Financial Contributions to 2016 Presidential Campaign in New Jersey

Introduction

Data about financial contributions to the 2016 US Presidential Campaigns for the state of New Jersey was download on the 1st of September 2015 from http://fec.gov/disclosurep/PDownload.do.

There are mulitple candiates for each party (Republican and Democrat) still in the runing for the party nomination, so there are more than two candidates at the moment.

For the past 6 elections (since 1992), New Jersey has voted Democrat, and it will be interesting to see how this may effect current political leanings.

Load the data and relevant libraries. Examine the data for summary information.

```
##     cmte_id   cand_id                cand_nm          contbr_nm
## 1 C00575795 P00003392 Clinton, Hillary Rodham STRINGER, KRISTINE
## 2 C00575795 P00003392 Clinton, Hillary Rodham     CROTTY, SHEILA
## 3 C00575795 P00003392 Clinton, Hillary Rodham      MITZMAN, THEA
## 4 C00575795 P00003392 Clinton, Hillary Rodham        YURT, NURAY
## 5 C00575795 P00003392 Clinton, Hillary Rodham      NICOLO, MARIA
## 6 C00575795 P00003392 Clinton, Hillary Rodham      TALLAJ, RAMON
##    contbr_city contbr_st contbr_zip contbr_employer     contbr_occupation
## 1 SOUTH ORANGE        NJ   70792116   SELF-EMPLOYED              ATTORNEY
## 2      CLIFTON        NJ   70121939             N/A          NOT EMPLOYED
## 3     CALDWELL        NJ   70071406             N/A             HOMEMAKER
## 4   PISCATAWAY        NJ   88544546        NOVARTIS              DIRECTOR
## 5   TITUSVILLE        NJ   85601724   SELF-EMPLOYED INFORMATION REQUESTED
## 6      PARAMUS        NJ   76525505   SELF-EMPLOYED             PHYSICIAN
##   contb_receipt_amt contb_receipt_dt receipt_desc memo_cd memo_text
## 1               250        12-Apr-15
## 2               100        27-Apr-15
## 3              2700        29-May-15
## 4              2700        27-Apr-15
## 5              2700        29-Jun-15
## 6              2700        30-Apr-15
##   form_tp file_num tran_id election_tp
## 1   SA17A  1015585  C19928       P2016
## 2   SA17A  1015585  C87019       P2016
## 3   SA17A  1015585 C176829       P2016
## 4   SA17A  1015585  C77059       P2016
## 5   SA17A  1015585 C292569       P2016
## 6   SA17A  1015585  C88649       P2016
```

```
##  [1] "cmte_id"          "cand_id"          "cand_nm"
##  [4] "contbr_nm"        "contbr_city"      "contbr_st"
##  [7] "contbr_zip"       "contbr_employer"  "contbr_occupation"
## [10] "contb_receipt_amt" "contb_receipt_dt" "receipt_desc"
## [13] "memo_cd"          "memo_text"        "form_tp"
## [16] "file_num"         "tran_id"          "election_tp"
```

```
## [1] 2435   18
```

```
## 'data.frame':    2435 obs. of  18 variables:
##  $ cmte_id          : Factor w/ 14 levels "C00458844","C00500587",..: 6 6 6 6 6 6 6 6 6 10 ...
##  $ cand_id          : Factor w/ 14 levels "P00003392","P20002721",..: 1 1 1 1 1 1 1 1 1 10 ...
##  $ cand_nm          : Factor w/ 15 levels "Bush, Jeb","Carson, Benjamin S.",..: 3 3 3 3 3 3 3 3 3 10
##  $ contbr_nm        : Factor w/ 1353 levels "ABDELAZIZ, AL",..: 1208 230 842 1344 895 1223 755 74 894
##  $ contbr_city      : Factor w/ 346 levels "ALLENDALE","ALLENHURST",..: 289 57 43 247 302 234 100 194
##  $ contbr_st        : Factor w/ 1 level "NJ": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contbr_zip       : int  70792116 70121939 70071406 88544546 85601724 76525505 70245022 70423025 71
##  $ contbr_employer  : Factor w/ 692 levels "","A&E STORES",..: 562 415 415 450 562 562 415 122 592 56
##  $ contbr_occupation: Factor w/ 445 levels "","ACADEMIC",..: 23 281 206 118 214 314 206 241 339 314
##  $ contb_receipt_amt: num  250 100 2700 2700 2700 2700 2700 2700 50 2700 ...
##  $ contb_receipt_dt : Factor w/ 117 levels "01-Apr-15","01-Jun-15",..: 40 98 109 98 107 110 28 95 28
##  $ receipt_desc     : Factor w/ 10 levels "","REATTRIBUTION / REDESIGNATION REQUESTED (AUTOMATIC)",..
##  $ memo_cd          : Factor w/ 2 levels "","X": 1 1 1 1 1 1 1 1 1 1 ...
##  $ memo_text        : Factor w/ 21 levels "","* EARMARKED CONTRIBUTION: SEE BELOW",..: 1 1 1 1 1 1 1
##  $ form_tp          : Factor w/ 3 levels "SA17A","SA18",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ file_num         : int  1015585 1015585 1015585 1015585 1015585 1015585 1015585 1015585 1015585 10
##  $ tran_id          : Factor w/ 2435 levels "A032126DA61AA40D699B",..: 491 1215 436 1167 900 1220 513
##  $ election_tp      : Factor w/ 2 levels "G2016","P2016": 2 2 2 2 2 2 1 2 2 2 ...
```

```
##       cmte_id          cand_id                          cand_nm
##  C00575795:1083   P00003392:1083   Clinton, Hillary Rodham  :1083
##  C00577130: 276   P60007168: 276   Sanders, Bernard         : 276
##  C00574624: 259   P60006111: 259   Carson, Benjamin S.      : 230
##  C00573519: 230   P60005915: 230   Cruz, Rafael Edward 'Ted': 230
##  C00458844: 208   P60006723: 208   Rubio, Marco             : 208
##  C00575449: 170   P40003576: 170   Paul, Rand               : 170
##  (Other)  : 209   (Other)  : 209   (Other)                  : 238
##                contbr_nm         contbr_city   contbr_st   contbr_zip
##  SACKS-WILNER, TOM :  19   PRINCETON  :  66   NJ:2435   Min.   :    7008
##  LORENZO, CAREY    :  15   HOBOKEN    :  54             1st Qu.:70783015
##  SPAIR SR, RICHAERD:  15   MONTCLAIR  :  49             Median :77242352
##  EDWARDS, DIANE    :  14   WEST ORANGE:  48             Mean   :74451818
##  HESS, CHARLES W.  :  14   MORRISTOWN :  47             3rd Qu.:80572352
##  STORCH, EVELYN    :  14   CHERRY HILL:  42             Max.   :89042725
##  (Other)           :2344   (Other)    :2129
##                                  contbr_employer
##  RETIRED                                   : 338
##  SELF-EMPLOYED                             : 214
##  N/A                                       : 207
##  NOT EMPLOYED                              :  96
##  INFORMATION REQUESTED PER BEST EFFORTS:  89
##  (Other)                                   :1489
##  NA's                                      :   2
##                               contbr_occupation contb_receipt_amt
##  RETIRED                                   : 444    Min.   :-5000.0
##  ATTORNEY                                  : 150    1st Qu.:   50.0
##  NOT EMPLOYED                              : 108    Median :  143.5
##  INFORMATION REQUESTED PER BEST EFFORTS:  80    Mean   :  669.4
##  HOMEMAKER                                 :  72    3rd Qu.: 1000.0
##  (Other)                                   :1580    Max.   : 5400.0
##  NA's                                      :   1
```

```
##   contb_receipt_dt
##  30-Jun-15: 189
##  29-Jun-15:  93
##  12-Apr-15:  82
##  23-Jun-15:  72
##  26-Jun-15:  68
##  12-Jun-15:  62
##  (Other)  :1869
##                                                receipt_desc  memo_cd
##                                                      :2402    :2378
##  Refund                                              :   9   X:  57
##  REATTRIBUTION / REDESIGNATION REQUESTED (AUTOMATIC):   5
##  REATTRIBUTION FROM SPOUSE                           :   3
##  REATTRIBUTION TO SPOUSE                             :   3
##  REDESIGNATION FROM PRIMARY                          :   3
##  (Other)                                             :  10
##                                                 memo_text     form_tp
##                                                      :2108   SA17A:2394
##  * EARMARKED CONTRIBUTION: SEE BELOW                 : 255   SA18 :  32
##  EARMARKED FROM MAKE DC LISTEN                       :  35   SB28A:   9
##  REATTRIBUTION / REDESIGNATION REQUESTED (AUTOMATIC):   5
##  REATTRIBUTION FROM SPOUSE                           :   3
##  REATTRIBUTION TO SPOUSE                             :   3
##  (Other)                                             :  26
##     file_num                       tran_id     election_tp
##  Min.   :1003942   A032126DA61AA40D699B:   1   G2016:  31
##  1st Qu.:1015509   A03612478EFDA491AB11:   1   P2016:2404
##  Median :1015585   A04059564B8CB422CA72:   1
##  Mean   :1015272   A06CBD04D2CBB4D29B7B:   1
##  3rd Qu.:1015585   A06F4FE70F5794854B7D:   1
##  Max.   :1015715   A0B430521A50B4B038B3:   1
##                    (Other)             :2429
```

What do the variables in the data mean ?

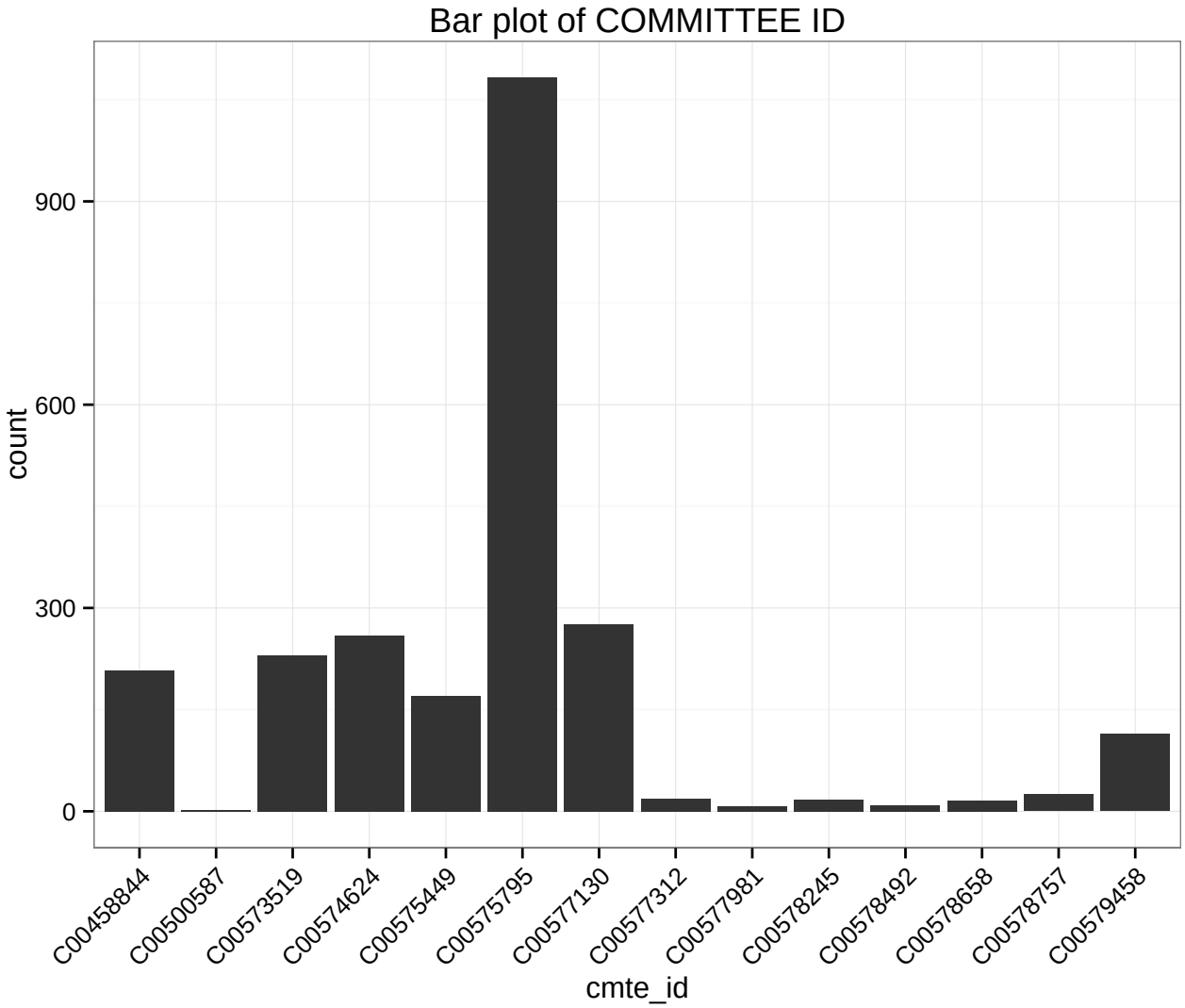
CMTE_ID = COMMITTEE ID
CAND_ID = CANDIDATE ID
CAND_NM = CANDIDATE NAME
CONTBR_NM = CONTRIBUTOR NAME
CONTBR_CITY = CONTRIBUTOR CITY CONTBR_ST = CONTRIBUTOR STATE CONTBR_ZIP = CONTRIBUTOR ZIP CODE CONTBR_EMPLOYER = CONTRIBUTOR EMPLOYER CONTBR_OCCUPATION = CONTRIBUTOR OCCUPATION CONTB_RECEIPT_AMT = CONTRIBUTION RECEIPT AMOUNT CONTB_RECEIPT_DT = CONTRIBUTION RECEIPT DAT RECEIPT_DESC = RECEIPT DESCRIPTION
MEMO_CD = MEMO CODE MEMO_TEXT = MEMO TEXT FORM_TP = FORM TYPE FILE_NUM = FILE NUMBER TRAN_ID = TRANSACTION ID ELECTION_TP = ELECTION TYPE/PRIMARY GENERAL INDICATOR

Analysis

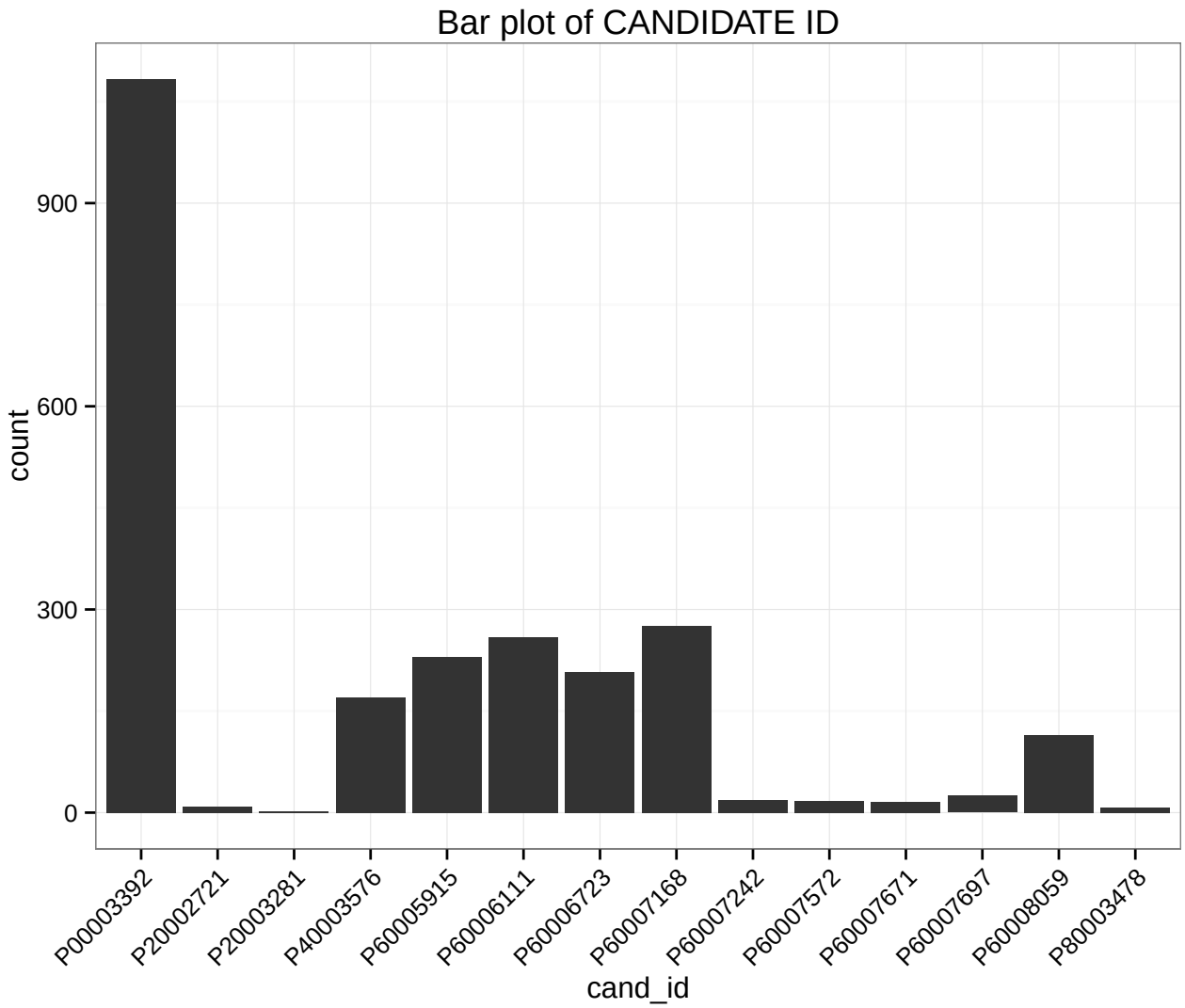
Univariate Analysis

I'm going to start by just getting to know the data, as I've never worked with it before. I find this easiest by plotting the variables and getting some summary informations for them.

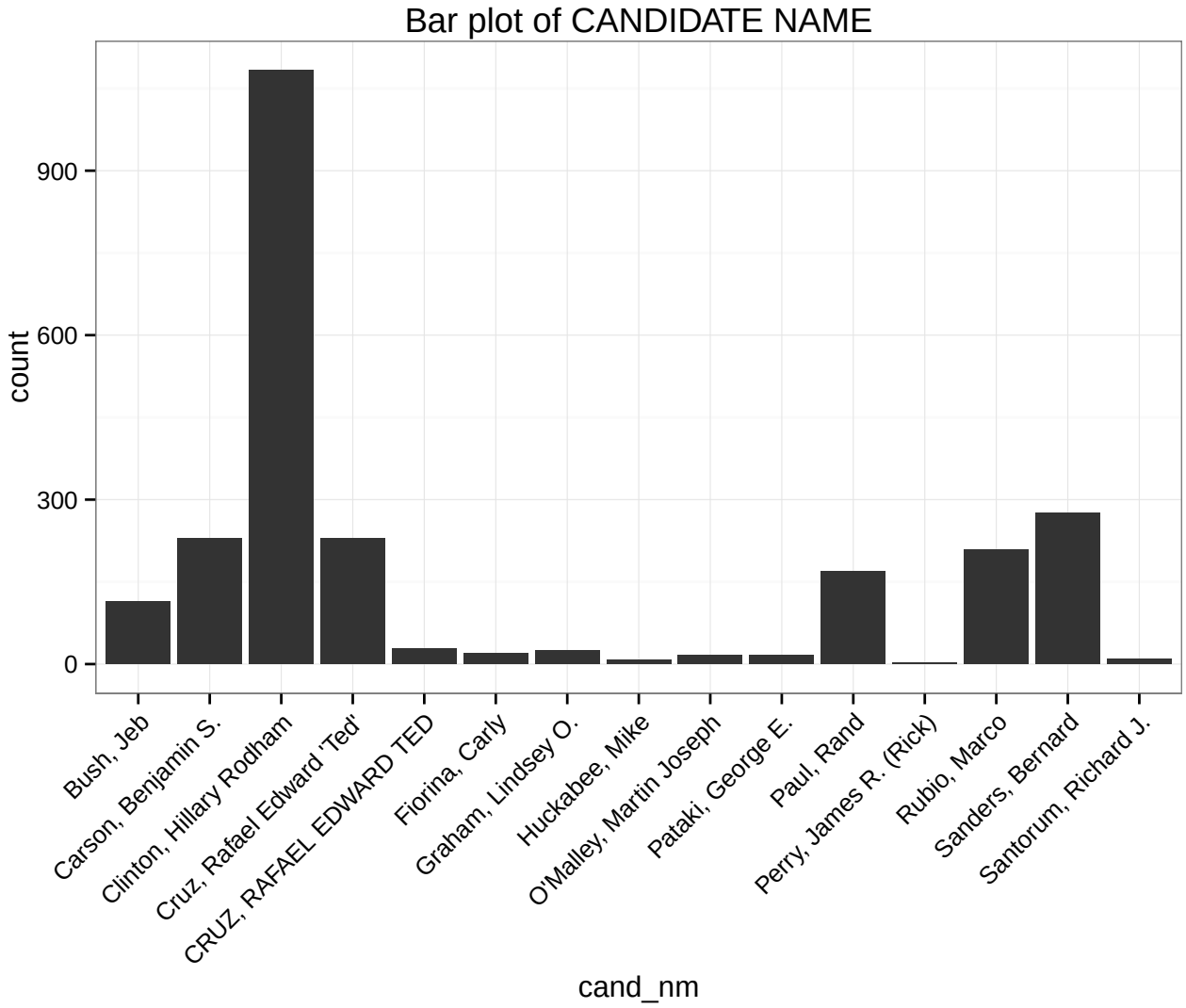
Examine the variable cmte_id



Most people are donating to one committee predominantly.



Most people are donating to one candidate predominantly.



