

## Exploration of Financial Contributions to United States Presidential Campaign of 2016 by the State of New Jersey by Madhu Shri Rajagopalan

### Data Loading and Cleaning

The data set to be loaded contains all the financial contributions to the united states presedential campaign of 2016 by the State of New Jersey.

This dataset contains 203,883 observations and 18 variables.

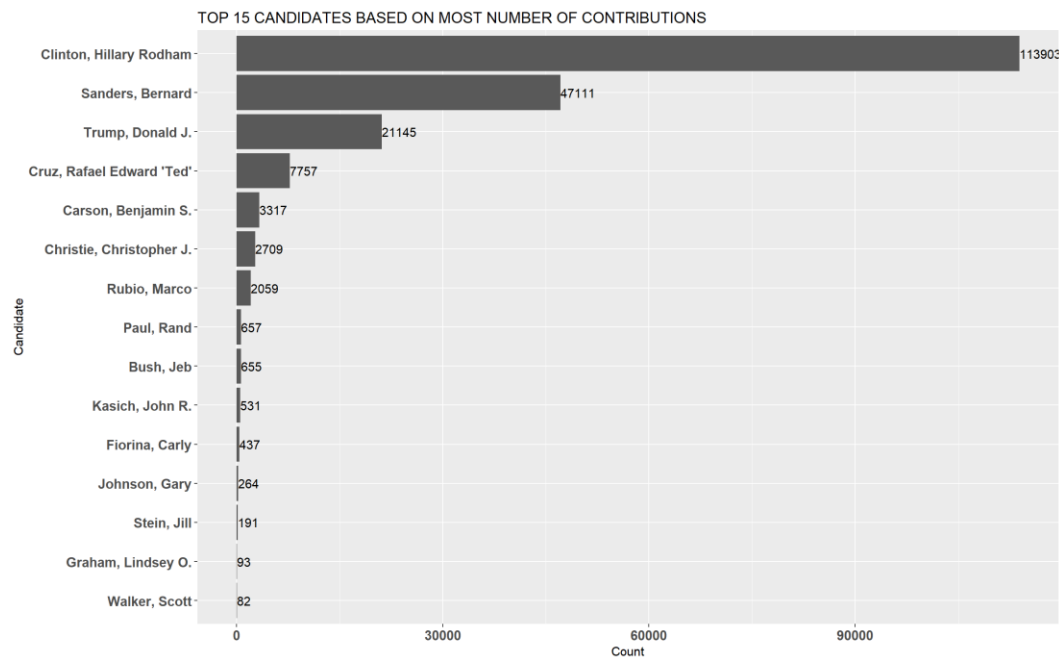
From the data summary, the minimum contribution is -7500 and the maximum is 10800.00 in terms of dollars. The negative contribution looks like a refund. When looking further into this, I found from Federal Election Commision website [FEC] ([www.fec.gov/help-candidates-and-committees/candidate-taking-receipts/contribution-limits-candidates/](http://www.fec.gov/help-candidates-and-committees/candidate-taking-receipts/contribution-limits-candidates/)) that the individual contribution limit is at 2,700\$ per individual and any other additional amount contributed will be refunded. To proceed further with the analysis, I'm filtering the data to include only the data that has a cotribution limit over zero and under 2700 which is the limit.

After filtering, there are 201166 observations and 18 variables.

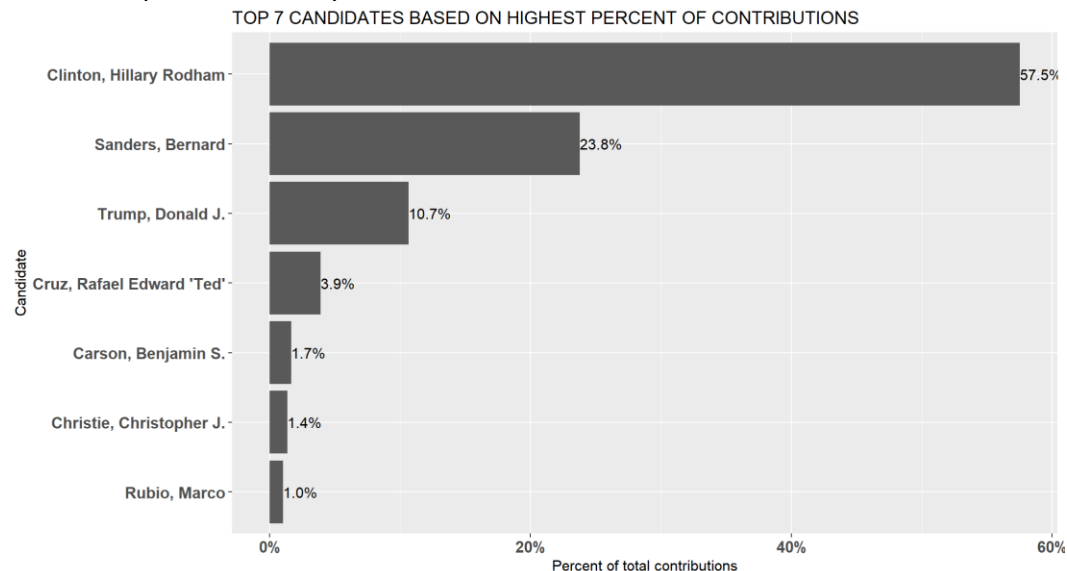
Before proceeding with further analysis and plotting, I would like to add variables to the dataset that would help with my analysis. The variables I would like to explore with are the Party (Democratic,Republican, Green, Liberatarian, Independent). gender (male/female), year and month variable, latitude and longitude for exploring contribution distributions over the map of state of New Jersey.

## Univariate Plots Section

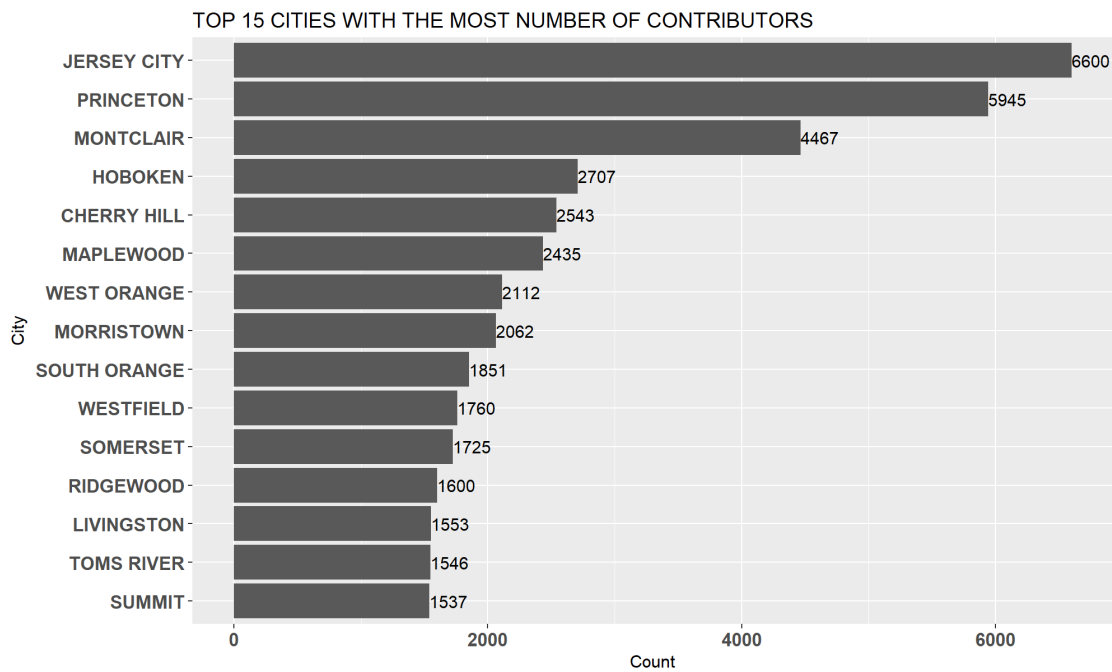
Now that the data set is pretty much clean and with all the variables needed for my exploration, I'm starting to plot the univariate variables to get an idea of the focus of the exploration in further steps.



