"High Turnout Equals Democracy?": Statistical Examination of Turkey Presidential Election Data

Robert Ke

University of Rochester

Abstract. This study presents a systematic examination of electoral data to identify potential anomalies and irregularities. Utilizing Benford's Law, the Turnout-Vote relationship, and the normality of logarithmic vote rates, we uncover evidence that casts doubt on the integrity of reported election outcomes. Despite initial findings, the paper underscores the necessity for further investigation in delineating the mechanisms behind these anomalies and quantifying the extent of potential electoral fraud. This research contributes to the ongoing dialogue on electoral integrity and proposes a foundation for future empirical inquiries into electoral processes.

Keywords: electoral anomalies, statistical testing, electoral integrity

1 Introduction

Democratic societies are founded on the concept of conducting elections that are both free and fair, ensuring that every citizen's vote has an equal value. Government interventions that undermine or deprive free and fair voting behaviors should be questioned, examined, and censured if necessary.

Turkey is known for its high turnout rate and a great frequency of political voting events. However, Turkey has also remained controversial in terms of being a free country. "Turkey could be on the brink of dictatorship," posted by the *Economist*. *Freedom House* also questioned Turkish freedom under the rule of President Recep Tayyip Erdoğan and his Justice and Development Party (AKP). The press gave Turkey a low 32 overall democracy index, and lamented with, "deepening economic crisis ... government new incentives to suppress dissent and limit public discourse."

Therefore, it could be doubtful that the incumbent government intervened in the allegedly "democratic" elections. The aim of this paper is to analyze whether the voting results form any statistical irregularities.

Contemporary statisticians shed light on how to detect fraud with ballot data.

Benford's law underwent a revival as a tool for potentially detecting election fraud. In its basic and simplest form, Benford's Law notes that in numerous real-life processes, the first significant digit's logarithm tends to be uniformly distributed. Any discrepancies from this pattern could suggest the presence of alternate, potentially dishonest government interventions [7] [4].

A different strategy for detecting fingerprints of election fraud is to examine the distribution of votes and turnout numbers. This paper refers to the method provided by this paper [8] as the "**turnout-vote nexus**". Authors also observed that ballot stuffing induces a high correlation between turnout and voting results – unusually high vote counts tend to co-occur with unusually high turnout numbers.

The third approach, referred to as "**suspicious number**", is based on the finding that in certain suspected fraudulent elections, a significant number of polling stations report a winning party's vote percentage that exactly divides by five – for instance, 50%, 65%, 75%[6] and [5]. Common sense suggests that in large voter populations, the chances of a party consistently securing such rounded-off vote percentages across numerous precincts are extremely slim without some kind of vote tampering.

It demonstrates that even perfectly legitimate election data can display seemingly anomalous patterns that might be wrongly interpreted as fraud, unless the process generating the data is thoroughly understood. Considering the increasing agreement within empirical studies that claims of election fraud frequently leads to violence after the election [10][3], it's crucial for methods of detecting fraud to place a significant emphasis on minimizing false positive errors. Thus, to minimize the false positive errors, the turkey voting data has to undergo all three proposed methods with a strict statistical threshold, and the conclusion will be relaxed.

This paper will be outlined in the following way:

2 Data Description

The source of the dataset can be found **??**, and the correctness of the dataset is monitored and censored by Prof. Cantay Caliskan. After the initial cleaning of the dataset, only two subsets are observed. 973 Provinces (Presidential voting units) are observed with two consecutive elections - the 2014 Presidential Election and the 2018 Presidential Election. The choice of two elections is based on the author's anticipation of the occurrence of oddities. The 2017 Turkey National Referendum marks a colossal political system change where the power of Erdogan (winner of both presidential elections) was enhanced. Therefore, studying the presidential election before and after the crucial change may give a chance to observe some unordinaries.

Table x, in appendix A marks the columns of the dataset used.

3 Exploratory Data Analysis

Turnout rates and winner's log vote rates are two major components in terms of a voting dataset, and it is valuable to browse in visuals (Statistical detection of systematic election irregularities). The winner of both elections is President Erdogan, the leader of AKP.

Here are the mathematical notations used in this analysis of the paper.

$$\begin{aligned} W_{it} &: \text{winner's votes in district } i \text{ at election year } t. \\ P_{it} &: \text{eligible voters in district } i \text{ at election year } t. \\ V_{it} &: \text{valid total votes in district } i \text{ at election year } t. \\ K_{it} &: \text{actual voters in district } i \text{ at election year } t. \end{aligned}$$

$$\begin{aligned} &\text{turnout}_{it} : \text{turnout rate in district } i \text{ at election year } t. \\ &\text{turnout}_{it} = \frac{K_{it}}{P_{it}}. \end{aligned}$$

$$\text{Erdogan's win rate} = \frac{W_{it}}{V_{it}}.$$

$$\text{Logarithmic win rate of the winner } (L_{it}) = \log\left(\frac{V_{it} - W_{it}}{W_{it}}\right).$$

Graphics of all the considering variables are drawn and recorded. This section displays a few major components, and the rest of the records may be traced in the Appendix.

The turnout rate for each district in the 2014 and 2018 Turkey presidential elections is visualized in Table x. A histogram and density plot could be drawn to show the turnout rate distribution for these two presidential elections. The plot shows that the turnout rate distribution in 2014 is approximately normal with a mean of around 0.82, while the turnout rate distribution in 2018 is left-skewed with a mean of around 0.9 which is much larger than the mean of the turnout rate in 2014. Since the regular turnout rate for presidential elections is supposed to be normally distributed according to **??**, it is reasonable to suspect that there may be intervention with citizen's voting behaviors in the 2018 Turkey Presidential election.