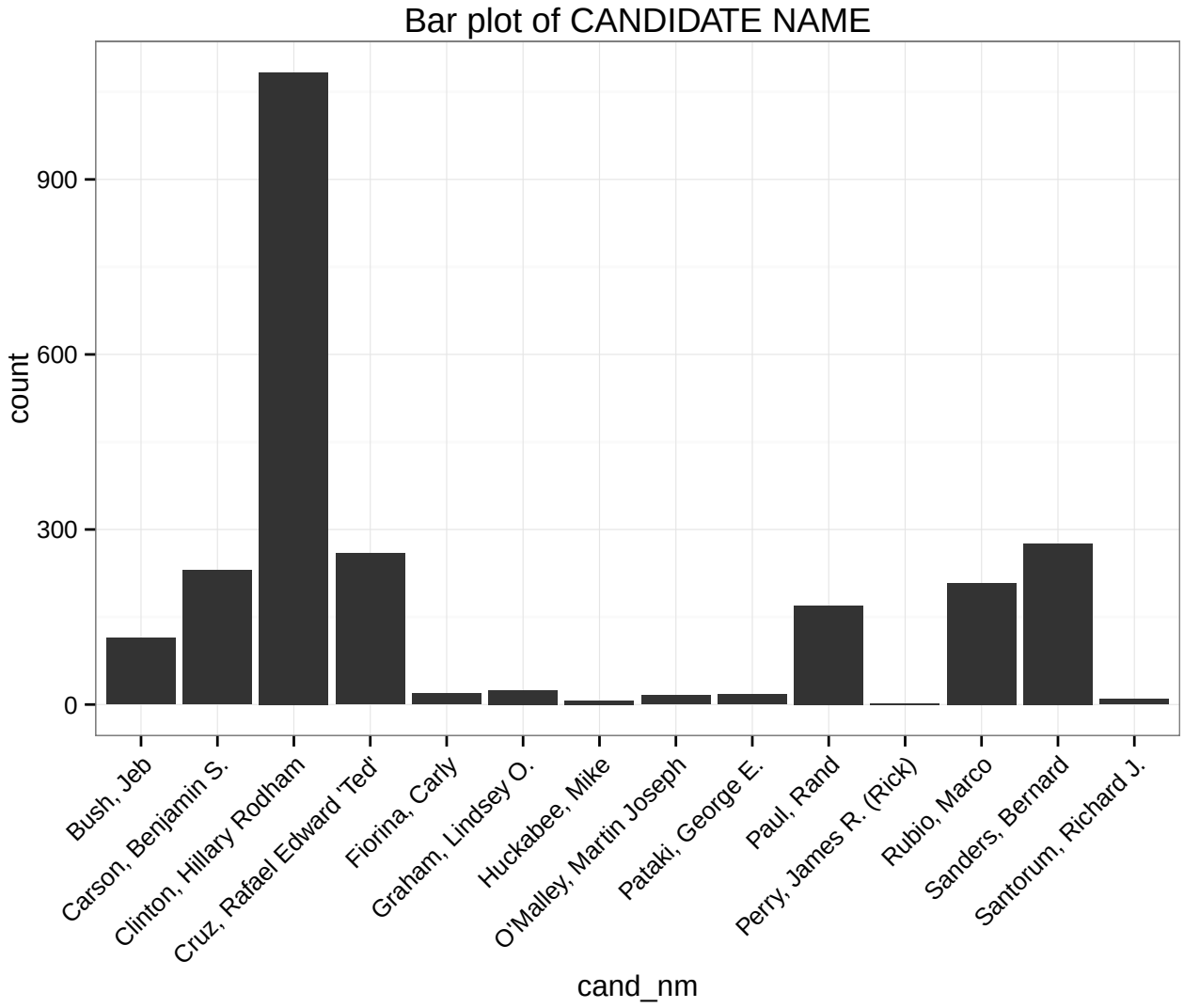
Most people are donating to Hilary Clinton (Democrat), which I think is the same information capture in the previous two plots.

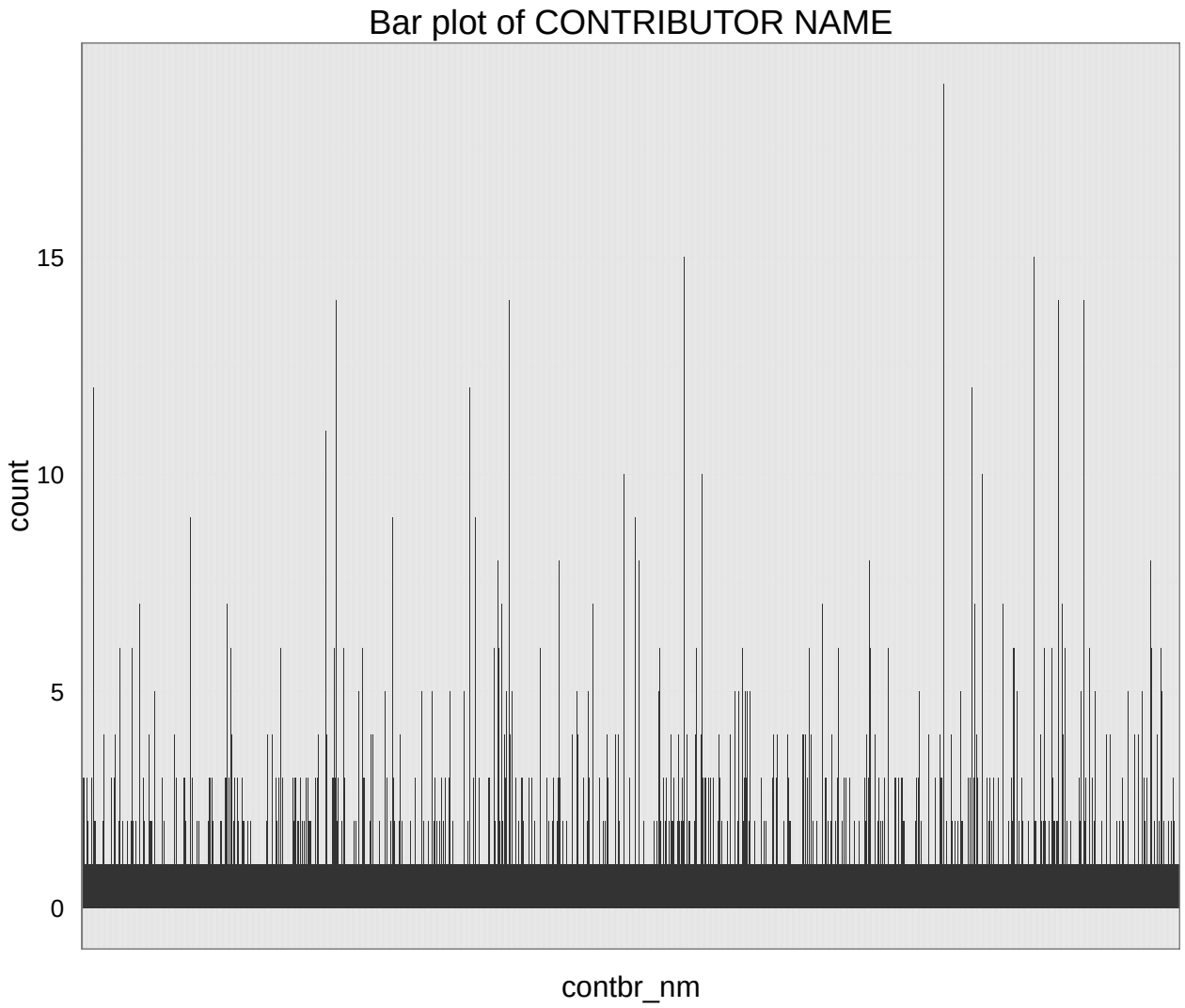
From the plot, I did noticed an issue with one of the candidates names. Ted Cruz is listed twice (once in all upper case). This will be problematic if not corrected as we would incorrectly make conclusions about the data. This can be easily fixed.

```
##
##                Bush, Jeb       Carson, Benjamin S.
##                      114                       230
##   Clinton, Hillary Rodham Cruz, Rafael Edward 'Ted'
##                     1083                       259
##   CRUZ, RAFAEL EDWARD TED            Fiorina, Carly
##                         0                        19
##        Graham, Lindsey O.            Huckabee, Mike
##                        25                         7
##   O'Malley, Martin Joseph         Pataki, George E.
##                        16                        17
##               Paul, Rand    Perry, James R. (Rick)
##                      170                         2
##             Rubio, Marco          Sanders, Bernard
```

```
##                         208                         276
##      Santorum, Richard J.
##                           9
```

Looks like it is fixed. Lets remake the plot.





There are many different contributors but there are some that contribute more than once. But this plot is too full to make much sense of the data. I'll count the number of times (frequency) each individual contributor name occurs (new variable called "count_NM").

```
## [1] 1353    2
```

How many unique individual contributors there are?

```
## [1] 929   2
```

How many have indivdiuals have made over 10 contributions?

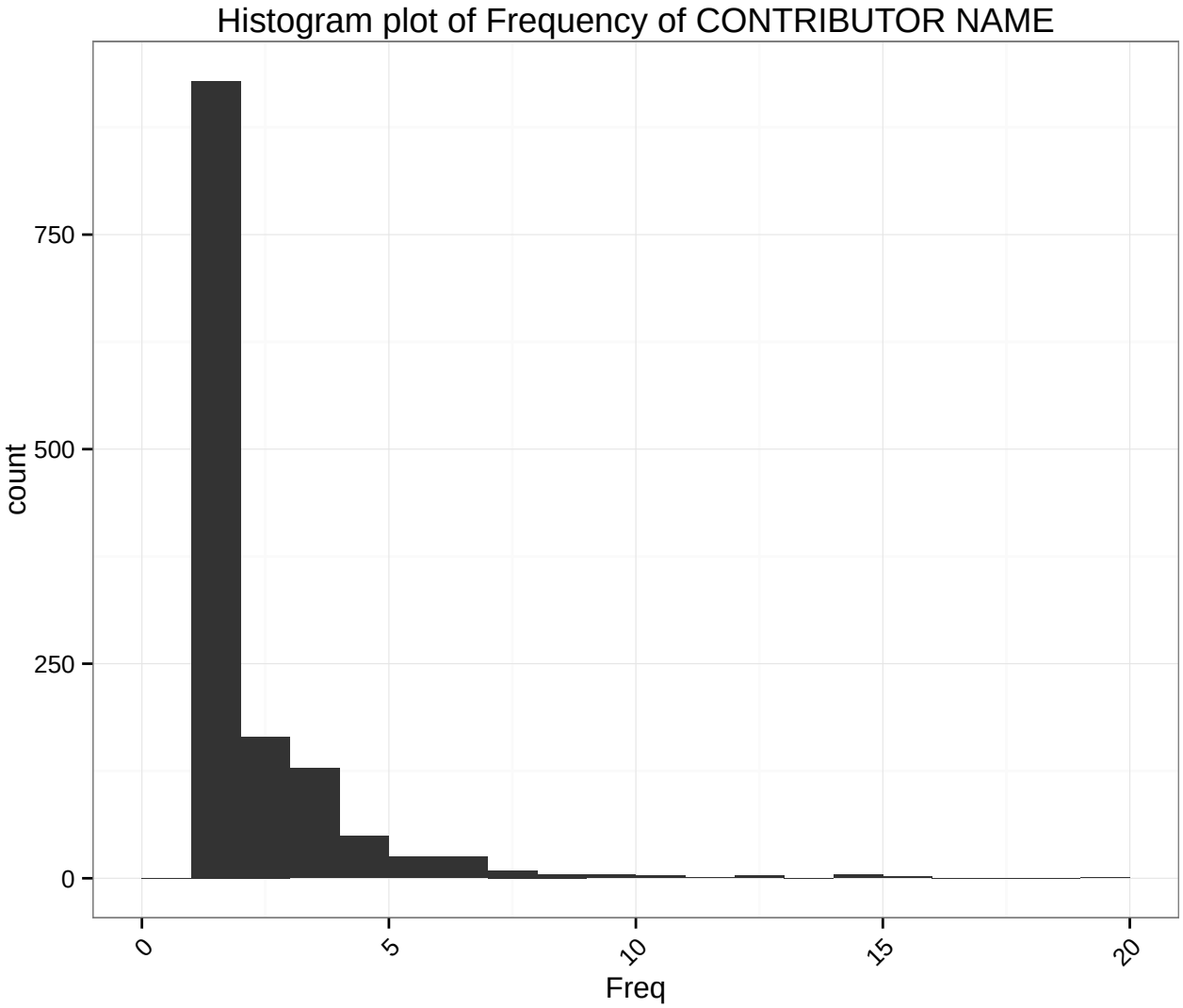
```
## [1] 11  2
```

Who is the most frequency contributors.

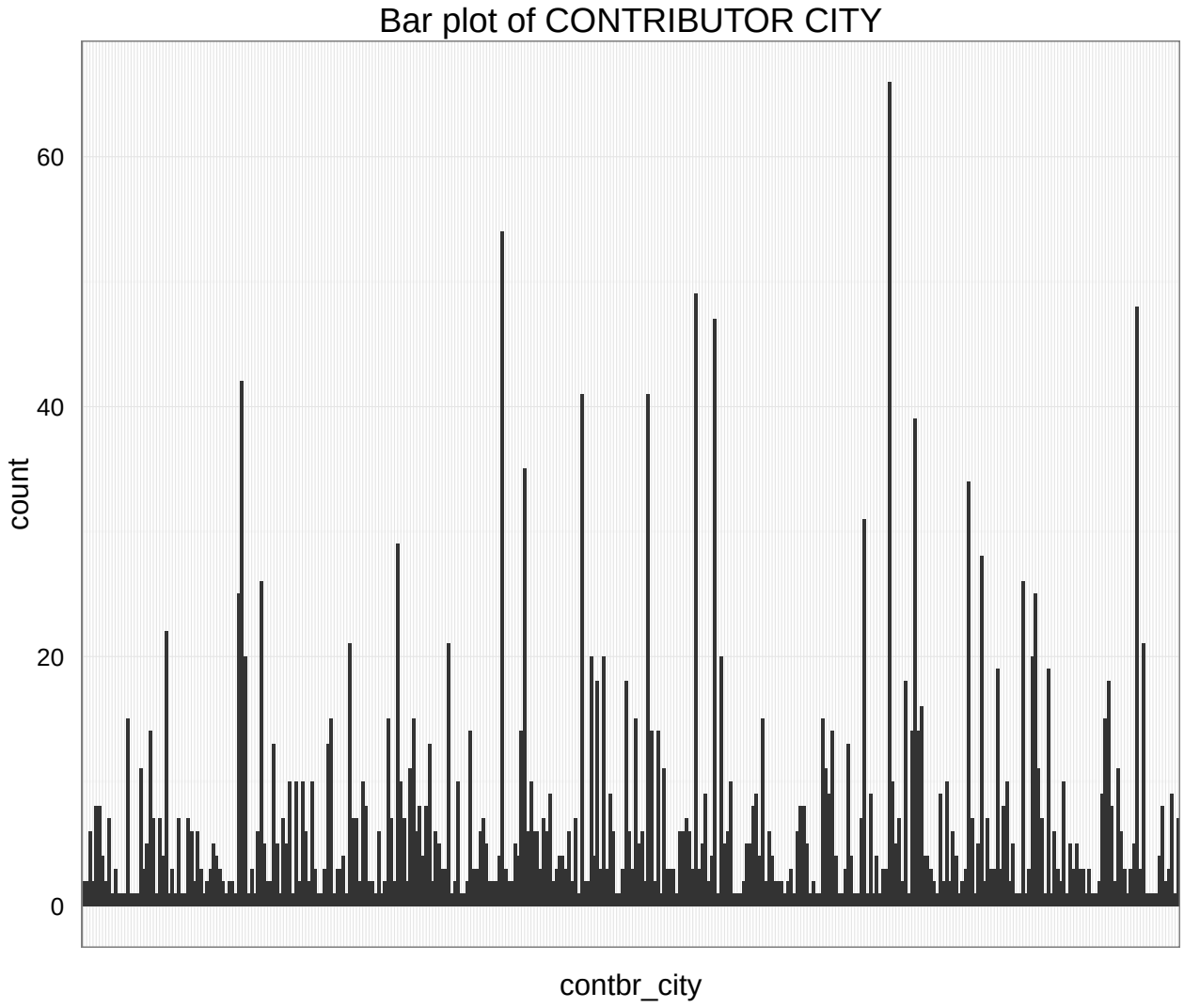
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     1.0     1.0     1.0     1.8     2.0    19.0
```

```
##                    Var1 Freq
## 1062 SACKS-WILNER, TOM   19
```

The new variable is a count of the number of times each individual contributor name occurs, and this can be plotted to looked at the distribution.



From the histogram for the number of times an indidivual donated, we can see that most people donate only once, and few donate more than 5 times. There are 1353 unique donators in the file of 2435 donations. Of those 929 have donated only once, with the maximum number of donations by a single person being 19 (listed as SACKS-WILNER, TOM). Only 11 individuals have donated more than 10 times.



The plot is too full to make much sense of it, but it is clear that some cities have more people making campaign donations than others. I'll again count the number of occurances of a city (new variable called "count_CITY").

```
## [1] 346   2
```

How many unique cities are listed?

```
## [1] 68  2
```

How many cities have made over 10 contributions?

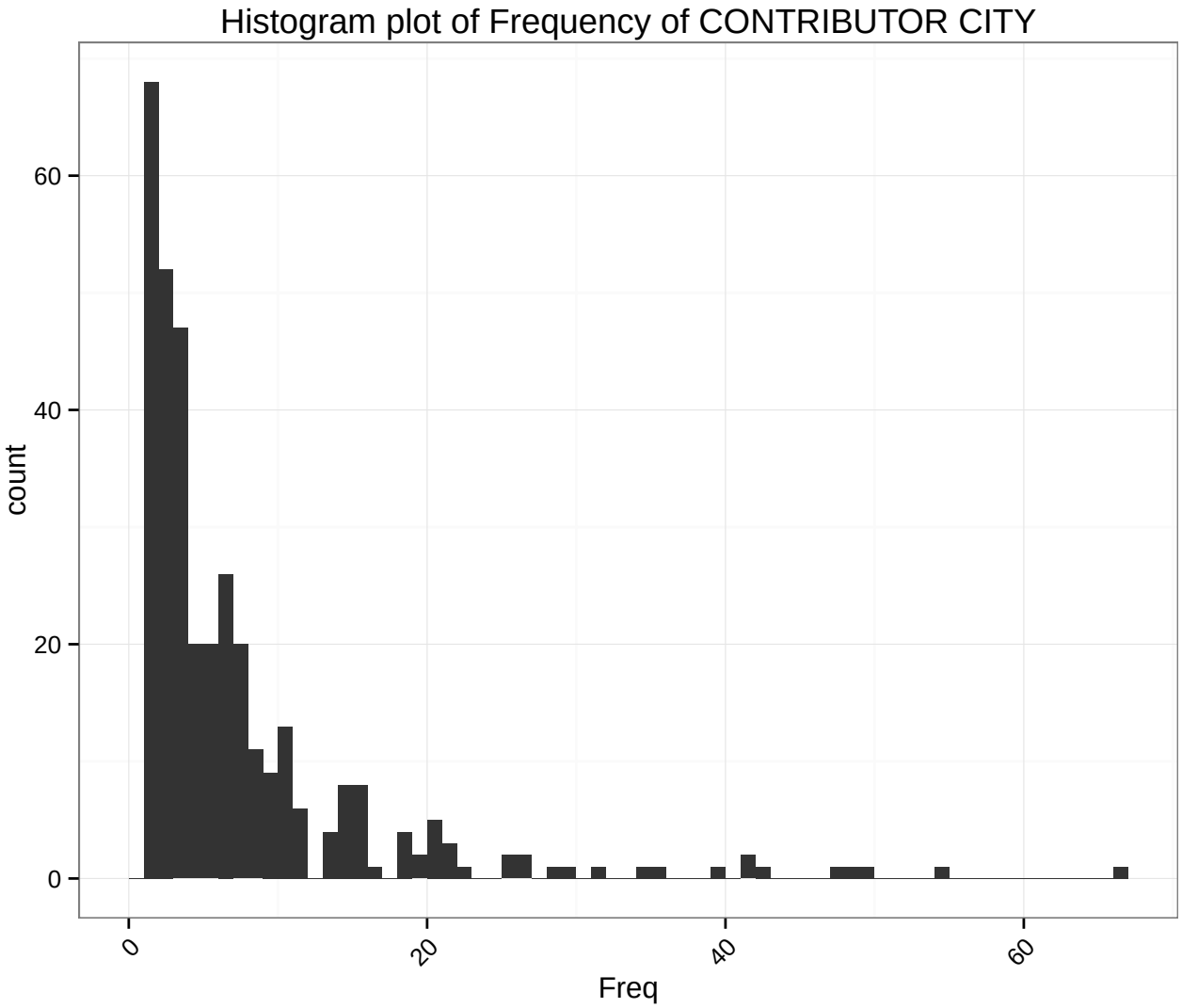
```
## [1] 60  2
```

Which city is the most frequently listed (city with most contributors)?

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   4.000   7.038   8.000  66.000
```

```
##          Var1 Freq
## 255 PRINCETON   66
```

This new variables can also be plotted to look at the distribution.



Many cities only have few people donating to campaigns, but there are very active cities, with the maximum number of people donating from a single city being 66 (city is listed as Princeton). There are 346 unique cities in the file of 2435 donations. Of those 68 are listed only once. 60 cities are listed more than 10 times.

Does the zip code variable give any additional information? I'll again count the number of observations of a zip code (new variable called "count_ZIP").

```
## [1] 1219    2
```

How many unique zip codes are listed?

```
## [1] 709   2
```

How many zip codes are listed in the dataset more than 10 times?

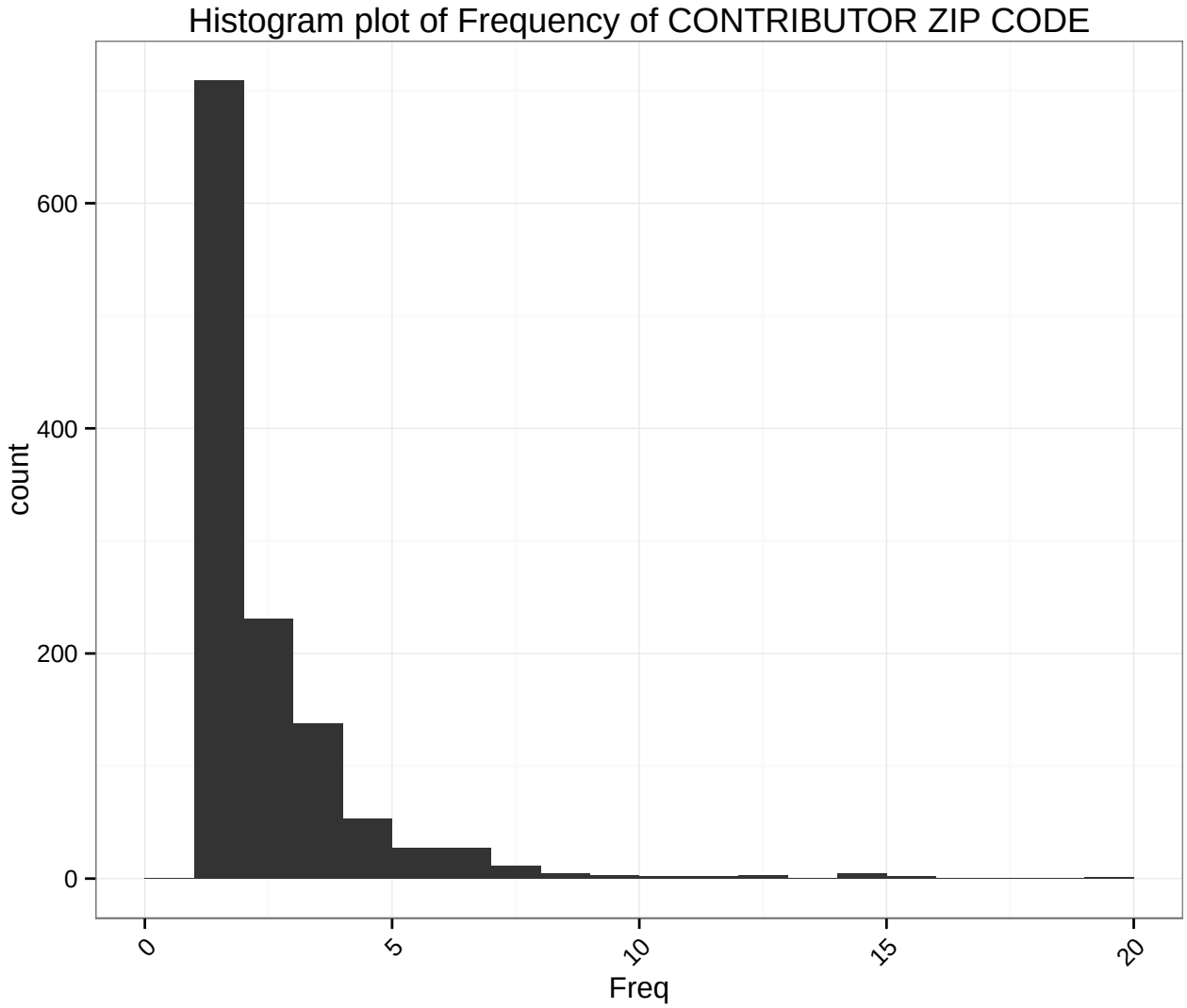
```
## [1] 13  2
```

Which zip codes is the most frequently listed zip codes (zip code with most contributors)?

