An Analysis of California Political Contributions to the 2012 Presidential Election by David Pankiewicz

In this analysis, I looked to see what were some of the common trends in regard to amounts of donations and total money donated based on other factors in the data.

First, I wanted to get a basic understanding of my dataset:
* How many observations am I working with?
* What types of data do I have and what are some basic summaries about this data?

```
## [1] 9503   23

##  [1] "committee.id"        "candidate.id"        "candidate"
##  [4] "name"                "city"                "state"
##  [7] "zip"                 "employer"            "occupation"
## [10] "amount"              "date"
"receipt.description"
## [13] "memo.code"           "memo.text"           "form.type"
## [16] "file.number"         "transaction.id"      "election.type"
## [19] "democrat"            "democrat.amount"     "republican"
## [22] "republican.amount"   "all.donations"

## 'data.frame':    9503 obs. of  23 variables:
##  $ committee.id       : Factor w/ 14 levels
..: 3 3 3 3 2 3 2 3 3 3 ...
##  $ candidate.id       : Factor w/ 13 levels
..: 12 12 12 12 13 12 13 12 12 12 ...
##  $ candidate          : Factor w/ 2 levels "Obama, Barack",..: 1 1 1
1 2 1 2 1 1 1 ...
##  $ name               : Factor w/ 9428 levels "ABADIR, BUSH Y.",..:
1367 2606 232 4194 1854 6626 8768 8728 4284 2955 ...
##  $ city               : Factor w/ 792 levels
.: 313 713 601 272 56 401 15 365 157 616 ...
##  $ state              : Factor w/ 1 level "CA": 1 1 1 1 1 1 1 1 1 1
...
##  $ zip                : int  919415722 93561 924231745 94547
30 926915939 910012606 900261208 95014 951483131 ...
##  $ employer           : Factor w/ 3927 levels "","[24]7 INC",..:
4 1984 3329 2451 2826 433 2770 3032 2134 1968 ...
##  $ occupation         : Factor w/ 2571 levels "","  TEACHER/ HOSPITAL
WORKER",..: 1442 2287 2007 1241 826 2053 1912 471 351 725 ...
##  $ amount             : num  25 35 14 25 35 10 30 14 19 50 ...
##  $ date               : Date, format: "2012-11-04" "2012-10-04" ...
##  $ receipt.description: Factor w/ 12 levels "","REATTRIBUTED",..: 1
1 1 1 1 1 1 1 1 1 ...
##  $ memo.code          : Factor w/ 2 levels "","X": 1 1 1 1 2 1 2 1 1
1 ...
##  $ memo.text          : Factor w/ 35 levels "","*","* EARMARKED
```

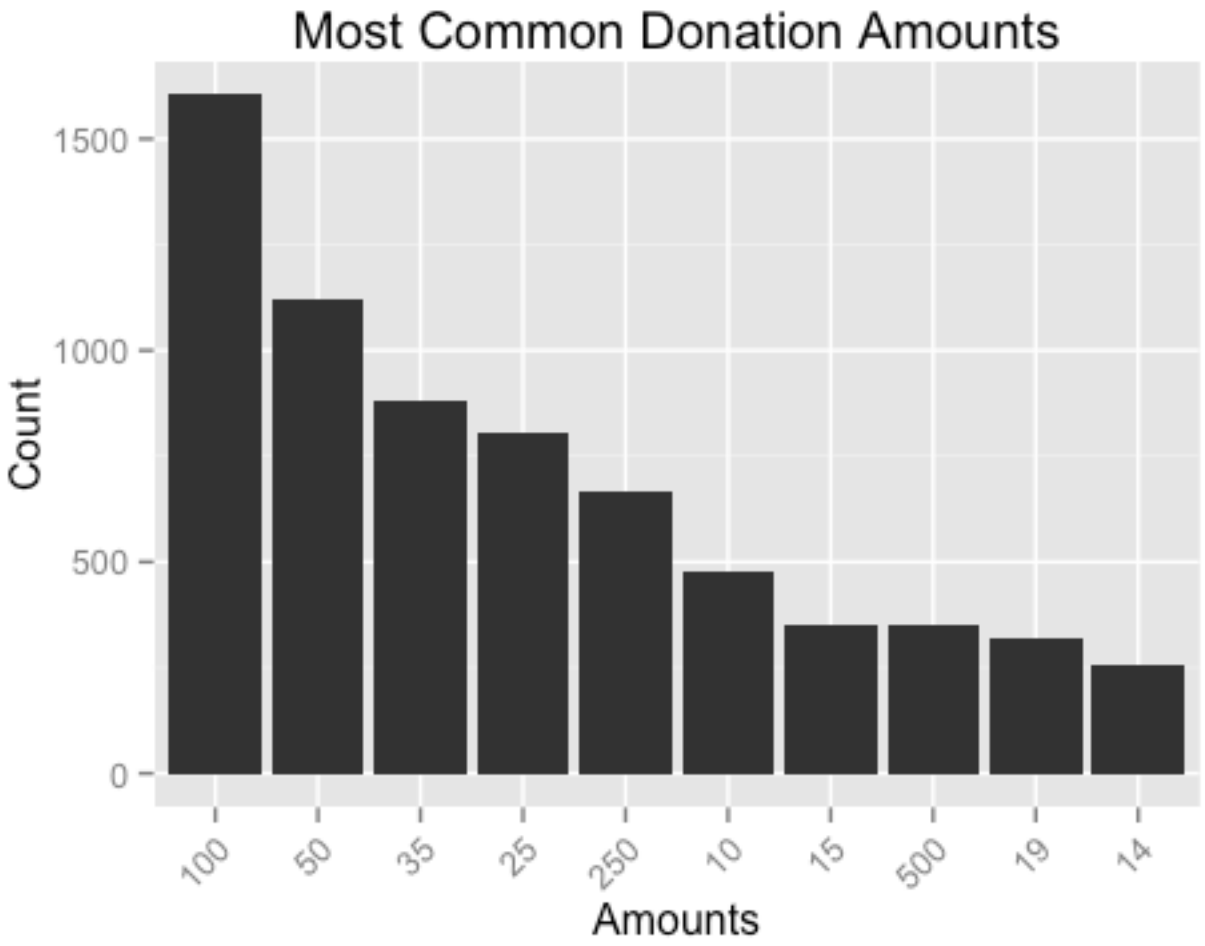
```
RIBUTION: SEE BELOW",..: 1 1 1 1 34 1 34 1 1 1 ...
##  $ form.type          : Factor w/ 3 levels "SA17A","SA18",..: 1 1 1
1 2 1 2 1 1 1 ...
##  $ file.number        : int  840327 846396 840327 846396 842943
27 933479 840327 806136 846396 ...
##  $ transaction.id     : Factor w/ 9997 levels "0850483-0001",..:
2 4727 6359 5634 9834 7494 9706 6671 1948 5209 ...
##  $ election.type      : Factor w/ 5 levels "G2008","G2012",..: 2 2 2
2 2 2 2 2 5 2 ...
##  $ democrat           : num  1 1 1 1 0 1 0 1 1 1 ...
##  $ democrat.amount    : num  25 35 14 25 0 10 0 14 19 50 ...
##  $ republican         : num  0 0 0 0 1 0 1 0 0 0 ...
##  $ republican.amount  : num  0 0 0 0 35 0 30 0 0 0 ...
##  $ all.donations      : num  1 1 1 1 1 1 1 1 1 1 ...

##     committee.id     candidate.id          candidate
##  C00431445:7471   P80003338:7471   Obama, Barack:7471
##  C00431171:2032   P80003353:2032   Romney, Mitt :2032
##  C00410118:   0   P00003608:   0
##  C00493692:   0   P20002523:   0
##  C00494393:   0   P20002556:   0
##  C00495622:   0   P20002671:   0
##  (Other)  :   0   (Other)  :   0
##                       name                 city      state
##  ADAM, MONIQUE          :   4   LOS ANGELES  : 722   CA:9503
##  SOBONYA, DOROTHY       :   4   SAN FRANCISCO: 565
##  VAN ZEE, ALI           :   4   SAN DIEGO    : 354
##  WESLEY, SANDI          :   4   OAKLAND      : 247
##  ACOSTA-RUBIO, MARX MR.:   3   SAN JOSE     : 218
##  ARGUELLO, JOSE         :   3   BERKELEY     : 176
##  (Other)                :9481   (Other)      :7221
##       zip                                            employer
##  Min.   :     2347   RETIRED                                :1979
##  1st Qu.:    94611   SELF-EMPLOYED                          :1370
##  Median :917115474   NOT EMPLOYED                           : 676
##  Mean   :627396820   INFORMATION REQUESTED                  : 214
##  3rd Qu.:941121342   INFORMATION REQUESTED PER BEST EFFORTS: 200
##  Max.   :961616629   (Other)                                :5062
##                      NA's                                   :   2
##                                    occupation       amount
##  RETIRED                                :2175   Min.   :   1.5
##  ATTORNEY                               : 298   1st Qu.:  25.0
##  TEACHER                                : 208   Median :  50.0
##  HOMEMAKER                              : 198   Mean   : 193.2
##  INFORMATION REQUESTED PER BEST EFFORTS: 195   3rd Qu.: 100.0
##  (Other)                                :6428   Max.   :5000.0
##  NA's                                   :   1
##       date
on
##  Min.   :2011-04-04                                        :9478
```

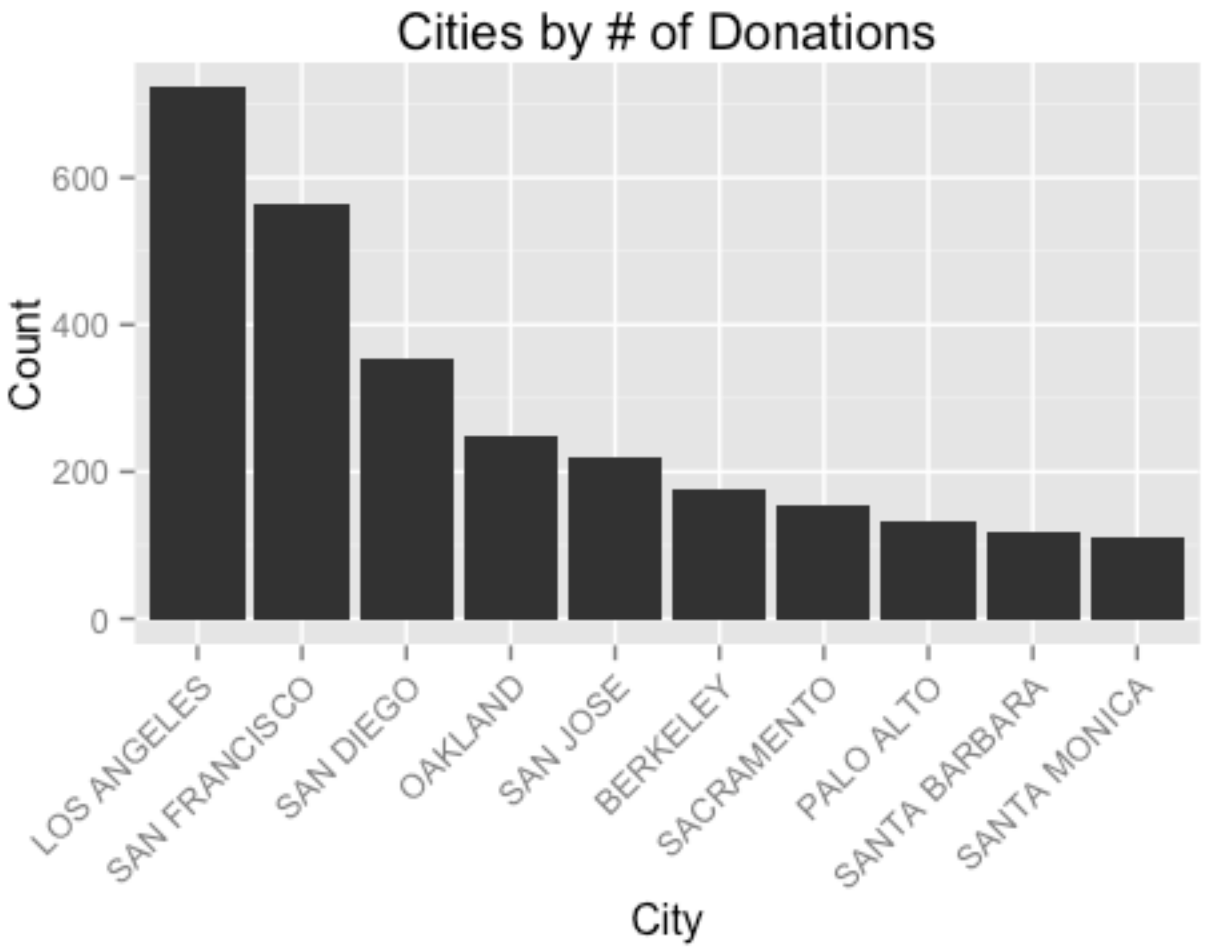
```
##  1st Qu.:2012-07-16   REATTRIBUTION / REDESIGNATION REQUESTED:   8
##  Median :2012-09-17   REATTRIBUTION FROM SPOUSE              :   5
##  Mean   :2012-08-09   REDESIGNATION FROM GENERAL             :   4
##  3rd Qu.:2012-10-17   REDESIGNATION FROM PRIMARY             :   4
##  Max.   :2012-12-04   SEE REATTRIBUTION                      :   3
##                       (Other)                                :   1
##  memo.code                                      memo.text    form.type
##   :7193                                              :7147   SA17A:7206
##  X:2310    * OBAMA VICTORY FUND 2012                 :1424   SA18 :2297
##            TRANSFER FROM ROMNEY VICTORY INC.         : 868   SB28A:   0
##            * EARMARKED CONTRIBUTION: SEE BELOW       :  12
##            REATTRIBUTION / REDESIGNATION REQUESTED:   8
##            EARMARK: ACTRIGHT                         :   5
##            (Other)                                   :  39
##   file.number                  transaction.id election.type
##  Min.   :756214   SA17.1675144        :   2   G2008:   0
##  1st Qu.:810684   SA17.2602881        :   2   G2012:5525
##  Median :821325   SA18.2407617.26.V007:   2   O2012:   5
##  Mean   :830762   C10976670           :   1   P2008:   0
##  3rd Qu.:840327   C10994901           :   1   P2012:3973
##  Max.   :944828   C10999858           :   1
##                   (Other)             :9494
##    democrat      democrat.amount    republican
t
##  Min.   :0.0000   Min.   :   0.0   Min.   :0.0000   Min.   :   0.00
##  1st Qu.:1.0000   1st Qu.:  10.0   1st Qu.:0.0000   1st Qu.:   0.00
##  Median :1.0000   Median :  35.0   Median :0.0000   Median :   0.00
##  Mean   :0.7862   Mean   : 103.8   Mean   :0.2138   Mean   :  89.39
##  3rd Qu.:1.0000   3rd Qu.: 100.0   3rd Qu.:0.0000   3rd Qu.:   0.00
##  Max.   :1.0000   Max.   :5000.0   Max.   :1.0000   Max.   :5000.00
##
##  all.donations
##  Min.   :1
##  1st Qu.:1
##  Median :1
##  Mean   :1
##  3rd Qu.:1
##  Max.   :1
##
```

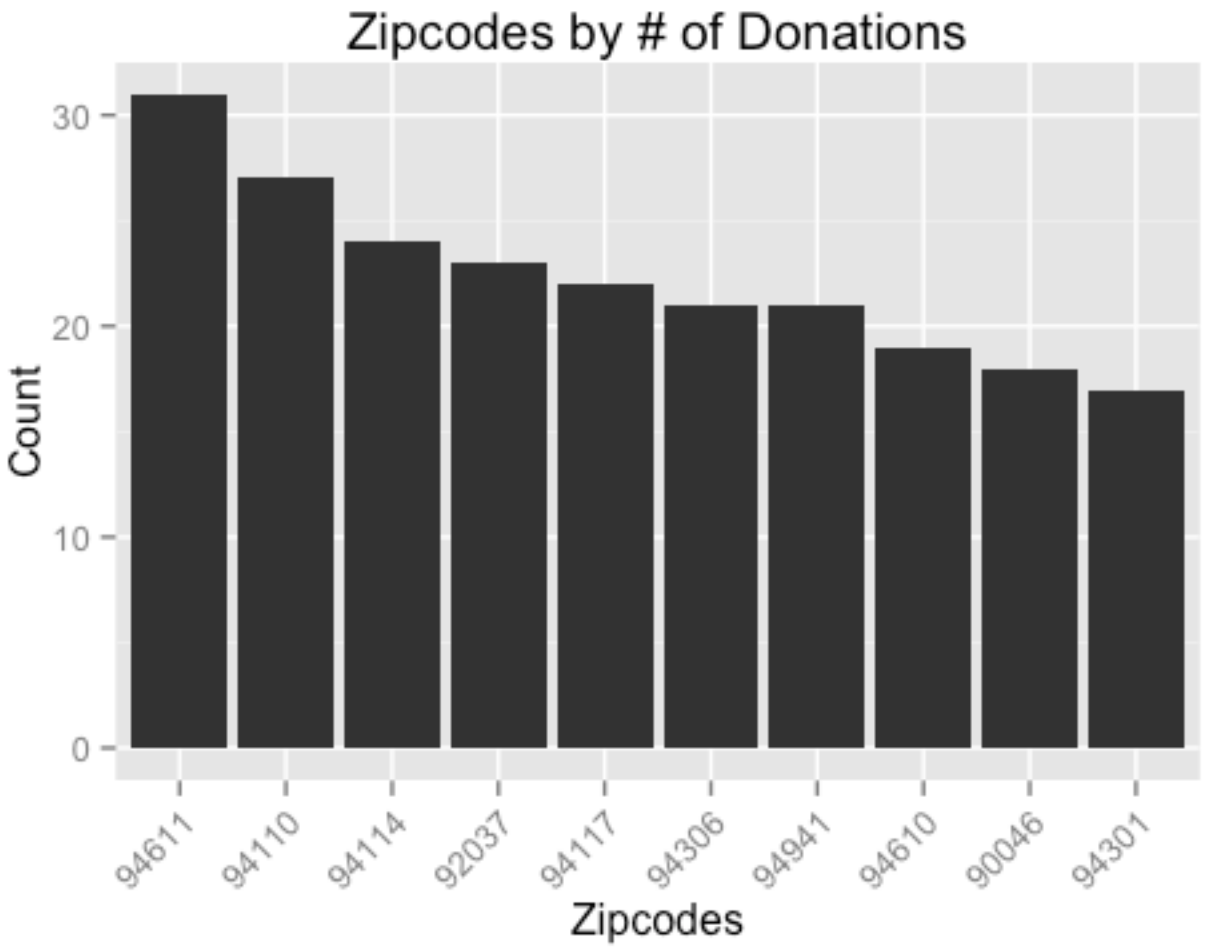
Based on this data summary alone, there is a lot of information about the dataset.
* Most donations go to Obama.
* Not surprisingly, the cities that send the most donations are the most populated cities.
* Many are retired or self-employed (more on this later).
* Of those employed by others, the most common occupation is attorney.
* The median amount of donation is $50 with a right skewed distribution.
* The median date is Oct 14th, 2012. Thus, 50% of all donations occur in the last month leading up to the election on Nov 6th, 2012.