To examine the interpretation that the difference between two consecutive elections is robustly huge. A pair-wise t-test is implemented where the null hypothesis is:

$$H_0 : \mathbb{E}(\text{Turnout}_{i,2014}) = \mathbb{E}(\text{Turnout}_{i,2018})$$

$$H_a : \mathbb{E}(\text{Turnout}_{i,2014}) < \mathbb{E}(\text{Turnout}_{i,2018})$$

The pairwise t-test result is displayed in Table 1: as the null hypothesis is firmly rejected with an extremely low p_value.

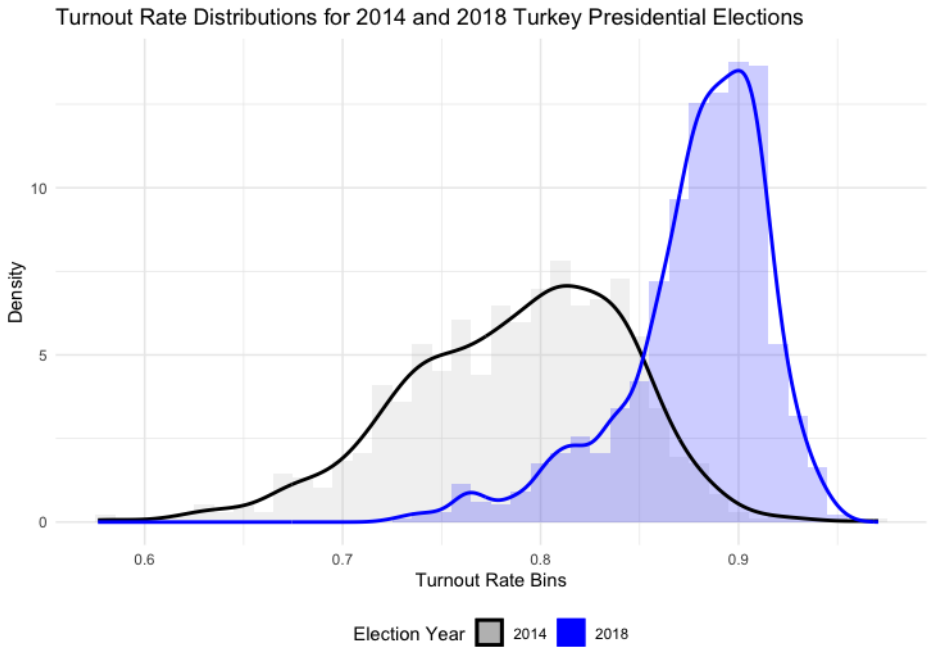

Fig. 1. Turnout Rates Across Districts (Turkey 2014, 2018 Presidential)

Statistic	Value
t-value	-71.652
Degrees of Freedom	972
p-value	$< 2.2 \times 10^{-16}$***
Alternative Hypothesis	true mean difference is less than 0
Confidence Interval	[-∞, -0.08848979]
Mean Difference	-0.09057094

Table 1. Paired T-Test Results for 2014 vs. 2018 Presidential Turnout

Figure x shows the vote percentage distribution of Erdogan in both 2014 and 2018 presidential elections across Tukey districts. Though obtaining a rather normal distribution, which is not a signal for fraudulance, according to [6], some questions could be raised. There is a bimodal trace in the histogram, where two modes are obtained by 0.6 and 0.7 vote rates. According to [5] and [7] who cast skepticism towards bimodal and rounded-up voting results.

4 Logarithmic Vote Rate and its Standard Normality

[6] and [2] found abundant evidence that the logarithmic vote rate forms perfect standard normal distribution. The mean is approximately zero with a relatively low standard error, and the standard deviation is 1. [6] found most west European nations achieve almost identical standard normal distribution and scale. They pointed out the skewness is around 0 and the Kurtosis score is around 3.

The logarithmic vote of Erdogan was both calculated in the 2014 and 2018 elections, with distributions across districts displayed in Figure x. The skewness are 0.612 and 0.310. The kurtosis in 2014 and the kurtosis in 2018 are 1.411 and 0.152. The kurtosis score is significantly lower than most west European countries and most elections. A low Kurtosis score may indicate a light-tail normal distribution. This may indicate the number of swing votes (a region split close to even votes for candidates) is less than most benchmarking Western European elections. The pro-Erdogan and the anti-Erdogan districts are starkly polarized (with a significant lead towards a specific candidate) with relatively larger amounts.

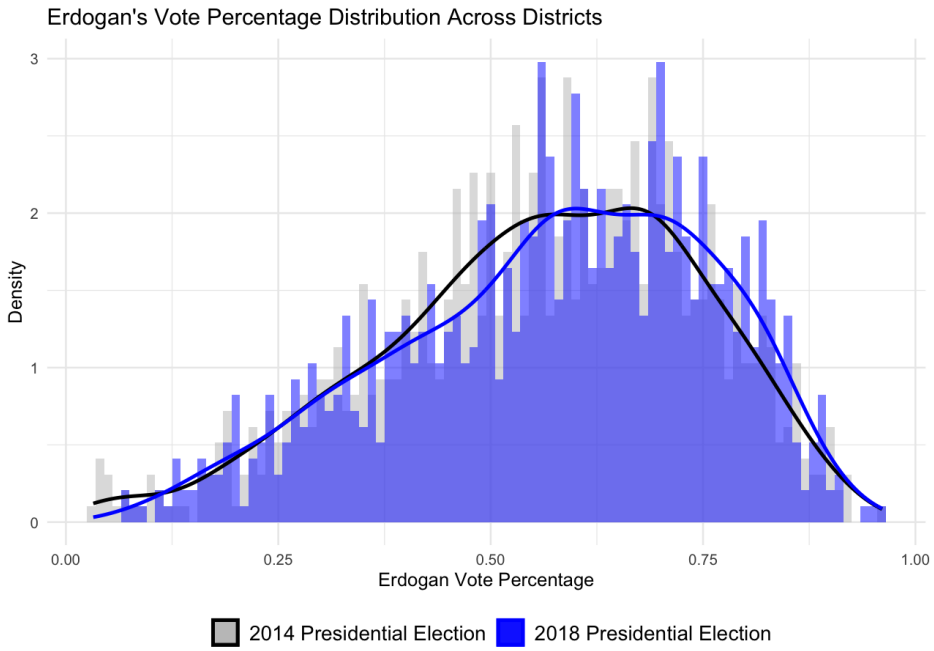

Fig. 2. Erdogan's Vote Distributed Across Districts (Turkey 2014, 2018 Presidential)