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   1.998   2.000  19.000
```

```
##         Var1 Freq
## 922 80559348   19
```

This new variables can also be plotted to look at the distribution.



It does. The tail does not go to as high a number as the city variable (max was 66, zip max is 19), suggesting that zip while obviously highly related to city does give slightly different information, with there being more zip codes in the dataset (unique zip codes = 1219) than cities (unique cities = 346). The zip code with the largrest number of donations is 08055-9348, which is for Medford, and that is different from the city with the most donations which was Princeton. Perhaps zip code offers a greater resolution to location of an individuals by giving their lcoation within a city as well as city.

I will assess employer information in the same way, creating a new variable called "count_EMPLOYER" which counts the frequency of each individual employer listed.

```
## [1] 692   2
```

Count the number of unique employers.

```
## [1] 446   2
```

How many employers are listed more than 10 times?

```
## [1] 18  2
```

List the most frequently listed employer from the contributors.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   3.516   2.000 338.000
```

```
##        Var1 Freq
## 529 RETIRED  338
```

There are 692 unique employers listed, most of which are single related to individuals who only gave a single donation (446). The maximum number of times the same employer is list is 338, which Retired. This variable may not be as relevant as CONTBR_OCCUPATION.

Occupation will be assessed the same as the others, creating a new variable called "count_OCCUPATION" which counts the frequency of each individual occupation listed.

```
## [1] 445   2
```

Count the number of unique occupations.

```
## [1] 224   2
```

How many occupations are listed more than 10 times?

```
## [1] 30  2
```

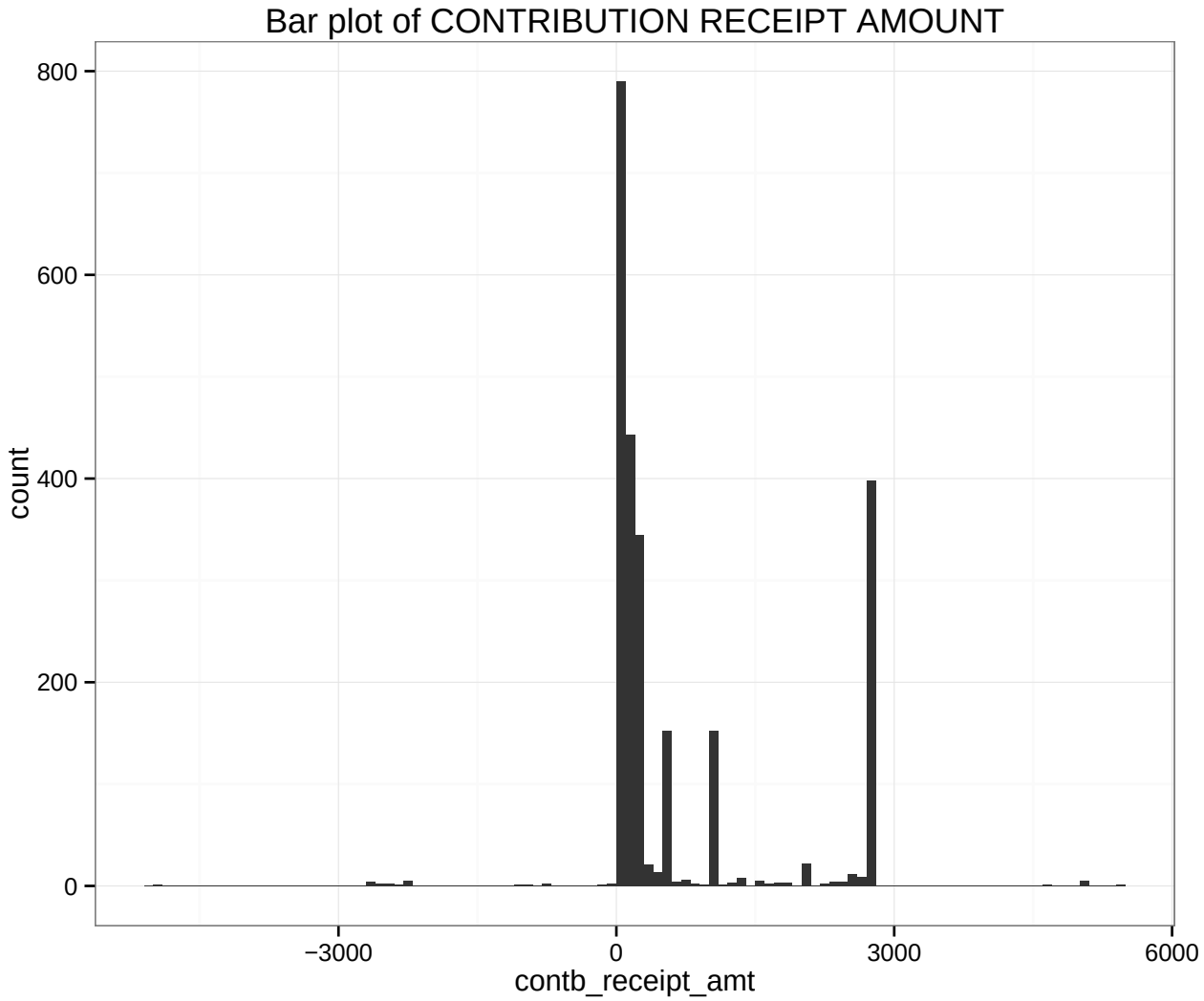
List the most frequently named occupation for the contributors, and the top 10 most frequent occupations listed.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    1.00    1.00    5.47    3.00  444.00
```

```
##        Var1 Freq
## 359 RETIRED  444
```

```
##                                      Var1 Freq
## 359                               RETIRED  444
## 23                               ATTORNEY  150
## 281                         NOT EMPLOYED  108
## 215 INFORMATION REQUESTED PER BEST EFFORTS   80
## 206                             HOMEMAKER   72
## 96                             CONSULTANT   64
## 214                INFORMATION REQUESTED   62
## 314                             PHYSICIAN   52
## 241                                LAWYER   41
## 49                                    CEO   38
```

Interestingly this variable list more donations coming from Retired individuals than employer. There are less unique occupations, but still a large number (445). It would be nice to see this broken down into even broader categories. Might think about how best to handle this information. Could possible investigate just the occupations with the greatest number of contributors (like say the top 10).



The histogram tells us that most people are giving small amounts, with some larger donations. There is also a peak just under $3000. This is likely $2,700 which is the limit an individual may give to an individual candidate, and thus the peak is signifying the maximum contribution. Two things are very interesting: (1) There are donations above the limit of $2700. (2) There are some negative amounts in the contributions.

```
## [1] 22 18
```

```
## [1]  7 18
```

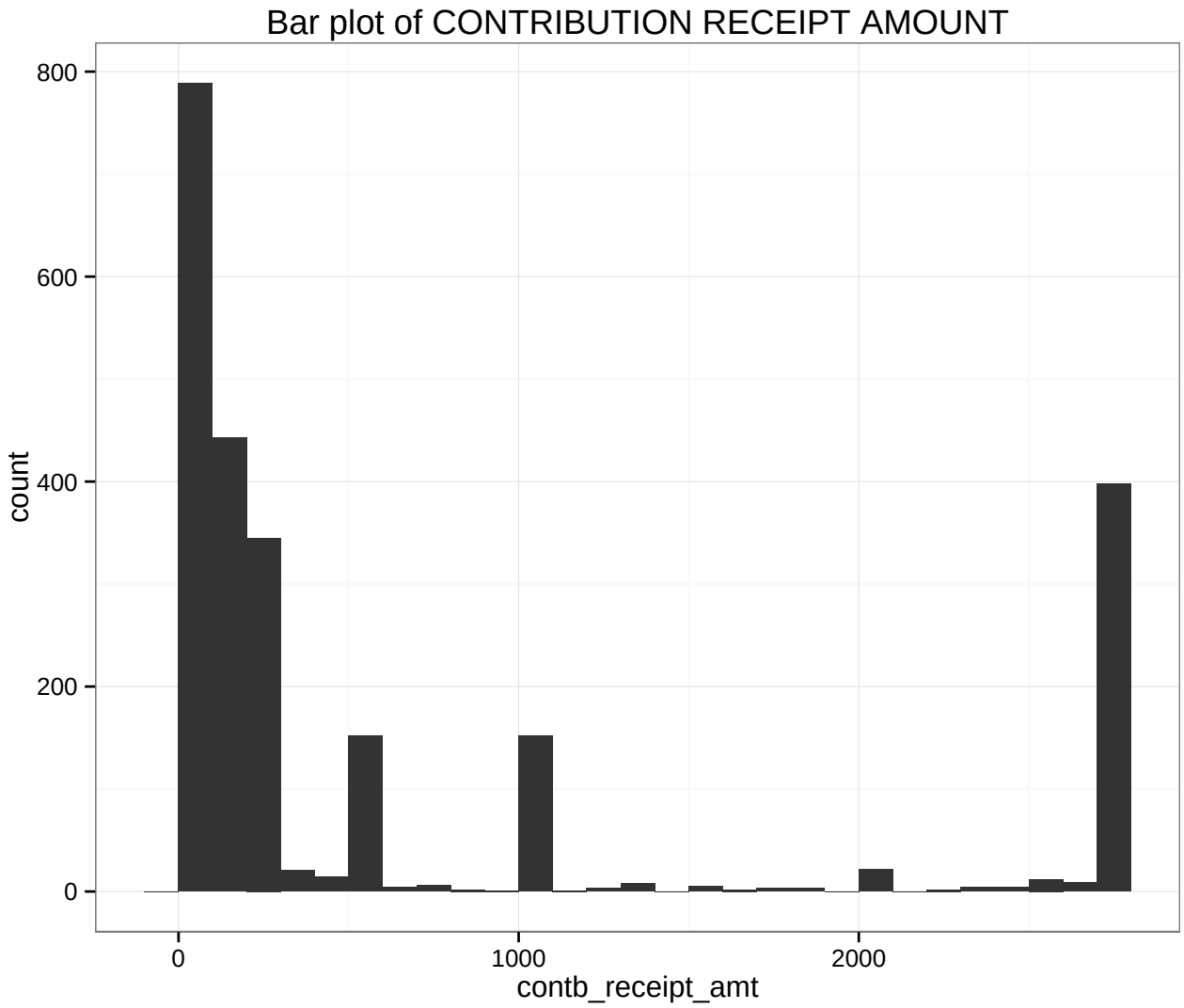
```
##                                                  Var1 Freq
## 1                                                       5
## 2  REATTRIBUTION / REDESIGNATION REQUESTED (AUTOMATIC)    0
## 3                            REATTRIBUTION FROM SPOUSE    0
## 4                              REATTRIBUTION TO SPOUSE    3
## 5                           REDESIGNATION FROM PRIMARY    0
## 6                 REDESIGNATION FROM SENATE GENERAL       0
```

```
## 7                            REDESIGNATION TO GENERAL    3
## 8                REDESIGNATION TO PRESIDENTIAL GENERAL    2
## 9                                              Refund    9
## 10                                  SEE REATTRIBUTION    0
```

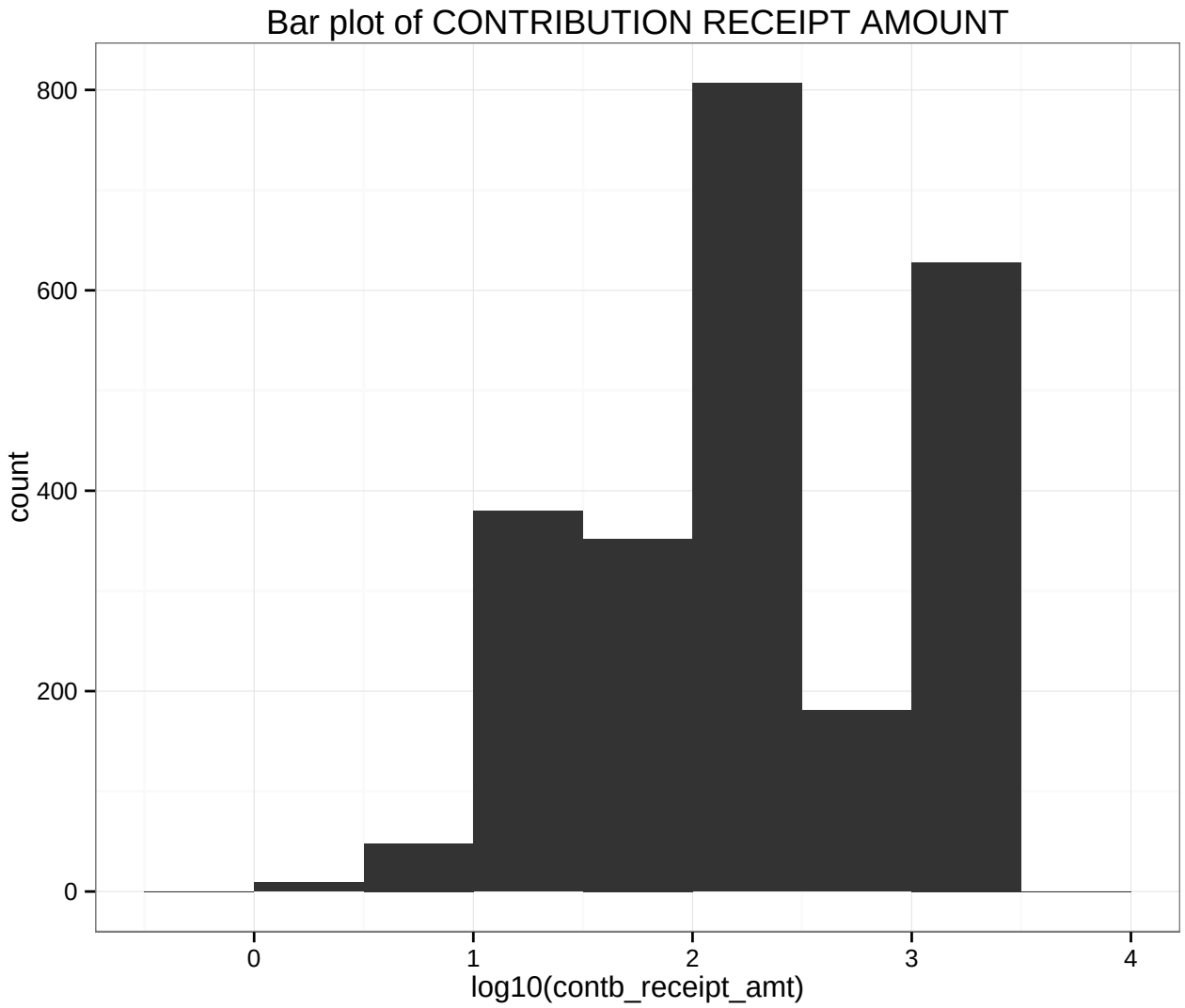
```
##                                                  Var1 Freq
## 1                                                       4
## 2  REATTRIBUTION / REDESIGNATION REQUESTED (AUTOMATIC)    1
## 3                              REATTRIBUTION FROM SPOUSE    0
## 4                                REATTRIBUTION TO SPOUSE    0
## 5                             REDESIGNATION FROM PRIMARY    0
## 6                        REDESIGNATION FROM SENATE GENERAL    0
## 7                              REDESIGNATION TO GENERAL    0
## 8                  REDESIGNATION TO PRESIDENTIAL GENERAL    0
## 9                                              Refund    0
## 10                                  SEE REATTRIBUTION    2
```

There are 22 donations below 0 and 7 above the federal set maximum limit of $2,700. Almost all the donations below 0 are refunds. The ones above $2,700 list reattribution, which means putting the donation potentially in some one else's name, but for the majority the receipt description is blank.

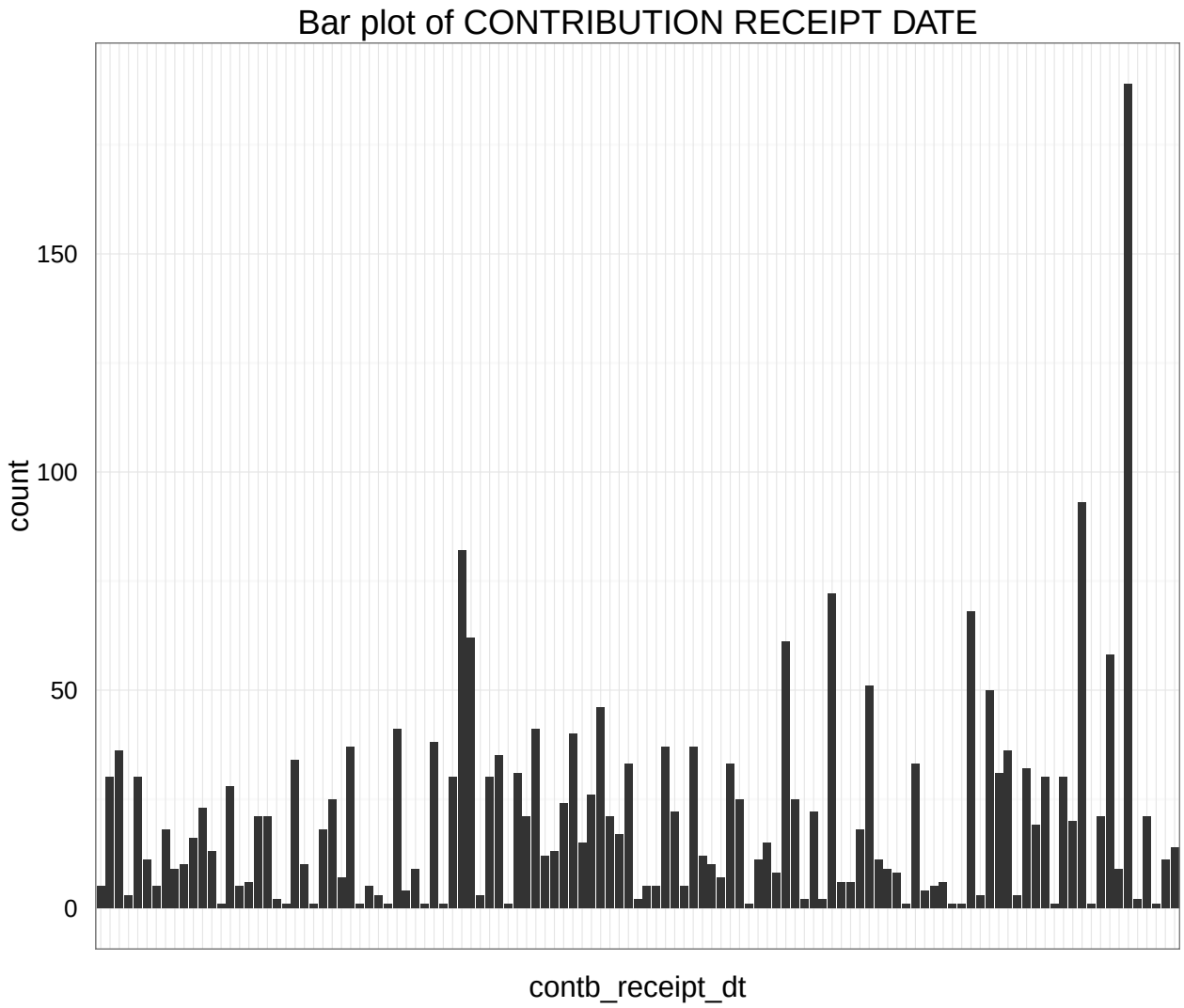
Would think to look at the data in the range of 0 to 2700, which is the allowable range for donations.



What does this look like transformed? Does the second peak go away.



Doesn't look that much better than the original - this is mostly likely due to the ceiling effect producing an odd peak no matter the transformation. I also tried log2 and square root, but they didn't look any better.



This is a busy bar graph, but can clearly see that there are certain dates that people donate on more than others, with one in particularly standing out by creating a new variable called "count_DATE" which counts the frequency of each individual date listed.

```
## [1] 117   2
```

Count the number of unique dates.