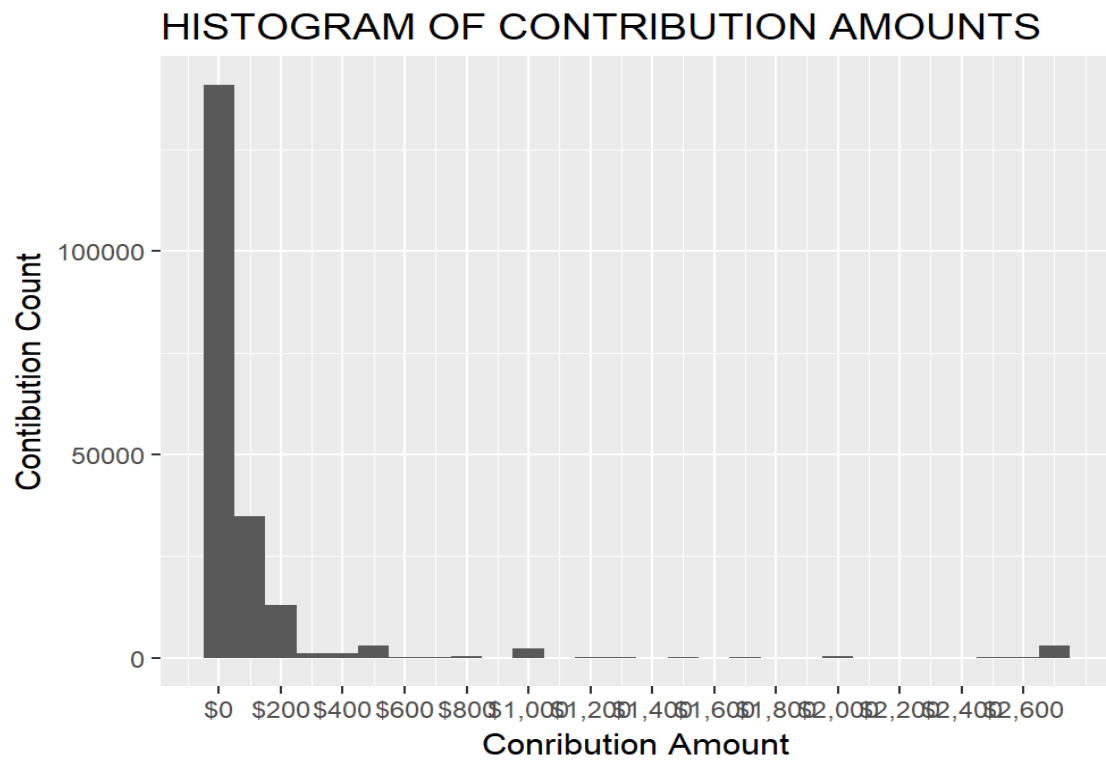
Hillary Clinton has got the most number of contributions from 113903 contributors followed by Bernie Sanders and Donald Trump although their contributor counts are pretty much very less when compared to Hillary.



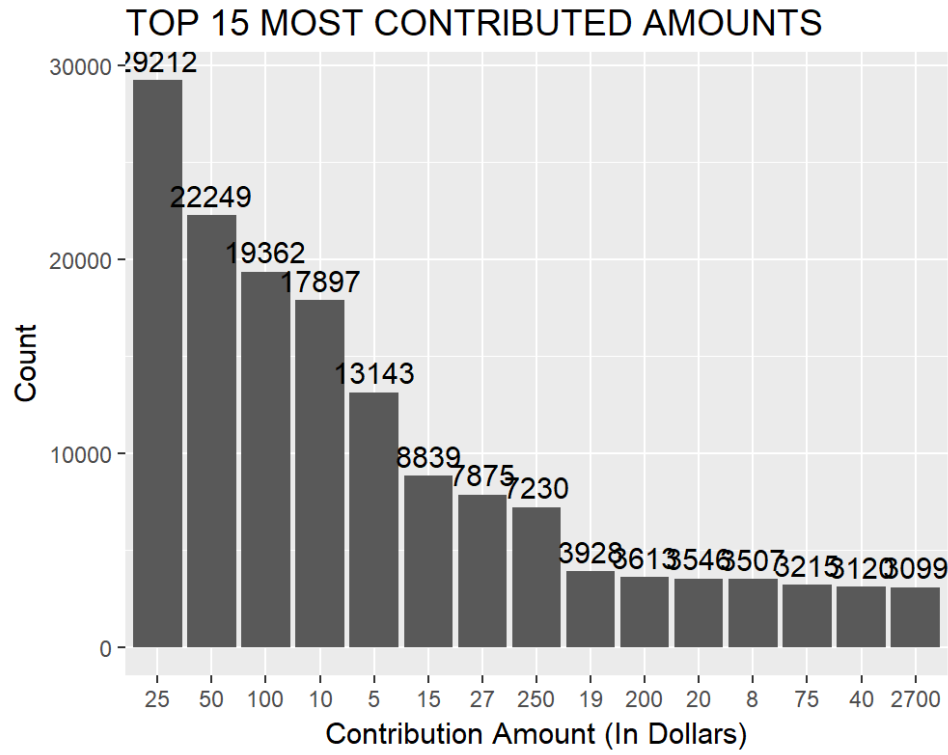
In terms of percetbage as well, Hilary clinton leads way ahead than other candidates with more than 50 percent than the other top candidates.



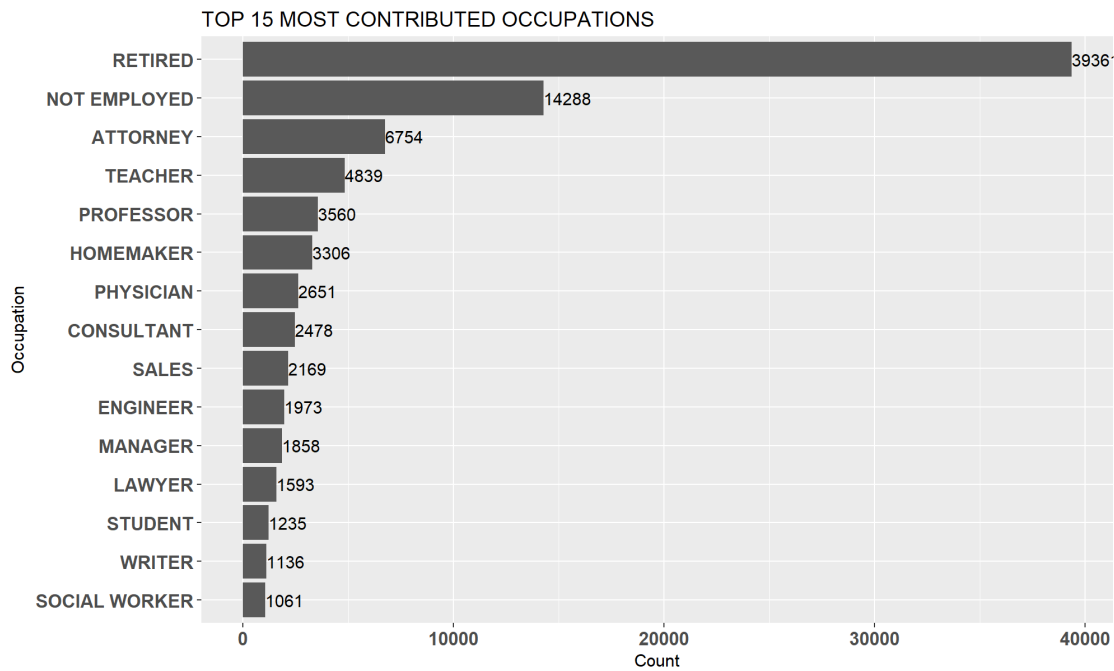
Jersey City has made the most number of contributions followed by Princeton.



