Campaign Contributions: Presidential Race in Pennsylvania, 2016

Tom Nececkas

Contents

Can campaign contributions predict an election's winner?

In 2016, campaign contributions seemed like a lousy predictor of success.

For example, in a March press statement, Bernie Sanders' campaign bragged about having 2.2 million more contributions than Hillary Clinton's campaign. Come July, Clinton was the Democratic Party's nominee.

Discussing the Clinton campaign's staggering $400 million in contributions as of September, the Atlantic wryly observed: "If Hillary Clinton loses in November, it won't be for lack of money." Come November, Donald Trump was the president-elect.

So what, if anything, do campaign contributions say about who's likely to win an election?

Pennsylvania seemed like a good place to start answering that question. It's the 5th most populous state and, more importantly, it's widely seen as a battle ground state, and some had predicted it could be the deciding factor in the 2016 election. Plus, Pennsylvania voters chose the same winners as the country.

The Dataset & Summary Statistics

The dataset of campaign contributions was made available by the Federal Election Commission. When I downloaded the dataset, an extra comma in the header row prevented it from loading. If you run into the same problem, you can simply open the .csv file and remove the comma.

I've included a basic summary of the dataset. Although the summary doesn't help answer our question, it's still useful to see what types of information are available to explore.

One important thing to note: Each of the dataset's roughly 244,000 observations represents one financial contribution. Each observation *does not* represent one individual, since an individual could give more than once.

```
##       cmte_id            cand_id                              cand_nm
##  C00575795:119288   P00003392:119288   Clinton, Hillary Rodham  :119288
##  C00577130: 59627   P60007168: 59627   Sanders, Bernard         : 59627
##  C00580100: 30828   P80001571: 30828   Trump, Donald J.         : 30828
##  C00574624: 15544   P60006111: 15544   Cruz, Rafael Edward 'Ted': 15544
##  C00573519:  9800   P60005915:  9800   Carson, Benjamin S.      :  9800
##  C00458844:  3094   P60006723:  3094   Rubio, Marco             :  3094
##  (Other)  :  5615   (Other)  :  5615   (Other)                  :  5615
##           contbr_nm            contbr_city     contbr_st
##  COMELLA, JOHN :   187   PHILADELPHIA: 32880   PA:243796
##  BETHEA, DAMON :   180   PITTSBURGH  : 21740
##  SHOVLIN, MARIE:   150   WEST CHESTER:  3646
##  ROSOFF, ANDREW:   142   HARRISBURG  :  3638
##  SHORT, CHRIS  :   139   LANCASTER   :  3620
##  LIBERTIN, MARY:   136   (Other)     :178271
##  (Other)       :242862   NA's        :     1
##    contbr_zip                     contbr_employer
##  Min.   :        0   N/A                  : 34161
##  1st Qu.:152322646   RETIRED              : 31524
##  Median :180173915   SELF-EMPLOYED        : 16257
##  Mean   :153092041   NONE                 : 12762
##  3rd Qu.:190951323   INFORMATION REQUESTED: 10342
##  Max.   :196125107   (Other)              :138488
##  NA's   :23          NA's                 :   262
##              contbr_occupation  contb_receipt_amt  contb_receipt_dt
##  RETIRED              : 56142   Min.   :-93308.0   Min.   :2014-07-17
##  NOT EMPLOYED         : 17391   1st Qu.:    15.0   1st Qu.:2016-03-11
##  INFORMATION REQUESTED: 10269   Median :    27.0   Median :2016-06-16
##  ATTORNEY             :  5928   Mean   :   102.8   Mean   :2016-05-29
##  PROFESSOR            :  5822   3rd Qu.:    80.0   3rd Qu.:2016-09-14
##  (Other)              :148175   Max.   : 10800.0   Max.   :2016-12-31
##  NA's                 :    69
##                      receipt_desc    memo_cd
##                            :241328    :193406
##  Refund                    :  1508   X: 50390
##  REDESIGNATION FROM PRIMARY:   177
##  REDESIGNATION TO GENERAL  :   176
##  REATTRIBUTION TO SPOUSE   :   134
##  REATTRIBUTION FROM SPOUSE :   130
##  (Other)                   :   343
##                                memo_text       form_tp
##                                     :160839   SA17A:193065
##  * EARMARKED CONTRIBUTION: SEE BELOW: 57902   SA18 : 49223
##  * HILLARY VICTORY FUND             : 23233   SB28A:  1508
##  *BEST EFFORTS UPDATE               :   335
##  EARMARKED FROM MAKE DC LISTEN      :   281
##  REDESIGNATION FROM PRIMARY         :   177
##  (Other)                            :  1029
##     file_num            tran_id       election_tp       X
##  Min.   :1003942   C1014901 :     2        :   452   Mode:logical
##  1st Qu.:1077916   C10164585:     2   G2016: 92431   NA's:243796
##  Median :1104813   C10165181:     2   G2106:     1
##  Mean   :1103468   C10179391:     2   O2016:    43
##  3rd Qu.:1133832   C1022774 :     2   P2016:150868
```

```
##  Max.   :1146285   C1022815 :     2   P2020:     1
##                    (Other)  :243784
```

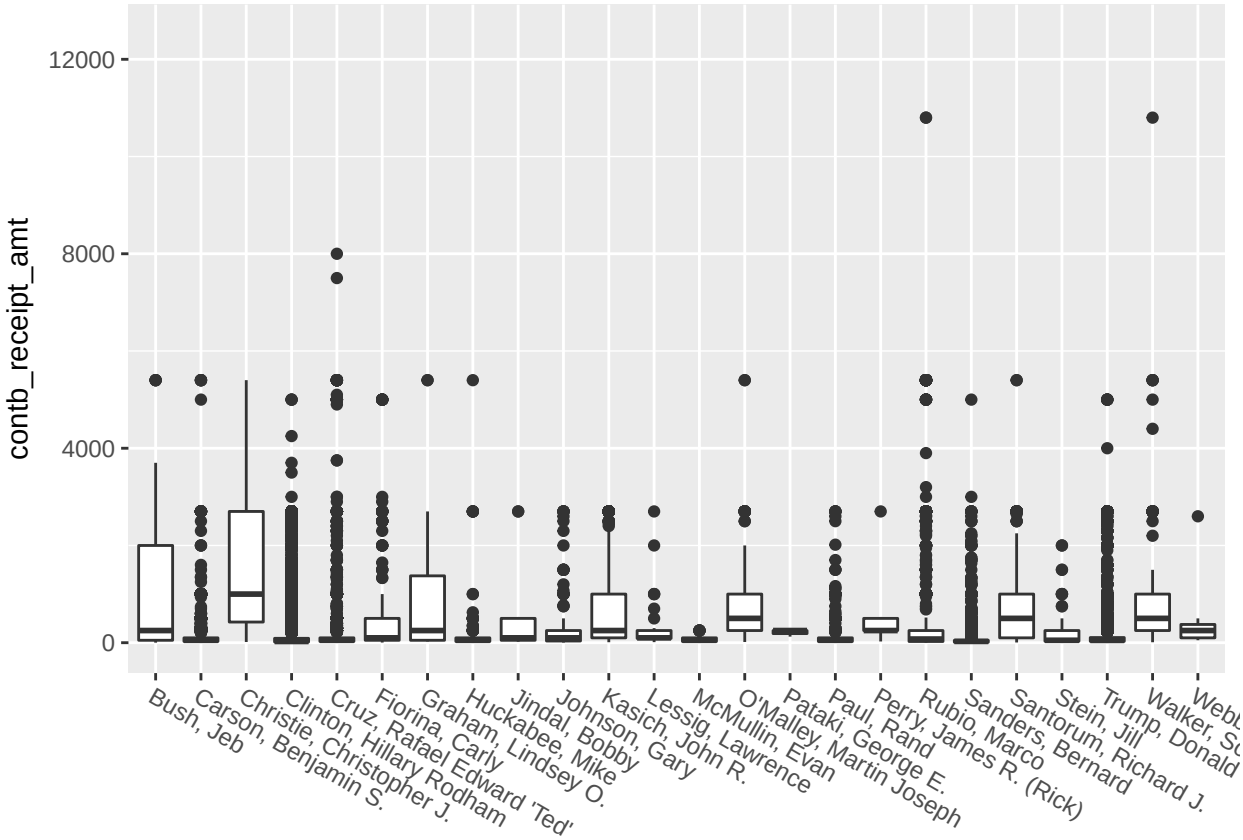
Exploration

Does the average size of a contribution matter?

Some campaigns rely on lots of donors who give small amounts, while other campaigns rely on fewer donors who give large amounts. Therefore, the average size of a contribution might say something about the types of people supporting a campaign, which might indicate a campaign is more or less likely to succeed.

I began by making a boxplot to look at the average contribution size for each campaign. I excluded negative values, which represent a campaign returning money (e.g. because the campaign ended, or because a contributor more than is allowed.)

```
## Warning: Removed 2746 rows containing non-finite values (stat_boxplot).
```