Next, I wanted to see what are the most common values of my categorical variables.

Histogram of Most Common Occurrences

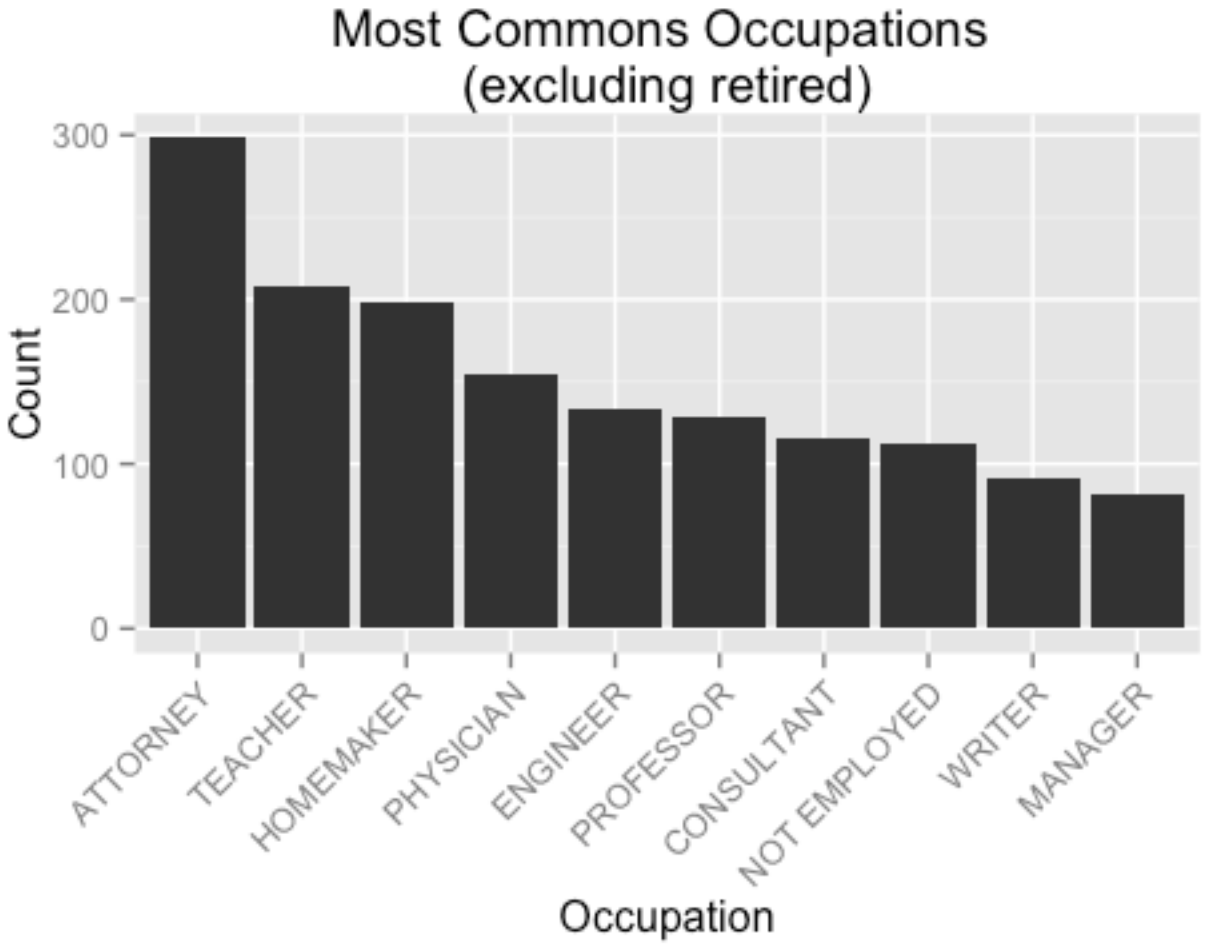



Cities by # of Donations
Count
600
400
200
0
LOS ANGELES
SAN FRANCISCO
SAN DIEGO
OAKLAND
SAN JOSE
BERKELEY
SACRAMENTO
PALO ALTO
SANTA BARBARA
SANTA MONICA
City

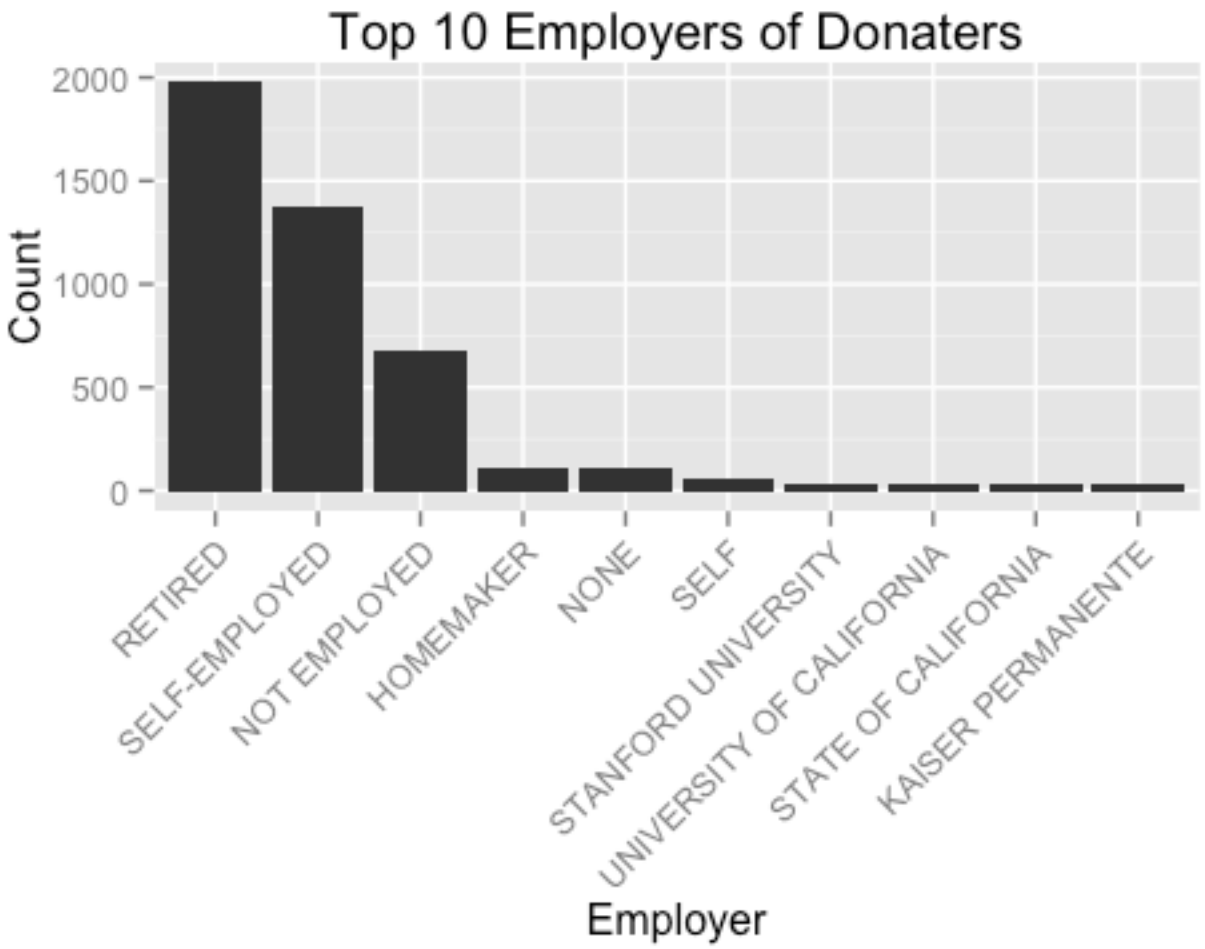


Nothing too suprising here. Generally, the cities and zips with the most donations tend to be the ones with the most amount of people.

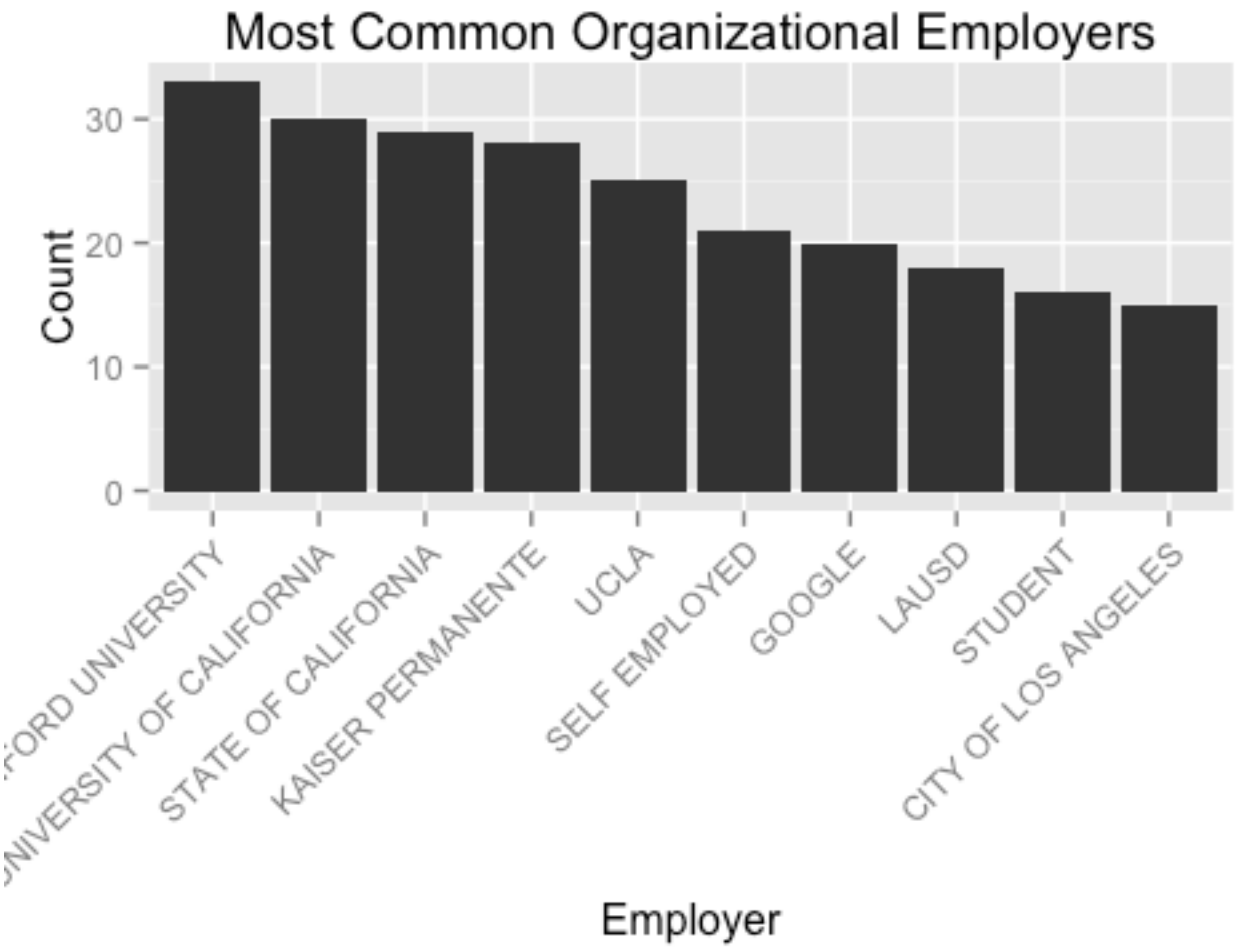
At first glance, this histogram seems strange. Do retired people really donate that much more often than everyone else? There is some credence to this, as older citizens tend to be the most politically active in terms of voting. However, this trend is better explained by the fact that retired people are probably not that different than other contributors, except for the fact that they are older and no longer work. In other words, ex-teachers and ex-attorneys are all lumped into one category, "retired". Thus, in reviewing the most commonly held occupations, it makes more sense to exclude retired people. Additionally, let's remove those who would not disclose their occupation.



Even from these results, it's still hard to draw any particular conclusions. We'd need data on the distribution of occupations in California to tell whether these particular occupations are more likely to contribute, or are just the most common amongst the population at large.



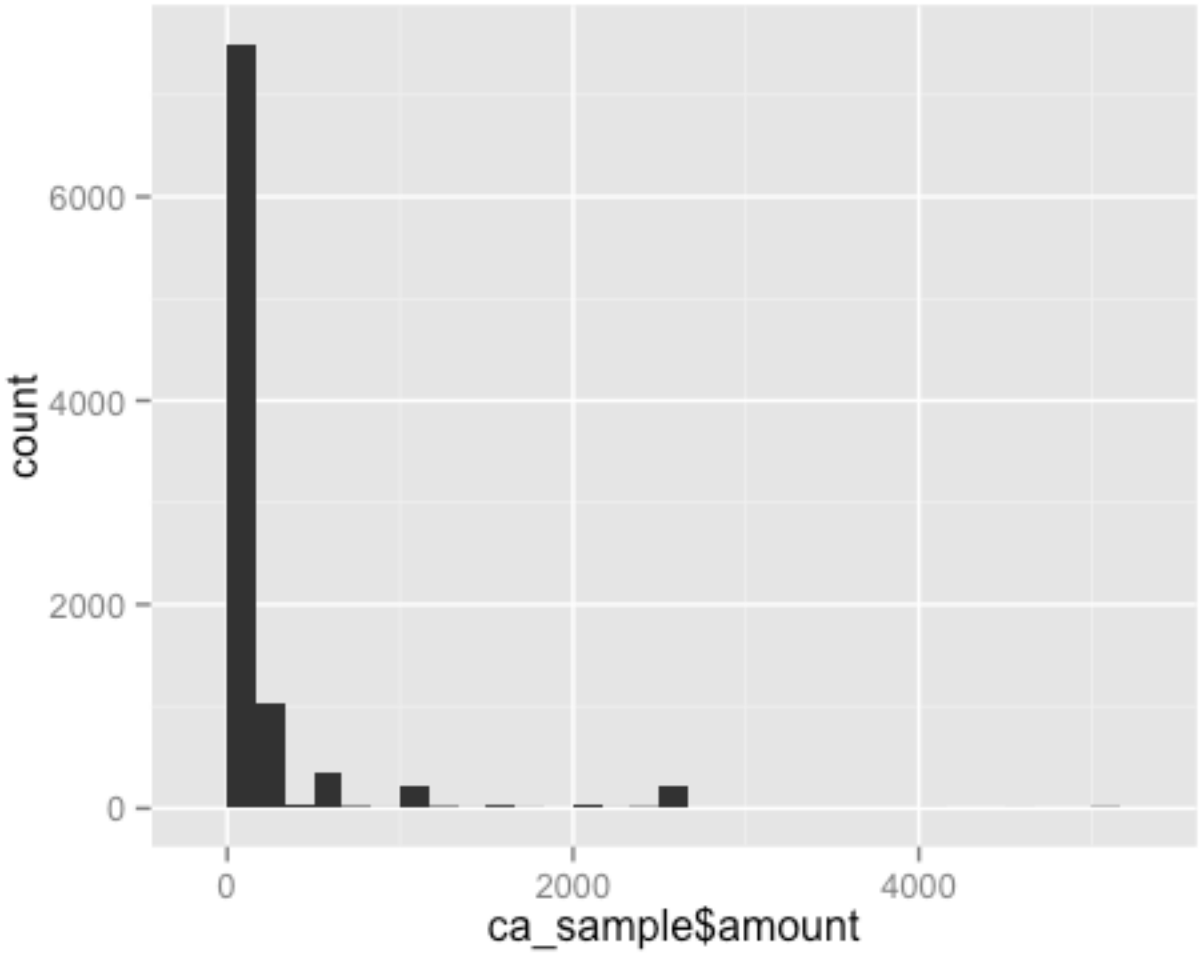
A few types of "employers" skew the distribution a bit. Let's remove non-organizational employers...

