

# Quantifying Power - Preliminary Analysis of Donations Made to Political Action Committees (PACs)

Bettina K. Arkhurst  
School of Mechanical Engineering  
Georgia Institute of Technology  
Atlanta, Georgia 30332-0250  
bettina@gatech.edu

**Abstract**—In a quest to quantify power within the United States, I took a preliminary view into Political Action Committee (PAC) donors, their estimated wealth, the most popular PACs and the political parties they are donating to. Due to the limited time for this project, only data sets from 2018 were used. Donors who gave to the top 20 PACs by amount in 2018 and paid at least \$200 were observed as well as 2018 IRS data and 2018 PAC contributions to political parties view the OpenSecrets website. More data analysis is needed to form conclusions on the correlation between wealth and PAC contributions, but from preliminary analysis, the higher one is in income, the more likely it is that they will donate to a PAC - with majority of donations coming from the top ~30% of incomes and exponentially decreasing with lowering income range. This preliminary view provides a framework by which I can begin the larger project of quantifying political power in the United States by observing more factors - financial or otherwise - contributing to an individual's or institution's power in our society.

**Index Terms**—political power; political action committees

## Introduction

### The Research Question

The overarching research question for this project is: Is there a way to quantify political power? The final goal would be to visualize the political power of various American groups, entities and individuals (thinking of political power as a constant value that is split up among these various groups for simplicity).

Given the scale of such a question, I decided to tackle a portion of it by seeking to identify what influence PAC spending has on the federal budget management. The first part of this search resulted in the question: "Where is PAC money coming from and where is it going?" This is the question I seek to answer in this paper. I will be looking at the overarching relationship between donors to the most prevalent PACs and where the money of said PACs is going. Power can be defined in a variety of ways, for example, as social influence, military force or even expertise. For the sake of this initial project, capital was used as the main metric of power.

### What is a Political Action Committee (PAC)?

A Political Action Committee, commonly referred to as a 'PAC', is a political committee that raises and spends money to elect or defeat candidates [1]. Most PACs represent businesses, labor unions or ideological interests, solicit funds from members and contribute to campaigns. There are rules governing PACs such as the fact that no individual may contribute over \$5000 to a PAC in an election cycle. Over the past 30 years, the amount of PAC contributions to political parties has steadily increased.

### The Data Sets

There are a few data sets that were used for this project. The first data set was information on PAC donors from Kaggle, which contained election cycles from 1990 [2]. Only the year 2018 was used. The second data set was from the Center for Responsive Politics OpenSecrets website [3]. It provided the 2018 breakdown of PAC contributions to various parties. Only the PACs donating over \$10M were included in this study. The final data set used for this project is the U.S. Internal Revenue Service's (IRS) 2018 Individual Income Tax ZIP Code Data [4]. This data provided a reference for the relative wealth coming from states and zip codes that the PAC donors belong to.

### Data Set Variables

The PAC donors during election cycles data set: This is a data set contains the location, occupation and sub-PAC donation amount from various donors to PACs in election cycles since 1990. The data come from the Center for Responsive Politics OpenSecrets website that the data set author scraped and curated.

The PAC donors during election cycles data set variables:

1. Contrib: the name of the contributor
2. Address: city, state and zip code
3. Occupation: donor's occupation
4. Date: date of donation
5. Amount: amount of donation
6. Parent: parent PAC organization
7. Sub-Pac: the PAC subgroup donor donated to

## 8. Election Cycle: election year

The PAC donations by industry data set: This data set looks at the money donated from various industry-related PACs to the Republican and Democratic parties. These data are also from the Center for Responsive Politics OpenSecrets website.

The PAC donations by industry data set variables:

1. Industry: the type of industries the PAC belongs to
2. Total Amount: sum of money PAC in the industry donated to both parties
3. To Democrats: sum of money PACs in the industry donated to Democratic party
4. To Republicans: sum of money PACs in the industry donated to Republican party
5. Lean: whether the donations were more heavily to the Democratic or Republican party (options are: Solidly Republican/Conservative, Leans Republican/Conservative, On the fence, Leans Democrat/Liberal, Solidly Democrat/Liberal)

Tax information by zip code data set: This data set contains tax information grouped by location. Although there are over 153 variables in this data set, only four variables were used for this study.