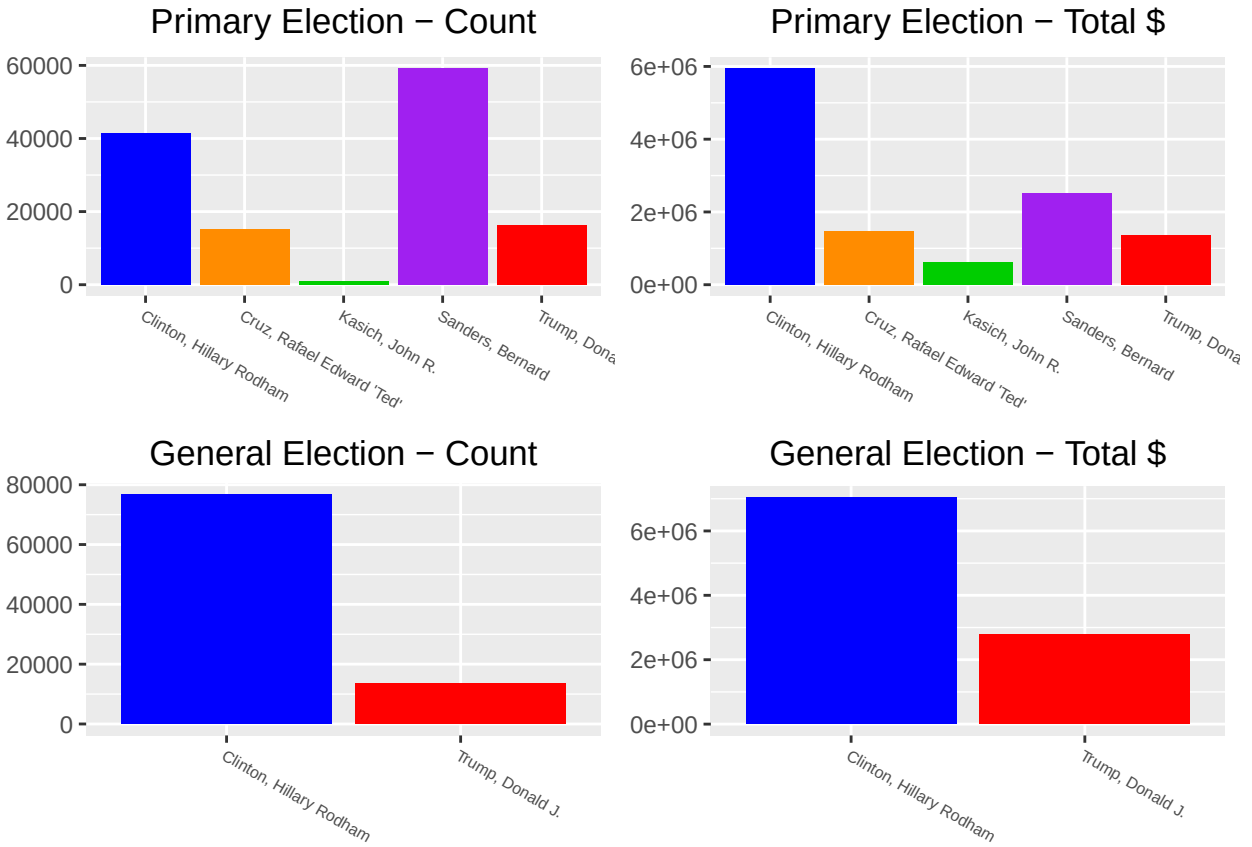


It was hard to tell much from the boxplot. The median contribution for the most successful candidates – Trump, Clinton, Sanders, and Cruz – was indeed lower than most other candidates. But that didn't help distinguish *between* Trump, Clinton, Sanders, and Kruz.

If we zoomed in on the most successful candidates, would we start to see that campaign contributions did predict success?

Does the number of contributions or total amount of contributions matter?

When we started, I noted that news coverage about *national trends* in campaign contributions made them seem like a lousy predictor of a candidate's success in 2016. If we focus on only Pennsylvania, will we find that campaign contributions are a better predictor at the *state level*?



So, when we look at Pennsylvania *state trends* in campaign contributions, do campaign contributions seem like a better predictor? Not really.

In Pennsylvania, Sanders received about 50% more contributions than Clinton, and he still lost the primary.

Clinton received several times the number and sum of contributions as Trump, and she still lost the general election.

If the counts and sums of contributions don't tell us much, can we learn anything from the timing of those contributions?

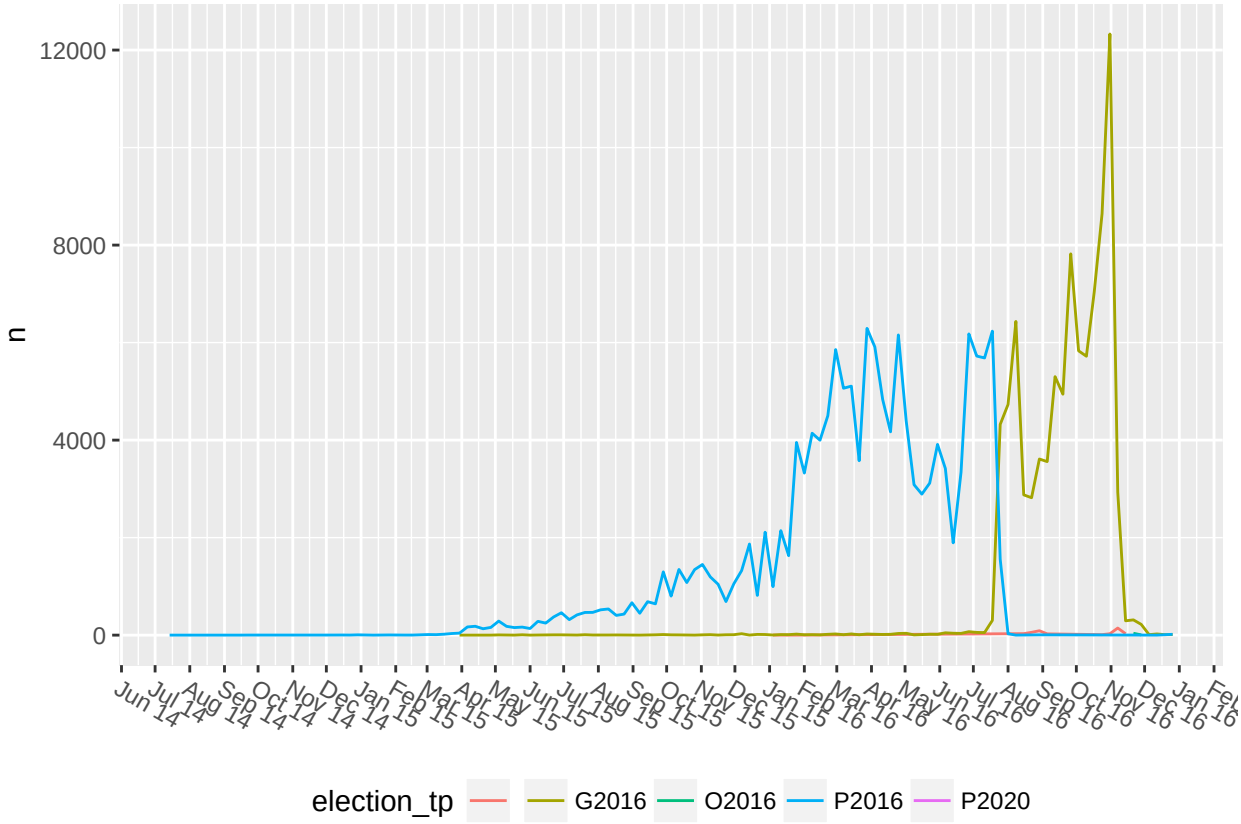
Does the timing of contributions matter?

Before delving into the data, I wanted to refresh myself on the relevant dates and get a baseline for *all* contributions.

Event	Date
First Caucus (Iowa)	February 1, 2016
Pennsylvania Primary	April 26, 2016
Last Primary for Republicans (South Dakota)	June 7, 2016
Last Primary for Democrats (D.C.)	June 14, 2016
Republican National Convention	July 18, 2016

Event	Date
Democratic National Convention	July 28, 2016
Election	November 8, 2016

Source: website of primary dates and just google results pages for other dates

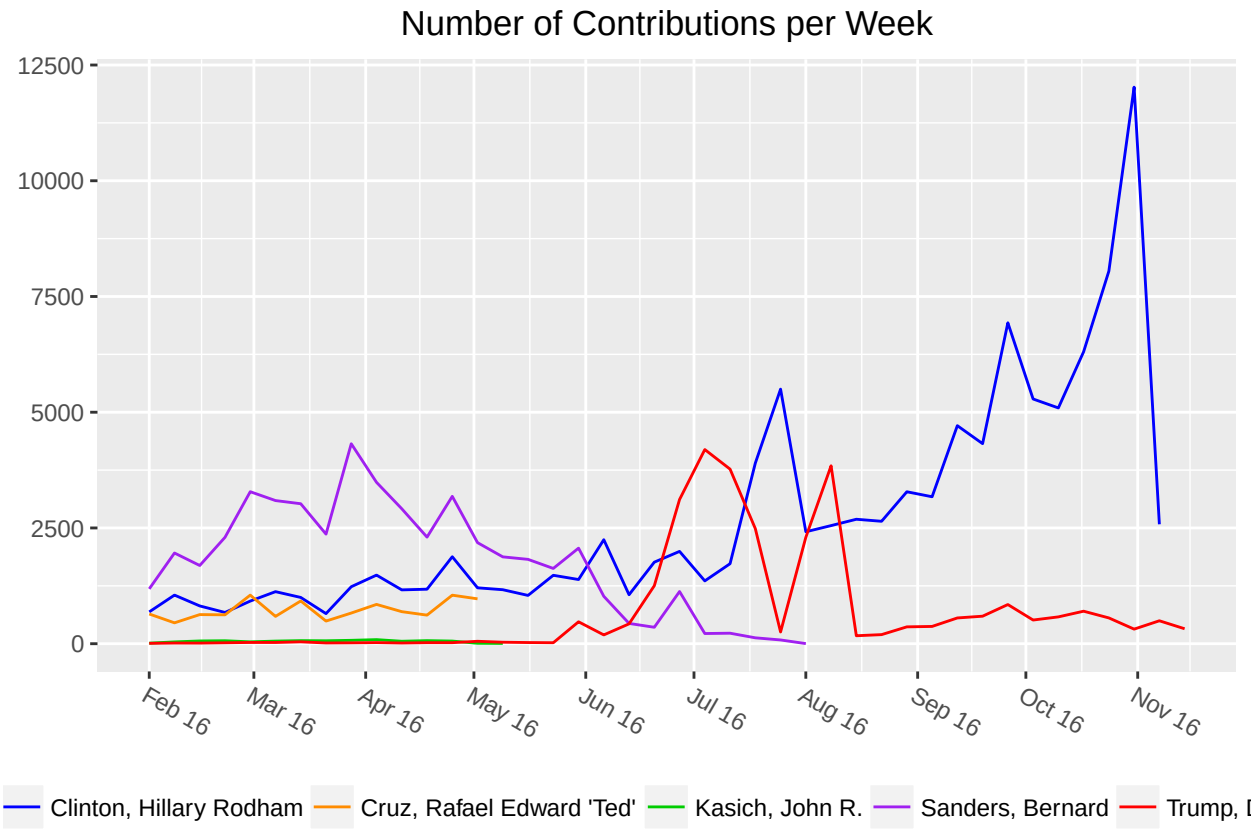


Contributions to the primary didn't pick up until early 2016, then dipped around May and June, 2016 – immediately following the Pennsylvania primary. Then they jupmed again in July, when both major parties held their conventions.

