

# Your Substantive Title Here

Your Name1\*

Your Name2†

2023-09-27

## Abstract

This is our informative abstract of fewer than 200 words. It describes what we investigate, how we investigate it, and what we find.

## 1 Introduction

In the ever-evolving landscape of political campaigns, understanding the intricate web of connections between individual donors, political action committees (PACs), and candidates is paramount. This paper presents a comprehensive data analysis utilizing R Studio to unravel the dynamics of political contributions and community structures. The initial step involved coding to classify data into distinct categories, distinguishing between individuals and PACs based on a designated column.

The dataset was divided into two subsets: ‘individuals\_df’ and ‘pac\_df.’ For ‘individuals\_df,’ a network analysis was conducted, establishing edges between candidates based on individual donations. Connections were weighted, reflecting the cumulative contributions from shared individual donors. A similar approach was employed for ‘pac\_df,’ where connections between candidates were determined by PAC donations, and weights represented the financial support provided.

Following the network analyses, the research delved into exploring candidate communities within the network. The investigation extended to examining these communities’ scorecard data, aiming to discern patterns and similarities among groups of candidates. This multifaceted approach offered a nuanced understanding of the relationships within and between different political entities.

The analytical phase included hypothesis testing, focusing on comparing the mean union scores of candidate communities. To evaluate whether communities exhibited similar or different means, ANOVA tests were employed, complemented by the use of standard error and other relevant statistical measures. Specifically, the paper delves into comparing means within communities with PACs, providing insights into potential disparities in political contributions.

Furthermore, the analysis goes beyond statistical comparisons, delving into the nuanced characteristics of different candidate groups. Descriptive methods are utilized to elucidate the distinctive traits and behaviors within these political communities, shedding light on the intricate interplay of individual and PAC contributions.

This scientific data analysis paper, rooted in R Studio, provides a robust exploration of political contributions and community structures. By combining network analyses, hypothesis testing, and descriptive approaches, the study offers a comprehensive perspective on the complex relationships that define contemporary political landscapes.

---

\*American University

†American University

## 2 Political Donations and Network Analysis

The last year has been marked by a growing number of coordinated labor movements and union action across the country.

There are a number of political implications of the expectations placed upon a candidate by a donor, including but not limited to ideological transparency and the degree of a representative agenda. Conversely, the actions that a congressperson takes to attract donations to further their own interests is also a deep topic of study. La Raja, Raymond J., and Brian Schaffner (2015) theorize that candidates who receive very large funding from individuals are very likely to be ideologically centrist, while candidates who seek funding from large amounts of small donations are more likely to be ideologically extreme. Additionally, the theory extends to corporations and business groups to claim that they will tend to favor moderate and conservative officeholders who will bargain over specifics rather than overarching ideological change. In our research, we expect to find that large donations from corporations will skew toward conservative networks in congress, while large donations from individuals will flow to centrist networks.

The academic literary body surrounding lobbying and political donations is dense, but there is a distinct difference in the post Citizens United v Federal Election Commission (2010) world. Citizens United further emphasized an already disproportionate influence in massive donations and dark money in elections. Hansen and Rocca (2019) draw a clear line in analysis of individual campaign contributors in the pre and post Citizens landscape to argue that a flood of extremely large donations to super PACs have contributed to a reduction in academic research surrounding political donations. Our investigation seeks to construct a web of large donations from available sources to understand the financial connections between current members of congress and highly active large donors. The primary difference in a post Citizens analysis is that many of the largest donations, largely from CEOs, is that the data has become obfuscated by inflated dark money values.

Graph theory is a powerful tool for analyzing complex systems of interacting agents in the political context. The precedent set by the research of Porter et al. (2005) to project bipartite networks onto one-mode networks, where the strength of connections can be quantified using the normalized interlock, in measuring social networks in congressional committees allows for an effective method of congressional network analysis. Our chosen method of clustering is the walktrap algorithm, discussed in greater detail in the following sections.

Here we go deeper into the intellectual debate, the political and social context of our investigation. To give the reader a clear sense of why we are writing this paper, we describe the relevant scholarly, technical, or popular literature. We give this section a meaningful *substantive* title; it is not entitled “Literature Review”, for example. We cite at least three published, peer-reviewed scholarly works. For example, we could cite @mooree20 or @moorav12<sup>1</sup>, which we discussed in class.<sup>2</sup> We only cite others’ work in our paper when it enhances the reader’s understanding of what we, the authors of this paper, are doing. We connect everything we cite to *our* investigation; this is our original research, not a book report or an annotated bibliography.