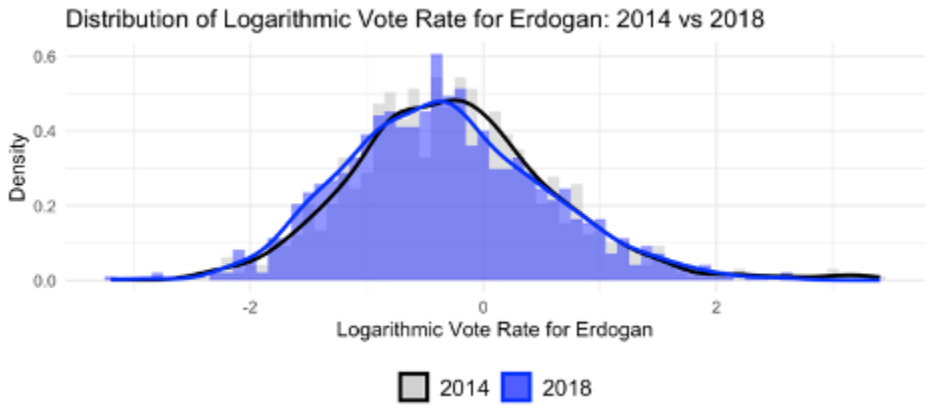

Fig. 3. Logarithmic Vote of Erdogan (2014, 2018 Turkey Presidential)

5 Benford's Law and its Exponentiality

Benford's law [1] states that in many naturally occurring collections of numbers, the leading digit is likely to be small. For example, the number 1 appears as the leading digit about 30% of the time, while 9 appears as the leading digit less than 5% of the time. The exact distribution predicted by Benford's Law for the leading digits 1 through 9 is logarithmic, given by the formula:

$$P(d) = \log_{10}\left(1 + \frac{1}{d}\right)$$

where $P(d)$ is the probability of a leading digit d.

Now we use *Fit Function* to fit our dataset into the Benford distributions function. The result is robustly positive. (The digit distributions in the election dataset of 2014, 2018 Turkey Presidential are perfect (with a $R\hat{2}$ score of 0.96) in terms of Benford's Law.)

Here is the visual of the Benford's Law:

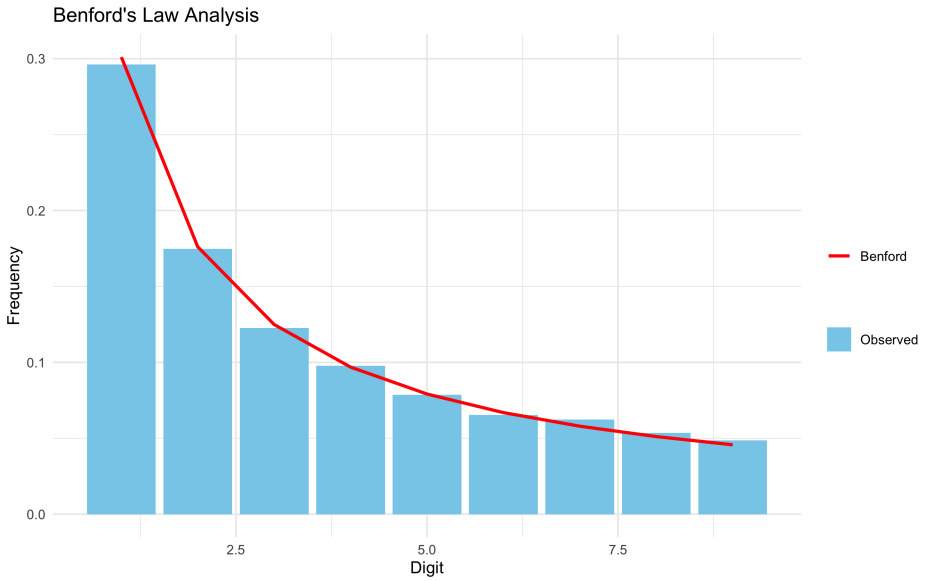

Fig. 4. Benford Law Fitting (2014,2018 Turkey Presidential)

6 Turnout-vote Nexus Analysis

[9] and [6] studies the correlation between turnout and vote support (i.e. the relationship between $turnout_{it}$ and w_{it}). Forensic analysis are repeated through different countries and a high correlation (an above 0.5 OLS coefficient) is found to be irregular.

Therefore, repetively, same OLS regression is casted on the dataset. The model of which is

$$W_{it} = \beta_0 + \beta_1 \times \text{turnout}_{it} + \epsilon$$

. A t robustness test is tested, and the result is showed in Table x.

For t = 2014,

- β_0 (Intercept) = 0.3632, with a standard error of 0.0833 and t value of 4.360, $p < 0.0001^{***}$.
- β_1 ($W_{2014,i}$) = 0.2477, with a standard error of 0.1055 and t value of 2.347, $p = 0.0191$.
- Multiple R-squared: 0.005643, Adjusted R-squared: 0.004619

For t = 2018,

- β_0 (Intercept) = 0.07072, with a standard error of 0.13827 and t value of 0.511, $p = 0.609$.
- β_1 ($W_{2018,i}$) = 0.57333, with a standard error of 0.15731 and t value of 3.645, $p < 0.0003$ ***.
- Multiple R-squared: 0.01349, Adjusted R-squared: 0.01248

As we can see, β_1 of the 2018 presidential election is significantly higher than 0.5 (benchmarking threshold), and 0.25 (2014 presidential election). The huge leap between 2014 election and 2018 may indicate mechanism of government intervention of people's voting behavior, which may give the incumbent government an edge of winning.