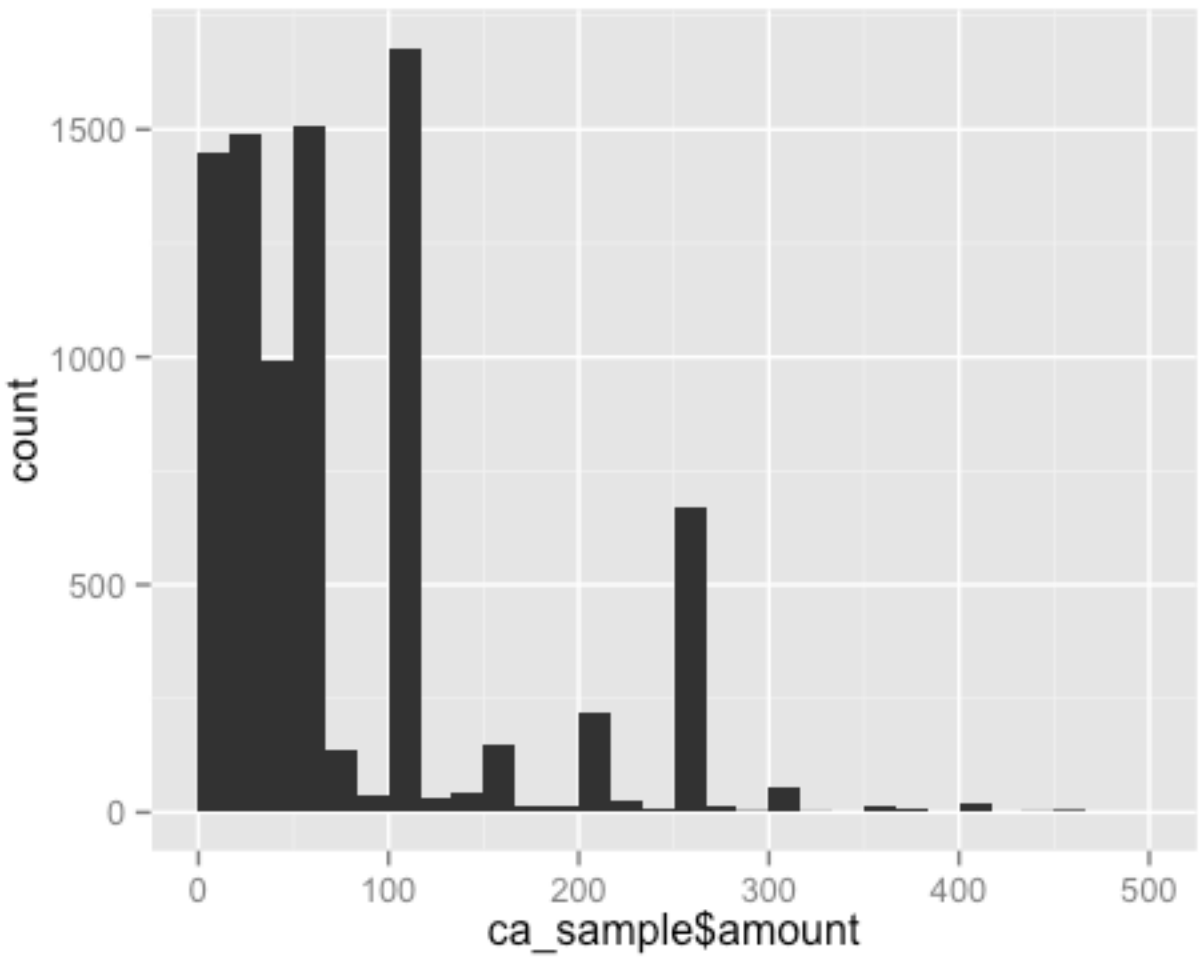


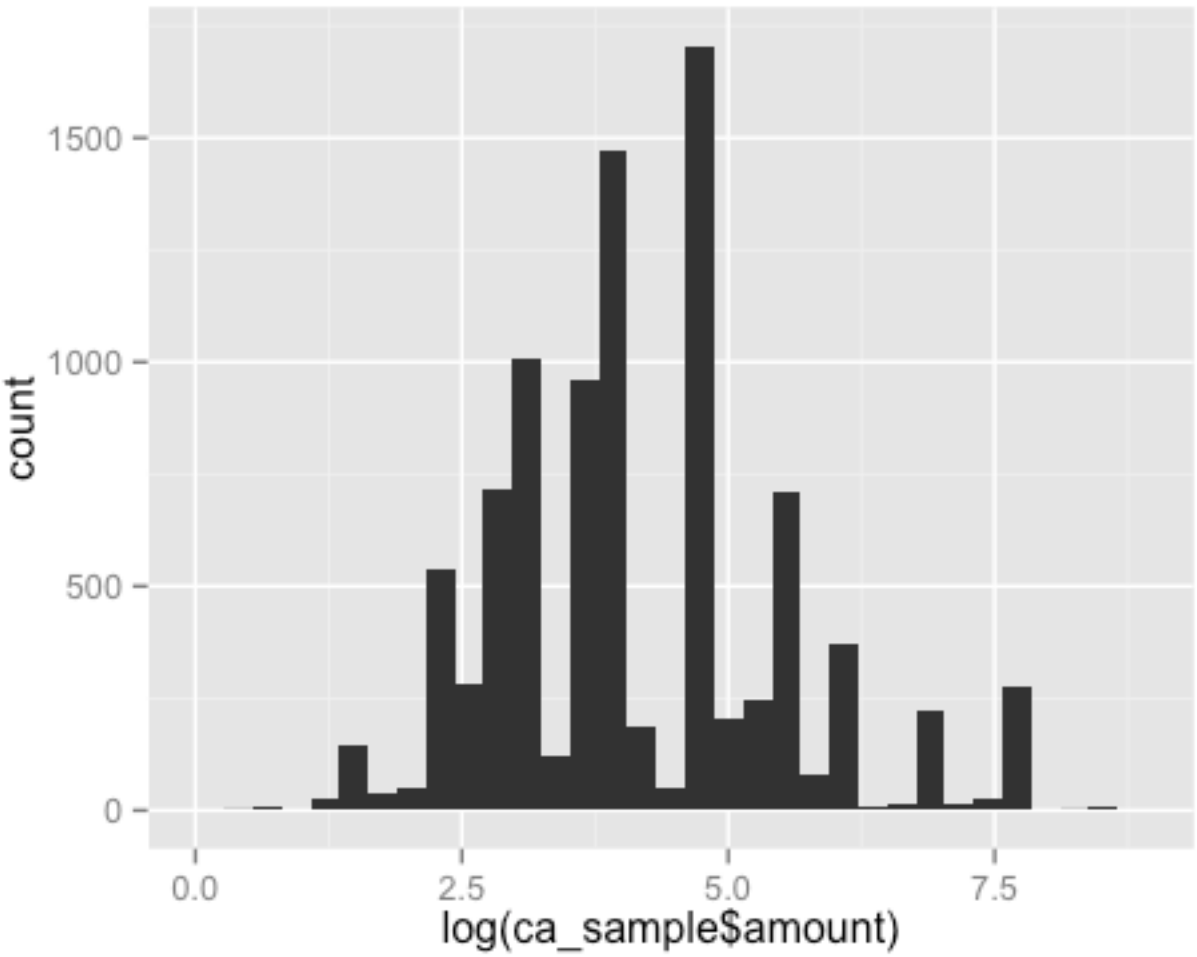
The most common employers tend to be the largest employers within the state of California. Again, we'd need to know more information about employer distributions to know if it any particular employer over or underindexes on rate of contribution.

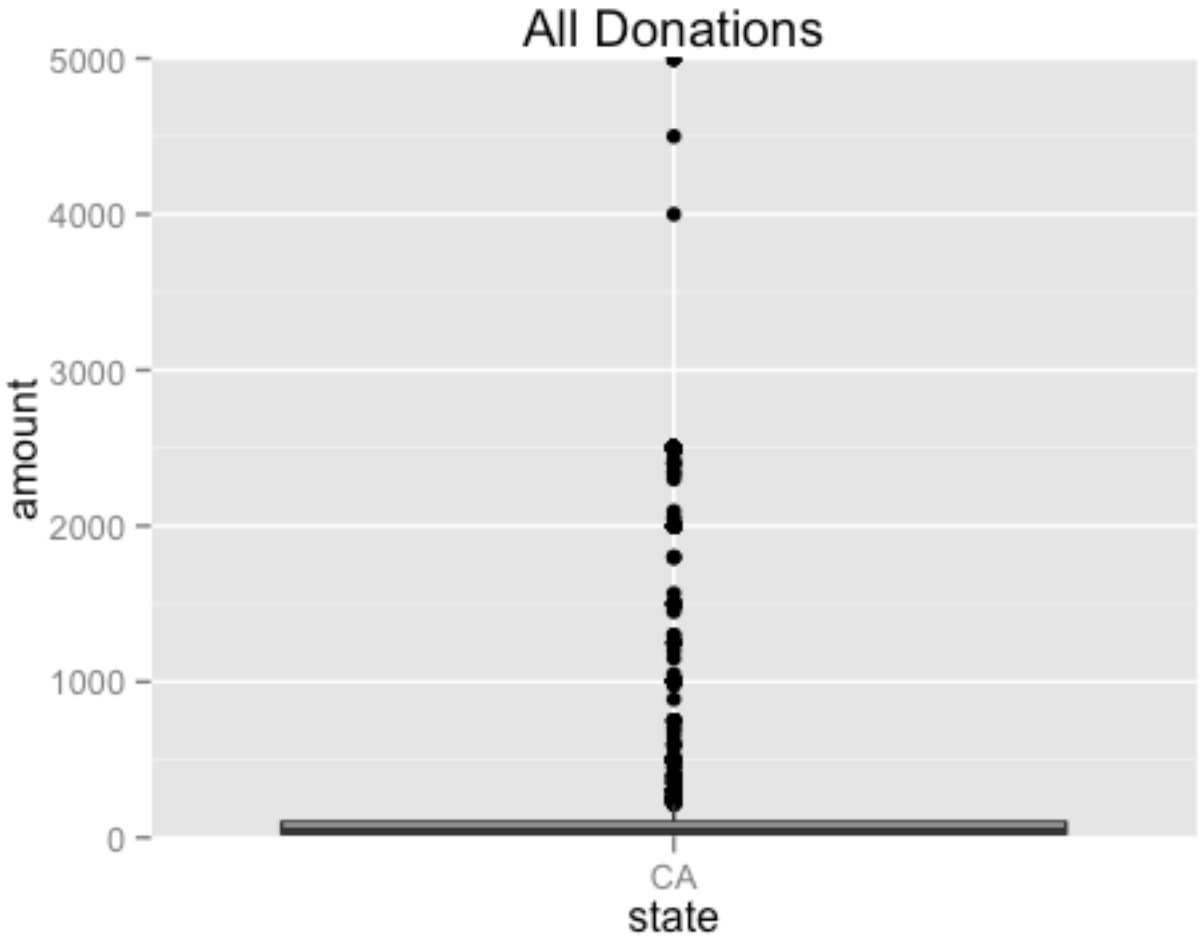
Analysis of Amount

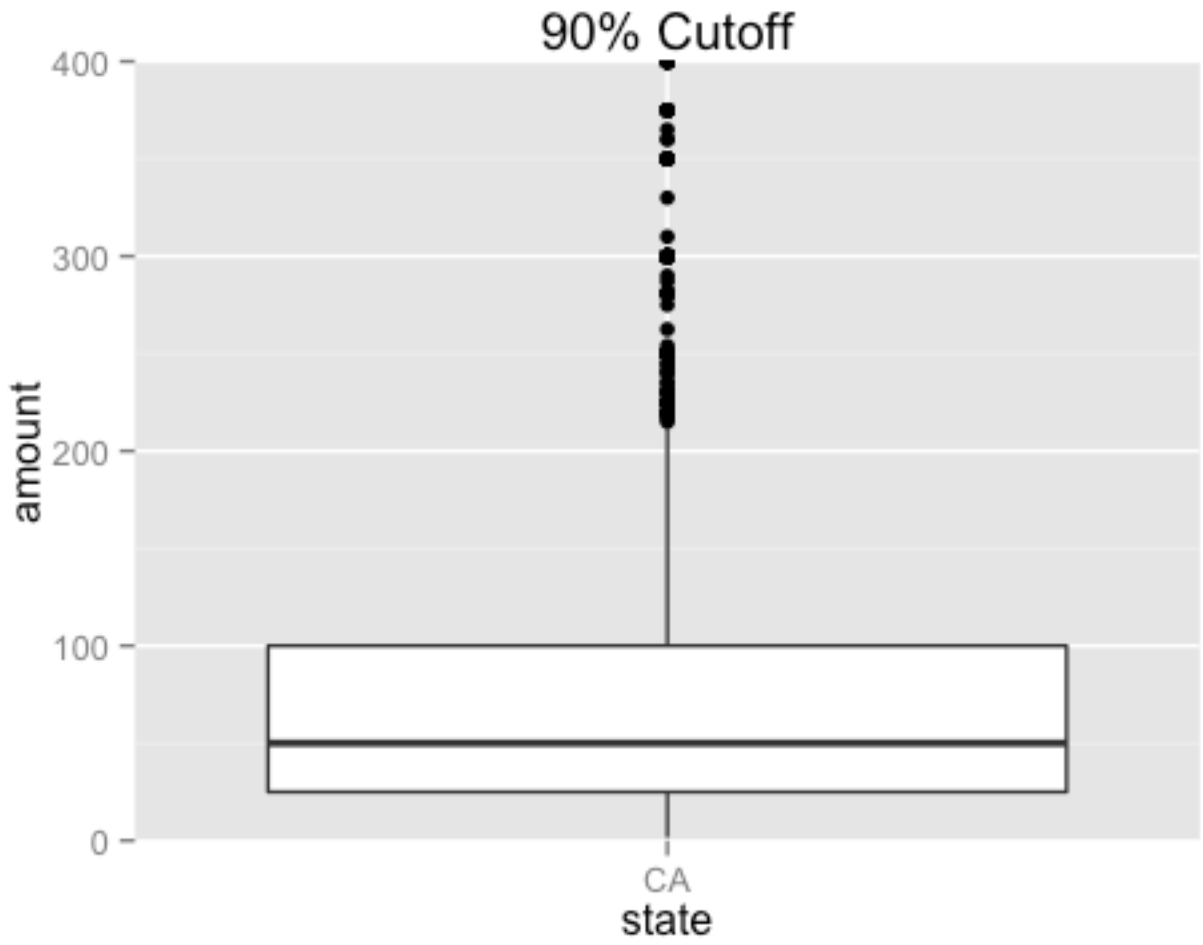
In analyzing amount, I wanted to know what the distributions looked like how they varied between parties.


6000
4000
2000
0
count
0
2000
4000
ca_sample$amount


1500
1000
500
0
count
0
100
200
300
400
500
ca_sample$amount


1500
1000
500
0
count
0.0
2.5
5.0
7.5
log(ca_sample$amount)