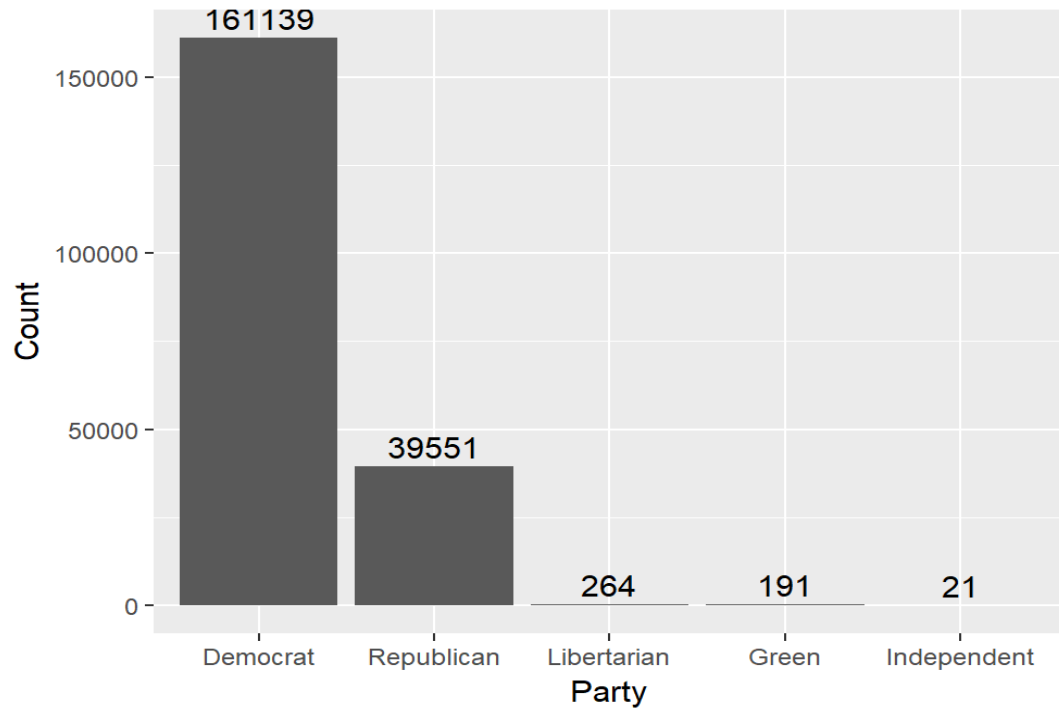
Most of the contributions made are less than \$250.00. Looks like the top contribution amounts were lesser than 50 dollars.



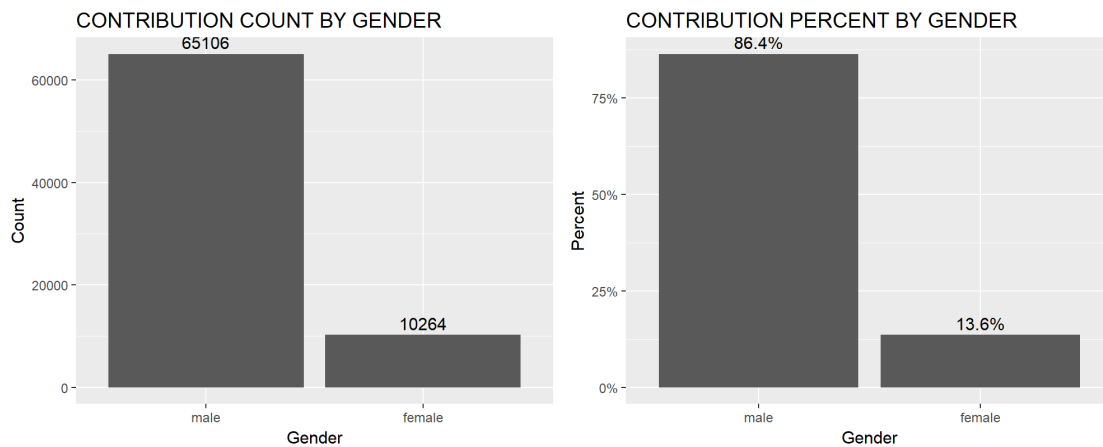
25 and 50 dollars are the top two contribution amounts made by most of the contributors.



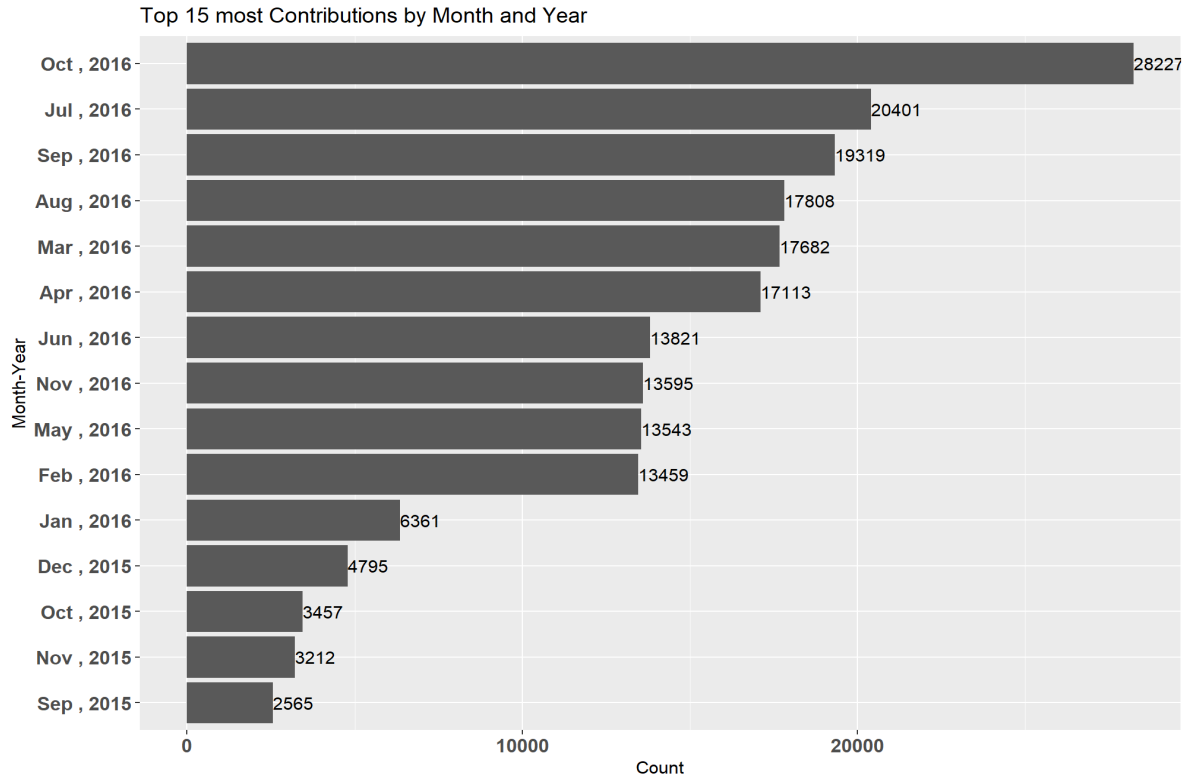
Top contributors are retired who made the most number of contributions compared with other occupations.



Democratic party has got the most number of contributions followed by the Republican party although the difference is huge.



Most of the contributions were made by males. with the available gender data, only 13% of the contributors were females which is significantly lower percentage.



October 2016, tops the most number of contributions made by months. It is the month closer to election date in November and second followed by July 2016, the month after the primaries and caucuses ended and the month, the nominees for the parties were decided.

## SUMMARY OF UNIVARIATE ANALYSIS

### Structure of the data set

After removing the negative contributions as well as the contributions exceeding the limit and adding new variables that will be helpful in the analysis, the structure of the data set is 201166 observations with 28 variables.

### Main features of interest in the dataset

From univariate plotting, I made the following observations

1. Hillary Clinton has got the most number of contributions followed by Bernie Sanders and Donald Trump.
2. Jersey City has made the most number of contributions followed by Princeton.
3. Most of the contributions made are less than \$250.00 with 25 and 50 dollars being the top two contribution amounts made by most of the contributors.
4. Top contributors are retired who made the most number of contributions compared with other occupations.