The tax information data set variables:

1. STATE: the state associated with the ZIP code,
2. ZIPCODE: 5-digit zip code,
3. AGI\_STUB: size of adjusted gross income,
4. A00100: adjusted gross income (AGI).

The AGI\_STUB designations can be found in table 1.

TABLE I  
Table of "AGI\_STUB" designations by IRS. Adjusted Gross Income (AGI) broken into sextiles

AGI STUB Key	AGI Range
1	under \$25,000
2	\$25,000 under \$50,000
3	\$50,000 under \$75,000
4	\$75,000 under \$100,000
5	\$100,000 under \$200,000
6	\$200,000 or more

This paper looks at the donors to the top 20 PACs by amount in 2018, the average income range for their zip codes and the amount and frequency by which members of particular income levels donate to PACs. I will first describe the related work in this area, techniques and packages used in R for this study, discuss the results of the data analysis and future work.

### Related Works

There exists a wealth of information in social science about various theories of power. Due to the condensed nature of this course, rather than forming new analytical methods based on several theories of power, I looked

for any existing examples of efforts to quantify power and/or influence. One example I found was a quiz from the Atlantic which sought to compare the quiz-taker's rough political power estimate to those of various influential groups/people [5]. The author broke various forms of power into six categories: money, ideas, force, crowds, governmental authority, and reputation.

Furthermore, their analysis held three "truths":

Truth No. 1: Power increases exponentially. Being 10 times richer than the next person can mean having 100 times more influence.

Truth No. 2: In the United States, money and the ability to gather crowds matter more than other forms of power.

Truth No. 3: There is a wide gap between your potential and applied civic power. Almost all Americans could do more to raise money, mobilize people, or get a message out than they in fact do.

Along with looking for potential examples, I had sought to understand the meaning of "political power". I got the following definitions from watch a HarvardX U.S. Govt Class through the John F. Kennedy School of Government on edx.org: Power: "The ability of an actor to influence policy or control the behavior of others." Politics: "The process through which society settles conflicts over resources and values."

Given my lack of political and sociological knowledge, I used Wikipedia to get an overview on the myriad theories that existed around power [6]. Similar to the HarvardX definition above, power (both social and political) was defined as "the capacity of an individual to influence the actions, beliefs, or conduct (behavior) of others." Power was also said to be "derived by the factors of interdependence between two entities and the environment" with the two overarching types of power to be "soft power" (the ability to attract and co-opt, rather than coerce... shaping preferences of others through appeal and attraction) and "hard power" (use of military and economic means to influence the behavior or interests of other political).

Given the subjective nature of soft power and the emphasis but on economic means in the Atlantic quiz, I decided to look more at financial ties and advocacy – starting with PAC donations.

### Techniques

I decided to do hypothesis testing using a linear regression model and look at correlations within the data. I chose this path because I sought to find any relations between or among income, donations and locations. The null hypotheses were formed by looking at PAC donor's donation amount, location, income, PAC of choice and PAC donations by industry to candidates. The three hypotheses focused on three different aspects of the data: wealth and donation amount, wealth and frequency of donations and geography and PAC recipients.

### Hypotheses

The hypotheses and null hypotheses are as follows:

Hypothesis 1. Those from the wealthiest zip codes collectively give more money to the largest PACS.

Null Hypothesis 1. Those from the wealthiest zip codes do not give more money to the largest PACS.

Hypothesis 2. Those from wealthier zip codes donate to PACs more frequently than those from poorer zip codes.

Null Hypothesis 2. There is no significant difference between the top sextile and bottom sextile of zip codes in terms of donating frequency.

Other potentially interesting hypotheses that I thought of but did not have time to do included:

Hypothesis A1. The donors give more heavily to PACs linked to the winning/dominant party of their state.

Null Hypothesis A1. The donors donate to PACs not linked to the winning/dominant party of their state.

Hypothesis A2. The PACs donors donate to PACs linked to popular industries in their areas.

Null Hypothesis A2. No link between industries in the area and PAC donations exists.

Hypothesis A3. Those from similar areas give to similar PACS (i.e., there is a relationship between the PACs people are donating to and their locations).

