

Financial Contribution Analysis in New York State of 2016 United States Presidential Campaign

Shilin Li

17, June, 2018

Abstract

The United States presidential election of 2016 was the 58th quadrennial American presidential election, during the course the series of presidential primary elections and caucuses took place between February and June 2016, staggered among the 50 states, the District of Columbia and U.S. territories. The presidential campaign finance data among the 50 states provides various valuable information for the prediction of the election. Current study focuses on 2016 US presidential campaign donations in the state of New York, which is the most influential region of today, not only in the United States but also all over the planet. New York State has been the most important source of political fund-raising in the United States for both major parties since 1988. Four of the top five zip codes in the nation for political contributions are in Manhattan. The top zip code, 10021 on the Upper East Side, generated the most money for the 2000 presidential campaigns of both George Bush and Al Gore. The data were downloaded from the Federal Election Commission, starting from October 11th, 2013, up to date until December 31, 2016.

According to the financial data set, key information such as demographic and geographic support for each party and candidate can be acquired, in addition, contributors's gender, occupation, etc., can be added to draw more comprehensive analysis. The patterns of the donation flow over time also reflect the current status of the candidates in the race. For this project, my analysis will be centered on the support for major two parties, namely, the democrats and republicans, in terms of parameters like demographic, geographic, gender, occupation, time, etc.

By understanding those questions, we can have a better view of 2016 presidential election in New York State, and possibly, a preliminary estimate for the national statistics.

This is an exploration of 2016 US presidential campaign donations in the state of New York.

Dataset

Preparations

This data sheet comes from Federal Election Commission, URL:

<https://classic.fec.gov/disclosurep/PDownload.do>

(<https://classic.fec.gov/disclosurep/PDownload.do>).

Data Wrangling

Adding New Variables

The first step of data analysis is wrangling, according to the profile of our data, it seems we are missing certain key components, such as gender, party affiliation as well as geographical coordinates (latitude and longitude). In order to perform profound analysis, I am hereby adding those variables by generic and external packages of Rstudio.

Cleaning Zipcodes

As we know there are only in total 25 candidates in the dataset, it is thus very easy to add their party affiliations. To acquire the coordinates, namely, longitude and latitude, we need to seek help from package zipcode. However, the format of data (zipcode) in that package only include 5-digit, as opposed to our zipcodes format which includes various types, such as 1, 2, 3, 4, 5 and 9. We then need to convert the 9 digit format to its 5 equivalent

(https://excel.tips.net/T002654_Shortening_ZIP_Codes.html)

(https://excel.tips.net/T002654_Shortening_ZIP_Codes.html) by reserving the left part, and for those 4 digit format, we can only deal with numbers between 1 and 4975 (as New York State zipcode ranges from 100001 to 14975) by adding number '1' at the beginning, for the rest it is impossible to extrapolate otherwise we get wrong data. For those numbers of 1 to 3 digit format, we have no choice but to discard.

Extracting the first name and subject to gender prediction after cleaning

To predict gender, the package of gender and its extensive database genderdata could be of help. The way it works is to predict gender from first names using historical data. The package includes 5 types of methods, we will compare the efficiency of each and select the one which yield the largest number of predictions. Prior to the prediction, we need to extract the first and last name and then clean the format of first name in order to make the gender predict function recognize them.

Data Cleaning

There are in total 7915 negative contributions in the data set, which shall be refunded. In order to make accurate analysis, I need to drop these observations.

More importantly, based on contributions regulation in 2016, the Limits had increased to \$2,700 per capita, per candidate, per election. So I also have to drop all contribution above the \$2,700 limit, not only because they violate the Federal Election Campaign Act and will be refunded, but also the affect validation of truth.

Data Analysis

Overview

```

## 'data.frame': 618112 obs. of 29 variables:
## $ cand_first      : chr "BENJAMIN" "BENJAMIN" "BENJAMIN" "BENJAMIN" ...
## $ contbr_first    : chr "WALTER" "DEBRA" "LAUREN" "LORETTA" ...
## $ zip             : chr "10001" "14450" "11557" "14591" ...
## $ cmte_id         : chr "C00573519" "C00573519" "C00573519" "C00573519" .
...
## $ cand_id         : chr "P60005915" "P60005915" "P60005915" "P60005915" .
...
## $ cand_nm         : chr "Carson, Benjamin S." "Carson, Benjamin S." "Cars
on, Benjamin S." "Carson, Benjamin S." ...
## $ contbr_nm       : chr "FISCHER, WALTER MR." "LUBBERTS, DEBRA" "PEASE, L
AUREN M. MRS." "CLARK, LORETTA M. MS." ...
## $ contbr_city     : chr "NEW YORK" "FAIRPORT" "HEWLETT" "WYOMING" ...
## $ contbr_st       : chr "NY" "NY" "NY" "NY" ...
## $ contbr_employer : chr "RETIRED" "MARION CENTRAL SCHOOLS" "HOMEMAKER" "R
ETIRED" ...
## $ contbr_occupation: chr "RETIRED" "ENGLISH TEACHER" "HOMEMAKER" "RETIRED"
...
## $ contb_receipt_amt: num 100 100 200 100 100 25 50 25 500 25 ...
## $ contb_receipt_dt : chr "2016-02-22" "2015-12-16" "2015-10-22" "2015-11-0
5" ...
## $ receipt_desc    : chr " " " " " ...
## $ memo_cd         : chr " " " " " ...
## $ memo_text       : chr " " " " " ...
## $ form_tp         : chr "SA17A" "SA17A" "SA17A" "SA17A" ...
## $ file_num        : num 1066643 1057553 1057553 1057553 1057553 ...
## $ tran_id         : chr "SA17.1313432" "SA17.1055330" "SA17.736178" "SA17
.847929" ...
## $ election_tp     : chr "P2016" "P2016" "P2016" "P2016" ...
## $ party           : chr "republican" "republican" "republican" "republica
n" ...
## $ latitude         : num 40.8 43.1 40.6 42.8 42.4 ...
## $ longitude        : num -74 -77.4 -73.7 -78.1 -76.8 ...
## $ contbr_gender   : chr "male" "female" "female" "female" ...
## $ cand_gender     : chr "male" "male" "male" "male" ...
## $ year             : num 2016 2015 2015 2015 2015 ...
## $ month            : chr "Feb" "Dec" "Oct" "Nov" ...
## $ day              : int 22 16 22 5 6 30 10 28 15 17 ...
## $ month_year       : chr "Feb, 2016" "Dec, 2015" "Oct, 2015" "Nov, 2015" .
...

```

The Financial Contribution Analysis in New York State of 2016 United States Presidential Campaign is based on data from Federal Election Commission website, it contains 618112 observations and 29 variables after wrangling and cleaning, each parameter represent one financial donation transaction

In order to dive deep into current dataset, I have added several variables accordingly, and meanwhile, drop some redundant as well as misleading data.

The Original Variables Meanings are Summarized Below:

- 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_employer: contributor employer
- contbr_occupation: contributor occupation
- contb_receipt_amt: contribution receipt amount
- contb_receipt_dt: contribution receipt date
- 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

New Variables Summary:

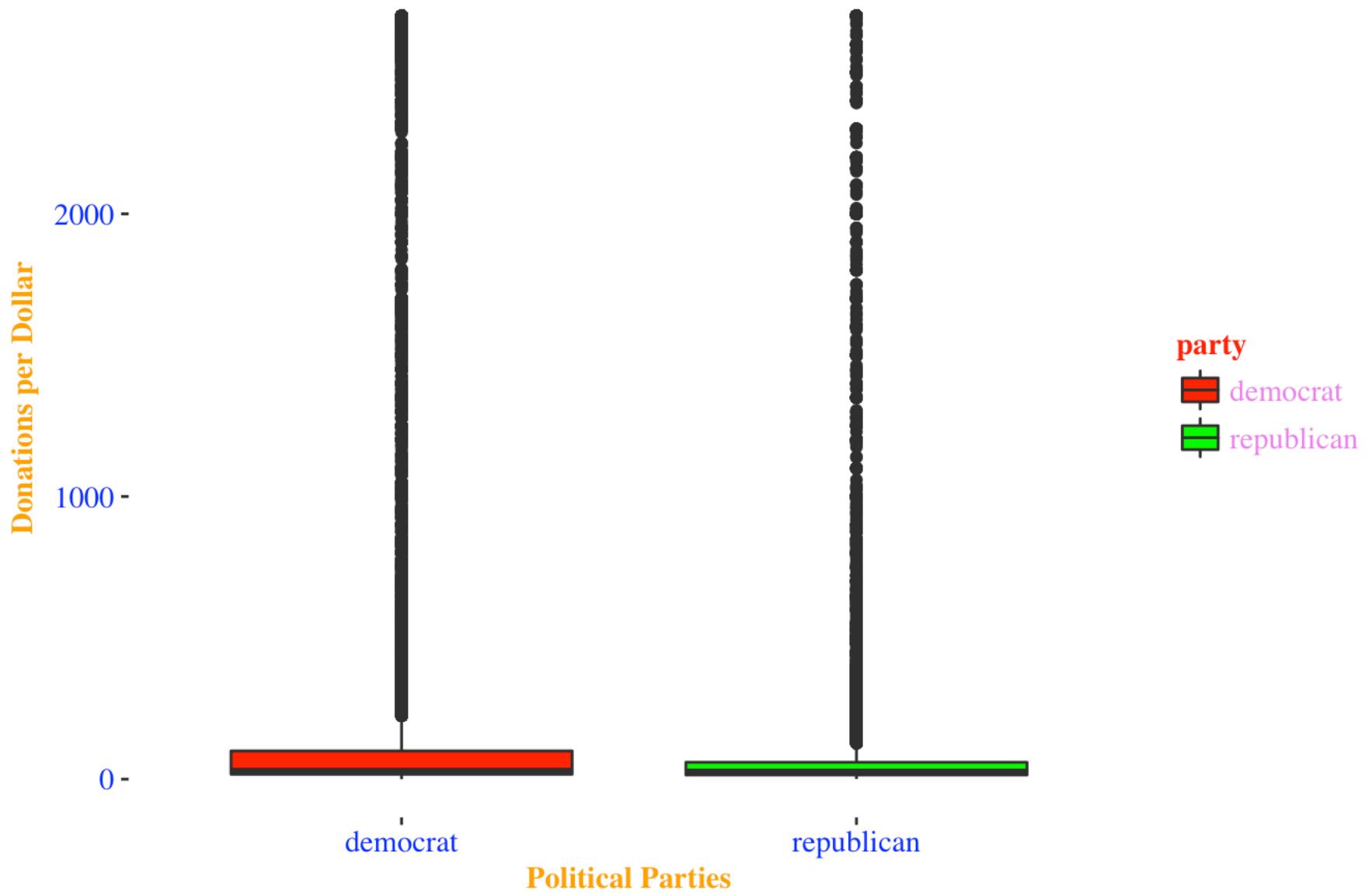
- zip: contributor's zipcode
- party: candidates' party affiliation
- latitude: contributor' geographic latitude
- longitude: contributor' geographic longitude
- contbr_first: contributor' first name
- cand_first: candidate' first name
- contbr_gender: contributor' gender
- cand_gender: candidate's gender
- date: contribution receipt date
- day: contribution receipt day
- year: contribution receipt year
- month: contribution receipt month

- month_year: contribution receipt date (year month format)

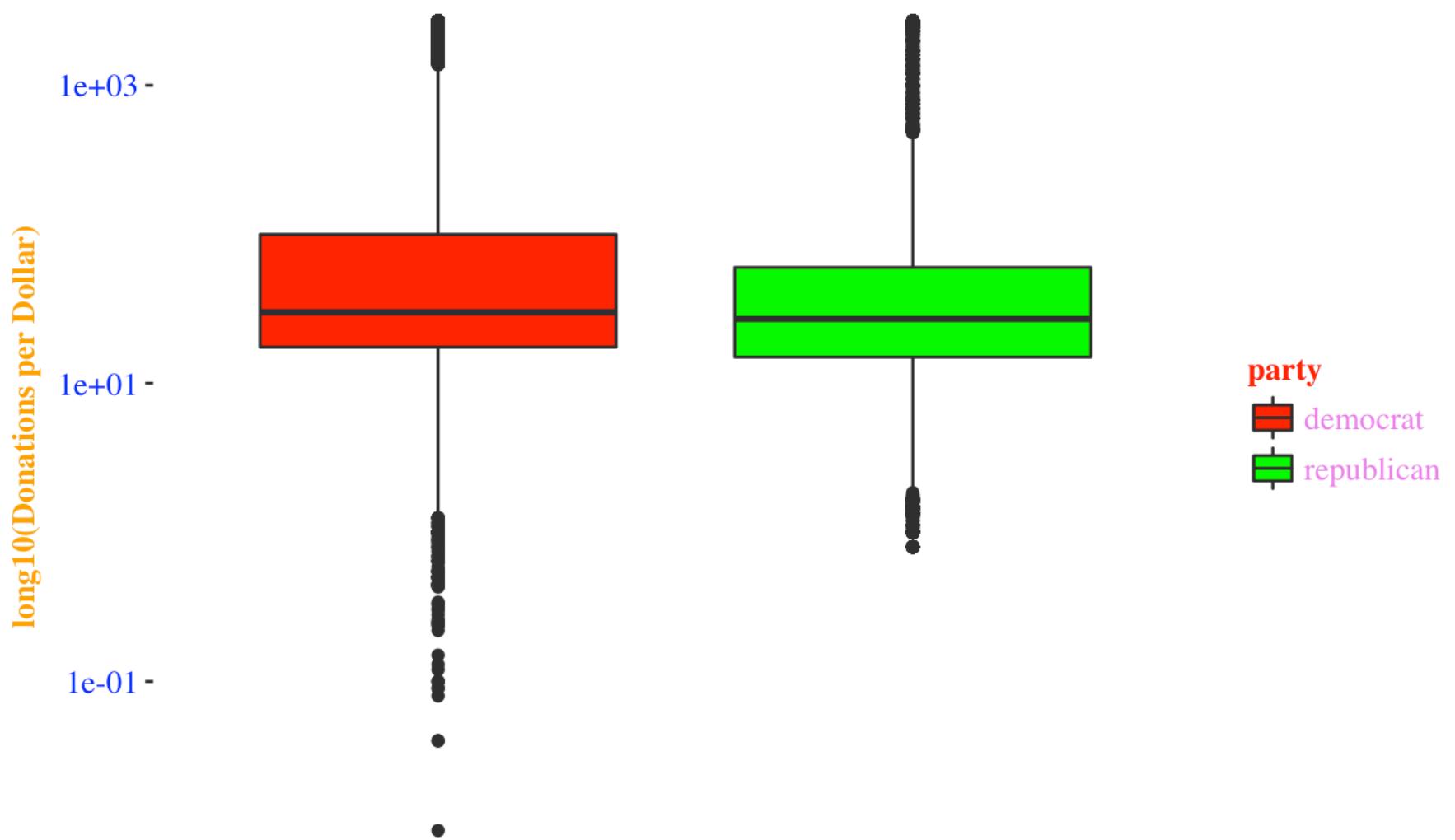
Boxplot of Contribution Receipt Amount

Firstly, I look at the pattern of donation received by parties in boxplot. The outliers greatly affect visualization, thus I have to perform the log transformation. Afterwards, it is easy to tell that donation received by democrats are more spread than that of republicans, that suggests bigger variance and standard deviation in democratic group.

Boxplot of Donations Received by Parties



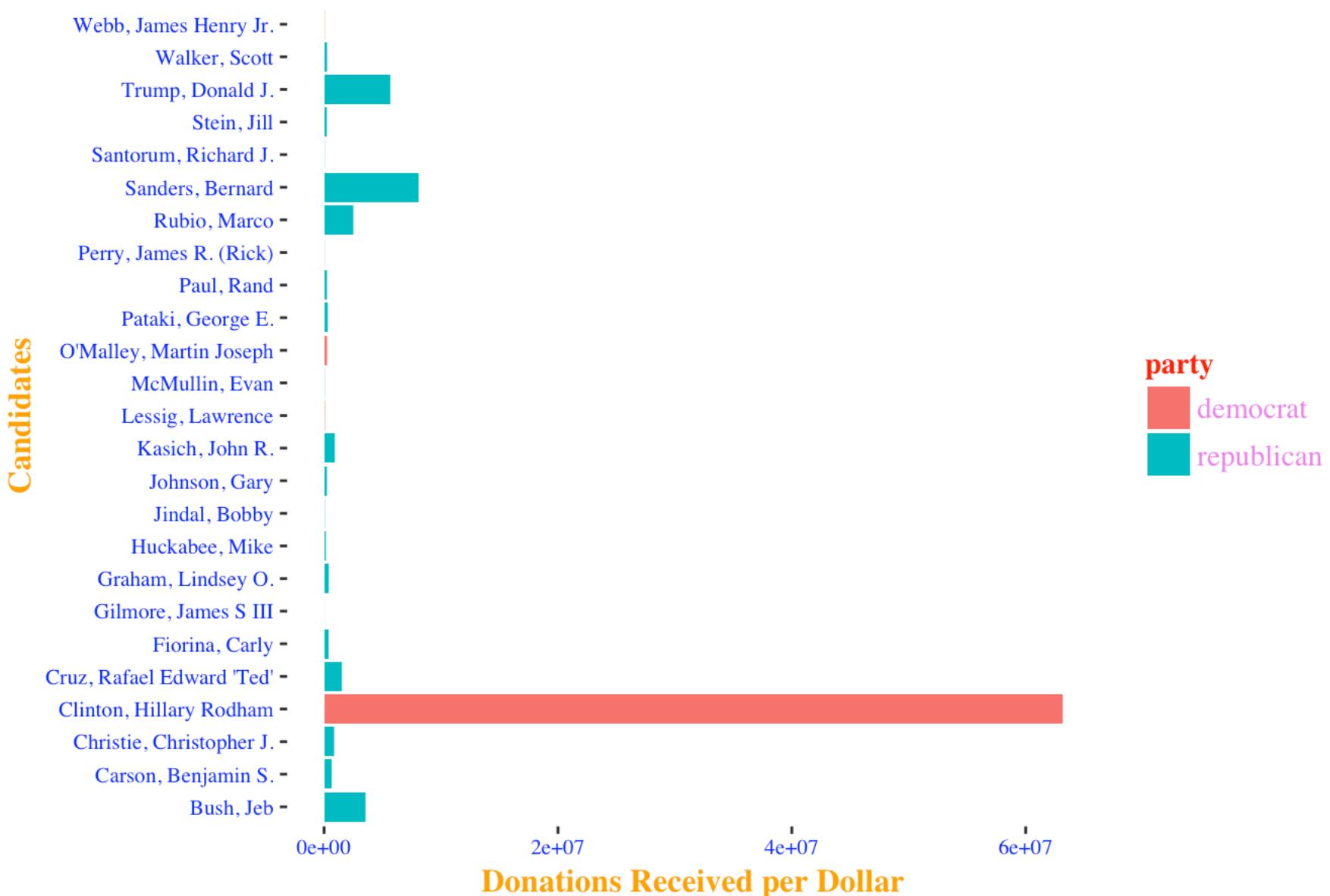
Boxplot of Donations Received by Parties



Histogram of Contribution Received by Candidates

From the histogram of “Donation Received by Candidates”, I found there are 25 candidates, 21 as republicans and 4 as democrats. Hilary Clinton was in a super lead in the race in New York State, she alone received more than 60 million dollars, much more than Bernard Sanders, who stood in second place right after Hilary and received around 10 million, even more than 5 times over the total of the rest candidates.

Donations Received per Candidates

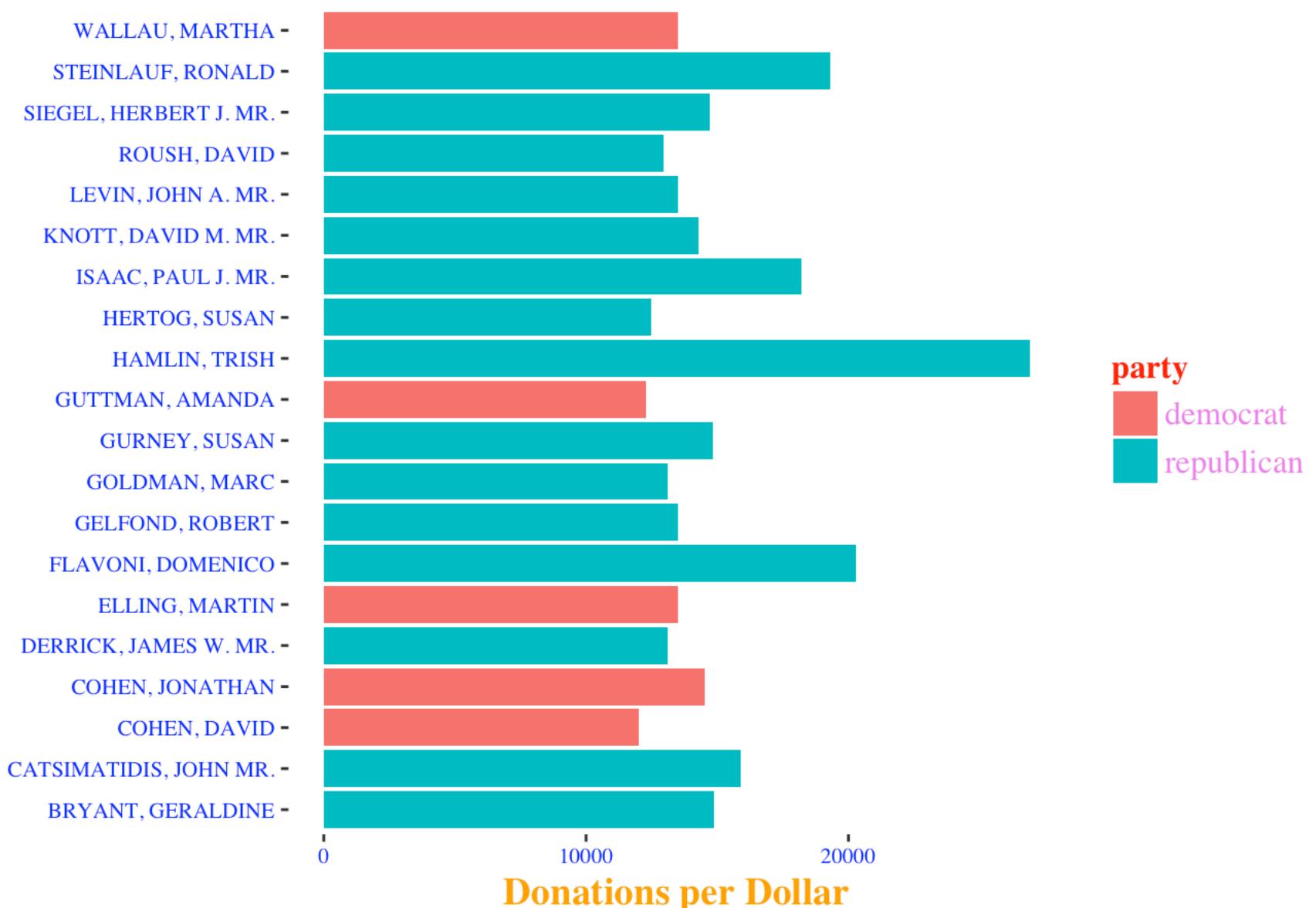


Histogram of Donations Offered by Contributor

From the histogram of “top 20 donations made by contributors”, I found somehow to some extent, this plot is similar with the histogram “Donation Received by Candidates”. That is more candidates from republican (15) received donations than democrats (5). That is fair because the candidates in New York State is consists of 21 republicans versus 4 democrats. The main difference from the previous histogram is the total amount of donation received by republican are much more than that from democratic contributors. Among all the top 20 contributors, Trish Hamlin made the most donations, more than 20000 dollars.