5. Democratic party has got the most number of contributions followed by the Republican party although the difference is huge.
6. Most of the contributions were made by males. with the available gender data, only 15% of the contributors were females.
7. Most number of contributions were made in October 2016, the month closer to election date followed by July 2016, the month the primaries and caucuses ended and the nominees for the parties were decided.

### Other features in the dataset that support further analysis

I'm more interested in exploring the contribution amounts with the variables. Does more number of contribution mean more donation? In which month did the contribution amount received was highest?

### Adding New variables from existing variables in the dataset

Since, the data is related with Presidential campaign, it is necessary to analyze it based on the party which I created with the Candidates name from the dataset as below.

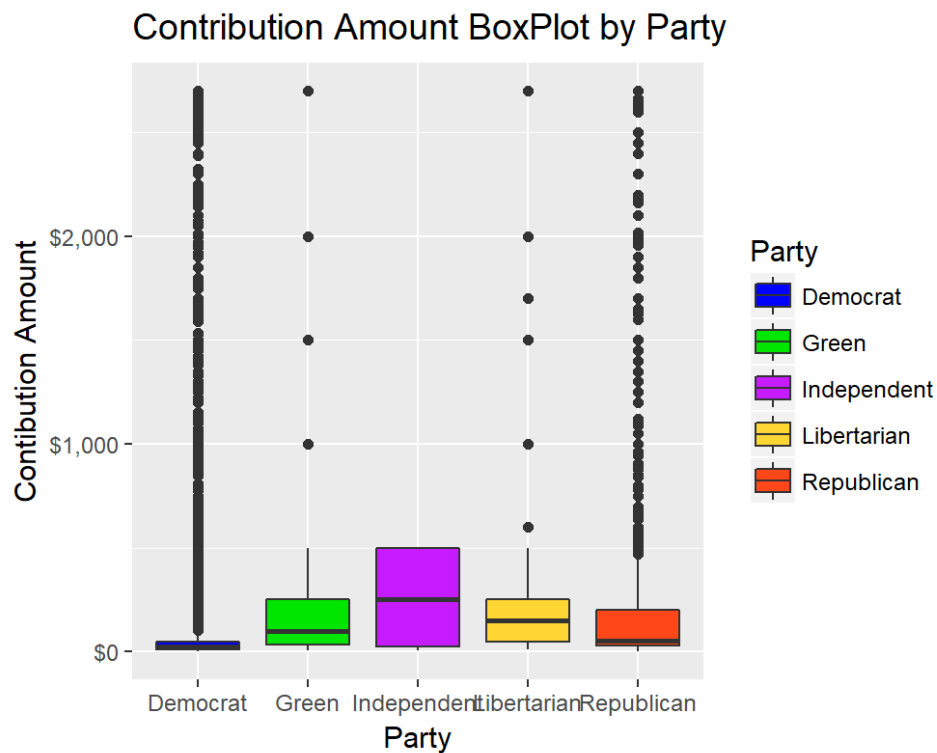
Democratic Party - Hillary Clinton, Bernie Sanders, Martin O'Malley, Lawrence Lessig, James Webb  
Libertarian Party - Gary Johnson  
Green Party - Jill Stein  
Independent Party - Evan McMullin

Also, since the exploration is for the state of New Jersey, I added the Latitude and Longitude variables from the Zipcode.

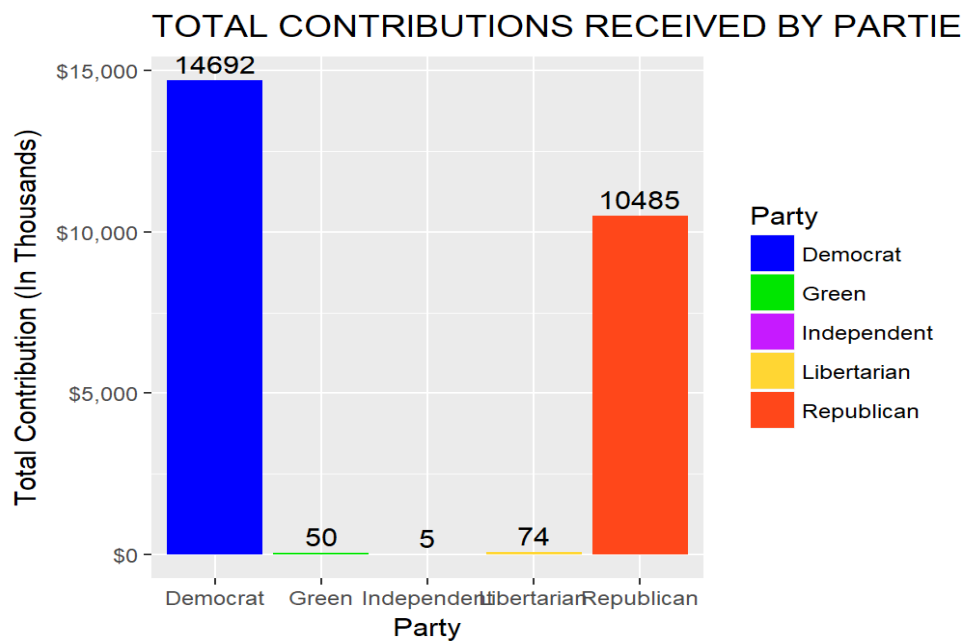
I was interested to see if woman candidates got more contributions from females. So I added the gender using the gender package. This helped to get just an idea of the distribution since there were names in the dataset those were not actual names like "A" or "4asted"

Finally from the date variable, I wanted the month and year variable separately to be helpful in plotting and also to see the contribution distribution in accordance with the election date which was November 8 for the year 2016.

## Bivariate Plots Section

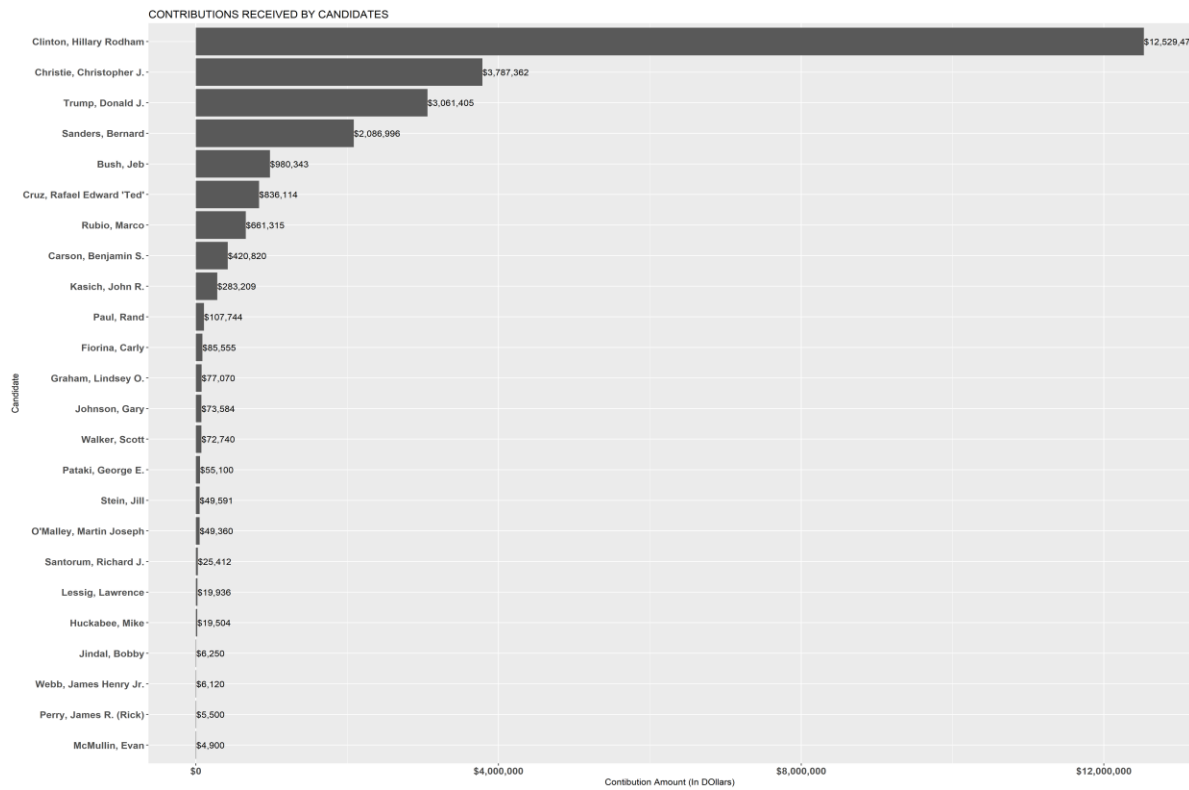


From the Box Plot, looks like the Democratic party has more number of outliers with a median contribution of \$25.00 which is lesser than the median contribution of the Republican Party which is 53.00 dollars. Democratic party has received more number of large contribution amounts from the contributors than small amount when compared with the Republican party.

