Campaign Donation Analysis

Eric Garcetti for Mayor

Patrick McCarthy

May 2013

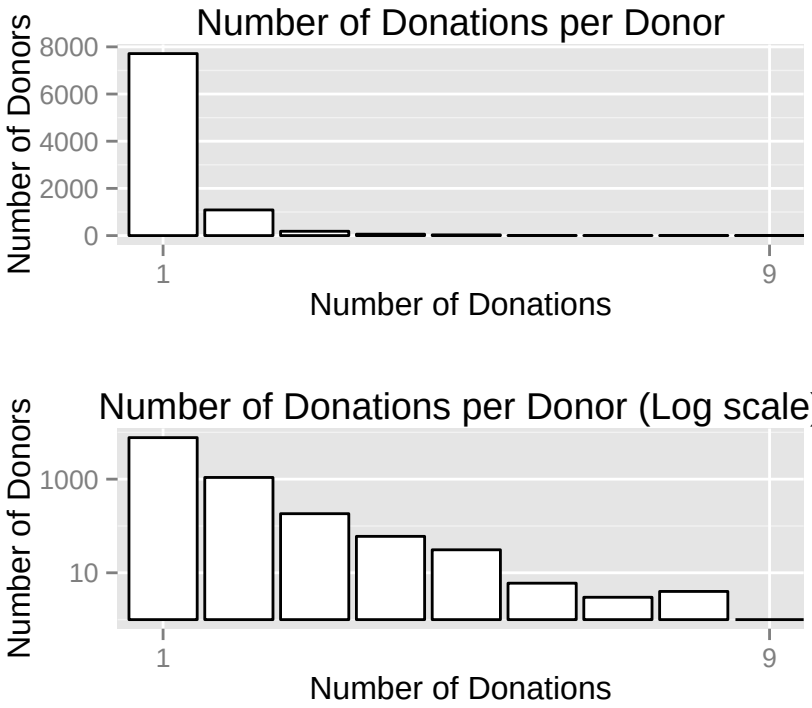
In late March 2013 I did a project for Eric Garcetti's campaign for mayor of Los Angeles, in which I attempted to classify and understand donor behavior. The objective was to help the campaign determine which existing donors could be approached again, how much to ask for, and to determine where new donors might be brought in.

With only two months left in the race (and no official standing with the campaign) I wasn't able to gather any data beyond what I was provided, but at the end of the analysis I list a few ideas that might provide interesting results.

Contents

What are the giving patterns?

Number of donors by frequency



If we want to get people to donate again, how often does that happen? Not very, it turns out. After the first gift, the number shrinks dramatically, and then dramatically again. It's a little interesting that with each donation, fewer than 2 in 5 (and sometimes 1 in 10) come back again.

