

# Bernie Sanders & Donald J. Trump Contributions.

*Kris Anaya*

*5/22/2017*

This report explores a dataset containing candidates and attributes for approximately 490,000 contributions.

## Univariate Plot Section

```
## 'data.frame': 493422 obs. of 12 variables:
## $ name : Factor w/ 107161 levels "& DREW BURKE, MELANIE",...: 1 2 3 3 3 4 5 5 5 5 ...
## $ candidate : Factor w/ 3 levels "cand_nm","Sanders, Bernard",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ city : Factor w/ 2251 levels "29 PALMS","ACAMPO",...: 1067 807 681 681 681 956 225 225 225 225 ...
## $ state : Factor w/ 2 levels "CA","contbr_st": 1 1 1 1 1 1 1 1 1 1 ...
## $ zip : num 9.55e+08 9.25e+08 9.29e+08 9.29e+08 9.29e+08 ...
## $ employer : Factor w/ 29481 levels "", "-", "- NONE -",...: 12217 10319 13216 13216 13216 3465 10227 ...
## $ occupation: Factor w/ 13683 levels "", "-", "----",...: 6653 6653 7000 7000 7000 8062 5233 5233 5233 ...
## $ amount : num 50 244 15 50 50 ...
## $ date : POSIXlt, format: "2016-04-17" "2015-09-04" ...
## $ receipt : Factor w/ 37 levels "", "* EARMARKED CONTRIBUTION: SEE BELOW REATTRIBUTION/REFUND PEND...: 1 1 1 1 1 1 1 1 1 1 ...
## $ lon : num -123 -117 -118 -118 -118 ...
## $ lat : num 39.4 34 33.8 33.8 33.8 ...

## name candidate
## WEIL, MONIQUE : 257 cand_nm : 0
## MCLENNAN, MARLYN : 240 Sanders, Bernard:407164
## AUSLENDER, LEONARD: 239 Trump, Donald J.: 86258
## ISERI, MARTIN : 220
## SPEAR, JOSEPH : 210
## DAVIDSON, LISA : 203
## (Other) :492053

## city state zip
## LOS ANGELES : 33799 CA :493422 Min. : 0
## SAN FRANCISCO: 32011 contbr_st: 0 1st Qu.:900683036
## SAN DIEGO : 17770 Median :926777949
## OAKLAND : 13124 Mean :759234179
## SAN JOSE : 11642 3rd Qu.:946091302
## BERKELEY : 10707 Max. :961628693
## (Other) :374369

## employer occupation amount
## NONE : 80793 NOT EMPLOYED :105838 Min. : -10500.00
## NOT EMPLOYED : 43044 RETIRED : 50674 1st Qu.: 15.00
## RETIRED : 37507 INFORMATION REQUESTED: 16746 Median : 27.00
## SELF : 33131 TEACHER : 11655 Mean : 68.16
## SELF EMPLOYED: 29749 ENGINEER : 8584 3rd Qu.: 50.00
## (Other) :268832 (Other) :299849 Max. : 10000.00
## NA's : 366 NA's : 76

## date
## Min. :2015-05-04 00:00:00
## 1st Qu.:2016-02-27 00:00:00
## Median :2016-04-06 00:00:00
```

```

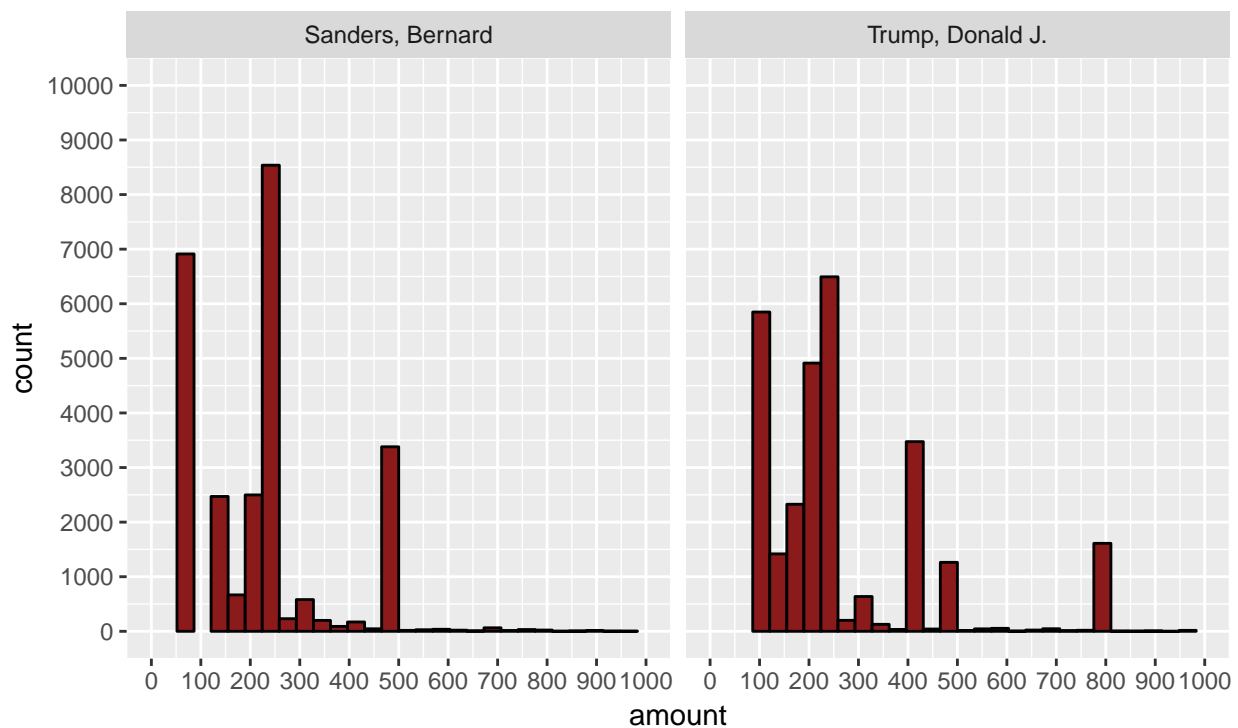
## Mean      :2016-04-09 14:49:04
## 3rd Qu.   :2016-05-31 00:00:00
## Max.      :2016-12-31 00:00:00
##
##
##                                     receipt
##                                     :489797
## Refund                                           : 3137
## * EARMARKED CONTRIBUTION: SEE BELOW REATTRIBUTION/REFUND PENDING: 431
## REATTRIBUTION/REFUND PENDING                    : 16
## REATTRIBUTION TO SPOUSE, TERRIE SCHULTZ.         : 4
## REATTRIBUTION FROM SPOUSE, DEDE WHITESIDE.       : 3
## (Other)                                          : 34
##
## lon      lat
## Min.     :-124.286 Min.     :23.15
## 1st Qu.   :-122.236 1st Qu. :34.05
## Median    :-120.572 Median  :36.68
## Mean      :-120.112 Mean    :36.00
## 3rd Qu.   :-118.235 3rd Qu. :37.80
## Max.      : 2.239 Max.    :48.88
## NA's      :2727 NA's    :2727

```

Within the summary you can see that largest city of contributors is Los Angeles. An overwhelming amount are Not Employed and most people made contributions around April of 2016. There are some strange things to point out like the minimum amount is listed as -10,500 dollars. This took me awhile to understand and with a little more investigating you can see in the receipt section there is a section called refunds were contributors may refund based on legalities of contributions in the state of california.