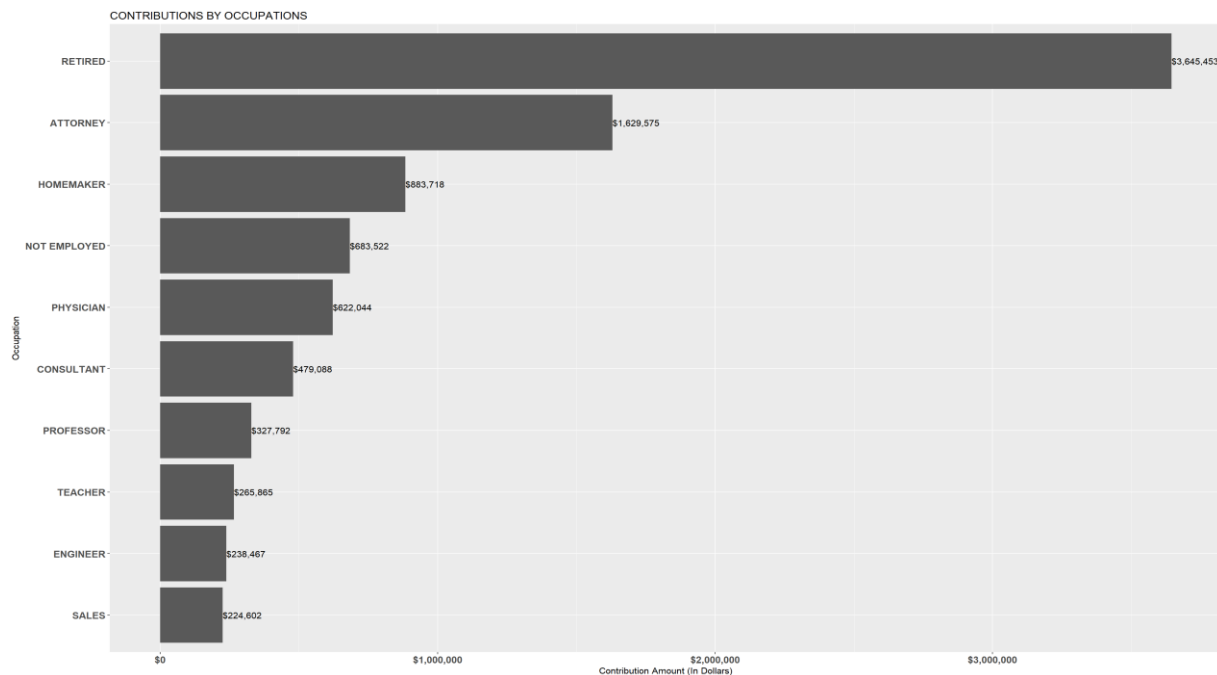




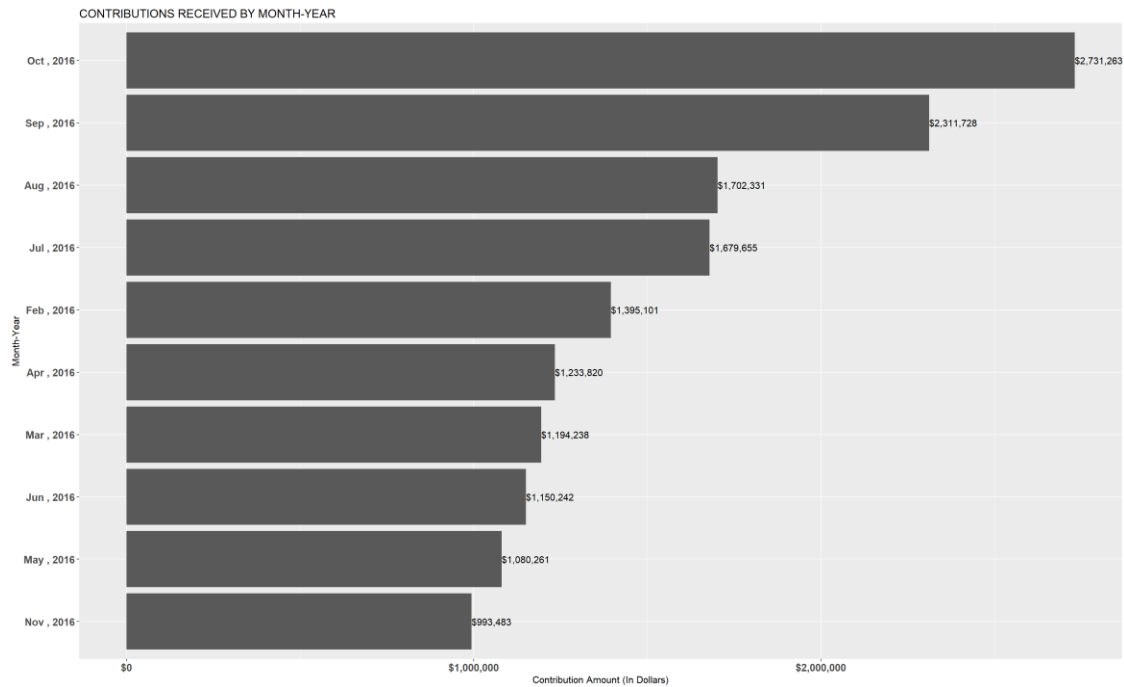
The Democratic Party has recieved the highest contribution amount of \$146,92,000 followed by the Republican Party with 104,85,000.



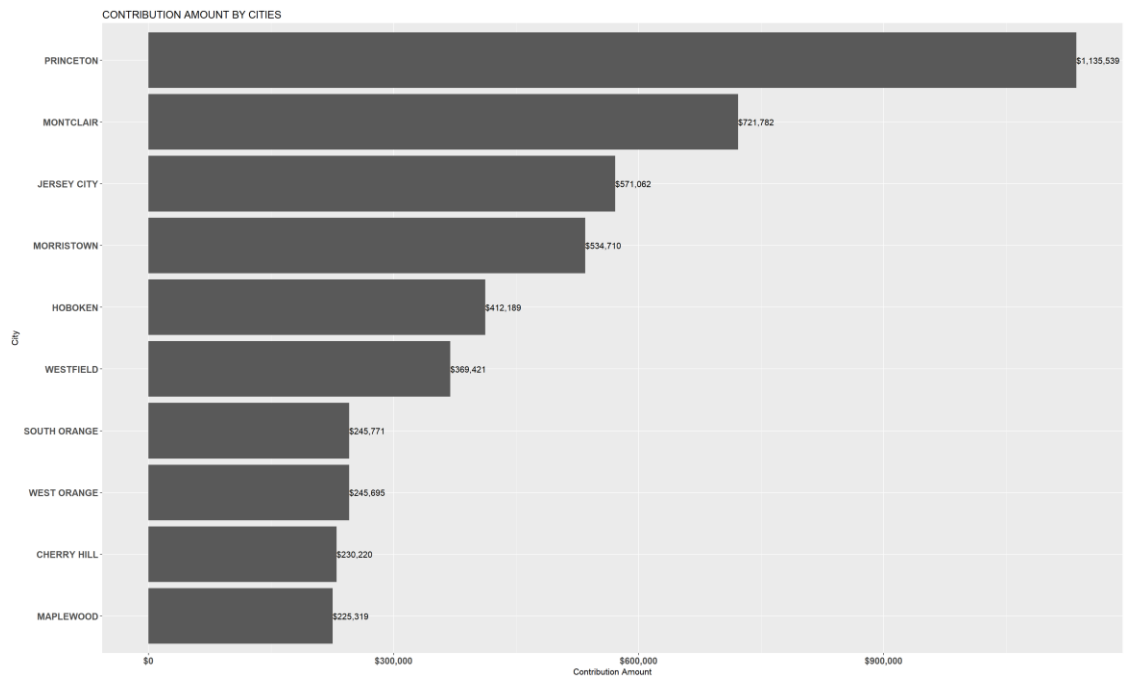
Hillary Clinton has received the highest amount of contribution of 125,29,471



Retired people have contributed the highest amount.