Top 20 Donations made by Contributors

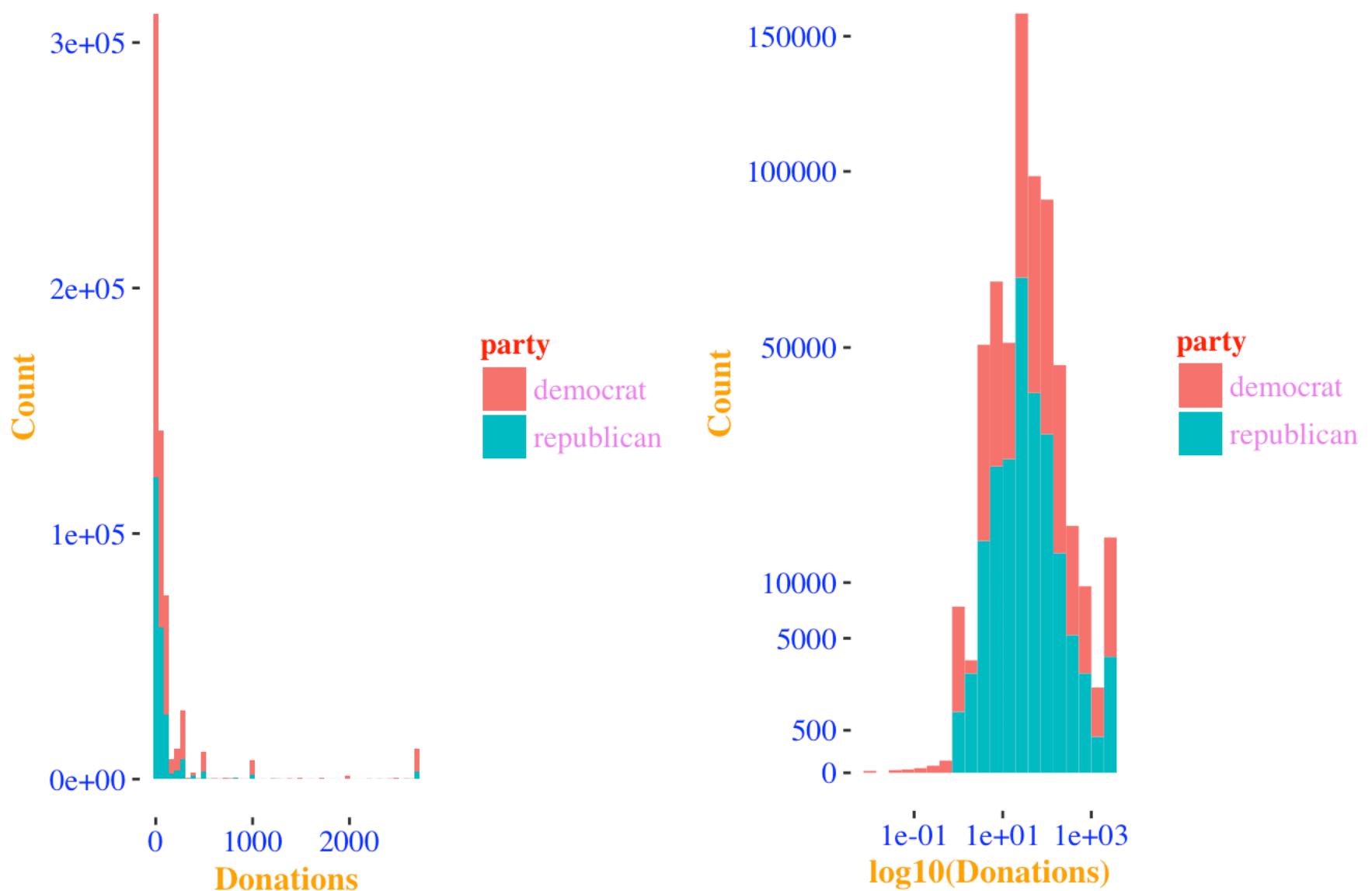
Contributors



Histogram of Distribution of Donations

In order to better understand the distribution of the donation, I also used histogram in addition to the previous boxplot. Again, log transformation made the visualization better. From the plot, most donations came from 0 to 500 dollars, even democratic contributors in total made more donation, however, the donation received also presented bigger variance.

Donations Received by Parties Donations Received by Parties in log Transformation



```
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
## 0.01     15.00   27.00 144.46 100.00 2700.00
```

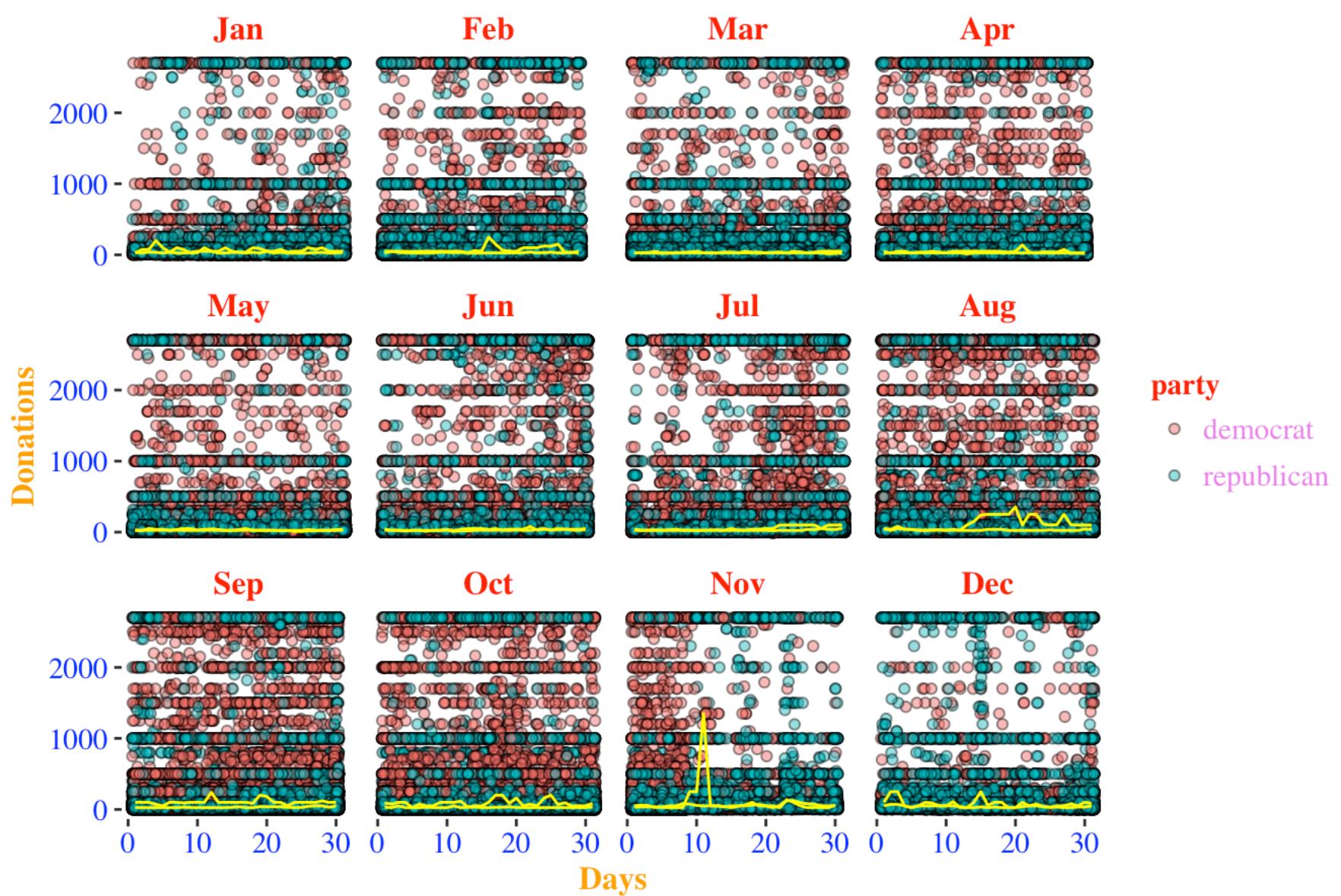
The distribution of donation mainly concentrates on the range of 0 to 500\$ as the plot shows, the statistics are summarized as above. Min donation is 0.01\$, median 27\$, 1st and 3rd quantile 15 and 100, respectively, max 2700\$. Therefore, the donations distribution skewed to the right, and log transformation made the visualization for better analysis.

Scatter Plot with Data

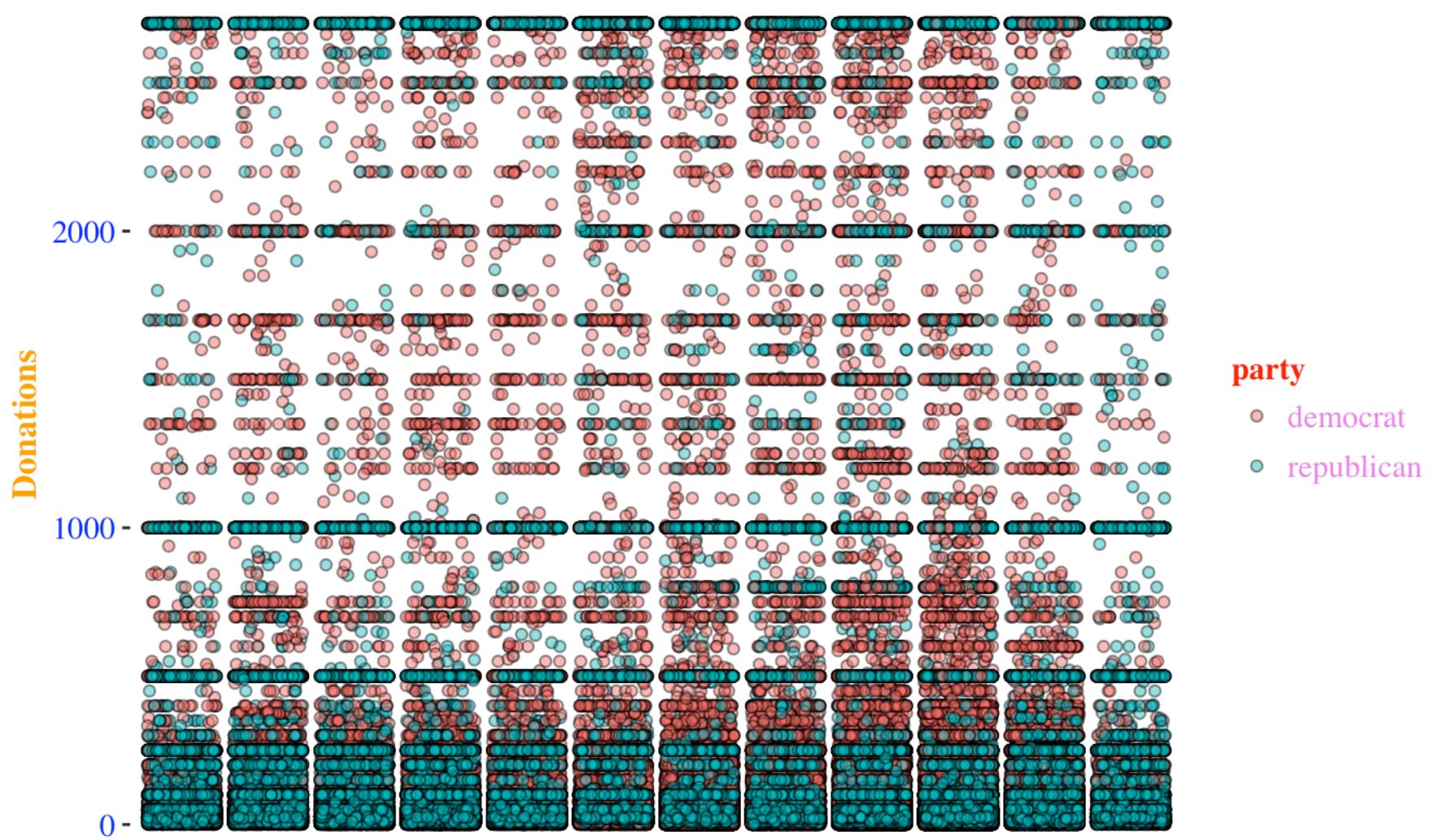
Plot by day and month with the distribution of donation, it is not easy to tell the pattern. I then switch to geom_line of mean donation with date, this time I found democratic party received a relatively higher donation (mean level of around 150 dollars per day) over days than that of republicans (mean level of around 100 dollars per day), the donation at the day level seems to be constant without big fluctuation. Over the month, the pattern of donation received by democrats was descending from 300 to 100 dollars, while republican ascending from 100 to 300 dollars.

Since the month_year information is not complete before 2016, and not that important as year of 2016, so I narrowed the timeframe only to 2016. There, I found highest mean donation happened in September, that might due to it was approaching the end of election. Meanwhile, the biggest standard error happened in January, I think that month was just at the end of Christmas Holiday, so there were less people made donations which caused that big standard error.

Donations Received by Days



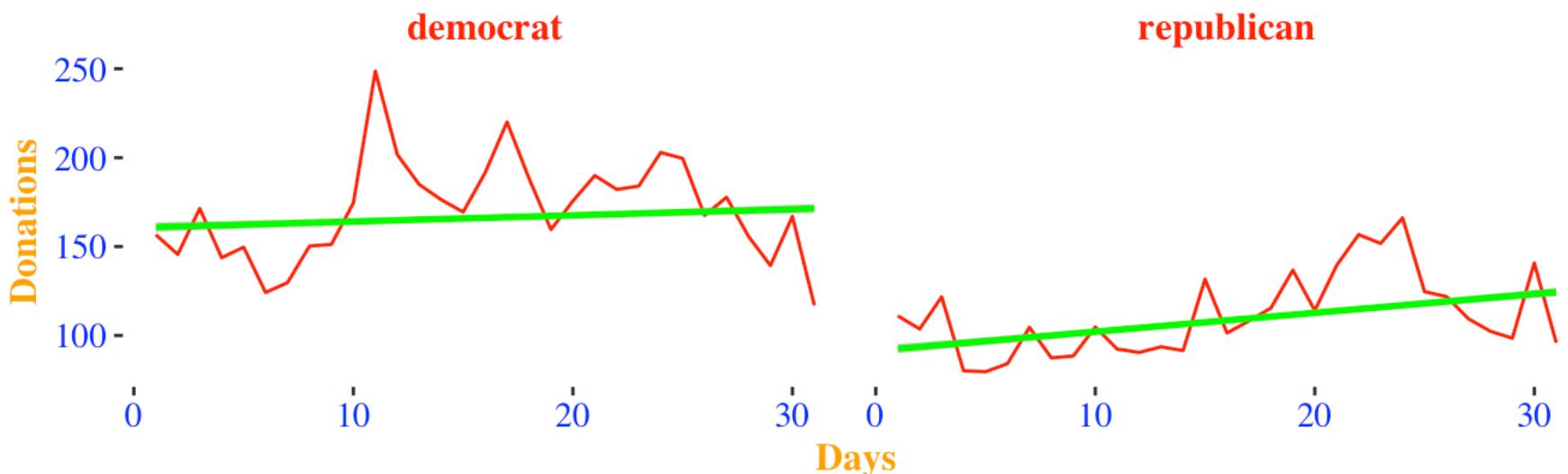
Donations Received by Months



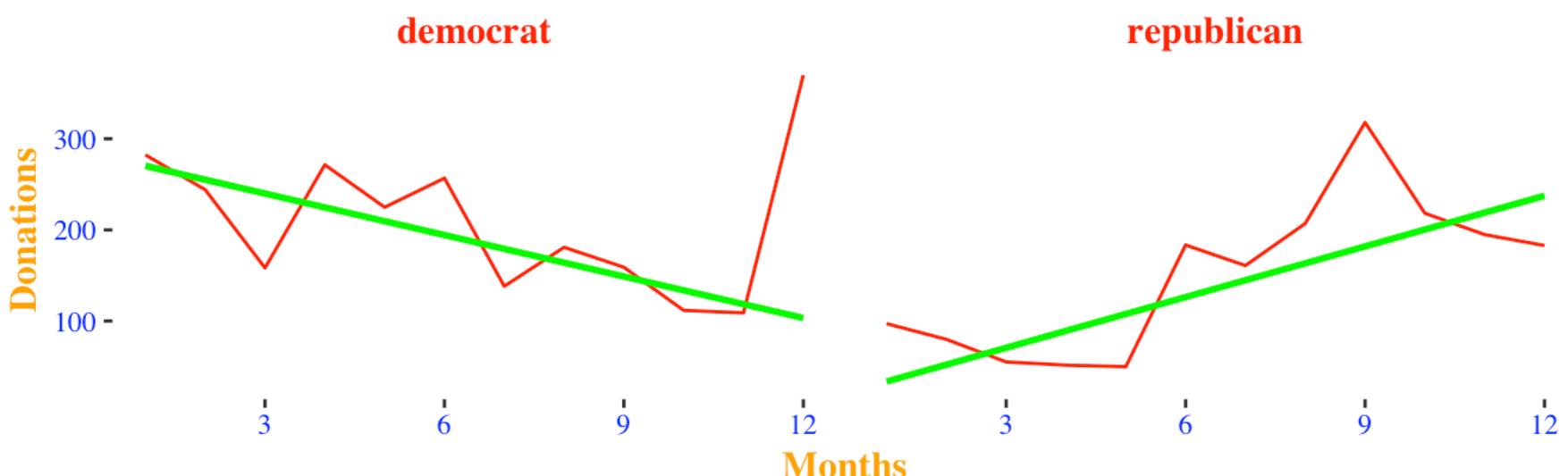
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

Months

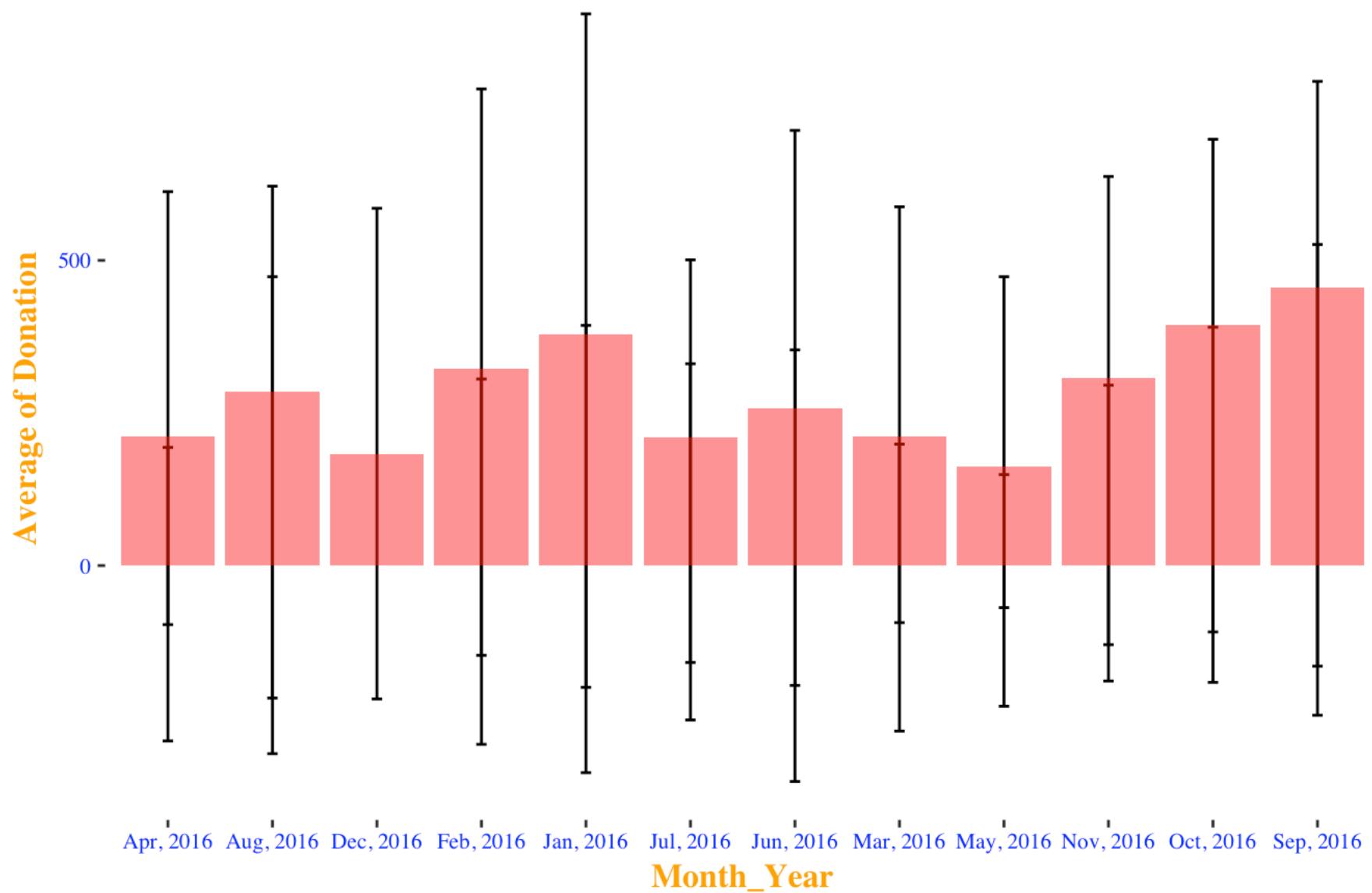
Average Donations Received by Days



Average Donations Received by Months



Statistics of Average Donation Over Each Month in 2016

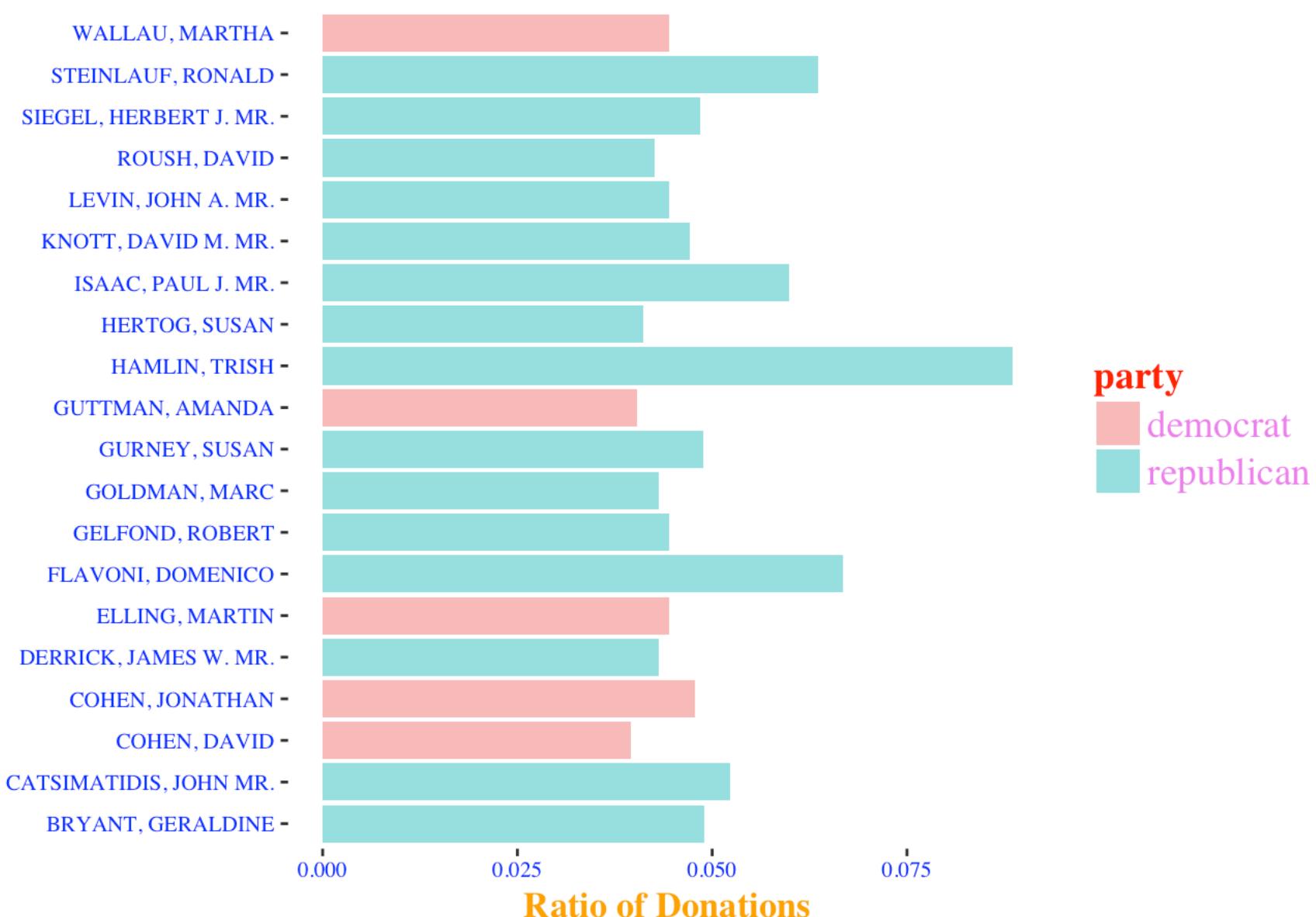


Ratio of Donation Made and Received by Parties

Diving deep into the pattern of donation, I found the top 20 contributors were more likely to support republican candidates, with Trish Hamlin, who supported republican made the highest contribution which possessed 7.5% of total donations among the top 20 contributors.