The two graphs below are the heatmap that visualize the correlation. These graphs show similarity traits as Russian 2011 and 2012 elections, which is deeply doubted by [6]. For the 2018 election graph, there are highly correlated points with large (almost 90%) turnout and large (almost 90% Erdogan's votes), which is extremely similar with what Klimek et. al's findings.

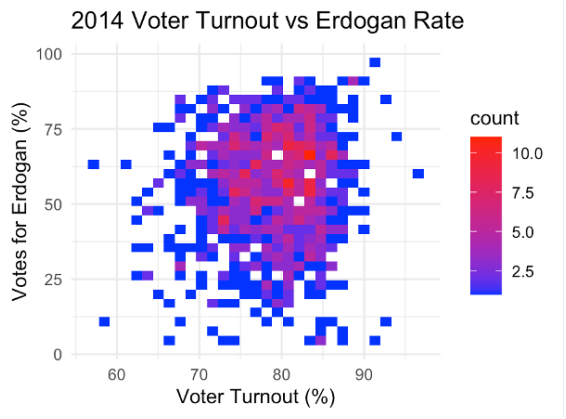

Fig. 5. Turnout-Vote Nexus Heatmap (2014 Turkey Presidential)

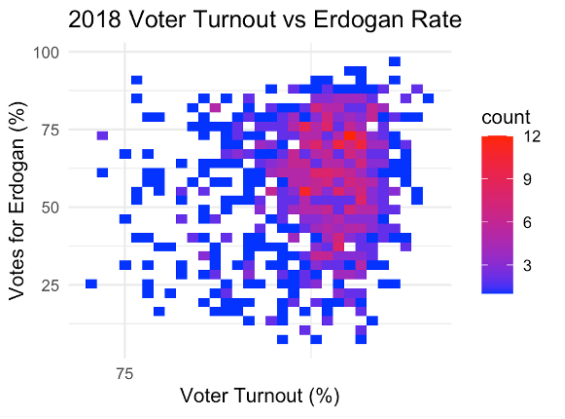

Fig. 6. Turnout-Vote Nexus Heatmap (2018 Turkey Presidential)

7 Conclusion

This manuscript endeavors to conduct a rigorous analysis of anomalies within electoral data. By employing Benford's Law, examining the Turnout-Vote relationship, and analyzing the normality of logarithmic vote rates, this study facilitates the identification of potential irregularities and causes for concern within election results. Despite these insights, further research is warranted in two principal areas:

1. **Mechanism and Causation:** An in-depth investigation into the nature of these irregularities from a political science perspective is necessary. It is crucial to elucidate the underlying mechanisms and causal factors, thereby contributing to the development of a comprehensive political science theory.
2. **Quantitative Estimation:** Inspired by the methodology outlined by [6], future studies should aim to quantify the extent of electoral fraud by estimating two key variables, f_e and f_i. These metrics represent the proportion of dishonest vote counts, encompassing scenarios where ballots from non-voters or invalid votes are illicitly attributed to the winning party.

This study raises significant concerns regarding sudden surges in voter turnout and the markedly increased correlation between turnout and the votes garnered by the winning party. Additionally, the investigation uncovers a pattern of binomial distribution with rounded numbers, which may indicate manipulation. These findings underscore the need for vigilance and further research to safeguard the integrity of electoral processes.

A Variables Names and Description Chart

Table 2: Election Data Description

Columns	Description	Type
Province	All province in Turkey	String
Eligible_2018_Presidency	The eligible voter in 2018 presidential election	Numeric
Voter_2018_Presidency	The actual voter in 2018 presidential election	Numeric
Valid_2018_Presidency	The valid votes in 2018 presidential election	Numeric
Invalid_2018_Presidency	The invalid votes in 2018 presidential election	Numeric
Ballot_2018_Presidency	The number of ballots in each district	Numeric
Ince_2018_Presidency	The votes for Ince in 2018 presidential election	Numeric
Aksener_2018_Presidency	The votes for Aksener in 2018 presidential election	Numeric
Erdogan_2018_Presidency	The votes for Erdogan in 2018 presidential election	Numeric
Demirtas_2018_Presidency	The votes for Demirtas in 2018 presidential election	Numeric
Karamollaoglu_2018_Presidency	The votes for Karamollaoglu in 2018 presidential election	Numeric
Other_2018_Presidency	The votes for other party in 2018 presidential election	Numeric
Eligible_2014_Presidency	The eligible voter in 2014 presidential election	Numeric
Voter_2014_Presidency	The actual voter in 2014 presidential election	Numeric
Valid_2014_Presidency	The valid votes in 2014 presidential election	Numeric
Invalid_2014_Presidency	The invalid votes in 2014 presidential election	Numeric
Ballot_2014_Presidency	The number of ballots in each district	Numeric
Erdogan_2014_Presidency	The votes for Erdogan in 2014 presidential election	Numeric
Demirtas_2014_Presidency	The votes for Demirtas in 2014 presidential election	Numeric
Ihsanoglu_2014_Presidency	The votes for Ihsanoglu in 2014 presidential election	Numeric