Number of donations by donor

Most Donations by an individual: 9

1	2	3	4	5	6	7	8	9
7715	1088	183	60	31	6	3	4	1

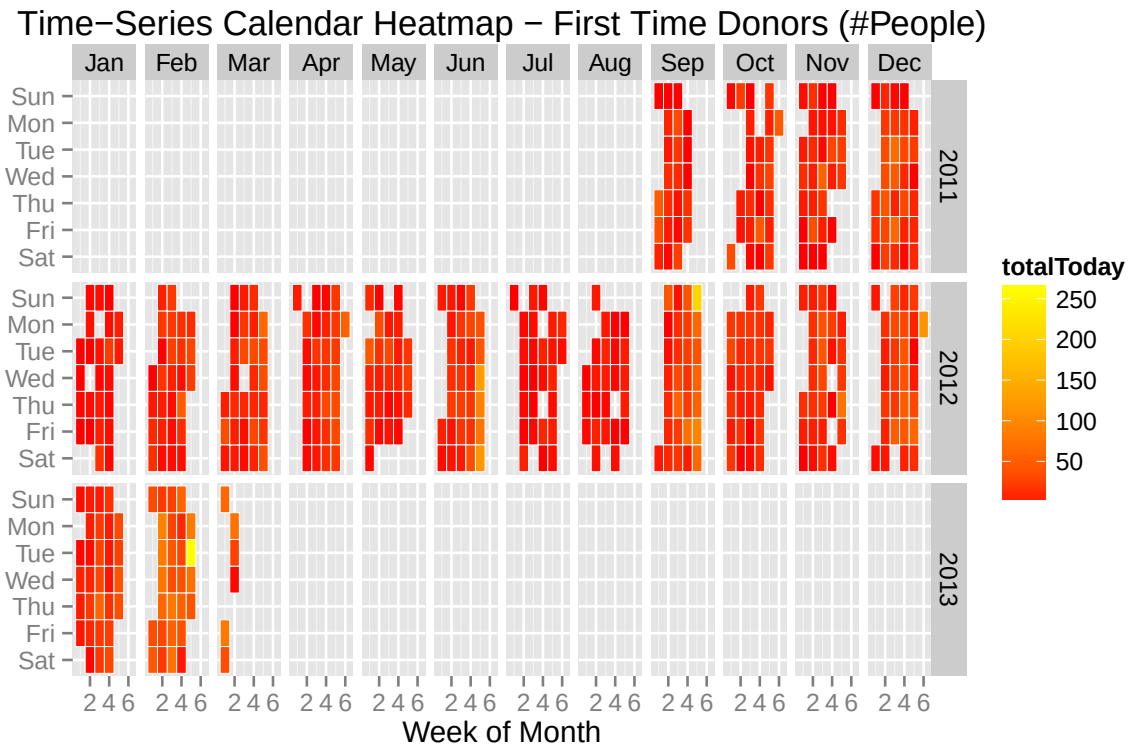
Table 1: Count of Donors by Donation Frequency

The max amount that can be donated by an individual is $1,300 during the primary (until 3/5/13) and another $1,300 during the general election until election day on 5/12/13. As I'm working with data from the full primary period, I can assume that everyone can legally donate up to another $1,300.

It might be interesting to try to predict how to make a 5-time donor contribute again, but there aren't very many of these. Instead to maximize output I focus on 1-to-2 and 2-to-3 conversions, covering 96% of the listed donors.

One easy thing to follow up on is if there are certain dates or events that bring in lots of donations. The campaign staff can reference this chart to find very effective fundraisers or rallies and try to replicate them.

Since far more people donate a first time and stop than ever repeat, the last question is "Where can we find more fresh people?". I don't know how new donors are found, but from this data I can show when people donate for the first time:



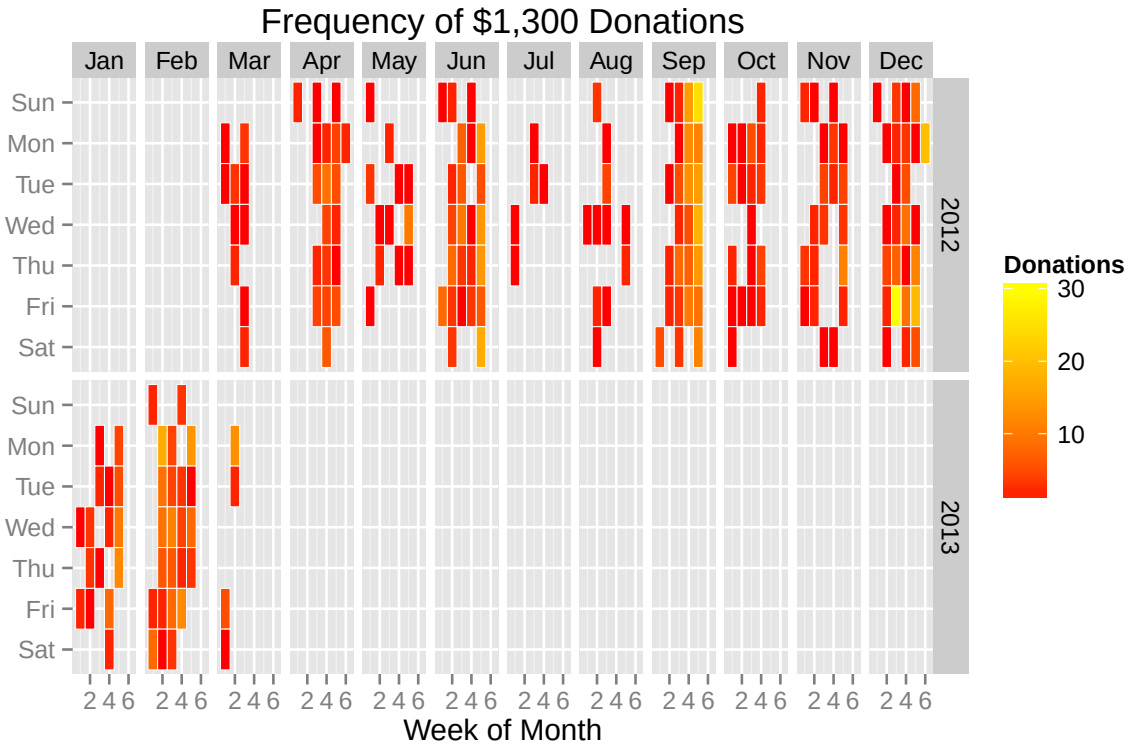
Characteristics of a next donation

If in a set of donors only 1 in 5 is going to come back, it's all the more important to try to profile why those who come back are different. Unfortunately, the data doesn't provide much about the individual (age, sex, employement status, salary...), leaving us only with donation behavior.

Since we want to maximize the donors we have, we want to use multiple donors to tell us how to solicit less frequent donors of similar characteristics.

Maxed out with one donation

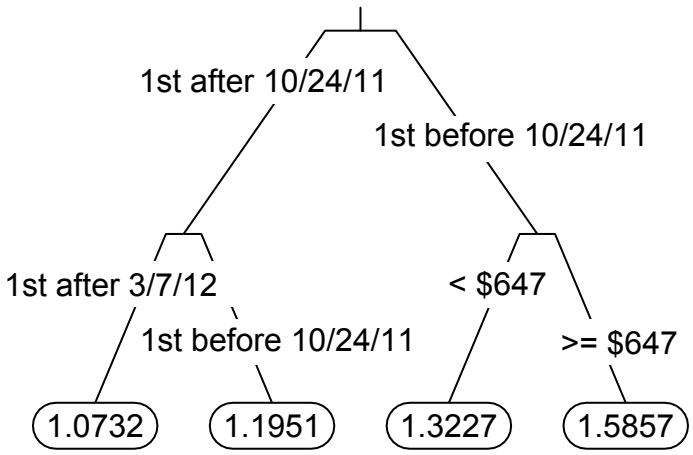
The first group of interest are those who donated the maximum during the primary with one donation. Hopefully they'll do it again in the general election.



The campaign staff can determine what's significant about these days and find the corresponding donors in the database.

What is a two-time donor?