Ratio of Donations made by Top 20 Contributors

Contributor Name

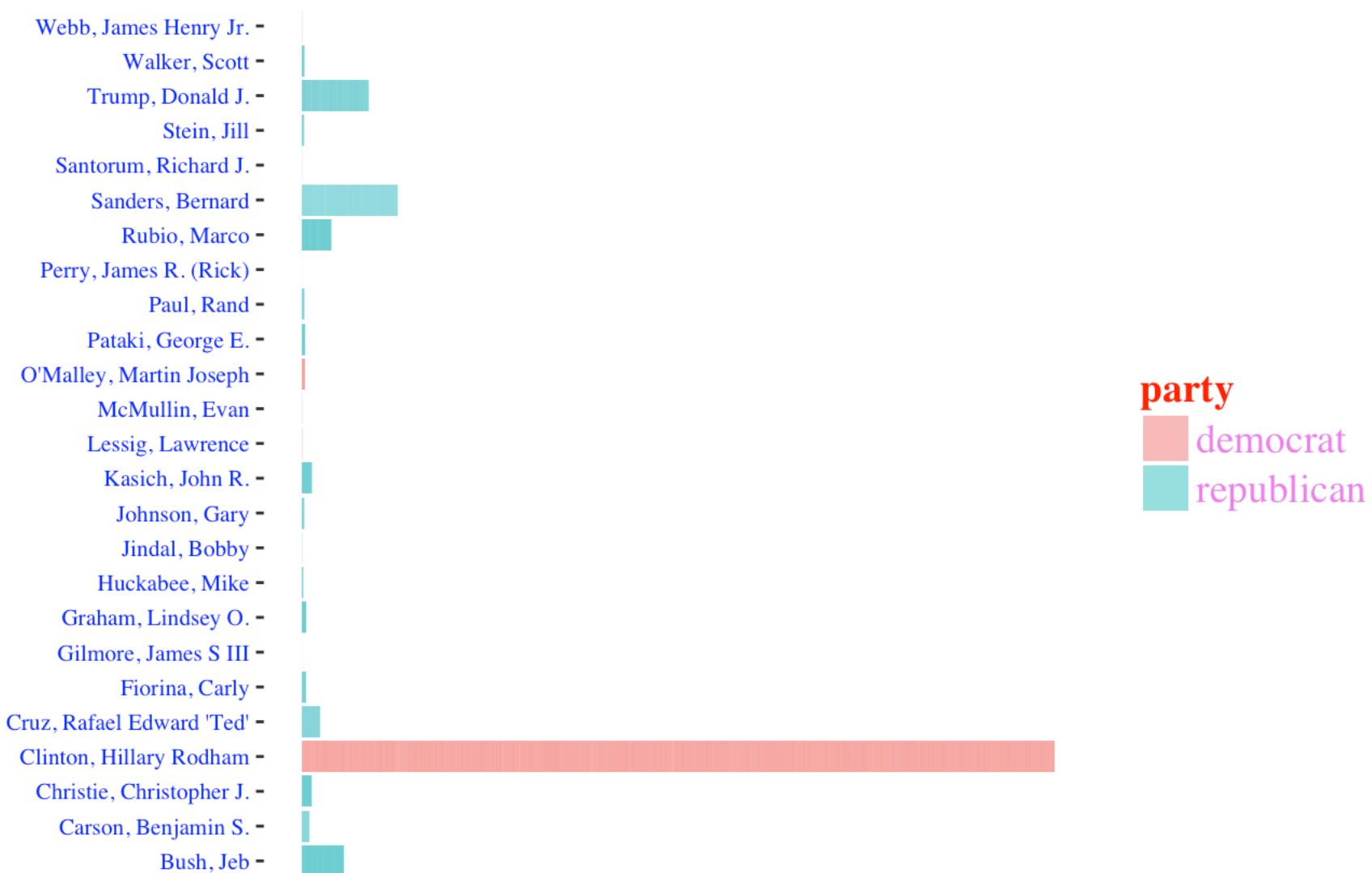


party

democrat
republican

Ratio of Donations Received by Candidates

Candidate Name



party

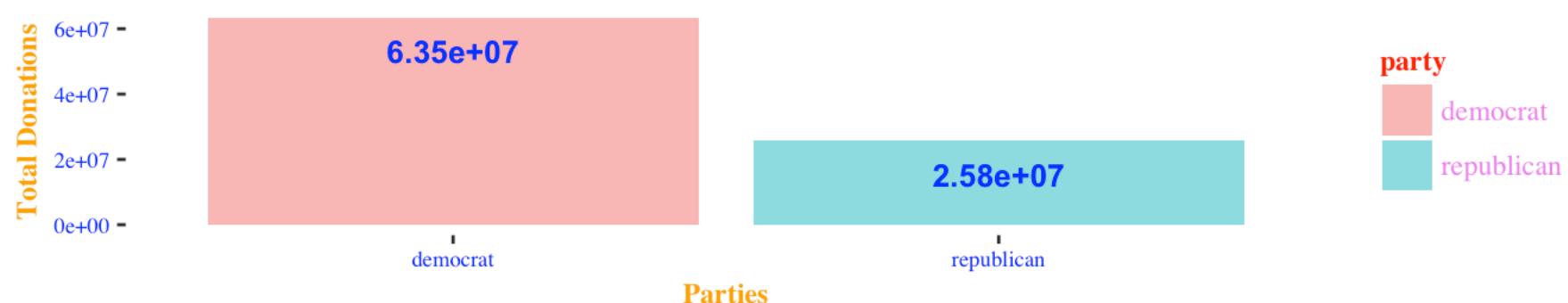
democrat
republican



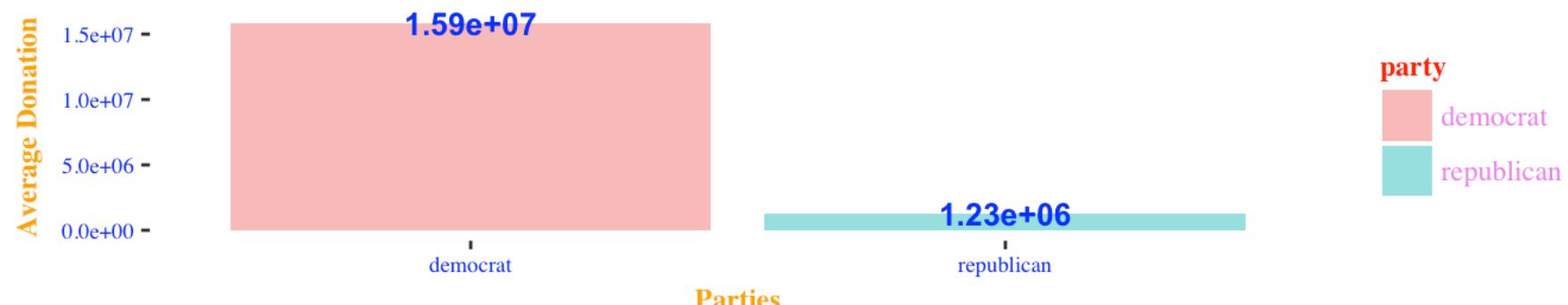
Total, Avg, Min and Max Donation

In total, democrats candidates received around 63.5 million dollars as opposed to around 25.8 million dollars received by republicans. The average donations received by candidates were around 15.9 million dollars per democratic candidate and around 1.23 million dollars per republican candidate. The average donations made by contributors to democrat were around 1020 dollars and around 476 dollars to republican. The max donation received by candidate is around 63.13 dollars, and we had already known it was definitely Hilary, whereas the min donation received was around 6000 dollars, and it was given to republican's candidate James S III, Gilmore. The max and min donation made by contributors are around 2700 and 0 dollars, they came from republicans, Trish Hamlin and Alfred Trotta, respectively.

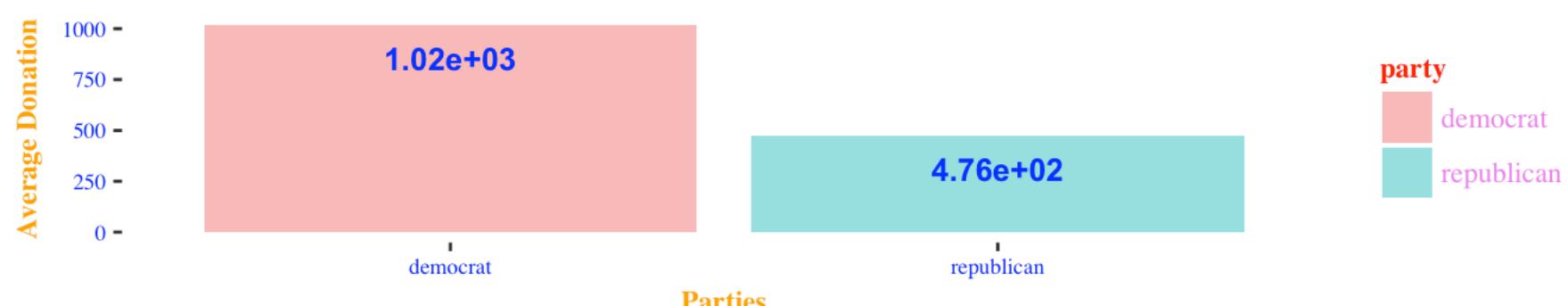
Total Donations Received by Parties



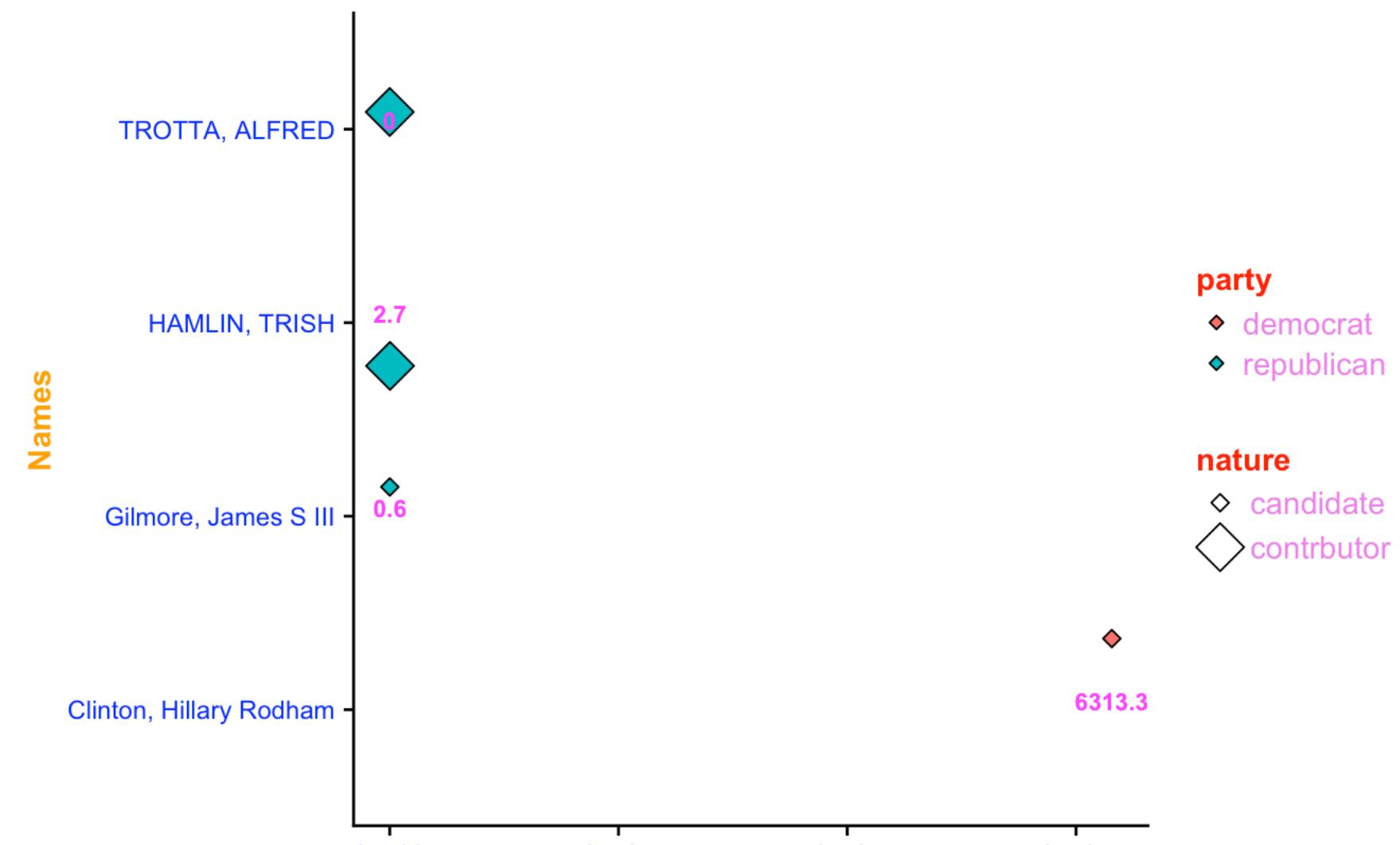
Average Donation Received by Parties



Average Donation Made by Contributors



Most and Least Amount of Donation Received by Candidates or Given by Contributors





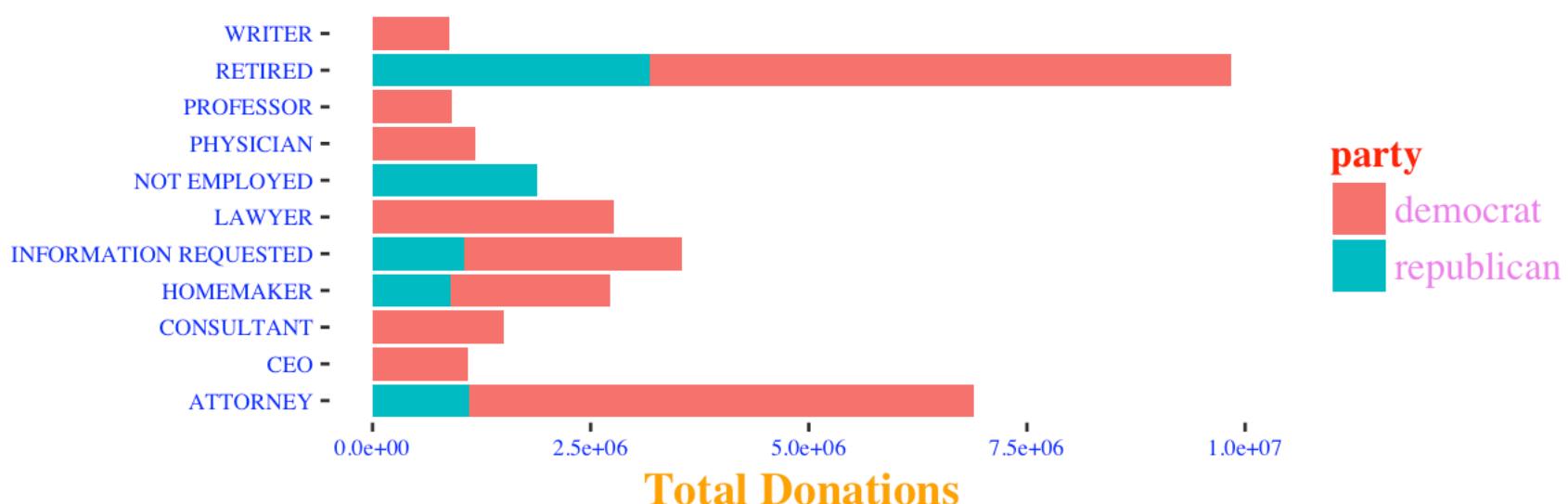
Top 15 occupation and employer and average contribution

The occupation of top contributor was from retired, who donated around 10 million dollars, but it gave little information regarding which occupation was most actively involved in political events. However, attorneys who made a second largest donation also contributed around 7 million dollars, at least it informed us that political related occupation was more likely to participate in such event. Among the top 15 occupations, more than half of the donation was given to democrats, which was in accordance with the high supporting rate of Hilary as well as democrats in financial parameters in New York State.

The top 15 employers suggested self employed folk was most interested in the presidential election, and most of them supported democrats. The average donation from top 15 occupations and employers followed the same pattern as the total donation mentioned above. In addition, from the scatter plot, democrats also tended to receive larger amounts of donations than Republicans in New York State.