```
## [1] 15  2
```

How many dates are listed more than 10 times in the dataset?

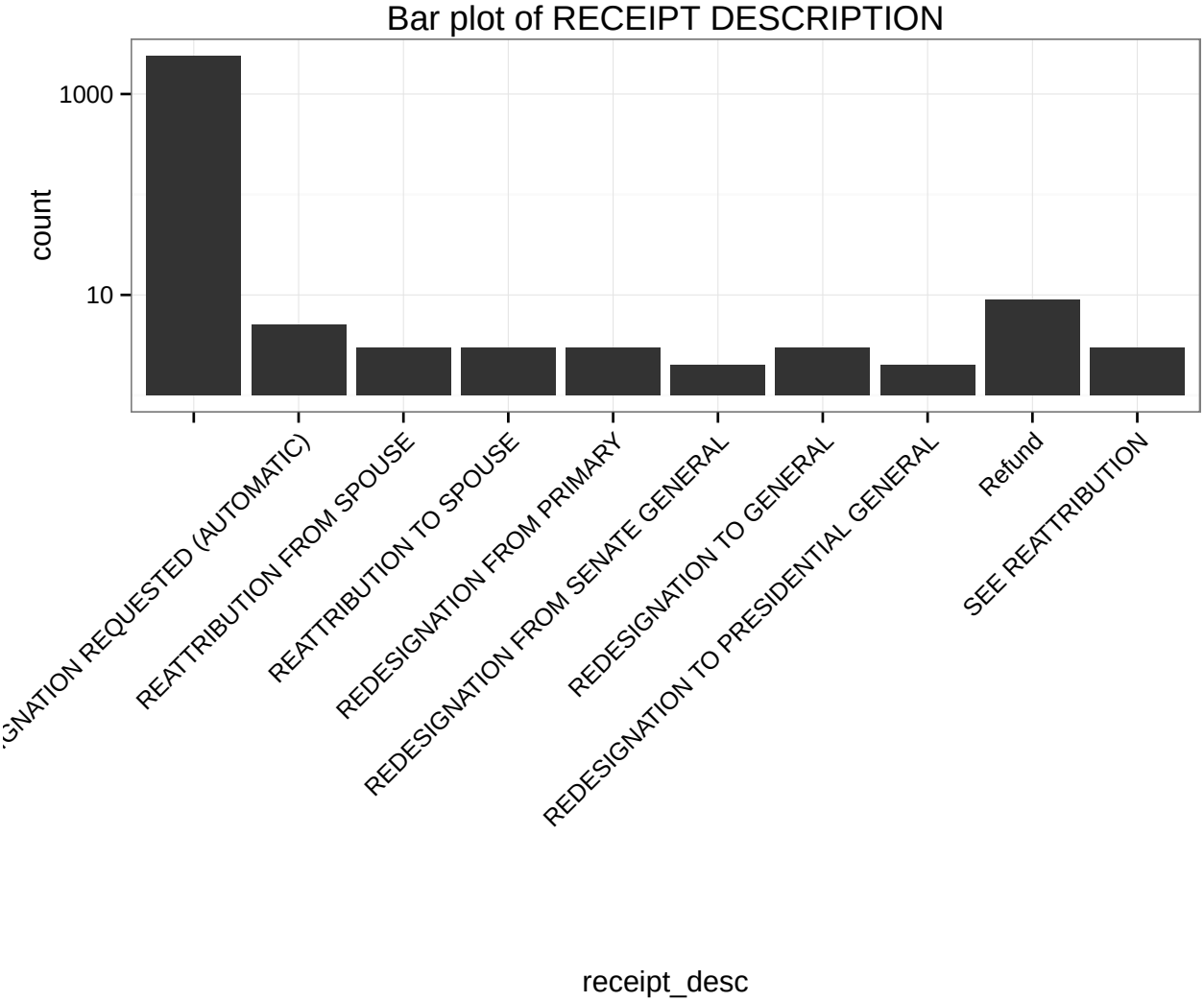
```
## [1] 67  2
```

List the most frequently listed dates in the dataset

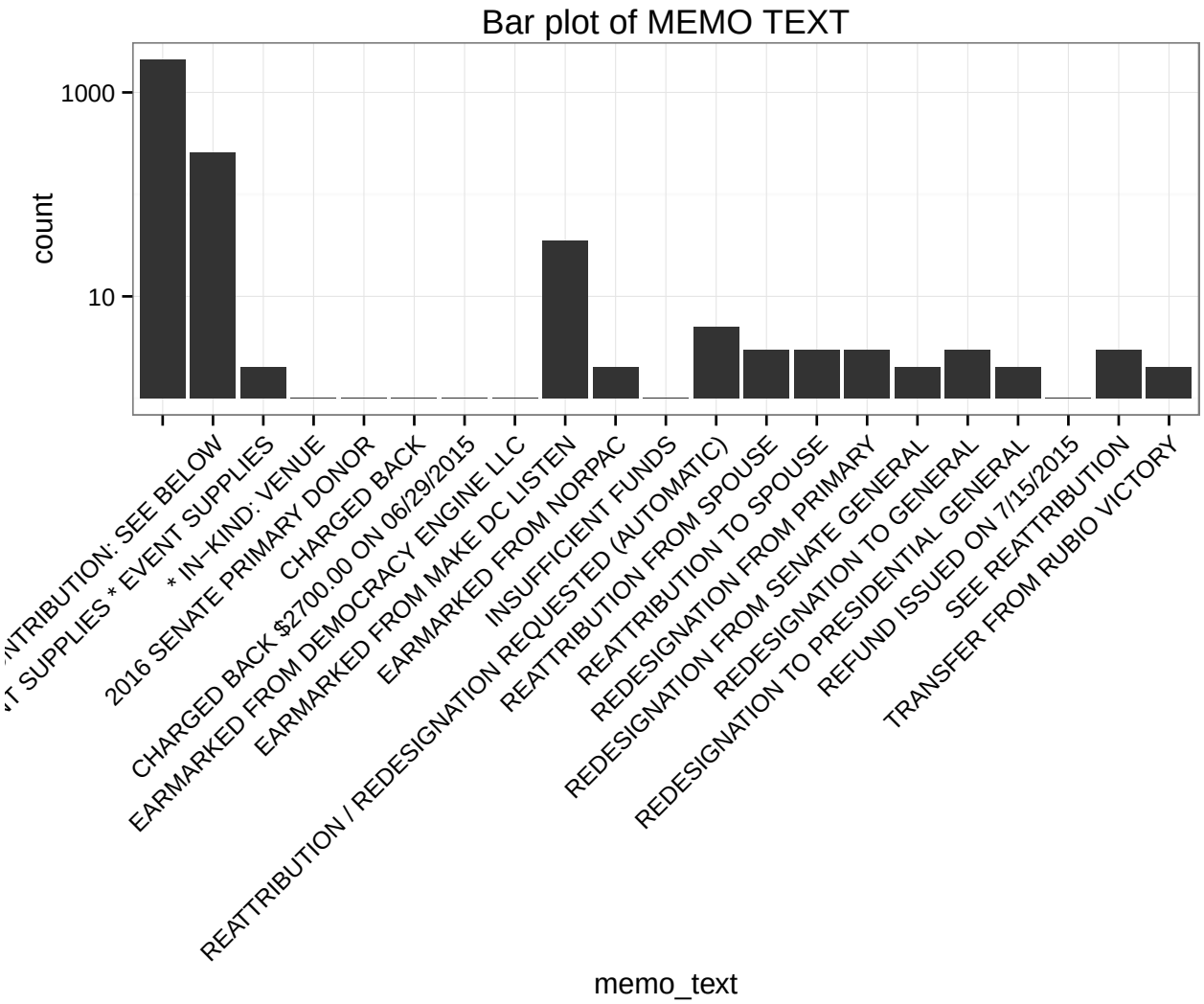
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    5.00   14.00   20.81   30.00  189.00
```

```
##          Var1 Freq
## 112 30-Jun-15  189
```

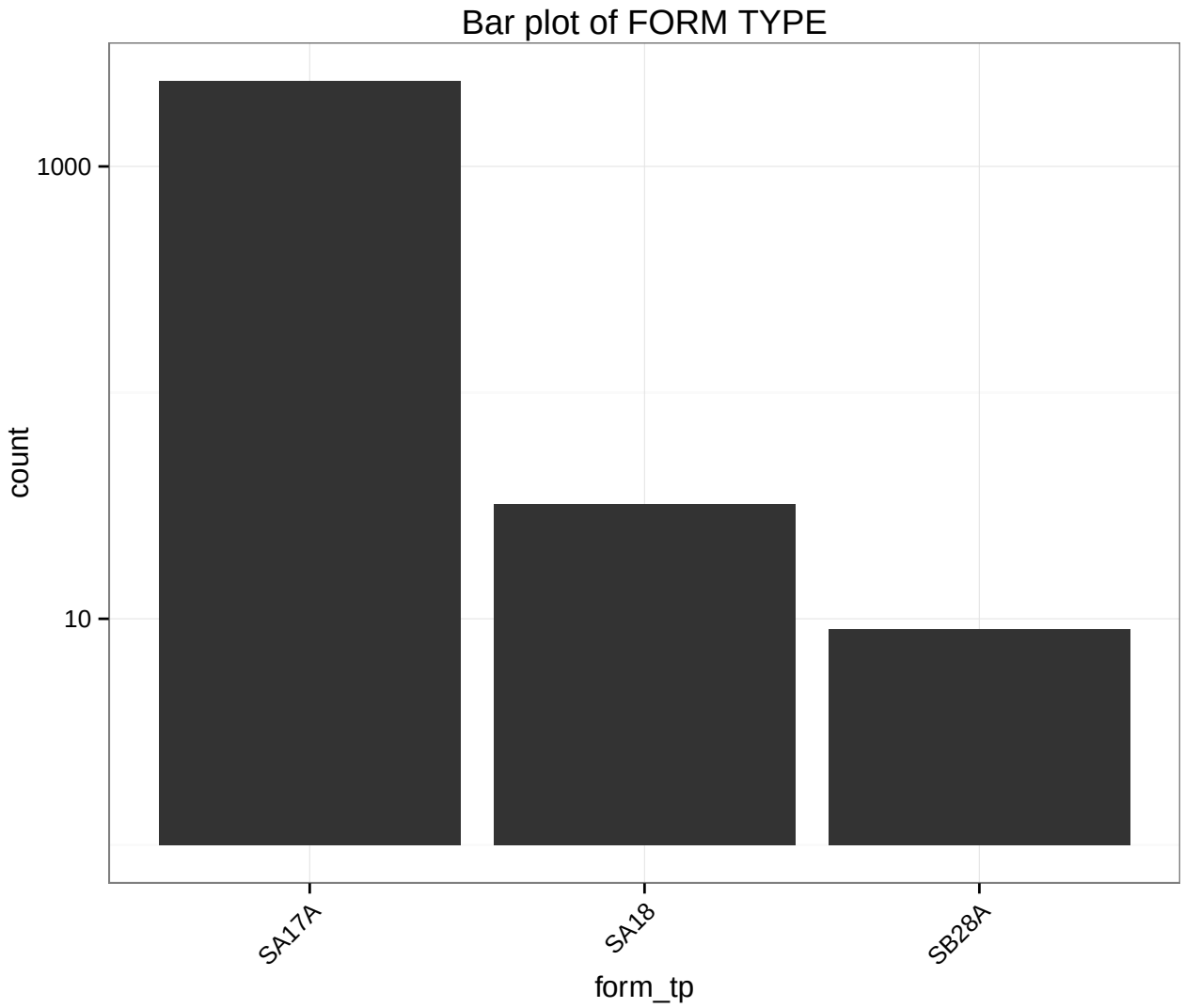
And that date is the 30th of June 2015 - with 189 donations made on that day. A quick google search tells me that on that day the governor of NJ (Chris Christie) declared his candidacy for the US presidential election, however Chris Christie was not one of the candidates named in the data for having donations - so maybe the announcement spurned people on to contribtue to rival campaigns. Will be interesting to see how the candidates campaigns received donations over time.



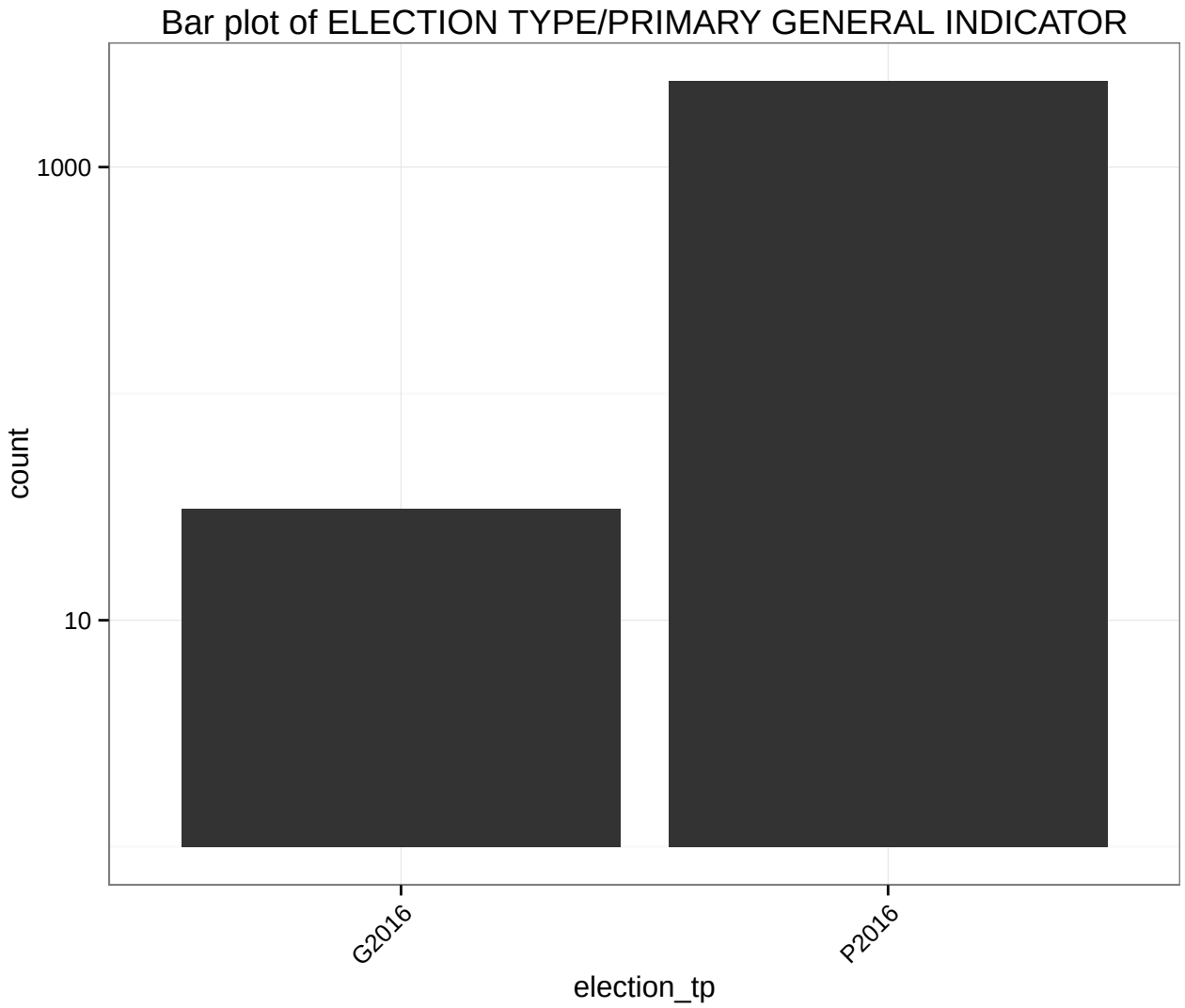
Most of these are relatively small in value.



Like the reciept description - most of these are relatively small in value.



SA17A is for individual contributions, which most of this data is. SA18 is transfers from other authorized committees, and SB28A are refunds to individuals.



There are two types of elections. Most are for P2016, which is Primary 2016, but there are a few that are designated for G2016 which is the General 2016 election, which is slightly confusing that people could contribute to that since the primaries are not yet completed and candidates have not been choosen for that election.

Make new variables

Since there are multiple candidates at the current moment, it would be good to know which candidate belongs to which party, so I'm going to make a new variable called cand_party. I'm going to table the frequency of each candidate name in the dataset.

```
##                         Var1 Freq
## 1                  Bush, Jeb  114
## 2        Carson, Benjamin S.  230
## 3    Clinton, Hillary Rodham 1083
## 4  Cruz, Rafael Edward 'Ted'  259
## 5    CRUZ, RAFAEL EDWARD TED    0
## 6             Fiorina, Carly   19
## 7         Graham, Lindsey O.   25
```

```
## 8             Huckabee, Mike    7
## 9    O'Malley, Martin Joseph   16
## 10         Pataki, George E.   17
## 11                Paul, Rand  170
## 12    Perry, James R. (Rick)    2
## 13              Rubio, Marco  208
## 14          Sanders, Bernard  276
## 15      Santorum, Richard J.    9
```

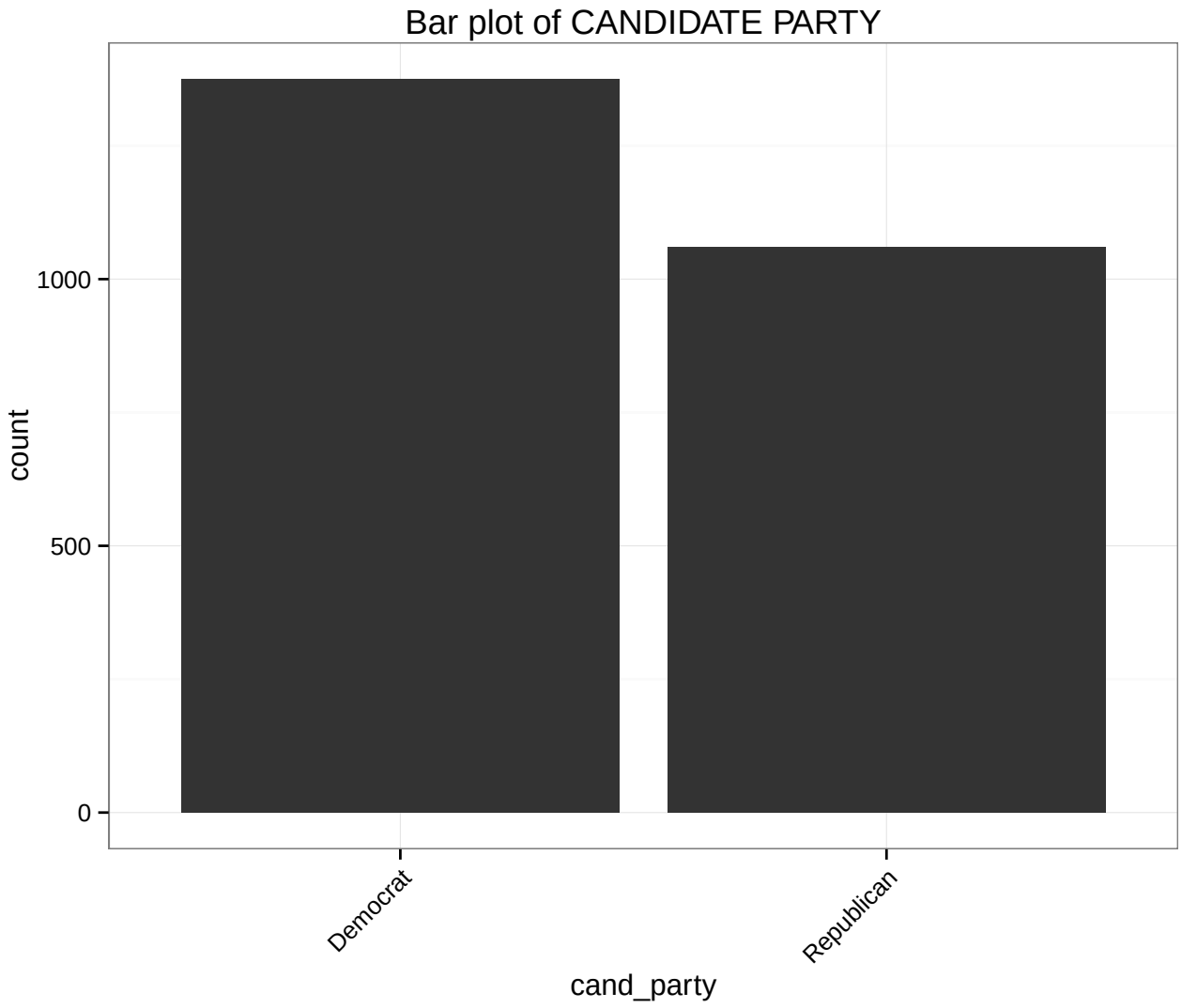
There are fourteen candidates (listed below) and their party affiliation (found via google): Bush, Jeb Republican Carson, Benjamin S. Republican Clinton, Hillary Rodham Democrat Cruz, Rafael Edward 'Ted' Republican Fiorina, Carly Republican Graham, Lindsey O. Republican Huckabee, Mike Republican O'Malley, Martin Joseph Democrat Pataki, George E. Republican Paul, Rand Republican Perry, James R. (Rick) Republican Rubio, Marco Republican Sanders, Bernard Democrat Santorum, Richard J. Republican

There are more Republicans than Democrats.

I'm going to make a new variable that distinguishes which party each candidate belongs to and call this "cand_party". Then I'm going to table this new variable to see how often they occur in the dataset.

```
##
##   Democrat Republican
##       1375       1060
```

Despite there being more Republican candidates, it appears that donations have occured for Democrats. I can plot this new variable.



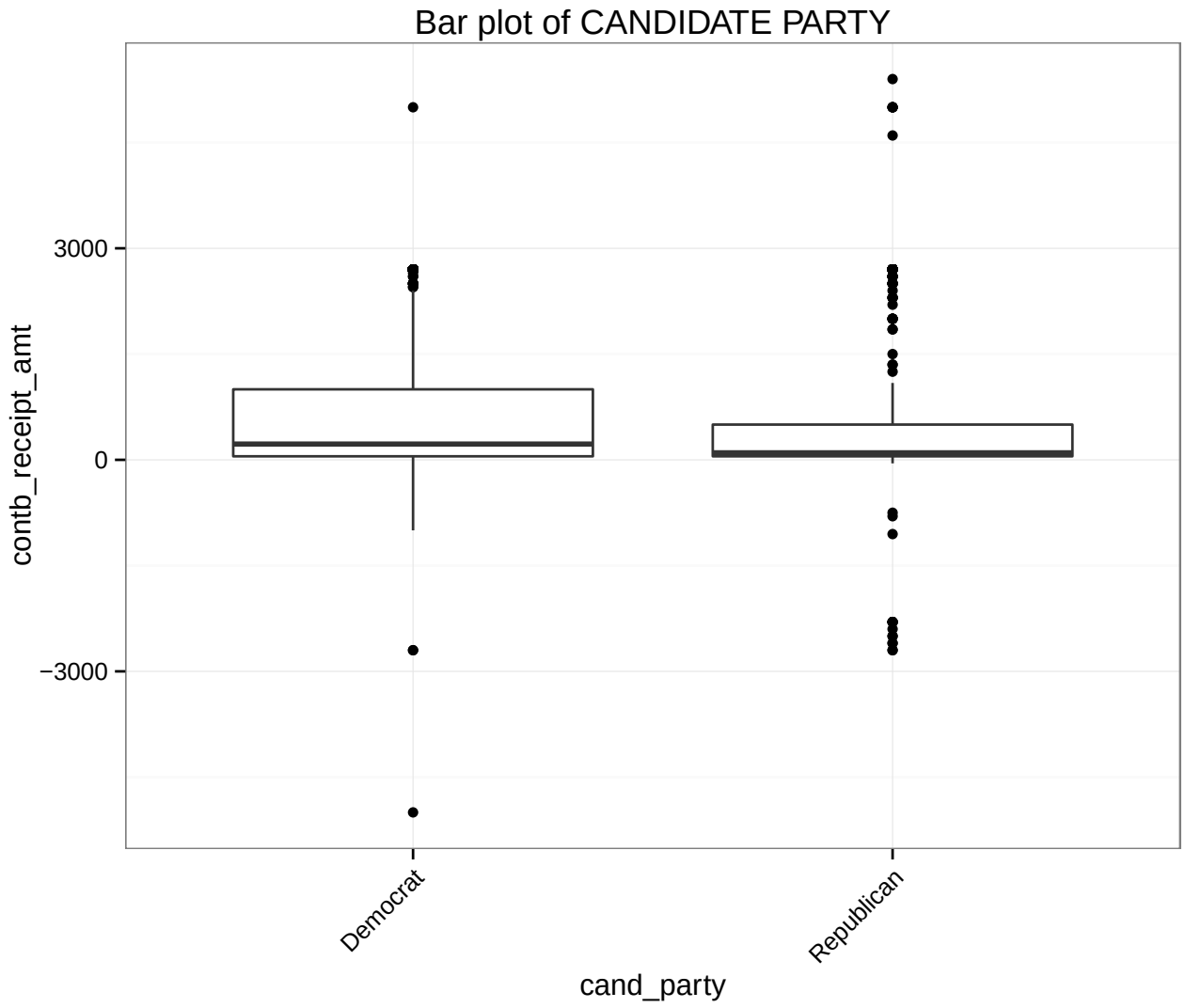
Univariate Summary

I've now looked at all the vairables individually from the dataset. This was a good way to get to know the data. Most people in NJ are making contributions to Democrats, with the most contributions going to Hilary Clinton. Most donations are small, but there is a peak at the ceiling of $2700 (the maximum allowed), however I did notice some amount above that and also negative numbers which seemed to reflect mostly refunds. Occupation data while interesting, was sparse, as not everyone had given this information. Furthermore, it was not broken into broad enough categories for it to be fruitful going forward (over 400 categories), however, a lot of them appear only once, and restricting to the top 10 occupations listed may still be of interest.