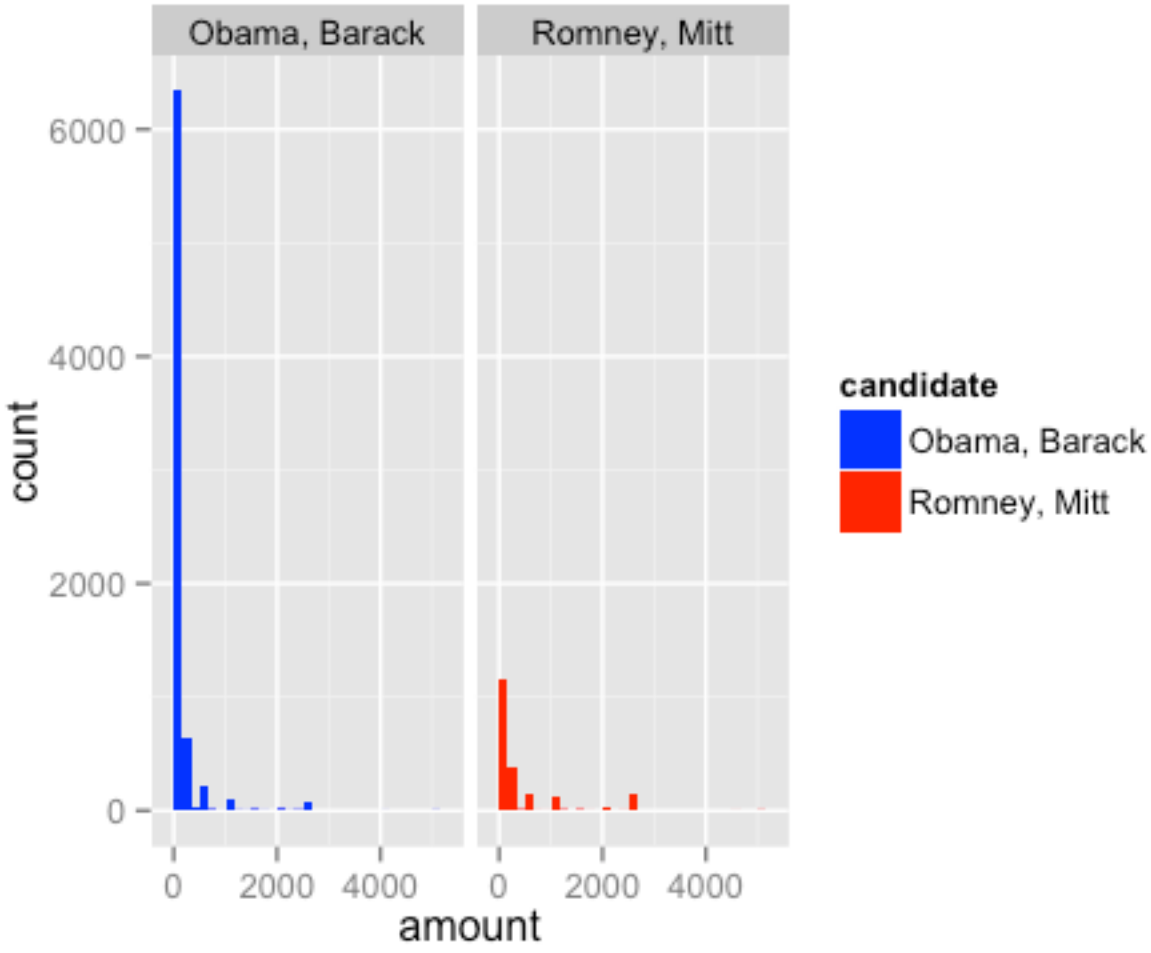

All Donations
5000
4000
3000
2000
1000
0
amount
CA
state

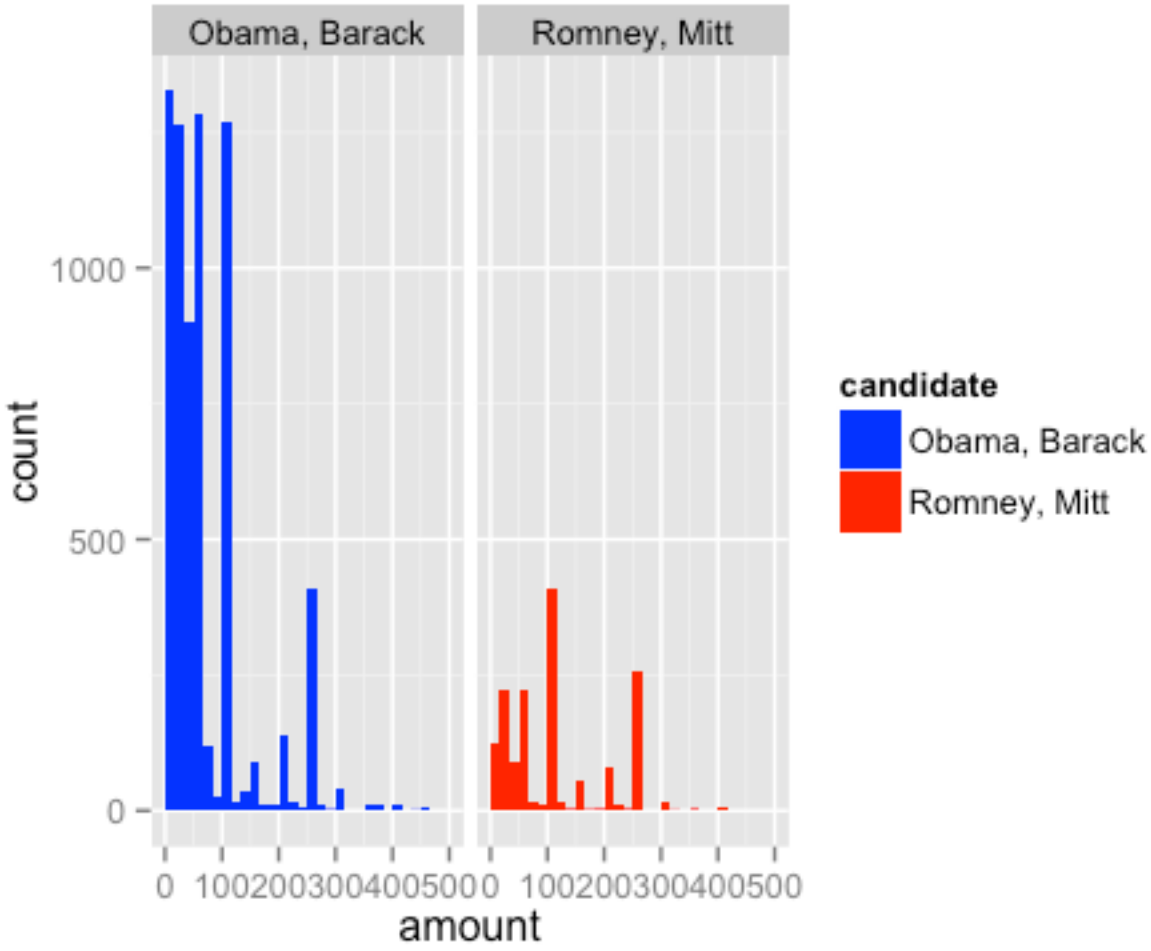


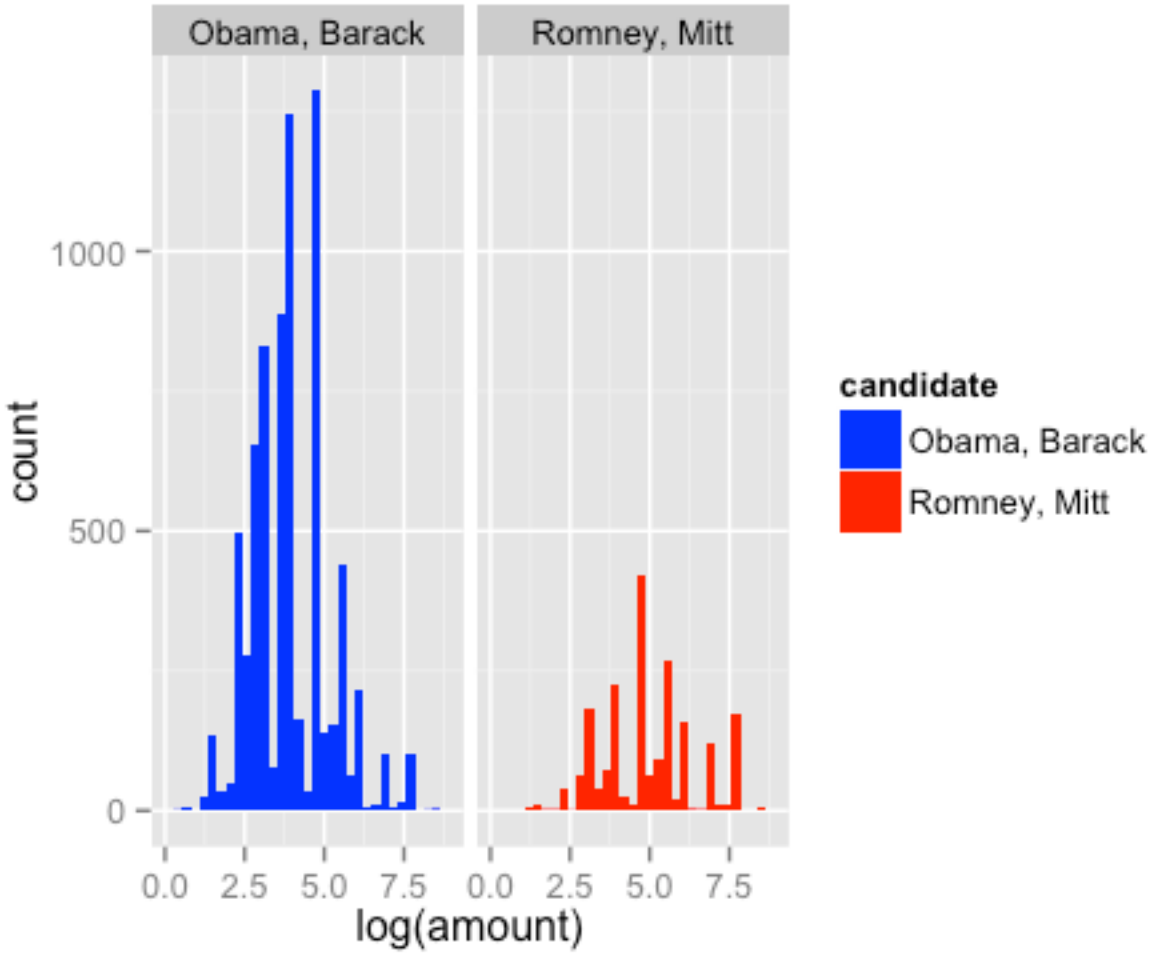

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     1.5    25.0    50.0   193.2   100.0  5000.0
```

Clearly, there are some large outliers in the values of amounts. Given that the initial histogram was right skewed, I first decided to take a log transformation to see whether it would then follow a normal distribution. It does look more normal, but not quite as perfectly as one might hope.

Next, I asked, is there a difference in distributions in donations by party?

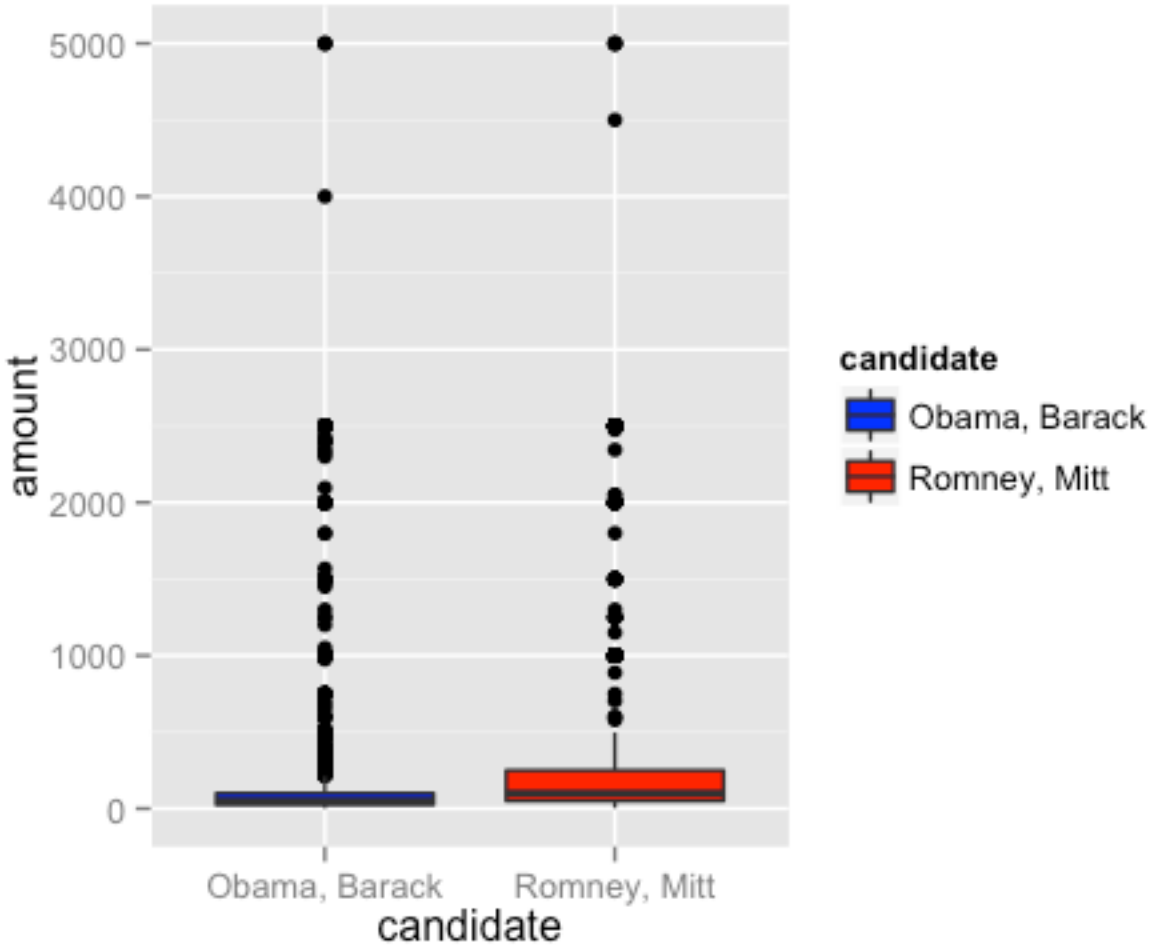

Obama, Barack
Romney, Mitt
6000
4000
2000
0
count
0
2000
4000
0
2000
4000
amount
candidate
Obama, Barack
Romney, Mitt

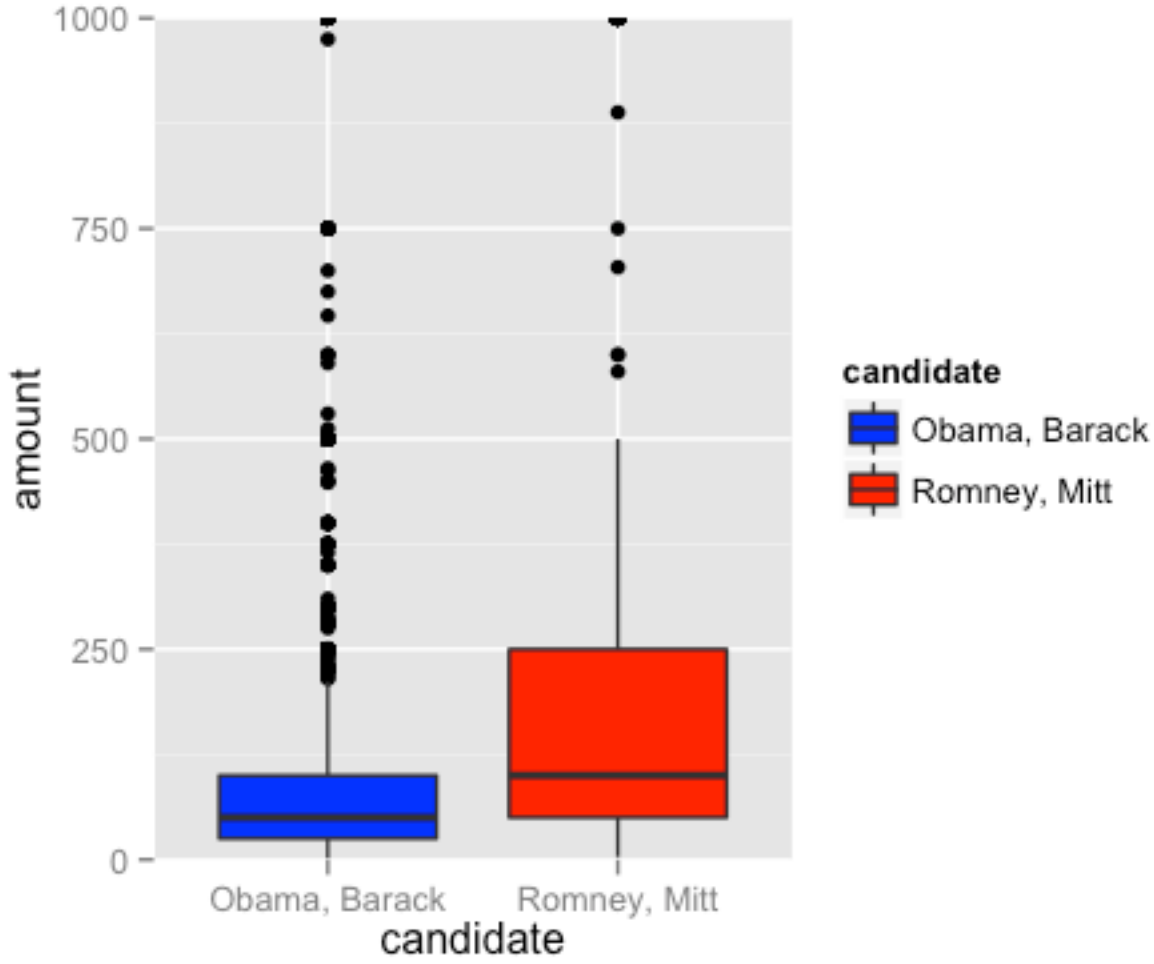

Obama, Barack
Romney, Mitt
count
1000
500
0
0 100200300400500
0 100200300400500
amount
candidate
Obama, Barack
Romney, Mitt



Based on these graphs alone, it's difficult to tell to precise details. Instead, I broke these out with boxplots and summary statistics.

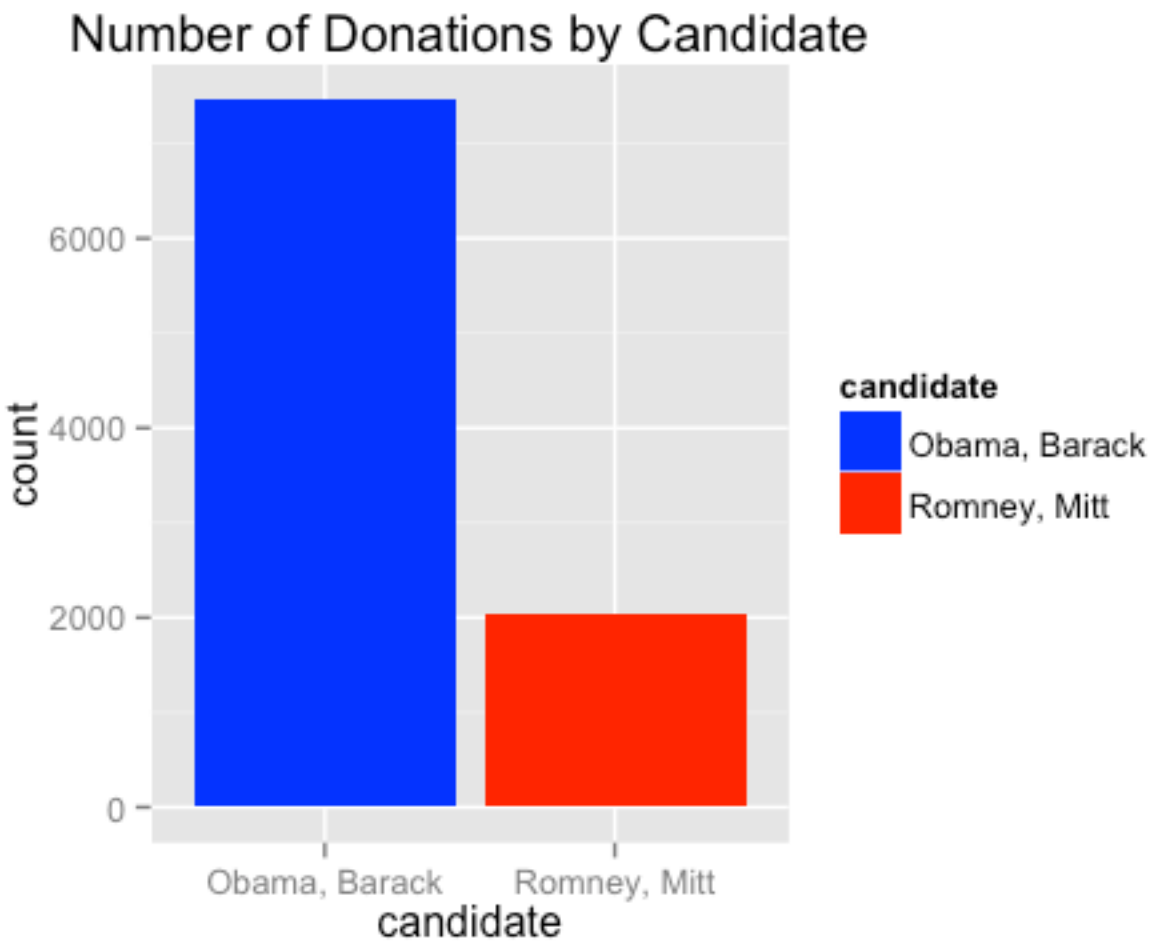
Summary of Amounts by Party

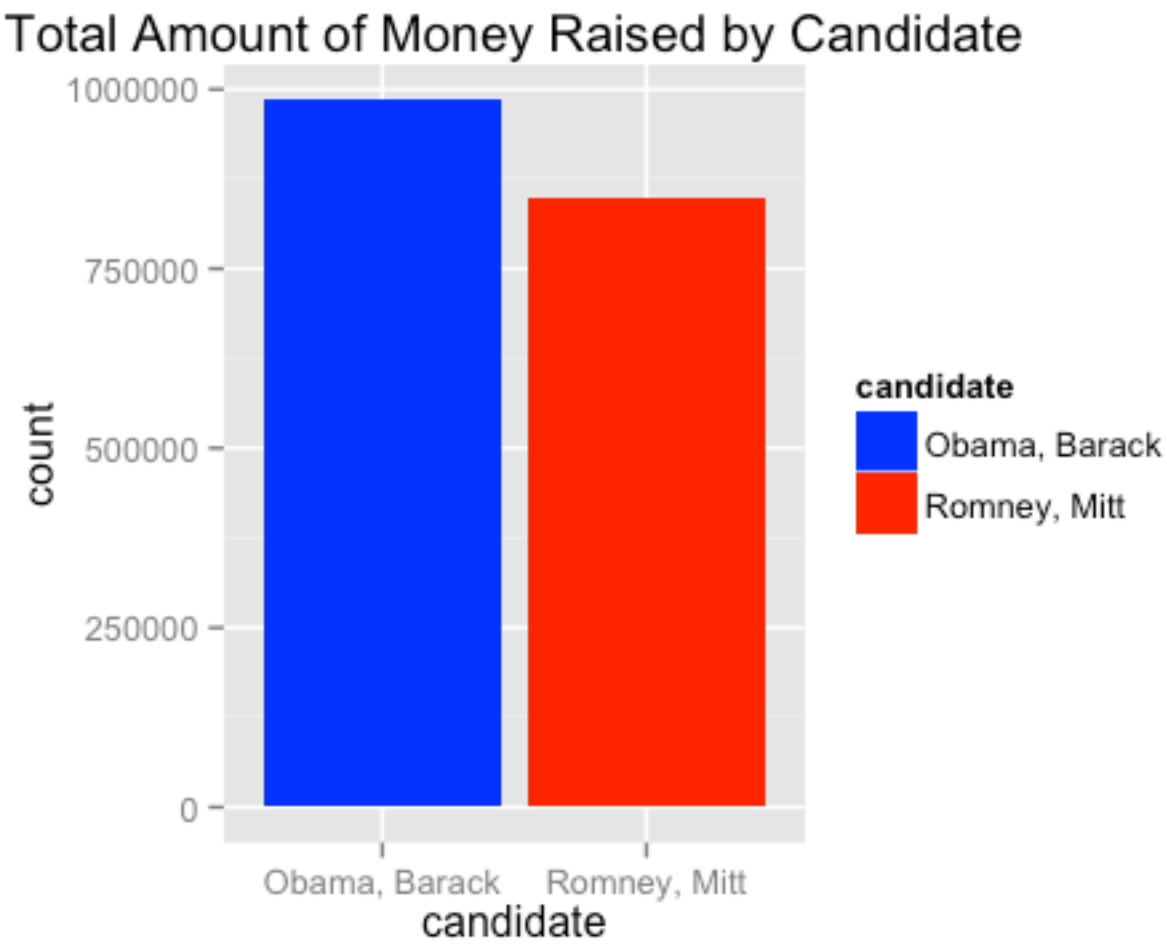

5000
4000
3000
2000
1000
0
amount
Obama, Barack
Romney, Mitt
candidate
candidate
Obama, Barack
Romney, Mitt