The highest contributions were made in the month of October, 2016, which was closer to the election date (Nov 8, 2016)



The city of Princeton made the highest amount of contribution. Although there were more number of contributors from Jersey city, it is ranked third in the contribution amount. This might mean that there were many small contributors from Jersey city.

## SUMMARY OF BIVARIATE ANALYSIS

### Relationships observed from Bivariate Plotting

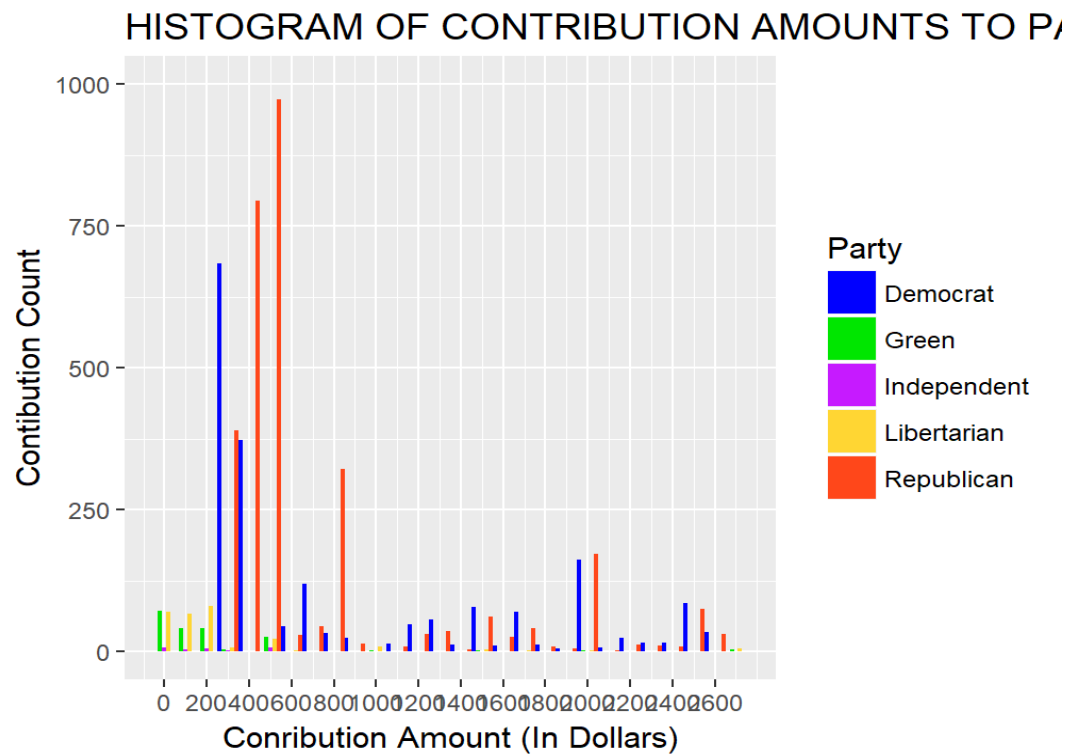
From Bivariate Plotting, I observed the following relationships.

1. From the Box Plot, looks like the Democratic party has more number of outliers with a median contribution of \$25.00 which is lesser than the median contribution of the Republican Party which is 53.00 dollars. Democratic party has received more number of large contribution amounts from the contributors than small amount when compared with the Republican party.
2. The Democratic Party has received the highest contribution amount of \$146,92,000 followed by the Republican Party with 104,85,000.
3. Hillary Clinton has received the highest amount of contribution of 125,29,471.
4. Retired people have contributed the highest amount.
5. The highest contributions were made in the month of October, 2016, which was closer to the election date (Nov 8, 2016)
6. The city of Princeton made the highest amount of contribution. Although there were more number of contributors from Jersey city, it is ranked third in the contribution amount. This might mean that there were many small contributors from Jersey city.

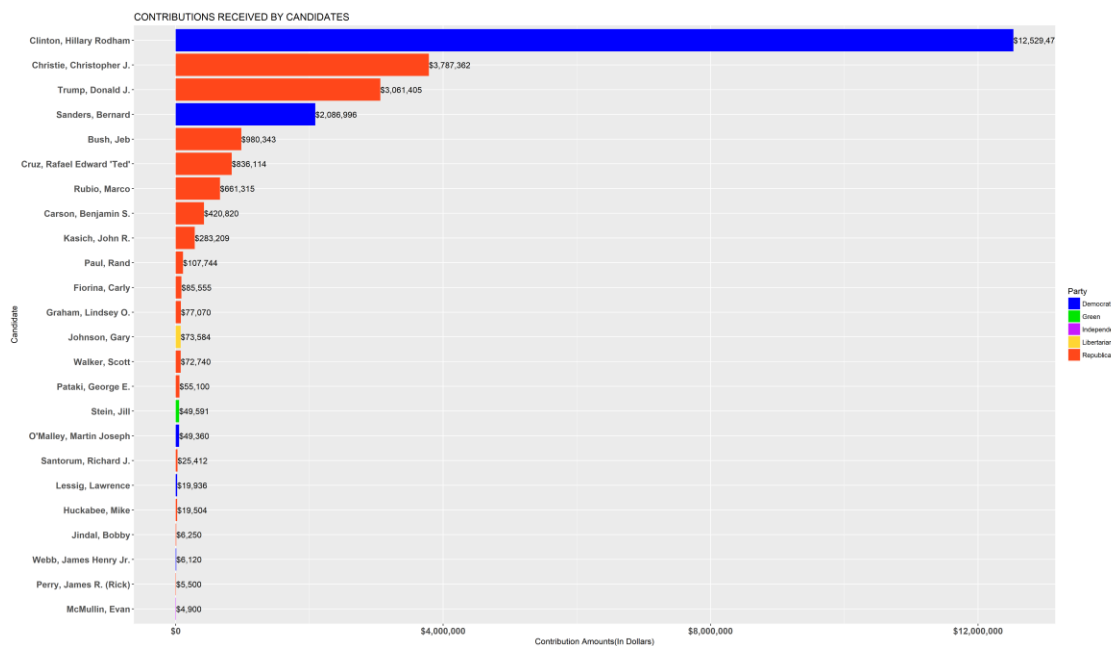
### Interesting Observation

The most interesting observation from the analysis so far is that Hilary Clinton has got more number of contributions as well as highest amount of contribution. Now, although Donald Trump and Bernie Sanders got the most number of contributions, in terms of the contribution amount, they stand third and fourth. This might be because they had more number of small contributors to fund. Christie Christopher on the other hand had large contribution amount from less number of High contributors.