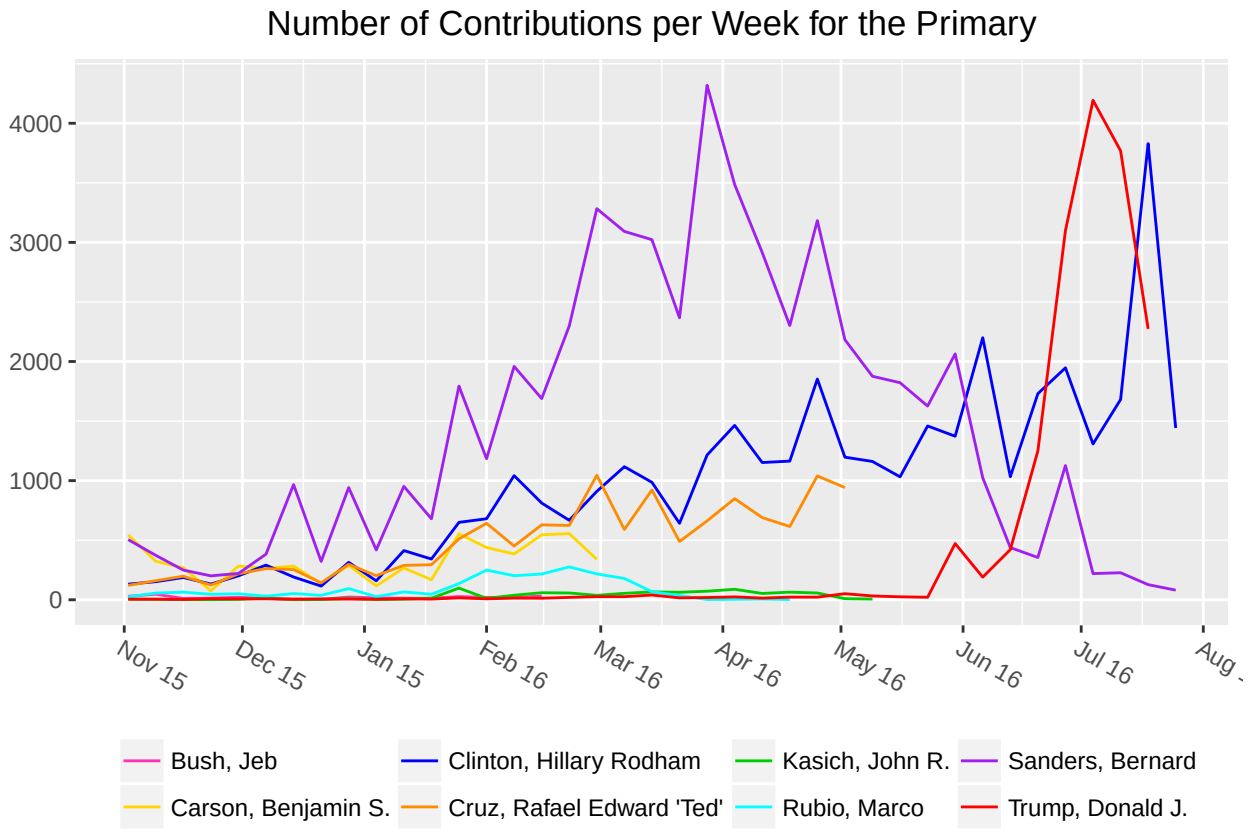
After the conventions, contributions switched to the general election. These peaked in early November, near election day.

There's nothing too surpring here. However, it's important to note that several "primary" contributions actually occured after the primary elections, meaning they couldn't have been useful in predicting the primary results.

So now we've got a baseline. But what does looking at contributions to *individual* candidates tell us about their odds of success?



This graph already says a lot, but I want to hone in on the primary and general election, so I'm going to limit the time frames.

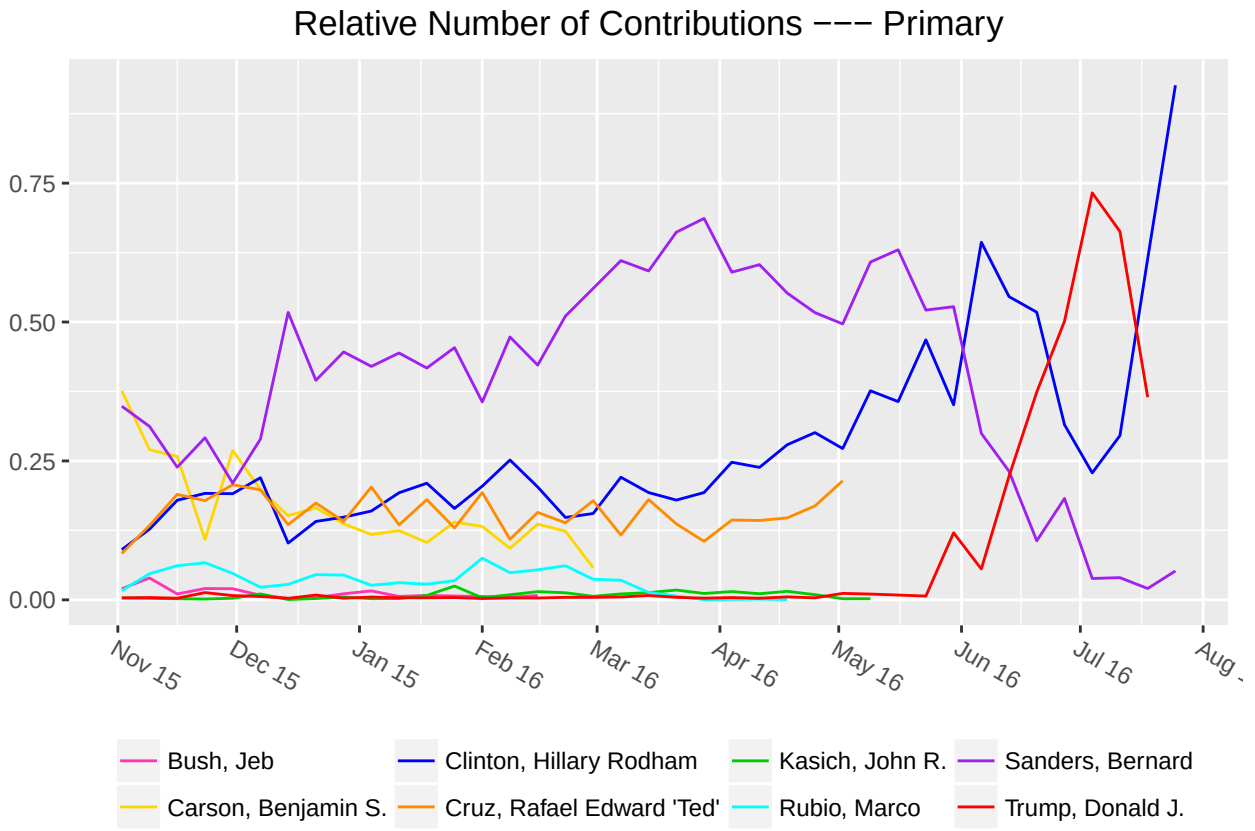


So does the *timing* of contributions seem like it could be a predictor of success for the primaries? No.

If anything, the timing makes the number of contributions look like an even worse predictor of success, since Clinton and Trump received many "primary" contributions in the lead-up to their party conventions, after they were already the clear winners. If we ignored those late-stage contributions, Sanders' lead over Clinton and Cruz's lead over Trump would be even more pronounced.

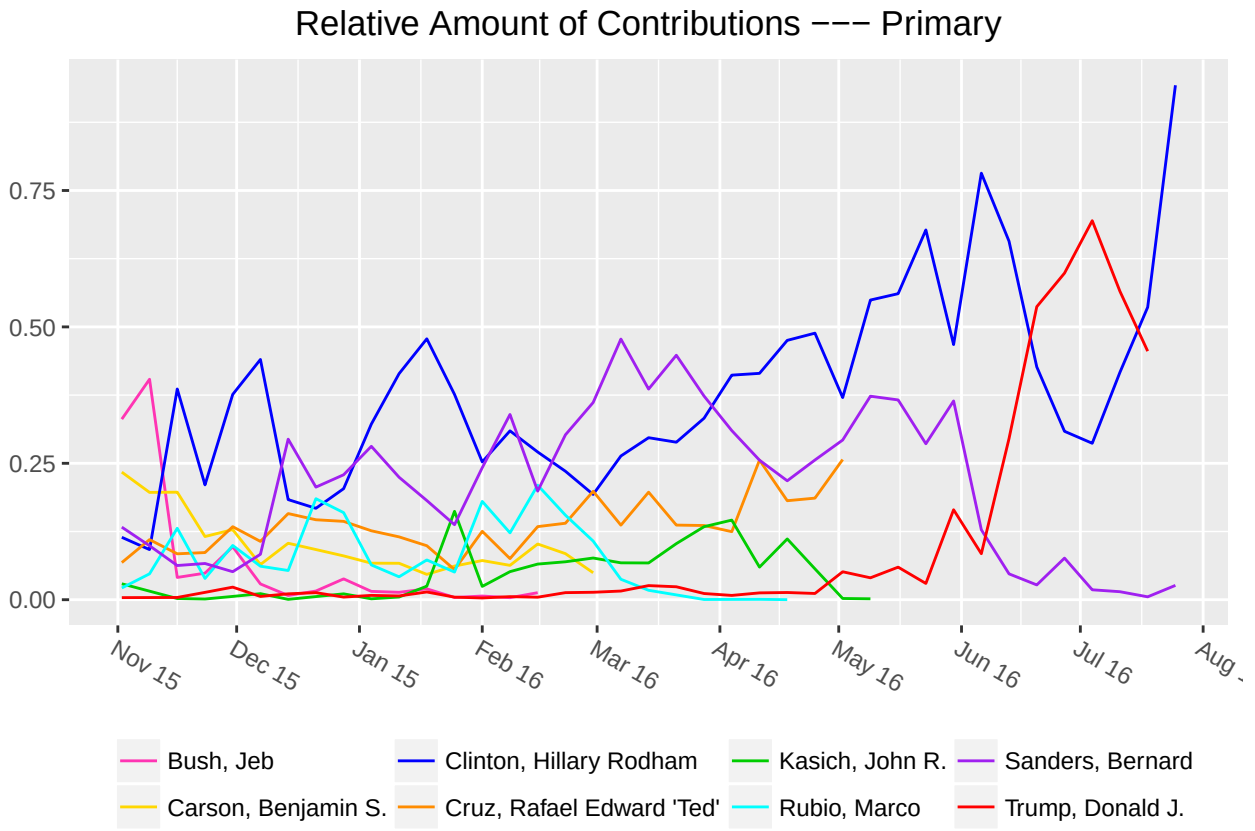
Trump's contributions didn't even climb noticeably above zero per week until ***after*** *the Pennsylvania primary.*

I thought that plotting the *relative* number of contributions would make these trends even more noticeable, by eliminating overall trends in contributions.