```
## $`Obama, Barack`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     1.5    25.0    50.0   132.1   100.0  5000.0
##
## $`Romney, Mitt`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       3      50     100     418     250    5000
```

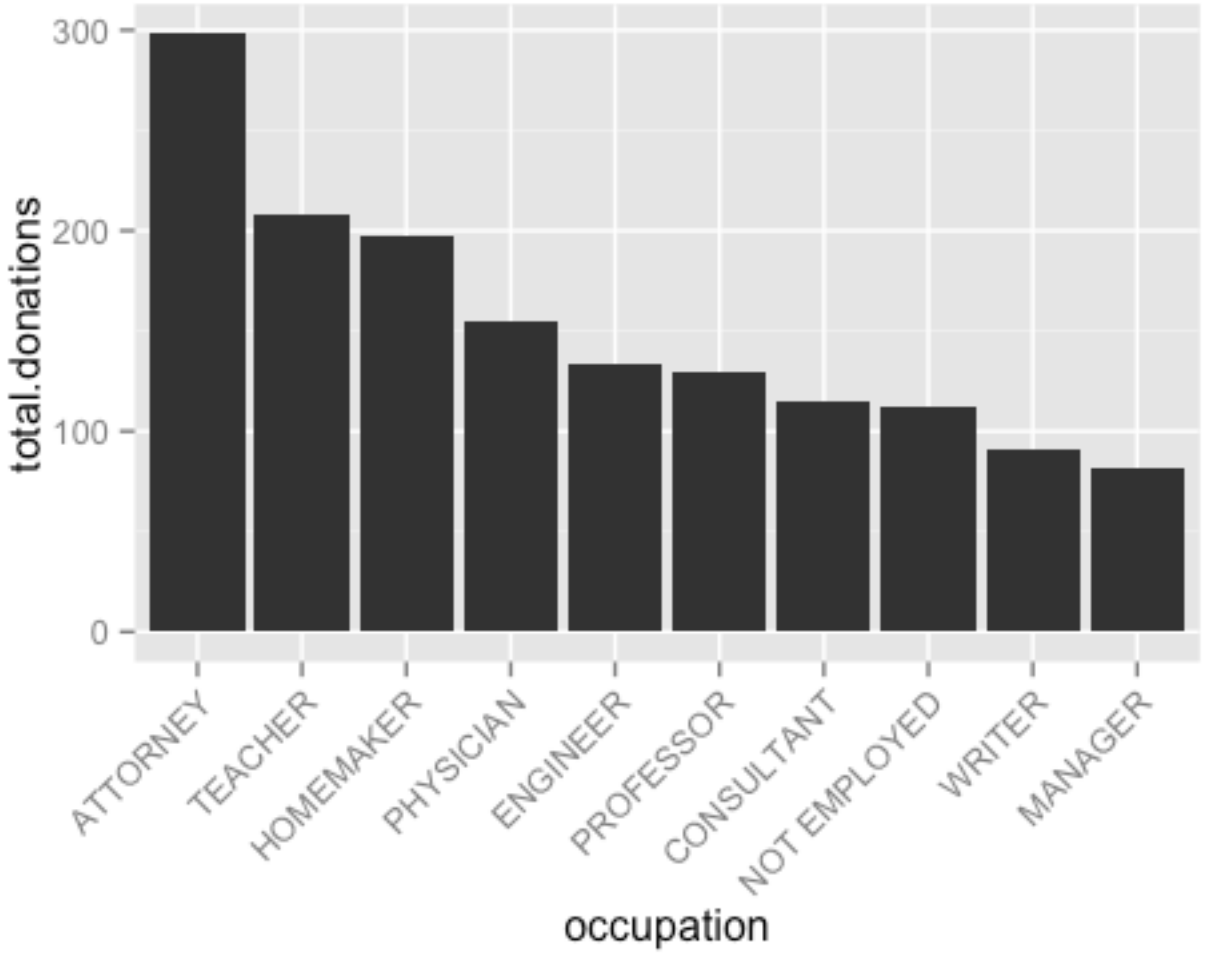
After zooming in our graph, it becomes very clear that the distribution of amounts is noticeably different between donations to Obama vs. Romney. The stark differences can be seen exactly in the numerical summary above. This would explain why although Romney has far fewer total contributors, his total amount of money raised is not as comparably low.

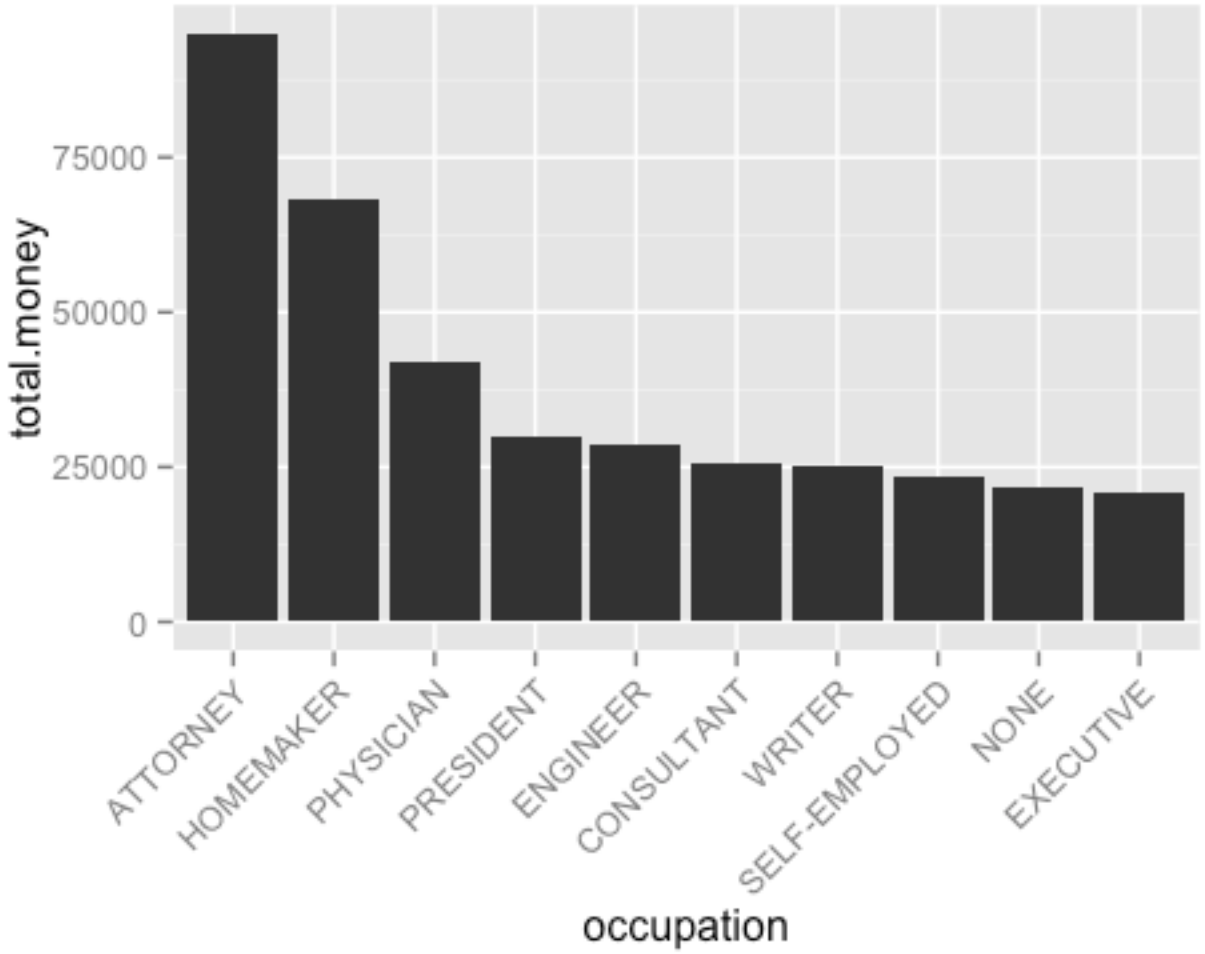

Number of Donations by Candidate
6000
4000
2000
0
count
Obama, Barack
Romney, Mitt
candidate
candidate
Obama, Barack
Romney, Mitt

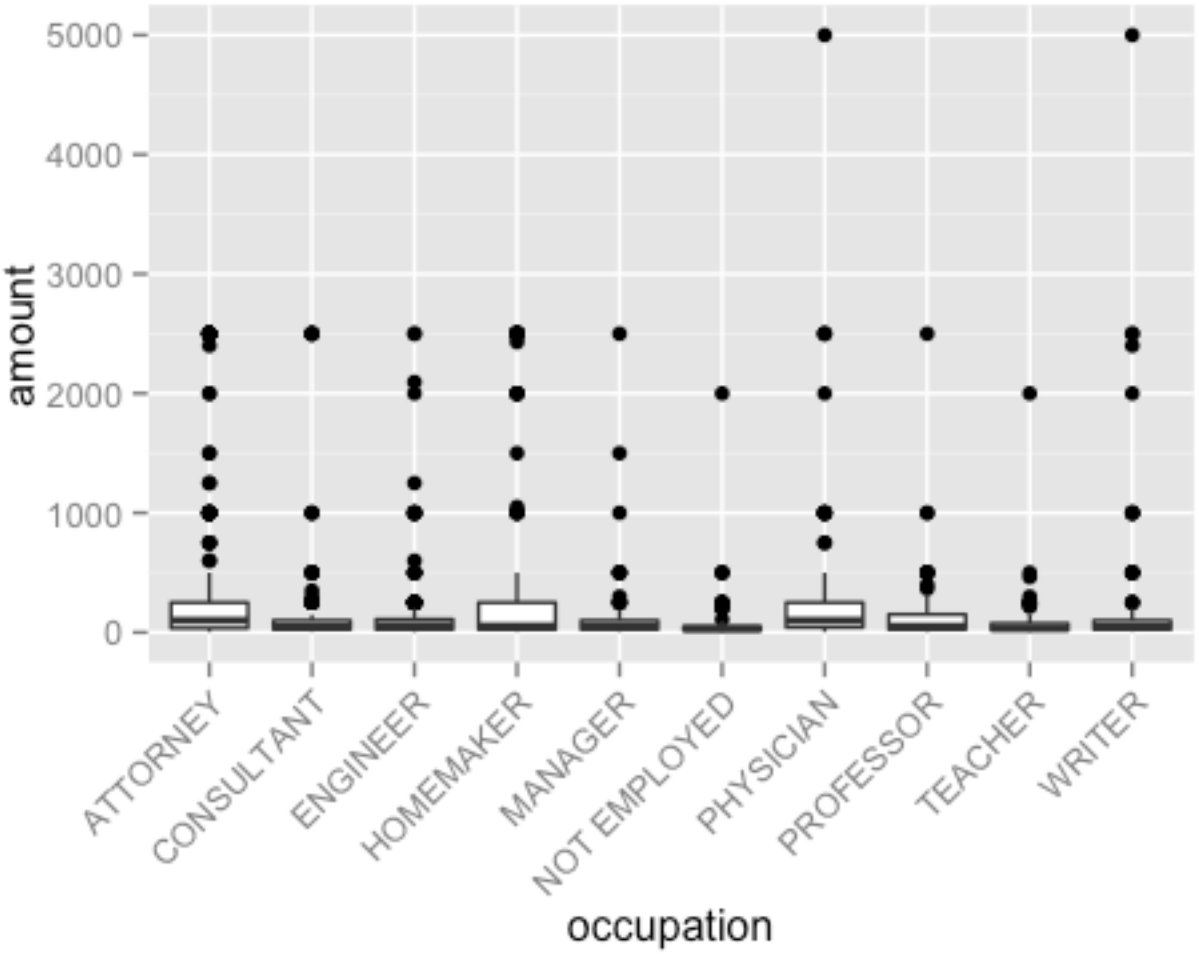


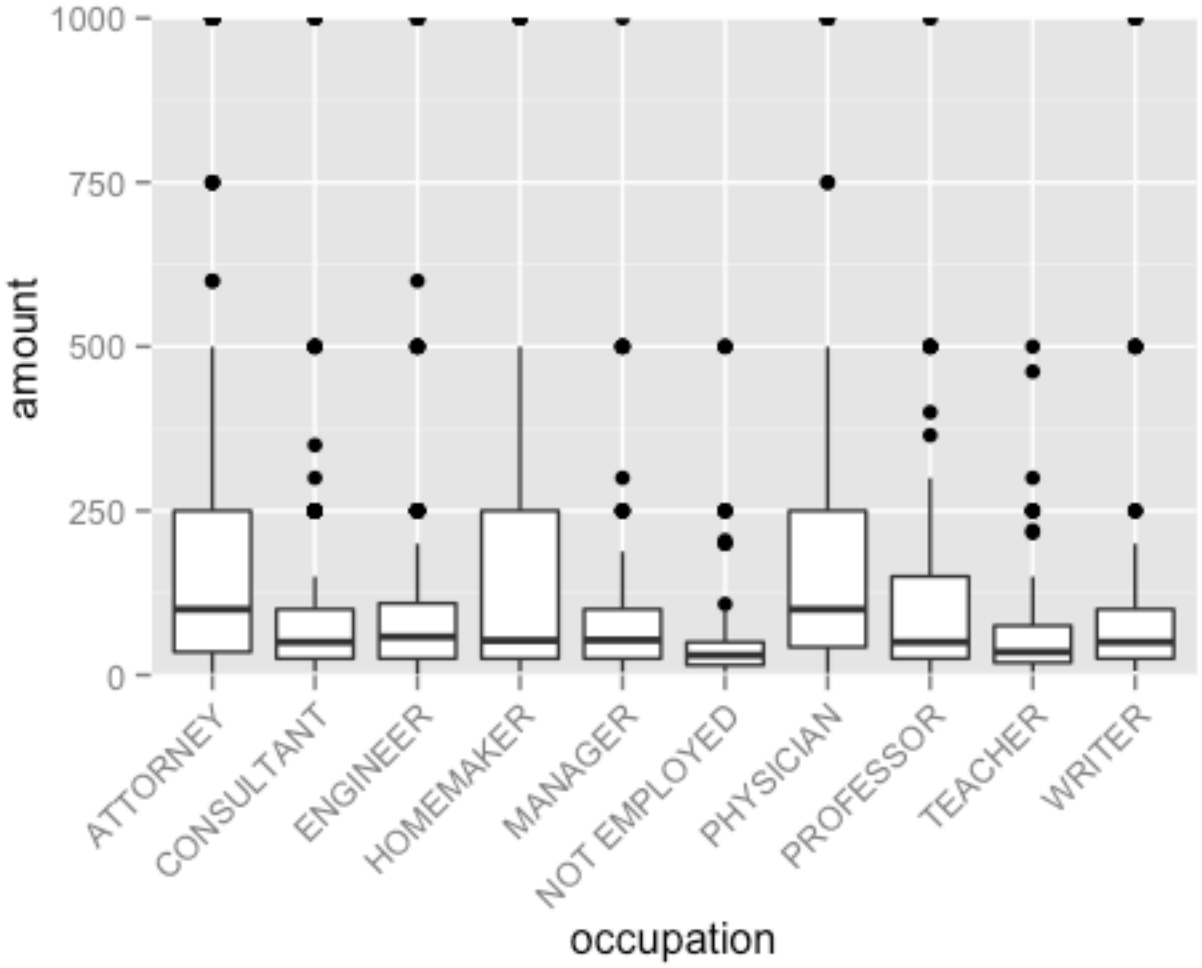