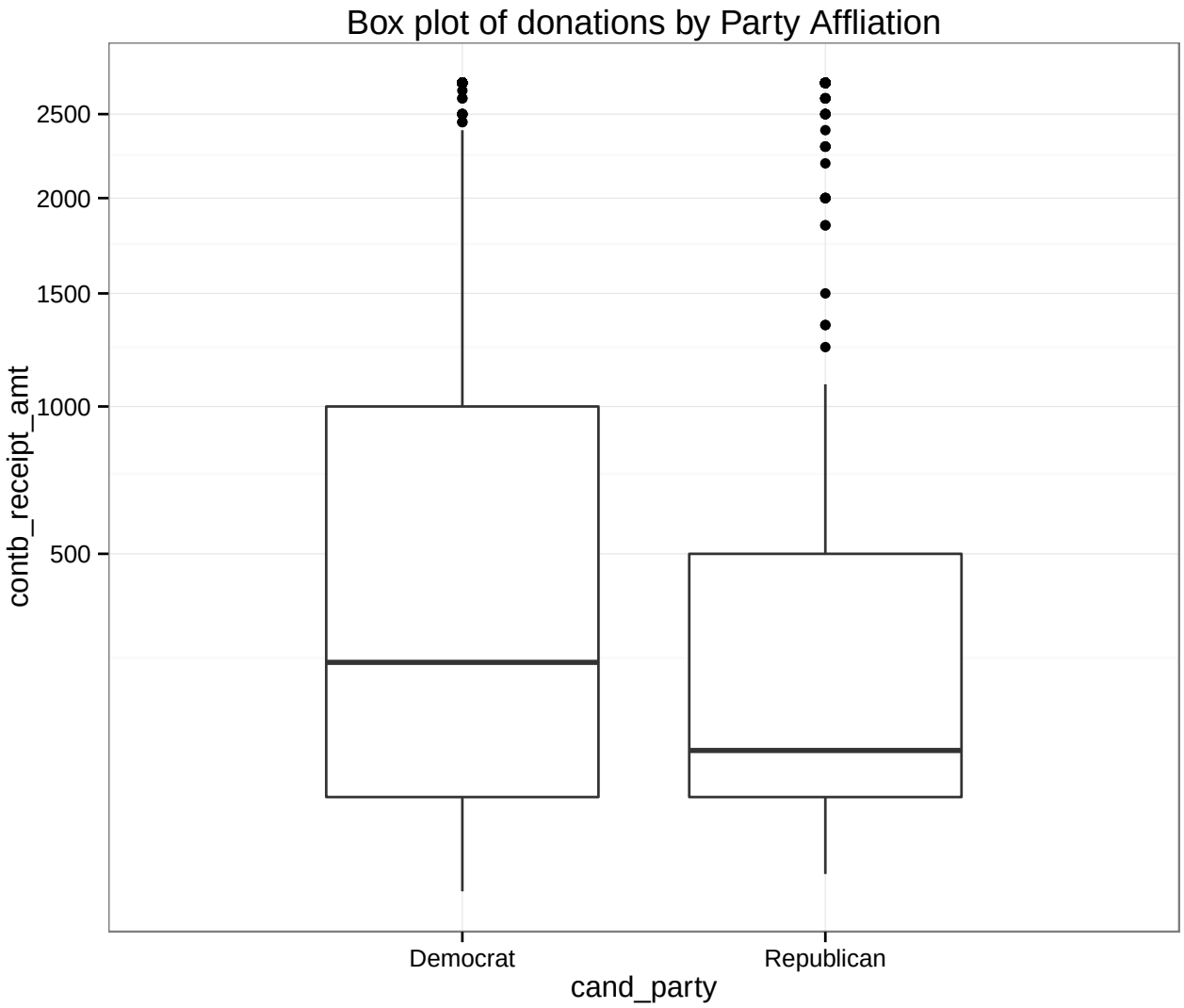
I'm mostly interested in seeing which candidates get the largest amount of money, if these is a difference in amounts by party affiliation, if there is difference in amount and candidaate by occupation or location.

Donation Amount by Candidate and Party.

Examine if the donation amounts differ by party affiliation and candidate. First look at party affiliation



Let's look at this without the donations we think are outliers.

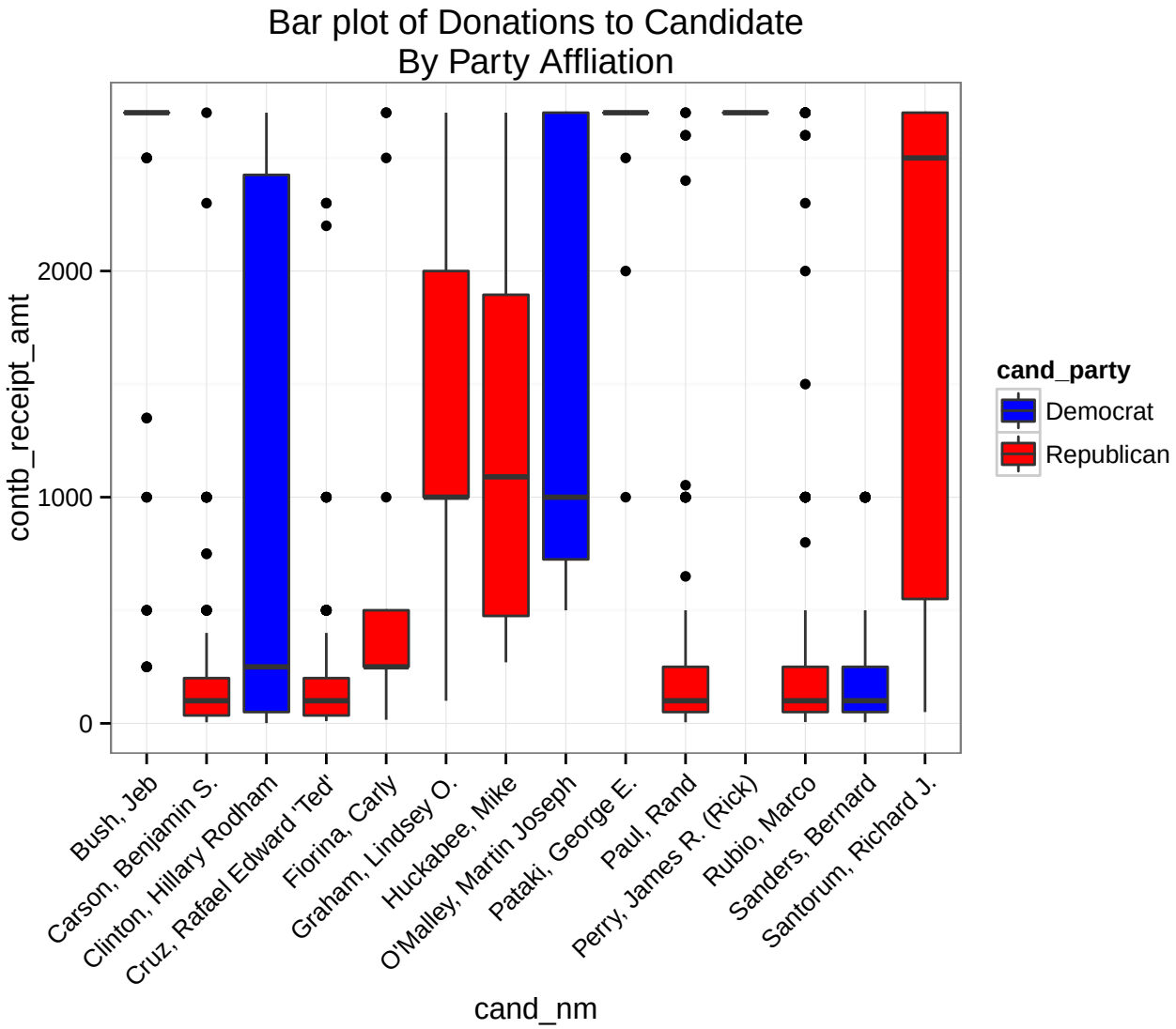


Democrats appear to get more donations than Republicans, but we can quantify this with the data. Let's calculate the mean (and standard deviation) of donation amount by party affliation and the total donation amount rasied by each party.

```
##   cand_party     mean       sd       sum
## 1   Democrat 755.1083 1057.661 1038273.9
## 2 Republican 558.1716 1021.008  591661.9
```

Democrats have a higher mean for donation amount and a higher total donation amount. So they are raising more money than the Republican party in New Jersey.

We can plot the donation amount for each individual candidate by pary.



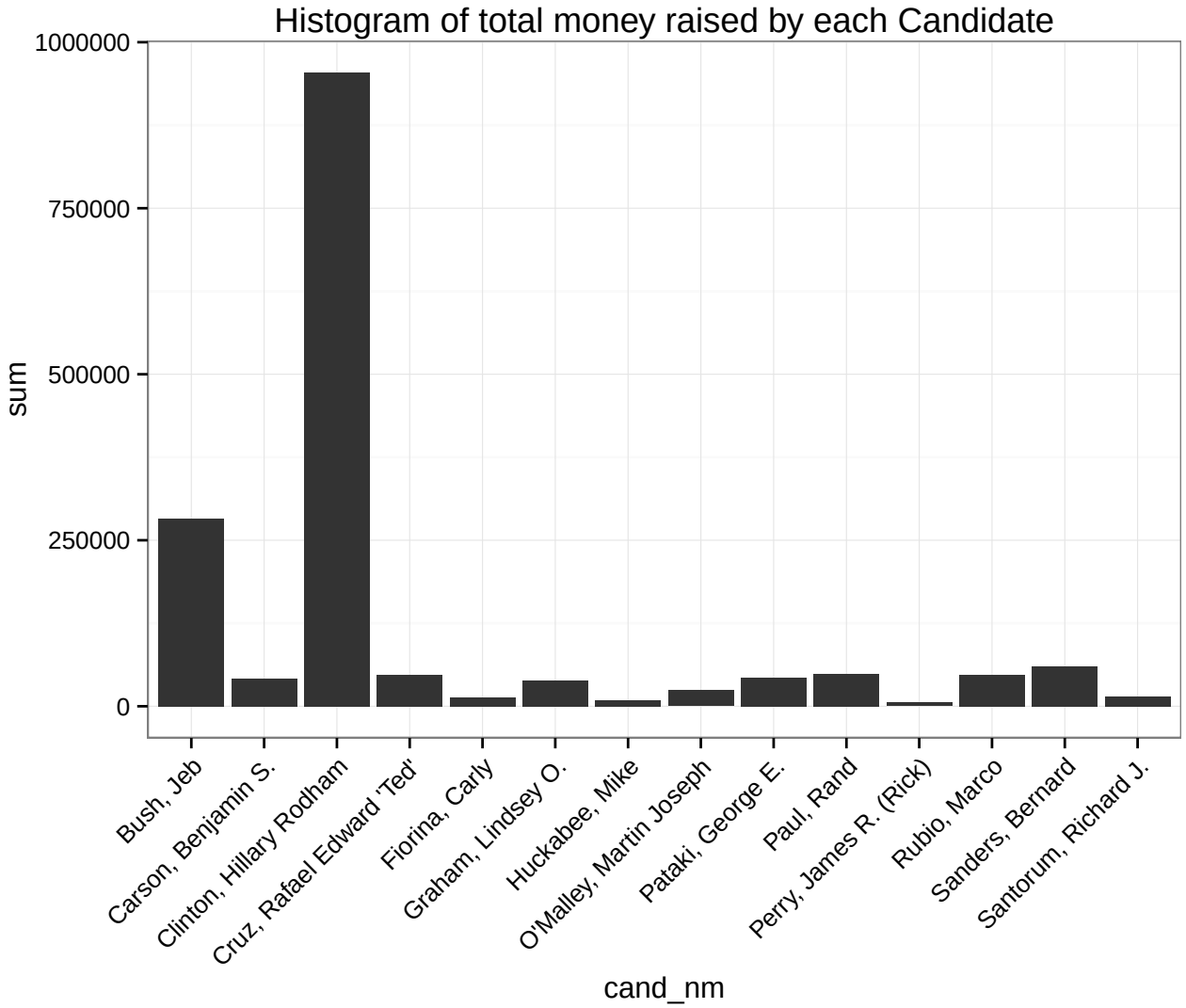
There is quite a spread among the Republican nominations with some receiving small donation amounts on average with some larger outliers (i.e. Carson, Cruz, Rand, Rubio), but there are also candidatest that just recieved large donations (but appears to be few) (i.e. Bush, Pataki, Perry), and there there are the candidates receiving a spread of donations (i.e. Graham, Huckabee, Santorum). This pattern aligns more with the Democrat candidates (i.e. Clinton, O'Malley) who also appear to have a large spread of donation amounts.

Total amount raised by each candidate

I will create a new dataset ("NJ_money_by_candidate") which will contain the mean, standard deviation and total sum of donation amounts by each individual candidate. This information will be plotted to show the total donation amount raised by each individual candidate.

```
##                     cand_nm      mean        sd       sum
## 1                 Bush, Jeb 2482.4561  635.0382 283000.00
## 2       Carson, Benjamin S.  182.8478  477.6293  42055.00
## 3    Clinton, Hillary Rodham  881.3203 1136.8932 954469.90
## 4  Cruz, Rafael Edward 'Ted'  182.1313  559.3187  47172.00
## 5            Fiorina, Carly  692.9474  893.0102  13166.00
```

```
## 6         Graham, Lindsey O. 1548.0000 1038.2758  38700.00
## 7             Huckabee, Mike 1257.1429 1034.3068   8800.00
## 8    O'Malley, Martin Joseph 1500.0000  979.1152  24000.00
## 9          Pataki, George E. 2547.0588  434.6229  43300.00
## 10               Paul, Rand  285.6818  488.7603  48565.90
## 11    Perry, James R. (Rick) 2700.0000    0.0000   5400.00
## 12             Rubio, Marco  226.2163  856.2588  47053.00
## 13          Sanders, Bernard  216.6811  255.7233  59803.97
## 14      Santorum, Richard J. 1605.5556 2301.6902  14450.00
```



Clinton (Dem) has raised the most money by far, with Bush (Rep) in second.

Donation Amount by Candidate and Party Over Time.

I have time data for donations, so the next question might be to look at patterns of donations over time, but first I need to ensure that the date information is being correctly recognized as a date.

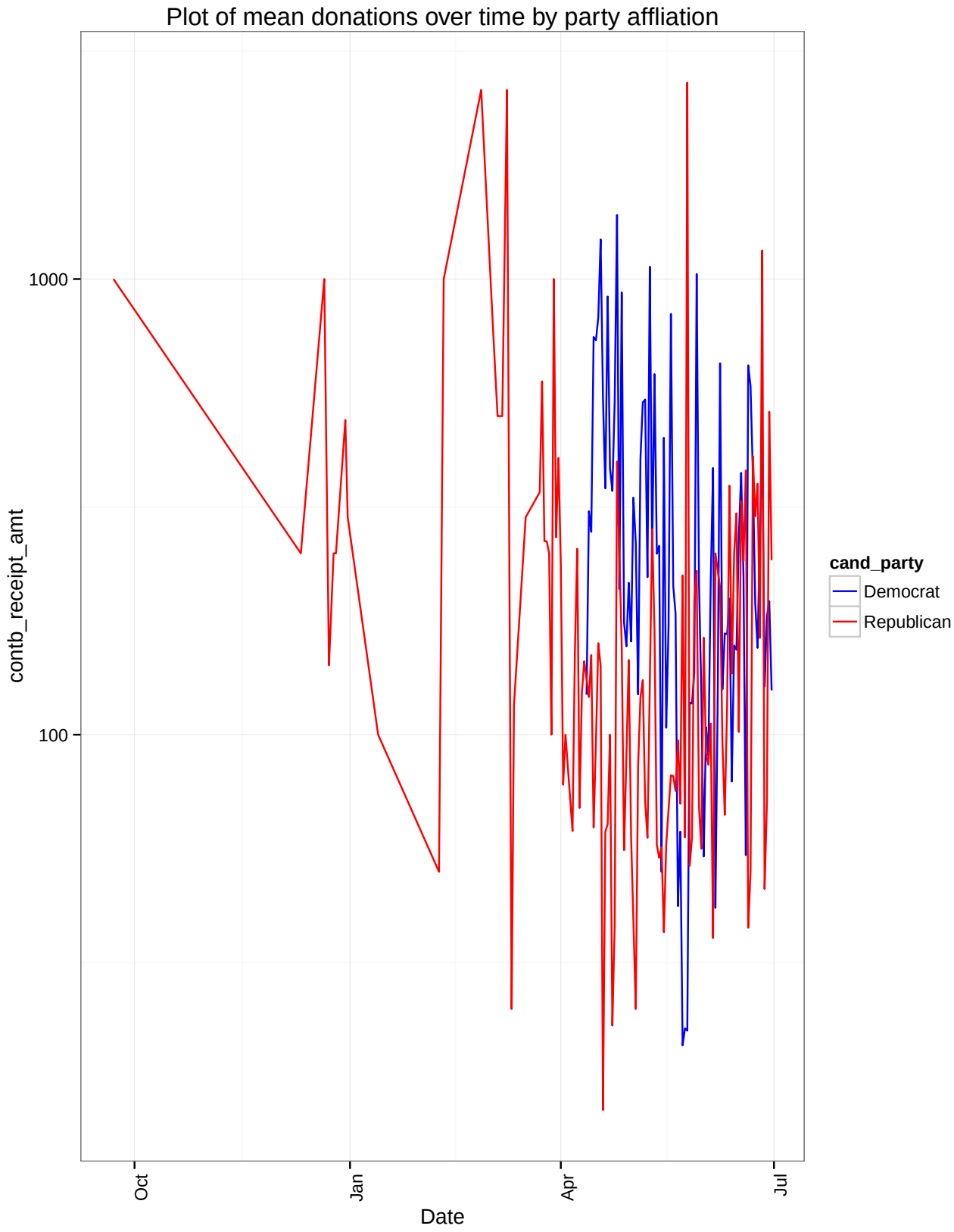
```
## 'data.frame':    2435 obs. of  19 variables:
##  $ cmte_id          : Factor w/ 14 levels "C00458844","C00500587",..: 6 6 6 6 6 6 6 6 6 10 ...
```

```
##  $ cand_id          : Factor w/ 14 levels "P00003392","P20002721",..: 1 1 1 1 1 1 1 1 1 10 ...
##  $ cand_nm          : Factor w/ 15 levels "Bush, Jeb","Carson, Benjamin S.",..: 3 3 3 3 3 3 3 3 3 10
##  $ contbr_nm        : Factor w/ 1353 levels "ABDELAZIZ, AL",..: 1208 230 842 1344 895 1223 755 74 894
##  $ contbr_city      : Factor w/ 346 levels "ALLENDALE","ALLENHURST",..: 289 57 43 247 302 234 100 194
##  $ contbr_st        : Factor w/ 1 level "NJ": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contbr_zip       : int  70792116 70121939 70071406 88544546 85601724 76525505 70245022 70423025 77
##  $ contbr_employer  : Factor w/ 692 levels "","A&E STORES",..: 562 415 415 450 562 562 415 122 592 56
##  $ contbr_occupation: Factor w/ 445 levels "","ACADEMIC",..: 23 281 206 118 214 314 206 241 339 314
##  $ contb_receipt_amt: num  250 100 2700 2700 2700 2700 2700 2700 50 2700 ...
##  $ contb_receipt_dt : Factor w/ 117 levels "01-Apr-15","01-Jun-15",..: 40 98 109 98 107 110 28 95 28
##  $ receipt_desc     : Factor w/ 10 levels "","REATTRIBUTION / REDESIGNATION REQUESTED (AUTOMATIC)",.
##  $ memo_cd          : Factor w/ 2 levels "","X": 1 1 1 1 1 1 1 1 1 1 ...
##  $ memo_text        : Factor w/ 21 levels "","* EARMARKED CONTRIBUTION: SEE BELOW",..: 1 1 1 1 1 1 1
##  $ form_tp          : Factor w/ 3 levels "SA17A","SA18",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ file_num         : int  1015585 1015585 1015585 1015585 1015585 1015585 1015585 1015585 1015585 10
##  $ tran_id          : Factor w/ 2435 levels "A032126DA61AA40D699B",..: 491 1215 436 1167 900 1220 513
##  $ election_tp      : Factor w/ 2 levels "G2016","P2016": 2 2 2 2 2 2 1 2 2 2 ...
##  $ cand_party       : chr  "Democrat" "Democrat" "Democrat" "Democrat" ...
```

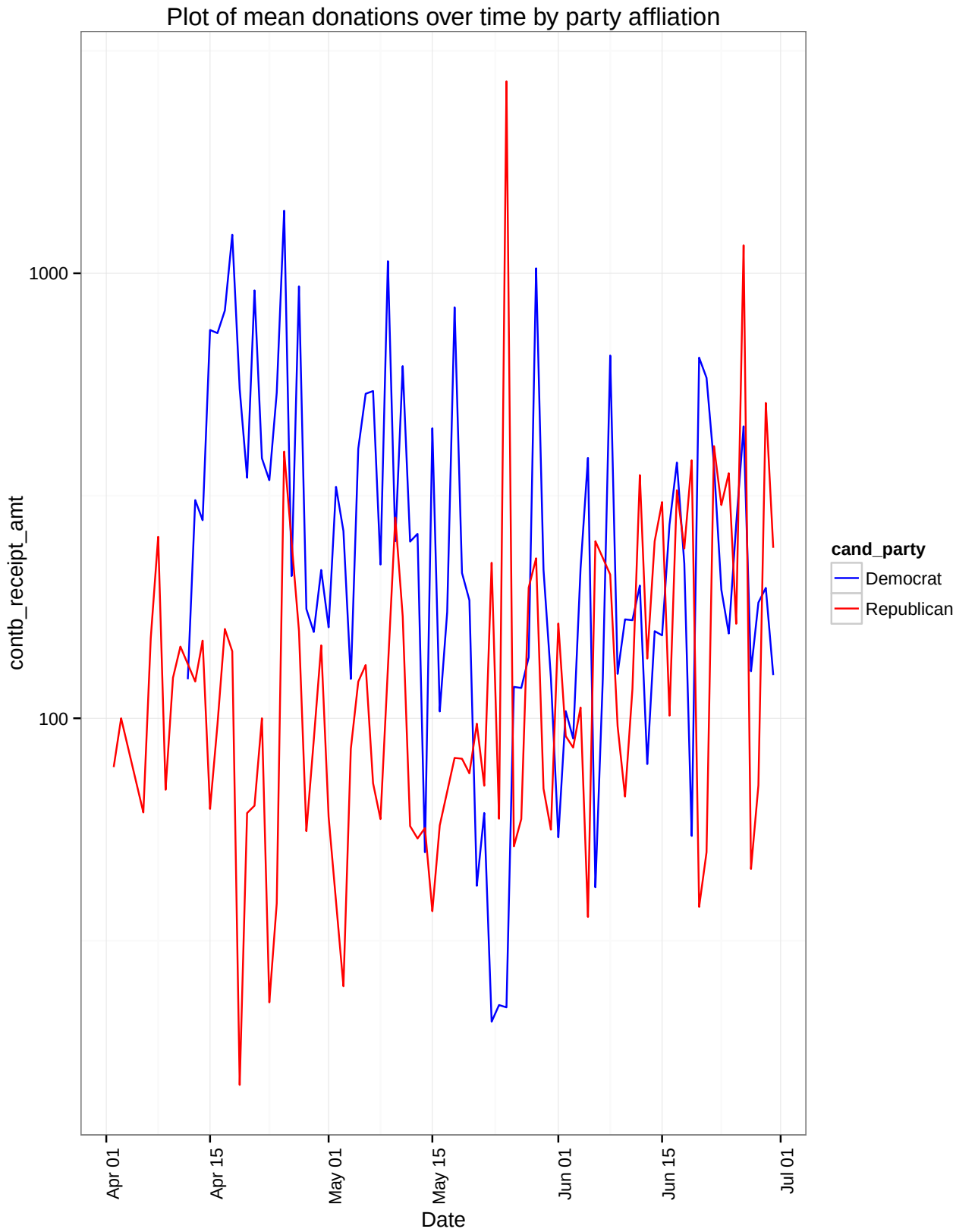
The 'contb_receipt_dt' variable is not as a date but instead a factor. This needs to be changed. I will create a new variable ("Date") that contains the date in the correct format.

```
## 'data.frame':    2435 obs. of  20 variables:
##  $ cmte_id          : Factor w/ 14 levels "C00458844","C00500587",..: 6 6 6 6 6 6 6 6 6 10 ...
##  $ cand_id          : Factor w/ 14 levels "P00003392","P20002721",..: 1 1 1 1 1 1 1 1 1 10 ...
##  $ cand_nm          : Factor w/ 15 levels "Bush, Jeb","Carson, Benjamin S.",..: 3 3 3 3 3 3 3 3 3 10
##  $ contbr_nm        : Factor w/ 1353 levels "ABDELAZIZ, AL",..: 1208 230 842 1344 895 1223 755 74 894
##  $ contbr_city      : Factor w/ 346 levels "ALLENDALE","ALLENHURST",..: 289 57 43 247 302 234 100 194
##  $ contbr_st        : Factor w/ 1 level "NJ": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contbr_zip       : int  70792116 70121939 70071406 88544546 85601724 76525505 70245022 70423025 77
##  $ contbr_employer  : Factor w/ 692 levels "","A&E STORES",..: 562 415 415 450 562 562 415 122 592 56
##  $ contbr_occupation: Factor w/ 445 levels "","ACADEMIC",..: 23 281 206 118 214 314 206 241 339 314
##  $ contb_receipt_amt: num  250 100 2700 2700 2700 2700 2700 2700 50 2700 ...
##  $ contb_receipt_dt : Factor w/ 117 levels "01-Apr-15","01-Jun-15",..: 40 98 109 98 107 110 28 95 28
##  $ receipt_desc     : Factor w/ 10 levels "","REATTRIBUTION / REDESIGNATION REQUESTED (AUTOMATIC)",.
##  $ memo_cd          : Factor w/ 2 levels "","X": 1 1 1 1 1 1 1 1 1 1 ...
##  $ memo_text        : Factor w/ 21 levels "","* EARMARKED CONTRIBUTION: SEE BELOW",..: 1 1 1 1 1 1 1
##  $ form_tp          : Factor w/ 3 levels "SA17A","SA18",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ file_num         : int  1015585 1015585 1015585 1015585 1015585 1015585 1015585 1015585 1015585 10
##  $ tran_id          : Factor w/ 2435 levels "A032126DA61AA40D699B",..: 491 1215 436 1167 900 1220 513
##  $ election_tp      : Factor w/ 2 levels "G2016","P2016": 2 2 2 2 2 2 1 2 2 2 ...
##  $ cand_party       : chr  "Democrat" "Democrat" "Democrat" "Democrat" ...
##  $ Date             : Date, format: "2015-04-12" "2015-04-27" ...
```

Now that it is in Date format, we can plot the data. This plot is the donation amount by time, with each line representing a differnet party (Republican or Democrat).



