows clearly that Boxer raised more money than Fong despite a huge relative deficit in major party contributions. This view also introduced a new outlier that had to be accounted for (the circle highest and most to the left). According to this visualization, this candidate had highest Total Receipts in 1998. I decided to mark it and watch for further information as I explored the source of Boxer’s advantage.

I quickly compared two more sources of contributions: one showing Corporate Contributions (fig. 4) and one showing Labor Contributions (fig. 5). Not surprisingly Fong far outpaced Boxer in Corporate Contributions and Boxer received much more from Labor (Fong got

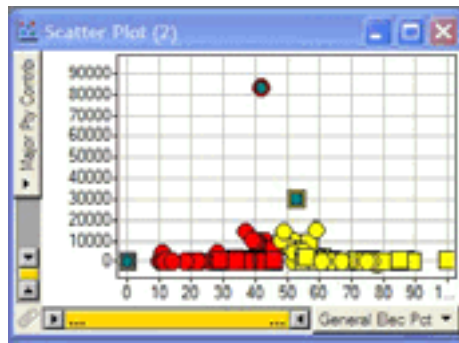


Figure 2. Fong is the green circle with a red rim.

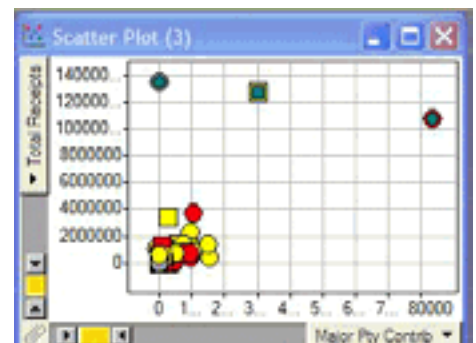


Figure 3. Boxer is the green square. The mysterious outlier is the green circle (left).

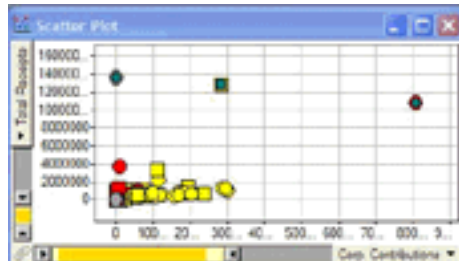


Figure 4.

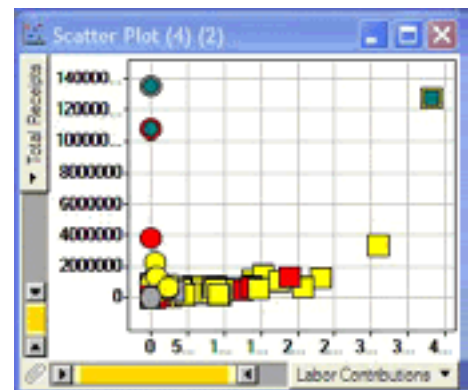


Figure 5.

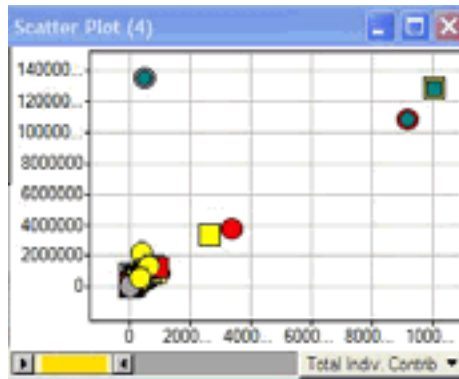


Figure 6.

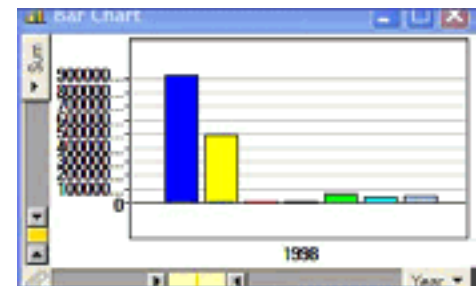


Figure 7.

zero from Labor). Yet neither of these contribution sources exhibited strong enough positive correlation with Total Receipts to account for Boxer’s advantage. Only when I compared Total Individual Contributions to Total Receipts did a strong positive correlation emerge (fig. 6). The instances make a relatively straight line up and to the right, and the slope of the line is much steeper than the previous examples. Since the (x,y) axes are approximately the same length, this shows that Individual Contribution is a greater



Figure 9. Out mysterious outlier moves off the y-axis.

determinant of Total Receipts than the other forms of contribution we looked at. Indeed, Boxer’s advantage in Individual Contributions appears proportionate to her advantage in Total Receipts. A review of the scale for each axis confirmed

this observation: Boxer had nearly 13M in Total Receipts and just over 10M of that came from Individual Contributions. In contrast, Fong's extreme relative position in Party Contributions accounted for only just over \$80,000. To make this point even more clear, I constructed a bar chart comparing total receipts (blue) to its major subsets of contributions (fig. 7). Only individual contributions (yellow) stands out. Boxer's small relative advantage in Individual Contributions equated to a significant dollar advantage – enough so to make negligible Fong's advantages from Corporate Contributions and Major Party Contributions.