We do not cite paper titles or journal names, unless our paper is about someone else’s paper or about the set of articles in a journal. We do not cite authors’ first, given names. We can refer to either what an author does, or what a paper does, but we should be consistent. For example, “@moorav12 argue that we should ...” refers to what the authors do; “@moorav12 argues that we should ...” refers to what the paper – @moorav12 – does.

---

<sup>1</sup>There should always be a space before the “(” in a citation date.

<sup>2</sup>To cite a paper within parentheses, use, e.g., [@moore12].

In order to integrate citations into the References section below, we add entries into our file `main.bib`. This is a plain-text file that we edit in RStudio. We store `main.bib` in the same folder as our paper's `.qmd` and `.pdf` files. Its entries are formatted so that they can be knit to `.pdf`; see <https://j.mp/2UzTXEZ> for example entries for articles, books, and miscellaneous. We can get these entries automatically from Google Scholar by turning on BibTeX in the Google Scholar Settings - Bibliography Manager. Perhaps we use a tool like free, open-source BibDesk to help us manage the `.bib` file.

### 3 Data and Methods

For this study, two primary datasets were utilized—Open Secrets donation data and Union Scores for the year 2022. The Open Secrets data underwent cleaning to isolate donors contributing over \$2000 to House and Senate members in 2022. The names were standardized, and the ‘candidate\_party’ variable was simplified to ‘D’ for Democrats, ‘R’ for Republicans, and ‘I’ for Independents. A new variable, ‘pac\_or\_individual,’ was introduced to categorize PACs (‘p’) and individuals (‘i’). Notably, we integrated two initially separate datasets for individuals and PACs.

Simultaneously, the Union Scores data for 2022 was processed, encompassing both yearly and lifetime scores. Titles preceding names were removed, and percentage-based score values were converted to vectors. The House and Senate data frames were merged, and encoding discrepancies between the Union Scores (latin1) and Open Secrets (utf-8) datasets were addressed.

We leveraged the `igraph` and `qgraph` packages in R to construct network graphs, visualizing connections between Congress members based on donor contributions. Nodes, representing Congress members, were sized proportionally to the total amount donated and color-coded to distinguish between individual and PAC donors. This graphical representation facilitated a comprehensive exploration of relationships within and between these donor groups.

A walkthrough function was employed to identify groups within the network. These groups were then transformed into a data frame, enabling the association of Union Scores with specific Congress members and their respective groups.

For statistical insights, we implemented code to perform ANOVA on the identified communities, assessing differences in mean Union Scores. Additional analyses, such as graphical representations of score and donation distributions, may be explored based on available time.

### 4 [Our Results Section Title Here]

Here, we explain and interpret our results. We try to learn as much as we can about our question as possible, given the data and analysis. We present our results clearly. We interpret them for the reader with precision and circumspection. We avoid making claims that are not substantiated by our data. We are careful about causality. When we describe associations, we avoid language like “effects” and “increases”; we only describe “effects” or “impacts” when we have a causally well-identified research design.

Note that this section may be integrated into Section 3, if joining the two improves the overall presentation.

## 4.1 Predicting Distance with Speed

Our results for the `cars` data include estimating the linear model

$$\text{Distance}_i = \beta_0 + \beta_1(\text{Speed}_i) + \epsilon_i.$$

Perhaps we start by plotting the data, as in Figure 1.