Null Hypothesis A3. There is no correlation between the types of PACs people are donating to and their location (state and/or zip code).

### Packages Used

The stats package's linear regression model was used for modeling. For my data set, I used a linear model calling "lm" [7]. The linear model is of the following form:

```
lm(formula, data, subset, weights, na.action, method = "qr", model = TRUE, x = FALSE, y = FALSE, qr = TRUE, singular.ok = TRUE, contrasts = NULL, offset, ...)
```

This linear regression model function can be used to carry out regression, single stratum analysis of variance and analysis of covariance. The model gives the user options to tailor the fit by using a custom formula, a subset of the data for fitting, weights of coefficients, what to do if an "NA" is found and so on. In the case of this data set, since exponential relations were found, log-log linera fits were used.

### Results

The results are broken into three sections. The first section gives an overview of the data analyzed, the second section looks at correlations in the data and the final section seeks to model observed correlations.

### Data Overview

A density function of the donation amounts to the top 20 PACs in 2018 by amount is found in Figure 1. As seen in the figure, the donations tend to skew towards the lower end of values with most donations below \$2000. There is also a spike at \$5000 - the maximum individual contribution.

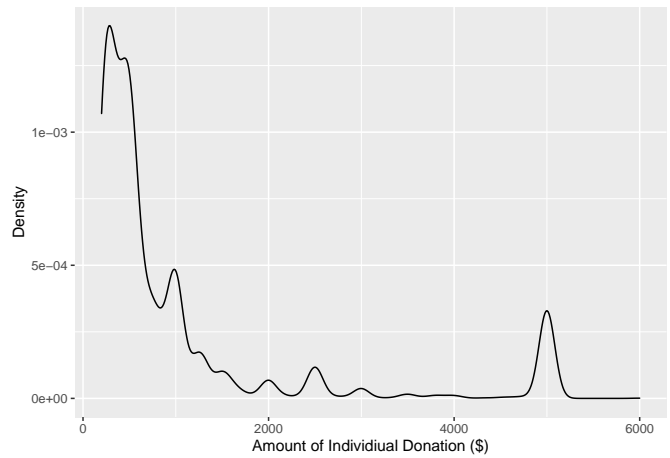


Fig. 1. Density of donations to top 20 PACs by amount from \$200 to \$5000 in 2018

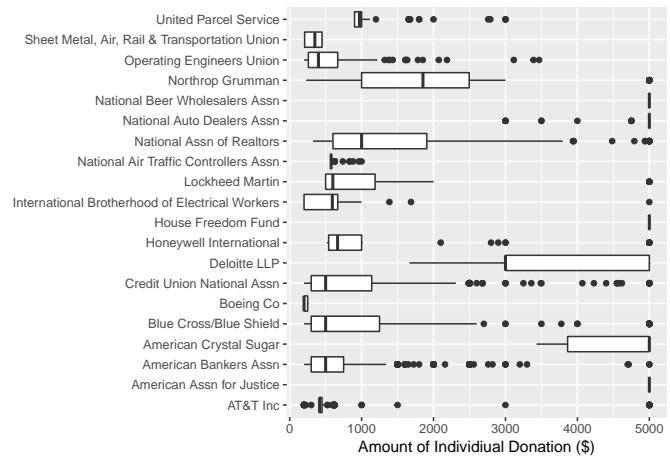


Fig. 2. Box plot of donations to top 20 PACs by donation amount

The overall amount range donated and frequency of donations to each of the 20 PACs is portrayed in Figure 2. For most PACs in this data set, the donations are averaging less than \$2000 with the exception of the National Beer Wholesalers Association, the House Freedom Fund, Deloitte LLP and American Crystal Sugar. It is possible that these PACs do, in fact, have many contributions on the lower end of the spectrum but since the data set starts from donations of \$200, such contributions are not captured.