```
## Obama, Barack  Romney, Mitt
##      986612.4      849464.4
```

Profession Level Analysis



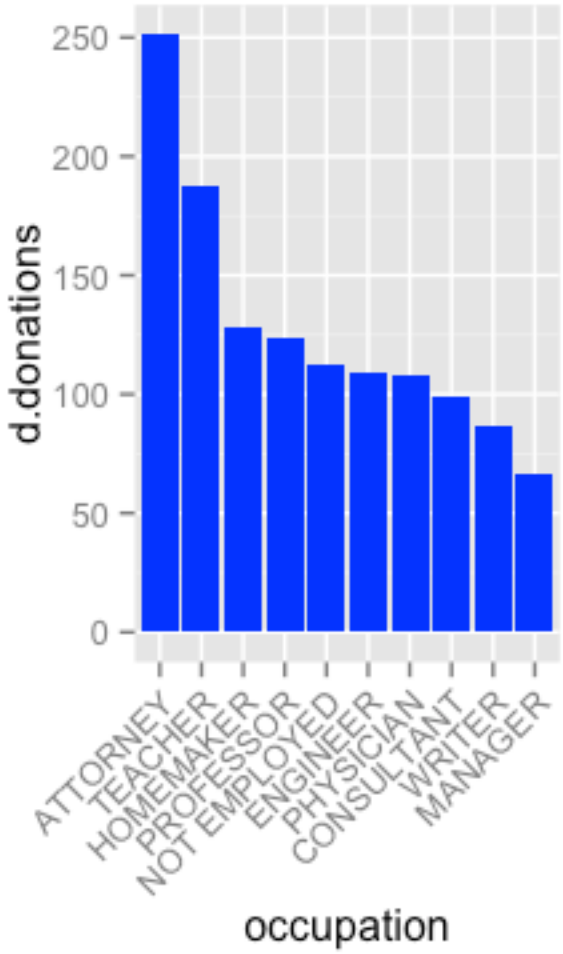

total.money
75000
50000
25000
0
ATTORNEY
HOMEMAKER
PHYSICIAN
PRESIDENT
ENGINEER
CONSULTANT
WRITER
SELF-EMPLOYED
NONE
EXECUTIVE
occupation

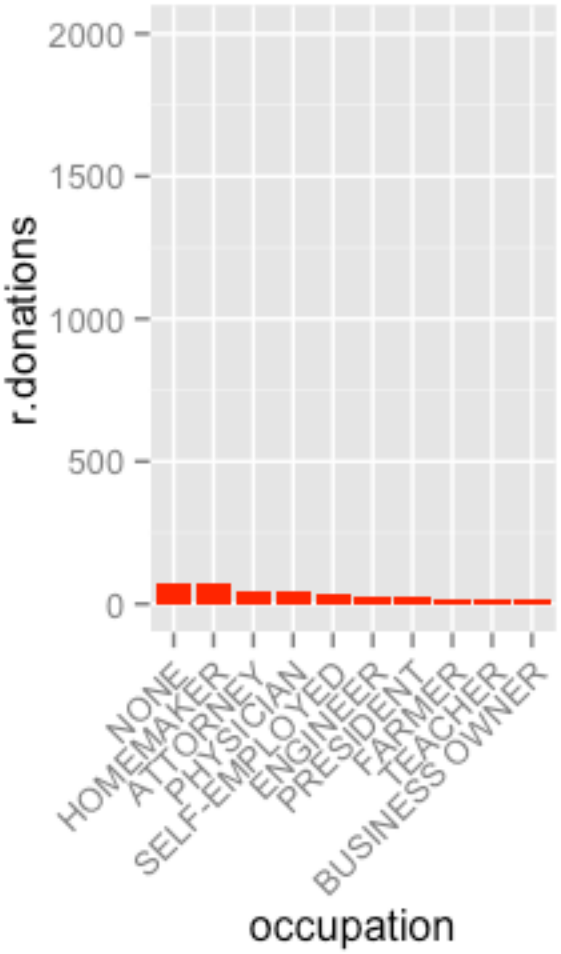

5000
4000
3000
2000
1000
0
amount
ATTORNEY
CONSULTANT
ENGINEER
HOMEMAKER
MANAGER
NOT EMPLOYED
PHYSICIAN
PROFESSOR
TEACHER
WRITER
occupation

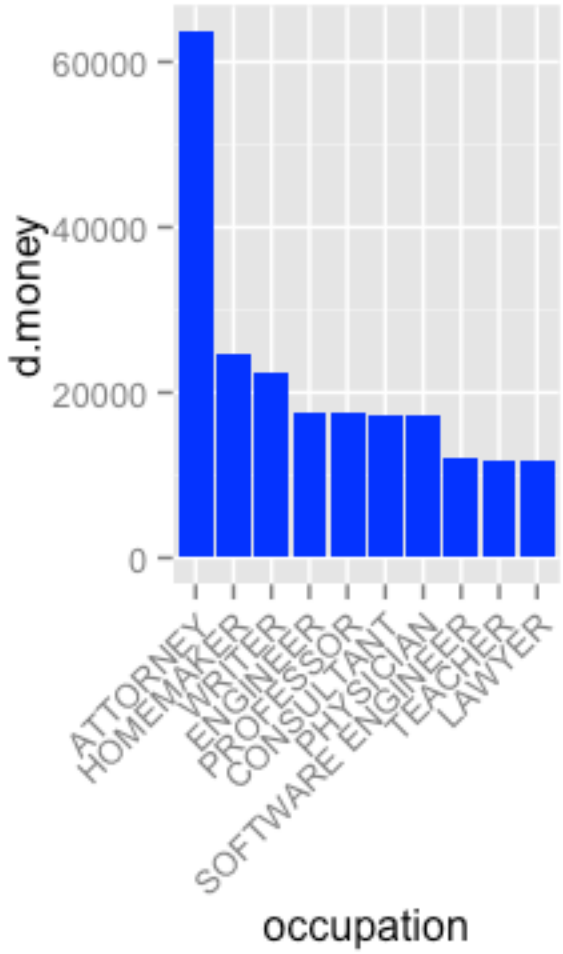


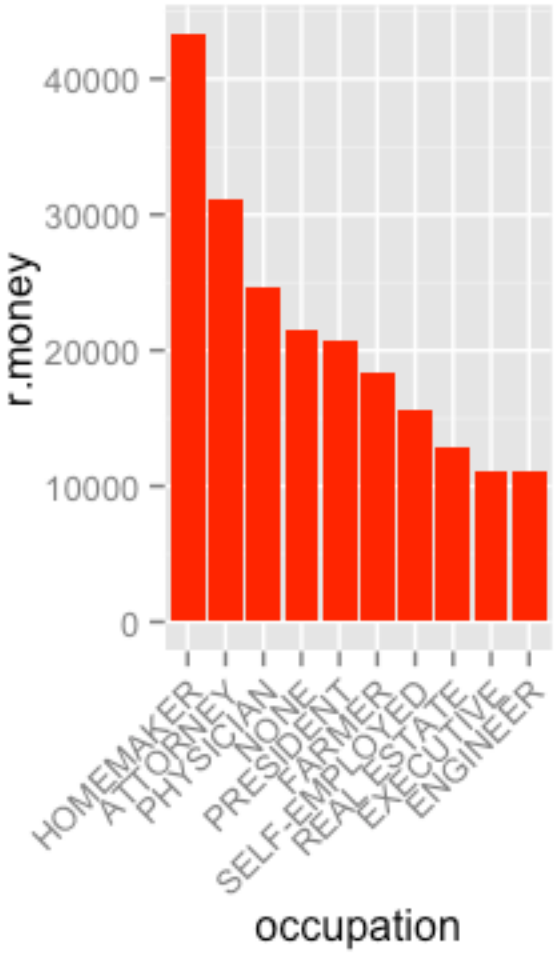
There are a few noticeable trends in occupation data. First is that expected occupation income seems to related with amount donated. For example, physicians and attorneys tend to donate more per person than teachers. The exception to this would be homeakers, but this makes sense upon further consideration. If you're a homemaker, your spouse probably must make enough money to support the entire household. Thus, households with a homemaker may have higher incomes on average, and thus be able to donate money from both partners at higher amounts.

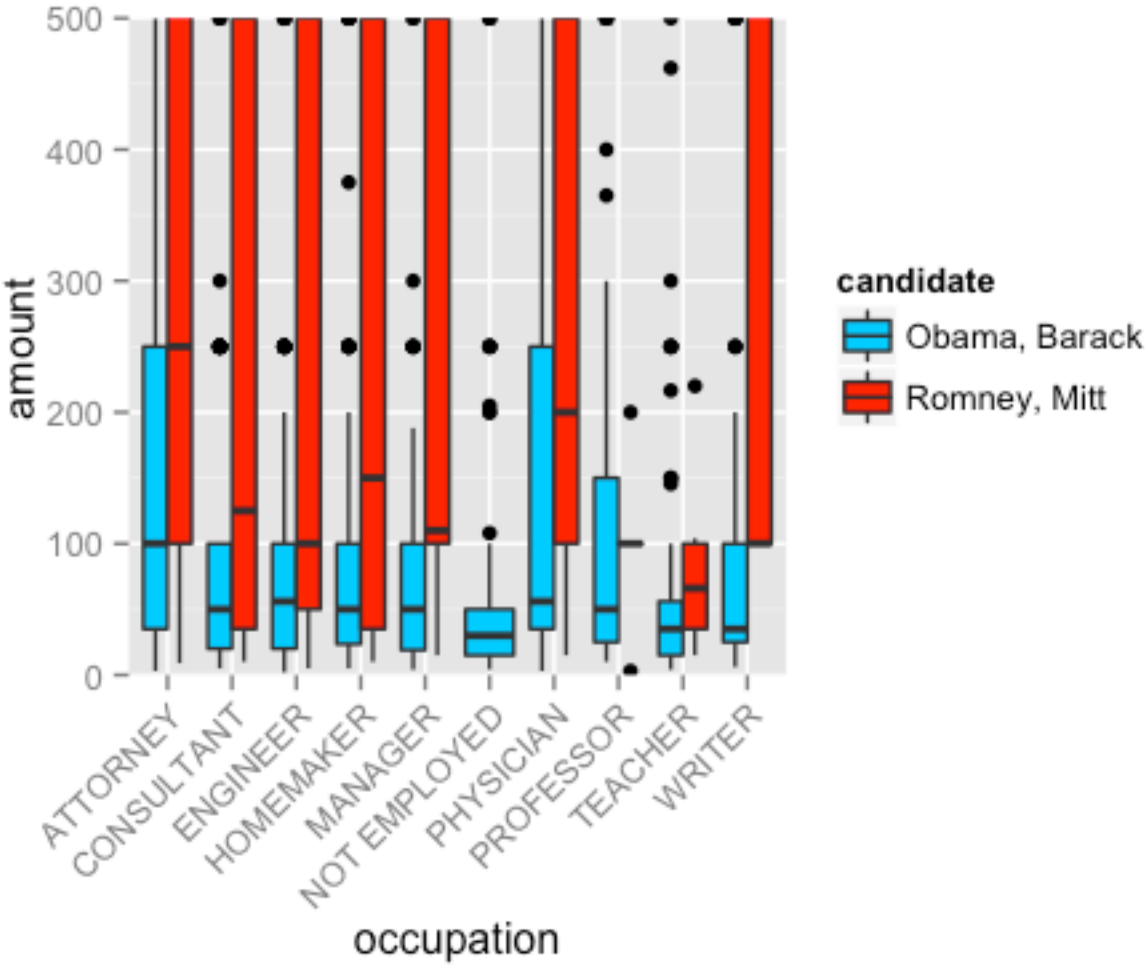
It is difficult to comment on the frequency of attorneys and physicians without outside data outlining their relative frequencies in the California population, but I'm willing to guess that they donate more often than average because specific political issues are likely to be very relevant to them given their field of work.


d.donations
250
200
150
100
50
0
ATTORNEY
TEACHER
HOMEMAKER
PROFESSOR
NOT EMPLOYED
ENGINEER
PHYSICIAN
CONSULTANT
WRITER
MANAGER
occupation


r.donations
2000
1500
1000
500
0
NONE
HOMEMAKER
ATTORNEY
PHYSICIAN
SELF-EMPLOYED
ENGINEER
PRESIDENT
FARMER
TEACHER
BUSINESS OWNER
occupation


d.money
60000
40000
20000
0
ATTORNEY
HOMEMAKER
WRITER
ENGINEER
PROFESSOR
CONSULTANT
PHYSICIAN
SOFTWARE ENGINEER
TEACHER
LAWYER
occupation

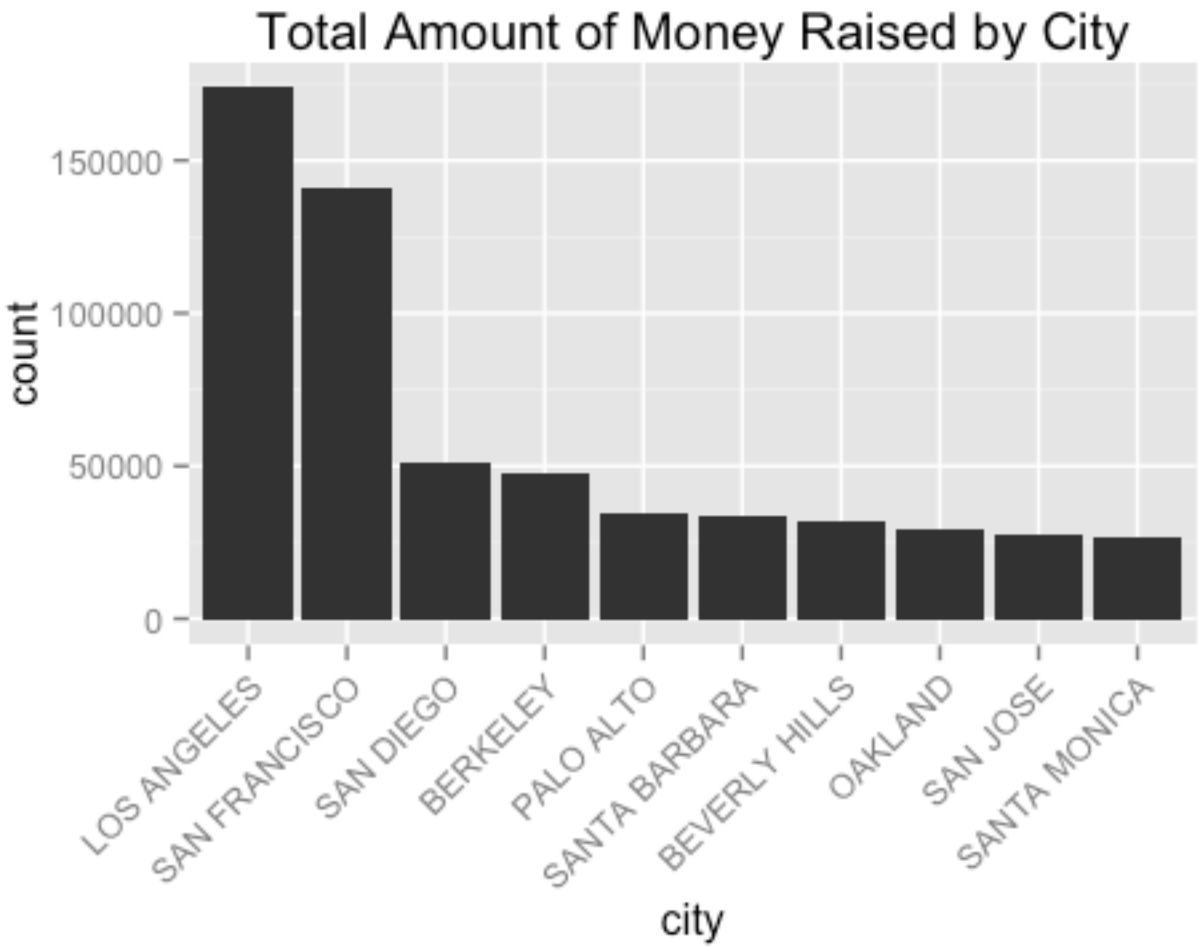

r.money
40000
30000
20000
10000
0
HOMEMAKER
ATTORNEY
PHYSICIAN
NONE
PRESIDENT
FARMER
SELF-EMPLOYED
REAL ESTATE
EXECUTIVE
ENGINEER
occupation

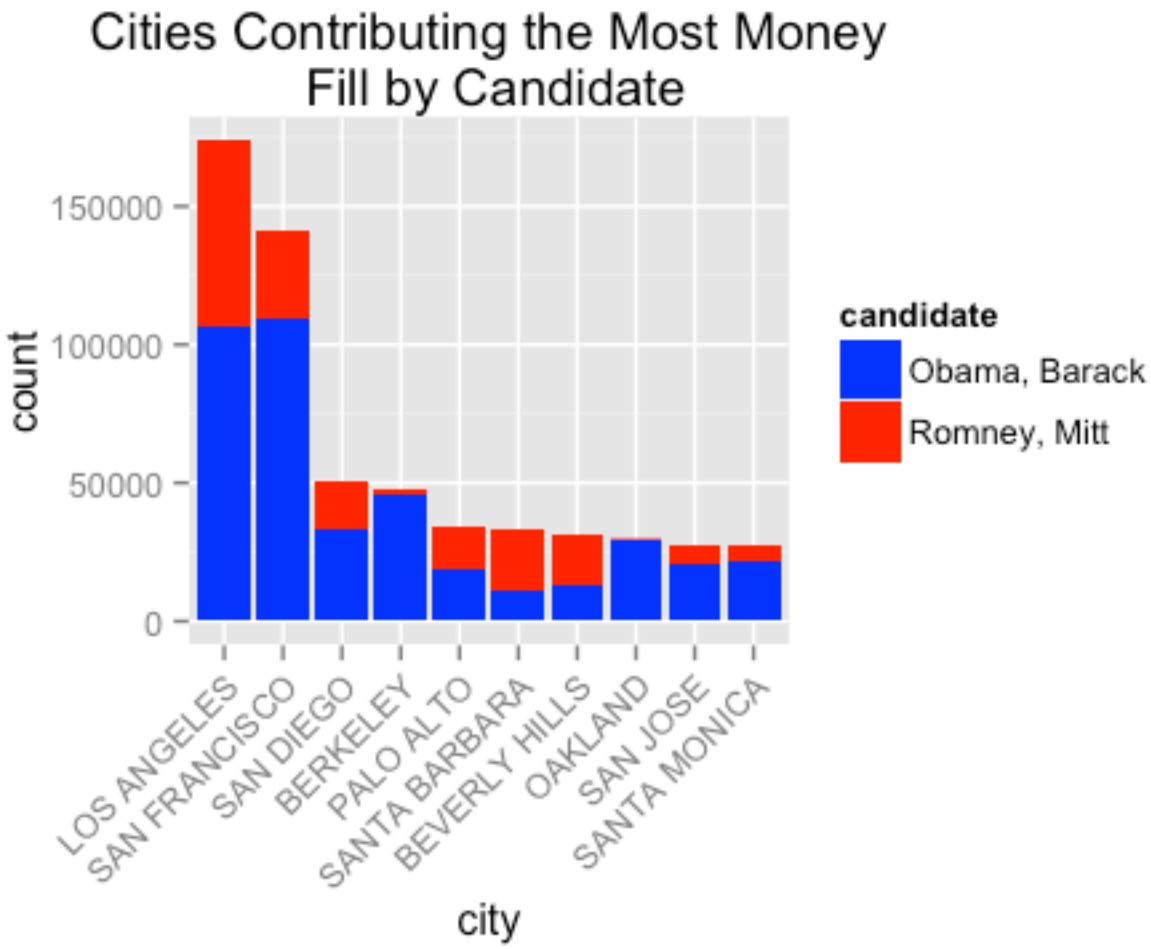