## Bernie v Trump Contributions

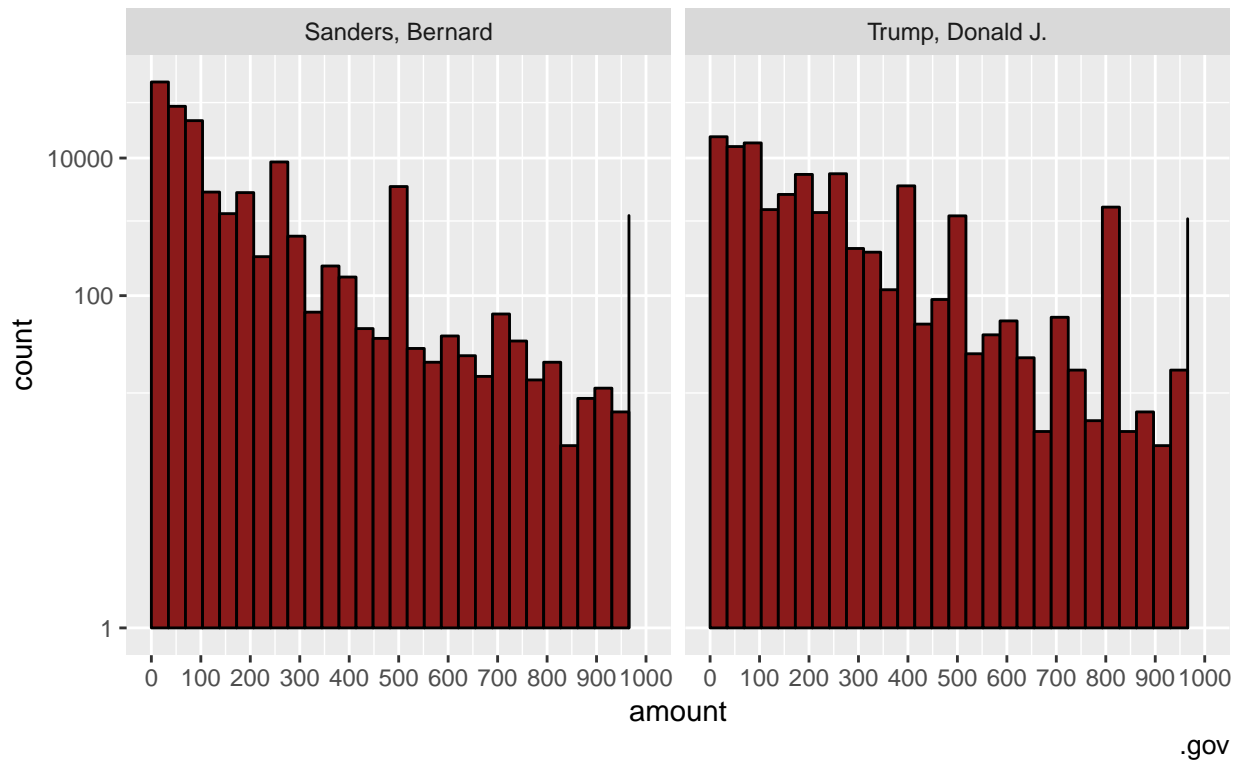
Count vs. Amount



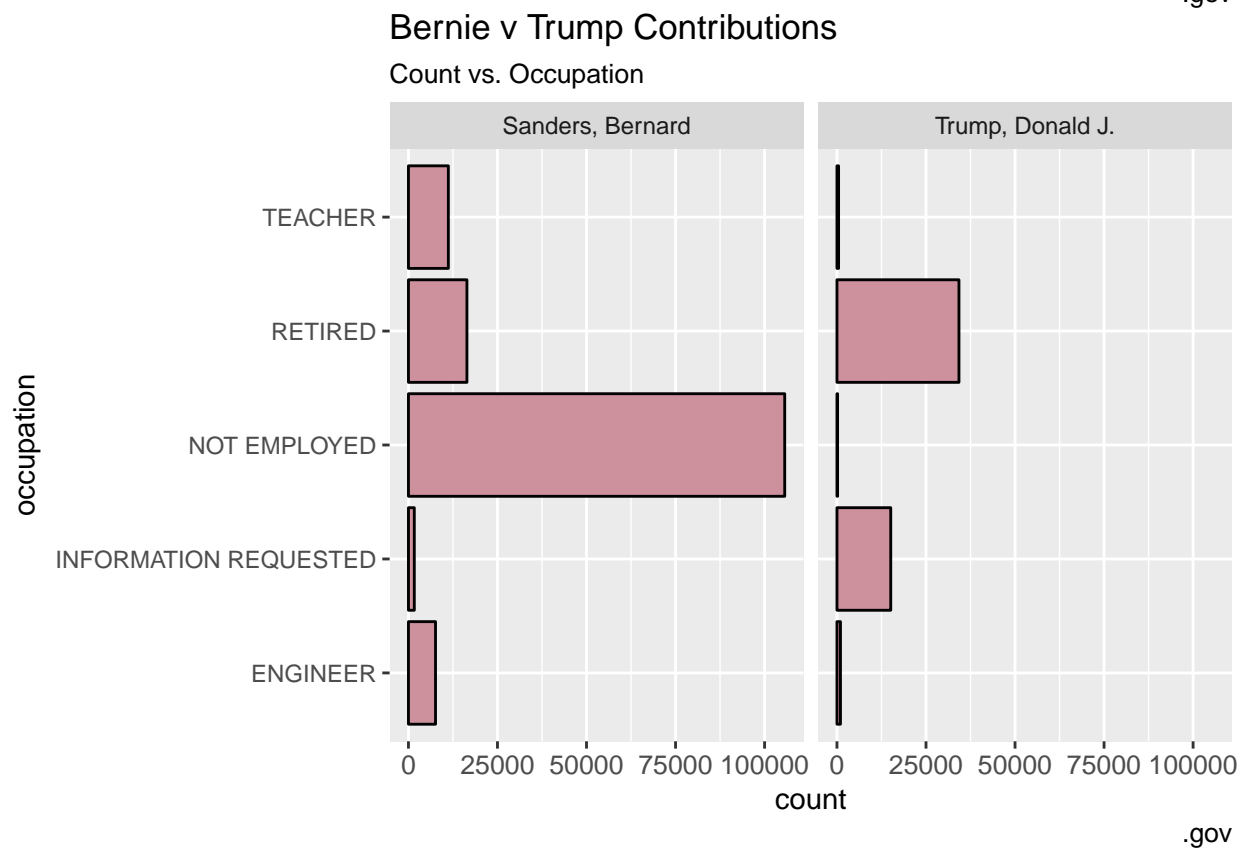
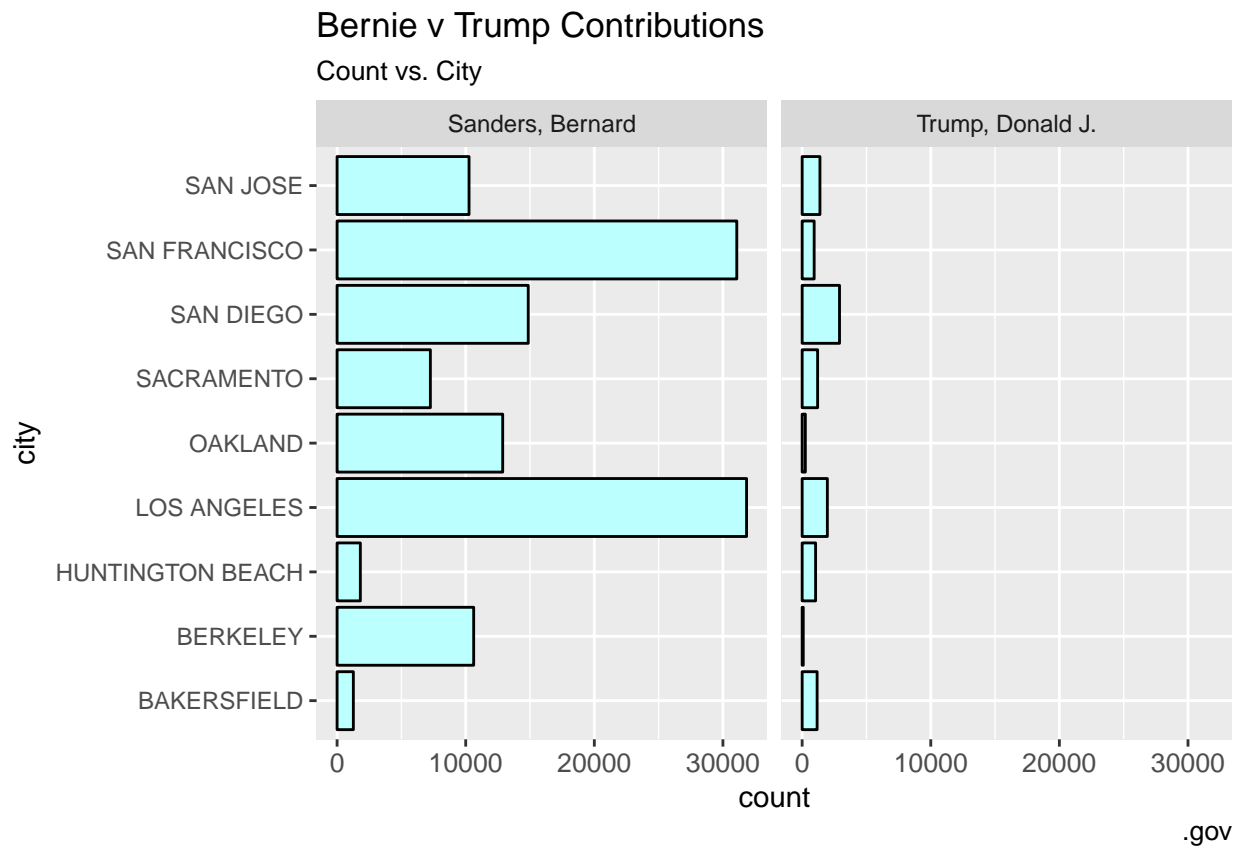
.gov

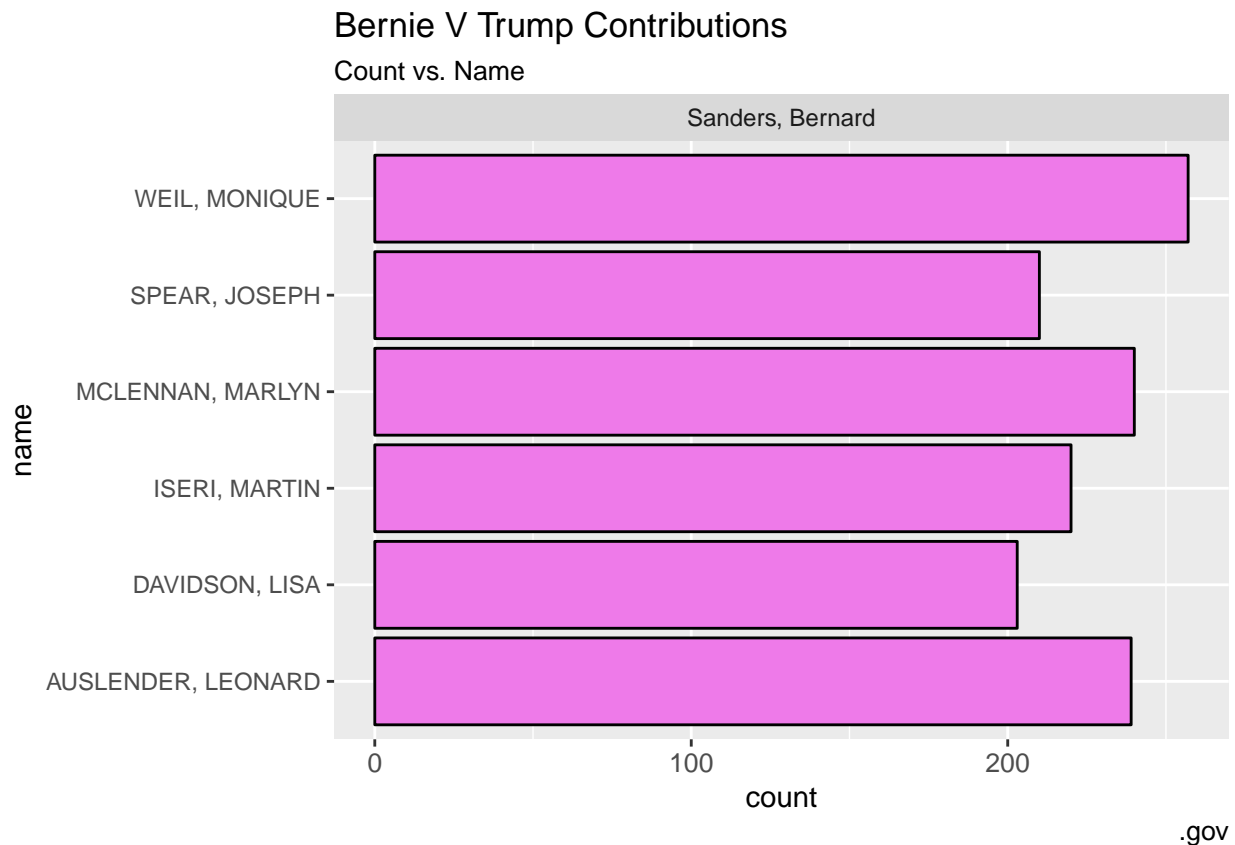
## Bernie v Trump Contributions Base 10

Count vs. Amount



Transformed both long tail-esque data to better understand the distribution of amount by each candidate. The transformed amount appears relatively similar with both candidates. Although, Sanders has a bit more contributions that are less than the fifty-dollar range peaking above 10 thousand, while Trump appears to have a larger count of amount around the nine-hundred dollar range. Let's investigate some other variables like cities, Occupation and name.

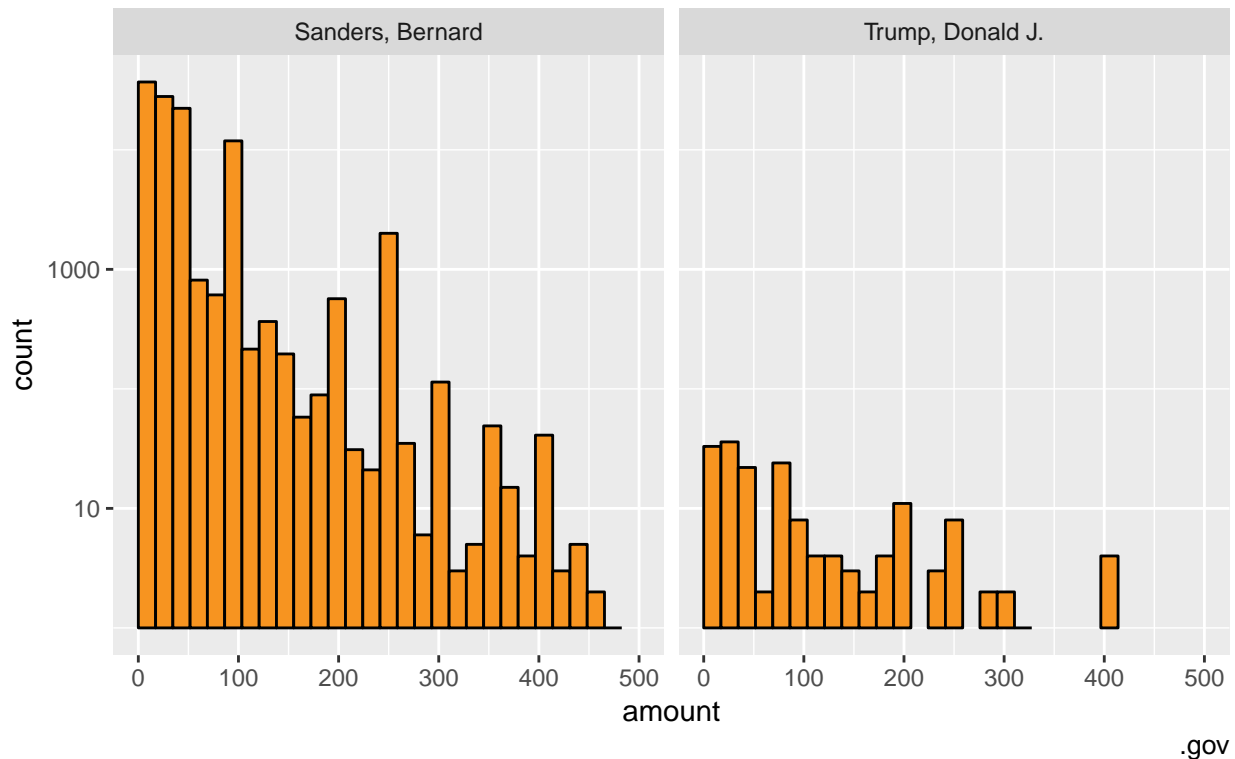




Most Major cities contributed to Bernie's campaign, with pretty smiliar contributions in Bakersfield and Hunnigton Beach. A majority of occupatations that contributed to Bernie's campaign are "Not Employed", while more in Trump's contributions are "RETIRED". Finally, that contributed more then two hundred contributions seem to support the Bernied campaign rather than Trump's. It looks like no large multiple contributions were made with the Trump campaign in California.