As shown in Figure 3, the top 5 PACs that most frequently received donations in this data set were the American Bankers Association, Blue Cross/Blue Shield, Operating Engineers Union, the Credit Union National Association, the National Association of Realtors and AT&T Inc. The top 6 contributors to political parties in 2018 were: the National Association of Realtors (\$3,444,276), the National Beer Wholesalers Association (\$3,433,500), AT&T Inc (\$3,113,400), Northrop Grumman (\$2,849,740), the National Air Traffic Controllers

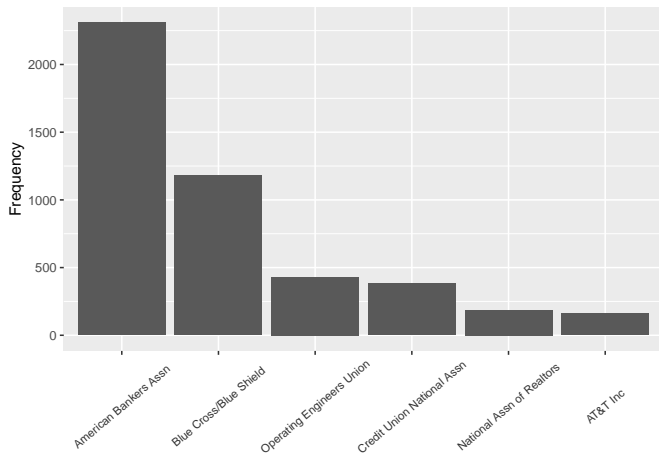


Fig. 3. PACs donors most frequently donated to in the 2018 data set



Fig. 5. Density of donors versus averaged AGI donation

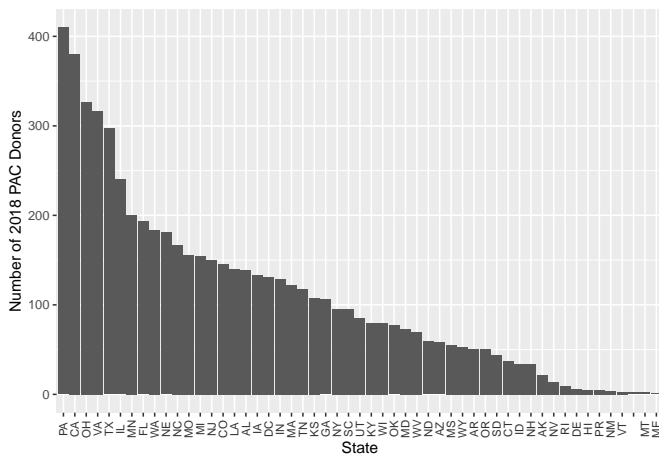


Fig. 4. 2018 donation frequency to top 20 PACs by state

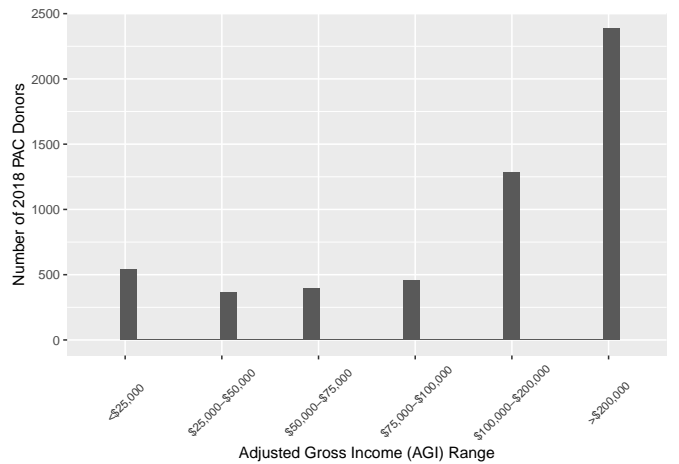


Fig. 6. Number of PAC donors by AGI range

Association (\$2,790,250) and the Sheet Metal, Air, Rail & Transportation Union (\$2,780,450). The frequency of donations by state is shown in Figure 4 with Pennsylvania, California and Ohio being the states with the most donors.

### Correlations

After a preliminary view of the data set, I sought to find correlations in income, donation amount and donation frequency. Since AGI values were average by zip code, these values were compared to individual contributions made in that same region.

The highest density of PAC contributions is seen at adjusted gross incomes below \$250K, but despite this increased density seen in Figure 5, most PAC contributions come from the two highest sextiles as seen in Figure 6.