As for the remaining mysterious outlier (who, despite Total Receipts of nearly 14M lost in the primary), I found that nearly his entire source of funding came in the form of a loan from himself (fig. 8). This

revenue source explains his large Total Receipts despite dismal performance in all other categories including Individual Contributions.

After this series of comparisons, I moved on to Tableau and decided to expand my scope out of California and consider the entire U.S. dataset. Like before with Spotfire, I spent some time investigating the program before delving into analysis. My goal with Tableau was to investigate my hypothesis about incumbent advantage. I spent some time comparing various views on the data before coming to a combination that I felt told the story quite well (fig. 9). In this visualization, like my initial find in the California election data, I compare the percentage of votes received in the general election to Total Receipts, distinguishing winners from losers by color. But this time I made small multiples of the graphic: one for in-

cumbents, one for challengers, and one for candidates who ran for an open seat. If the three visualizations were put together we would expect something very similar to what we saw for California in (fig. 1). But when stratified by these criteria we see some interesting and clear divisions. Most notable is the difference between incumbents and challengers: they almost perfectly divide the bell curve, with incumbents occupying the vast majority of instances on the right side of the curve (more than 50% of the vote and hence victory in most cases) and challengers occupying the majority of the left side (less than 50% of the vote and hence defeat in most cases). Each dips slightly into the other's territory, suggesting that occasionally a challenger will beat an incumbent but will do so by a small margin. The third visualization, where the seat was open, shows a pattern similar to what we'd expect if the data

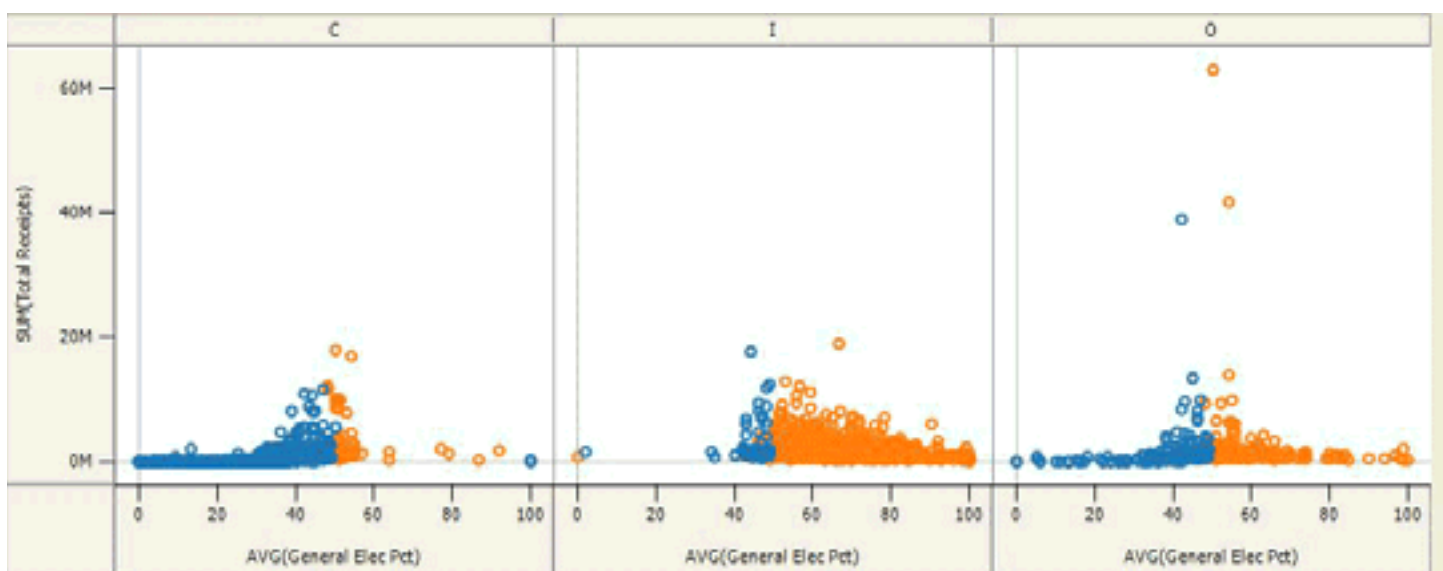


Figure 9. (left to right) Challenger, Incumbent, and Open Seat. Percentage of general election vote on x-axis; Total Receipts on y-axis. Color shows win vs. loss.

weren't stratified. The only major difference being that the number of instances is much smaller.