Top 15 Donations by Contributors' Occupations

Occupations



Total Donations

Top 15 Donations by Contributors' Employers

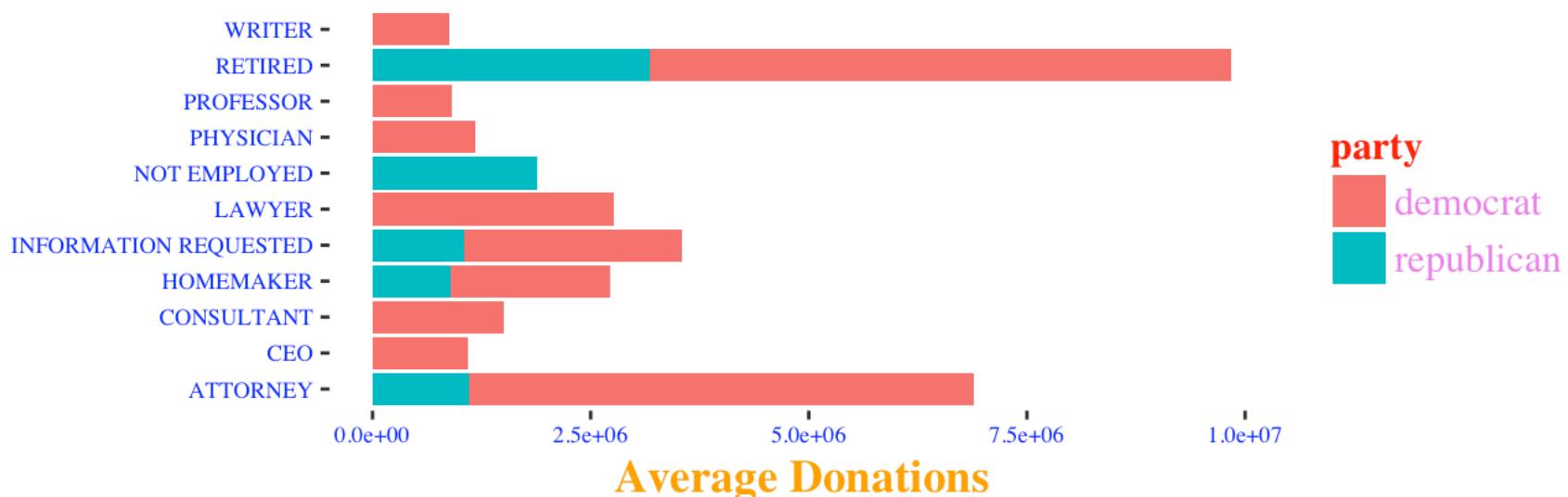
Employer



Donations

Top 15 Average Donations by Contributors' Occupations

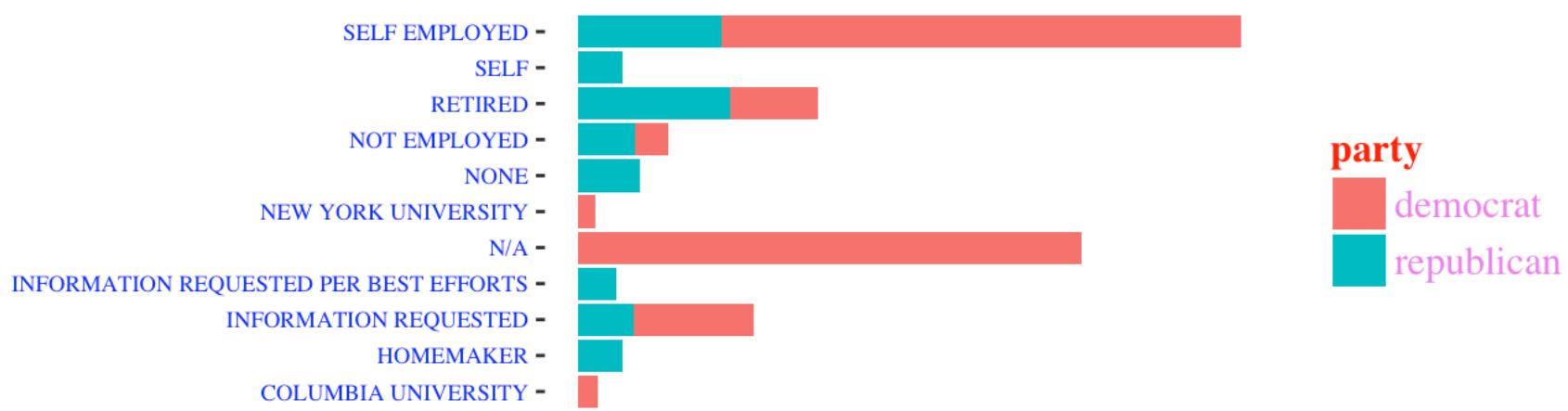
Occupations



Average Donations

Top 15 Average Donations by Contributors' Employers

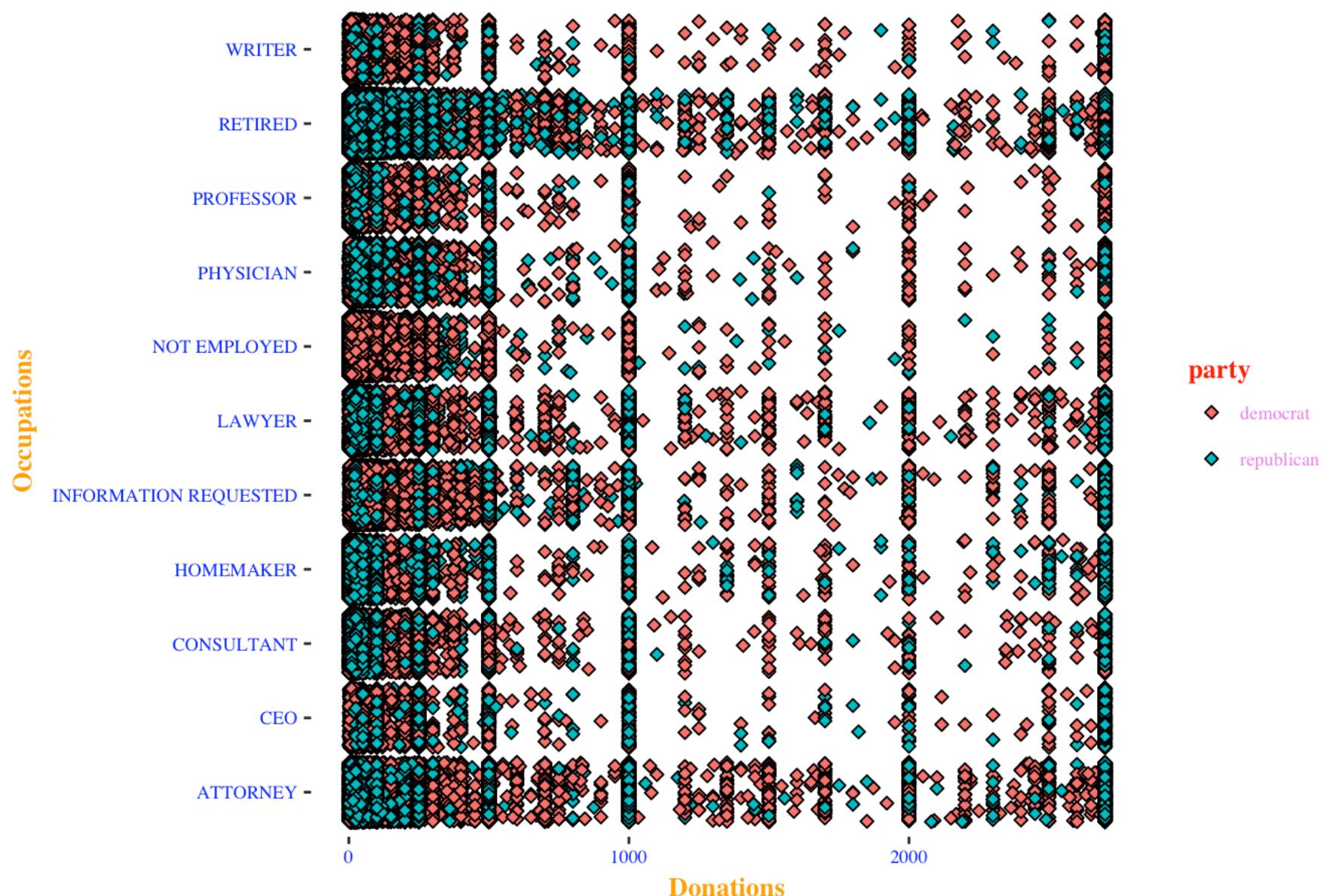
Employers

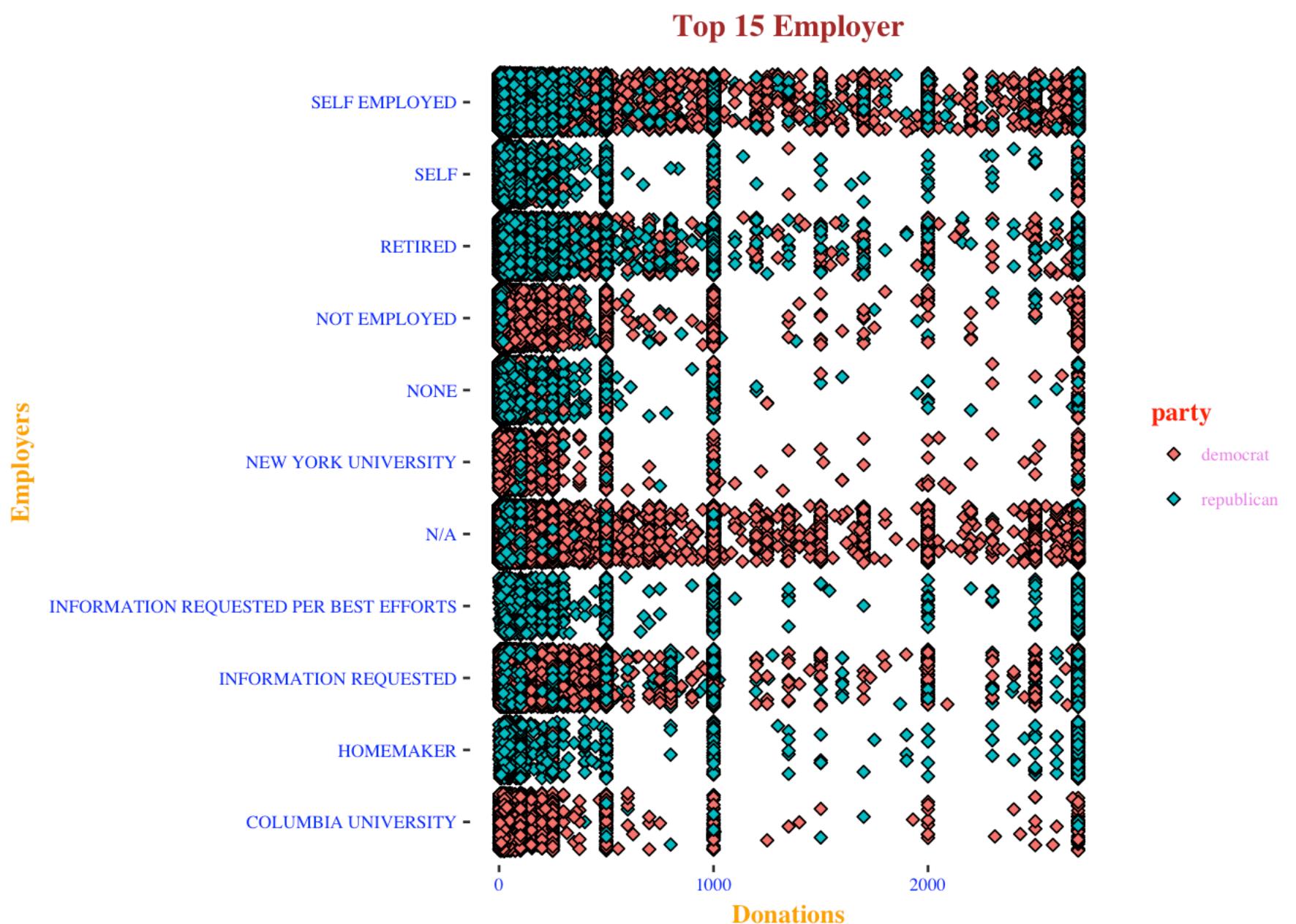


0e+00 5e+06 1e+07

Average Donations

Top 15 Occupations





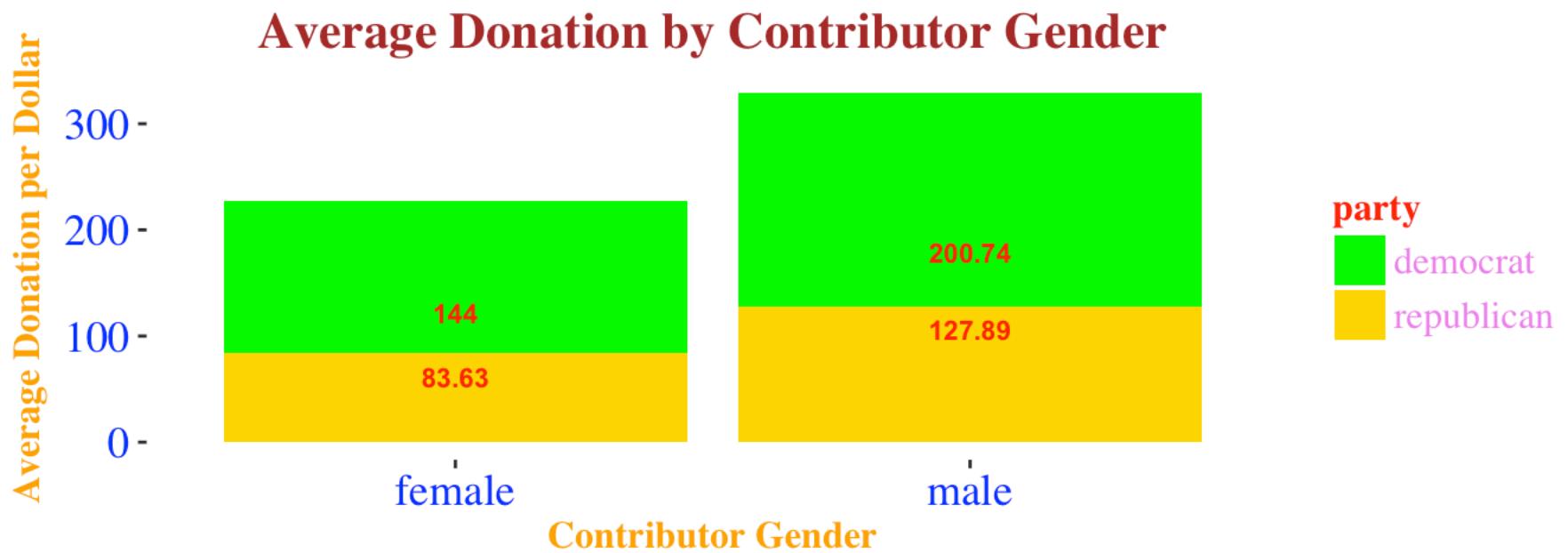
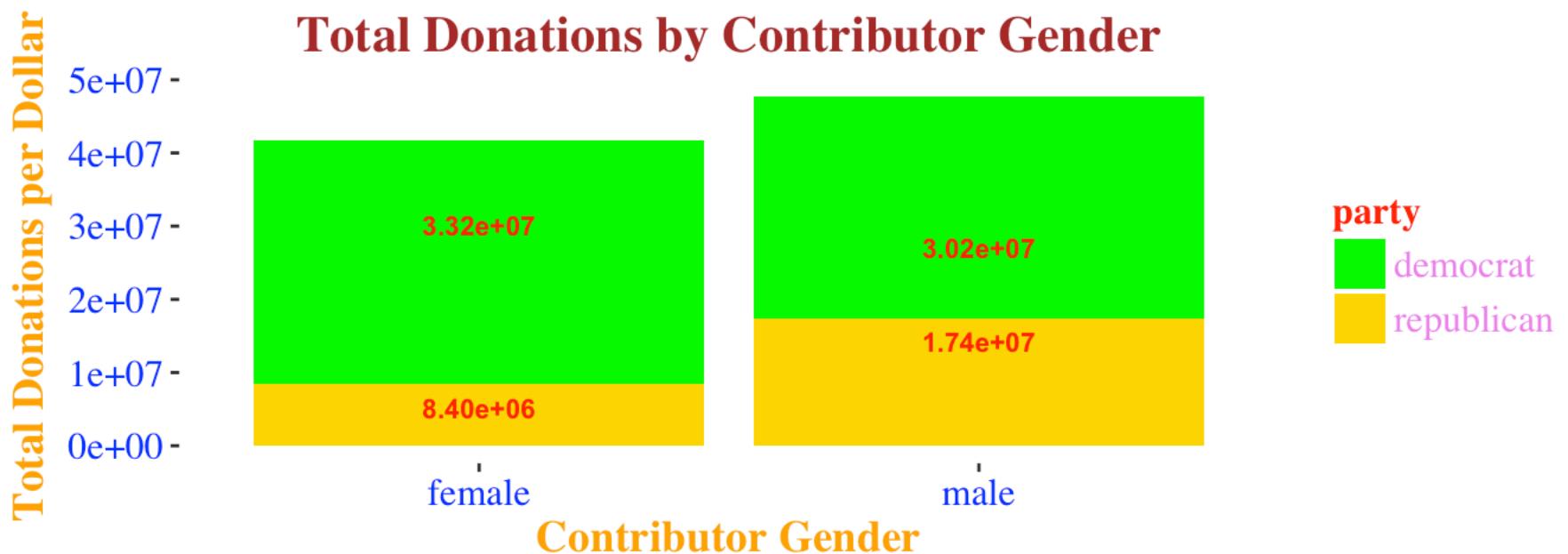
Gender and Contribution

From New York Times article, Crowdpac research indicated “the wage and wealth gap between men and women plays a role [in fund-raising gap]”, and all over the world not only United States where the gender income equality was smaller, more of politicians’ big donors are surprisingly coming from women. And it continues to emphasize on that “women give more to liberals and to other women.” According to the New York presidential campaign financing data, I had managed to predict each donor’s gender by their first names using “genderdata package”. Hence, I was able to verify if the conclusion of the research was true:

Did females tend to donate to female candidates? Did this reflected by the amount donated to political candidates? Were females in favour of donating democratic candidates and/or to women candidates?

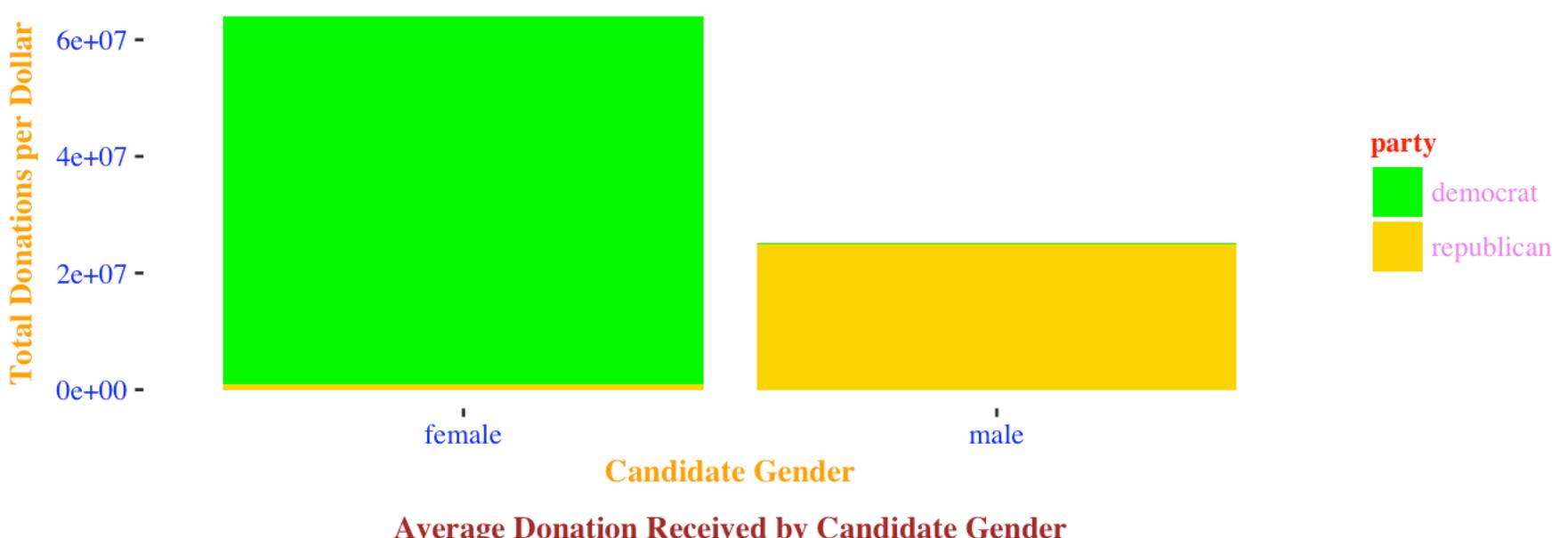
The result suggested that in total and by the average male contributed more than female, t-test also confirmed that female donated less than male in total in such president campaign. It may due to female income equality was smaller. However, the female was more likely to donate to democrats. Surprisingly, male followed the same pattern, which was they were in favor of

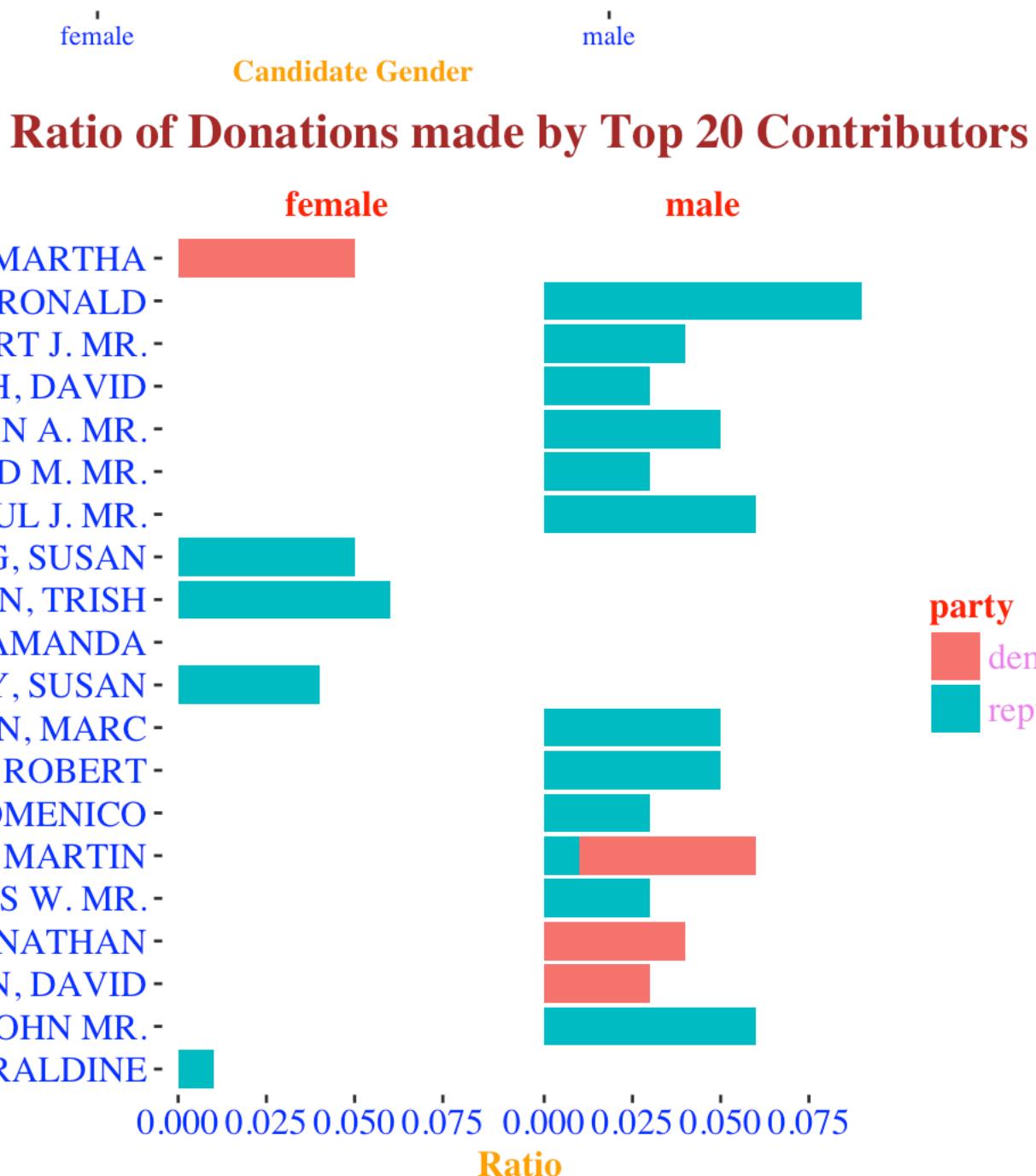
democrats as well and such actions were reflected by in total and average donation. Even in total, male contributed more than female, however, female candidate as well as democrats received more donation. Donation by top 20 contributors denoted female was slightly fond of republican, but that was not a typical representation of the whole female contributors in New York State. Ratio of donation received by candidates told us democrats received almost 80% of total donations, however, as we know it was mostly concentrated on one candidate, Hilary. A scatter plot of “Distribution of Donations” informed us democrats received much more larger amount of donations, in the range of 1000 to 2000 dollars, in addition, they also received more donations in the range of 0 to 1000 dollars.



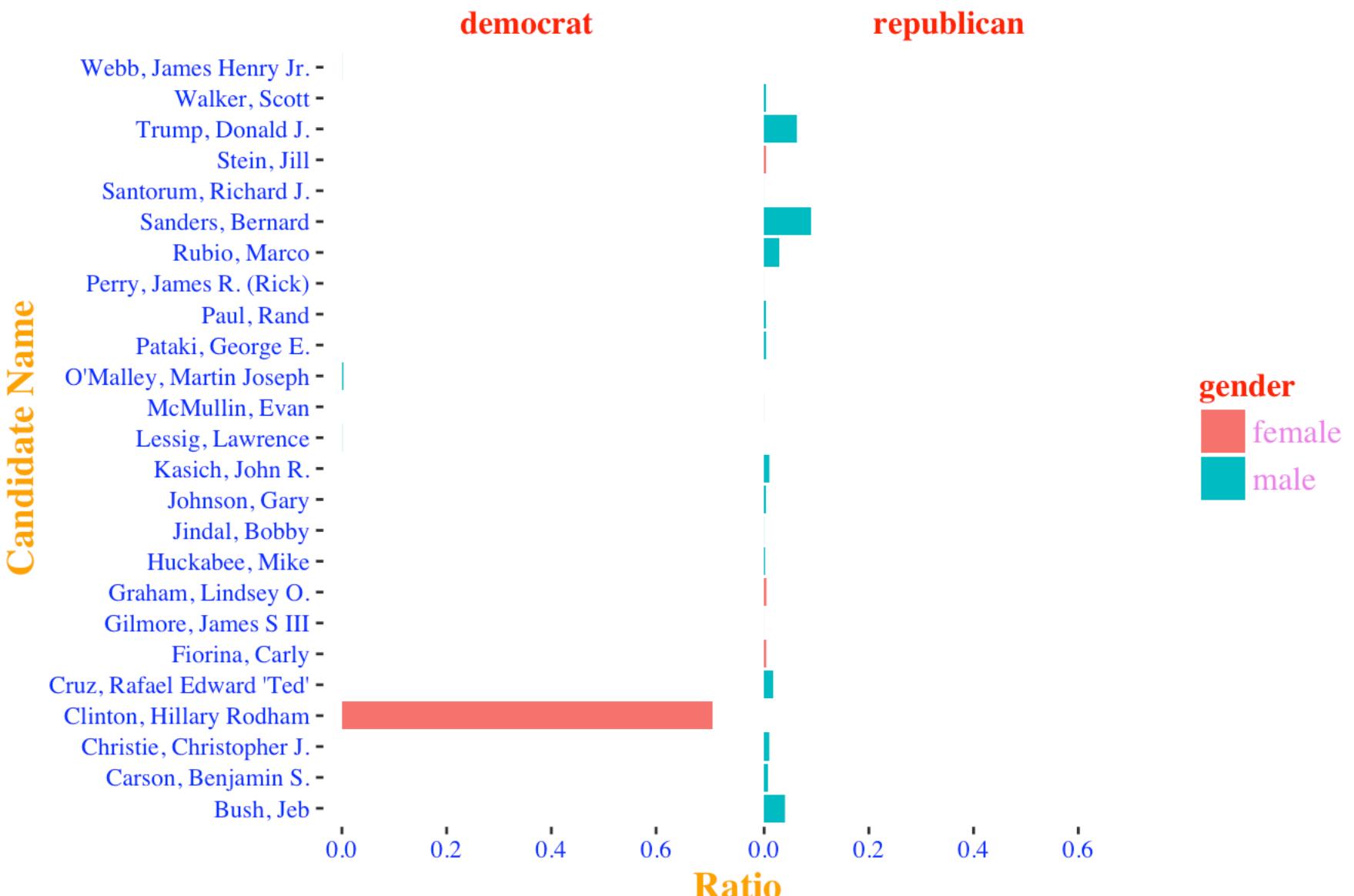
```
##  
## Welch Two Sample t-test  
##  
## data: female$contb_receipt_amt and male$contb_receipt_amt  
## t = -36.682, df = 555950, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
##       -Inf -38.64715  
## sample estimates:  
## mean of x mean of y  
## 125.6927 166.1542
```