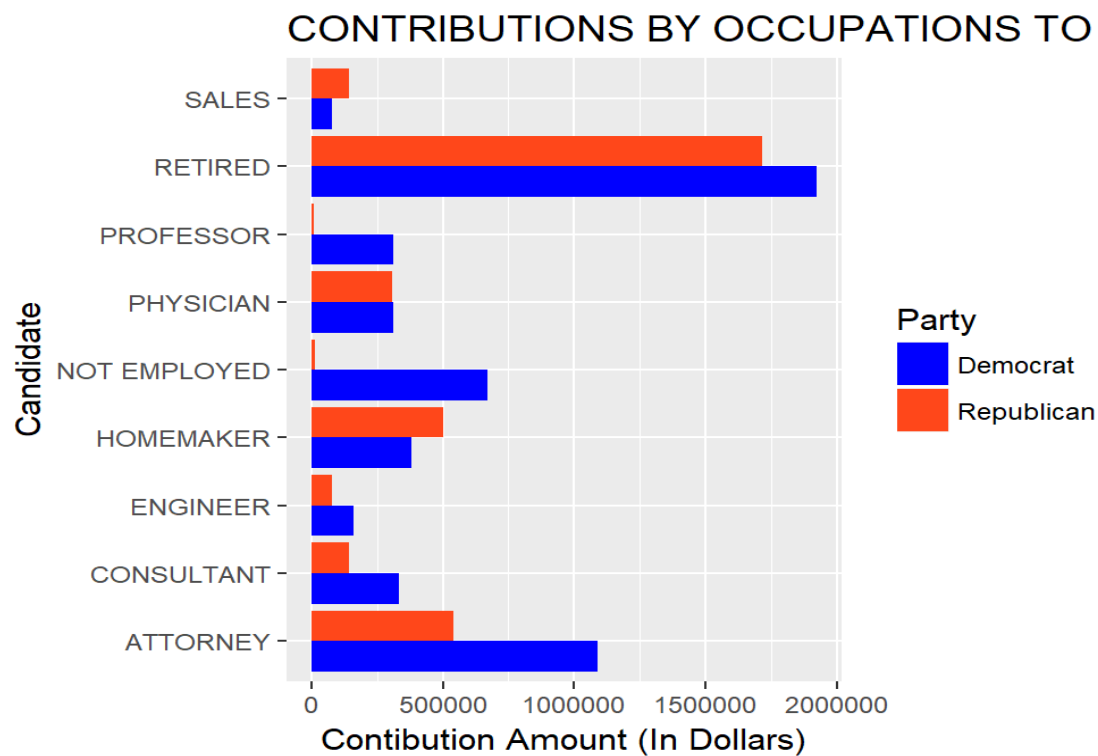
## Multivariate Plots Section



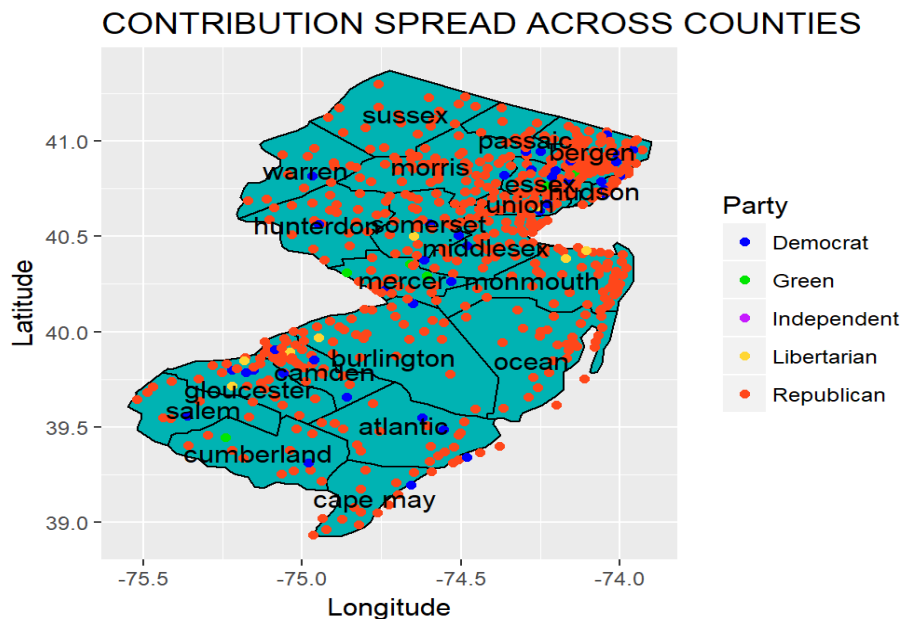
From the above histogram, the leading parties - Democratic and Republican have a spread of various contribution amounts from small to large. The other three parties - Independent, Libertarian and Green party have mostly smaller contribution amounts.



From the Bar chart, Hilary Clinton and Bernie Sanders of the Democratic Party and Donald Trump and Chris Christie of the Republican Party are the top four candidates who received the highest amount of total contributions.

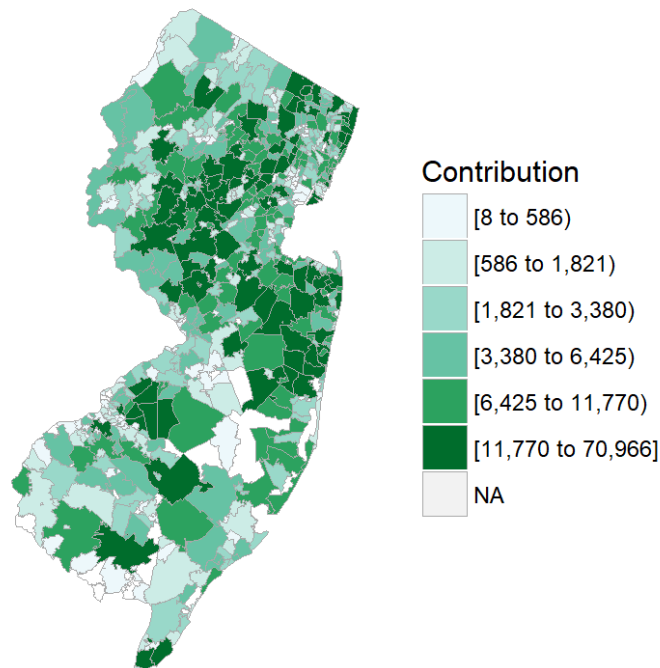


The top occupations that contributed the most seems to have leaned towards contributing more to the Democratic Party except for Sales and Homemakers in which the trend shows that contributions made were more for the Republican Party



Bergen, Essex, Hudson, Camden counties seems to have made more number of contributors when compared to other counties and also the contributors to the republican party seems much more widespread compared to the other parties.

## TOTAL CONTRIBUTIONS BY COUNTY



somerset, Morris, Hunterdon, Monmouth counties seems to have highest contribution amounts.

## Multivariate Analysis

### Summary of Observations

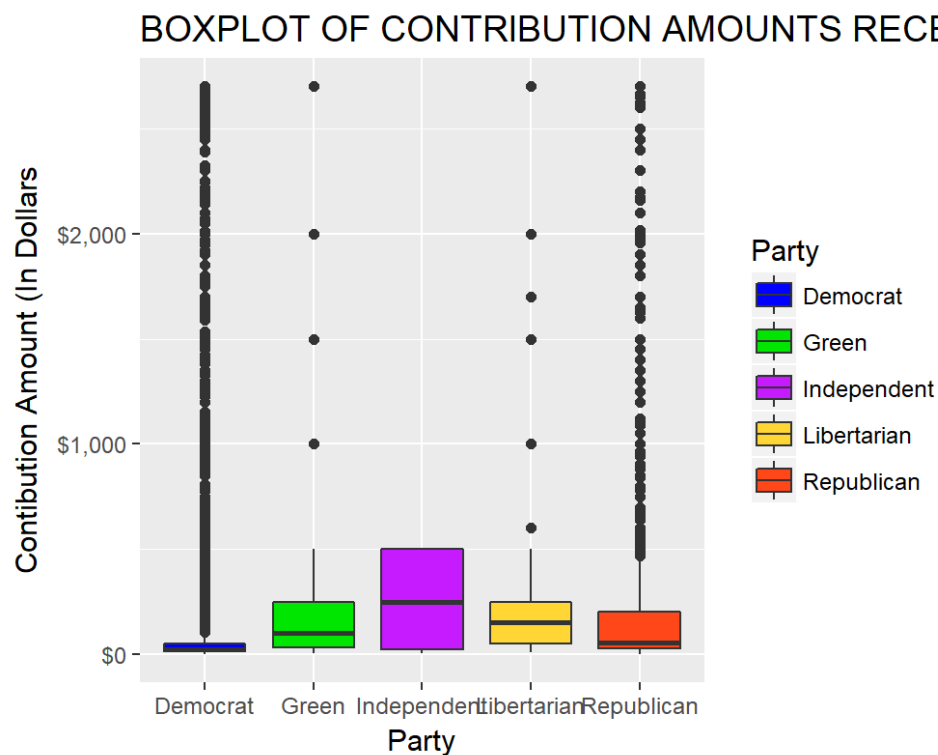
From multivariate plotting, the following are the relationships I observed.