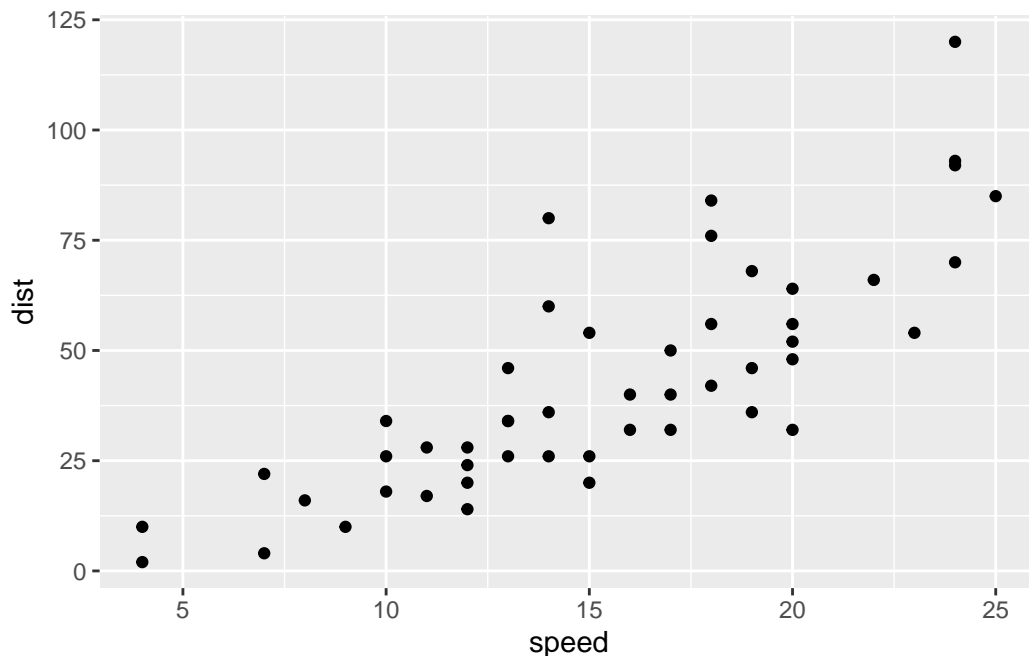


Figure 1: Distance on Speed

The data may be roughly linear, though there may be some non-linearity we should incorporate.

Below we show the model estimates. The first table uses `xtable()`, the second uses `stargazer()` [Hlavac18].

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-17.58	6.76	-2.60	0.01
speed	3.93	0.42	9.46	0.00

Table 1: Our Informative Caption

Using the `cars` data, we find that each unit of speed is associated with 3.9 more units of distance. We draw out what this really means, and what it implies. For example, if a typical difference among our observations is 7 units of speed, then our model estimates that a typical difference in distance among our observations is  $7 \times 3.9 = 27.3$  units of distance. We describe the substantive relevance of this number.

We do not report estimates like  $p = 3.242\text{e-}15$ , since these are computational zeros. Instead, we write  $p < 0.001$  or  $p \approx 0$ , as appropriate.

Table 2: Our Informative Title

	Outcome
	dist
speed	3.93*** (0.42)
Constant	-17.58** (6.76)
Observations	50
R <sup>2</sup>	0.65
Adjusted R <sup>2</sup>	0.64
Residual Std. Error	15.38 (df = 48)
F Statistic	89.57*** (df = 1; 48)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

We do not report quantities to unhelpful degrees of precision. Although there were 112,030,874 votes cast from voting-eligible population of 242,690,810 in the U.S. in 2022, it is not helpful to report turnout as 46.1619762%; writing 46.2% suffices.

## 4.2 Comparing Distances between High- and Low-Speed Cars

To report the results of a  $t$ -test, we do so in text, and perhaps in a well-formatted table as well, such as Table 3. Here, as above, we report the important details in text. For example, when we define “high-speed” cars as those traveling above the mean speed, the difference between the high-speed and low-speed group means is 32.25, with a 95% confidence interval that covers (20.6, 43.9).

Table 3: Distance by Speed Group

Parameter	Group	hs = 0	hs = 1	Difference	95% CI	t(42.87)	p
dist	hs	27.50	59.75	-32.25	(-43.90, -20.60)	-5.58	< .001

If I have tests of two outcomes from the same data, I can bind them together, as in Table 4:

Table 4: Distance and Square-root Distance by Speed Group

Parameter	Group	hs = 0	hs = 1	Difference	95% CI	t	df	p
dist	hs	27.50	59.75	-32.25	(-43.90, -20.60)	-5.58	42.87	< .001
dist_sqrt	hs	4.99	7.60	-2.61	(-3.49, -1.74)	-5.99	47.82	< .001

## 5 Discussion

We remind the reader what this paper was about, why it was important, and what we found. We reflect on limitations of the data or methods. If we have specific advice for someone picking up where we leave off, we provide that guidance. We avoid making trite statements like “more research should be done”.

## 6 References