Looking at the relative numbers, it's clear that Sanders received about **twice** as many contributions per week as Clinton from late 2015 to May 2016. But, again, those contributions didn't mean he would win.

It's possible that the timing of contributions might still be important, but that we should be looking at contribution amounts. For example, if a candidate received a greater amount of contributions closer to the election, maybe that indicated support for their campaign was growing. What do the relative amounts of contributions tell us about the growing or waning support for a campaign?

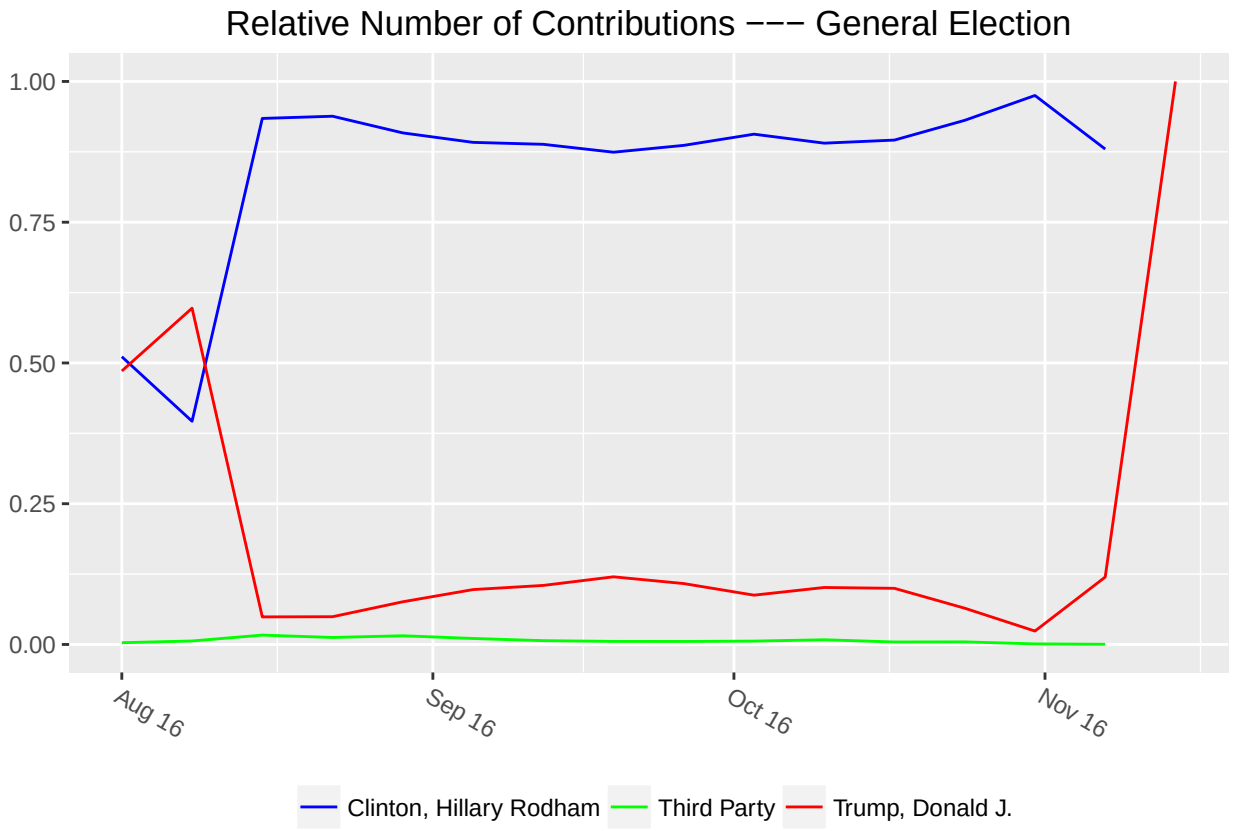


When we look at the timing of when contribution amounts were received, do we see a different story? No – Sanders received a greater amount of contributions than Clinton in the months immediately before the primary. The same is true for Cruz versus Trump.

There's another possible interpretation of timing. Clinton received a higher amount of contributions for most of November 2015 through February 2016. We might tentatively posit that receiving greater amounts when Clinton did – *before* primaries began – was a predictor of later success.

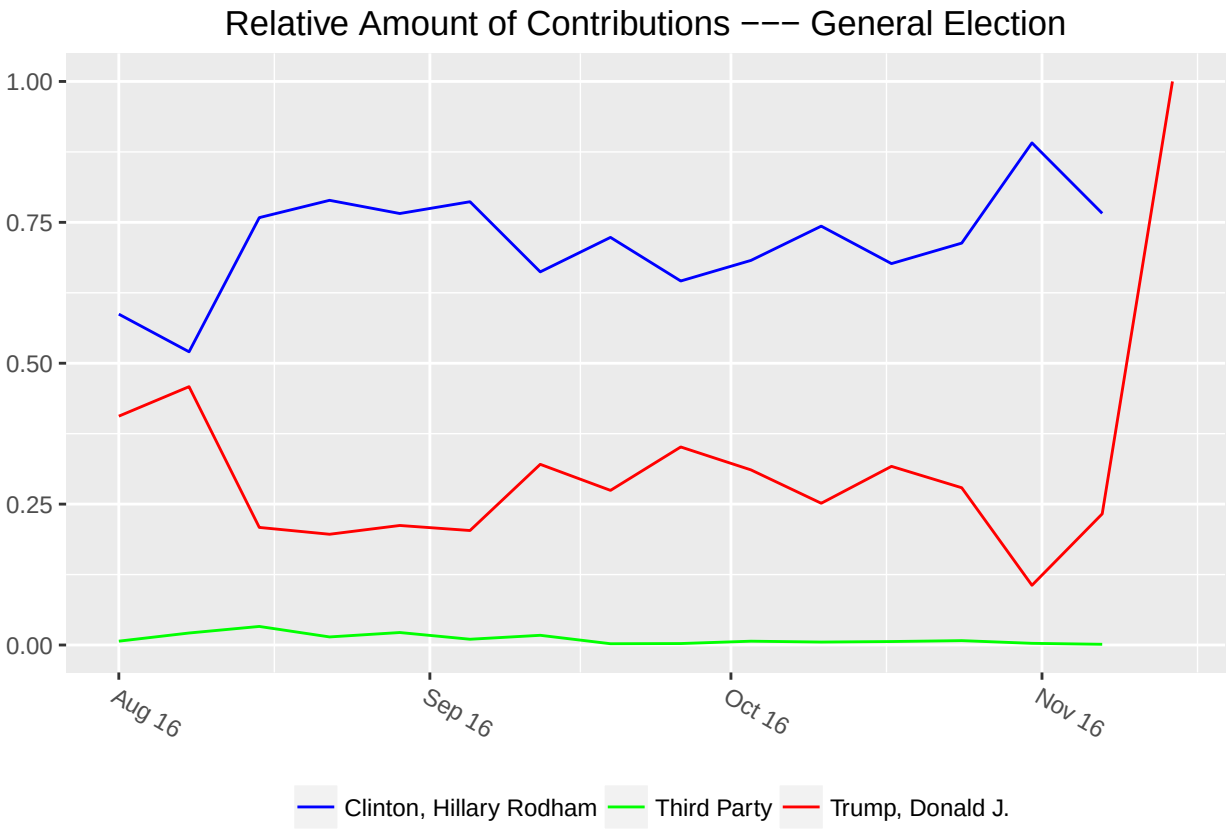
But that doesn't hold true on the Republican side, where Cruz, Carson, and Rubio all led Trump for late 2015 and early 2016.

Is it possible that the timing of contributions isn't a good predictor during the primaries, but that it's a better predictor of the general election?



Clinton absolutely dominated Trump in the number of contributions received per week from mid-August to the election day. It'd be difficult for there to be a clearer indication that the number of contributions did not predict the winner.

As for the amount of contributions?



The gap in contribution amounts was consistently closer, but Clinton still received more than twice – and sometimes three times – the amount of contributions as Trump.

So it didn't matter *when* a candidate received their contributions. Did it matter *who* was giving the contributions?

Do the occupations of contributors matter?

Perhaps some contributors more reliably indicate the eventual winner. Some people might be more influential in an election's outcome, or a certain group of people might be a good bellweather for how their neighbors will vote.

I didn't have a detailed profile on each contributor, but I did have information about their occupation and their geographic location. I decided to first look at contributors' occupations.

Which candidate did better with nurses? With CEO's? And were contributions from one group possibly more predictive than support from another group?

To answer these questions, I wanted to look at the number of unique contributors with each occupation. If John, a truck driver, contributed four times, I still only wanted to count him as one truck driver.

To identify unique contributors, I decided that two contributions that shared the same contributor name, zip code, and occupation were probably from the same person.

```
## [1] 61381
```

Woah! There are only 61,381 unique contributors. That's far less than I expected, given that there were more than

```
## [1] 241028
```

241,000 positive-value contributions. This is worth investigating more!

But first, I still wanted to look at occupations.

There's a lot to unpack here.

For one, Kasich's main supporters were executives, presidents, and CEOs – though since he didn't lead among these occupations, Kasich's loss doesn't say anything about the influence of these wealthy contributors.