This illustration of incumbent candidate advantage led me to investigate the phenomenon in more detail. I eventually landed on an examination of how Beginning Cash might affect how much money a candidate raised and whether or not s/he won (fig. 10). This visualization shows that incumbents with a starting cash advantage almost never lose. Interestingly, this appears to be a more important factor than the Total Receipts raised. The losers are distributed along a range of zero to over 10M in Total Receipts, yet nearly all share the quality having started with little-to-zero Beginning Cash. By contrast, many winning candidates fell well short of 10M in Total Receipts, yet many had noticeable amounts of Beginning Cash.

By this point I had a good hunch that my third hypothesis would be difficult to prove from the data. I had hypothesized that Senate candidates who spend more of their own money would likely win. I now knew that the data made an aggregate comparison of races difficult. And I knew that individual contribution was not usually a major component in total receipts. The most telling features I discovered when I visualized this factor (plotting per-

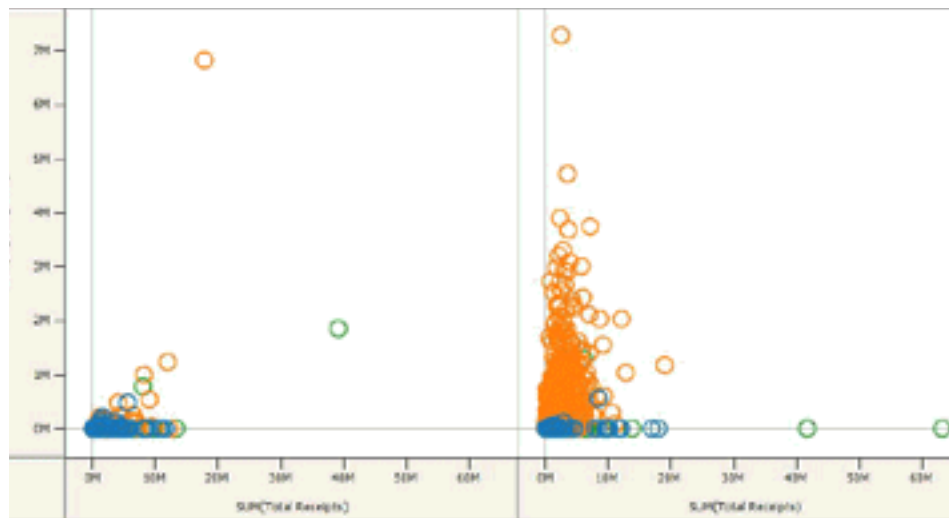


Figure 10. Total Receipts on x-axis. Beginning Cash on y-axis. Left chart shows losers. Right chart shows winners. Orange=incumbents, blue=challengers, green=open seat.

centage of vote on the x-axis and both Contributions from the Candidate and Loans from the Candidate on the y-axis) are that significant contributions from individual candidates are rare, and they tend to occur in tight races – sometimes winning and sometimes losing. This point was driven home time and time again: close races seem to produce an “arms race” mentality in which both winners and losers spend much more than candidates in non-close races. Raising a lot of money alone doesn't guarantee victory, but rather it ups-the-ante to the opponent.

Appendix

What follows are a few notes on features that were either good or problematic in each program:

- In Spotfire, I couldn't find an easy way to change the visualization format of the current graph (e.g. from scatter-plot to bar graph). The process I ended up using was to start a new visualization and then set each axis from scratch.
- I like the way Tableau truncates large numbers (e.g. 60M) when labeling tick-marks. Spotfire would show all the digits (which can be hard to count) or use a ... approach.
- I found it hard to control zooming with Tableau. I would use the "focus" tool (which is not intuitive) but then I couldn't zoom-out on my data. The only way I could take a broader view on the data was by double clicking on an axis and then resetting its value in a pop-up box. This is a very cumbersome way to switch between detail and contextual views.
- I like the simplicity by which you can view the underlying tabular data in Tableau.
- Tableau wants to keep me constrained to one view of the workspace at a time. When I minimized a view and tried to make a new view for comparison, the two views did not support brushing and linking with one another.
- Similarly, if I make the window smaller in Tableau I don't get scrollbars and therefore lose whatever falls outside the window. This occurs if I adjust window size by dragging on the border of the bounding box (not if I use the toolbar buttons for bigger and smaller view).