## Bernie V Trump Contributions

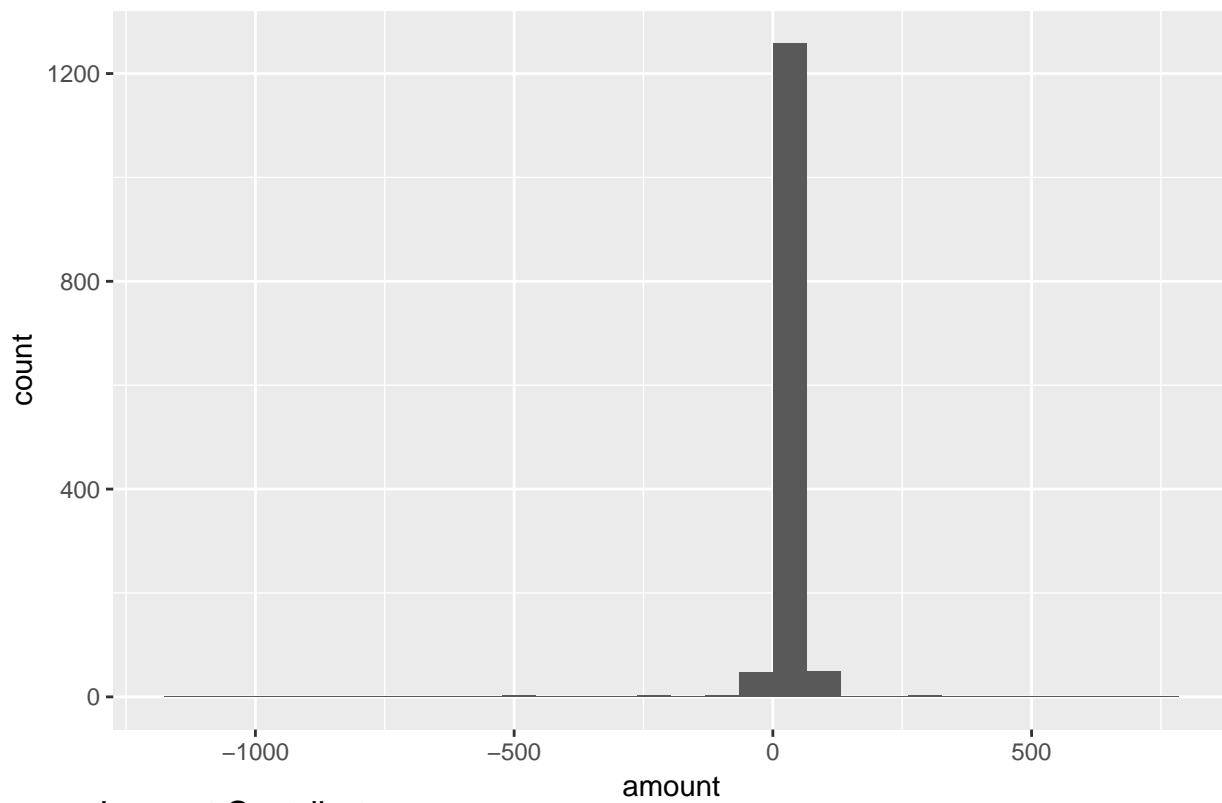
Count vs. Name



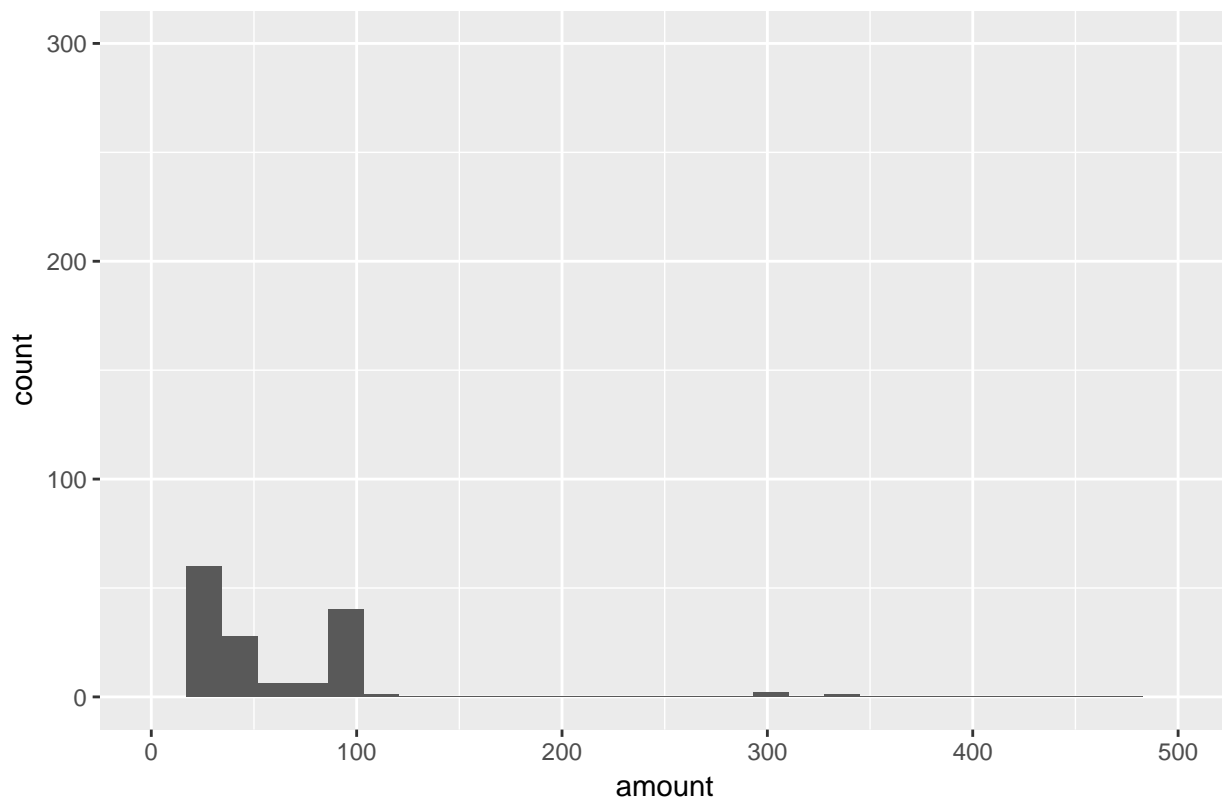
```
## not_employed_df$state: CA
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2300.00   15.00   27.00   50.16   50.00  5000.00
## -----
## not_employed_df$state: contbr_st
## NULL
```

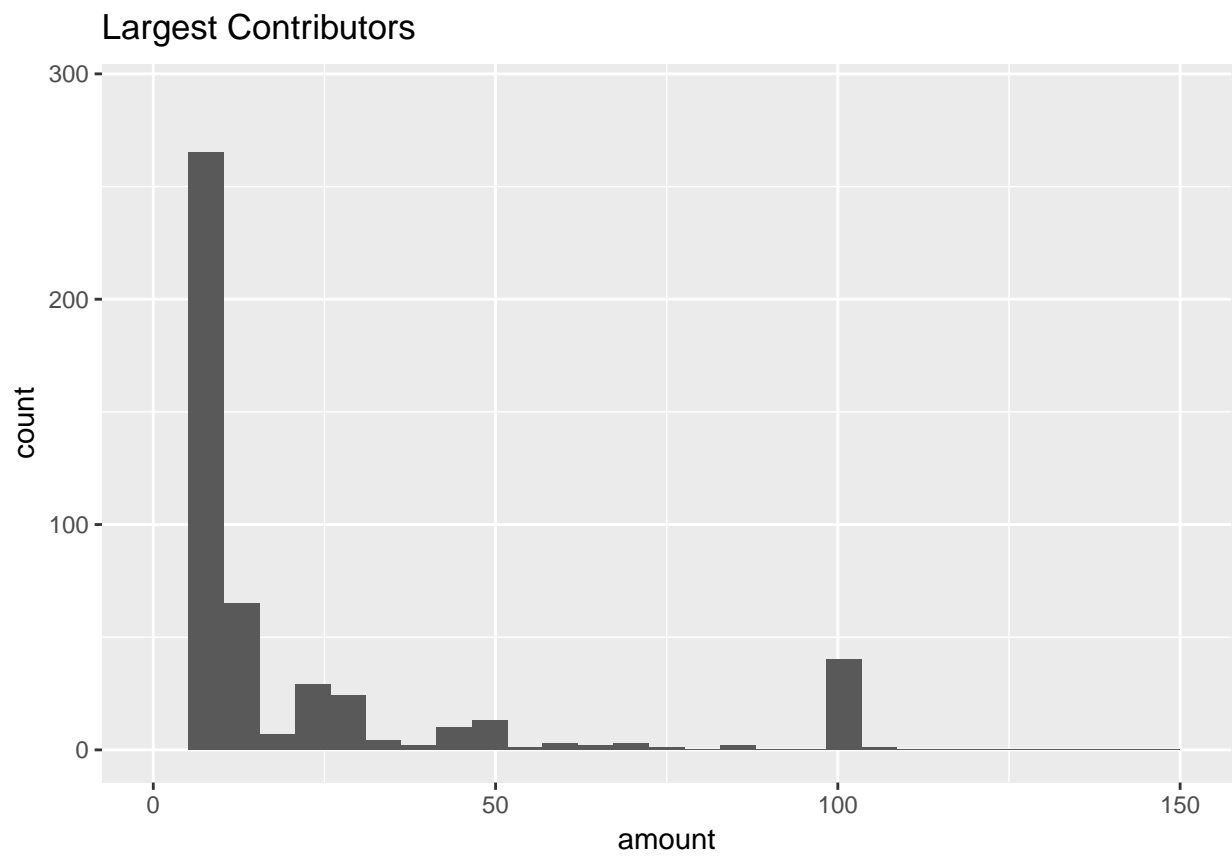
If we consider the unemployment between two candidates it would seem that mister Sanders has the largest. Most poeple who are unemployed fall into the zero to hundred dollar amount while we see a converging pattern as the amount increases. Within the summary we see that the median is exactly twenty-seven dollars the same amount which Mr. Sanders proclaimed in his campaign speeches.

Largest COntributors



Largest Contributors

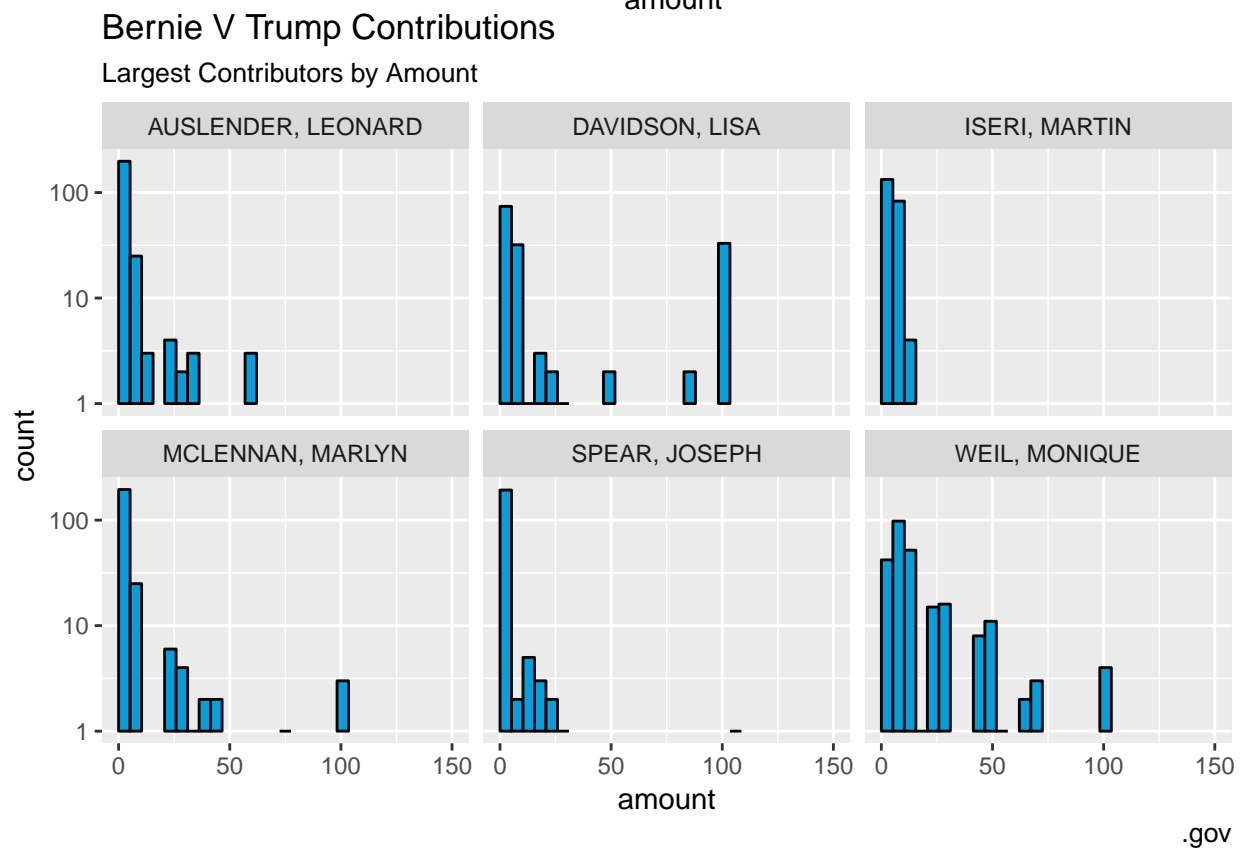
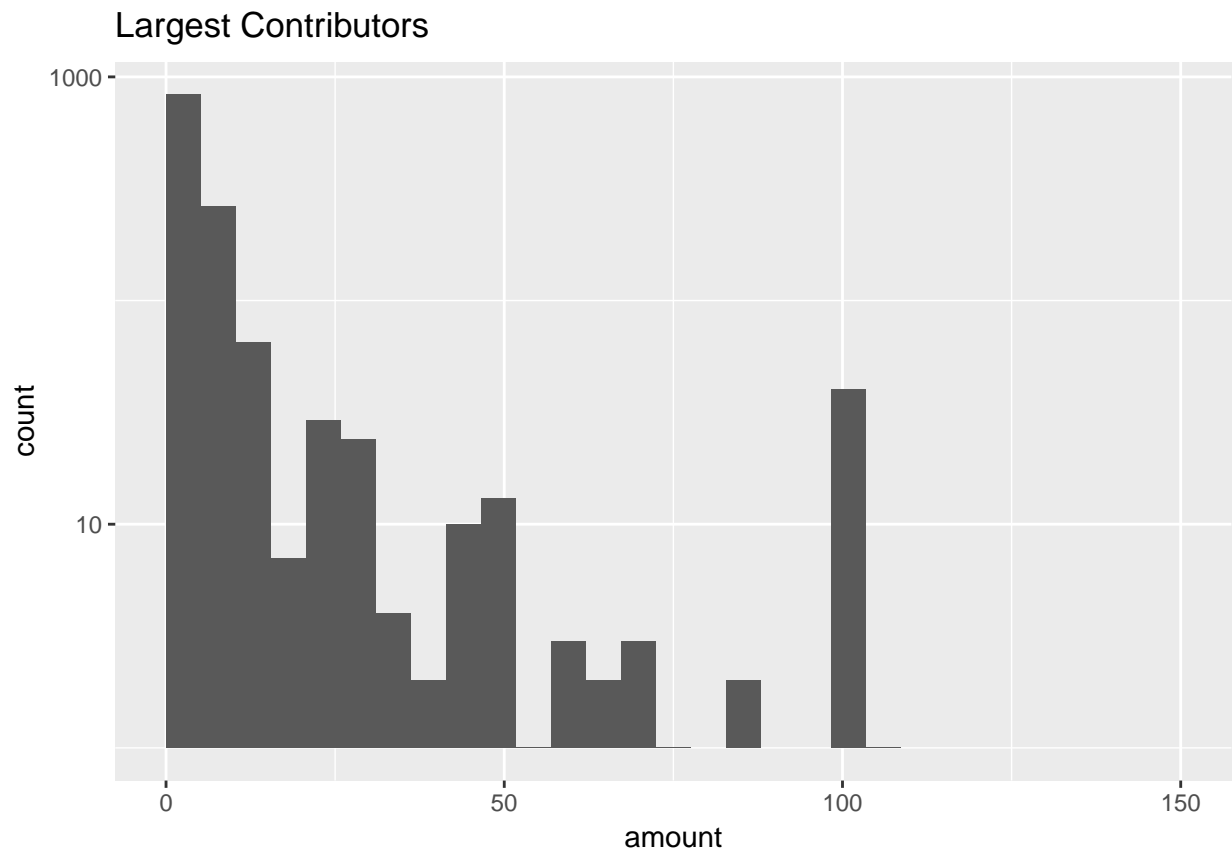




```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1153.00   2.70    3.00    8.03  10.00  742.40

##      Mode  FALSE  TRUE  NA's
## logical   150   1219     0
```