1. Looking at the histogram of contribution amount by Party, The leading parties - Democratic and Republican have a spread of various contribution amounts from small to large. The other three parties - Independent, Libertarian and Green party have mostly smaller contribution amounts.
2. The top candidates of the leading party, Hilary Clinton and Bernie Sanders of the Democratic Party and Donald Trump and Chris Christie of the Republican Party have received the highest amount of contributions.
3. The top occupations that contributed the most seems to have leaned towards contributing more to the Democratic Party except for Sales and Homemakers in which the trend shows that contributions made were more for the Republican Party

4. The counties - Bergen, Essex, Hudson, Camden seems to have made more number of contributors when compared to other counties and also the contributors to the republican party seems much more widespread compared to the other parties.
5. The counties - Somerset, Morris, Hunterdon, Monmouth seems to have highest contribution amounts.

## Final Plots and Summary

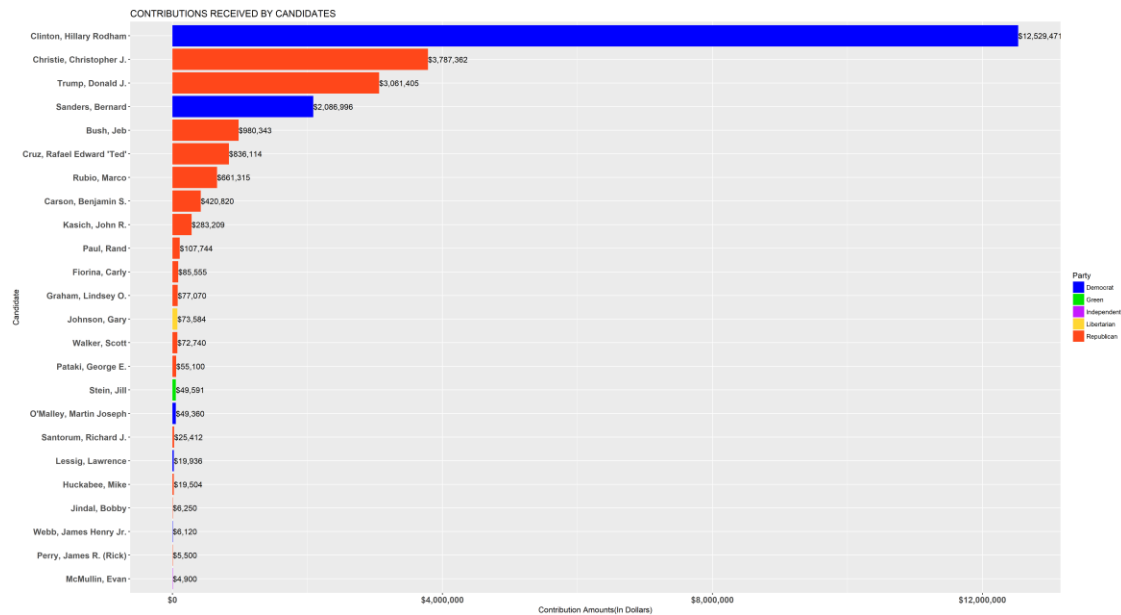
### Box Plot of Contribution Amounts for each Party



### Description

The Democratic party has a median contribution of \$25.00 which is lesser than the median contribution of the Republican Party which is 53.00 dollars. The Democratic Party has more number of outliers which implies that the party has received more number of large contribution amounts from the contributors than small amount. Next to this is the Republican party which has a better spread than the Democratic party with more number of small contributors. The Republican party also has outliers, the large contributors but it is lesser compared to the Democratic Party.

## Contribution Amount Received by the Candidates



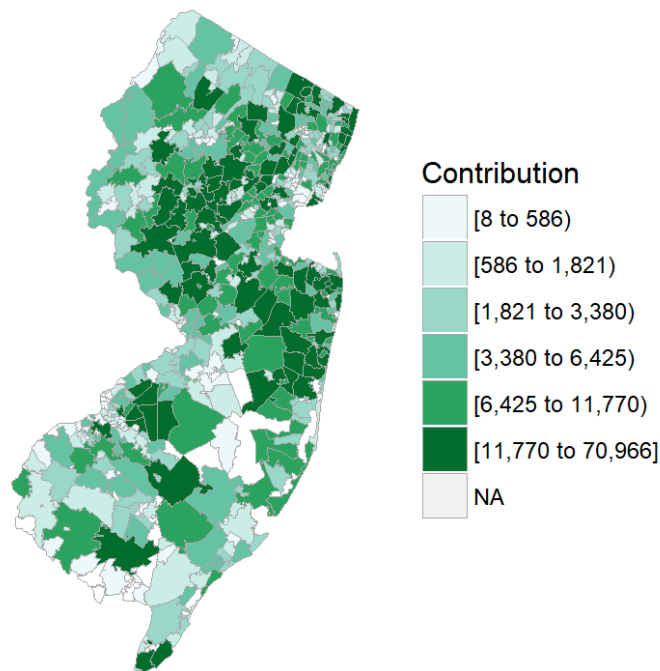
## Description

The top five candidates by the amount of contribution received are Hillary Clinton, Chris Christie, Donald Trump, Bernie Sanders and Jeb Bush. All of them are the top candidates of the leading parties - Democratic and Republican. This Plot also explains that Hillary Clinton has got the highest amount of contribution leap from the total contributions that the Democratic Party got. Top Candidates from Republican party on the other hand, have almost same spread of total contribution amounts. One more interesting observation to note here is that although Donald Trump has got more number of contributors, Chris Christies seems to lead by the contribution amount. This might be because Chris Christie has small number of contributors but have contributed huge amounts.



## Total Contribution Spread in Counties

### TOTAL CONTRIBUTIONS BY COUNTY



## Description

The counties - Somerset, Morris, Hunterdon, Monmouth, Mercer seems to have highest contribution amounts although these were not the top counties with high number of contributors. It is interesting to note that these counties are ranked top with high median household income. There are also high contribution notices in counties such as Ocean, Middlesex. This might be because contributors from these counties would have supported a particular candidate or a party and would have made high amount of contributions.

---

## Reflection

I chose to explore the city of New Jersey because that is the city I live in. Although New Jersey is a blue state/Democratic state, I thought it would be interesting to explore the financial contribution by the people of New Jersey to the candidates and party.

The way of exploring the project splitting it into Univariate, Bivariate and Multivariate was a very helpful method to think about questions and answer them in the flow of the project. Initially when performing the Univariate analysis, I was focussed on the number of contributions made for different parties and candidates, number of contributions made based on contributors city, occupation, gender and of course the number of contributions made based on contribution

amounts. When looking at these variables and counts, question such as if number of contributions were related to contribution amounts and also the spread of contributions across the state of NJ. To answer these questions, the Bivariate analysis helped me figure out a good flow to get the best way to analyze these variables. Finally Multivariate analysis helped put together the variables from univariate and bivariate analysis and yield a final explanation of the contribution spread.

After the exploration, it indeed was true that there were more contributors to the Democratic Party.

Since this was my first project exploring data with R, I enjoyed learning new ways to explore this dataset. Learning about the Choropleth and gender package was very interesting. I had difficulties in getting the code run when coding with these packages but help from internet blog posts were good guiding factors to complete this Project.

Future analysis would include looking at other states and also to analyse by choosing a particular candidate and looking at the contributions received by them.