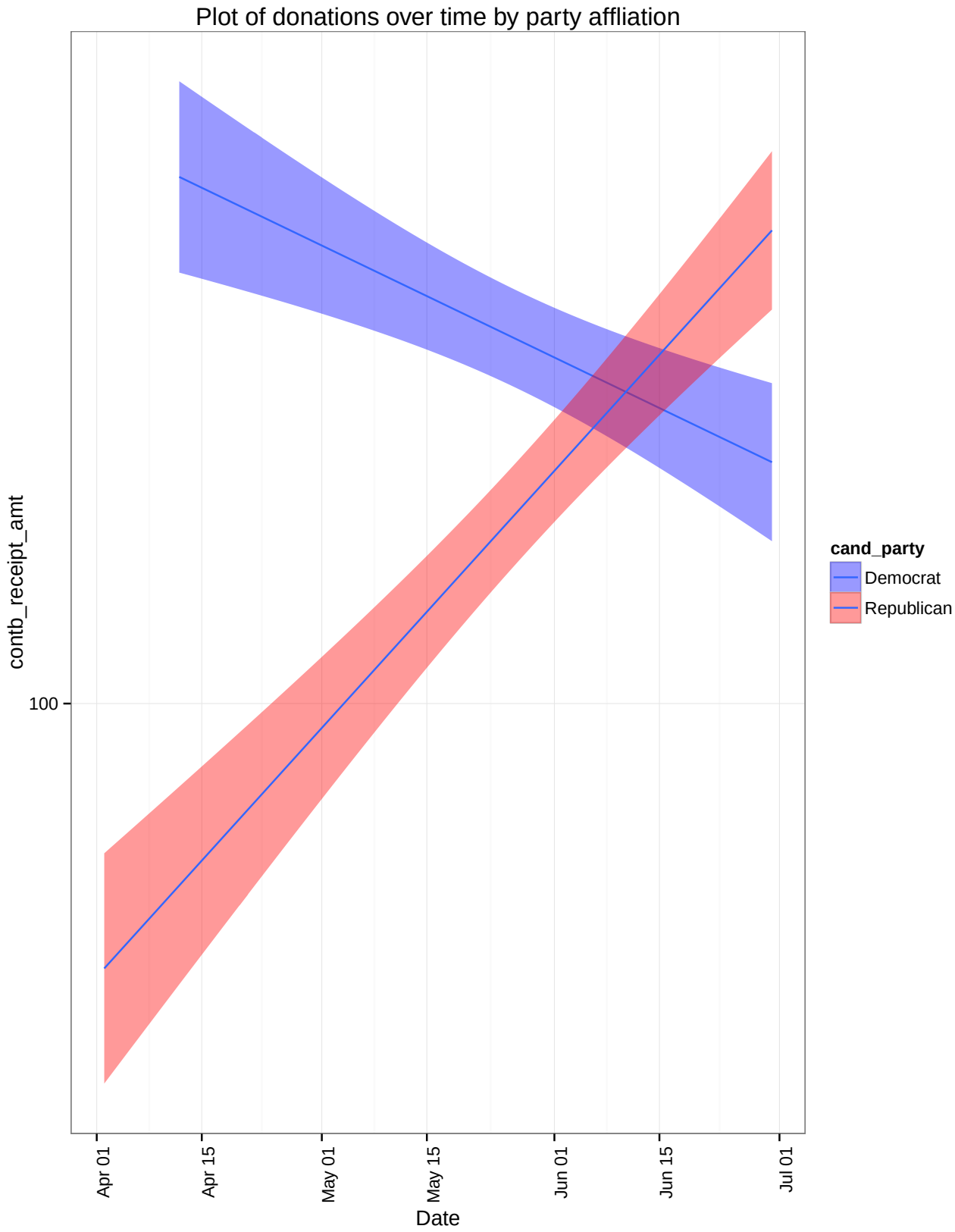
The first people to get donations was a Republican, and it looks like Democrats did not start to receive donations till about April 2015. Let's subset this plot to just look at time points after April.



Democrats around mid April to early June were receiving larger donations than Republicans however around mid June this trend seems to fade with them receiving equal donation amounts. There is one spike early for Republicans in late May, but that looks more like an outlier than a true increase. Republicans over this

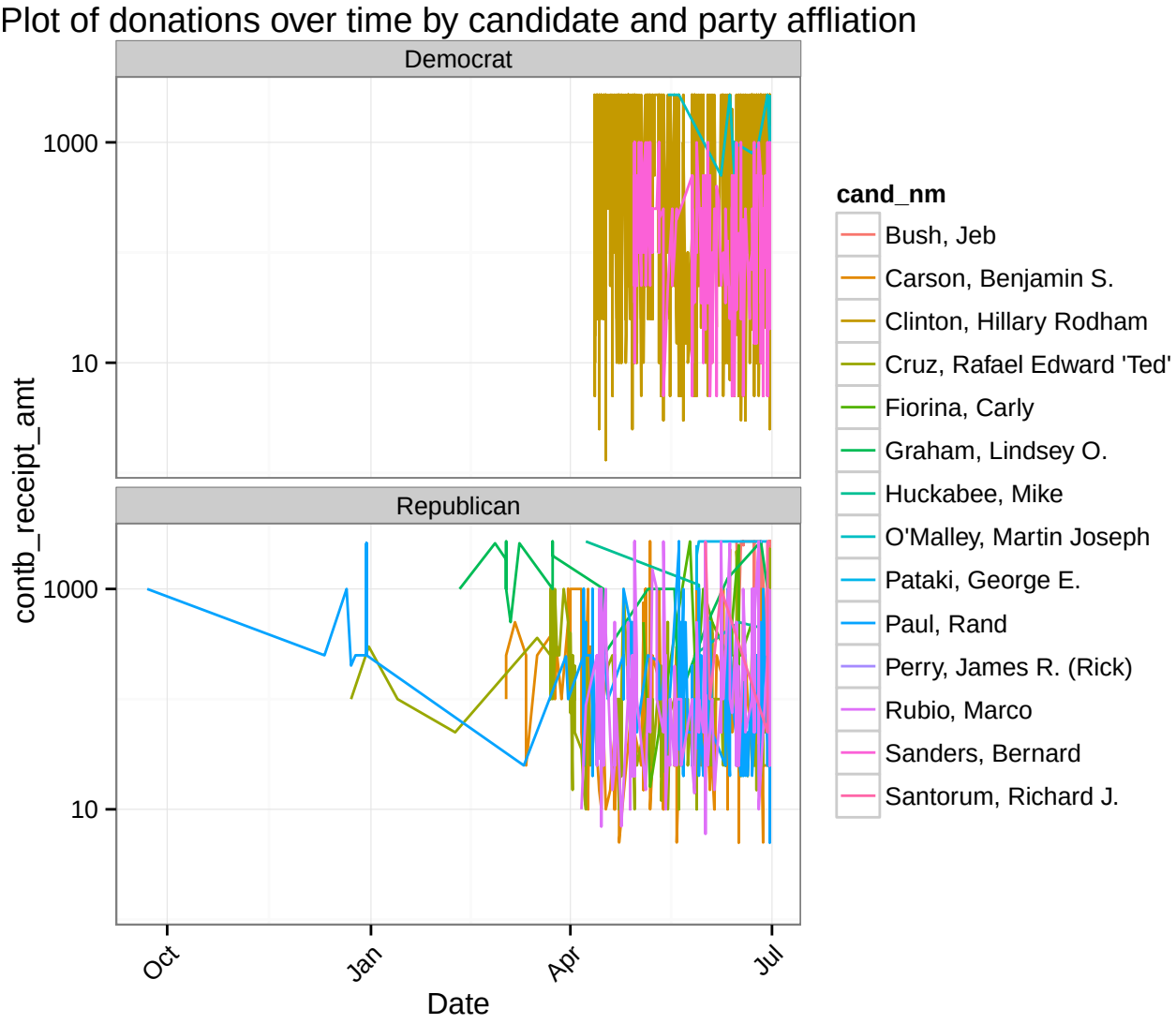
period of time appear to be growing in donation amounts, where the Democrats (while fluxuating) seems to remain mostly the same.

If geom_smooth is used instead, the plot might reflect this general pattern.



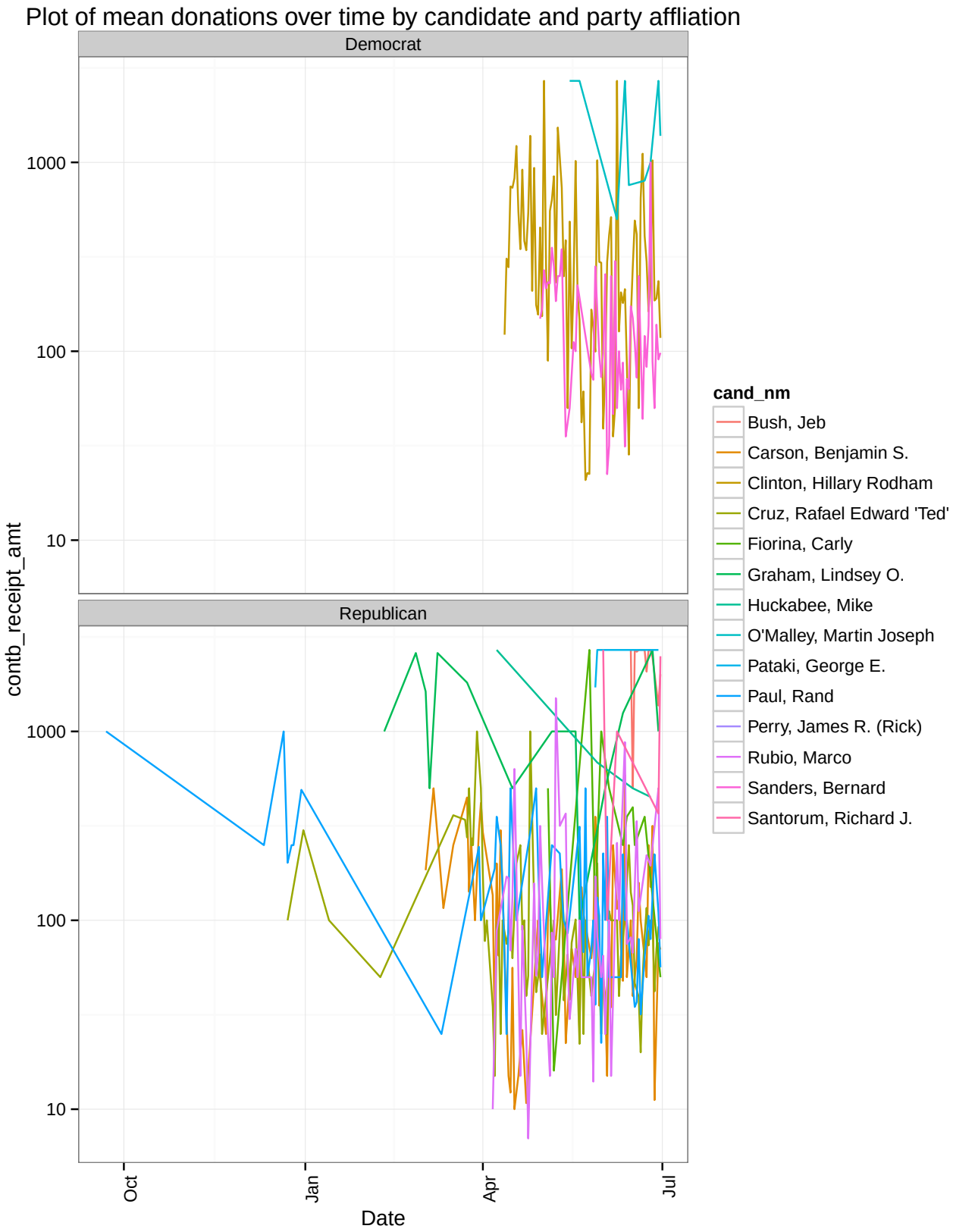
This plot loses a lot of information about contributions, but does show that the trend for increase in donations to Republicans, and the steady (maybe a slight decrease) in donation to Democrats over time.

When can also look at this information for each individual candidate separated by party affiliation.

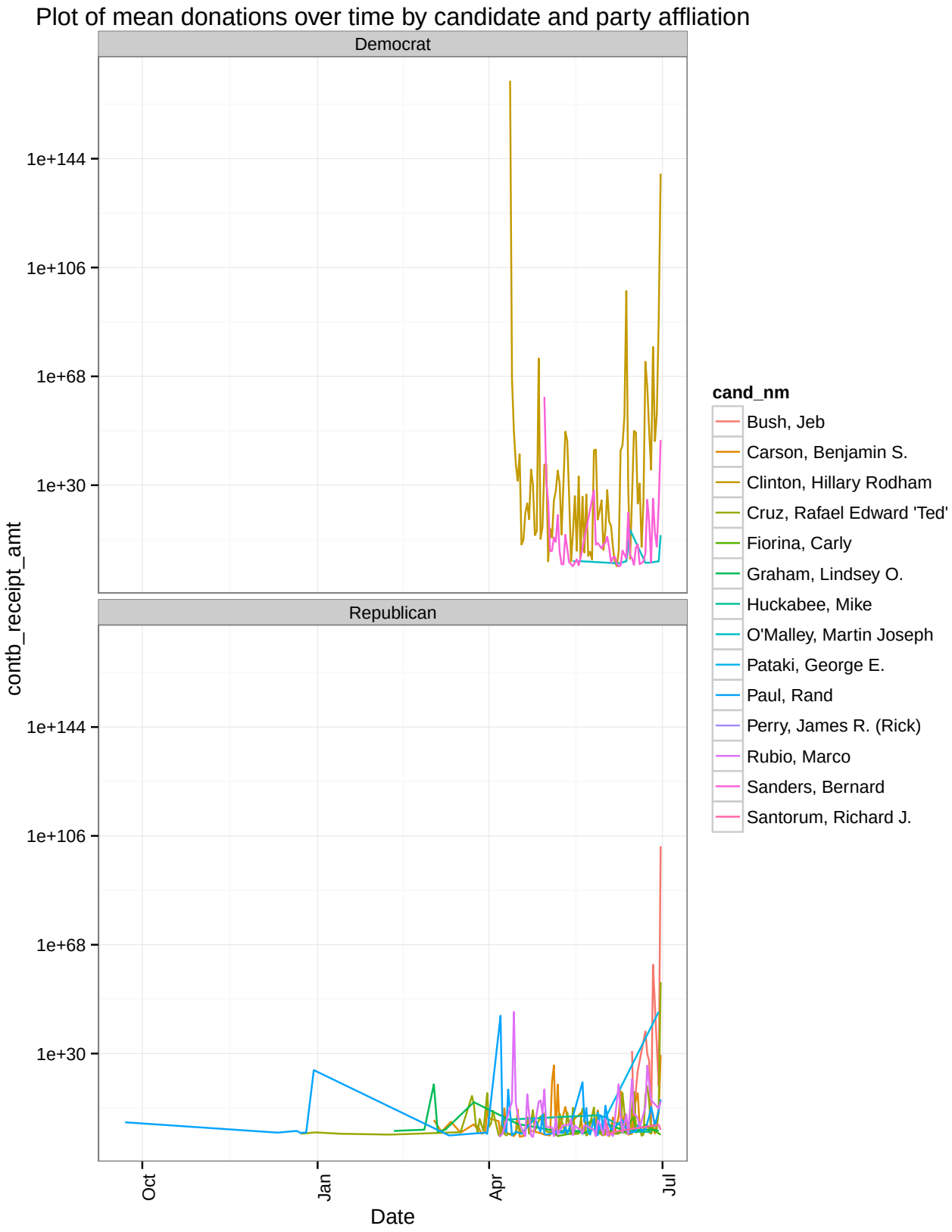


Interesting. The plot is too messy to make too much of it, but it is clear that people came into the race at different time points. Republicans were the first to come forward to declare they were running for President, with Rand declaring very early and then Cruz coming in second. Democrats don't declare for almost a full 6 months later, with Clinton being the first to recieve donations. Around April time, we see that a lot of Republican candidates start receiving contributions, suggesting they all declared their intentions around this time.

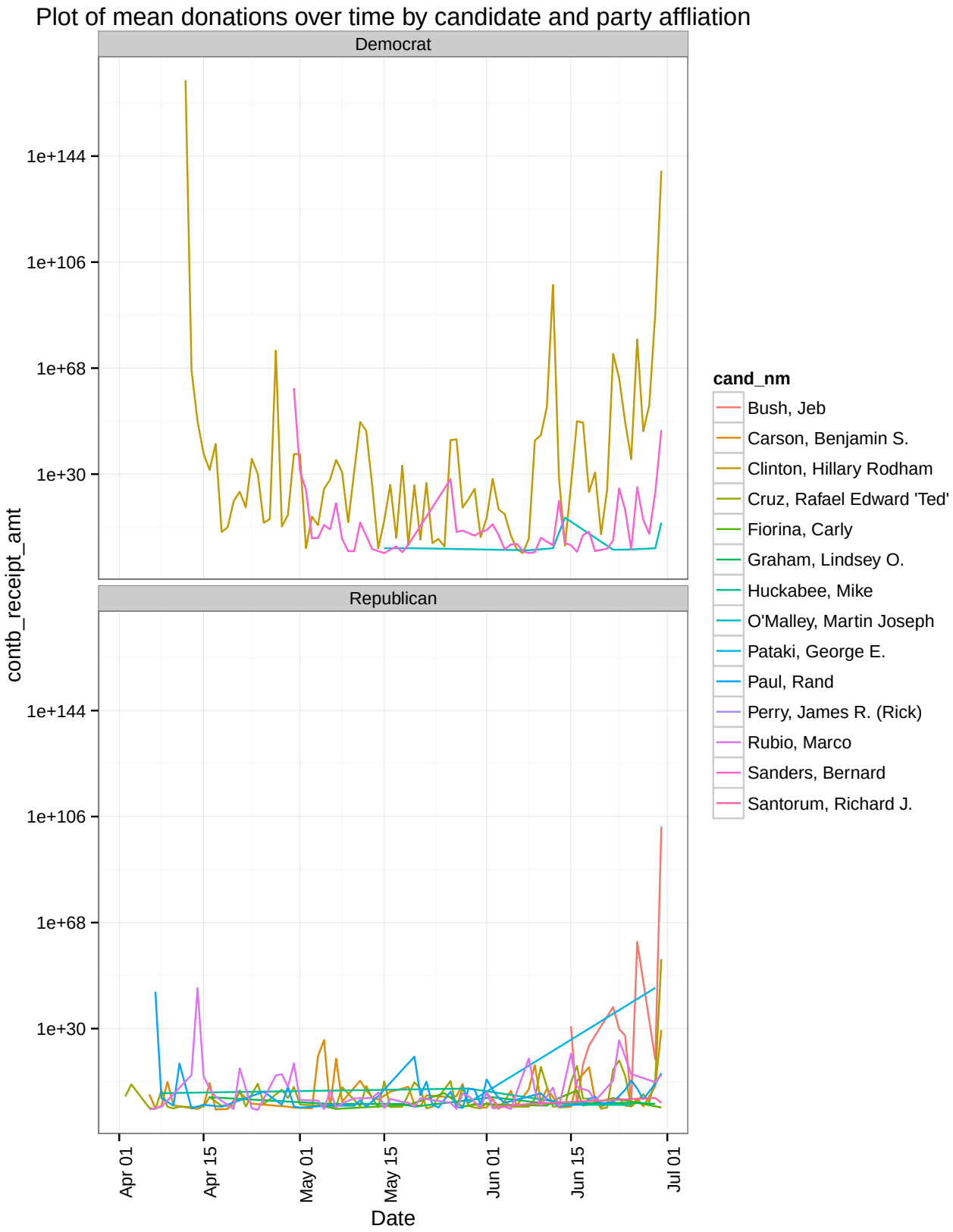
However, the plot is too messy to make any sense of the data, but maybe if we plot the mean value of the donations rather than just the donations the plot might be a bit smoother.



This is a slightly better plot of the data, if I plot the sum instead of the mean does this look a bit better as a plot



This is good, but I think I'm going to subset it to get a better look at the donations starting from April, where a lot of the data is ploting and keep it as the sum instead of the mean.