Ratio of Donations made by Contributor Gender

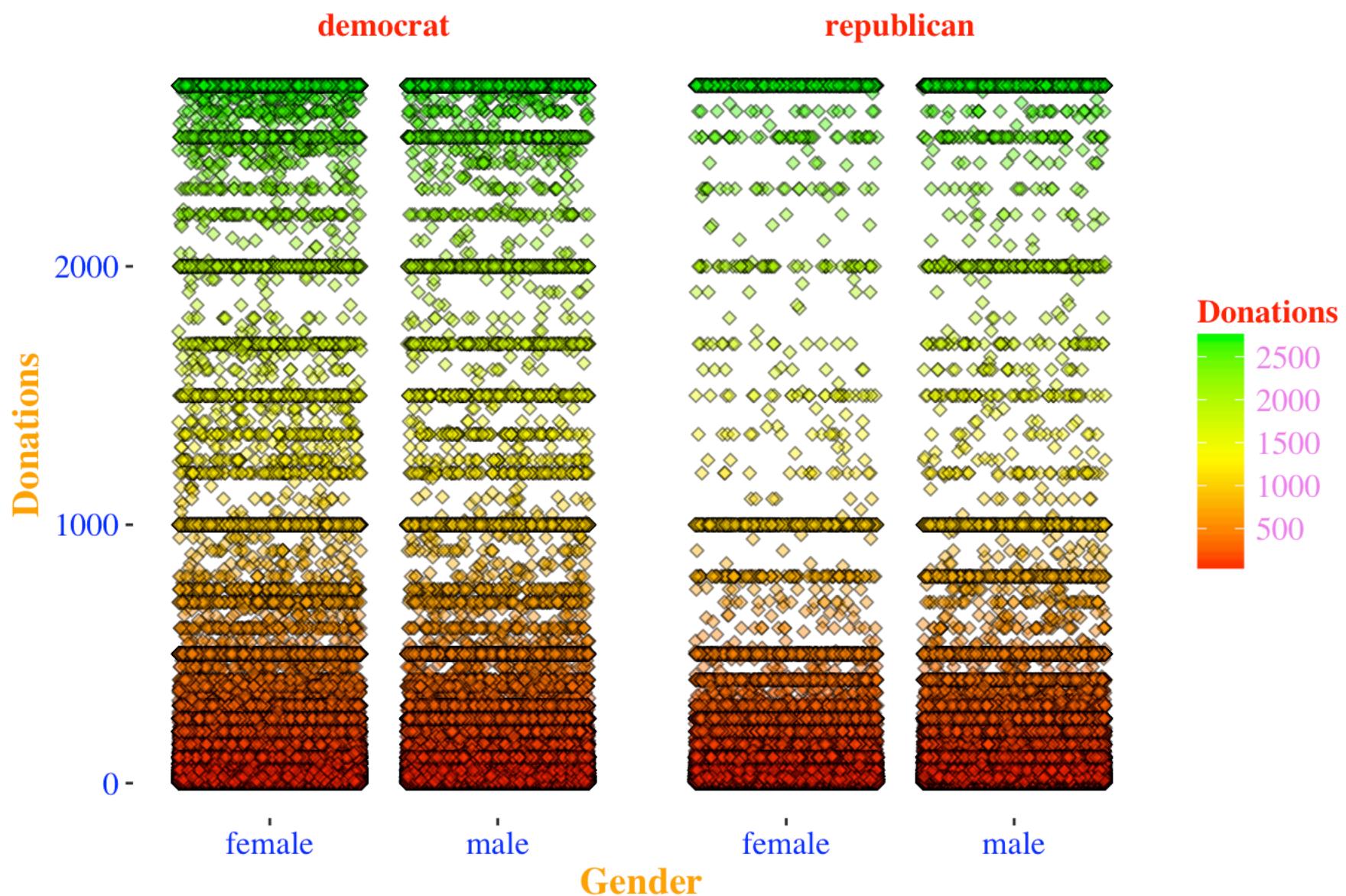




Ratio of Donations Received by Candidates



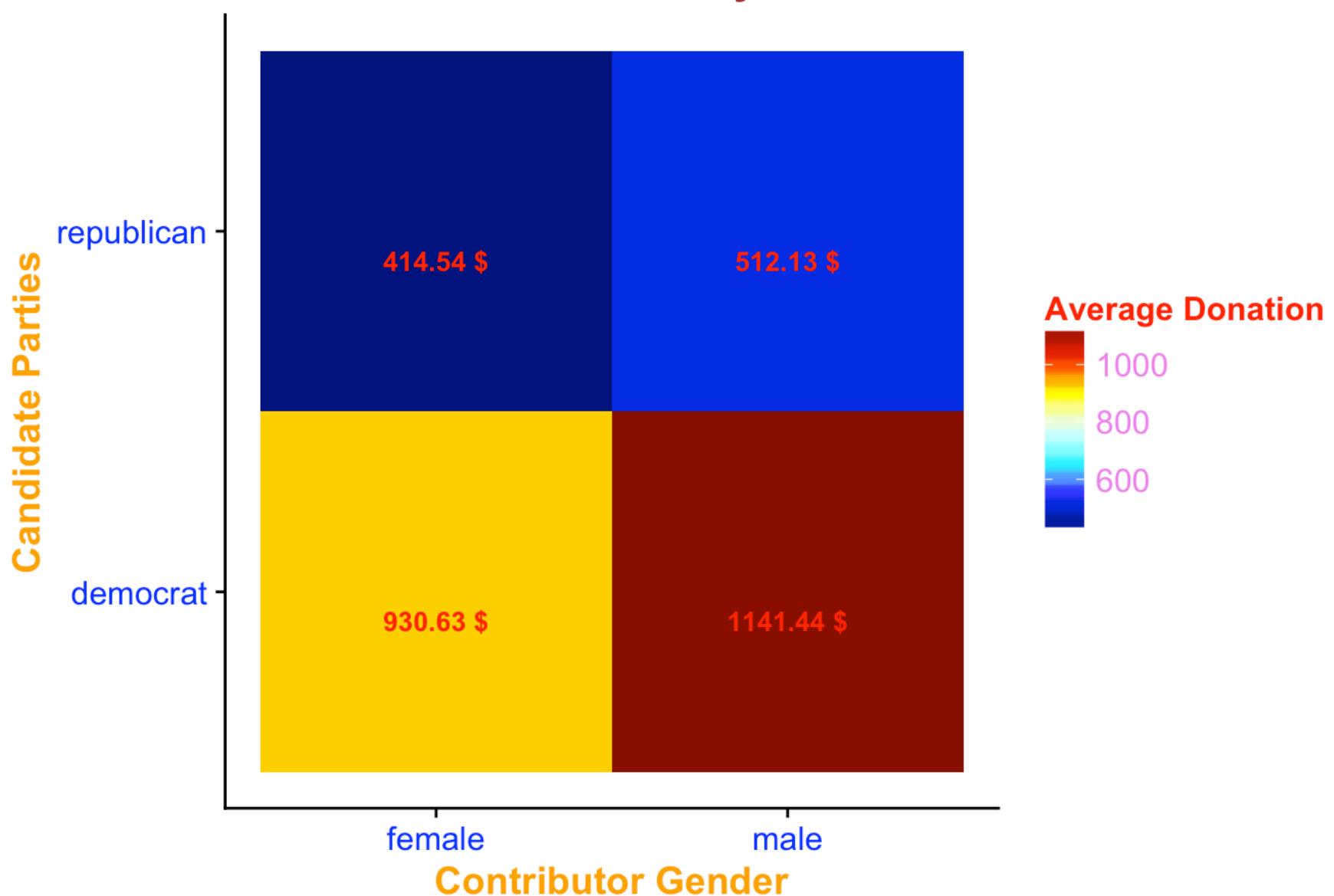
Distribution of Donations by Gender



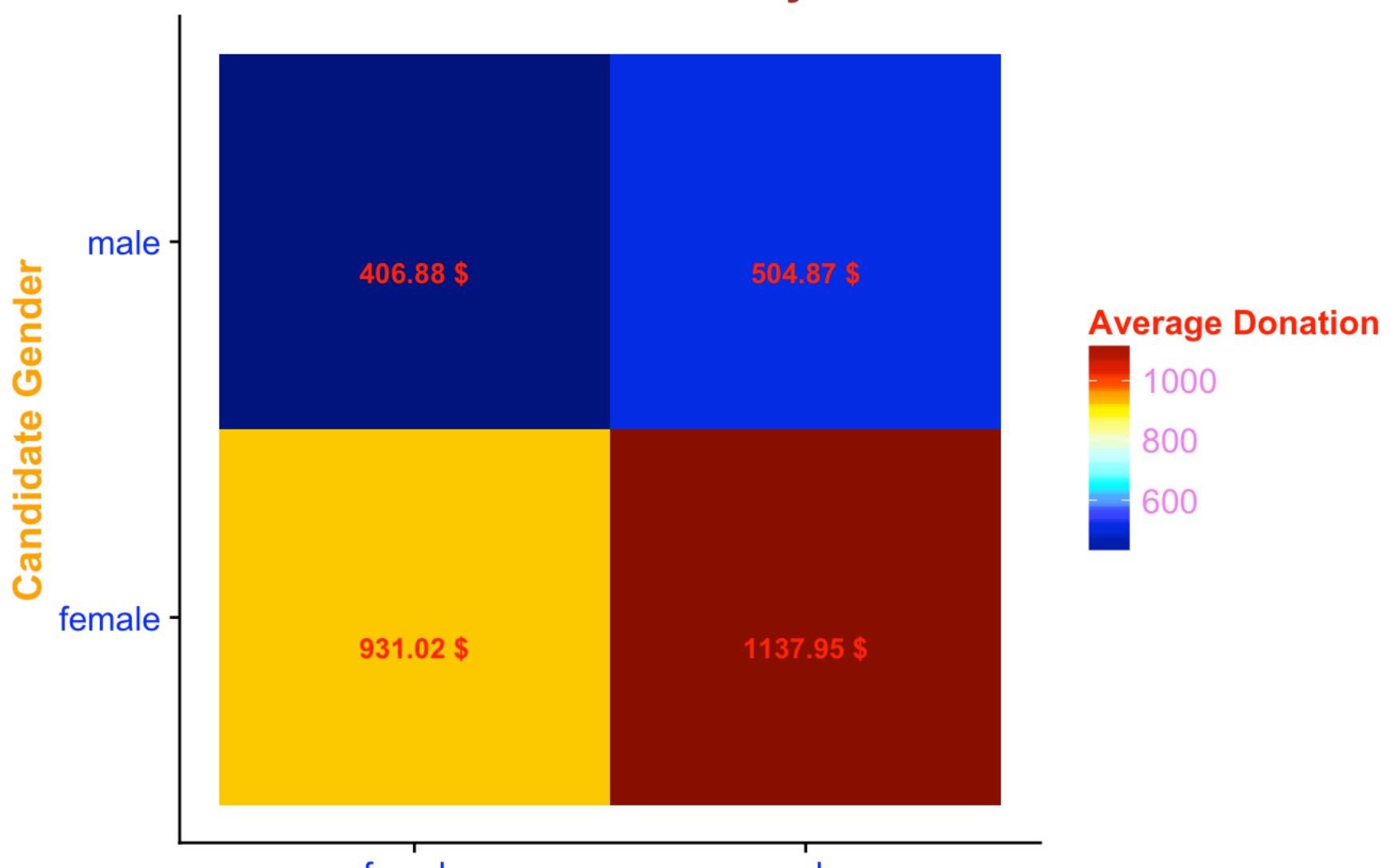
Do women donate to democrats and female candidates?

From previous result, it is not clear whether women tend to donate more to women candidates. The heatmap said female averaged more to democrat than to republican, 930.63\$ versus 414.54\$, more than doubled. In addition, men followed the same pattern, 1141.44\$ versus 512.14\$, those findings were in accordance with our previous results. When it comes to the question whether women donate more to female candidates, I found the answer is “yes”. Female donors tend to donate more on average (931.02\$ to female candidates versus 406.88\$ to male candidates) to female candidates, while men donated more to male candidates more(1137.95\$ to female candidates versus 504.88\$ to male candidates), and cross-gender donation amounts tend to be lower than same-sex. Though the result is clear, when I look at the gender distribution of candidates, realized that female candidates are only 4 out of 25, that means donations among republicans are much more competitive than democrats. More important, Hilary alone possessed almost 80% of total donations in New York State, thus such result is less convincing if women or even men are prone to vote for women as well as if they are more fond democrats in New York State.

Donation to Parties by Gender



Donation to Candidates by Gender



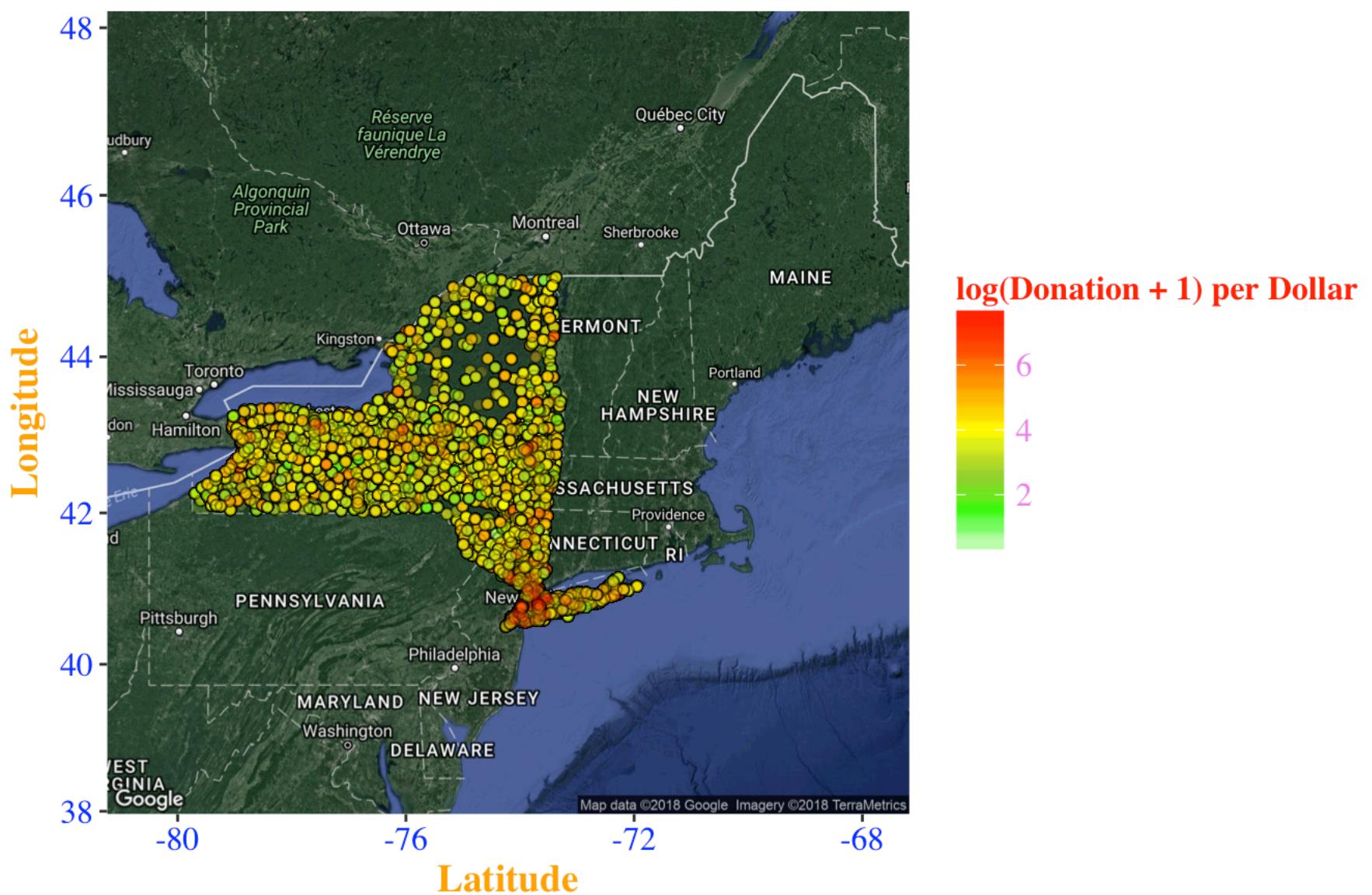
female male
Contributor Gender

```
## # A tibble: 2 x 2
##   cand_gender gender_num
##   <chr>          <int>
## 1 female           4
## 2 male            21
```

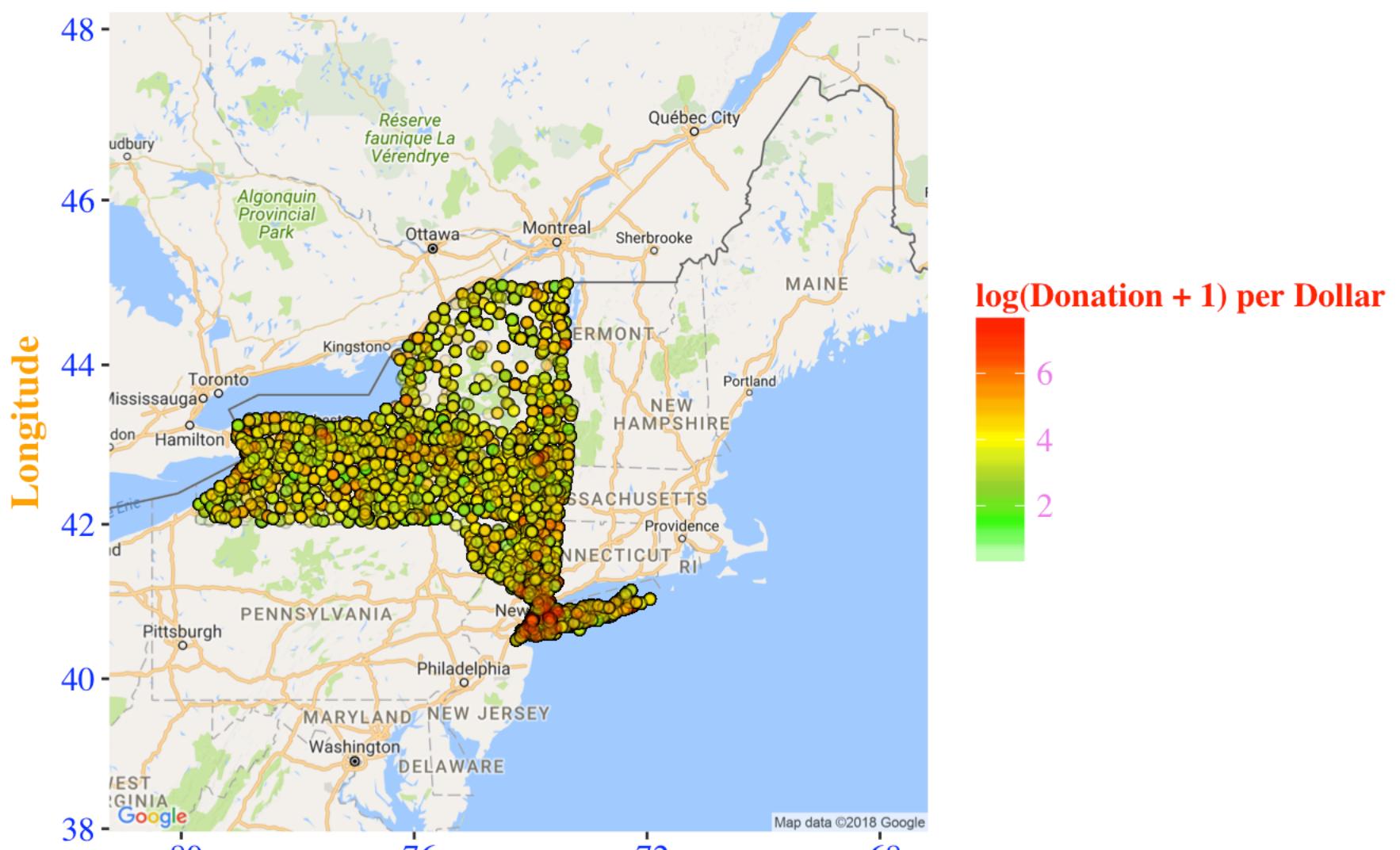
Where did the money came from

Prior to know the distribution of donors, I need to prepare different types of maps, and select the those with best visualization. ggmap package has provided various types of maps with specific visualization, including hybrid, terrain, satellite, tonner, etc. I tried each of them, and decided to improve my result with hybrid, terrian_labels and roadmap. From the distribution of donors on the map, it is not surprising to see that most of the donors as well as high donation are concentrated in the big cities, especially more concentrated on long island. But what about the donation amounts by geographical locations, do big cities tend to donate more per capita than smaller ones? I need to aggregate donation amounts and donation per capita by counties for comparison.