The areas with the most donations are the top two sextiles followed by the bottom sextile. This unexpected phenomenon may be due to error introduced through averaging zip code AGIs and exclusions in the IRS data as discussed in the conclusion.

To better identify trends in the data, the donations were normalized by income and compared over the range of AGI values from \$0 to \$1M as shown in Figure 7. From this plot, an exponential decay appeared between the percentage of income given to PACs and the averaged adjusted gross income in a donor's zip code. The curve is most prevalent where the donation amount is \$5000 - the maximum contribution. This figure shows that despite lower averaged income, members of the bottom 4 sextiles were still willing to contribute the maximum amount.

After identifying the correlation in Figure 7, I sought to test the aforementioned hypotheses and better model the amount donated and frequency with which each AGI stub donated to these PACs.

### Modeling

Two models came from this analysis - both resulting from log-log linear regression. The first model, shown in Figure 8, looked at the normalized donation amount as a function of the donor's zip code's averaged adjusted gross income in 2018. The model was of the form:  $y =$

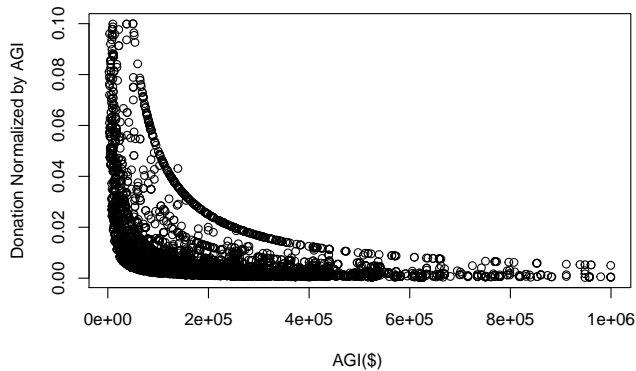


Fig. 7. Donations normalized by average AGI as a function of AGI

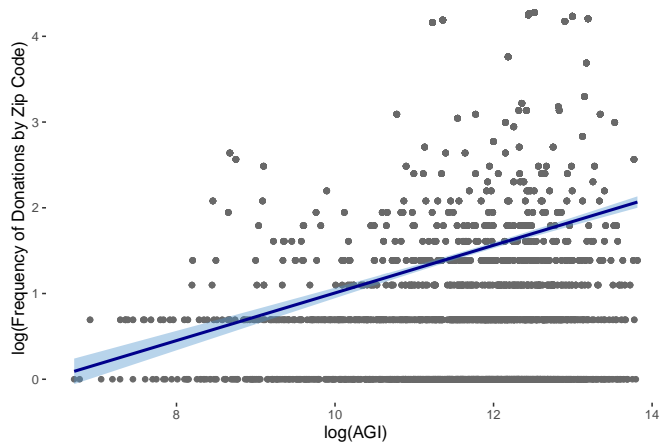


Fig. 9. Log-Log plot of donation frequency and AGI

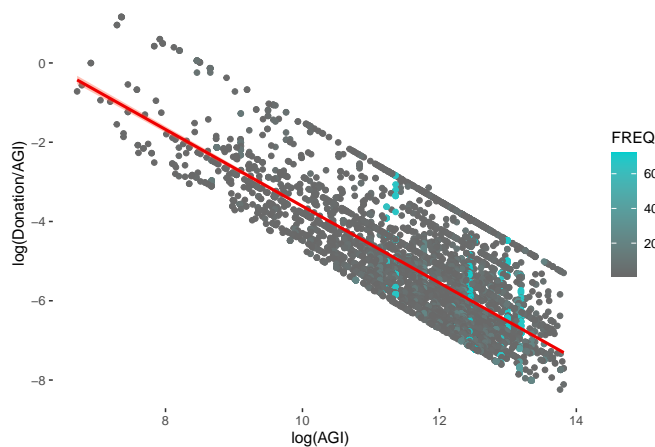


Fig. 8. Log-Log plot of normalized donation amounts as a function of AGI

$6.06730 - (0.96812)x$  with a p-value  $< 2.2e-16$  and an adjusted R-squared of 0.6213. Furthermore, the frequency of donations was also analyzed with those from areas of higher AGIs more frequently donating to these PACs.