B Summary Statistics, Boxplots, and Histograms

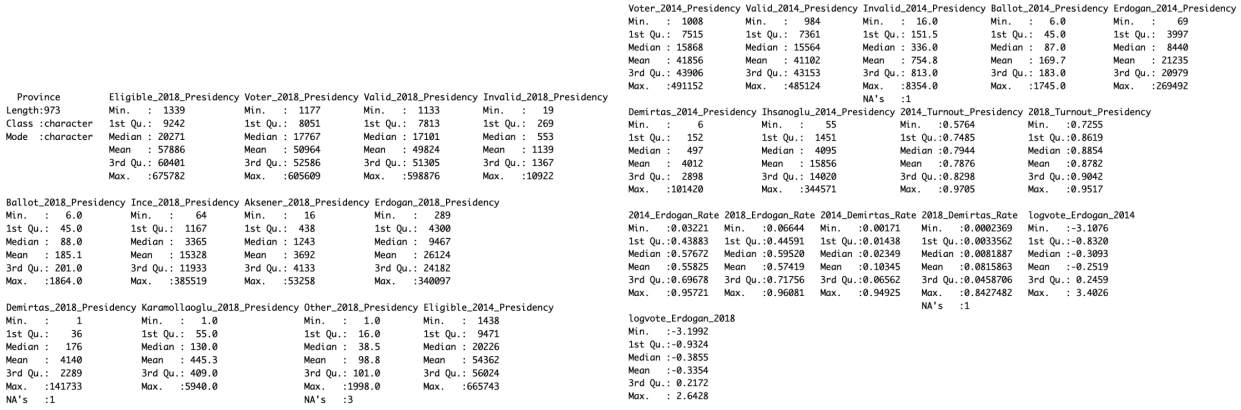

Fig. 7. Summary Statistics for all variables

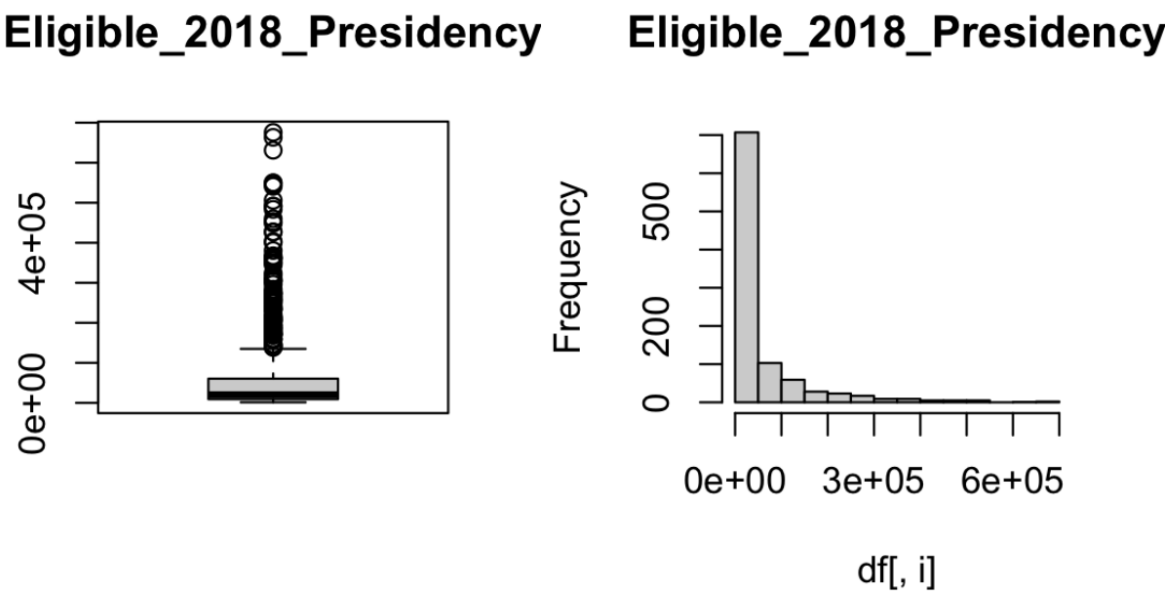

Fig. 8. Boxplot and Histogram