I wanted to do a step by step approach of the dataset of the names who contributed the most. Within the

first three plots I had noticed an error by checking the data set within it's quantiles. It was brought to my attention in the third plot that 1200 observations between 0 and 30 are not identified in the first three plots. This can always be problematic in dealing with data sets. Fortunately, I was able to remedy the problem by moving the boundary line back to zero and rescaling the data set to show a more accurate description of the original plot. All of these names contributed to the Bernie Sanders campaign as was visualized above. When we zoom in and look closer at the data and separate the names in each column while taking a log transformation of the slightly skewed chart we can get a very clear visualization of how the contributors spread out there payments. I didn't want to get into the refund payments just yet in the data set because it requires multiple variables to investigate further, but as you can see from the chart above most of the contributions to the Sanders Campaign were paid around a wide variety of ranges.

```
## 'data.frame': 493422 obs. of 12 variables:
## $ name : Factor w/ 107161 levels "& DREW BURKE, MELANIE",...: 1 2 3 3 3 4 5 5 5 5 ...
## $ candidate : Factor w/ 3 levels "cand_nm","Sanders, Bernard",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ city : Factor w/ 2251 levels "29 PALMS","ACAMPO",...: 1067 807 681 681 681 956 225 225 225 225 ...
## $ state : Factor w/ 2 levels "CA","contbr_st": 1 1 1 1 1 1 1 1 1 1 ...
## $ zip : num 9.55e+08 9.25e+08 9.29e+08 9.29e+08 9.29e+08 ...
## $ employer : Factor w/ 29481 levels "", "-", "- NONE -",...: 12217 10319 13216 13216 13216 3465 10227 ...
## $ occupation: Factor w/ 13683 levels "", "-", "----",...: 6653 6653 7000 7000 7000 8062 5233 5233 5233 ...
## $ amount : num 50 244 15 50 50 ...
## $ date : POSIXlt, format: "2016-04-17" "2015-09-04" ...
## $ receipt : Factor w/ 37 levels "", "* EARMARKED CONTRIBUTION: SEE BELOW REATtribution/REFUND PEND",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ lon : num -123 -117 -118 -118 -118 ...
## $ lat : num 39.4 34 33.8 33.8 33.8 ...

## name candidate
## WEIL, MONIQUE : 257 cand_nm : 0
## MCLENNAN, MARLYN : 240 Sanders, Bernard:407164
## AUSLENDER, LEONARD: 239 Trump, Donald J.: 86258
## ISERI, MARTIN : 220
## SPEAR, JOSEPH : 210
## DAVIDSON, LISA : 203
## (Other) :492053

## city state zip
## LOS ANGELES : 33799 CA :493422 Min. : 0
## SAN FRANCISCO: 32011 contbr_st: 0 1st Qu.:900683036
## SAN DIEGO : 17770 Median :926777949
## OAKLAND : 13124 Mean :759234179
## SAN JOSE : 11642 3rd Qu.:946091302
## BERKELEY : 10707 Max. :961628693
## (Other) :374369

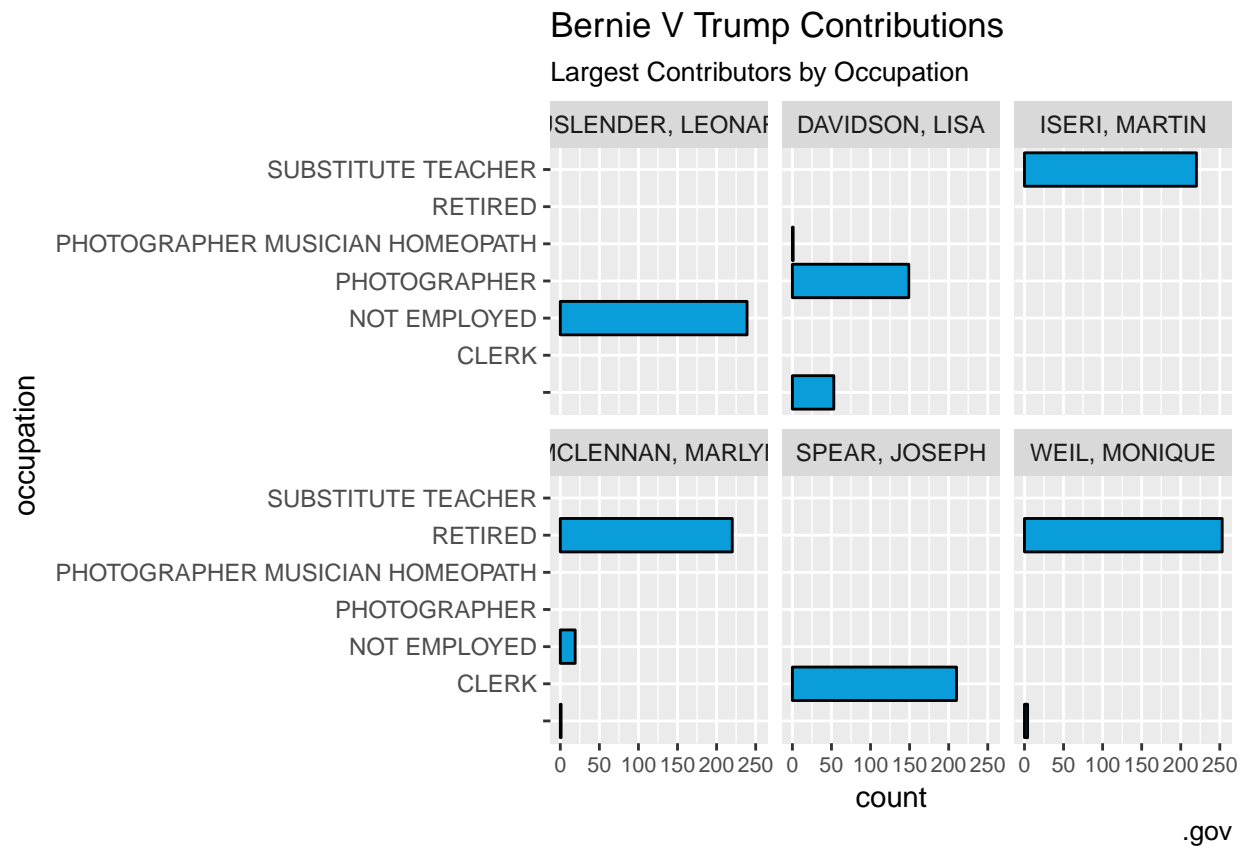
## employer occupation amount
## NONE : 80793 NOT EMPLOYED :105838 Min. : -10500.00
## NOT EMPLOYED : 43044 RETIRED : 50674 1st Qu.: 15.00
## RETIRED : 37507 INFORMATION REQUESTED: 16746 Median : 27.00
## SELF : 33131 TEACHER : 11655 Mean : 68.16
## SELF EMPLOYED: 29749 ENGINEER : 8584 3rd Qu.: 50.00
## (Other) :268832 (Other) :299849 Max. : 10000.00
## NA's : 366 NA's : 76

## date
## Min. :2015-05-04 00:00:00
## 1st Qu.:2016-02-27 00:00:00
## Median :2016-04-06 00:00:00
## Mean :2016-04-09 14:49:04
## 3rd Qu.:2016-05-31 00:00:00
```

```

## Max.      :2016-12-31 00:00:00
##
##
##                                receipt
##                                :489797
## Refund                                : 3137
## * EARMARKED CONTRIBUTION: SEE BELOW REATTRIBUTION/REFUND PENDING: 431
## REATTRIBUTION/REFUND PENDING          : 16
## REATTRIBUTION TO SPOUSE, TERRIE SCHULTZ.      : 4
## REATTRIBUTION FROM SPOUSE, DEDE WHITESIDE.    : 3
## (Other)                                   : 34
##
## lon      lat
## Min.     :-124.286 Min.     :23.15
## 1st Qu.  :-122.236 1st Qu. :34.05
## Median   :-120.572 Median   :36.68
## Mean     :-120.112 Mean     :36.00
## 3rd Qu.  :-118.235 3rd Qu. :37.80
## Max.     : 2.239 Max.     :48.88
## NA's     :2727 NA's     :2727

```



```

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1153.00  2.70    3.00    8.03  10.00   742.40

```

In this second graph concerning occupation you can see that people had standard employments such as Teacher, Not Employed, Retired, Clerk and Photographer. What struck me odd in this barplot was some people changed there occupation with donations for example MCLENNAN, MARLYN was listed as RETIRED and NOT EMPLOYED.

## Univariate Analysis

### What is the structure of your dataset?

This report explores a dataset containing candidates and attributes for approximately 490,000 contributions with 12 features (name, candidate, city, state, zip, employer, occupation, amount, date, receipt, lon, lat). The variables name, candidate, city state, employer, occupation are factors.

Other observations: \* cities with most contributors are leaning liberal \* largest contributor occupation is not employed \* largest count of contributions are Bernie supporters \* largest refund is more than allotted 10k \* biggest contributions is the upperbound of 10k

### What is/are the main feature(s) of interest in your dataset?

the main features of the dataset are amount, city and occupation. I'd like to determine which features are voting for each respected candidates the most and what is their median or max contributions for each candidate as well as top locations.

### Did you create any new variables from existing variables in the dataset?

They only feature I truly changed within the data set is the datetime which I re translated it to a more formal method so I may view it during my bivariate plot section. Other then that most variables were just re written to express a more readable form.

### Of the features you investigated, were there any unusual distributions? Did you perform

any operations on the data to tidy, adjust, or change the form of the data? If so, why did you do this?

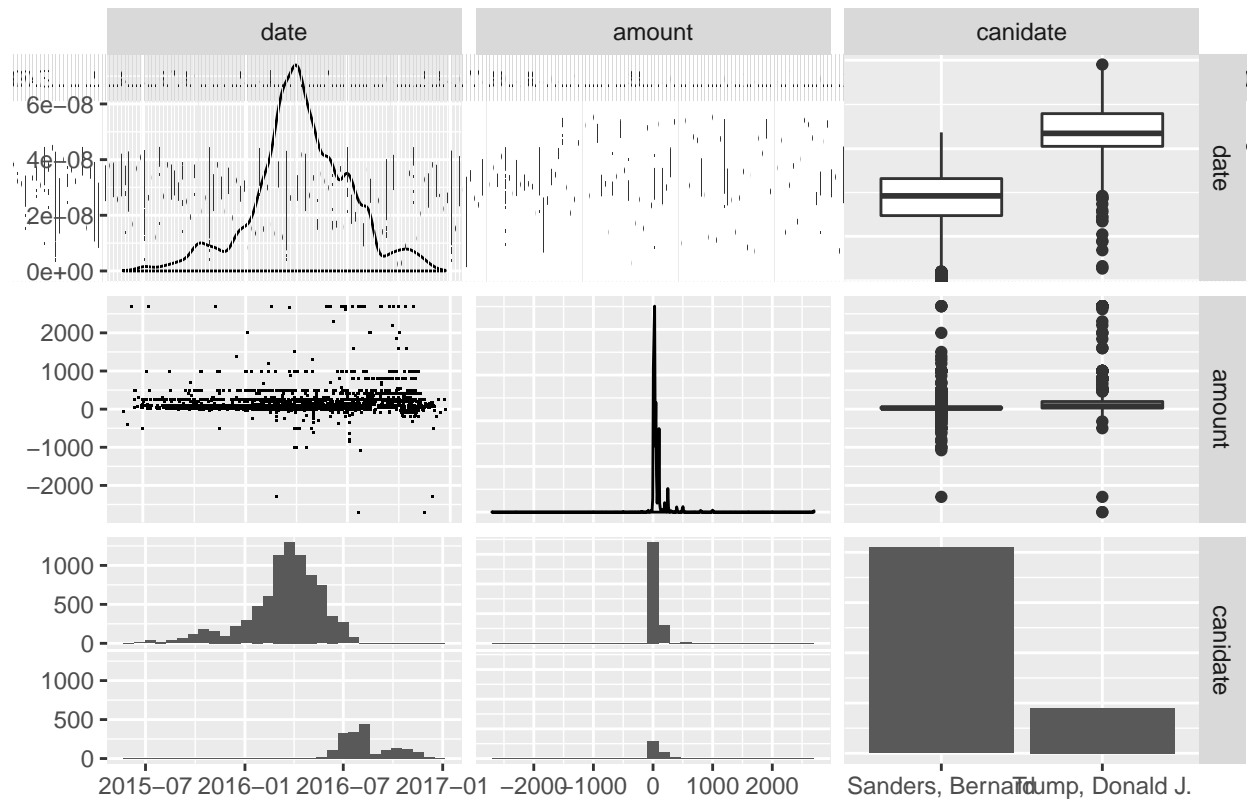
I log-transformed the right skewed of the amount and count of the contributions in the candidate. While the price still has a similar skewedness I wanted to to investigate a deeper understand of the dataset for which it gave me a clear picture and understand of how the data set functioned. Primarily, we can see within the amount that the Sander's group tends to contribute more in the below hundred dollar range while Trump's group tends to follow a higher contribution pattern. Most of the distrobutions had okay patterns. They only real unusual visualization I saw was employment being change by person in the largest contributions by occupation graph.