To better model frequency, of donations, a log-log plot of the frequency of donations versus the adjusted gross income was used. Figure 9 shows how those from a higher income level donate more frequently than those from lower income zip codes. In this case, the linear model took the form:  $y = -1.77205 + (0.27787)x$  with a p-value  $< 2.2e-16$  and an adjusted R-squared of 0.06278.

### Conclusion

Members of wealthier zip codes were more likely to donate to PACs when compared to those of less wealthy zip codes and this likelihood increased with a decrease in the ration of donation amounts to adjusted gross income. Due to the low p-values, both null hypotheses can be rejected but it should be noted that the data are so dispersed that the R-squared values are below 0.7 for the

correlation to donation amount and less than 0.07 for the correlation to frequency.

### Sources of Error

As in the quest for any model, sources of error will be introduced in the process of simplifying the data and/or model. In the case of this study, potential sources of error include:

1. The averaging of adjusted gross incomes within an area code. Within each area code are vastly different household incomes, so information was lost by averaging the incomes to represent wealth in a given zip code.
2. Limited data samples. These data sets were limited given the wealth of information provided regarding PAC donations, so working with only a subset of that data may not fully tell the whole story.
3. Exclusion of data from the IRS database. The IRS follows particular guidelines in order to keep data anonymous, which potentially led to some income data being excluded from the data set used.
4. Simplified models. Only regression models were used in this study. For future work, more work may be needed to determine more accurate models.

### Future Work

I seek to look more into the question of political power by incorporating money from other entities like lobbyists, looking at federal budget breakdown, looking at the bills passed in congress and social media engagement. And that would just be looking at the power breakdown in Congress. I would also need to look at the executive and legislative branches before looking at lower levels of government, such as at the state and county levels. It is beneficial to start large because there is a wealth of data at the federal level and working through federal data can provide preliminary information about the state and lower levels.

In the future, other potential data set sources for this project can include: the Pew Research Center Political

Datasets, the Twitter's API (using a package such as rtweet), USAspending.gov- which has the breakdown of all federal government spending for every fiscal year since 2008, and ProPublica's Politics Data Set, which contains a data set on all bills introduced in Congress since 1973. These sources can be used to better assess the discourse and sentiments of the general public, view influences of soft power and so on, but would require more advanced techniques for analysis such as text analysis (such as this "Text Analysis Markup" tool) or ML or AI tools.

#### Acknowledgment

I would like to thank the instructors of GT 4801 for their aid and guidance throughout this process, OpenSecrets.org for making the majority of the information in this paper available and Kaggle user SlideRuleMath for scraping the PAC donor data used in this paper.

#### References

- [1]T. C. for R. P. A 501tax-exempt, charitable organization 1300 L. S. NW, and S. 200 Washington, "Political Action Committees (PACs)," OpenSecrets. <https://www.opensecrets.org/political-action-committees-pacs/2018> (accessed Jan. 23, 2021).
- [2]"PAC Donors for Historic Election Cycles." <https://kaggle.com/sliderulemath/pac-donors-for-historic-election-cycles> (accessed Jan. 24, 2021).
- [3]"Industry Breakdown • OpenSecrets." <https://www.opensecrets.org/political-action-committees-pacs/industry-breakdown/2018> (accessed Jan. 24, 2021).
- [4]Internal Revenue Service, "SOI Tax Stats - Individual Income Tax Statistics - ZIP Code Data (SOI) | Internal Revenue Service." <https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi> (accessed Jan. 25, 2021).
- [5]Eric Liu, "How Much Political Power Do You Have? - The Atlantic," Jul. 2015. <https://www.theatlantic.com/magazine/archive/2015/07/power-quiz/395273/> (accessed Jan. 24 2021).
- [6]"Power (social and political)," Wikipedia. Jan. 27, 2021, Accessed: Jan. 23, 2021. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Power\\_\(social\\_and\\_political\)&oldid=1003179057](https://en.wikipedia.org/w/index.php?title=Power_(social_and_political)&oldid=1003179057).
- [7]J. M. Quick, "R Tutorial Series: Simple Linear Regression | R-bloggers," Nov. 26, 2009. <https://www.r-bloggers.com/2009/11/r-tutorial-series-simple-linear-regression/> (accessed Feb. 10, 2021).