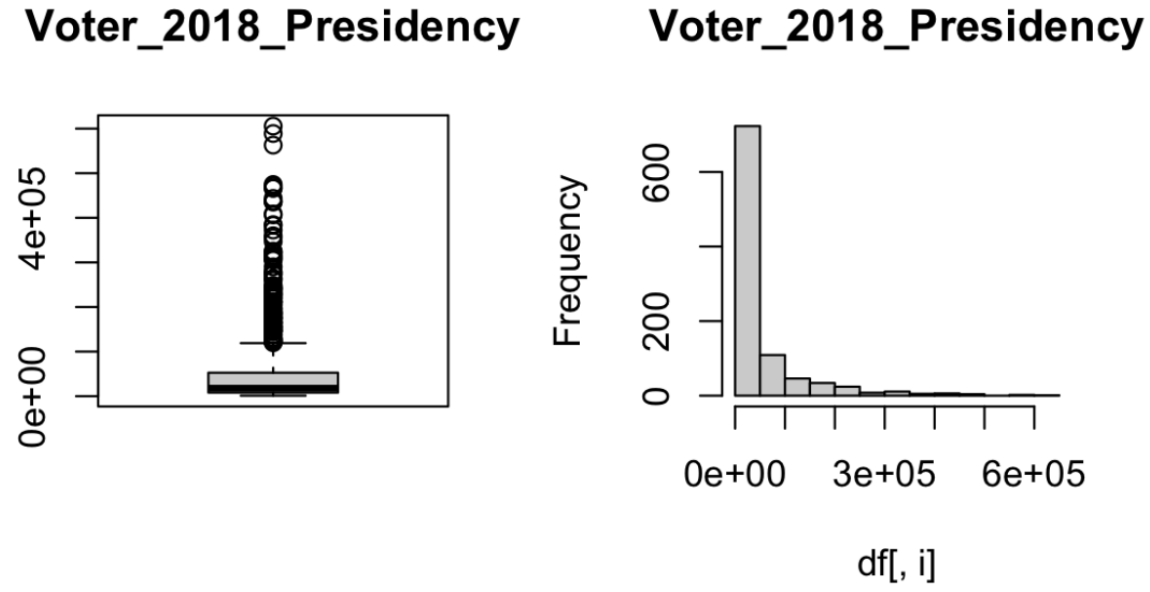

Fig. 9. Boxplot and Histogram

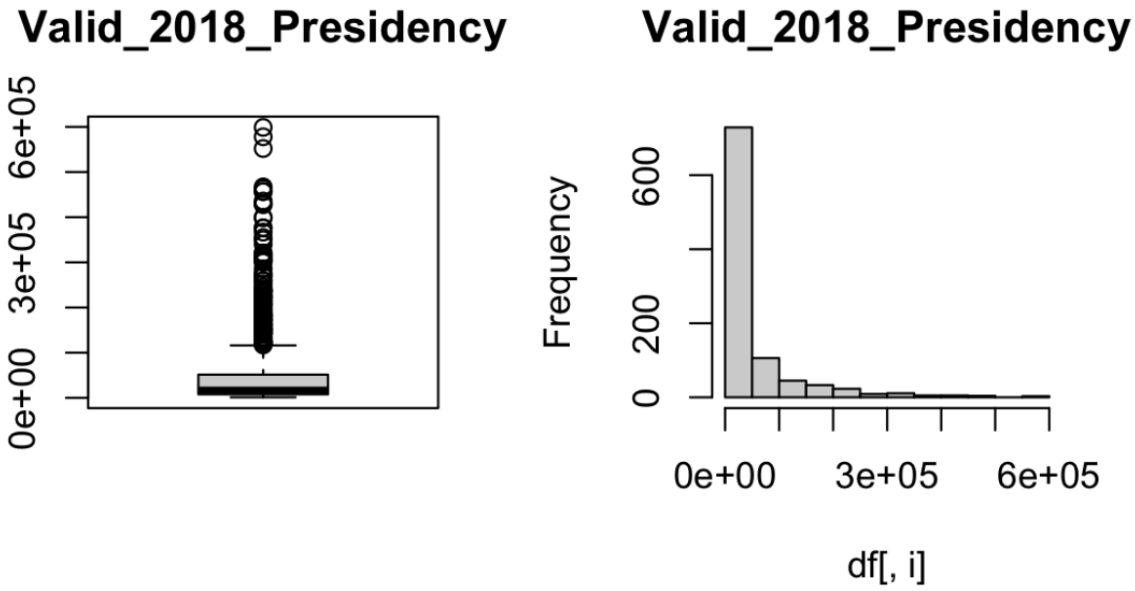

Fig. 10. Boxplot and Histogram

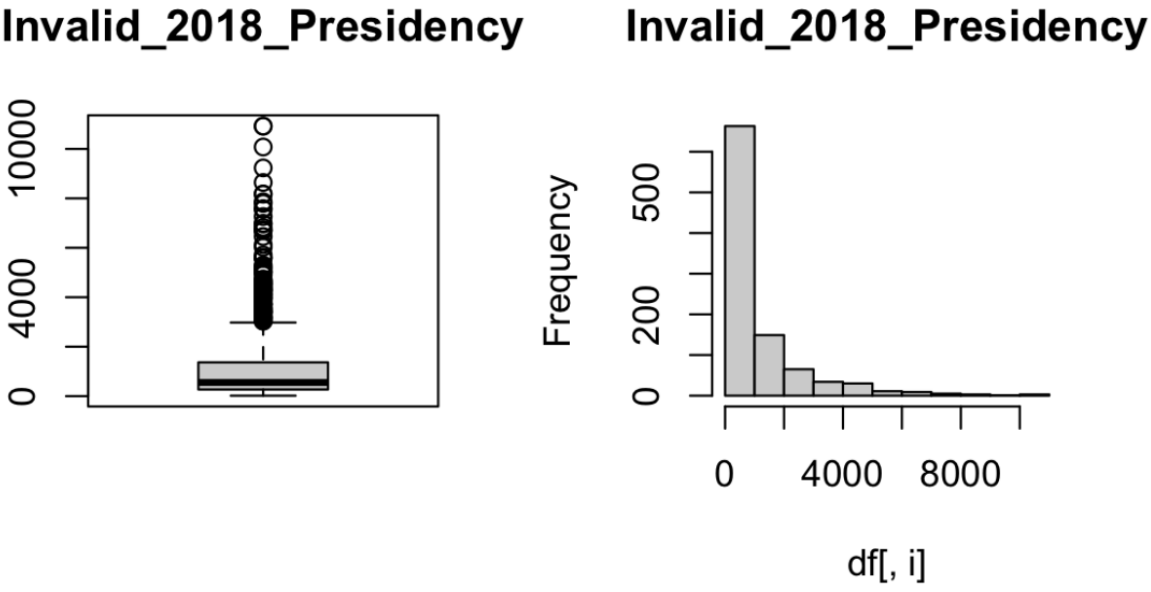

Fig. 11. Boxplot and Histogram