## Bivariate Plots Sections

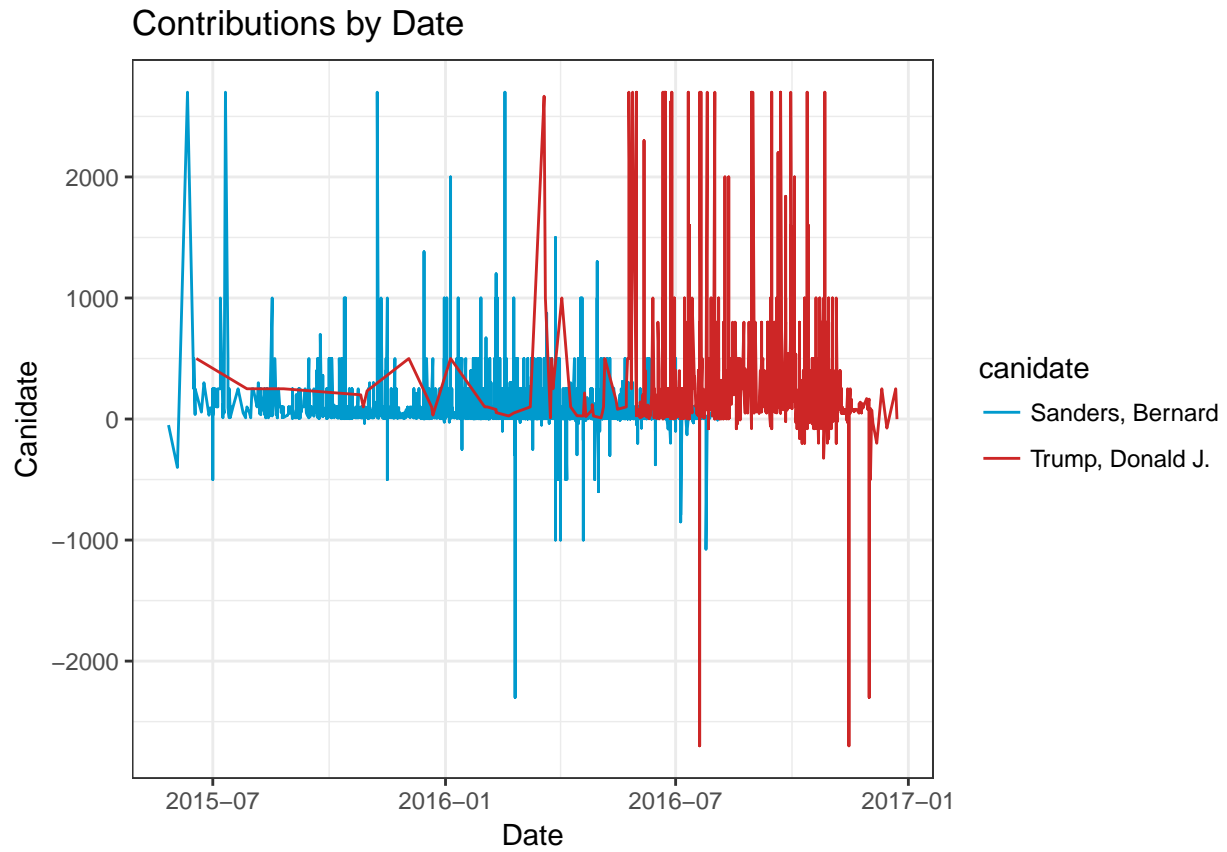
```
##           date                amount
##  Min.      :2015-05-04 00:00:00  Min.      : -10500.00
## 1st Qu.:2016-02-27 00:00:00 1st Qu.:    15.00
## Median :2016-04-06 00:00:00 Median :    27.00
## Mean   :2016-04-09 14:49:04 Mean    :    68.16
## 3rd Qu.:2016-05-31 00:00:00 3rd Qu.:    50.00
## Max.    :2016-12-31 00:00:00 Max.     : 10000.00
##           candidate
## Sanders, Bernard:407164
## Trump, Donald J.: 86258
##
```

```
##
##
##
```

Matrix plot of Date, Amount from Canidates.

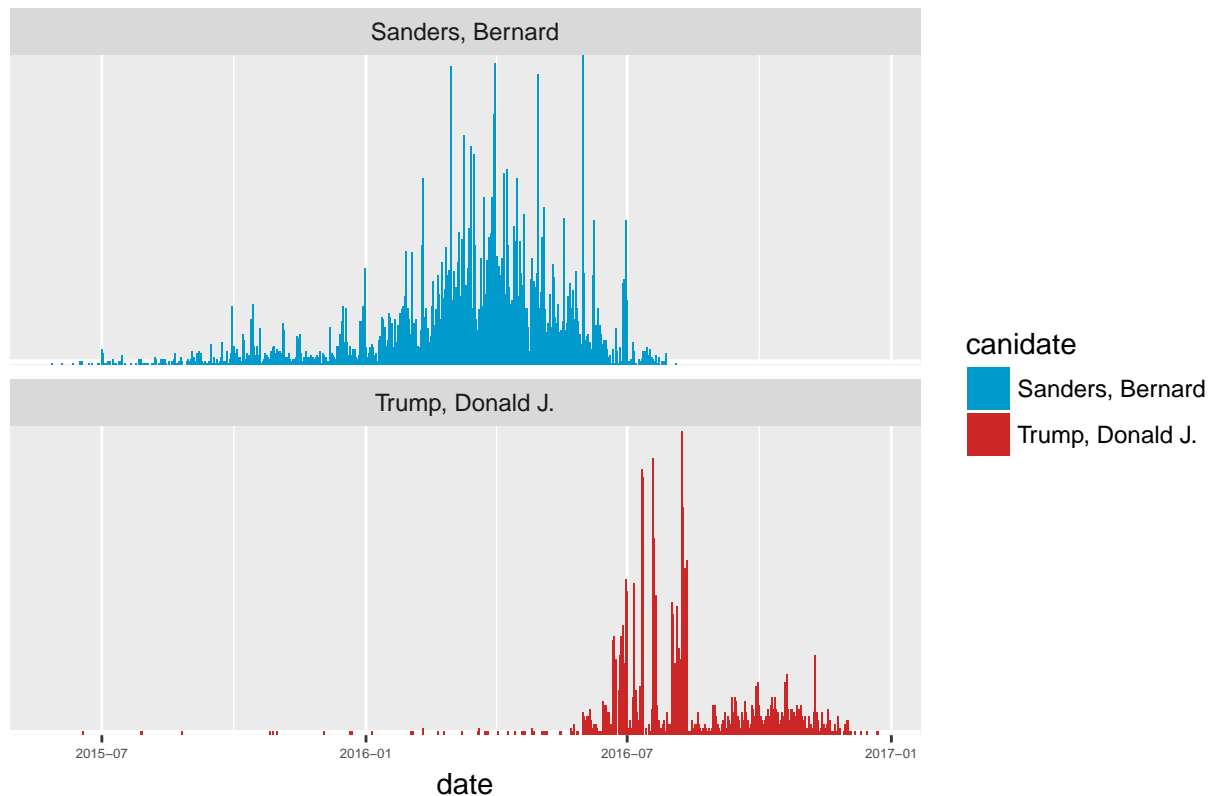


I created a subset of the data with amount, candidate and date. The problem I have in this dataset is you don't have real numerical analysis but primarily location and amount contributions. But, you still can see interesting results in this matrix plot for instance, the candidates by date you see some noramility in one of the candidates. Secondly, if you look at the date time between Jan and July you see a large growth in contributions, let's investigate this further.



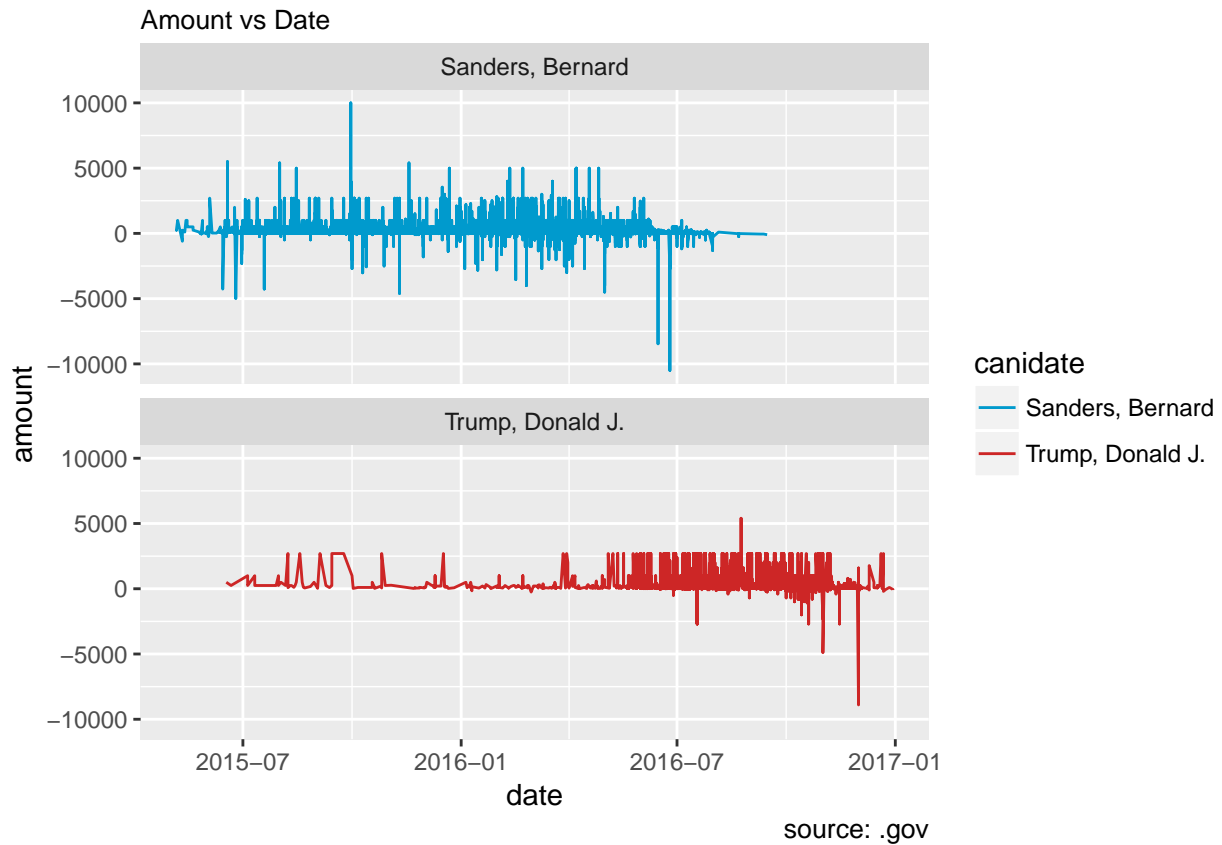
As the date moves into the election you see that donald Trump contributions are growing and Bernies are decreasing. This is due to the fact that Donald Trump went on to become the Repulican party canidate and Bernie Sanders had to drop out of the race. Let's look at this in a more normal distrubtion.