The Geographical Location of Donors



The Geographical Location of Donors



-80

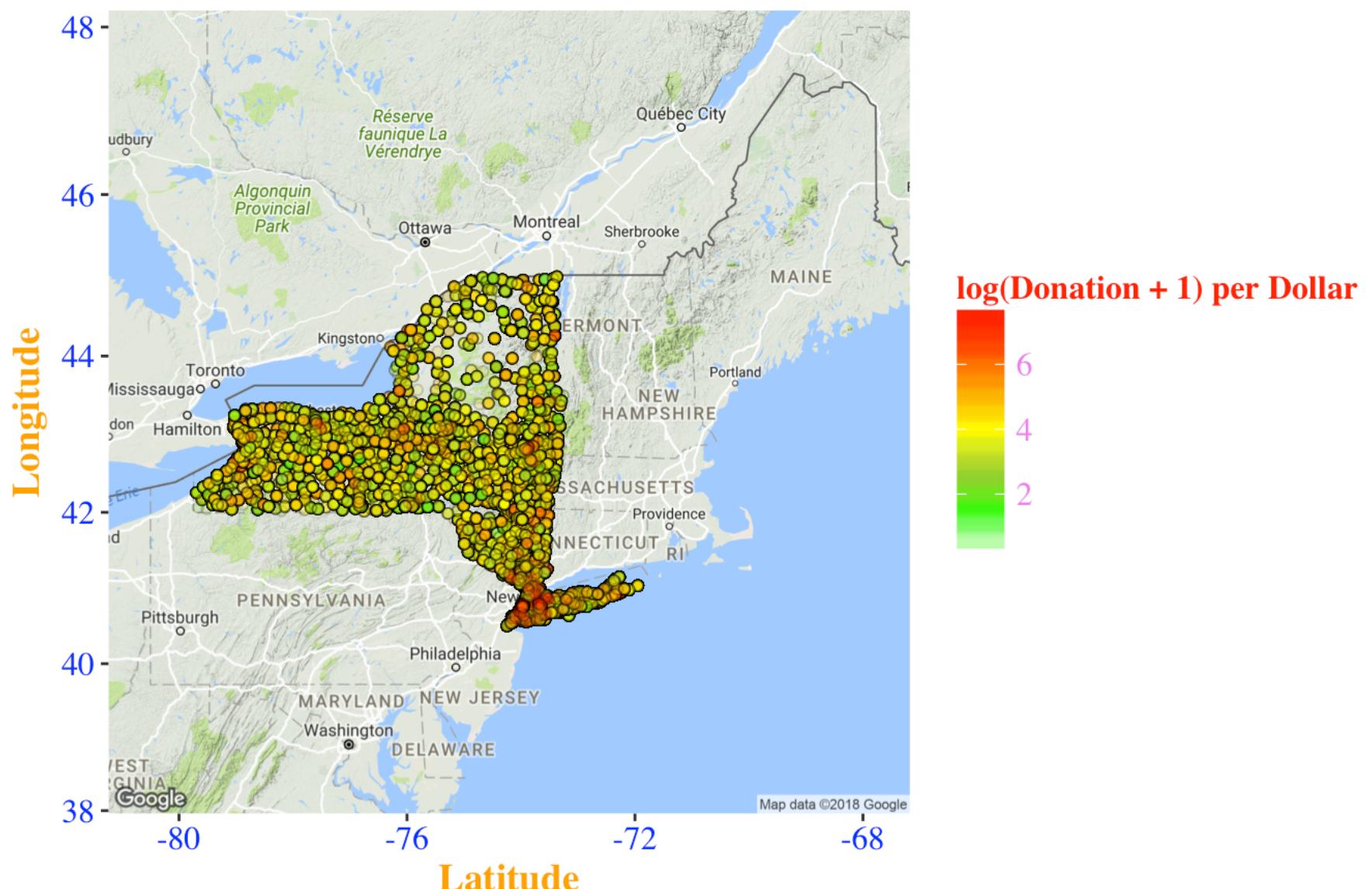
-76

-72

-68

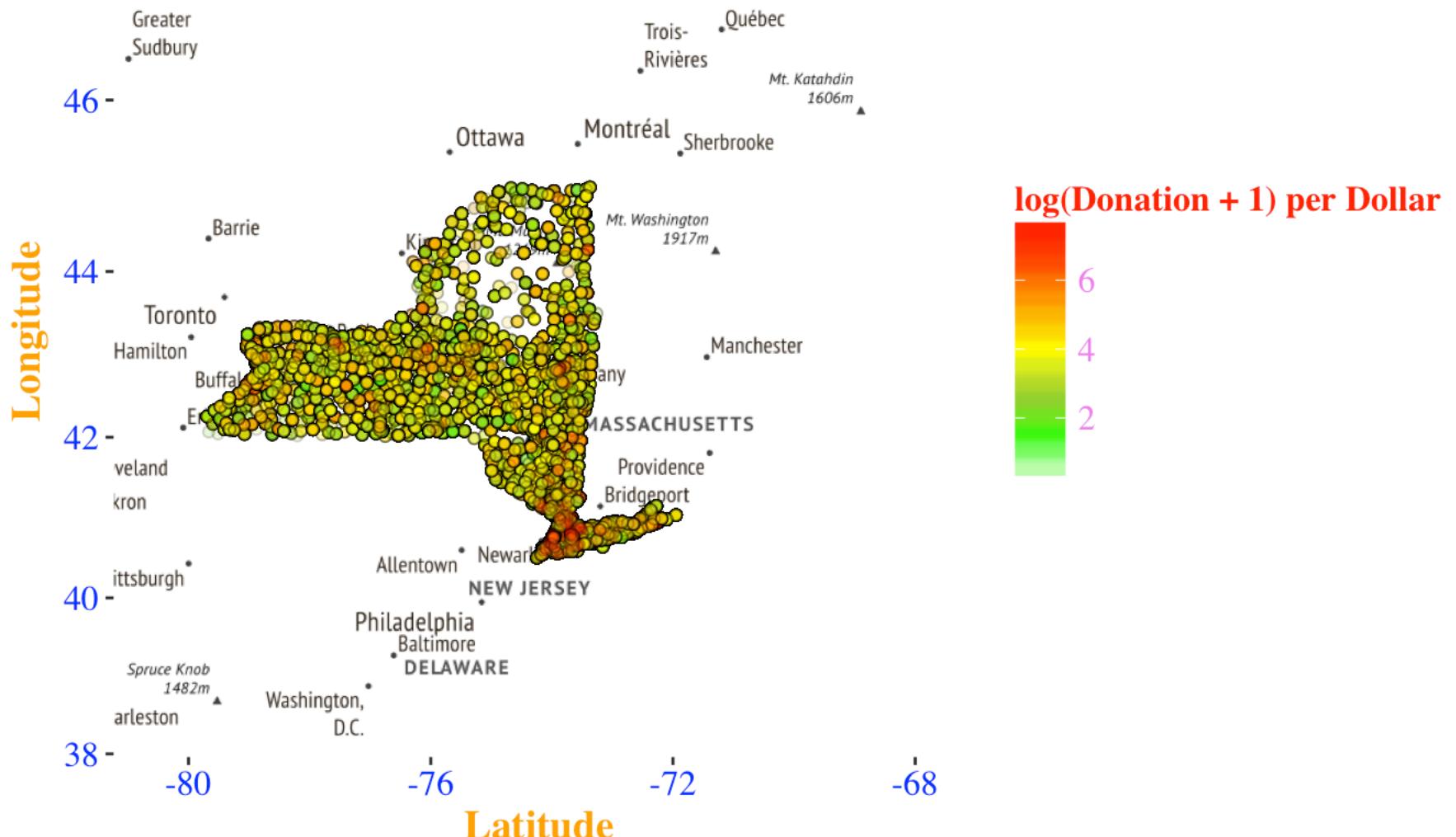
Latitude

The Geographical Location of Donors

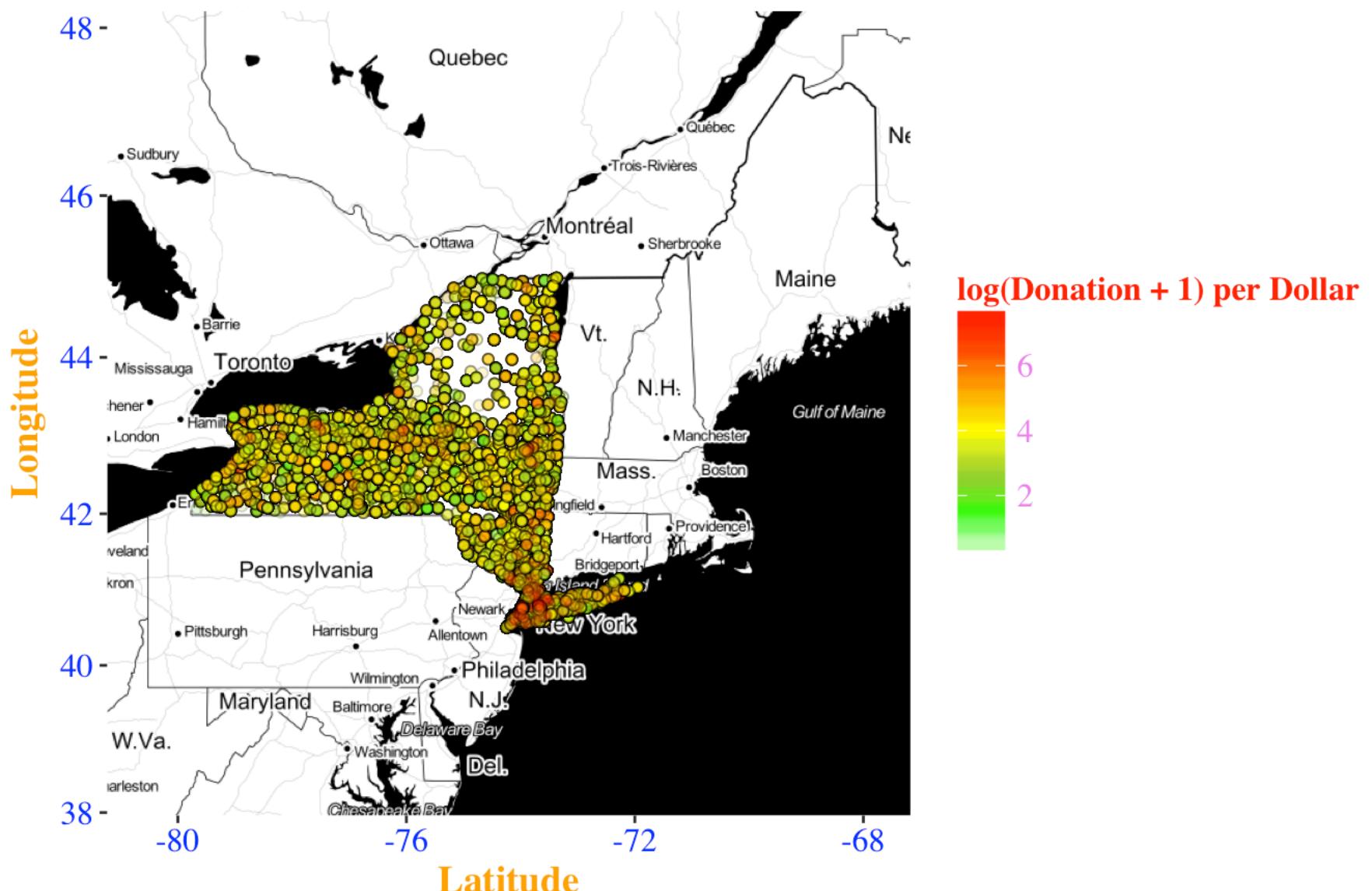


The Geographical Location of Donors

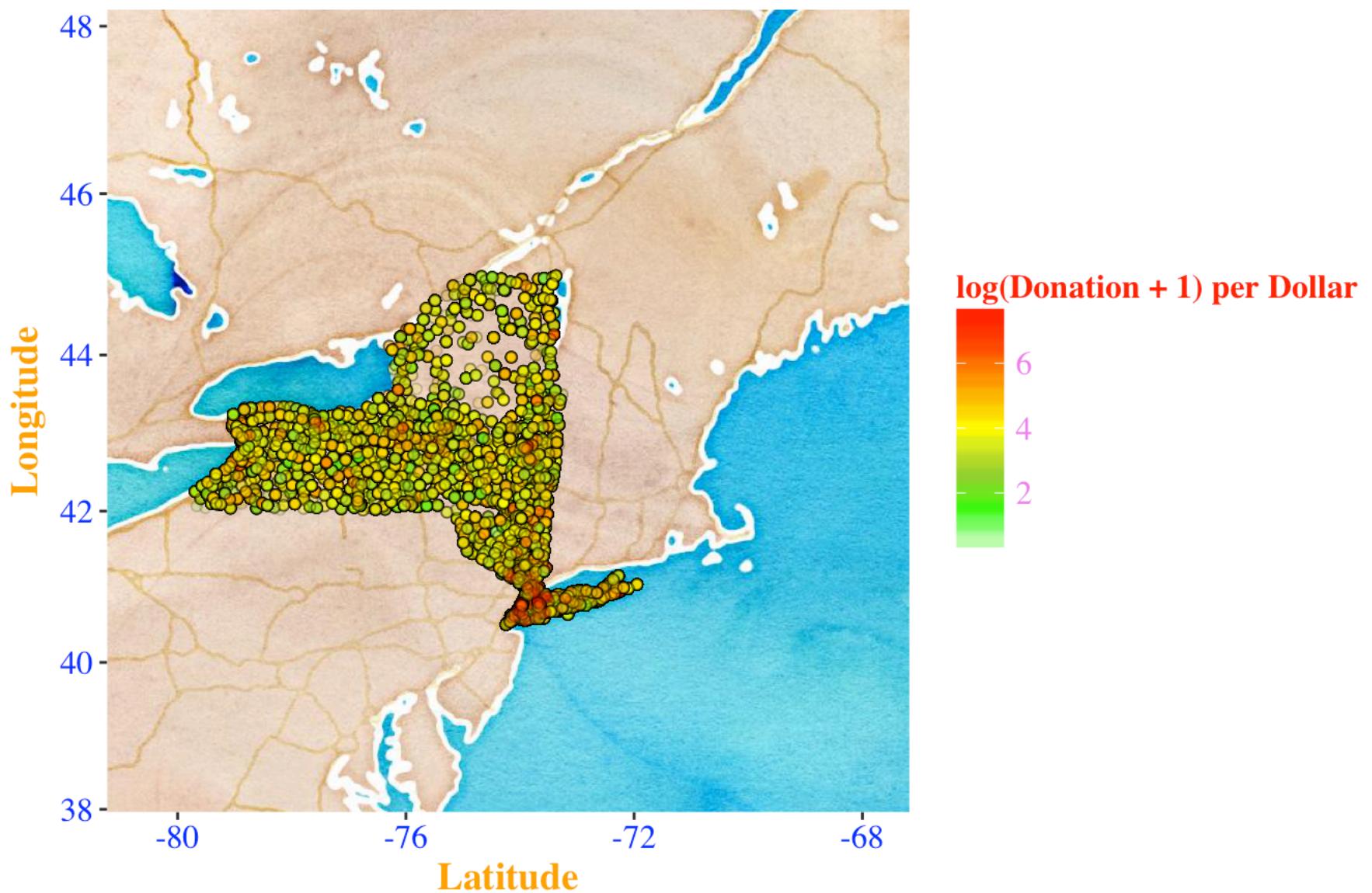
48 -



The Geographical Location of Donors



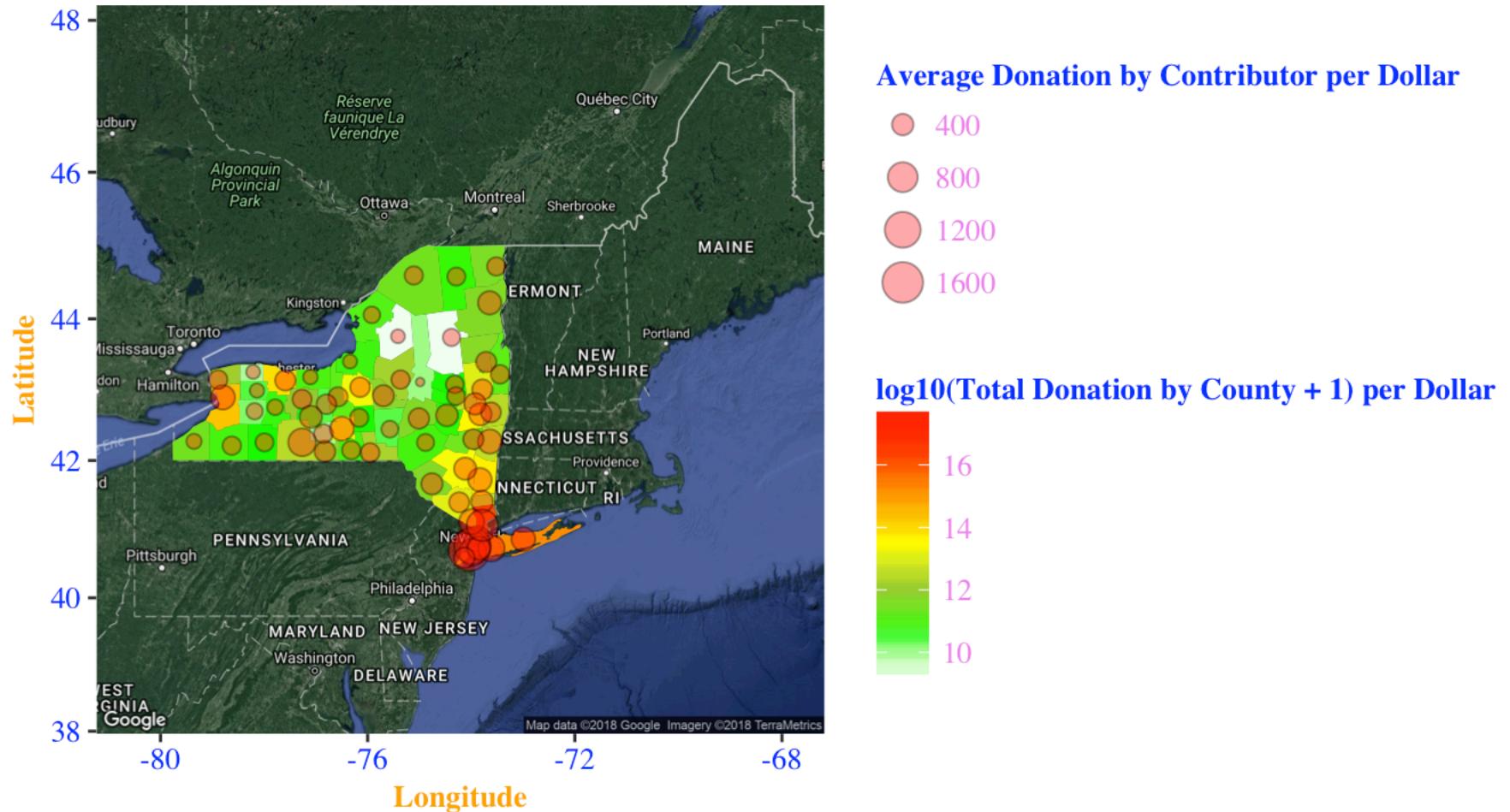
The Geographical Location of Donors



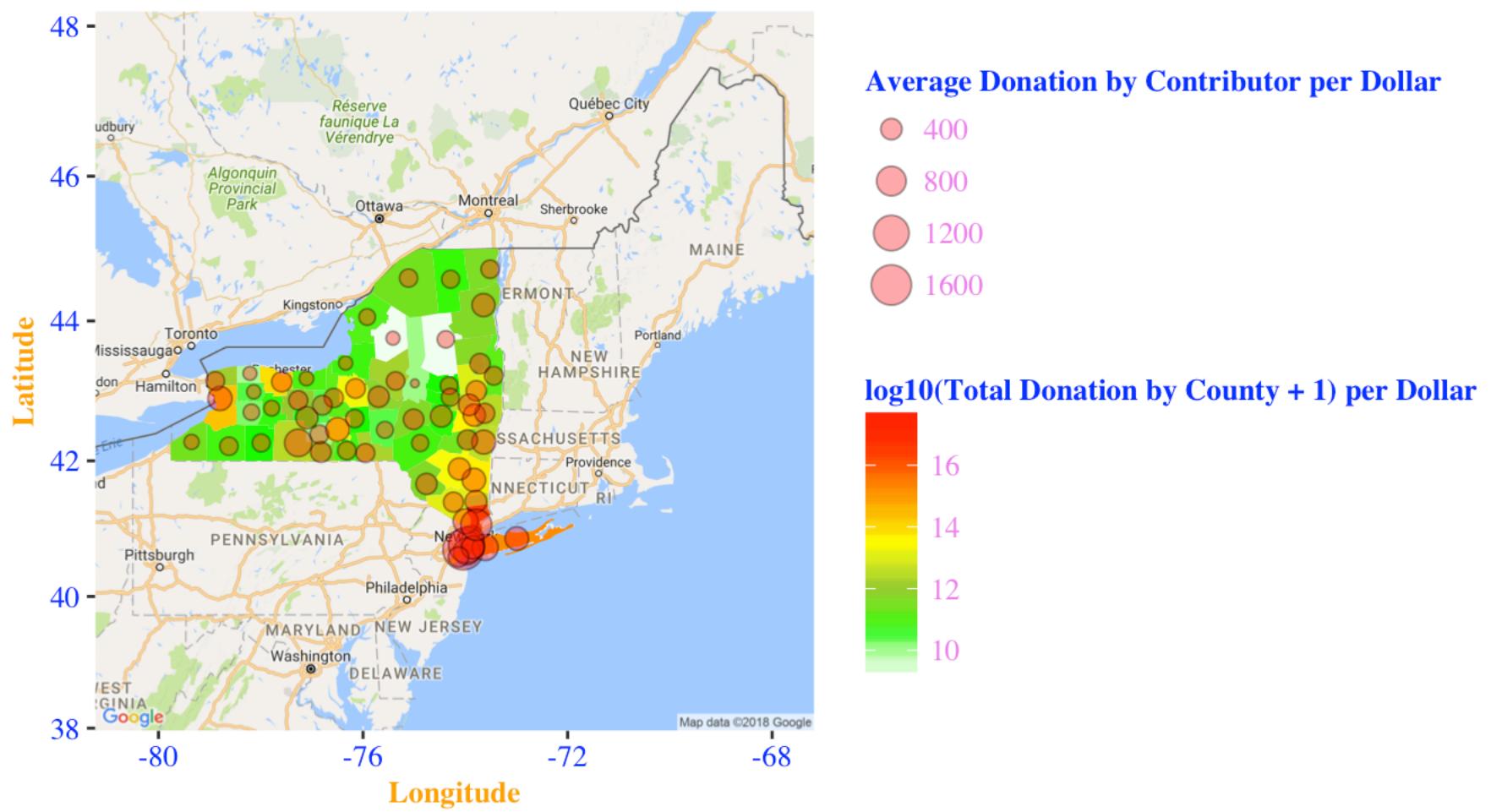
Total and average donation by county

Donations tended to concentrate on a more prosperous counties. The total donation amount is the highest in the long island as mentioned above, as those areas lived the most rich people as well as are most densely populated region in New York State, however, more data is needed to determine this.