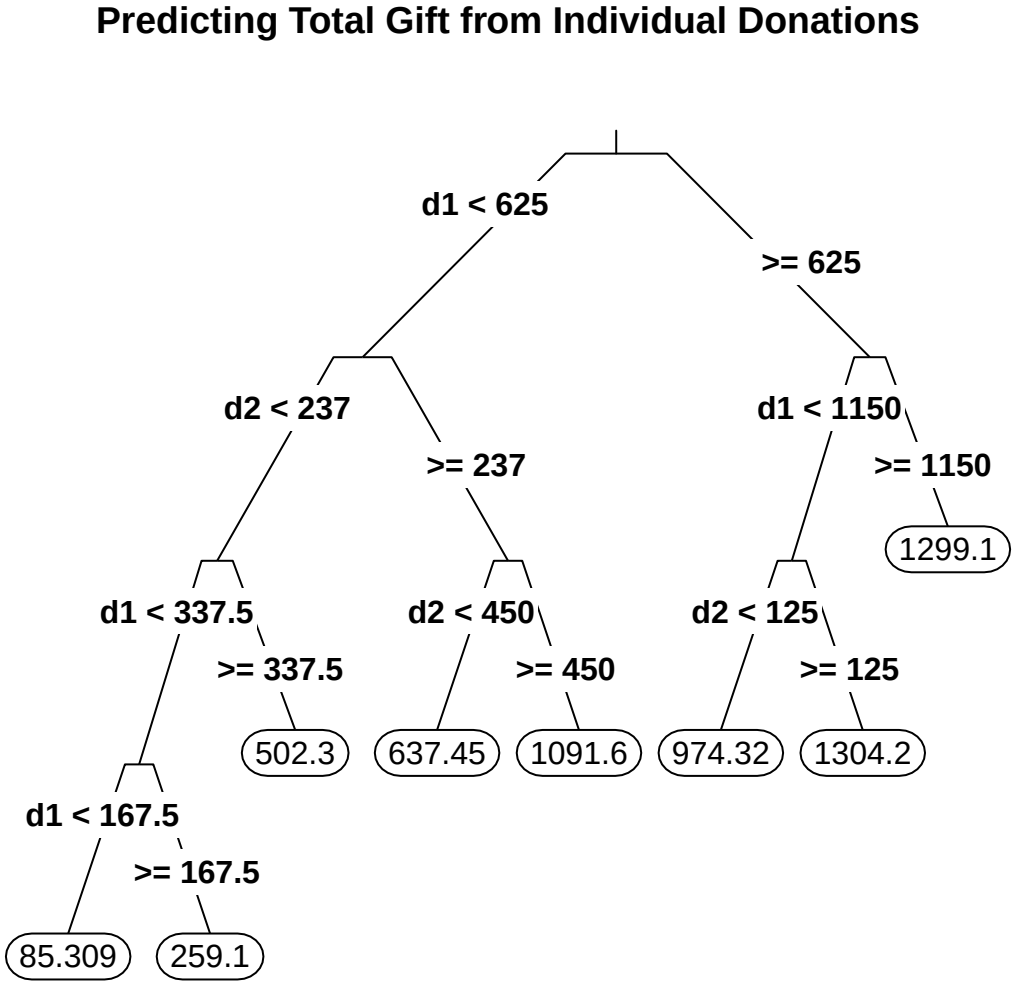
To distinguish characteristics that may indicate propensity to donate twice, a regression tree was applied to the only variables visible for one-shot donors, amount donated and the date of the donation.



On this chart, the 'leaves' at the bottom of the tree indicate the expected number of donations if the criteria above are fulfilled. From this, donors who invested a lot (¿$647) and early (before 10/24/2011 for a primary 3/5/2013) appear to have the highest expected rate at 1.5 donations.

Sorting on the lowest branches produced 1,881 high-likelihood candidates with an average possible remaining donation of $300 each. Of two-time donors fulfilling these criteria, 93% donated exactly $300 the second time. A cursory overview of the corresponding records shows that most of these one- and two-time donors are professionals (attorneys, consultants, etc.) who donated $1,000 in the first round.

Classifying several donations



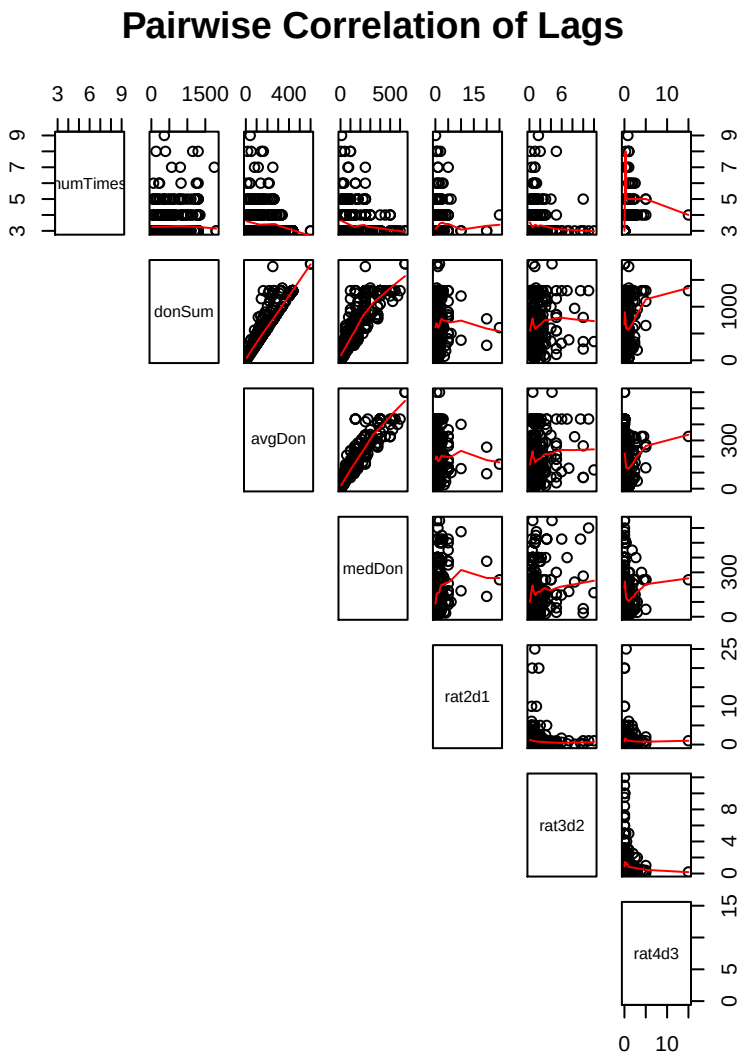
This model is fit to the first donation amounts and their corresponding dates, as the sample size for five or more donations isn't sufficient for a robust analysis. This time modeling the total donation amount (max $1,300), it seems only the first and second donations are significant enough to make a prediction. Neither the timing of the donations nor the amount of time between them appeared in the model, suggesting we don't need to track our timing, just the amounts.