## Contributions by Date



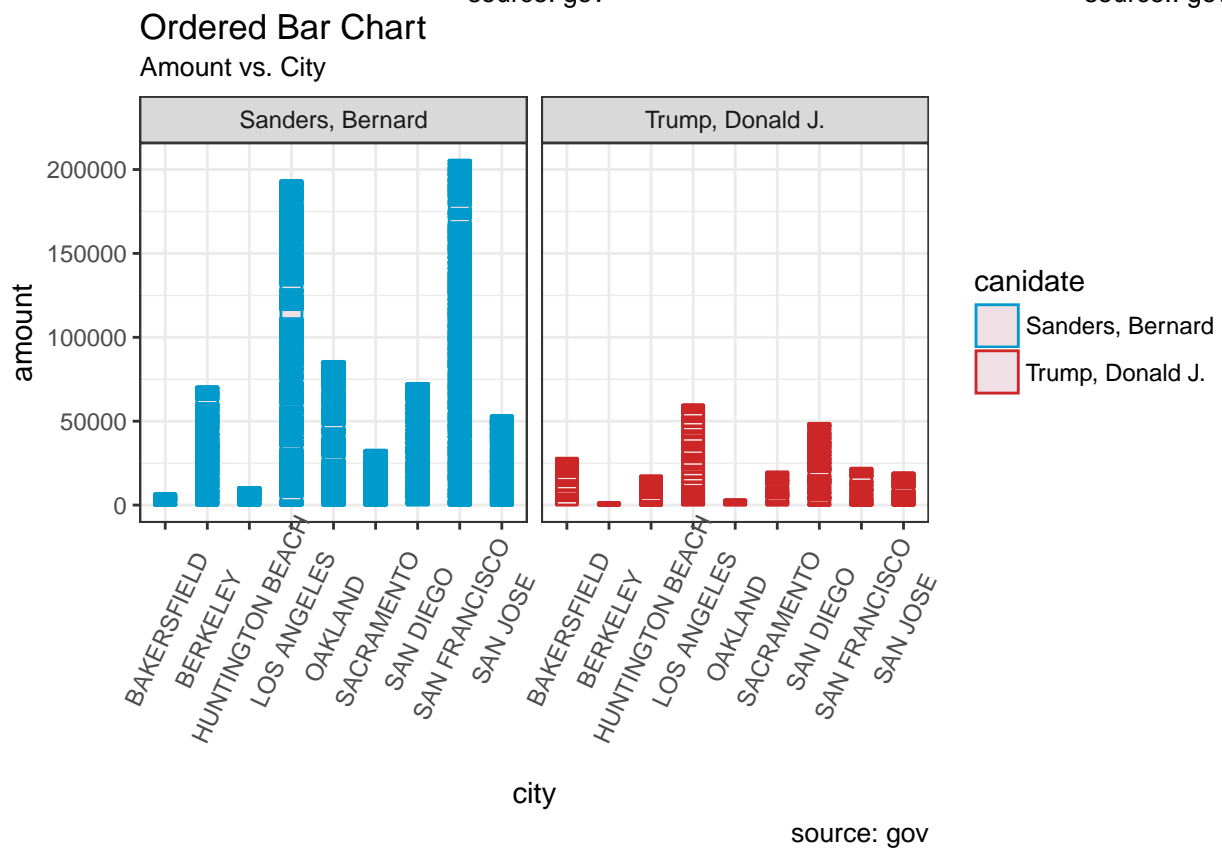
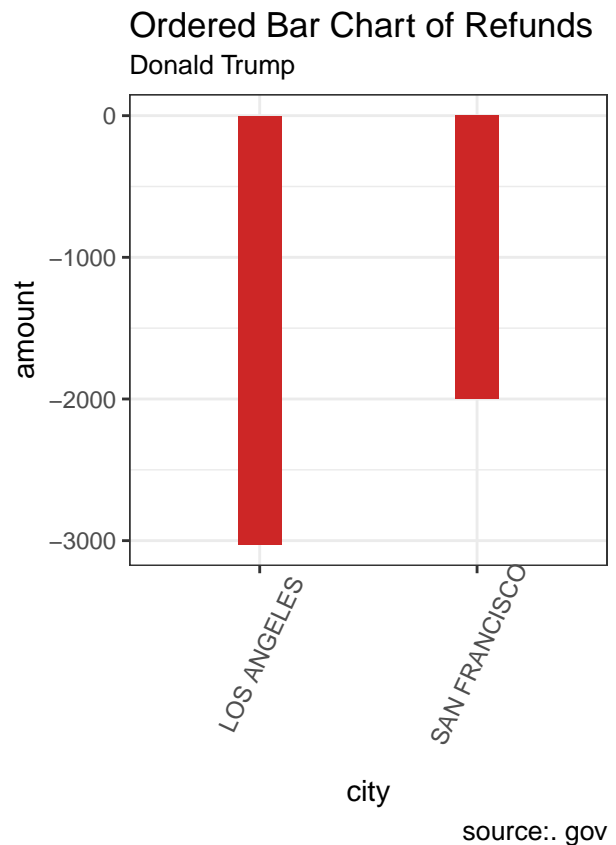
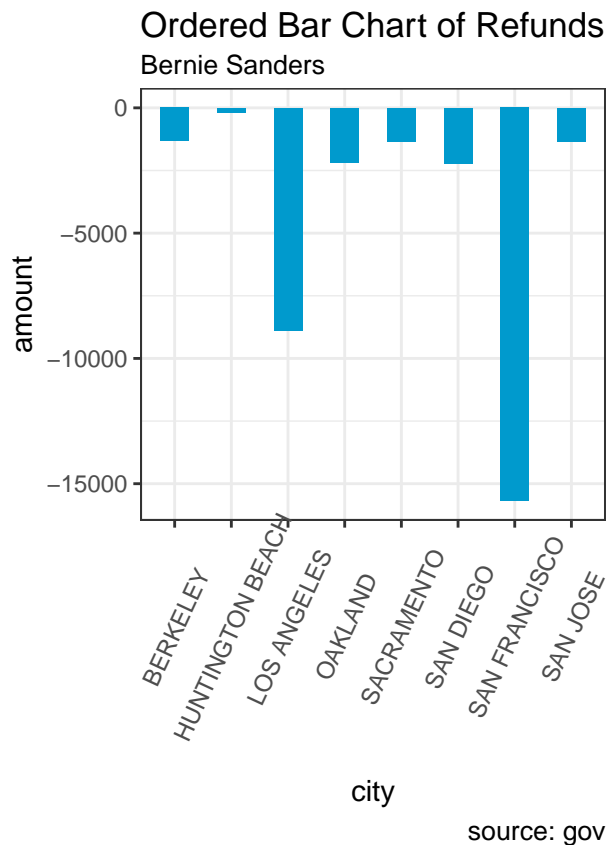
```
## merge_samp$canidate: Sanders, Bernard
##           Min.           1st Qu.           Median
## "2015-05-27 00:00:00" "2016-02-13 00:00:00" "2016-03-25 00:00:00"
##           Mean           3rd Qu.           Max.
## "2016-03-14 15:50:05" "2016-04-30 00:00:00" "2016-08-04 00:00:00"
## -----
## merge_samp$canidate: Trump, Donald J.
##           Min.           1st Qu.           Median
## "2015-06-18 00:00:00" "2016-07-06 00:00:00" "2016-08-02 00:00:00"
##           Mean           3rd Qu.           Max.
## "2016-08-07 09:56:08" "2016-09-12 00:00:00" "2016-12-23 00:00:00"
```

Clearly we see that Sanders' campaign is increasing as the months go on peacking around the third or forth month of 2016. We don't see large contributions in Donald Trump until the June mark. This could be due to teh fact that he was nominated aroundn this time or it's possible that California was still not interested in Donald Trump being the Republican canidate and still was hoping for a more moderate conservative like Jeb Bush or John Kasich. These plots do paint a very well detailed portrait of the canidates time running for president.



As you may view from above the trend of the line graph by date between Sanders and Trump. Some observations to view are Sander's data set is not a consistent as Trump's. Meaning, his data set tends to peak and fall between high donations and refunds, while Trump's data set seems to look more uniform in nature. Could this have something to do with anxiety in the election cycle? One observation to be noted is the negative amount in the y axis. Let's dive into this a bit closer.

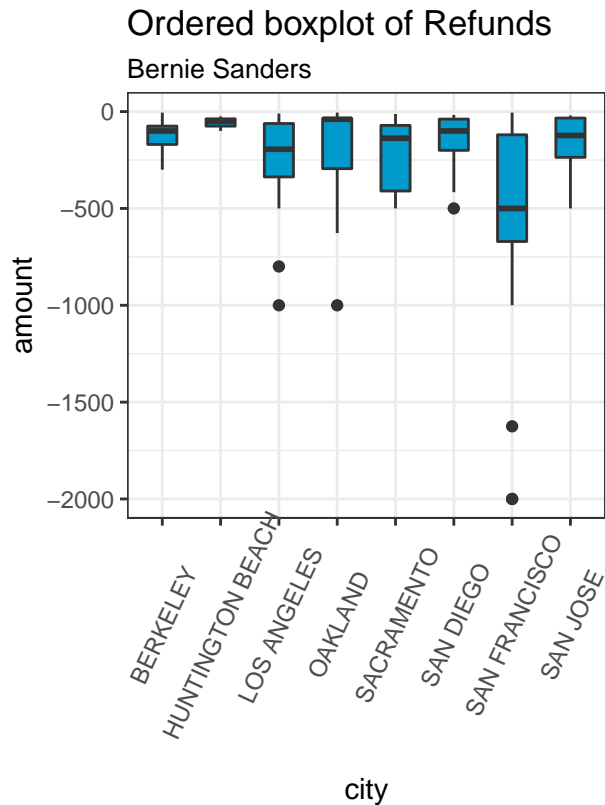




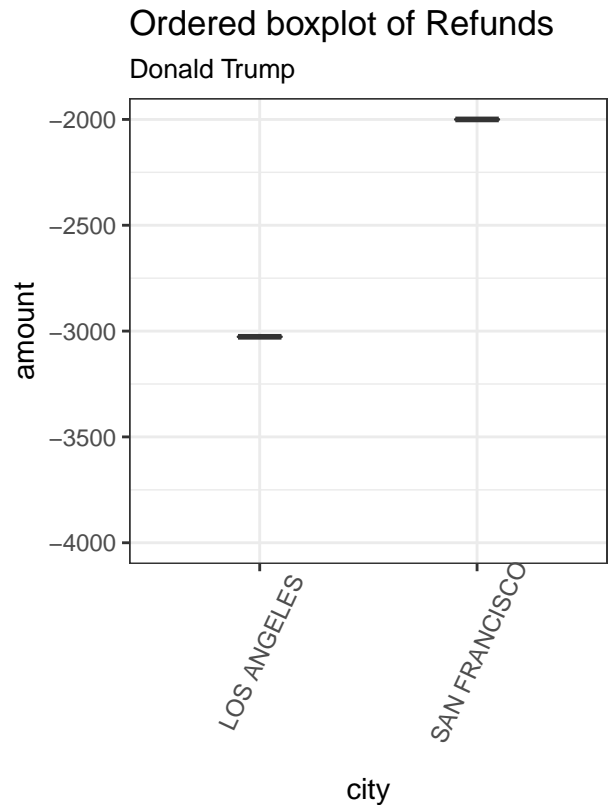
as you can see above most of the refund that were return were by Bernie Sanders campaign. While some

refunds were made by Donald Trump it seems that Bernie Sanders had the most. However, one thing to keep in mind is that Bernie's dataset is larger so it could account for just more datapoints in the plot. if we look at amounts we can see that Bernie sanders has more contributions in the postive direction as well. Let's look at these in more depth statistically.

```
## bernie_city$city: BERKELEY
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -300.00 -169.50 -100.00 -119.80  -75.00   -5.27
## -----
## bernie_city$city: HUNTINGTON BEACH
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -100.00  -75.00  -50.00  -58.33  -37.50  -25.00
## -----
## bernie_city$city: LOS ANGELES
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -2280.0  -362.4  -214.0  -329.1   -71.0   -10.0
## -----
## bernie_city$city: OAKLAND
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -1000.00 -294.60  -39.59  -244.00  -35.00   -5.00
## -----
## bernie_city$city: SACRAMENTO
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -500.00 -410.00 -137.50 -222.90  -71.25  -12.50
## -----
## bernie_city$city: SAN DIEGO
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -500.00 -200.00 -100.00 -158.70  -38.75  -17.50
## -----
## bernie_city$city: SAN FRANCISCO
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -2700.0  -757.8  -500.0  -653.2  -142.2   -5.0
## -----
## bernie_city$city: SAN JOSE
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -500.00 -236.20 -123.00 -171.80  -33.36  -20.23
## -----
## trump_city$city: LOS ANGELES
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   -3027   -3027   -3027   -3027   -3027   -3027
## -----
## trump_city$city: SAN FRANCISCO
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   -2000   -2000   -2000   -2000   -2000   -2000
```



source: gov



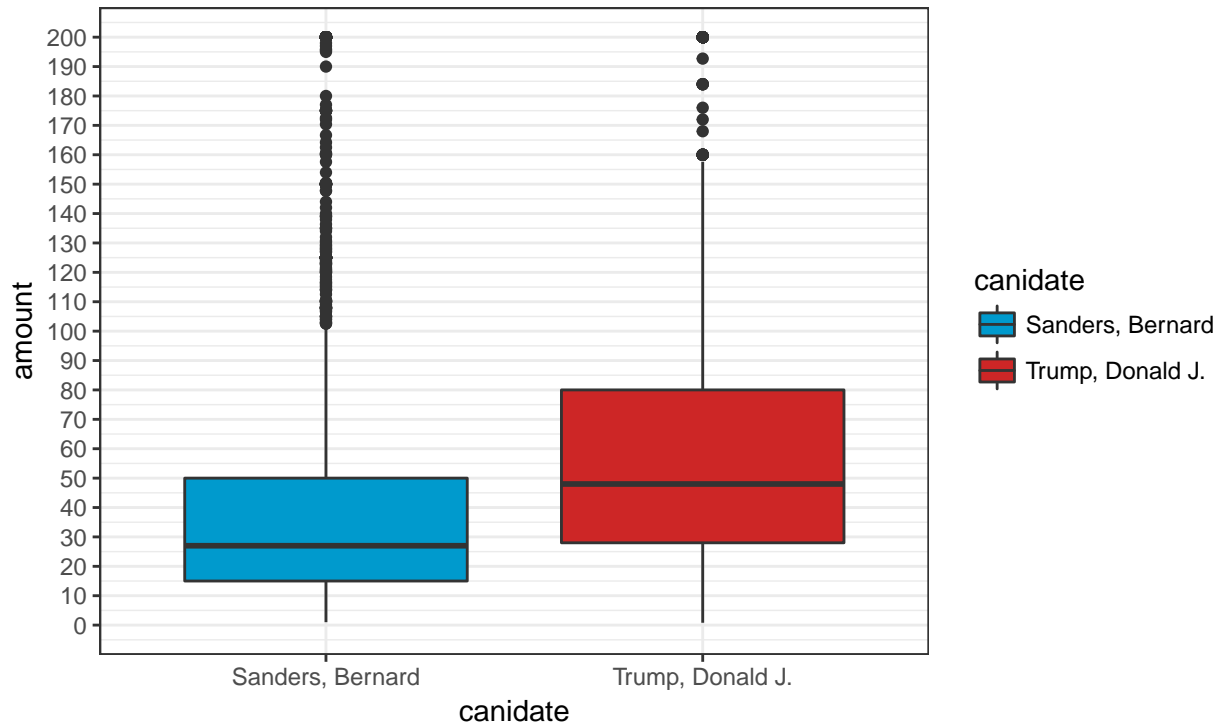
source: . gov

```
## largest_amount_bernietrump$city: BAKERSFIELD
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.0   20.0   37.5   130.0   100.0  2700.0
## -----
## largest_amount_bernietrump$city: BERKELEY
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00   15.00   27.00   60.17   50.00  2600.00
## -----
## largest_amount_bernietrump$city: HUNTINGTON BEACH
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.0    15.0   35.0    89.9   100.0  2400.0
## -----
## largest_amount_bernietrump$city: LOS ANGELES
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.80   15.00   27.00   67.72   50.00  5400.00
## -----
## largest_amount_bernietrump$city: OAKLAND
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00   15.00   27.00   61.11   50.00  2700.00
## -----
## largest_amount_bernietrump$city: SACRAMENTO
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00   15.00   27.00   58.78   50.00  2000.00
## -----
## largest_amount_bernietrump$city: SAN DIEGO
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00   15.00   27.00   61.76   50.00  2700.00
```

```
## -----
## largest_amount_bernietrump$city: SAN FRANCISCO
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.00  15.00   27.00   64.55  50.00 2700.00
## -----
## largest_amount_bernietrump$city: SAN JOSE
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.00  15.00   27.00   55.96  50.00 2000.00
```

## Ordered Box Plot

Amount vs. City



source: .gov

```
##
## Pearson's product-moment correlation
##
## data: merge_bs_dt_df$zip and merge_bs_dt_df$amount
## t = -150.4, df = 493420, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2120382 -0.2067023
## sample estimates:
##           cor
## -0.2093718
```

We can now see that most of the the regunds fall into a median point for Sanders less then negative five hundred dollars. While, the Trump team seems to have small strips. If we view the postive amount, meaning the contributions, we see that by city the median rate is in twenty-seven mark for Bernie Sanders (which was a campaign footnote) while Trump's falls in the forty-nine to fifty range.