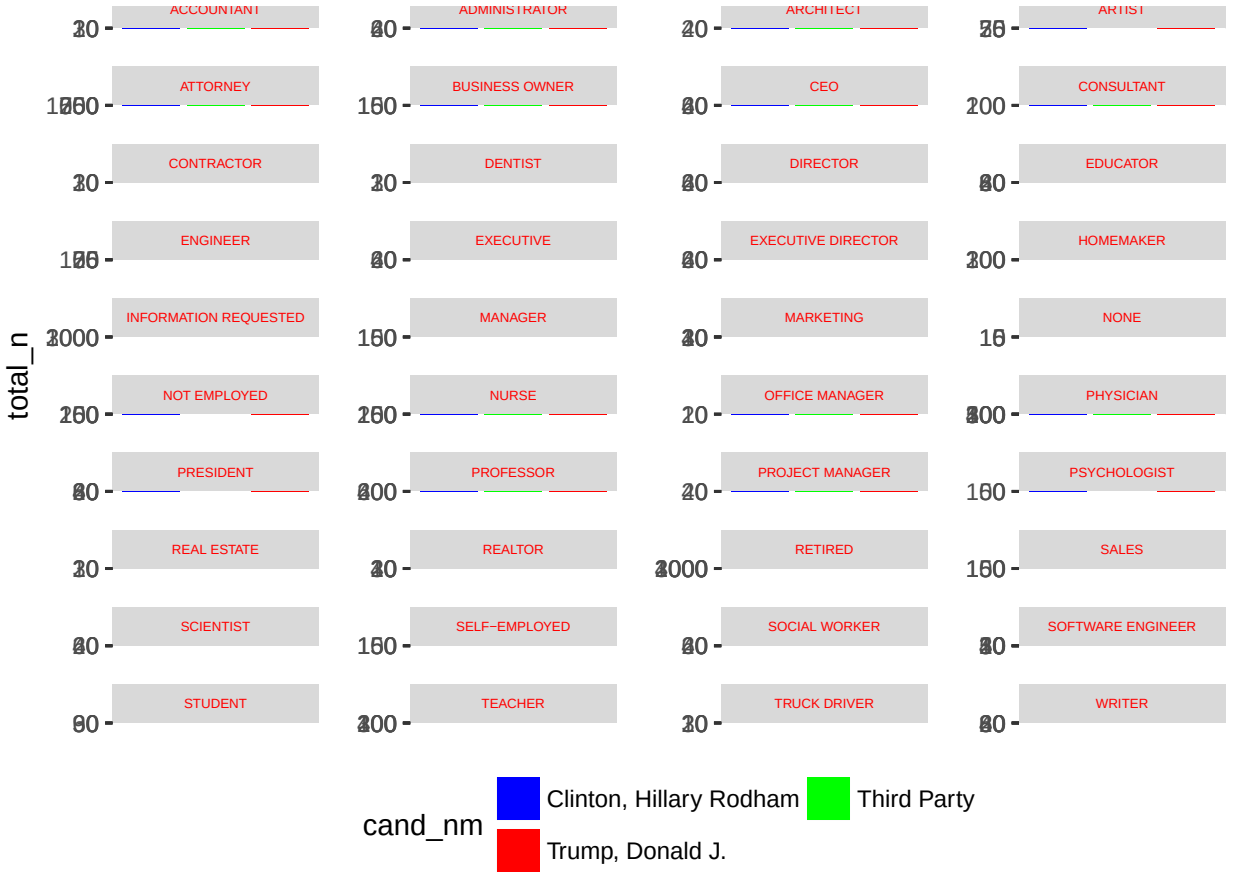
There were two occupations that are difficult to interpret – why did some people list their occupation as "None," and why were these mostly Cruz contributors? And why did Trump receive the most contributors for whom their occupation was "Information Requested"? It's possible these people were unemployed, but it's also possible these people chose not to disclose their occupation, which begs the question *Why?* The dataset doesn't have enough information to answer this question.

As for Sanders, he fared well among contributors of several occupations, many of which also favored Clinton. There were a few exceptions: Sanders and Trump split the support of engineers, and Sanders led all the candidates among software engineers and the not employed.

Compared to Sanders, Clinton had an uncontested lead among several more common occupations. This is really the first time that Pennsylvania's campaign contributions begin to display some predictive power. But is that because we're looking at occupations, or because we're looking at **unique** contributors?

The same question arises when we look at Trump — who either led or held even with the other Republicans in almost every category.

But before looking at unique contributors more generally, let's look at unique contributors to the general election.



When shifting focus from the primaries to the general election, we can ask an interesting question: *Did Sanders' supporters switch to Clinton or Trump?* Among occupations where Sanders clearly led in the primary — software engineers and the not employed — Clinton now leads Trump. These results are suggestive, but certainly not definitive, since we're not necessarily tracking the same individuals.

Then there's the larger question: *Could focusing on contributors' occupations have helped us predict the general election outcome?* It's hard to say, but we can rule out some things about occupations being predictive:

Clinton led among several occupations that might be called intellectuals or well-educated – such as attorneys, scientists, students, professors, nurses, physicians, teachers, educators, software engineers, and writers – and the support of the well-educated did not translate into a win.

The not employed and difficult-to-understand "None" category also overwhelmingly supported Clinton — and these groups' support did not translate into electoral success.

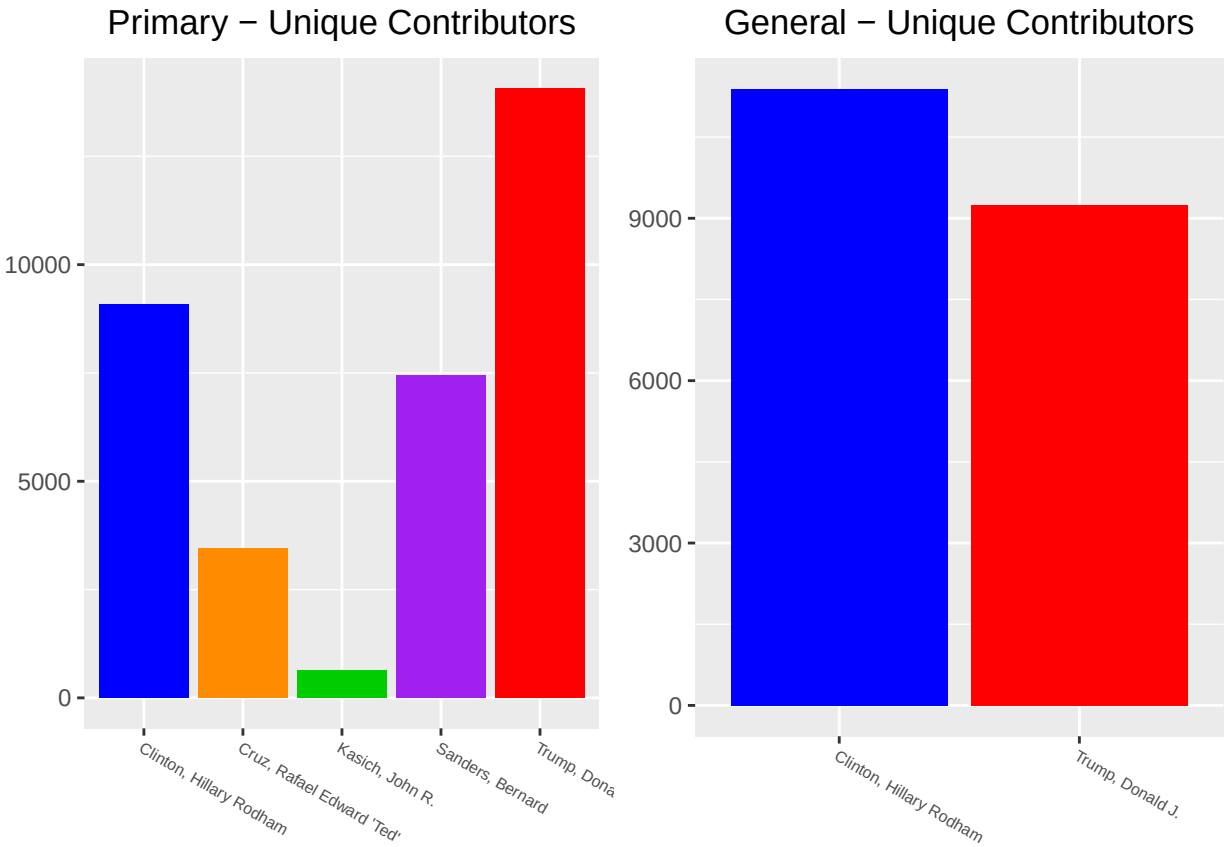
It becomes harder to make even tentative conclusions when looking at the occupations of Trump's supporters. Trump led among some blue collar workers — like truck drivers and contractors — but we can't say these supporters predicted his success, only that we can't rule it out.

But let's now turn to a more promising question. Since about...

```
## [1] 0.7453366
```

75% of contributions were made by repeat contributors, what happens when we begin looking at the number of *unique contributors*, not the number of contributions?

Does the number of unique contributors matter?

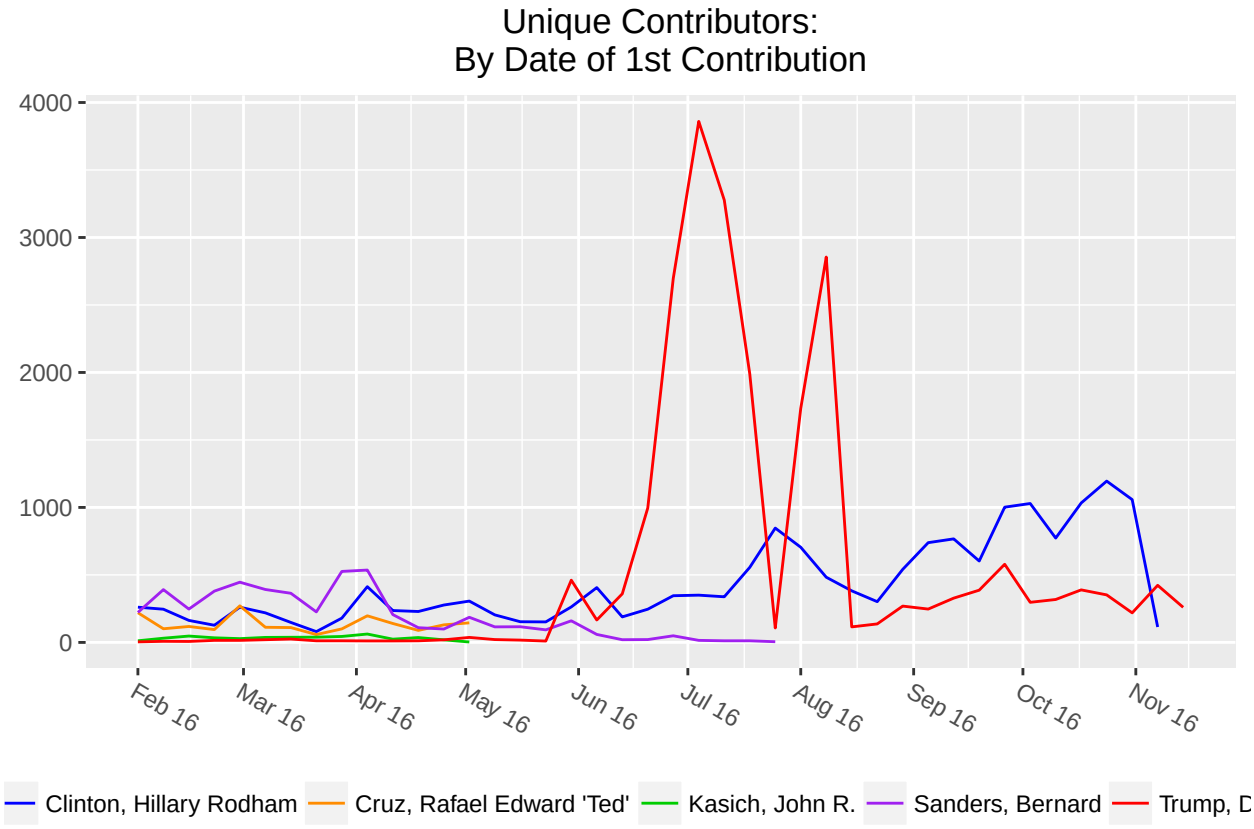


Looking at unique contributors, we begin to see the possibility that campaign contributions might be predictive of a candidate's success!