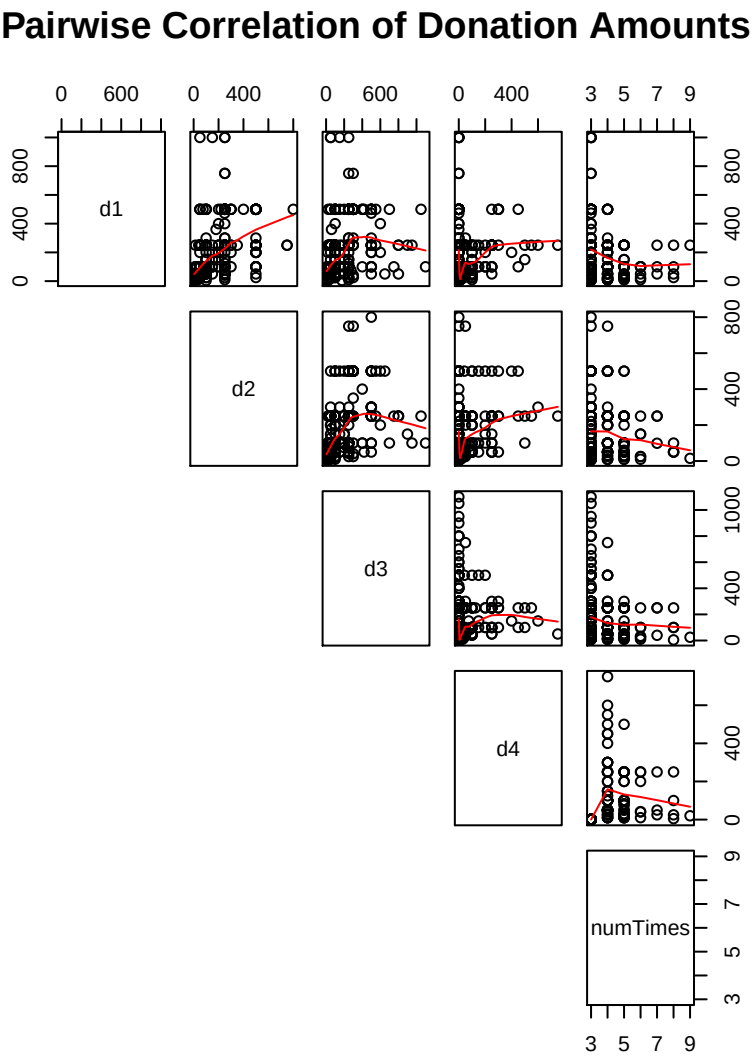
From this model, attention should be focused on first time donors who gave more than $625, and second time donors whose second donation was more than $237. Nobody else comes within $500 of their maximum potential.

Of course, these are only hints in the right direction, and far from gospel truth. In fact, it's worth mentioning at this point that these results from the primary may not transfer to the general election at all. Despite this, it's worthwhile to look for these patterns.