Once agian, we see that candidate has a strong relationship across the top occupations sampled. In these charts, we also see that professors tend to lean more towards donating to the democratic candidate, whereas occupations related to business functions (sales, manager, president, real estate, etc.) tend to lean toward donating to the republican canidate.

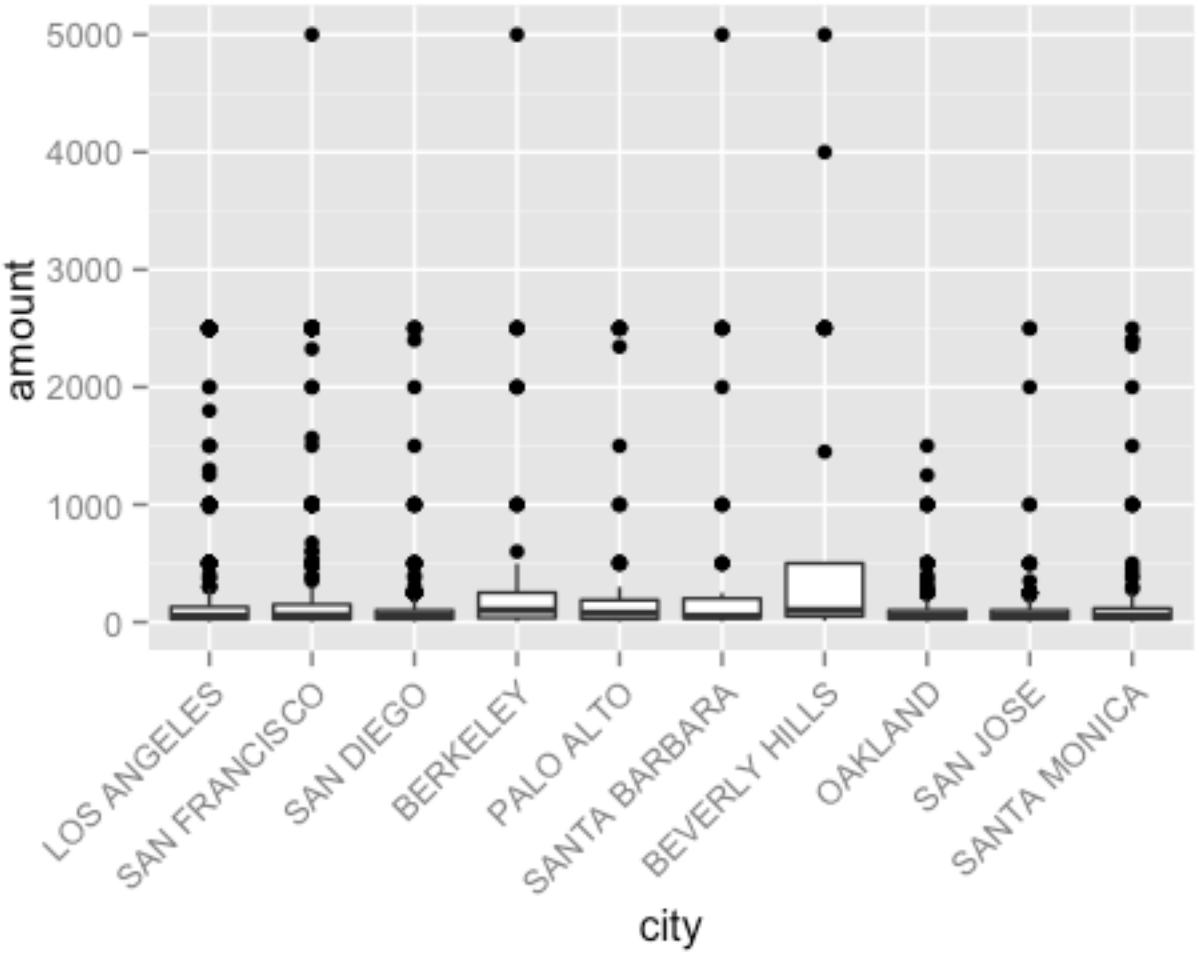
City Level Analysis

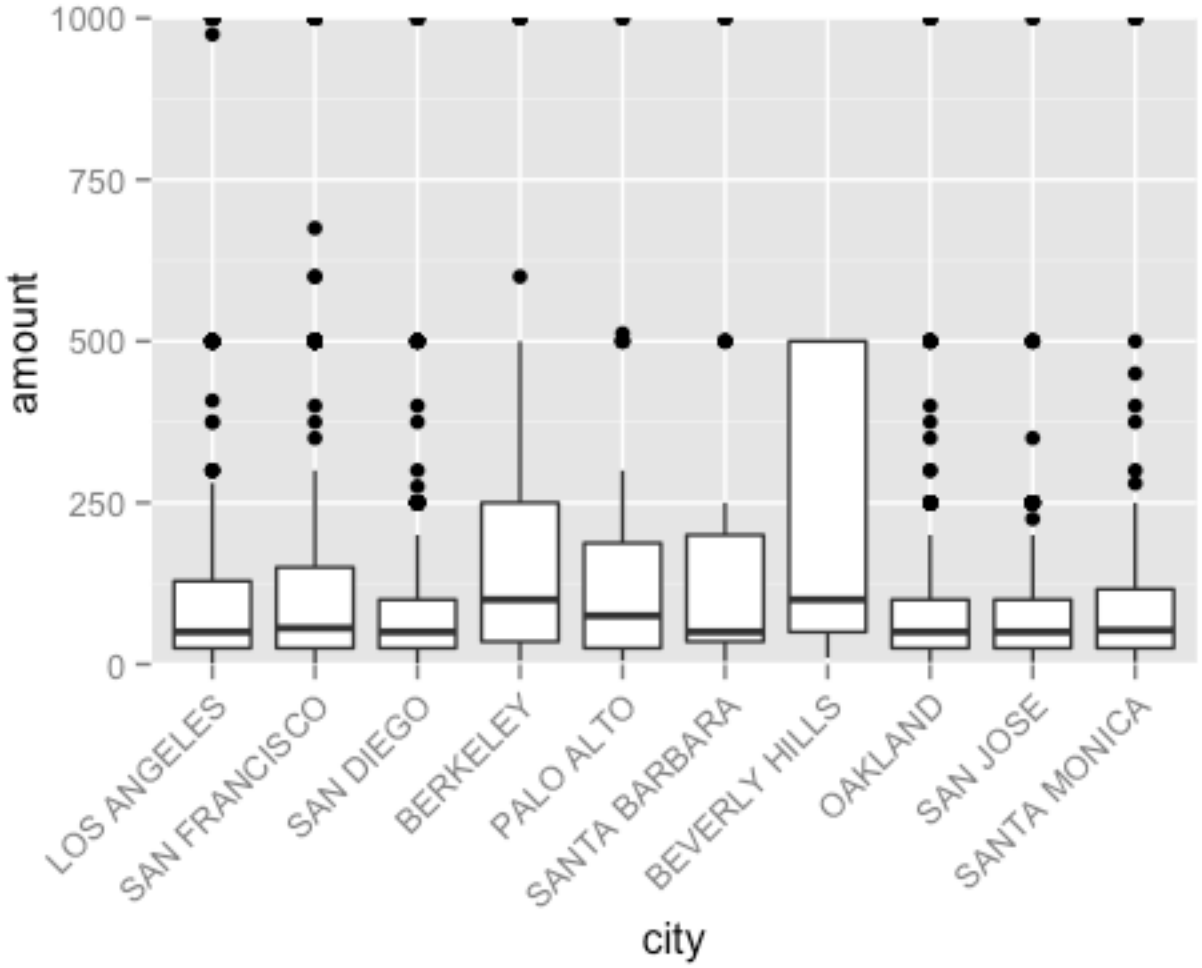
Given what we know about amounts across all of California, I looked to explore city-level data.

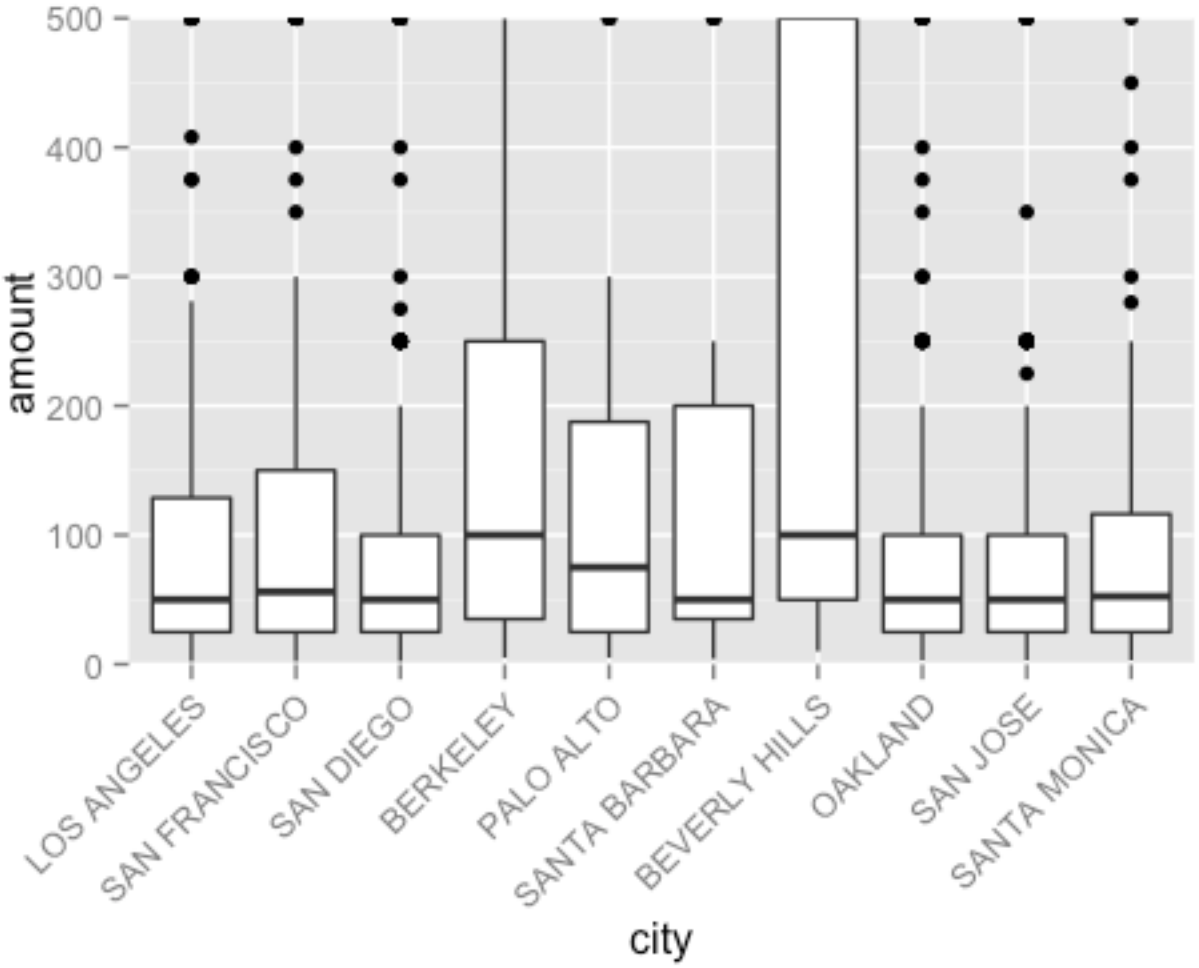

Total Amount of Money Raised by City
count
150000
100000
50000
0
LOS ANGELES
SAN FRANCISCO
SAN DIEGO
BERKELEY
PALO ALTO
SANTA BARBARA
BEVERLY HILLS
OAKLAND
SAN JOSE
SANTA MONICA
city



Some of this chart is as expected. The largest cities like Los Angeles and San Francisco donate the most money just through their sheer population size. But some smaller cities also make this list. Let's separate out the effects of population by looking at boxplots of the amounts by cities...

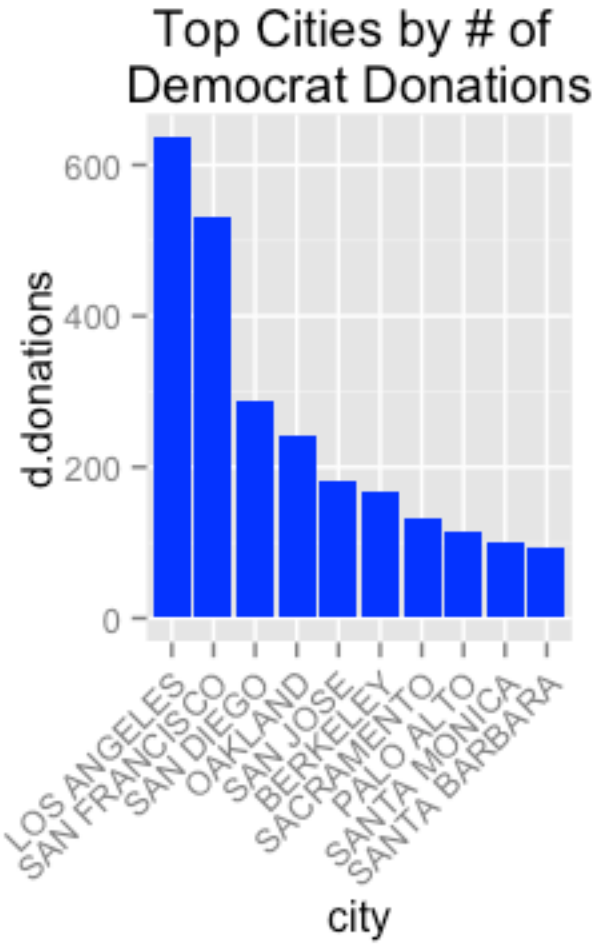

5000
4000
3000
2000
1000
0
amount
LOS ANGELES
SAN FRANCISCO
SAN DIEGO
BERKELEY
PALO ALTO
SANTA BARBARA
BEVERLY HILLS
OAKLAND
SAN JOSE
SANTA MONICA
city

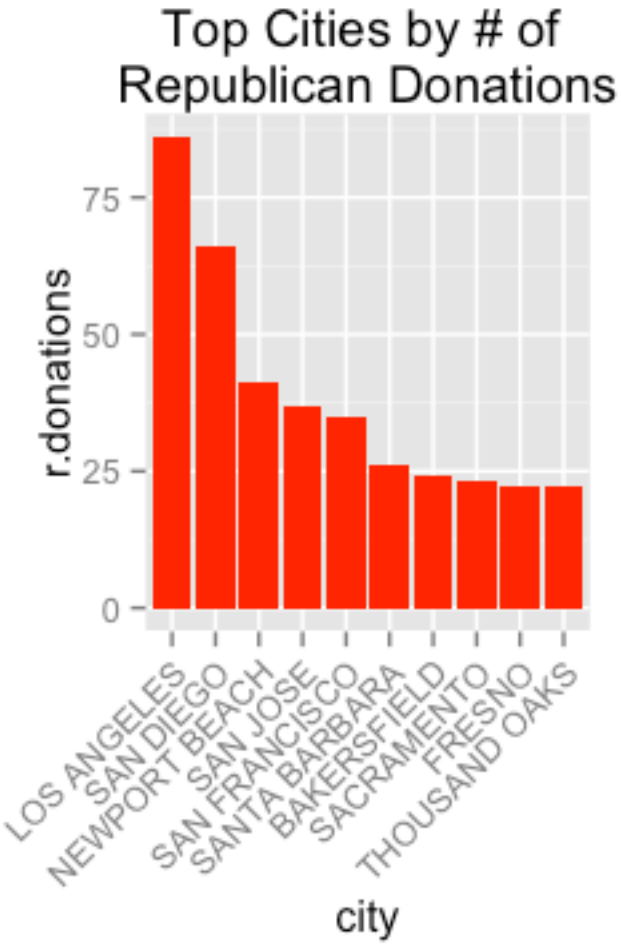

1000
750
500
250
0
amount
LOS ANGELES
SAN FRANCISCO
SAN DIEGO
BERKELEY
PALO ALTO
SANTA BARBARA
BEVERLY HILLS
OAKLAND
SAN JOSE
SANTA MONICA
city

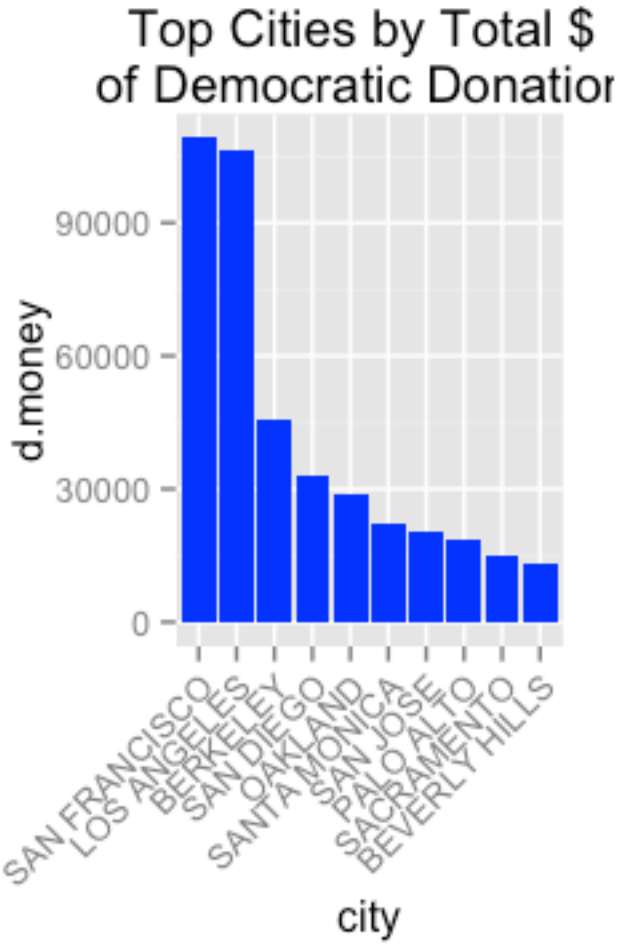


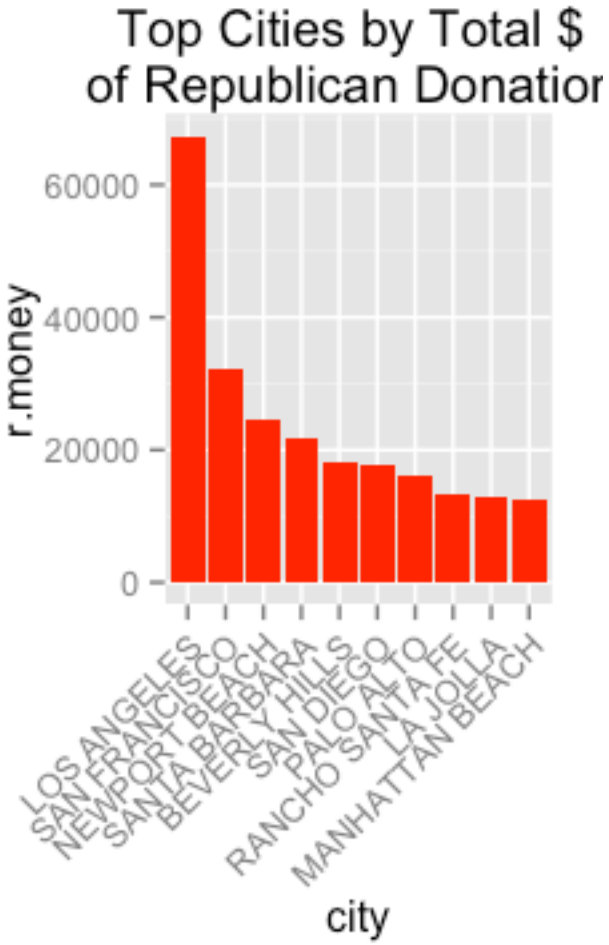
Clearly, some cities (like Newport Beach) donate at a rate far above those of other cities, hence why it makes the list despite only having a population of rouhly 85,000 (as opposed to Sacramento, population of approximately 480,000).

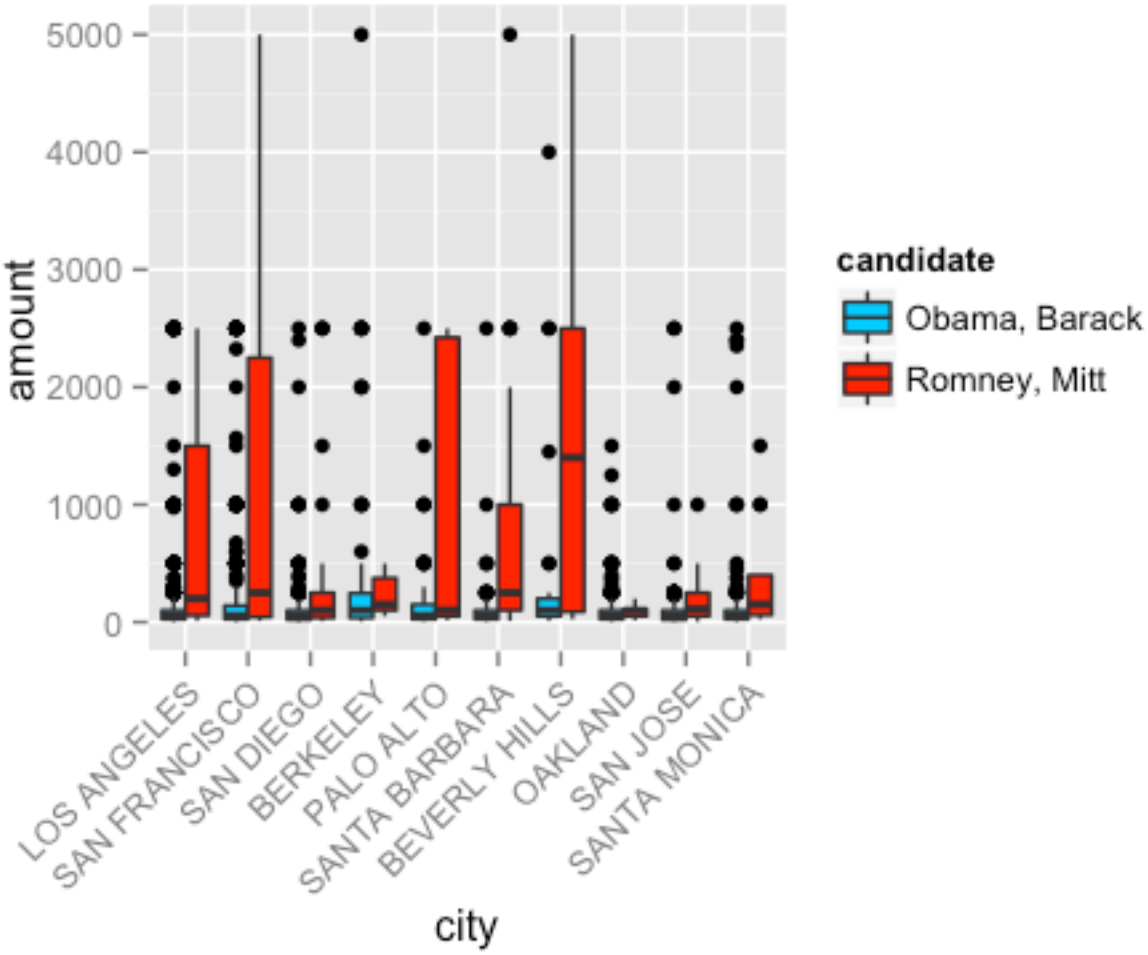
Next, I asked: What are the most partisan cities in terms of dollars donated and number of donations?

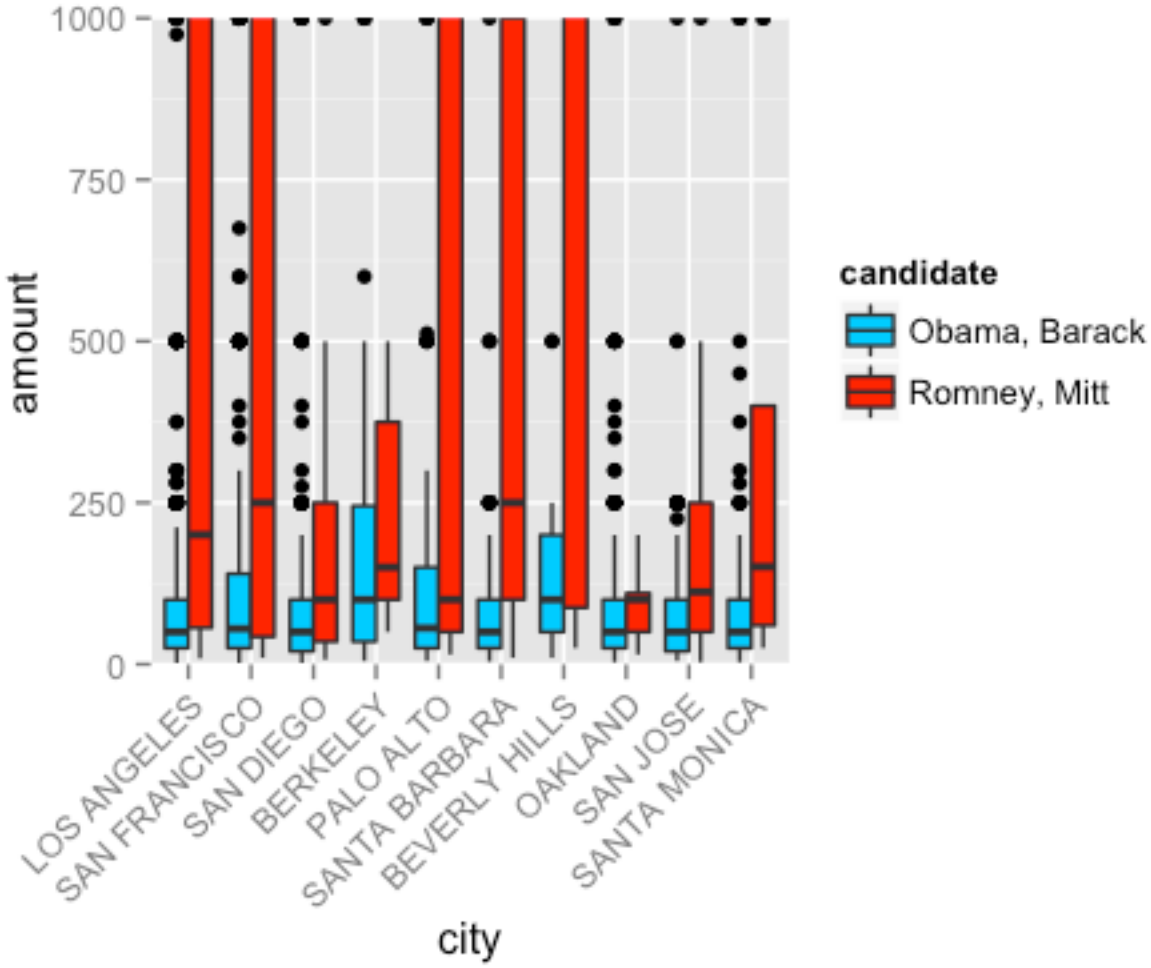

Top Cities by # of Democrat Donations
d.donations
600
400
200
0
LOS ANGELES
SAN FRANCISCO
SAN DIEGO
OAKLAND
SAN JOSE
BERKELEY
SACRAMENTO
PALO ALTO
SANTA MONICA
SANTA BARBARA