Clinton had more unique contributors than Sanders in the primary, and she won PA's democratic primary. Trump had more unique contributors than Kasich or Cruz in the primary, and he won PA's Republican primary. Unique contributors looks like it might be predictive of who's going to win a party's primary.

But the predictive power falls away when we look at the general election. Although Clinton's lead has narrowed — as compared to contribution numbers or total dollar amounts — she still had more unique contributors in the general election and lost to Trump.

But what if we look at the timeline of when unique contributors first gave to a campaign? Will we find a way to fine tune our analysis?

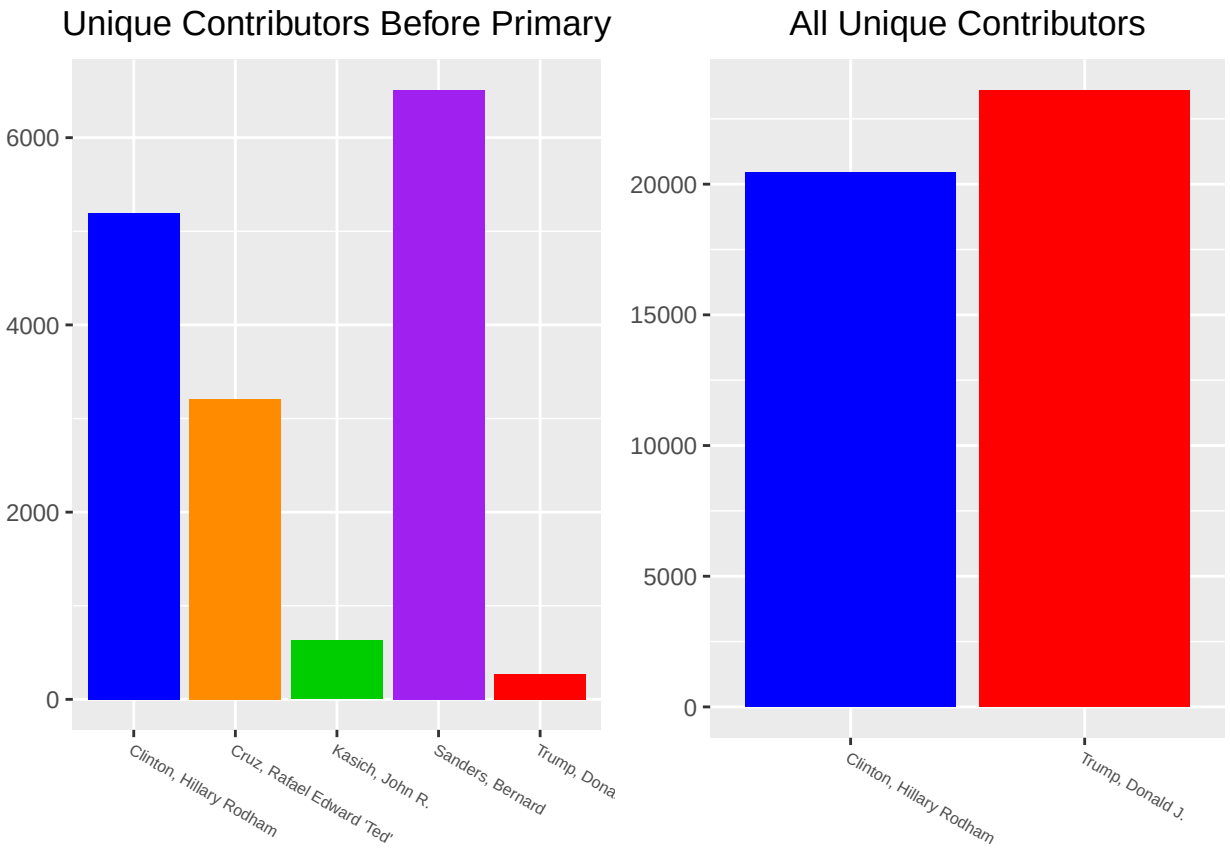


Now that we're looking at unique contributors, we get a **VERY** different perspective on campaign contributions during the presidential campaign.

Trump had two huge spikes in the number of first-time contributors in July, before the Republican National Convention, and in August, *after* the convention. Maybe when considering whether first-time contributors might predict the *general election* winner, we should look at unique contributors during both the primary and general election.

Also of note, Cruz was beating Trump in first-time contributors until the beginning of May, right after the Pennsylvania primary. If first-time contributors were as predictive of a *primary's* winner, shouldn't we focus on contributions before the primary took place? Otherwise, we're just seeing contributors **react** to the winner.

What happens when we incorporate these lessons into the bar graphs above?



Now, it seems like our most recent — and tentative — hypotheses have been turned upside down.

The number of unique contributors was NOT predictive of a campaign's success in the primary. Before the PA primary, Sanders had more unique contributors than Clinton, and he still lost the primary. Before the PA primary, Cruz had many times more unique contributors than Trump — and even Kasich had about twice as many contributors as Trump — and both Cruz and Kasich lost.

On the other hand, the number of unique contribtors may *possibly* have been predictive of the general election results. Trump had more unique contributors than Clinton throughout the *entire* election season.

This is the first time that some portion of campaign contribution data might be useful in predicting general election results! Certainly, this is tentative, but it's still a potentially useful result that's worth exploring more.

Now let's look at one more variable: location. Much has been made of the divide between rural and city voters. What happens when we look at these voters using our newfound awareness of the potential predictive power of unique contributors?

Do the locations of contributors matter?

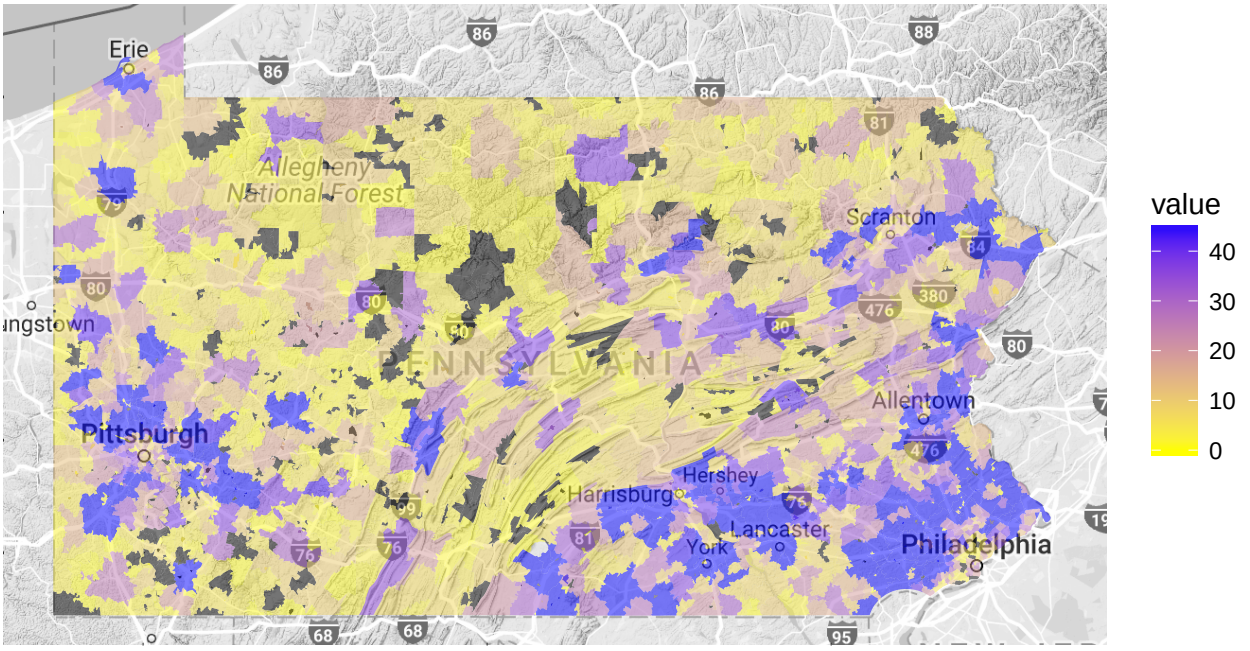
Before answering this question, I needed to do a little data cleaning. When I used zip codes from the campaign contributions to look up latitude, longitude, city, and state; I was surprised to find that some zip codes were not in Pennsylvania.

```
## [1] 27
```

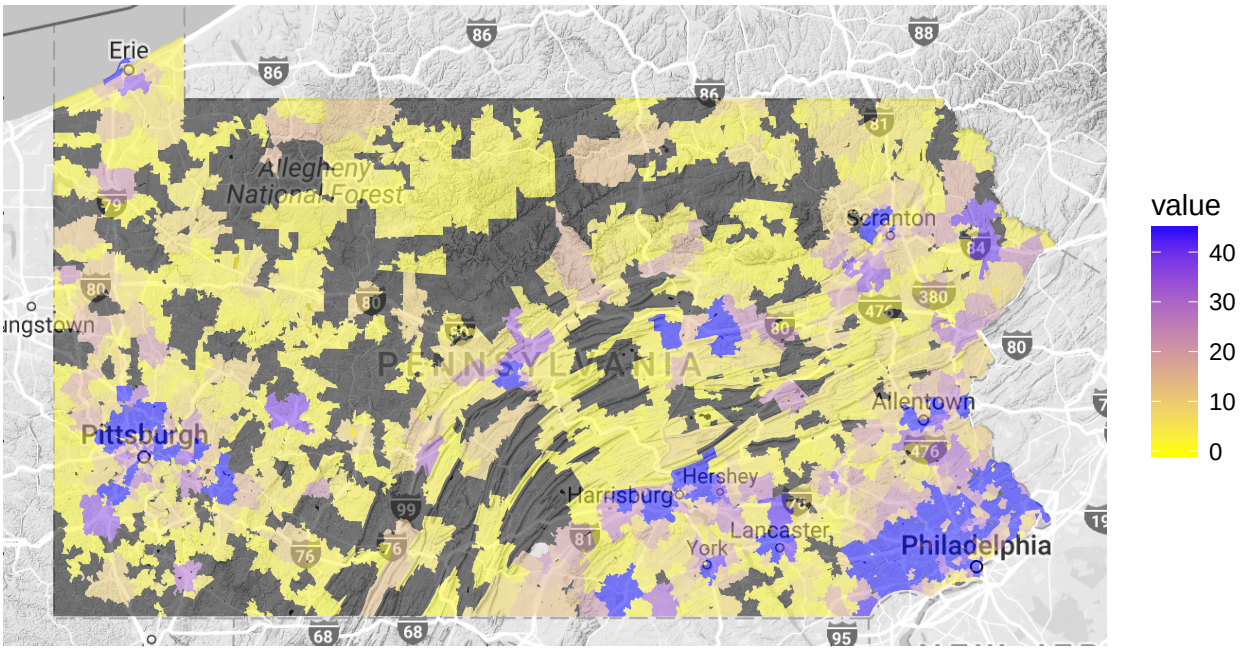
Since there were only 27 unique contributors outside Pennsylvania, out of about 61,000 contributors, I decided to simply exclude these contributors and continue with the analysis.