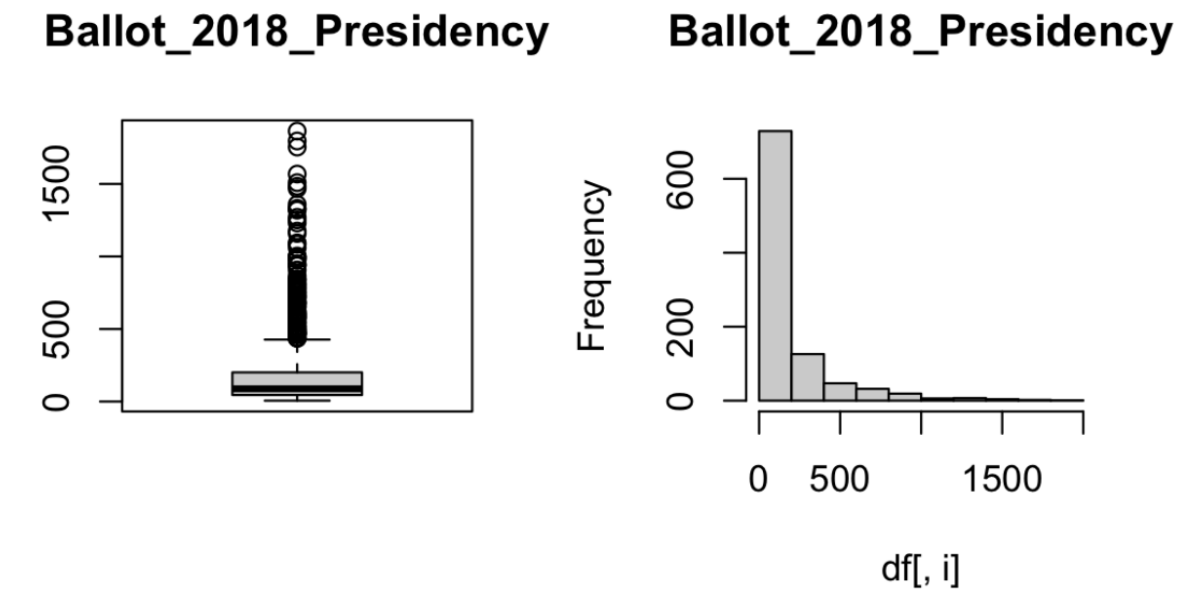

Fig. 12. Boxplot and Histogram

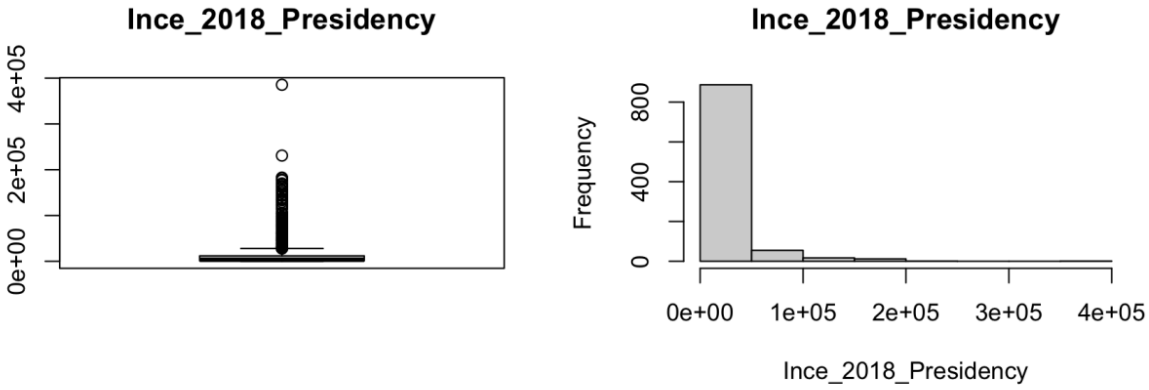

Fig. 13. Boxplot and Histogram

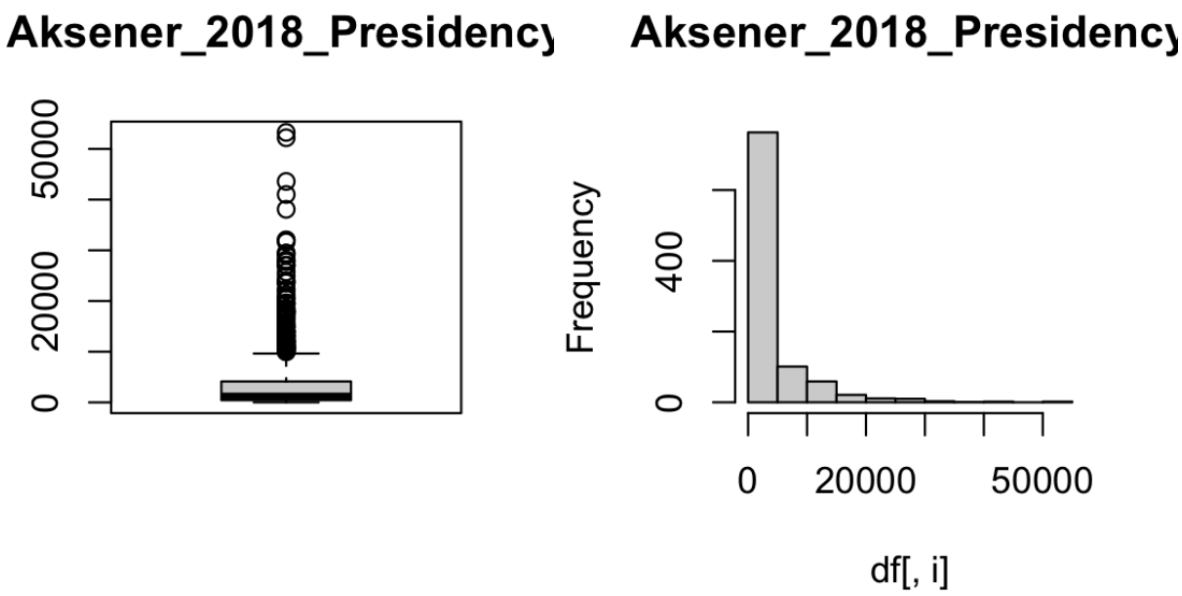

Fig. 14. Boxplot and Histogram