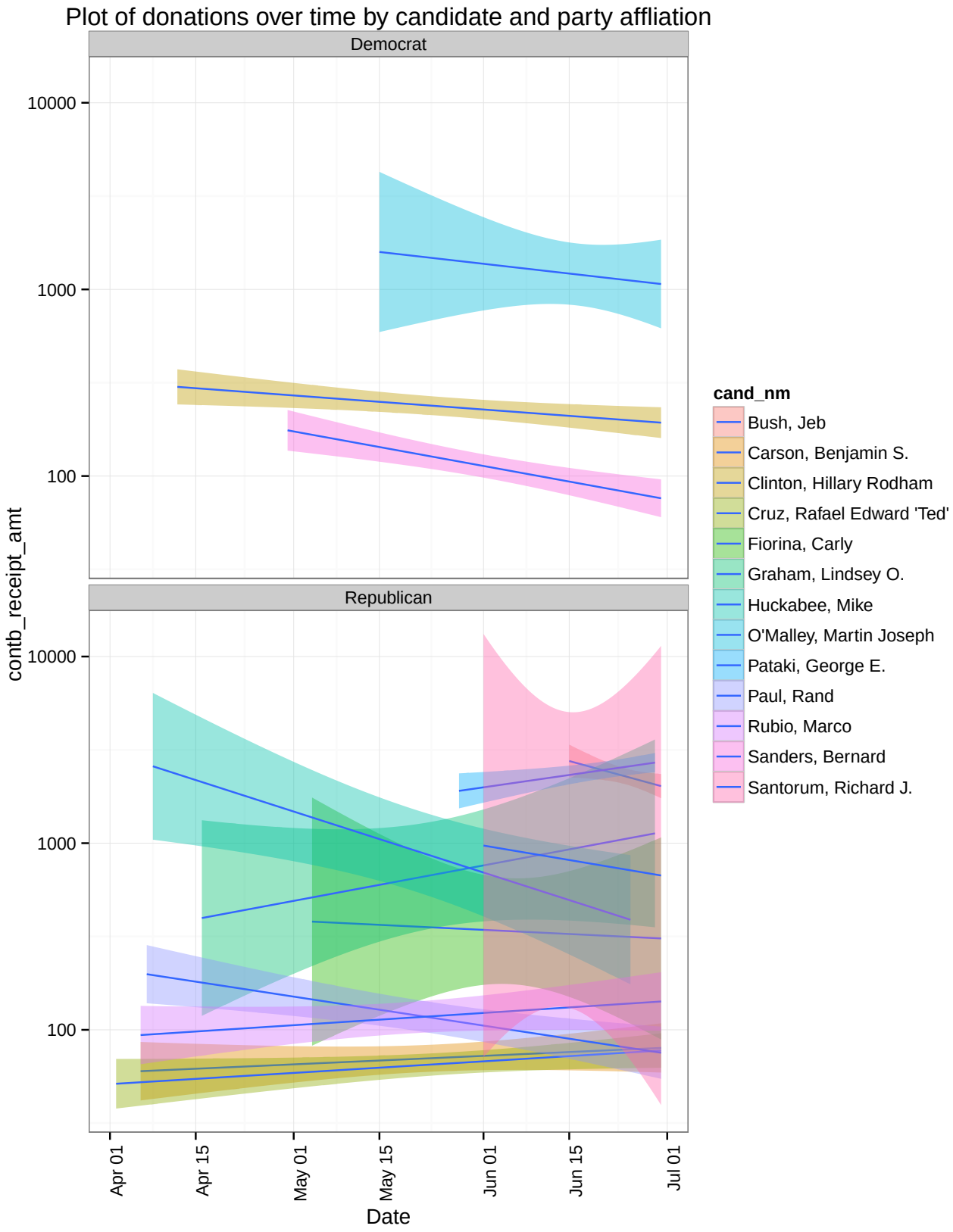


There are late people in the Republican side that receive high donation amount on average (i.e Bush and Pataki), but they have little data as they only started receieving donations around June (when they must have declared their intentions to run for President), so the amount of time they could have received donations

is less than that of Clinton or Rand.

For the Democrats, Clinton is consistently recieveing larger donations than Sanders, with O'Malley varying more than the other two. The Republican data is a little more inconsistent than the Democrat data, and this could in part be because they have a larger number of candidates for donations to be spread out among. But there is one indivdiuals who seems to be on a down trend (receiving lower amounts of donations over time) (Huckabee), while another appears to be increasing the donation amount with time (Graham).

Again using geom_smooth to give a general idea about the overall pattern of the data might be a good idea.



This isn't as clear as it was for the party affiliation, and this is mostly due to the large number of candidates for the Republican party. The Democrats are mostly stable lines for Clinton and Sanders, while O'Malley has a little decrease and wider variance (standard error) than the other two candidates.

For the Republicans, a lot of individuals have wide SEs. There are trends for people going up and down as hypothesed above. Interestingly, Bush in this graph has a flat line with wide SEs, not appearing to be increasing like the previous plot. This may be due to low number of data points for Bush that are highly variable.

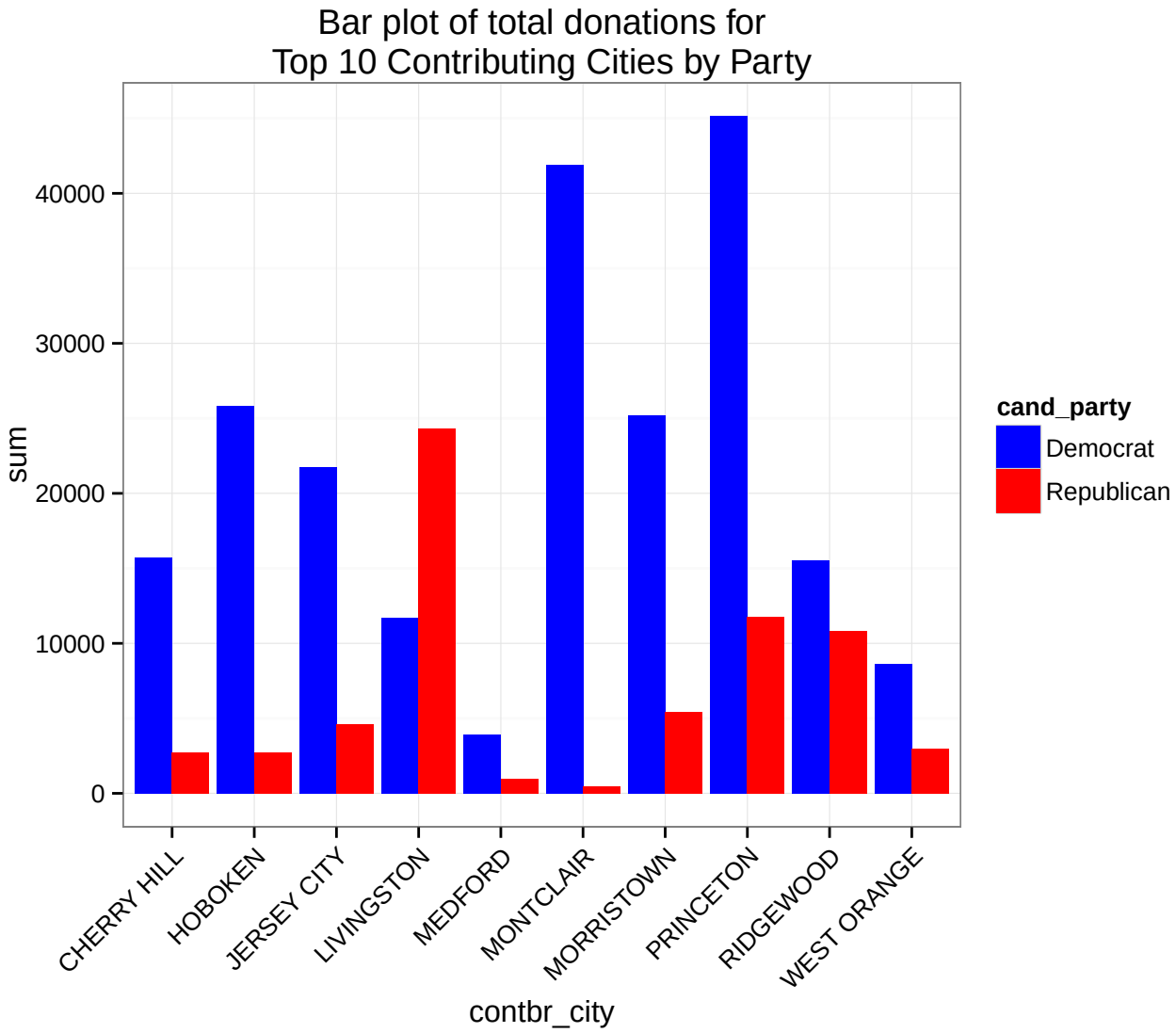
Donation Amounts by City

There are a lot of cities in the dataset, so I'm only going to look at the cities that have the most donations (the top 10 cities for frequency in the dataset). Earlier I made a new dataset set with the counts of the number of occurance for each city (called count_CITY). I'm going to use this information to create a new dataframe ("NJ_top_cities") which only contains the top 10 cities that occur most frequently in the data (have the most contributing indivduals). Then using this new dataframe I will calculate the total amount of donations for each city for each candidate by party affiliation.

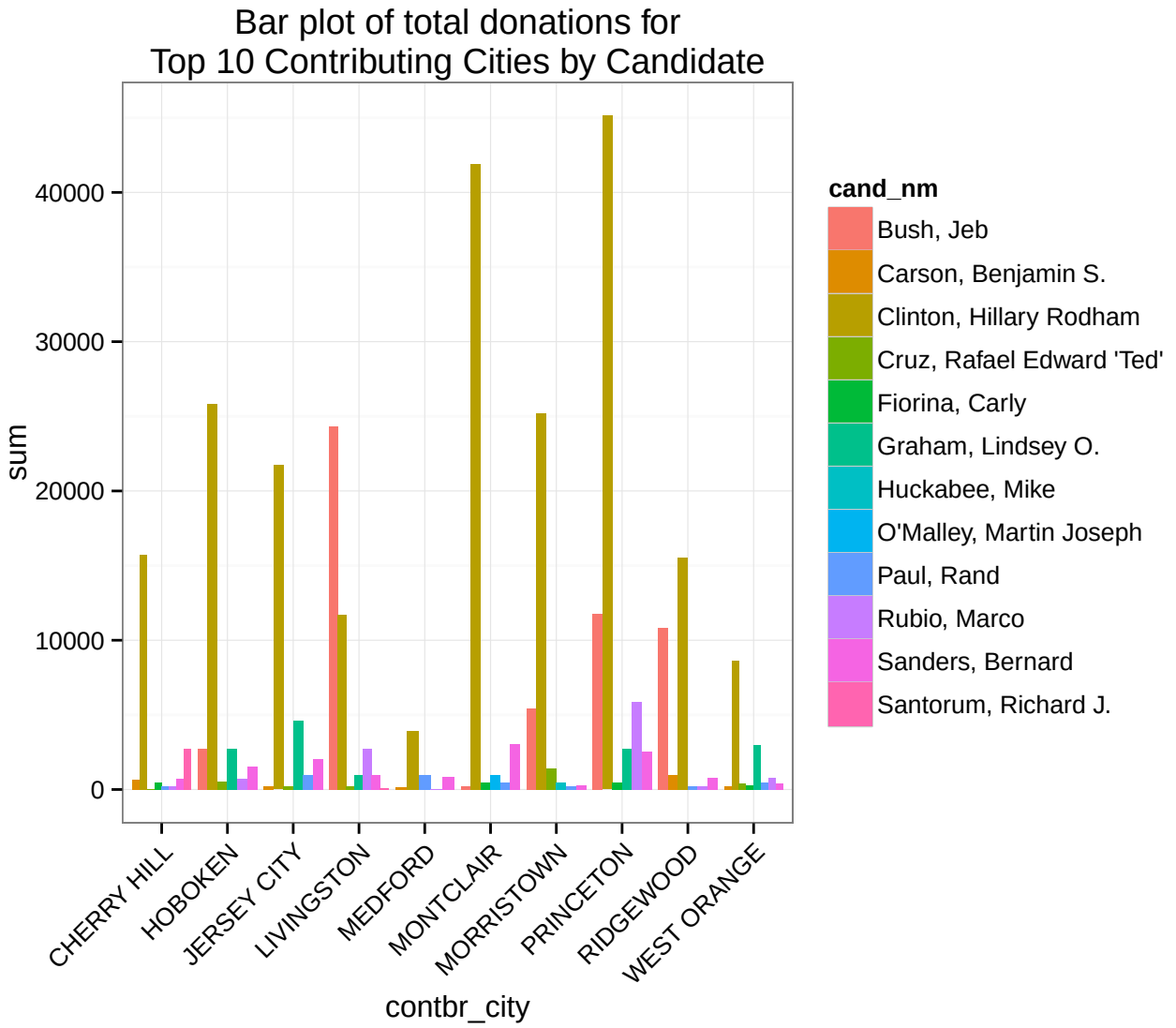
```
##    contbr_city                   cand_nm cand_party      sum
## 1  CHERRY HILL       Carson, Benjamin S. Republican   650.00
## 2  CHERRY HILL   Clinton, Hillary Rodham   Democrat 15717.80
## 3  CHERRY HILL Cruz, Rafael Edward 'Ted' Republican    20.00
## 4  CHERRY HILL            Fiorina, Carly Republican   500.00
## 5  CHERRY HILL                Paul, Rand Republican   250.00
## 6  CHERRY HILL              Rubio, Marco Republican   250.00
## 7  CHERRY HILL          Sanders, Bernard   Democrat   746.88
## 8  CHERRY HILL      Santorum, Richard J. Republican  2700.00
## 9      HOBOKEN                 Bush, Jeb Republican  2700.00
## 10     HOBOKEN   Clinton, Hillary Rodham   Democrat 25844.64
## 11     HOBOKEN Cruz, Rafael Edward 'Ted' Republican   520.00
## 12     HOBOKEN        Graham, Lindsey O. Republican  2700.00
## 13     HOBOKEN              Rubio, Marco Republican   700.00
## 14     HOBOKEN          Sanders, Bernard   Democrat  1510.00
## 15 JERSEY CITY       Carson, Benjamin S. Republican   250.00
## 16 JERSEY CITY   Clinton, Hillary Rodham   Democrat 21721.00
## 17 JERSEY CITY Cruz, Rafael Edward 'Ted' Republican   250.00
## 18 JERSEY CITY        Graham, Lindsey O. Republican  4600.00
## 19 JERSEY CITY                Paul, Rand Republican  1000.00
## 20 JERSEY CITY          Sanders, Bernard   Democrat  2060.00
## 21  LIVINGSTON                 Bush, Jeb Republican 24300.00
## 22  LIVINGSTON   Clinton, Hillary Rodham   Democrat 11736.75
## 23  LIVINGSTON Cruz, Rafael Edward 'Ted' Republican   222.00
## 24  LIVINGSTON        Graham, Lindsey O. Republican  1000.00
## 25  LIVINGSTON              Rubio, Marco Republican  2700.00
## 26  LIVINGSTON          Sanders, Bernard   Democrat  1000.00
## 27  LIVINGSTON      Santorum, Richard J. Republican   100.00
## 28     MEDFORD       Carson, Benjamin S. Republican   150.00
## 29     MEDFORD   Clinton, Hillary Rodham   Democrat  3951.80
## 30     MEDFORD                Paul, Rand Republican  1001.60
## 31     MEDFORD              Rubio, Marco Republican    36.00
## 32     MEDFORD          Sanders, Bernard   Democrat   855.00
## 33   MONTCLAIR                 Bush, Jeb Republican   250.00
## 34   MONTCLAIR   Clinton, Hillary Rodham   Democrat 41881.96
## 35   MONTCLAIR            Fiorina, Carly Republican   500.00
## 36   MONTCLAIR   O'Malley, Martin Joseph   Democrat  1000.00
## 37   MONTCLAIR                Paul, Rand Republican   500.00
```

```
## 38   MONTCLAIR          Sanders, Bernard   Democrat  3055.00
## 39  MORRISTOWN                Bush, Jeb Republican  5400.00
## 40  MORRISTOWN   Clinton, Hillary Rodham   Democrat 25233.90
## 41  MORRISTOWN Cruz, Rafael Edward 'Ted' Republican  1425.00
## 42  MORRISTOWN           Huckabee, Mike Republican   450.00
## 43  MORRISTOWN               Paul, Rand Republican   195.00
## 44  MORRISTOWN          Sanders, Bernard   Democrat   251.88
## 45   PRINCETON                Bush, Jeb Republican 11800.00
## 46   PRINCETON   Clinton, Hillary Rodham   Democrat 45137.41
## 47   PRINCETON           Fiorina, Carly Republican   500.00
## 48   PRINCETON        Graham, Lindsey O. Republican  2750.00
## 49   PRINCETON             Rubio, Marco Republican  5860.00
## 50   PRINCETON          Sanders, Bernard   Democrat  2572.00
## 51   RIDGEWOOD                Bush, Jeb Republican 10800.00
## 52   RIDGEWOOD       Carson, Benjamin S. Republican  1000.00
## 53   RIDGEWOOD   Clinton, Hillary Rodham   Democrat 15507.25
## 54   RIDGEWOOD               Paul, Rand Republican   201.60
## 55   RIDGEWOOD             Rubio, Marco Republican   250.00
## 56   RIDGEWOOD          Sanders, Bernard   Democrat   757.81
## 57 WEST ORANGE       Carson, Benjamin S. Republican   250.00
## 58 WEST ORANGE   Clinton, Hillary Rodham   Democrat  8644.41
## 59 WEST ORANGE Cruz, Rafael Edward 'Ted' Republican   385.00
## 60 WEST ORANGE           Fiorina, Carly Republican   266.00
## 61 WEST ORANGE        Graham, Lindsey O. Republican  3000.00
## 62 WEST ORANGE               Paul, Rand Republican   474.66
## 63 WEST ORANGE             Rubio, Marco Republican   776.00
## 64 WEST ORANGE          Sanders, Bernard   Democrat   402.00
```

First plot the cities by total donation amount by party affiliation.



Next plot the total donation amounts by city for each individual candidiate.



The top 10 contributing cities are contributing more the the Democrats than to Republicans with most of the donations to Clinton (Dem). However some cities support other candidates with a larger majority (i.e. Livingston and Bush), but it appears the Clinton gets more total donations in these cities than other candidates.

Donation Amounts by Occupation

Going to examine the occupation information in the same way I just looked at cities. Again, beucase there are a lot of occupations in the dataset, I'm only going to look at the occupations that have the most donations (the top 10 occupations for frequency in the dataset). Earlier I made a new dataset set with the counts of the number of occurance for each city (called count_OCCUPATION). I can use this to create a new dataframe called "NJ_top_occu" which contains only the top 10 most frequently list occupations. With this new dataframe, I will calculate the total donation amount for each candidate and party affiliation by occupation.

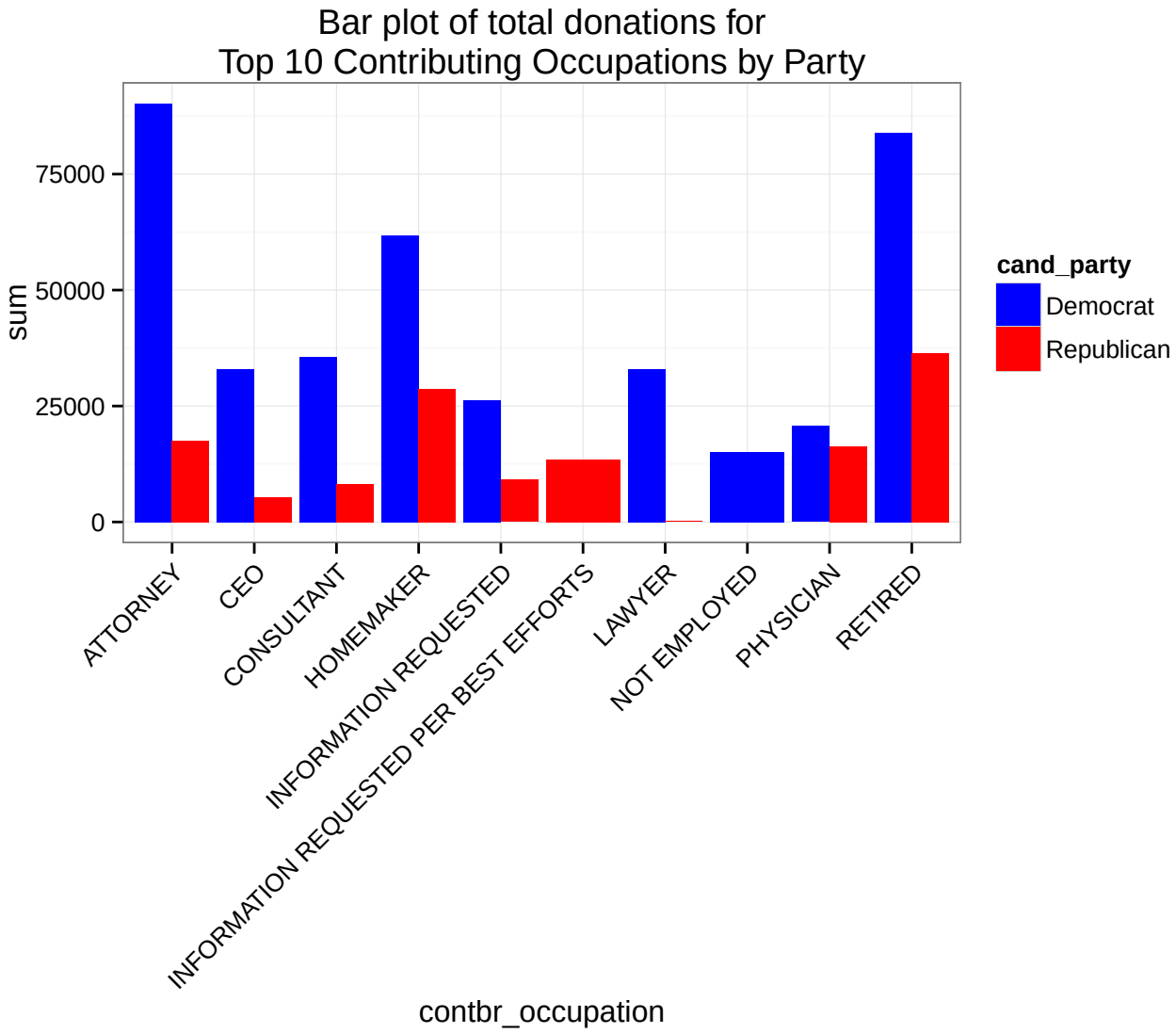
```
##                         contbr_occupation                   cand_nm
## 1                               ATTORNEY                 Bush, Jeb
## 2                               ATTORNEY       Carson, Benjamin S.
## 3                               ATTORNEY   Clinton, Hillary Rodham
```

```
## 4                                    ATTORNEY Cruz, Rafael Edward 'Ted'
## 5                                    ATTORNEY            Fiorina, Carly
## 6                                    ATTORNEY        Graham, Lindsey O.
## 7                                    ATTORNEY                Paul, Rand
## 8                                    ATTORNEY              Rubio, Marco
## 9                                    ATTORNEY           Sanders, Bernard
## 10                                   ATTORNEY      Santorum, Richard J.
## 11                                        CEO                 Bush, Jeb
## 12                                        CEO       Carson, Benjamin S.
## 13                                        CEO   Clinton, Hillary Rodham
## 14                                        CEO Cruz, Rafael Edward 'Ted'
## 15                                        CEO            Huckabee, Mike
## 16                                        CEO   O'Malley, Martin Joseph
## 17                                        CEO                Paul, Rand
## 18                                        CEO              Rubio, Marco
## 19                                        CEO           Sanders, Bernard
## 20                                 CONSULTANT                 Bush, Jeb
## 21                                 CONSULTANT   Clinton, Hillary Rodham
## 22                                 CONSULTANT                Paul, Rand
## 23                                 CONSULTANT           Sanders, Bernard
## 24                                  HOMEMAKER                 Bush, Jeb
## 25                                  HOMEMAKER       Carson, Benjamin S.
## 26                                  HOMEMAKER   Clinton, Hillary Rodham
## 27                                  HOMEMAKER Cruz, Rafael Edward 'Ted'
## 28                                  HOMEMAKER        Graham, Lindsey O.
## 29                                  HOMEMAKER            Huckabee, Mike
## 30                                  HOMEMAKER   O'Malley, Martin Joseph
## 31                                  HOMEMAKER         Pataki, George E.
## 32                                  HOMEMAKER              Rubio, Marco
## 33                      INFORMATION REQUESTED   Clinton, Hillary Rodham
## 34                      INFORMATION REQUESTED                Paul, Rand
## 35                      INFORMATION REQUESTED           Sanders, Bernard
## 36 INFORMATION REQUESTED PER BEST EFFORTS                 Bush, Jeb
## 37 INFORMATION REQUESTED PER BEST EFFORTS       Carson, Benjamin S.
## 38 INFORMATION REQUESTED PER BEST EFFORTS Cruz, Rafael Edward 'Ted'
## 39 INFORMATION REQUESTED PER BEST EFFORTS            Fiorina, Carly
## 40 INFORMATION REQUESTED PER BEST EFFORTS              Rubio, Marco
## 41 INFORMATION REQUESTED PER BEST EFFORTS      Santorum, Richard J.
## 42                                    LAWYER   Clinton, Hillary Rodham
## 43                                    LAWYER              Rubio, Marco
## 44                                    LAWYER           Sanders, Bernard
## 45                               NOT EMPLOYED   Clinton, Hillary Rodham
## 46                               NOT EMPLOYED           Sanders, Bernard
## 47                                  PHYSICIAN                 Bush, Jeb
## 48                                  PHYSICIAN       Carson, Benjamin S.
## 49                                  PHYSICIAN   Clinton, Hillary Rodham
## 50                                  PHYSICIAN Cruz, Rafael Edward 'Ted'
## 51                                  PHYSICIAN         Pataki, George E.
## 52                                  PHYSICIAN                Paul, Rand
## 53                                  PHYSICIAN              Rubio, Marco
## 54                                  PHYSICIAN           Sanders, Bernard
## 55                                    RETIRED                 Bush, Jeb
## 56                                    RETIRED       Carson, Benjamin S.
## 57                                    RETIRED   Clinton, Hillary Rodham
```