To map contributions, I used the choroplethr package. It was mostly straight-forward, but this stack exchange post was particularly useful in helping understand choroplethr's zipcode functionality.

Trump: Unique Contributors



Clinton: Unique Contributors



A couple things to note about the above choropleths: (1) zip codes with no contributors are black and (2) some zip codes have values above 45 — however, those zip codes have the same color scheme as counties with 45 contributors

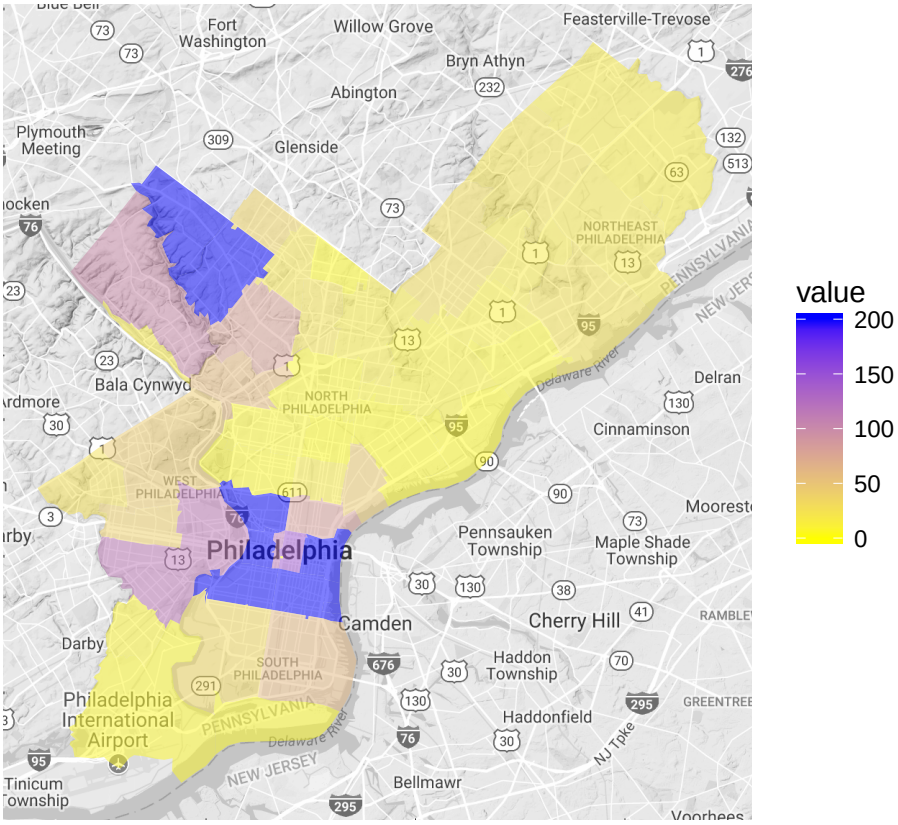
Now that that's out of the way, what do the above maps tell us about rural Pennsylvanians versus city Pennsylvanians?

As for rural areas, Trump certainly had more contributors than Clinton. However, Clinton had a small

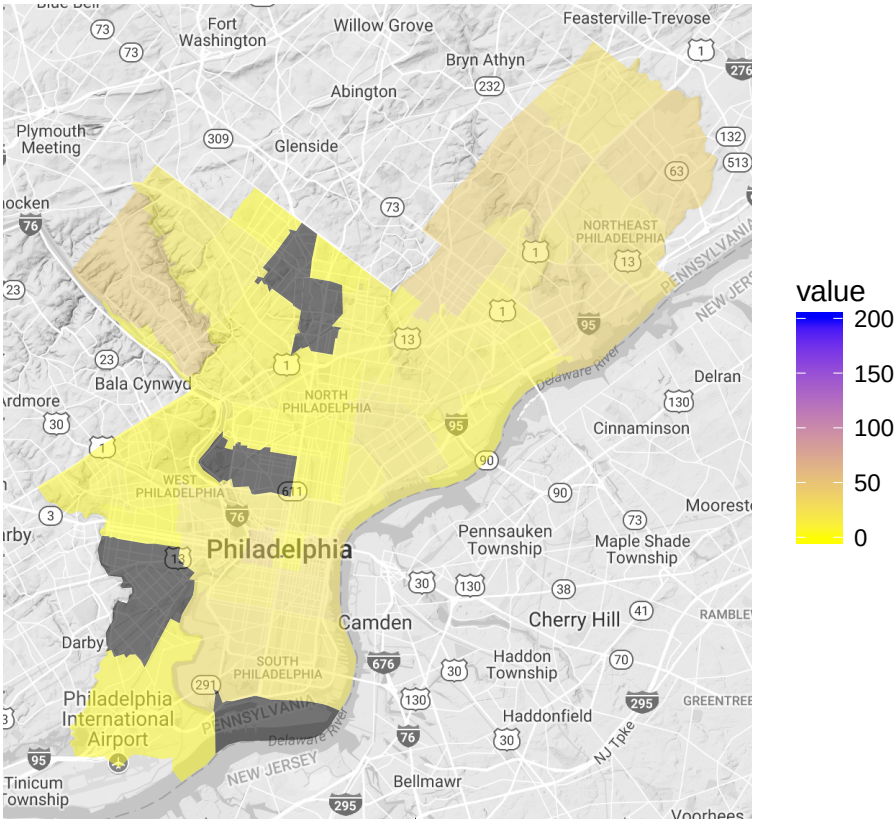
number of contributors in many rural zip codes.

As for cities, especially Philadelphia, the above maps don't helps us distinguish between the candidates. Because there were so many more contributors in urban zip codes, in order to see differences in rural areas, I needed to limit the max value on the scale. What can we learn about contributors in Philadelphia by focusing on zip codes in Philadelphia and changing the scales?

Clinton: Unique Contributors



Trump: Unique Contributors

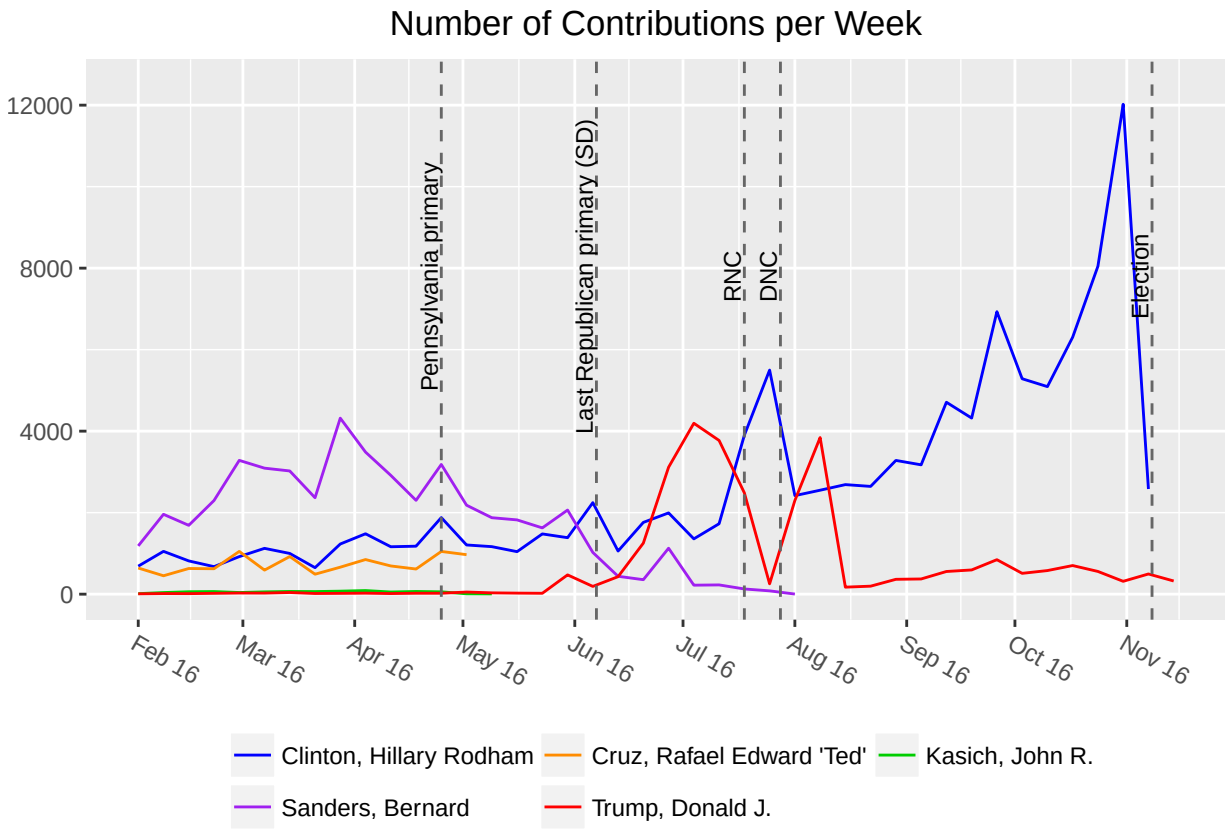


Looking at unique contributors in Philadelphia, we find the converse of rural areas. Clinton certainly had more contributors than Trump, though Trump had a small number of contributors throughout most of the city.