The Geographical Location of Donors by County and Average Donations

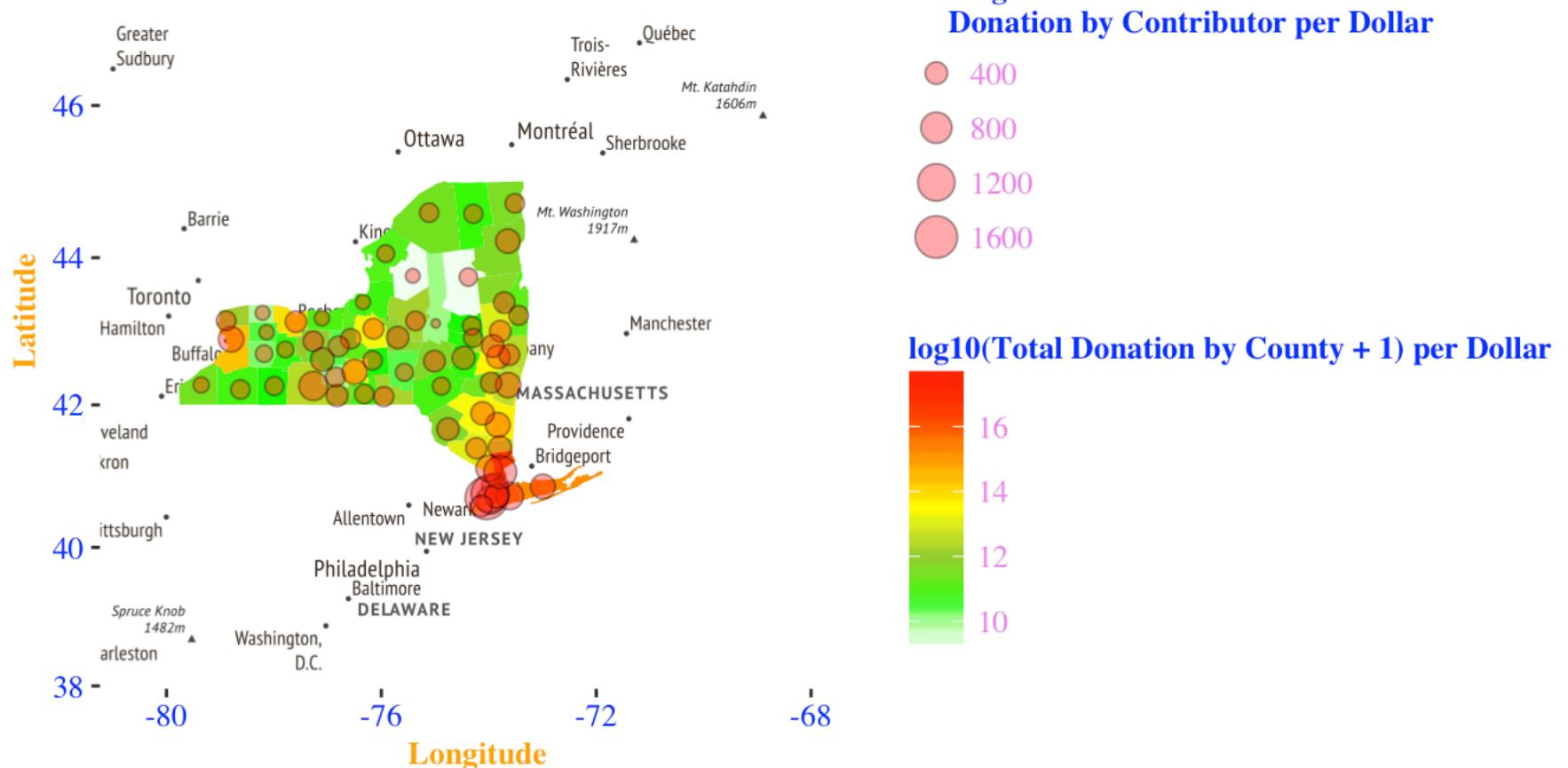


The Geographical Location of Donors by County and Average Donations



The Geographical Location of Donors by County and Average Donations

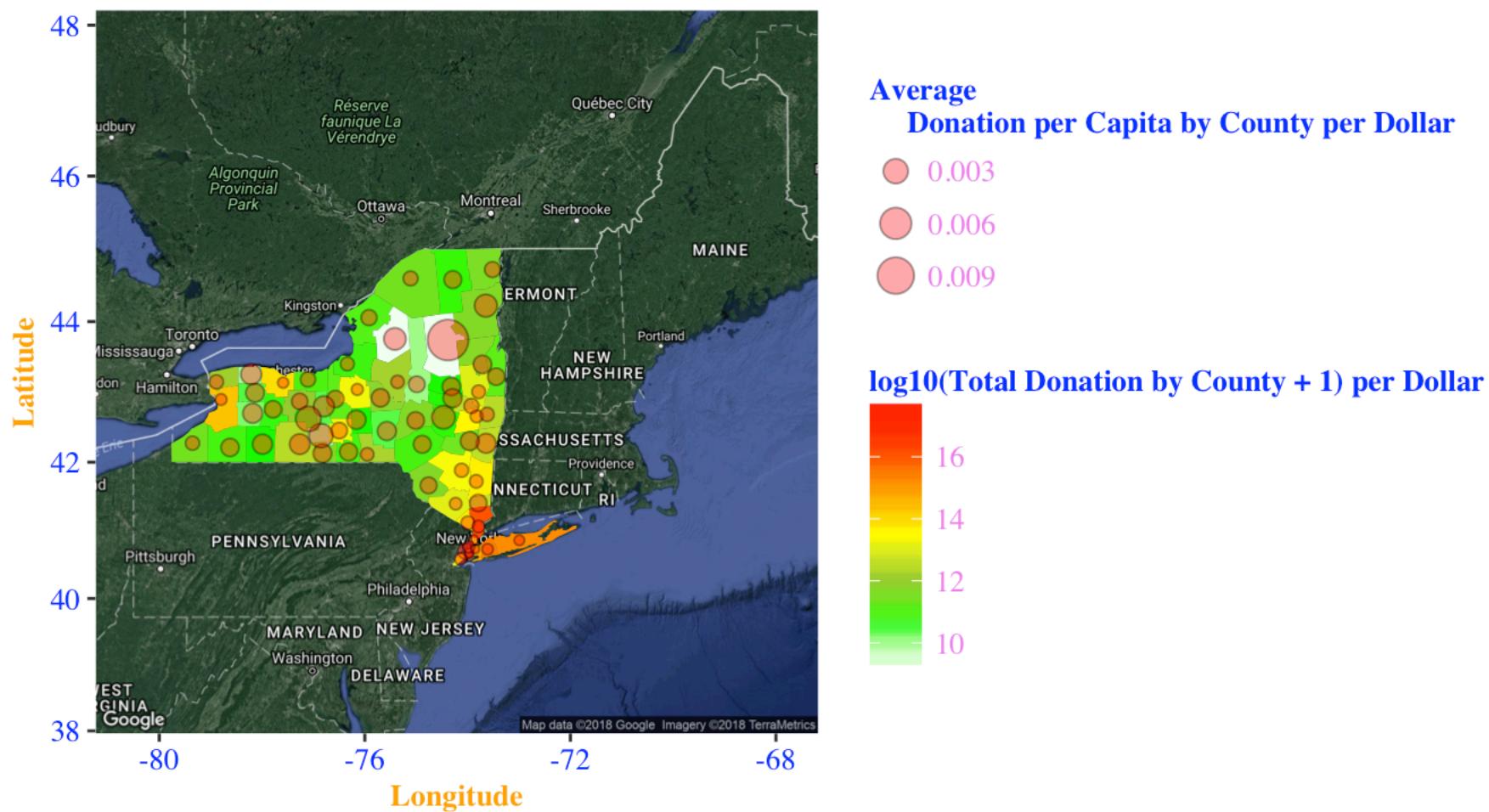
48 -



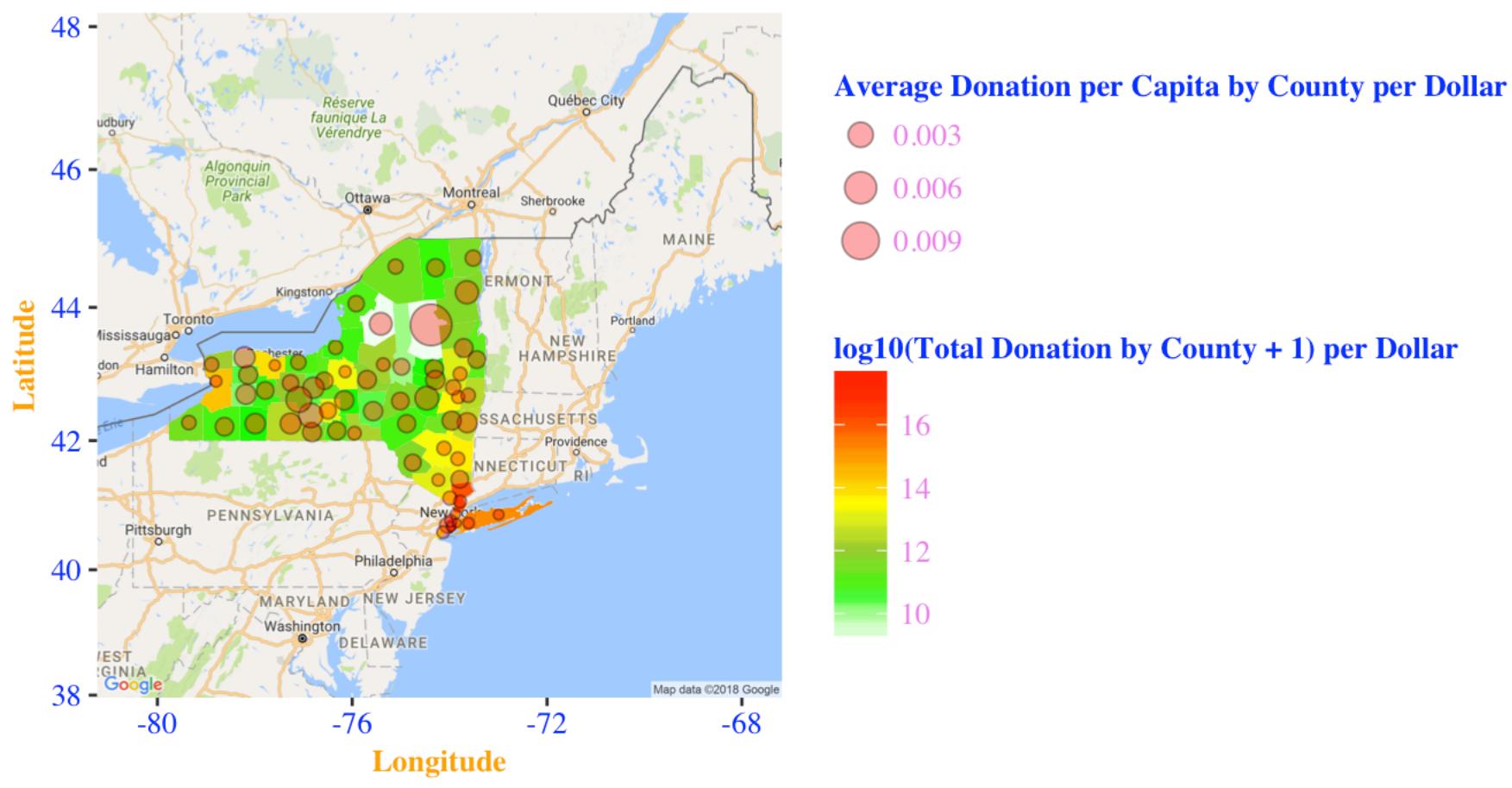
Total and average demographic donation by county

Looking at the average demographic donation by county, I have different findings, that the highest donation per capita was quite different from that of average donation per contributor. The highest number of donation per capita occurred in counties of Lewis and Hamilton. According to the census data in 2016, their population in Hamilton and Lewis were only 4542 and 26865, respectively. There are in total 62 counties in New York State, and the population ranks of Hamilton and Lewis were 62th and 59th, respectively, that explains why those two states had higher average donation per capita.

The Geographical Location of Donors by County and Average Donations per Capita

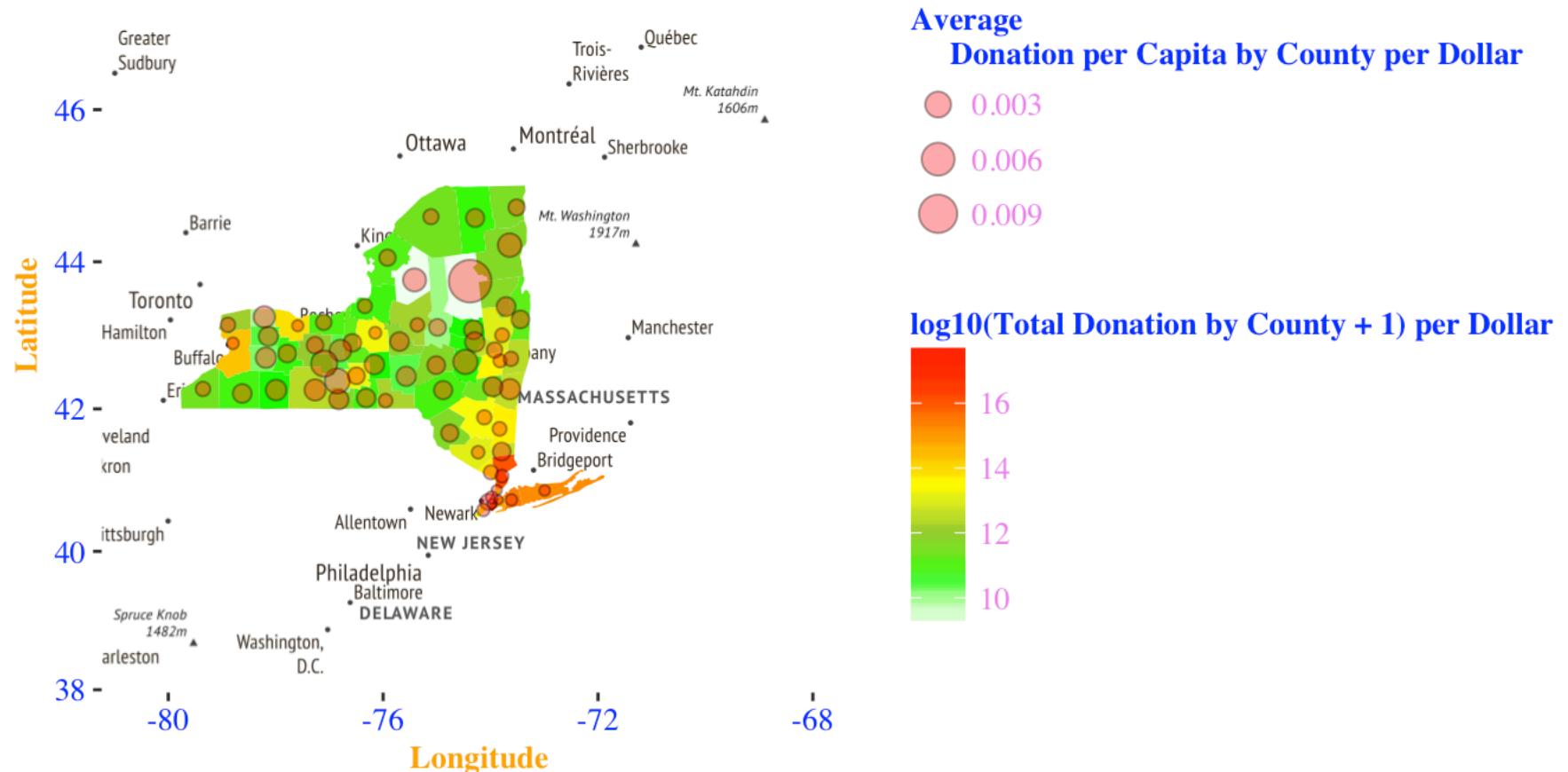


The Geographical Location of Donors by County and Average Donations per Capita



The Geographical Location of Donors by County and Average Donations per Capita

48 -

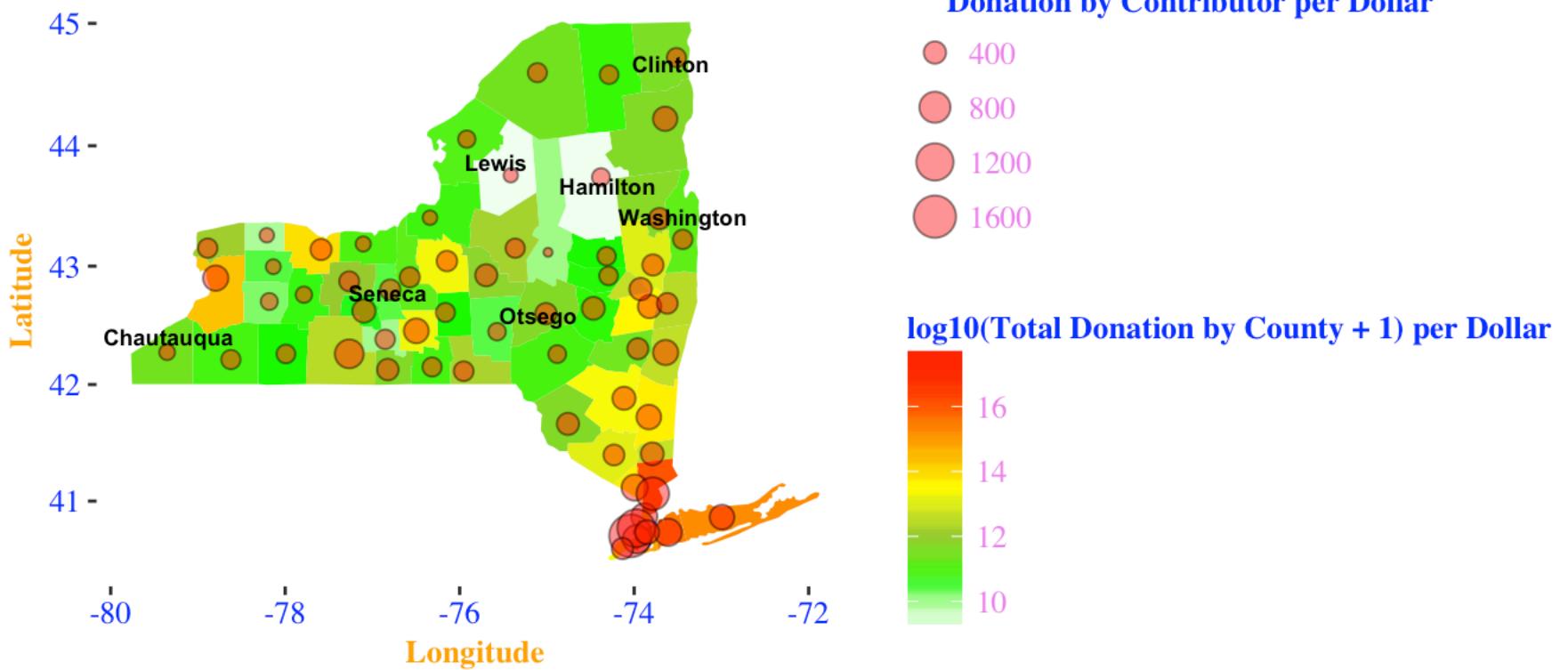


Reconstruct the map with CRAN packages to zoom in and see more details

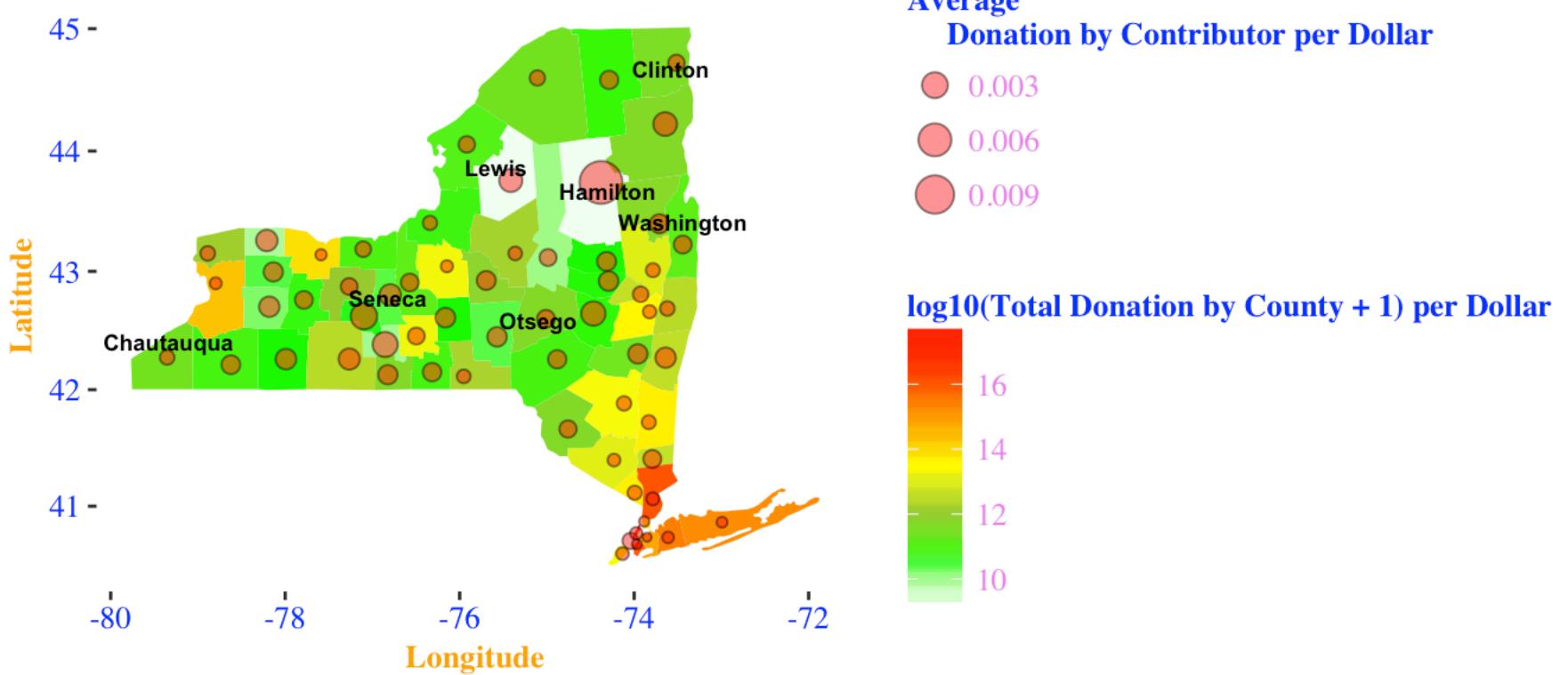
GGmap provides vivid visualizations, however, all the map images it can provide are static, that means for specific location such as my project which focus on New York State, it is either bigger or smaller according to specific requirements. In addition, if we crop the bigger size map, substantial resolution would be lost. Thus, if I'd like to see more details from the map, it is inevitable to recreate the map. Luckily, with zipcode package and geom_polygon function, such purpose can be achieved.

Average donation by contributor tended to occur in prosperous counties, whereas average donation per capita largely depends on the population, such as, even the average donations by contributor were quite small (around 400 per dollar) in Hamilton and Lewis, however, their average donations per capita were even the largest among all the counties in New York State. As mentioned above, that was due to the small population numbers in those two counties, and thus such data was not reliable as compared to average donation per contributor.