city

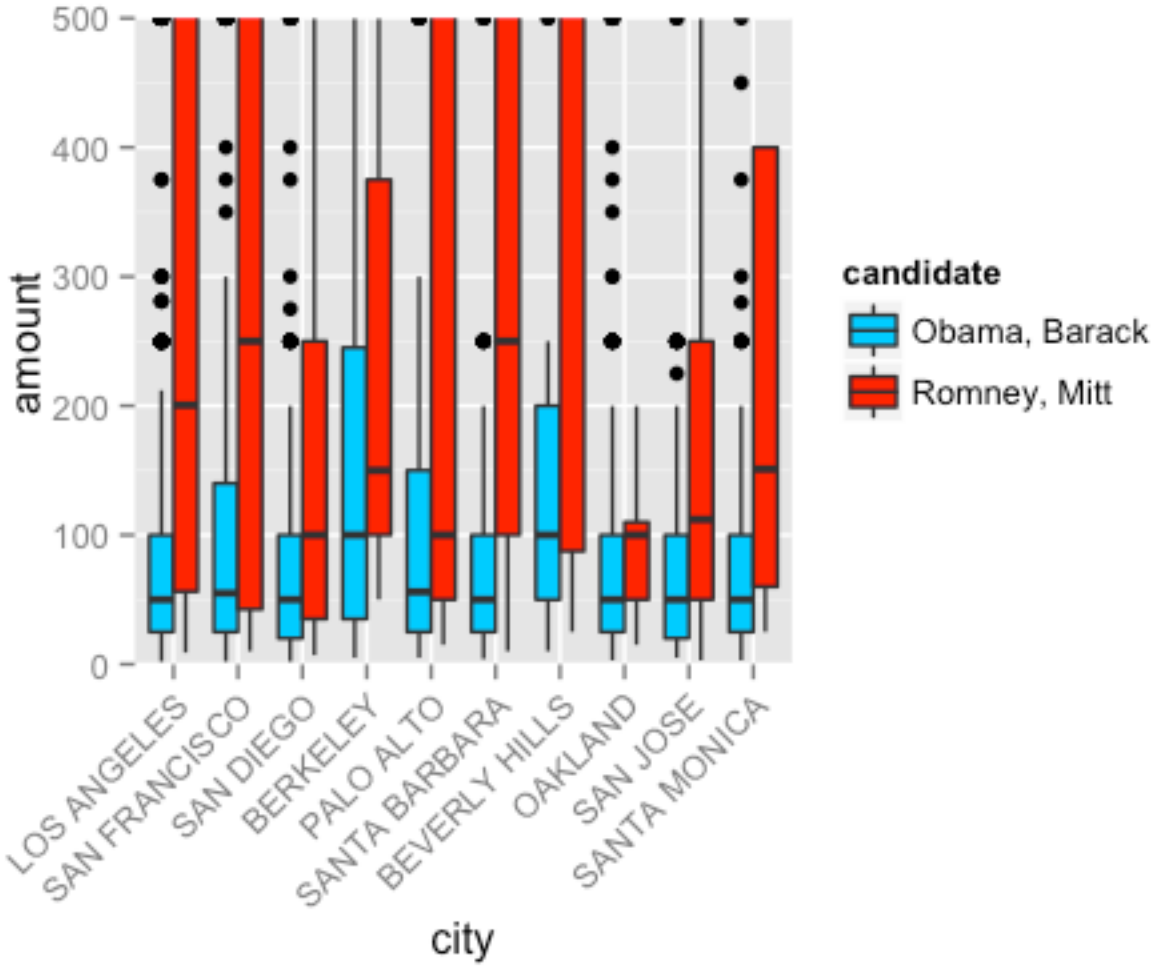

Top Cities by # of Republican Donations
r.donations
75
50
25
0
LOS ANGELES
SAN DIEGO
NEWPORT BEACH
SAN JOSE
SAN FRANCISCO
SANTA BARBARA
BAKERSFIELD
SACRAMENTO
FRESNO
THOUSAND OAKS
city


Top Cities by Total $
of Democratic Donatior
d.money
90000
60000
30000
0
SAN FRANCISCO
LOS ANGELES
BERKELEY
SAN DIEGO
OAKLAND
SANTA MONICA
SAN JOSE
PALO ALTO
SACRAMENTO
BEVERLY HILLS
city


Top Cities by Total $
of Republican Donatior
r.money
60000
40000
20000
0
LOS ANGELES
SAN FRANCISCO
NEWPORT BEACH
SANTA BARBARA
BEVERLY HILLS
SAN DIEGO
PALO ALTO
RANCHO SANTA FE
LA JOLLA
MANHATTAN BEACH
city

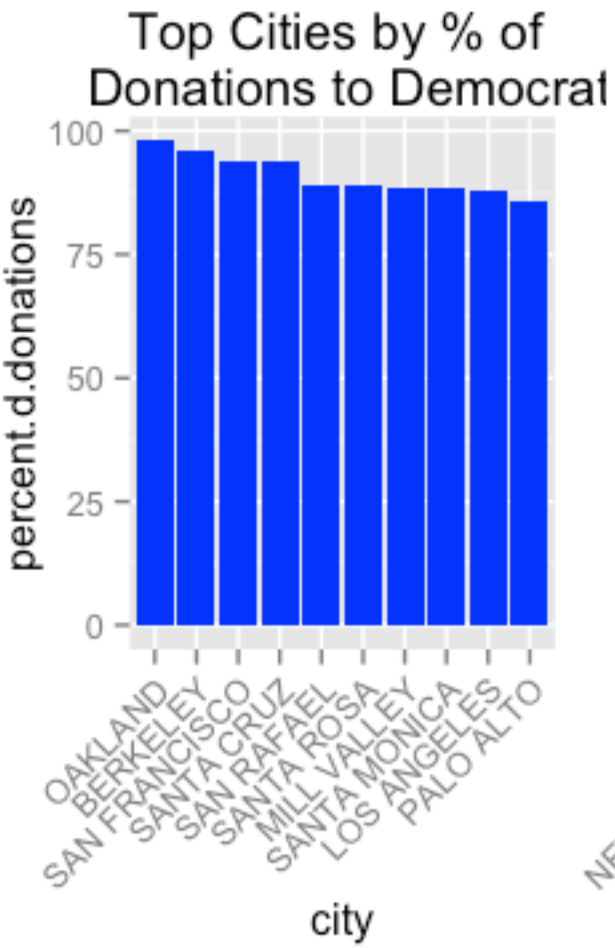

5000
4000
3000
2000
1000
0
amount
LOS ANGELES
SAN FRANCISCO
SAN DIEGO
BERKELEY
PALO ALTO
SANTA BARBARA
BEVERLY HILLS
OAKLAND
SAN JOSE
SANTA MONICA
city
candidate
Obama, Barack
Romney, Mitt

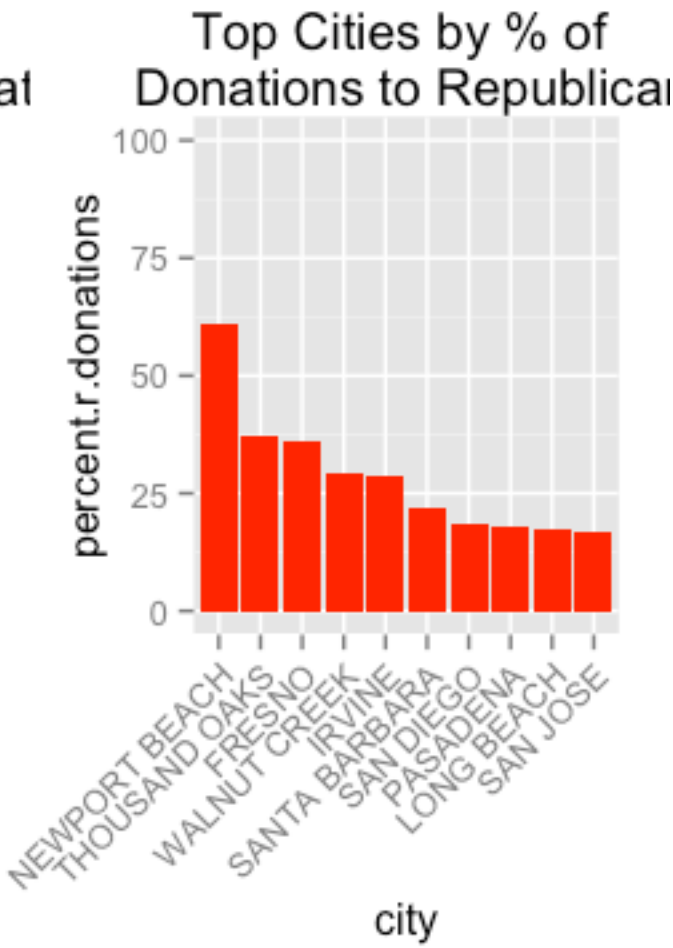

1000
750
500
250
0
amount
LOS ANGELES
SAN FRANCISCO
SAN DIEGO
BERKELEY
PALO ALTO
SANTA BARBARA
BEVERLY HILLS
OAKLAND
SAN JOSE
SANTA MONICA
city
candidate
Obama, Barack
Romney, Mitt

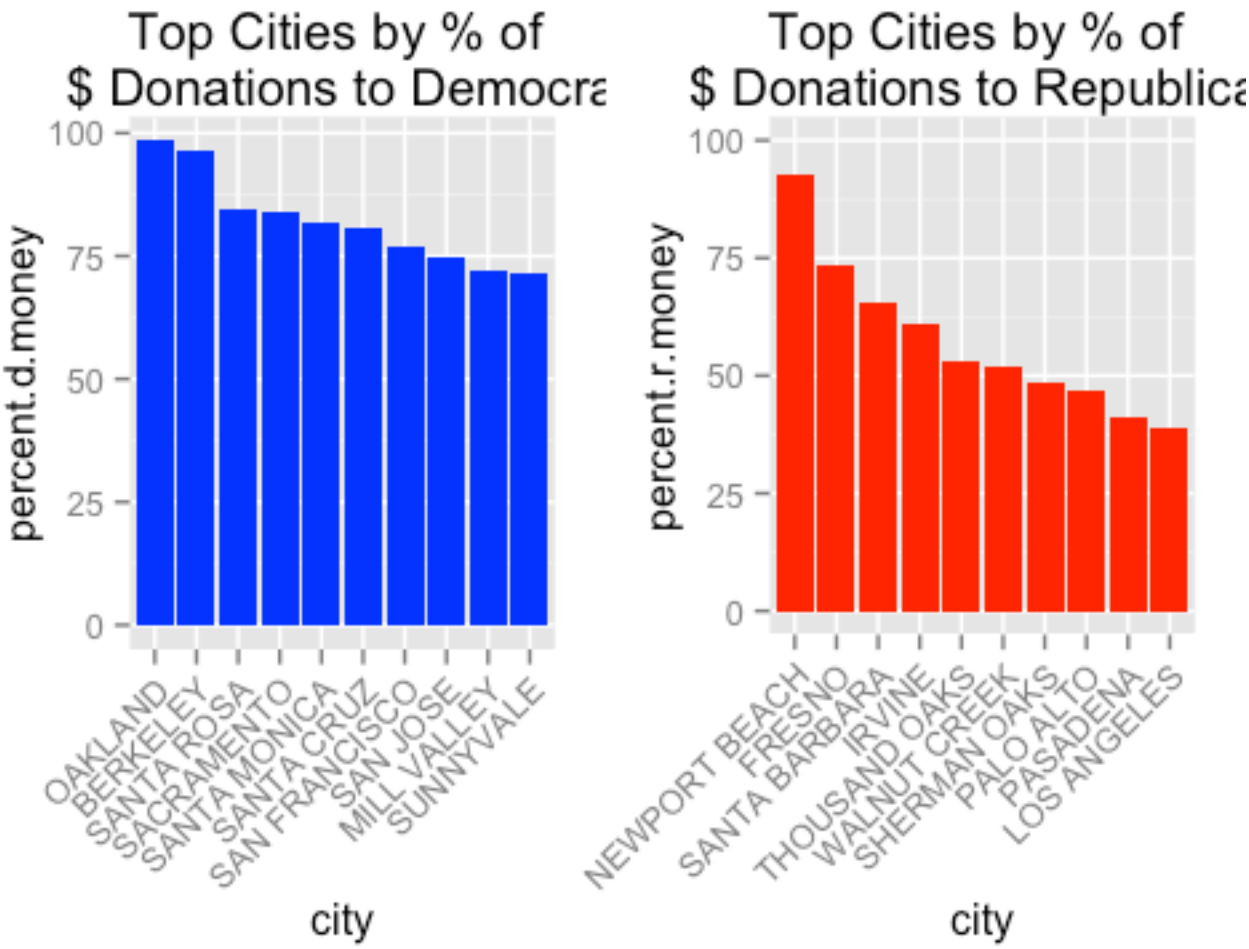


Clearly, city and candidate can have a strong relationship to amount.

Finally, I wanted to look at: what are the most partisan cities by percentage of contributions and by percentage of total money contributed?


Top Cities by % of Donations to Democrat
percent.d.donations
100
75
50
25
0
OAKLAND
BERKELEY
SAN FRANCISCO
SANTA CRUZ
SAN RAFAEL
SANTA ROSA
MILL VALLEY
SANTA MONICA
LOS ANGELES
PALO ALTO
city


Top Cities by % of Donations to Republica
percent.r.donations
100
75
50
25
0
NEWPORT BEACH
THOUSAND OAKS
FRESNO
WALNUT CREEK
IRVINE
SANTA BARBARA
SAN DIEGO
PASADENA
LONG BEACH
SAN JOSE
city



Note that for the analysis on cities above, a mimimum threshold was set at 50 donations for a city to be included in partisan city analysis.

Based on city data, it's clear that Republicans do best in small, affluent communities in Southern California. Democrats do best in Berkeley and small cities nearby in the Bay Area. A notable exception to this trend is Hollywood, a Democrat stronghold.