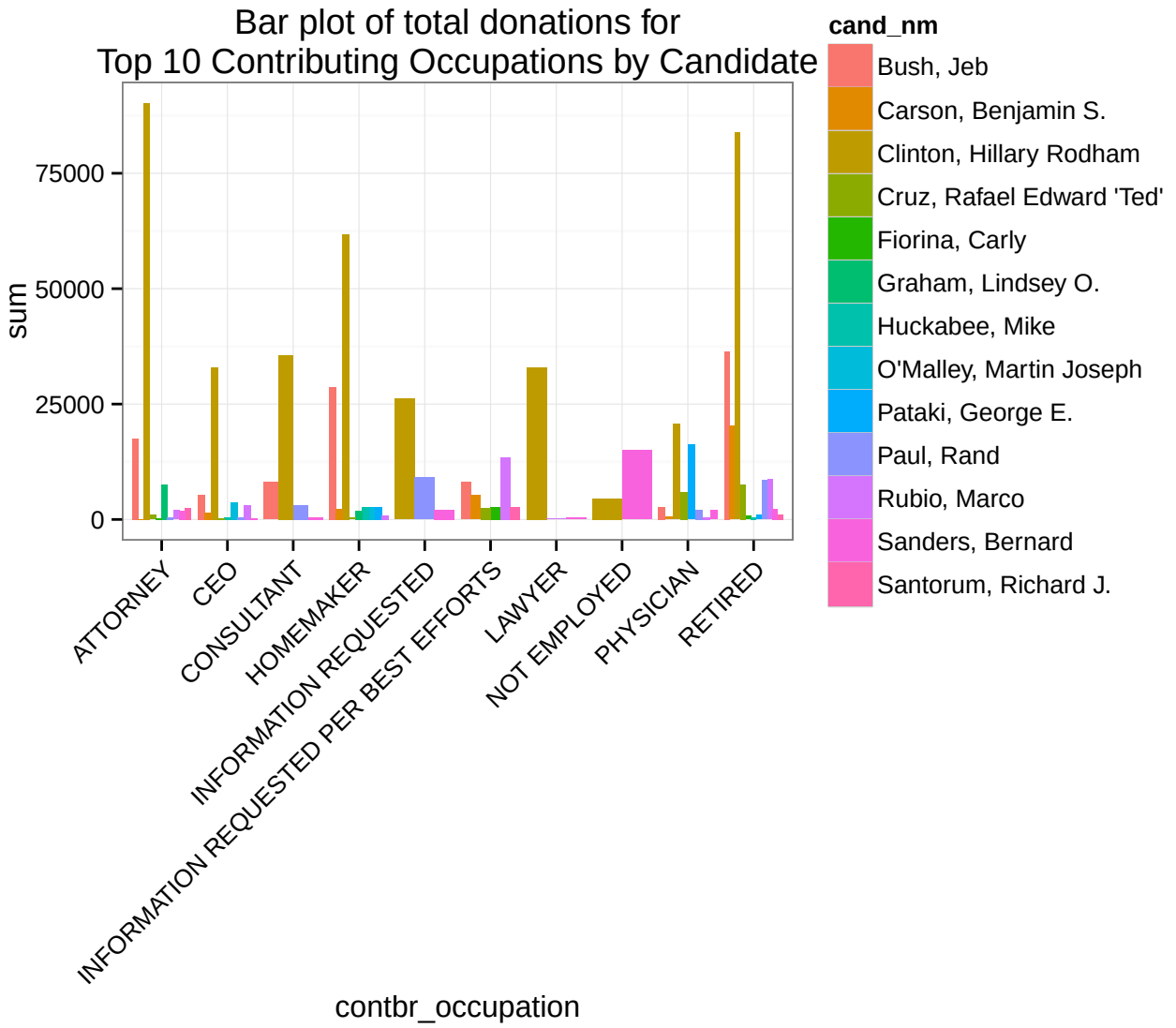
```
## 58                                  RETIRED Cruz, Rafael Edward 'Ted'
## 59                                  RETIRED            Fiorina, Carly
## 60                                  RETIRED            Huckabee, Mike
## 61                                  RETIRED         Pataki, George E.
## 62                                  RETIRED                Paul, Rand
## 63                                  RETIRED              Rubio, Marco
## 64                                  RETIRED          Sanders, Bernard
## 65                                  RETIRED      Santorum, Richard J.
##    cand_party      sum
## 1  Republican 17550.00
## 2  Republican    50.00
## 3    Democrat 90253.68
## 4  Republican   975.00
## 5  Republican   266.00
## 6  Republican  7600.00
## 7  Republican   474.66
## 8  Republican  2026.00
## 9    Democrat  1946.88
## 10 Republican  2500.00
## 11 Republican  5400.00
## 12 Republican  1495.00
## 13   Democrat 32935.00
## 14 Republican   250.00
## 15 Republican   500.00
## 16   Democrat  3700.00
## 17 Republican   495.00
## 18 Republican  3100.00
## 19   Democrat   250.00
## 20 Republican  8100.00
## 21   Democrat 35538.05
## 22 Republican  3101.60
## 23   Democrat   500.00
## 24 Republican 28650.00
## 25 Republican  2300.00
## 26   Democrat 61752.55
## 27 Republican   350.00
## 28 Republican  1850.00
## 29 Republican  2700.00
## 30   Democrat  2700.00
## 31 Republican  2700.00
## 32 Republican   863.00
## 33   Democrat 26120.00
## 34 Republican  9070.00
## 35   Democrat  2070.00
## 36 Republican  8100.00
## 37 Republican  5275.00
## 38 Republican  2485.00
## 39 Republican  2700.00
## 40 Republican 13520.00
## 41 Republican  2700.00
## 42   Democrat 32918.41
## 43 Republican   200.00
## 44   Democrat   402.00
## 45   Democrat  4483.00
```

```
## 46   Democrat 14977.05
## 47 Republican  2700.00
## 48 Republican   600.00
## 49   Democrat 20638.00
## 50 Republican  5950.00
## 51 Republican 16200.00
## 52 Republican  2153.20
## 53 Republican   500.00
## 54   Democrat  2000.00
## 55 Republican 36350.00
## 56 Republican 20235.00
## 57   Democrat 83794.48
## 58 Republican  7627.00
## 59 Republican   750.00
## 60 Republican   450.00
## 61 Republican  1000.00
## 62 Republican  8520.66
## 63 Republican  8711.00
## 64   Democrat  2330.33
## 65 Republican  1050.00
```

First plot the total donation amount by occupaton and party affiliation.



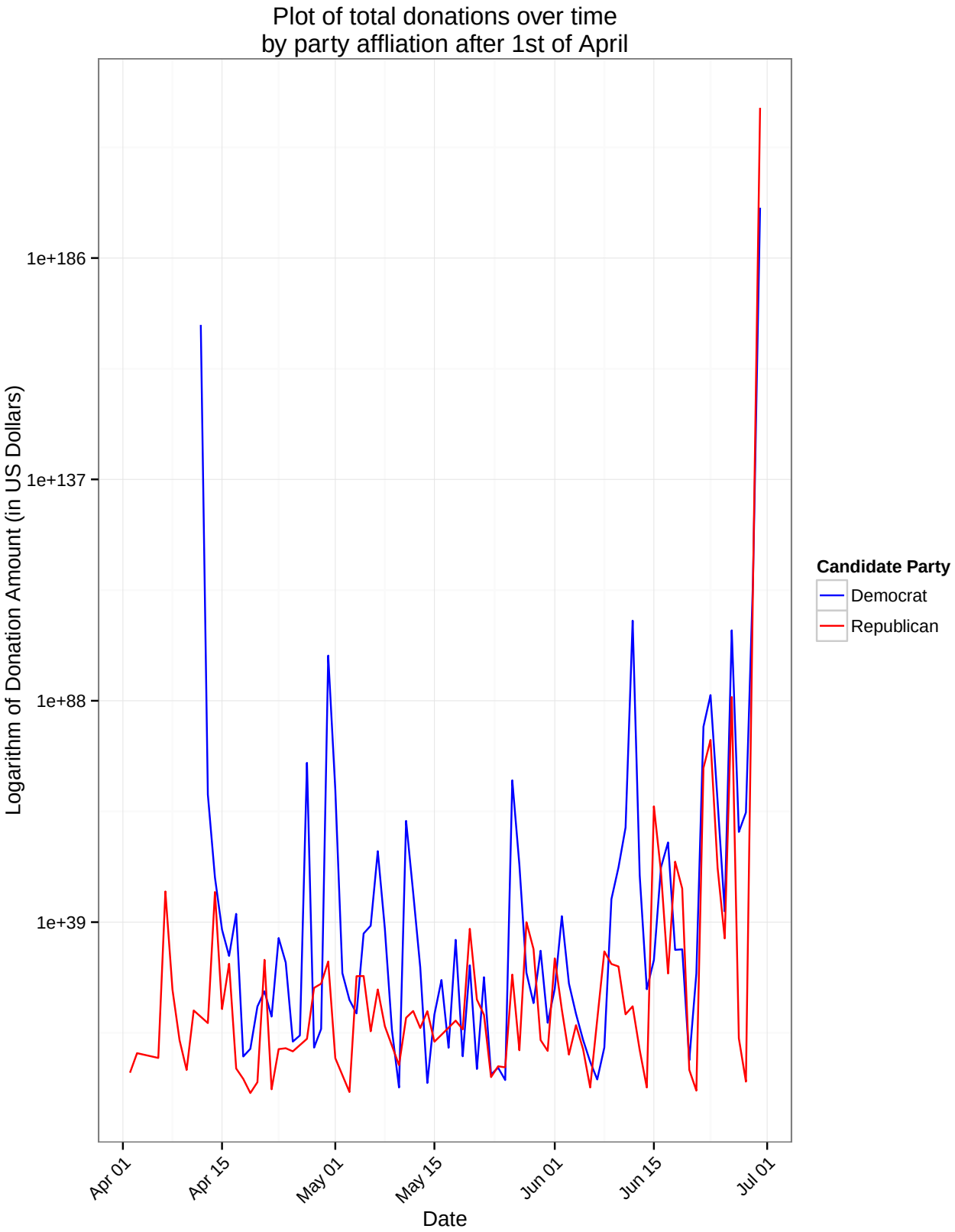
Next plot the total donation amount by occupation and individual candidate.



Again, because Democrats have more donations overall, the plots show that Democrats are getting more donations from the top 10 occupations, but there are some interesting patterns. Lawyers and Not Employed include only donations to Democrats and none to Republicans, whereas those that did not disclose what they did (Information requested per best efforts) only donated to Republicans and not Democrats. Also for Retired, Physicians and Homemarkers, the split appears to be much closer to 50/50 than other occupations. Also, the candidate receiveing the most money for almost all occupations is Clinton (Dem), except for those that did not disclose what they did (Information requested per best efforts) wher Rubio (Rep) received the largest amount, and for not employed indivdiuals who gave more money to Sanders (Dem) than Clinton (Dem).

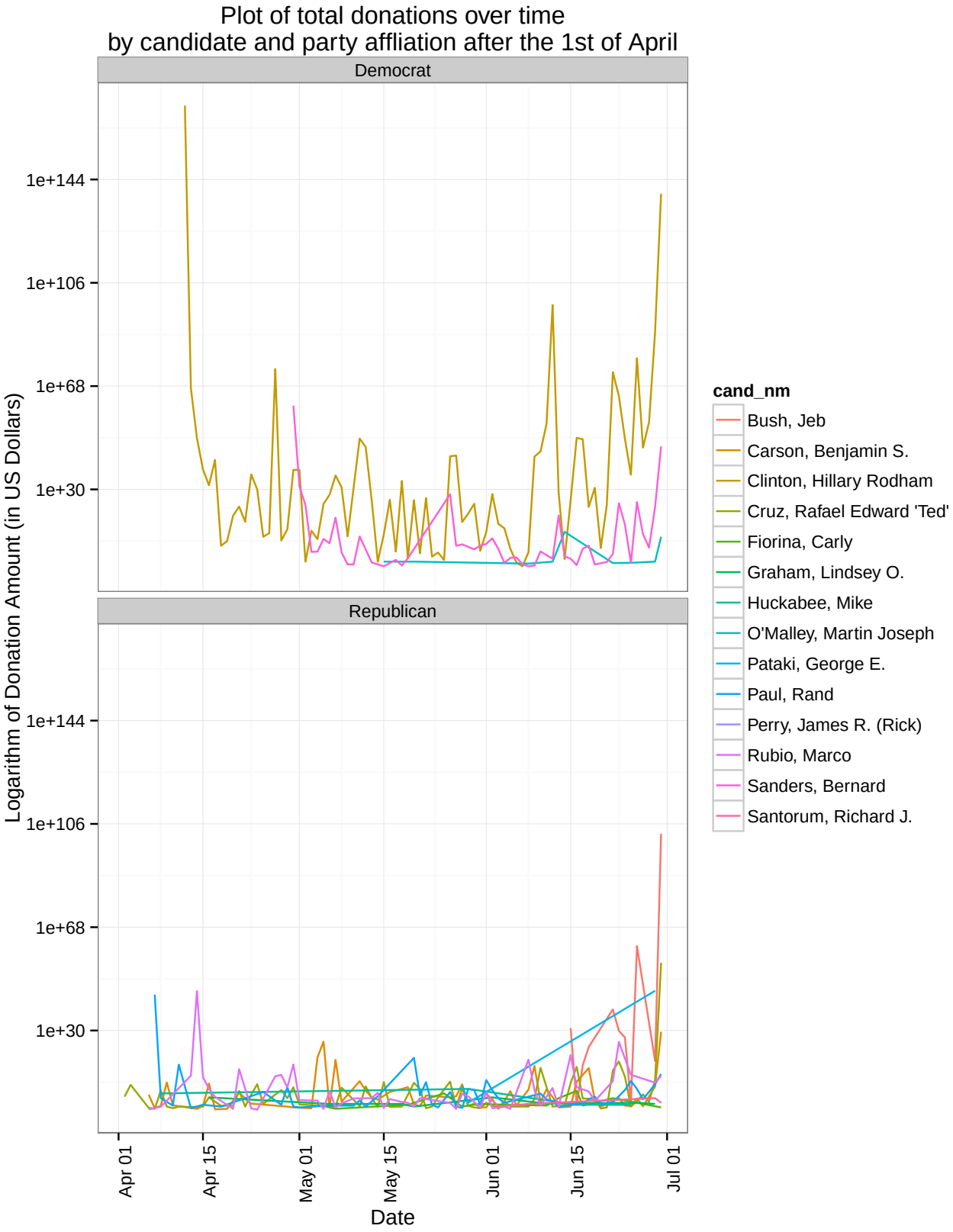
Final Plots

The first plot is shows the donation amounts received over time to each party (Republican or Democrat). This plot shows the differences in donations made, particularly to the Republican party as it appears to be gaining in donation amount over time. It would be interesting to re-assess this data in 6 to 12 months time to see if that trend continues.



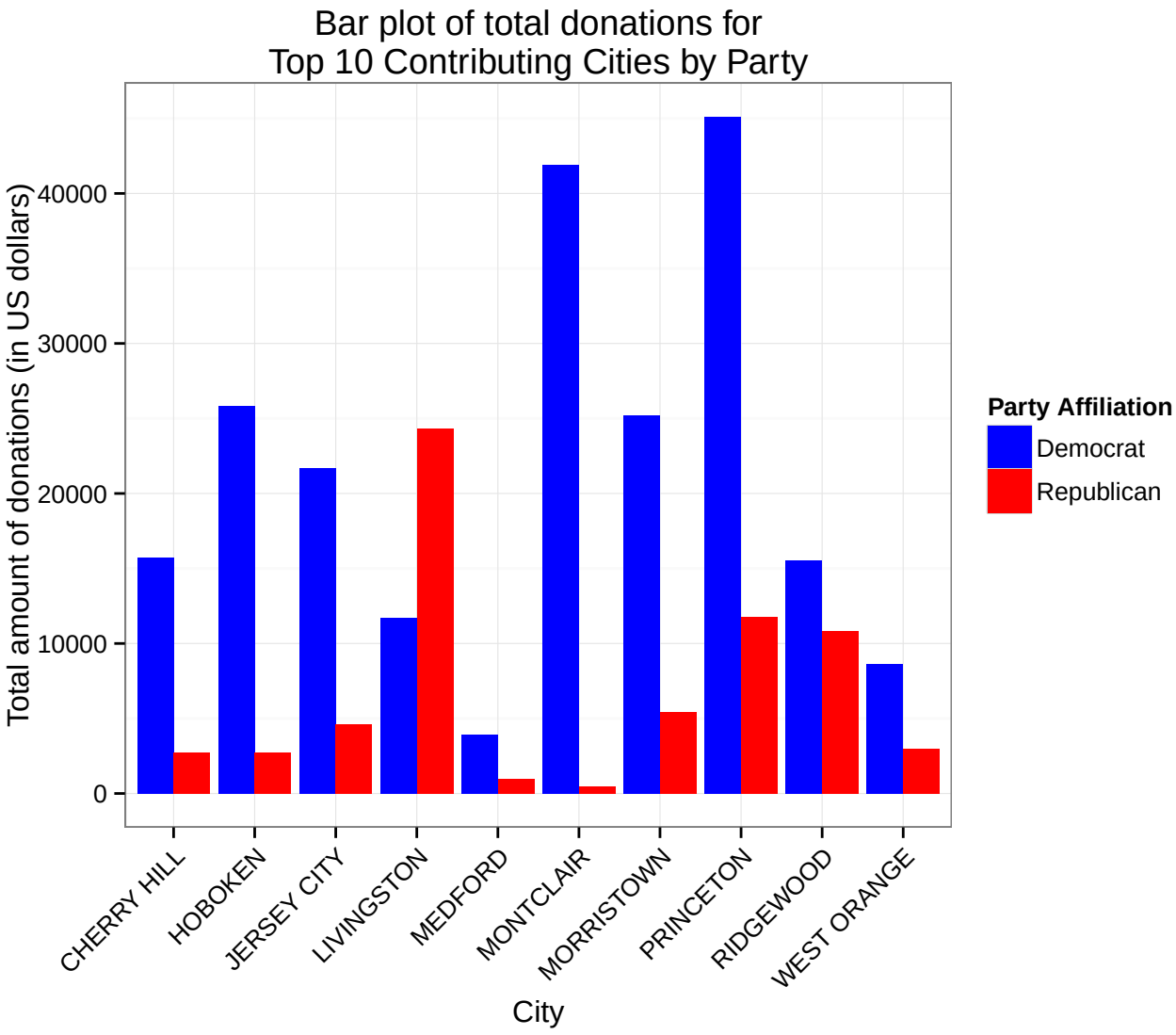
Furthermore, the plot which shows the donations amounts to indivdiaul candidates by party affliation is also interesting. This shows that Clinton recieves a consistent amount of donations over time which is higher than Sanders, but really the Repubican side of this plot is more interesting than the Democrats. Clinton has

received the most money over time even in comparison with the Republicans who all recieve smaller amounts over time than Clinton does. The Republicans have more candidates, and this plot shows when some declared their intentions to run for President (as this is approximately near the time they start receiving donations). Bush for the Republicans is one of the last to start recieving donations and his line is increasing at teh very end of the plot, but on average they are large, and this might in part explain the upward trend seen in the previous plot.



The last plot choosen is the bar plot of total donations made from the top ten contributing cities by party affiliation. While most show that donations are largely to Democratic candidates, Livingston is intersting as they have donated more to Republicans than Democrats. It also reflects differences between cities in New

Jersey and their political leanings. While most (9 out of 10) appear to most heavily support (via monetary donations) a Democratic candidate, one town favors Republicans, showing that location and potential through that socio-economic status or other factors linked to location, play a role in political support.



Reflection

This is the first time that I have worked with this type of data, and I found it very interesting.

There were some anomalies in the data, particularly with regards to donations amounts. There were some in the data that are over the legal limit ($2700), and these are obviously in violation of the law. Having done an internet search on this, it appears these are normally refunded (as some donations in this dataset were), or part of the amount transfered to be a donation in a spouses name (as was also the case for some of the data in this dataset). For this reason, I choose to only focus on donations that I felt were valid, meaning falling within the values of $1 and $2700.

It was difficult to get to know the data, and all the various columns of information. The date information needed to be reformatted to work correctly. Other information was lacking, like the employer and occupation information was only there for a subset of individuals. Cleaning needed to be done for candidate names, which is a little worrying. This needed to be updated to give accurate reflections of the donations.

I also gave the individual candidates party affiliation, to assess differences between the two main US parties in how individuals form NJ were donating. Not overly suprisingly, Democratic candidates were more successively at raising money in NJ than Republican counterparts. I think this is because NJ as a state in recent elections (the past 6) votes for Democratic candidates, suggesting Democratic leanings to the NJ populations.

Most people were donating to Clinton above all other candidates, but the time series plots showed that there were some increases on the Republican side which might be due to individuals making contributions to Bush who entered the race later than most other individauls. Re-assessment of this data in 6 months time may show slightly different patterns if people continue to donate to Bush to the same extent as they had started here.

Overall I found this incredibly interesting, and will most likely download the data in a few months time to see how it has changed.