Personal donation trends

After a big first donation, is a typical donor more likely to give a second gift that's big or small? Is there a trend at all? To investigate this, I generated for each person the ratios between their nth and $n+1$th donations, and analyzed the relationships between these values and total, average, and median donation amounts to look for trends.





The scatterplots are not promising. Where a relationship exists, the red trend line should be different from flat, and should pass through somewhat dense areas of points, but here this does not happen often. Strong trends appear in common sense places (*e.g.* total donated and average donation are correlated) or are thrown off by one outlying point, but generally are not different than the flat line of no effect.

To find the impact of different donation levels on the expected amount, I regressed the amounts of a person's first four donations (if any) and the number of times donated against the total contribution. Because these values are very correlated, I orthogonalized the columns for linear independence. The advantage of this approach is that it provides the most clear picture of the relationships of the variables, whereas the disadvantage is that you can't take the effects literally, only relatively.

The largest positive coefficient in the model reinforces what the tree diagrams told us above, that the first donation is the most important for determining total returns, with the others much less so. The negative coefficient on 'number of donations' tells us that we want donors who give a lot in a few bursts, rather than those who trickle their way toward the cap with several small chunks.

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	0.01	0.00	29.34	0.00
1st Donation	0.68	0.03	26.92	0.00
2nd Donation	0.12	0.01	10.77	0.00
3rd Donation	0.06	0.01	5.60	0.00
4th Donation	0.03	0.01	2.67	0.01
Total Number of Donations	-0.58	0.02	-26.18	0.00

Table 2: Regression of Orthogonalized Donation Amounts on Number of Donations

How much to ask for

A targeted regression model might be used to determine a precise amount that a given person may be willing to give, however the limited resources of a campaign often require painting with a broader brush. If you can only send out direct mailings asking for five or ten different amounts, what should those amounts be?

Clustering of donation amounts

Is a donor giving $50 more like one who gave $5 or one who gave $100? It's important to find these groupings so you can set an upper bound to ask for, as not to leave money on the table.

The following are k-means clusters of donation amounts, which is to say the algorithm attempted to determine which amounts were most like the others. This should be a better estimate than arbitrarily dividing $0-$1,300 into equal-sized bins.

	Minimum	Median	Average	Maximum	Count
1	0.01	50.00	70.70	160.00	4172
2	174.67	250.00	255.06	375.00	2934
3	400.00	500.00	498.34	600.00	1478
4	650.00	750.00	764.95	935.00	172
5	950.00	1000.00	1131.89	1300.00	2177

Table 3: Automatic Clustering of Donation Amounts - 5 Clusters

	Minimum	Median	Average	Maximum	Count
1	0.01	50.00	70.70	160.00	4172
2	174.67	250.00	254.14	300.00	2906
3	325.00	400.00	381.66	433.00	71
4	450.00	500.00	500.11	575.00	1419
5	587.00	700.00	661.75	720.00	53
6	750.00	800.00	779.69	800.00	128
7	850.00	900.00	904.38	950.00	8
8	961.50	1000.00	1000.59	1050.00	1211
9	1100.00	1200.00	1177.89	1202.03	17
10	1250.00	1300.00	1298.99	1300.00	948

Table 4: Automatic Clustering of Donation Amounts - 10 Clusters

Insignificant tests

In addition to these approaches, there are several avenues which didn't pan out. I was hoping to be able to discern trends between businesses and individuals, to do classification based on telephone area code and exchange, and to distinguish between people with email addresses from different services, such as Gmail and aol.com. Though initially promising, none of these produced any significant results.

Ideas for next time

Having been brought in late there was little to do but to work with the data that was available. Were I to join a new campaign, however, I'd suggest enhancing understanding (and hopefully returns) in a few ways:

Track sources Did a donation come in at random from the internet, from canvasers, from a fundraising dinner? This would help to segment the data and hone in on ways to optimize each revenue stream.

A/B testing Find similar neighborhoods (or even blocks within a single neighborhood) and test variations in literature, canvasing times, etc. Design experiments to examine the role of particular issues in communities or to isolate, understand and control the impact of particular demographic factors.

Examine old data Campaign finance data like that examined here is a matter of public record. Although each race is different (Antonio Villaraigosa was a strong incumbent with some differences in core supporters) it would be worthwhile to try to learn more about how campaign fundraising works in LA in general.