The numbers of unique contributors in different parts of a state seems like a possible predictor of how voters in those areas will vote.

Final Plots and Summary

After exploring campaign contributions to the 2016 presidential campaigns in Pennsylvania, one of the most important take-aways is how potentially misleading contributions can be as a predictor of candidate's success.



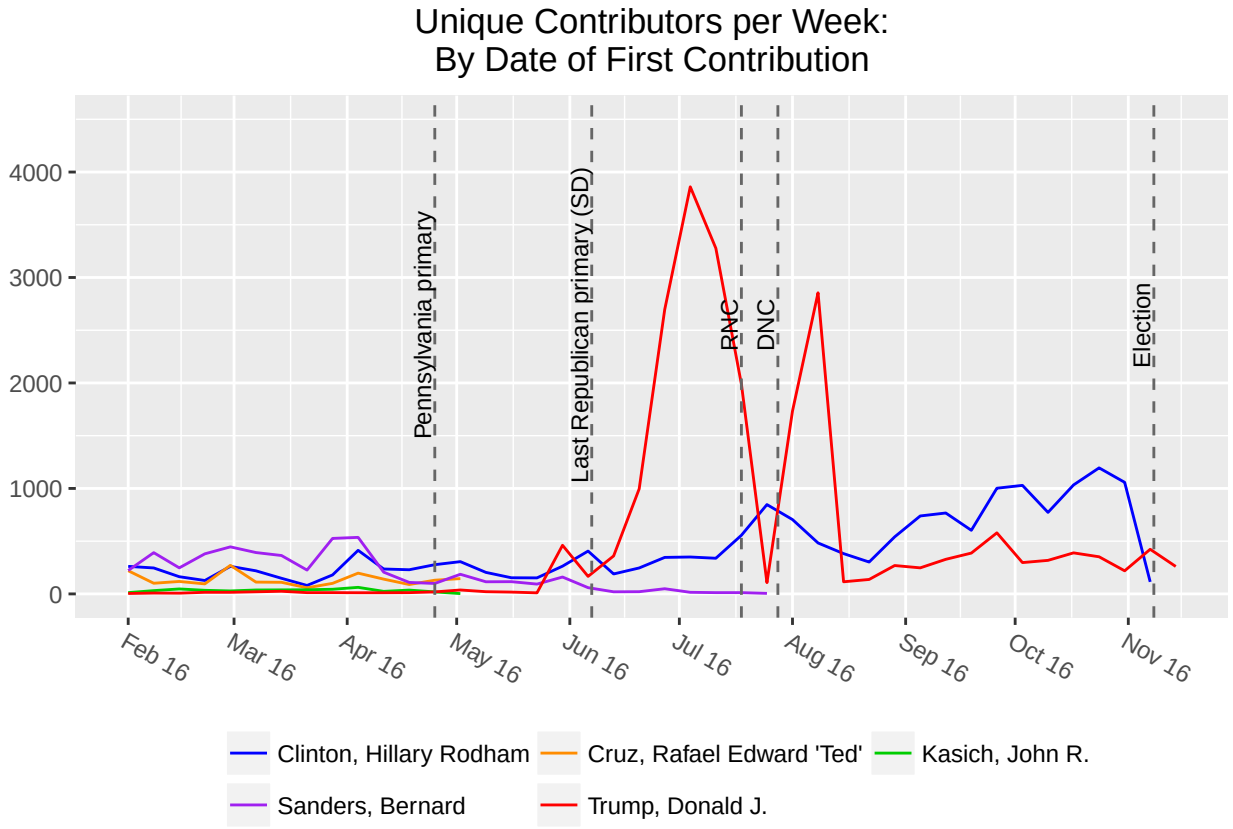
The above plot shows the number of contributions received by a candidate during each week of the election season. If we relied on the number of contributions as a predictor of a candidate's success, we would have called every race wrong:

Sanders led Clinton in contributions before the primary, but Clinton prevailed. Kruz led Trump in contributions before the primary, but Trump prevailed. Most strikingly, Clinton received many times more contributions per week than Trump during most of the primary and general election season, but Trump prevailed.

Data can sometimes be used to predict the future, but it's just as important to uncover how data can mislead us about the future. In upcoming elections, we should perhaps be wary of media reports that suggest a correlation between the number of contributions received by a candidate and that candidate's odds of success.

Of course, this tentative conclusion relies a single state's dataset during a single election – an election which many observers would consider unusual. We cannot rule out that the number of contributions received by a candidate is generally a good predictor of success, but we should be skeptical of that relationship pending analysis of more datasets.

However, another metric of campaign contributions seems to hold promise as a predictor of candidate's success.



The above plot shows the number of unique contributors to a candidate during each week of the election season. If a paricular contributor gave to a campaign more than once, only their first contribution is included in the plot.

By focusing on unique contributors, an entirely different picture of Trump's campaign emerges. Though Trump received fewer contributions, many more of those contributions came from unique contributors.

The number of new contributors giving to his campaign peaked twice: (1) after Trump had secured his party's nomination and (2) after the Democratic National Convention.

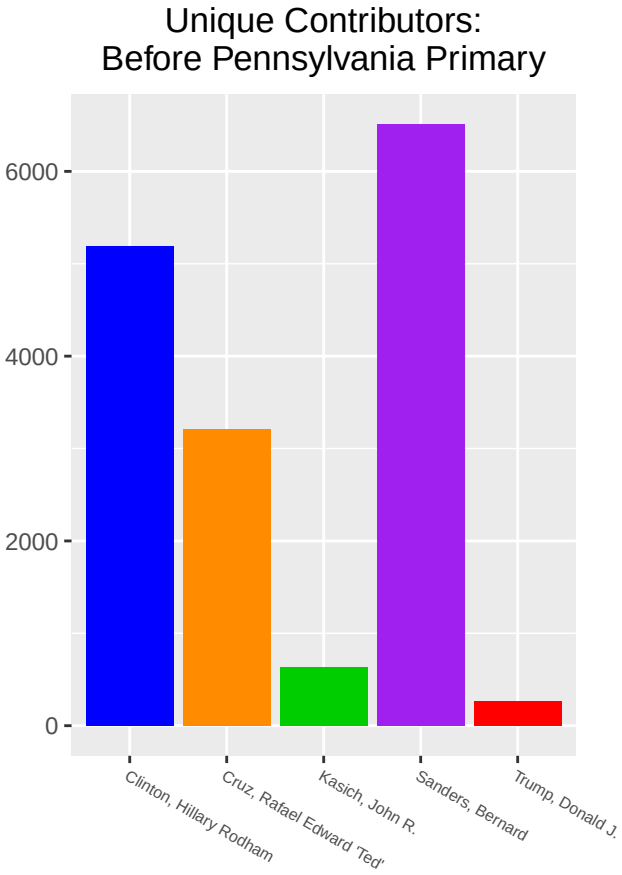
We don't know what motivated so many first-time contributors to give at these specific times. Were they excited that they're preferred candidate had secured the Republican party's nomination? Were they dismayed that a candidate they disliked had secured the Democratic party's nomination? But we can say that the peak in July resulted from contributors reacting to the identities of the presumptive nominees, rather than contributors trying to influence the primary.

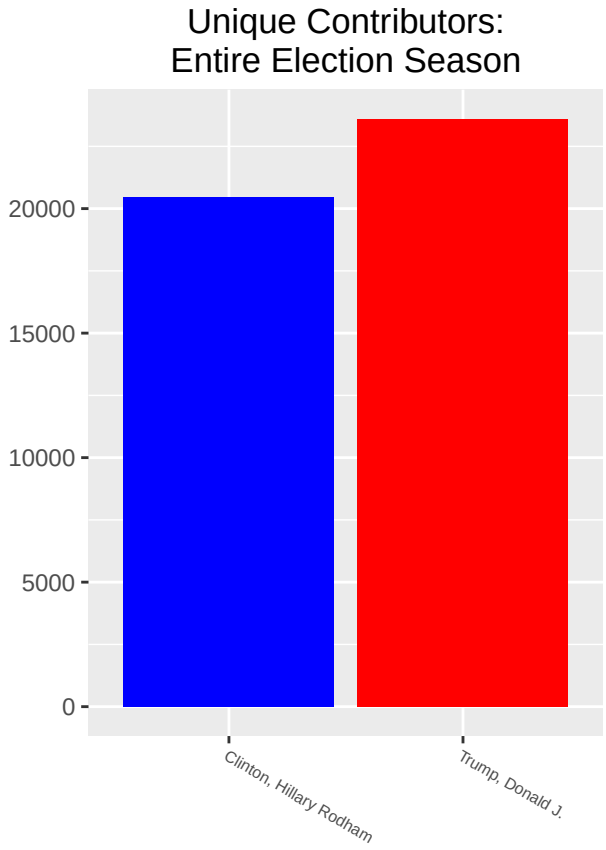
We can also make some inferences based on the weeks when new contributors to Trump did not peak.

First, in mid-August, the number of new contributors to Trump fell off, and he then trailed Clinton until the election in November. Mostly, the fall-off is notable for what it didn't signify — it didn't signify that support for Trump had disappeared. Since Trump ultimately won the election, we can hypothesize that the number of unique contributors during an election season matters — and that the exact timing of their contributions doesn't matter.

Second, given the huge peaks for Trump's campaign in July and August, it's easy to miss that his campaign had very few new contributors before Pennsylvania's primary in late April.

That's what brings us to our final, simplest plot.





The number of unique contributors to a candidate does **not** say much about that candidate's likelihood of success in a state's primary. Here, again, Sanders led Clinton before the Pennsylvania primary, and Clinton prevailed in the primary. Cruz and even Kasich led Trump, and Trump prevailed in the primary.

However, the number of unique contributors to a candidate may possibly be an indicator of that candidate's success during a general election. In this one area, Trump led Clinton, and ultimately Pennsylvanians selected Trump during the November presidential election.

I can't stress enough: we cannot conclude that the number of unique contributors can predict election results. This is one data point, based on looking at one state during one election.

But it's also the most interesting data point here.

When I began exploring this dataset, campaign contributions seemed like a lousy way to predict the likelihood of a candidate's success. Yet there's a chance that campaign contributions might have some predictive power after all. To investigate whether that's the case, the next step would be to investigate multiple elections in multiple states, and then to test whether there is a correlation between the number of unique contributors to a candidate and that candidate's likelihood of success.