## Bivariate Analysis

**Talk about some of the relationships you observed in this part of the investigations.**

How did the features of interest vary with other features in the dataset.

the only true correlation I could do on this dataset was amount and zip\_code using pearson and it found no real true correlation (something closer to zero) or around twenty percent negative correlation. Which means it the price of contributions is pretty all across the board in each zip code. I must stress again that the datasets that were both gathered on the government website are skewed in size, Bernie's being higher, but taking that into account there is a strong notion that Bernie's footnote campaign talk about the average american spending roughly twenty-seven dollars in this contributions seems to be right on the money. However, Sanders did not explain the large amount of refunds he was getting from contributors. Trump's, while low, seems to follow a normal uniform pattern in campaign contributions which just shows he's steadily reaching the finish line.

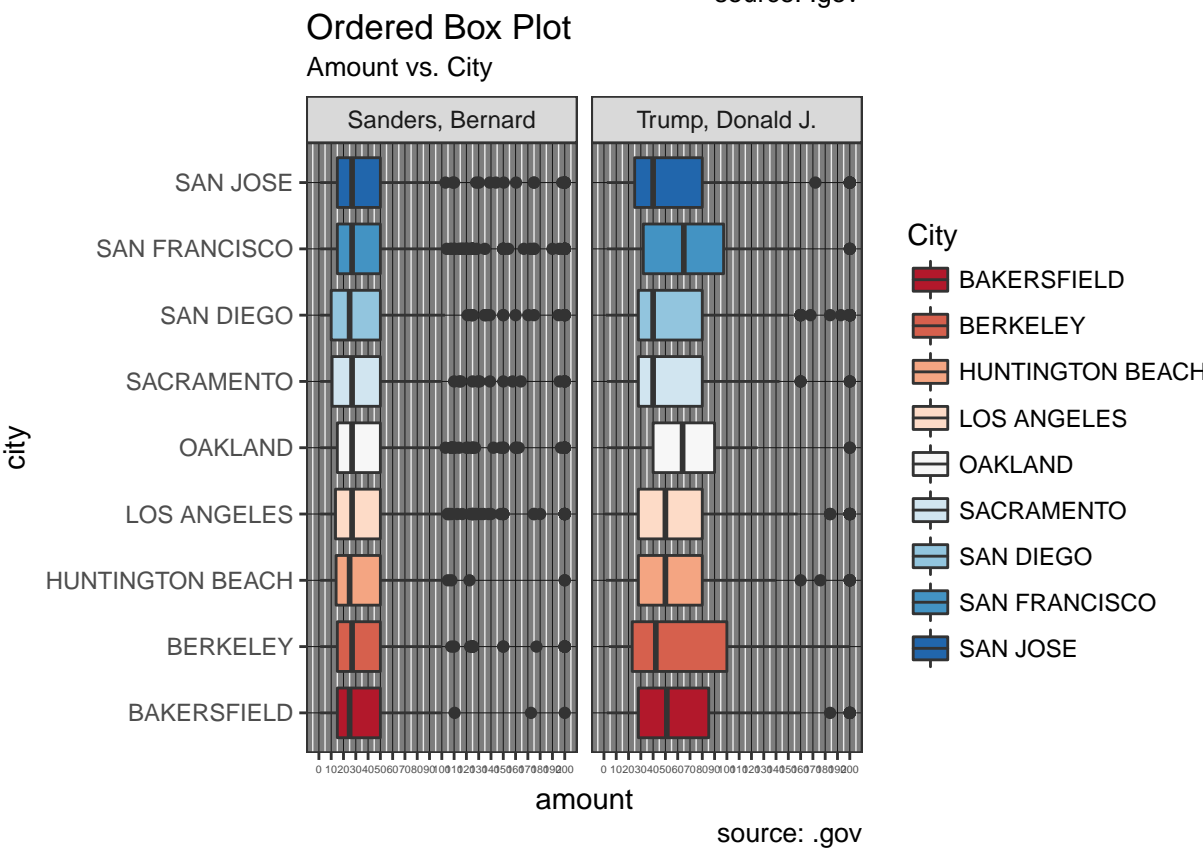
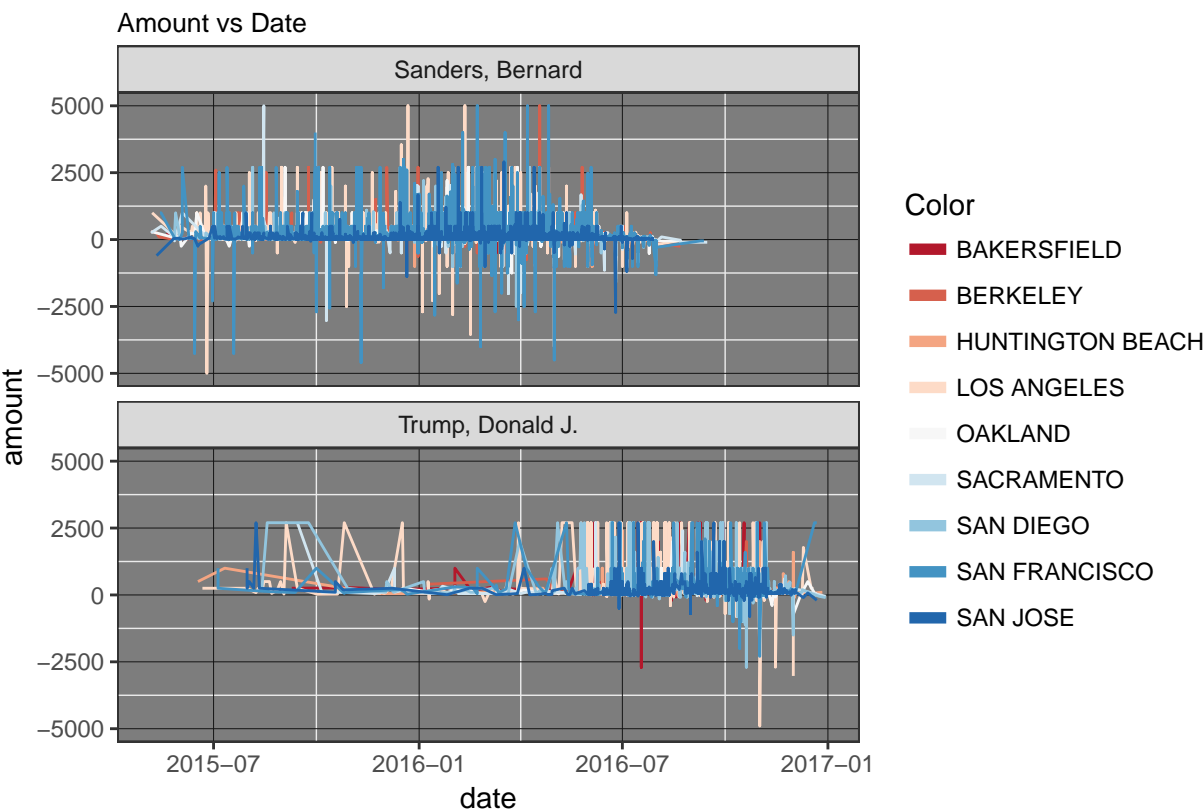
**Did you observe any interesting relationships between the other features?**

There isn't really any true relationship across the board in this dataset it really just outlines that the Sanders campaign had many contributors who seemed to be up and down along the campaign while Trump's supports were steady and consistent.

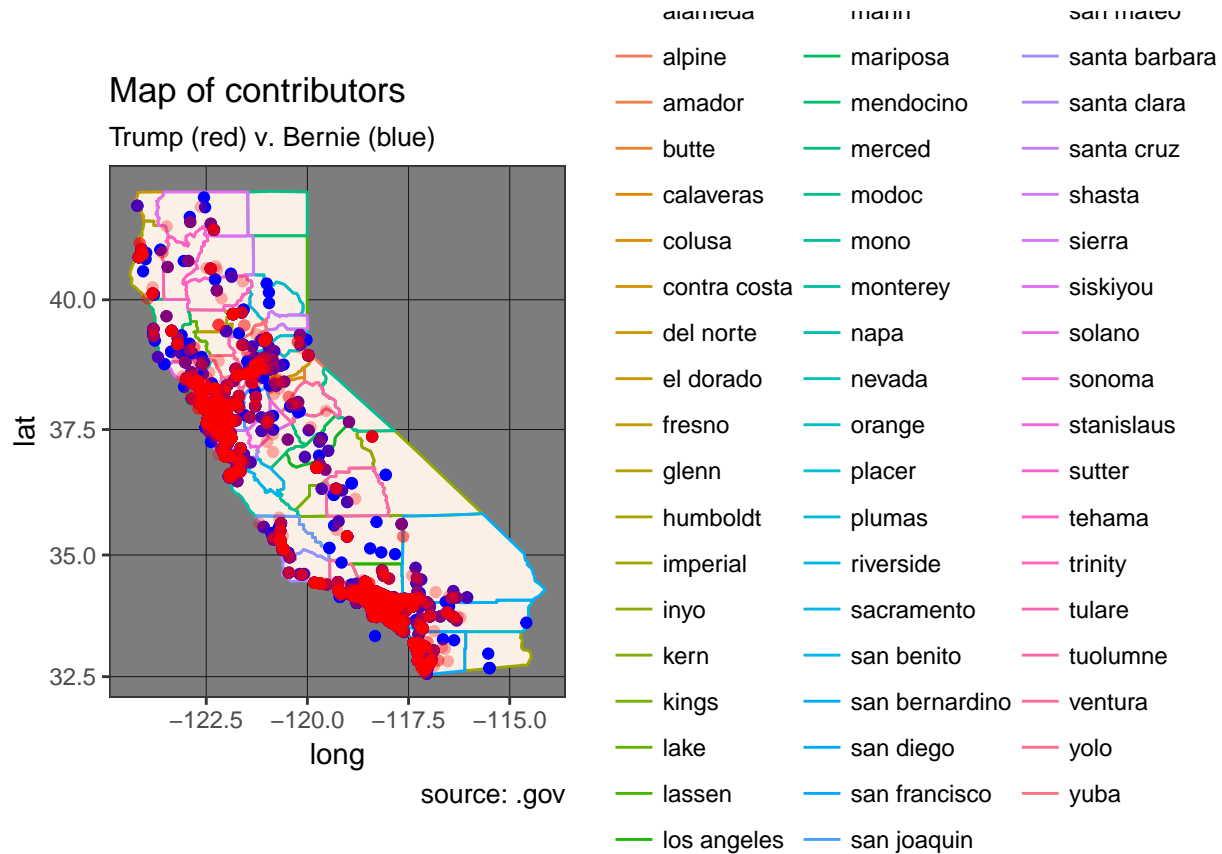
**What is the strongest data point you found?**

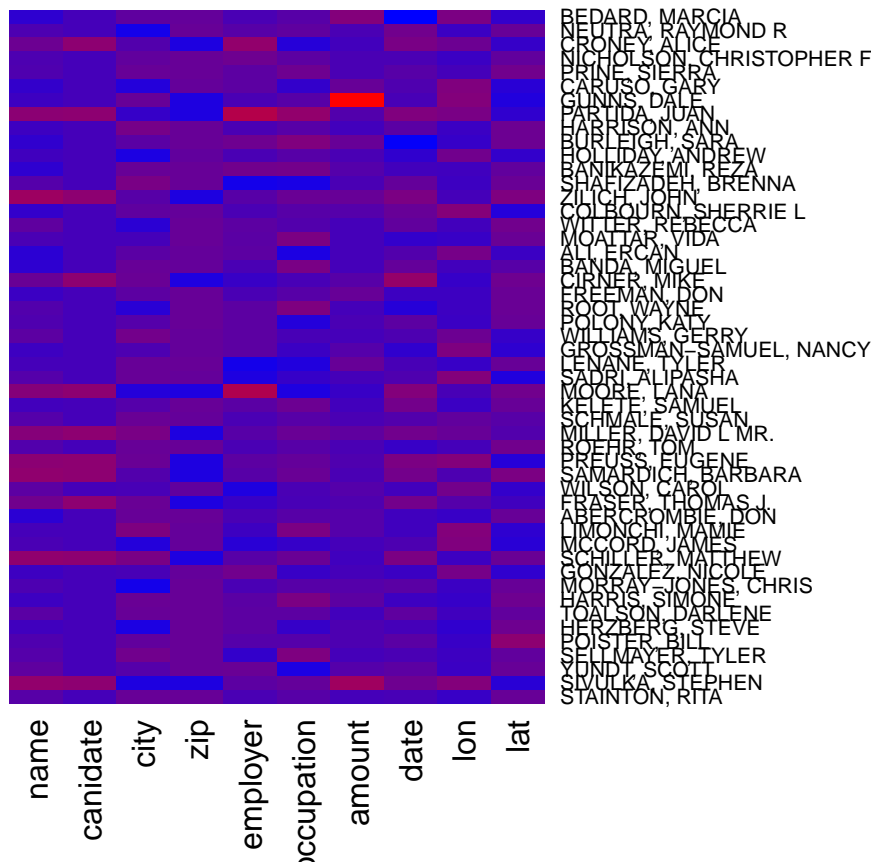
the strongest thing I found was Bernie Sanders footnote of twenty-seven dollars. I thought that was pretty wonderful.