The Geographical Location of Donors by County and Average Donations



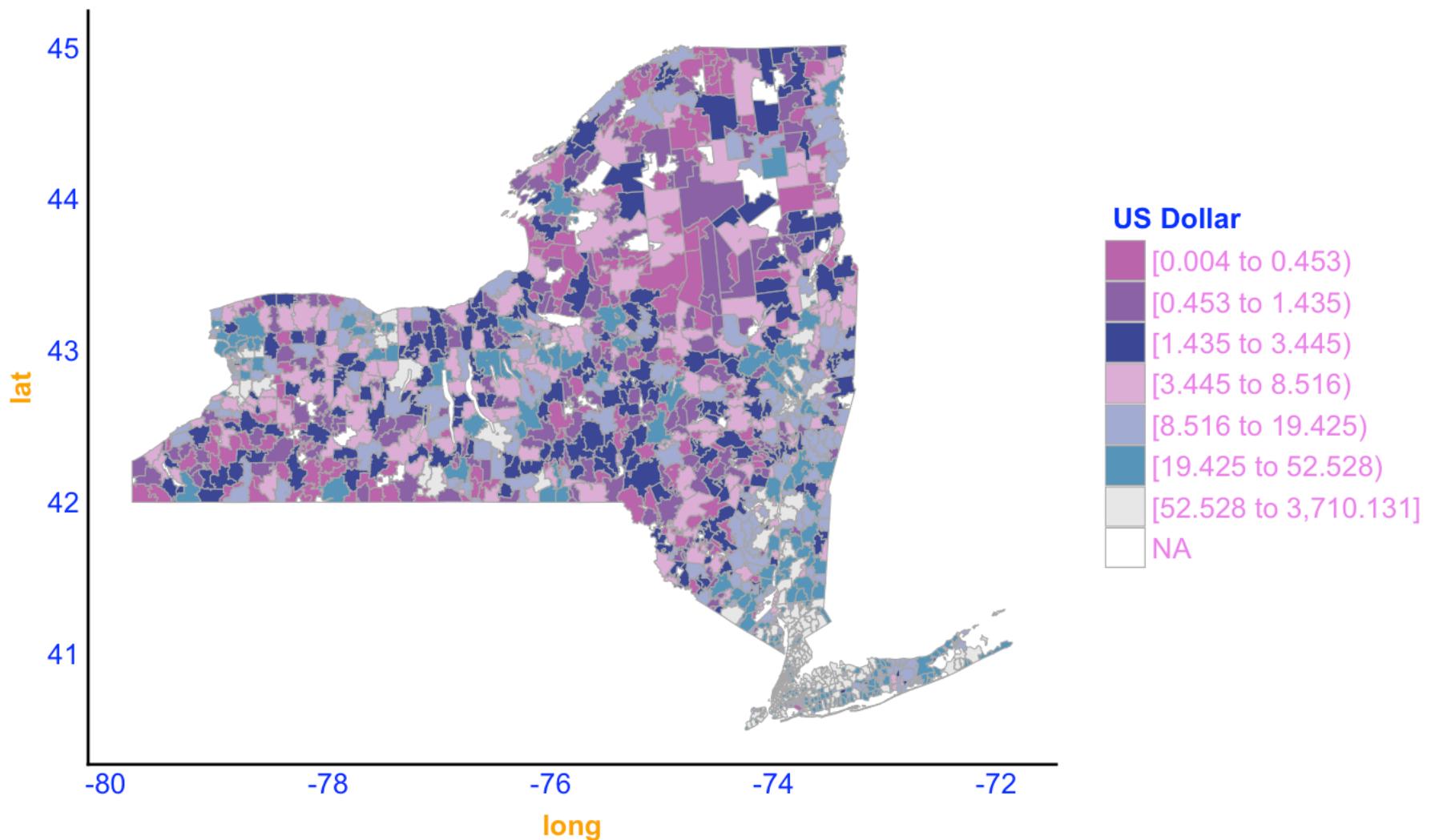
The Geographical Location of Donors by County and Average Donations



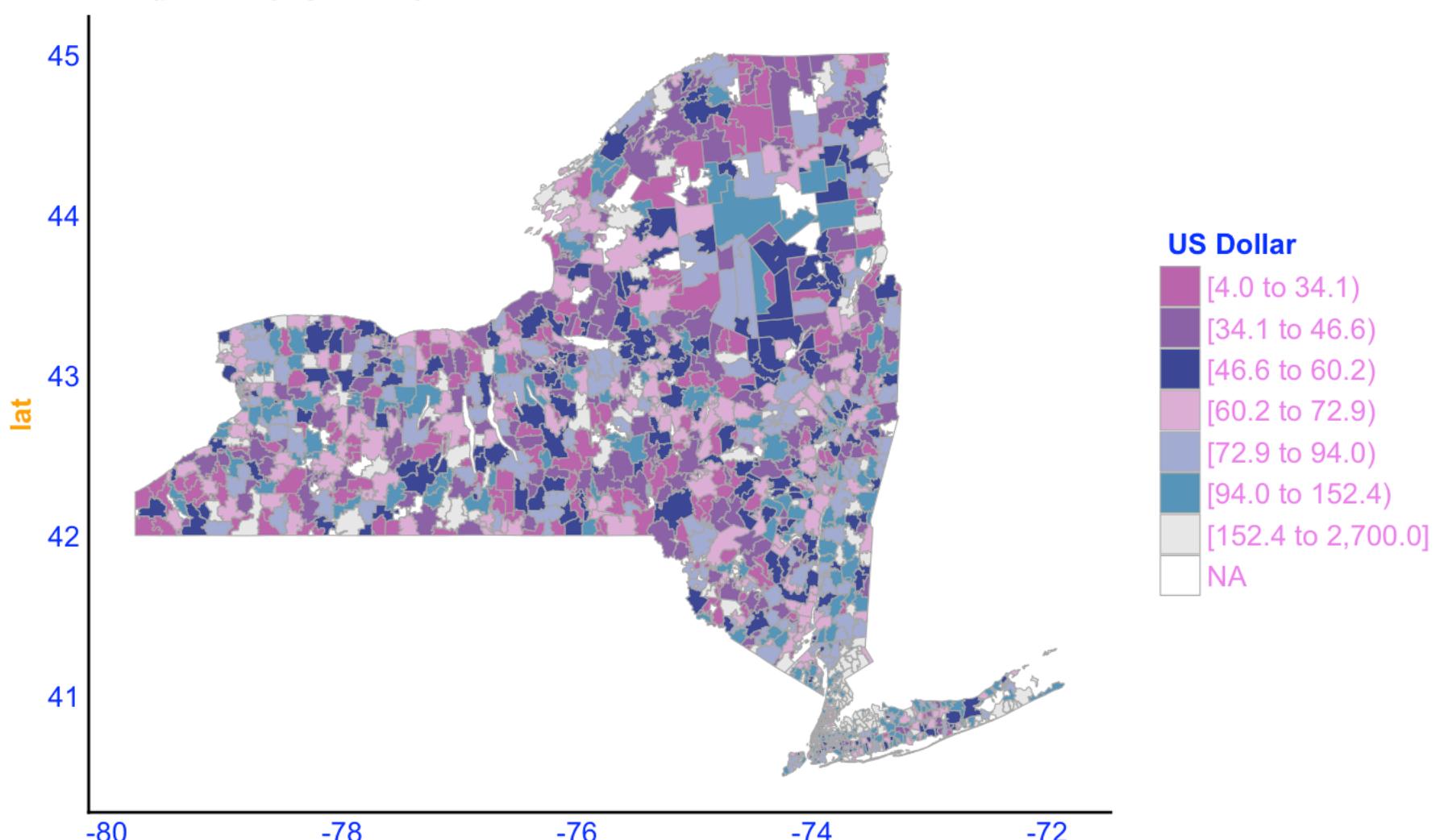
Better Visualization

In order to maximumly visualize the features of donations in New York State, I decide to apply an external package called choropleth. A choropleth map displays divided geographical areas or regions that are coloured, shaded or patterned in relation to a data variable. It allows to study how a variable evolves along a territory.

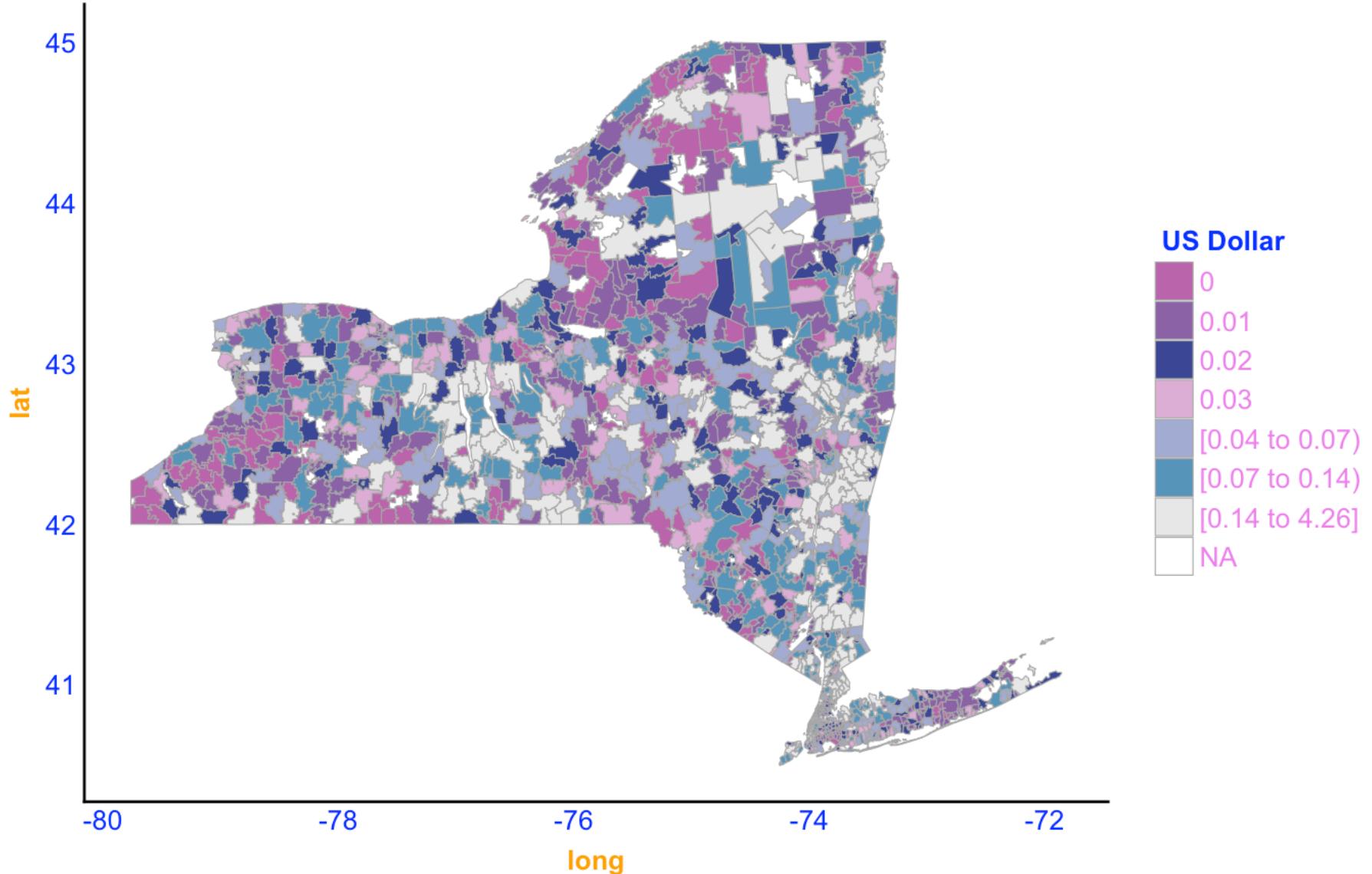
**Total Donation
(thousands USD) by County in New York State, 2016 Presidential Election**



**Average Donation per Contributors
(per USD) by County in New York State, 2016 Presidential Election**



Average Donation per Capita (per USD) by County in New York State, 2016 Presidential Election



Final: Shall we trust the average donation per contributors or capita?

As previously discussed, even total donations in the counties of sparse population such as Hamilton and Lewis were low, however, their donations per capita were surprisingly high among all the counties in New York. That was due to Simpson's paradox, which would misleading analysis. In this final plot, I'd like to confirm this point in a densely populated area, Manhattan.

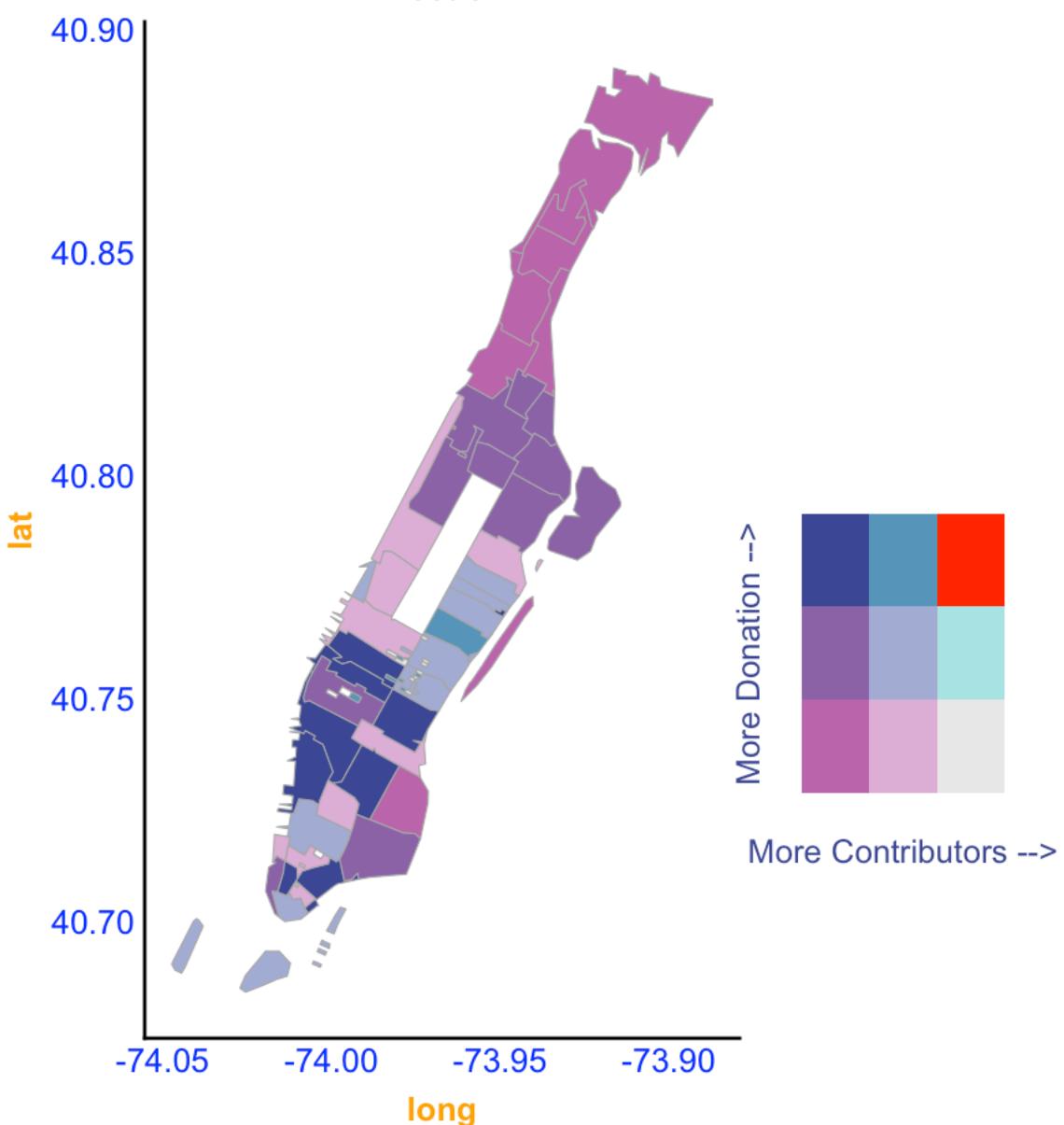
Prior to that, I need to use some technical assistance to transform the legendary label from pure bins of numbers to the relationship, such as average donation versus contributors or population.

It seems ggplot doesn't really have the capacity for a bivariate legend. However, Joshua Stevens shows nice plots from his website by applying a dedicated graphic composition program. We can definitely cobble something similar up in R with the help of 'cowplot' package' which allows for the creation of a layered plotting canvas.

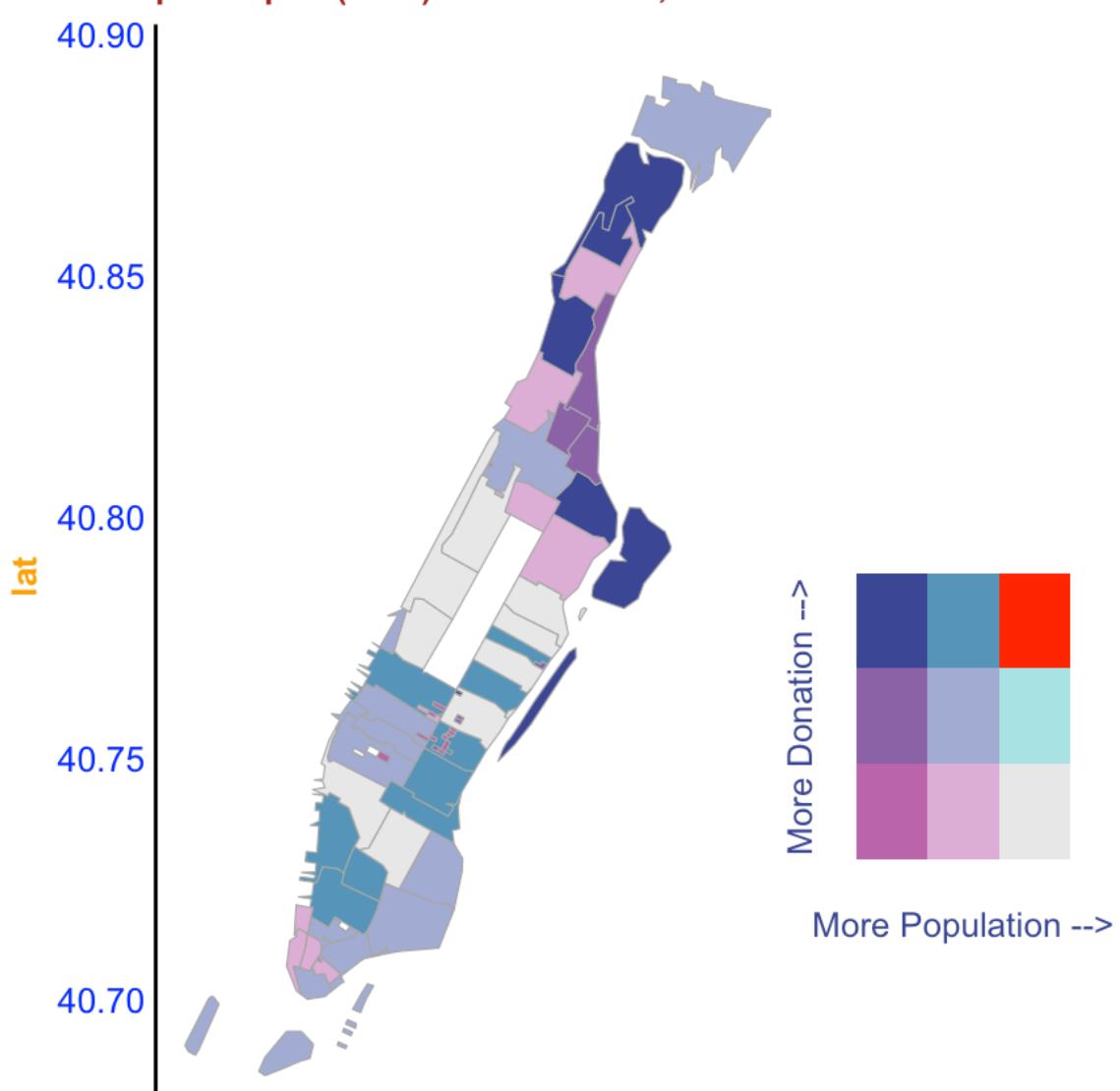
Now we have the idea on how to overlay multiple plots in arbitrary positions and sizes, since Choropleth provides excellent visualization compared to all the rest methods above, in that manner, I would be able to observe the donations in Manhattan which is the world's most famous and prosperous area.

The choropleth produced a high quality image with excellent visualization, I am thus allowed to observe even tiny details under such map. From the comparison between average donation versus contributor and capita, I had realized average donation per capita is quite lower than that of contributor in Manhattan. Referring to the facts in Hamilton and Lewis counties, selecting population as a measurement parameter in the current study is therefore problematic, it is not as reliable as contributor which at least was in a 1 on 1 pattern paired with a donation. If we really want to consider the impact of population on average donation, at least additional variables should be added, such as family numbers, income level, occupation types, that is a more profound and complicated analysis.

Average Donation per Contributors (USD) in Manhattan, 2016 Presidential Election



Average Donation per Capita (USD) in Manhattan, 2016 Presidential Election



Conclusion

After those substantial analysis, it is clear that democrats triumph in New York State. However, such analysis is only from one state, based on which we still can't extrapolate the national level poll. According to current study, I list my main findings as followings:

1. Democrats received much more funds than republicans in New York State, the ratio is even higher than 5: 1. Among all the candidates, Hilary alone possessed almost 80% of the donation.
2. There were in total 25 candidates, 5 were democrats and the rest were republicans. Among those candidates, 4 were female and 21 were men.
3. Top 1 contributor was Trish Hamlin, who donated more than 2000\$ to republican.
4. Distribution shows dnoation mainly came from the range of 0 to 500\$.
5. Democrats on average received more donation over time on the scale of day, month, and month_year.
6. Democrats tended to receive a larger donation in the range of 1000 to 2000\$ as well as small donations in the range of 0 to 1000\$ than republicans.
7. Top 20 contributors were more likely to support republican candidates.
8. The top 2 occupations of contribution are retired and attorney, whereas the top 1 employer of contribution is self-employed.
9. The average donations received by candidates were around 15.9 million dollars per democratic candidate and around 1.23 million dollars per republican candidate. The average donations made by contributors to democrat were around 1020 dollars and around 476 dollars to republican. The max donation received by candidate is around 63.13 dollars, and we had already known it was definitely Hilary, whereas the min donation received was around 6000

dollars, and it was given to republican's candidate James S III, Gilmore. The max and min donation made by contributors are around 2700 and 0 dollars, they came from republicans, Trish Hamlin and Alfred Trotta, respectively.

10. Female and Male were tended to donate to female and democrats in New York States, this analysis is not reliable as most donations went to top1 candidate, Hilary.

11. Donations tended to concentrate on a more prosperous counties, average donation per capita is much less reliable than the average donation per contributor, because population fluctuation among the counties might cause Simpson's paradox.

Reflection

There are so many challenges while moving on with this project, this is by far the most difficult and time consuming project I've ever dealt with. Especially, plot with map is where most pain came from since such content was not covered in the Udacity courses. I started from scratch and learned different kinds of materials over and over again. In total, I've read over 100 links, including the tutorial courses and stack overflow to accomplish this project. I believe there is still space to improve my work.

I tried to get best visualization with map, however, different methods have different drawbacks. GGmap provides vivid image, that is great, but it can't provide the perfect size of a specified location. So the map is either bigger or smaller, while cropping would cause resolution loss. Beside, ggmap sets 2000 limits for query per day, generally speaking that is less than 7 maps in my case.

Reconstruct the map with R could solve the size problem but the result is less vivid for the visualization. Choropleth, an external package around 60MB size, is a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map, such as population density or per-capita income. Choropleth maps provide an easy way to visualize how a measurement varies across a geographic area or show the level of variability within a region. It gave more details than any of the above methods, moreover, it also output even higher quality map than ggmap.

The only flaw is it relies on ZIP Code Tabulation Area (ZCTA) for the mapping, but ZCTA is not perfect matched with Postal Service (USPS) ZIP codes. Postal Service (USPS) ZIP codes developed to deliver mail. ZCTAs are polygons whose location and area are well defined by shapefiles for use in mapping and GIS applications. USPS ZIP codes are not polygons but a set

of lines and points typically not forming polygons. USPS ZIP codes correspond to a set of roads, streets and specific addresses. The USPS does not provide shapefiles nor other latitude-longitude representations for USPS ZIP codes. In current study, zipcode package only provides USPS ZIP codes, so when converting to ZCTA codes, some unmatched points would be lost, further leading to incomplete mapping. According to the warning message, it seems most of the USPS ZIP codes were successfully transformed to ZCTA, considering the size of our sample (over 0.6 million) the plot is relatively reliable.

In addition, there are some bad data in the set, like the 1 to 3 digit zipcodes, it would be impossible to extrapolate, luckily they are just several hundred in total.

It should be noted that, the dataset I have is for New York State only, though it helps me to understand the election event there, I am not able to extrapolate the conclusions nationwide.

Lastly, I am looking forward for the updates of ggmap as well as choropleth package. If ggmap will allow choosing specific location with satisfactory resolution or choropleth improve its pairing between USPS ZIP and ZCTA, mapping with coordinates shall become more visualized, which would in turn lead to high quality analytical work.

Thanks for evaluating my work! All errors are my own and I am not trying to make any political points at all. I am just a data science dabbler so critiques of the code, methods and conclusions are all welcome!

Resources

<https://rpubs.com/gary7135/udacity-2016-campaign-finance>
(<https://rpubs.com/gary7135/udacity-2016-campaign-finance>)

[https://en.wikipedia.org/wiki/Elections_in_New_York_\(state\)](https://en.wikipedia.org/wiki/Elections_in_New_York_(state))
([https://en.wikipedia.org/wiki/Elections_in_New_York_\(state\)](https://en.wikipedia.org/wiki/Elections_in_New_York_(state)))

<https://s3.amazonaws.com/udacity-hosted-downloads/ud651/GeographyOfAmericanMusic.html>
(<https://s3.amazonaws.com/udacity-hosted-downloads/ud651/GeographyOfAmericanMusic.html>)

<https://stackoverflow.com/questions/17723822/administrative-regions-map-of-a-country-with-ggmap-and-ggplot2> (<https://stackoverflow.com/questions/17723822/administrative-regions-map-of-a-country-with-ggmap-and-ggplot2>) http://www.nickeubank.com/wp-content/uploads/2015/10/RGIS3_MakingMaps_part1_mappingVectorData.html
(http://www.nickeubank.com/wp-content/uploads/2015/10/RGIS3_MakingMaps_part1_mappingVectorData.html)

<https://arilamstein.com/open-source/choroplethrzip/creating-zip-code-choropleths-choroplethrzip/> (<https://arilamstein.com/open-source/choroplethrzip/creating-zip-code-choropleths-choroplethrzip/>)