# Multivariate Plots Section and Summary



These last two plots show the relationship that the candidates have by major cities of the date of time and the amount that is spent. As you can see clearly Sanders contributions in these states are more sporadic as the months go on while Trump's tend to fall in a more uniform pattern (this is not to say that this is either good or bad). While, in the box plot we see a clearer picture of amounts received where Sanders falls into the median range of around \$27 dollars while Trump's median range per city now seems to have more sporadic scaling. I personally believe the box plot shows a clear image of what's going on in these cities.





These last two graphs show the overall range of where people are contributing. As you have noticed in the map you see a lot of red within the San Diego region and more across the ocean. However, these regions do have some dots for Bernie. Purple represents counties and cities where you have equal amounts of contributions. However, this is more in-depth with a sample size of 50 contributors with the heatmap. Within the heatmap you see most of the areas are converging to purple with some leaning blue and other leaning more red. This gives us a good representation of how the overall distribution of how voters were contributing in California.

## Multivariate Analysis

**Were there features that strengthened each other in terms of looking at your features of interest.**

Essentially, what the multivariable analysis taught me is there really isn't a left-leaning or right-leaning contribution within these two data sets. Primarily, most of the contributions seem to be equal in strength. While, Bernie's dataset is larger than Trump's, the considerable spread across the state between Democratic or Republican seems to be pretty equal here and has no direct correlation between one and another.

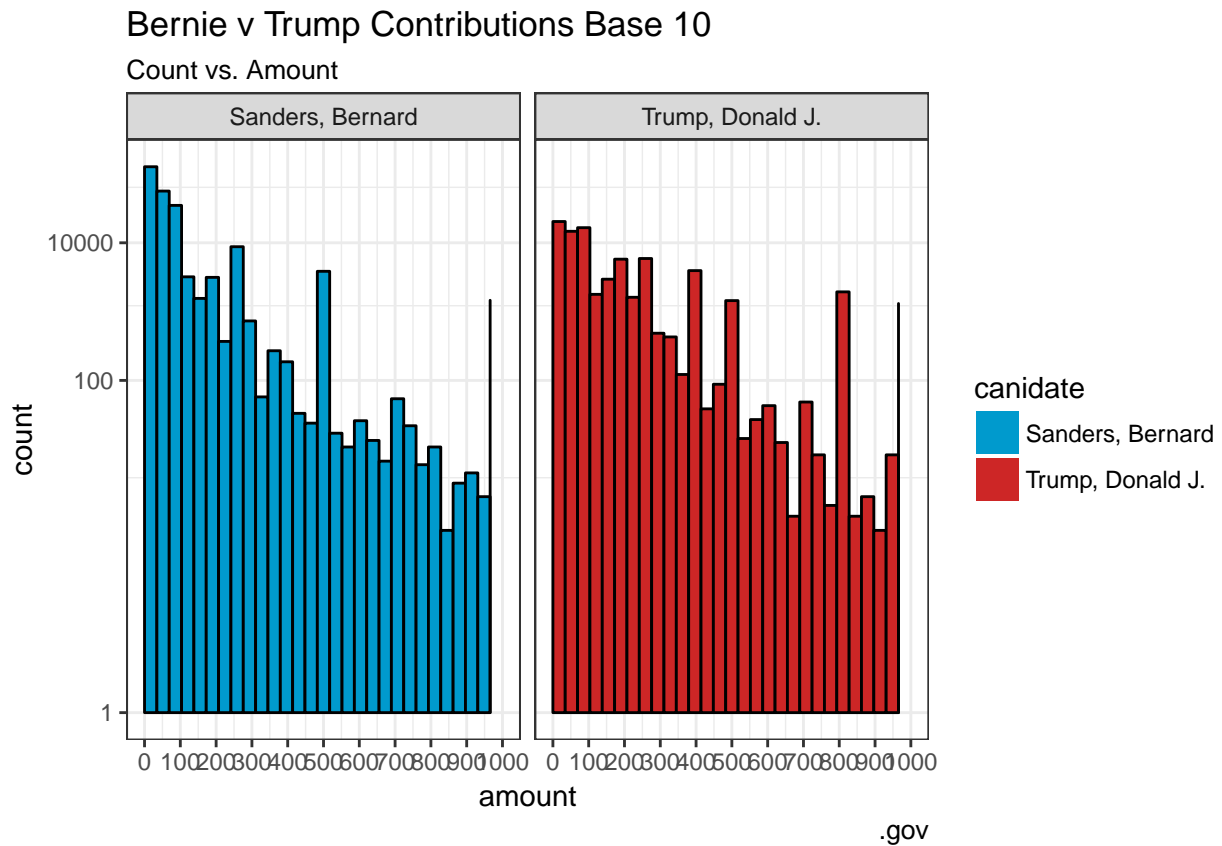
**Were there any interesting or surprising interactions between features?**

I think the most surprising thing is that it's NOT more blue in the data sets and it's roughly spread fairly equally. I would have thought more people to support Bernie Sanders way more than Donald Trump and while they did support Sanders more I thought it would be more extreme.



## Final Plots and Summary

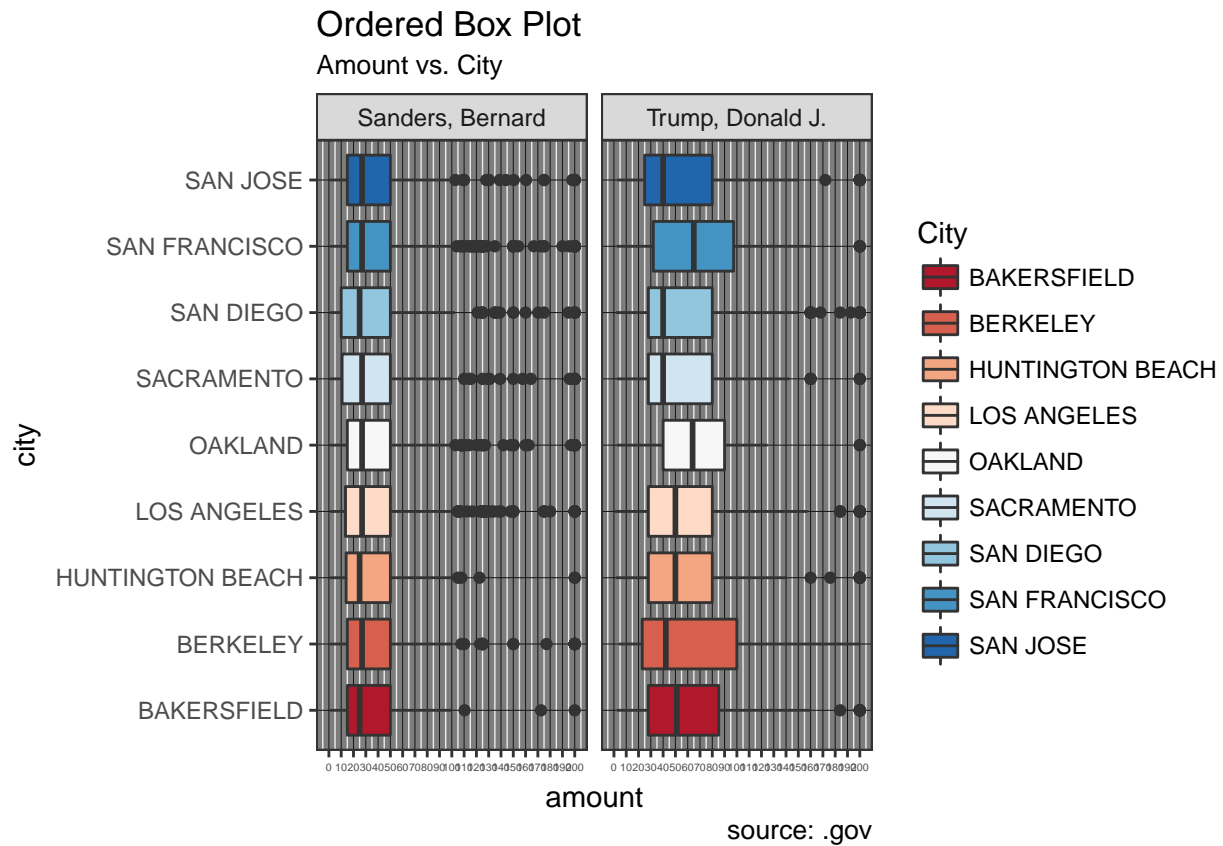
### Plot One



### Description One

The distribtuion of amount by count is on a log scale and is still skewed. However, this does give us a first impression of equal distribtuion between Sanders and Trump. Sanders, falling higher in scale below the <100 mark and Trump falling higher in the =800 mark.

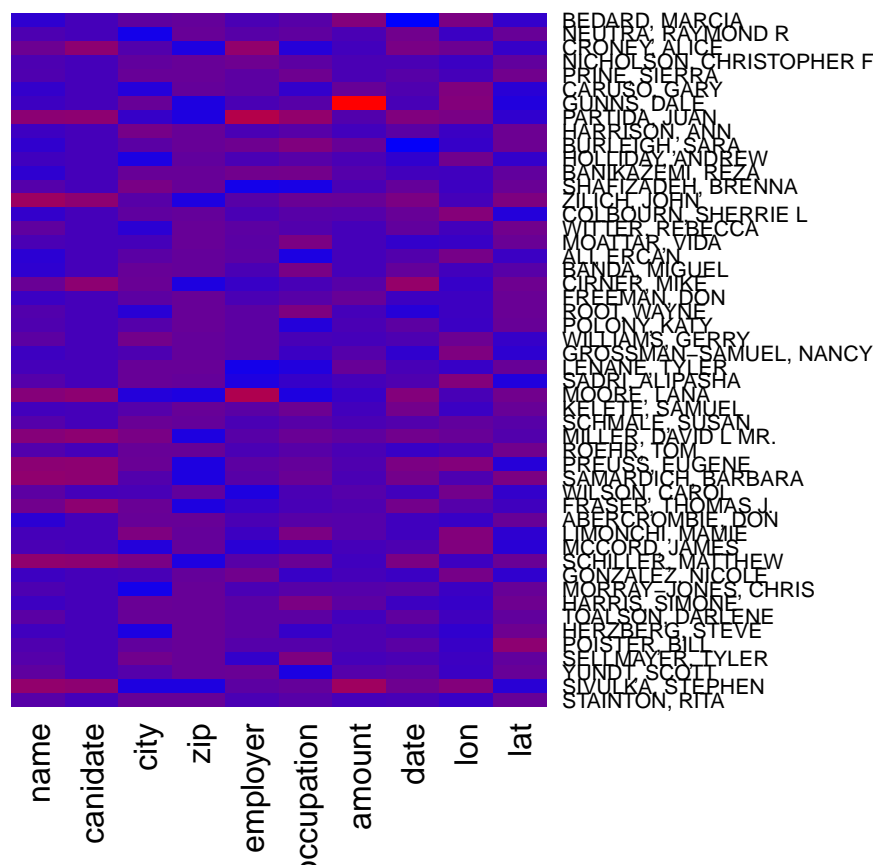
## Plot two



## Description two

This plot indicates how major cities contributed to the campaign. As you can see sanders has a median of roughly 27 dollars while Trump hwas a median around 50 dollars depending on the city. This shows a more uniform steady contribution for the sanders campaign while Trump's is more all over the place. However, it looks to me like larger contributions were made for trump within the third quartile, while Sanders has heavier outliers.

## Plot three



## Description three

This heat map shows a sample of 50 contributors for Bernie and Trujmp which shows a realtive equal pattern between contributions. While some are more left leaning and dmore right a majority fall into the purple category.

## Reflection

This report explores a dataset containing canidates and attributes for approximately 490,000 contributions in California during the 2016 election for President of the United States. Most of the explination within the data set was done with amount, date, canidate, occupation, name and location. I wanted to explore if certain regions were more impacted by Sander's campaign or Trump's campaign. And my conclusion would be that, yes, there were more contributions made with the Sanders campaign, However, I was surprised to see that there were still contributions made for the Trump campaign as well.

When I explored the data set the first thing that suprisred me was the amoutn of refunds that were carried about by the Sander's campaign, one of the largest be -10,500 dollars. While I know there are alot of rules pertaining to campaign contributions I wonder if it was a true rerfund or something that wasn't done properly based on California contribtuiion rules.

Sander's campaign median level of donation fell around the 27 dollar mark, while Donald Trump's campaign fell around the 50 dollar mark. I was pleased to see that Sanders was correct in using his footnote for his speeches. The thing that struck me the most about these contributions is that a large majority were done by

people who were unemployed, which makes me question, do people get more involved in election cycles when they are unemployed? Do people tend to not care about contributions when they are in fact employed? Does the upper middle class seem to not care about elections because they are comfortable with their stage in life? It's an interesting analysis to approach.

Finally, I believe doing an analysis of a more red state and comparing Sander's to Trump would be interesting, or a swing state. However, I choose this state because it is in fact where I live.