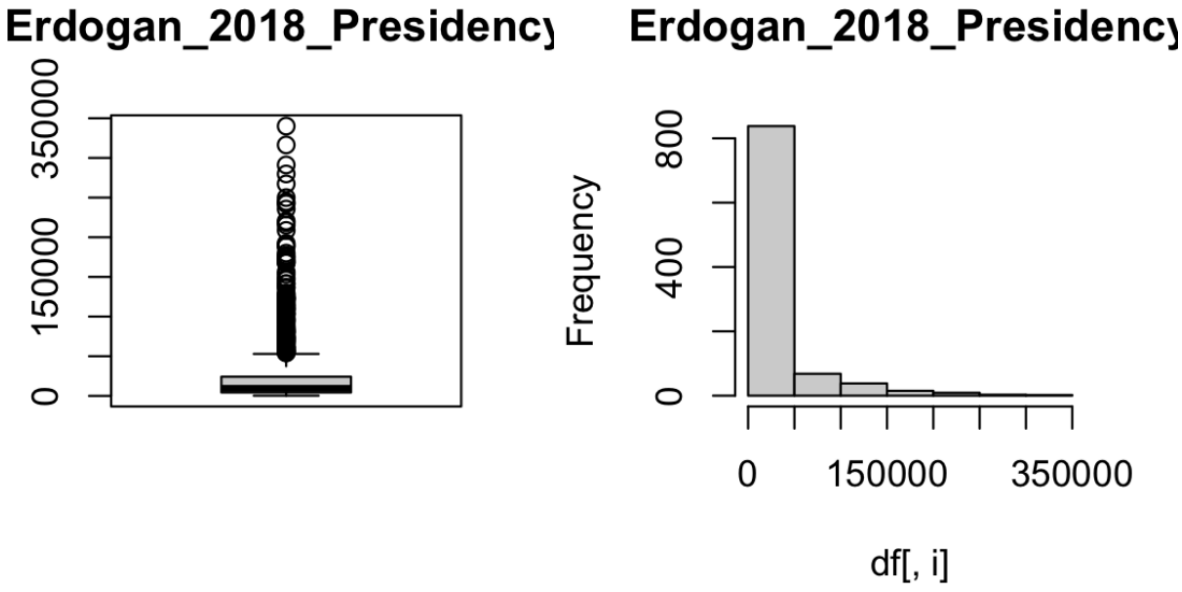

Fig. 15. Boxplot and Histogram

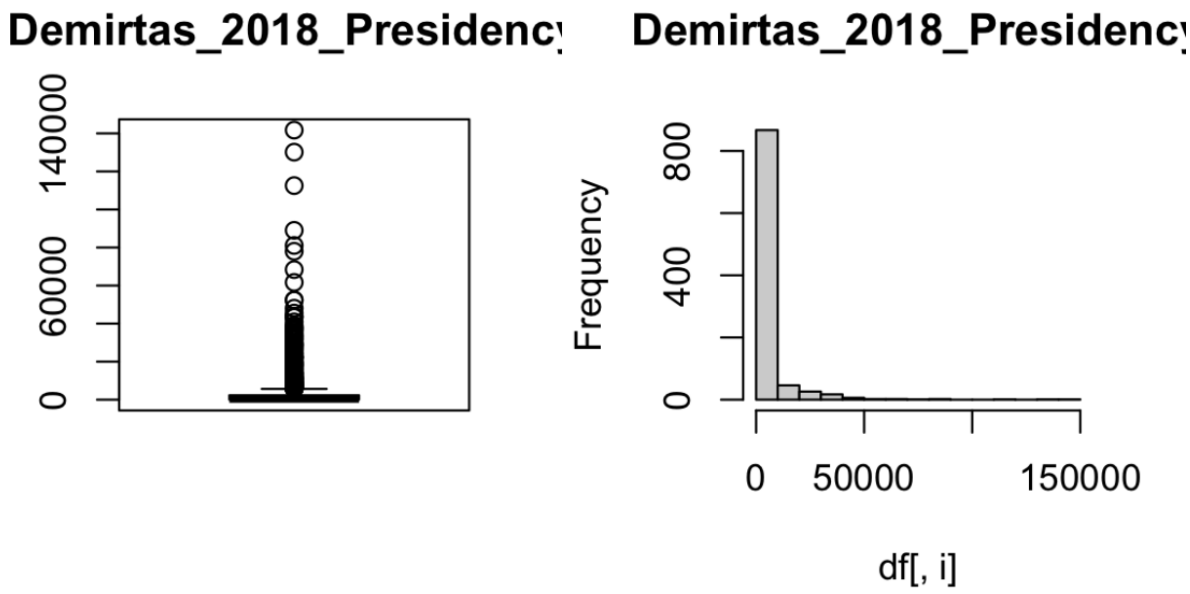

Fig. 16. Boxplot and Histogram

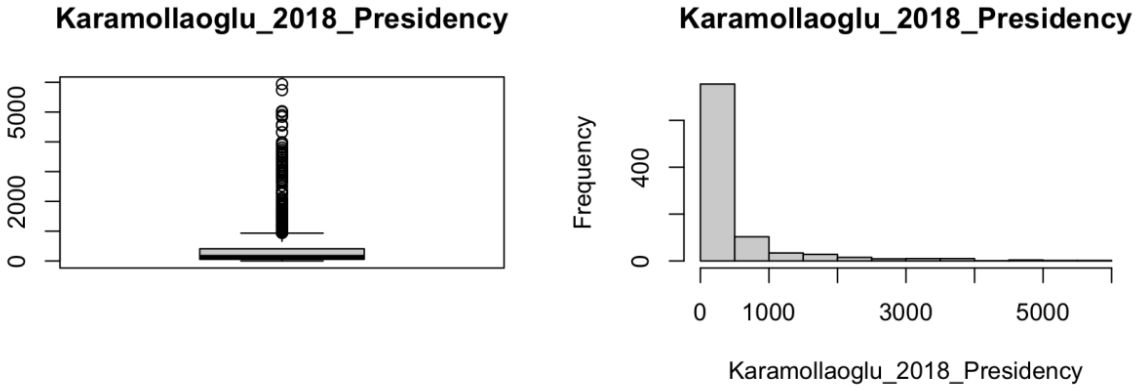

Fig. 17. Boxplot and Histogram

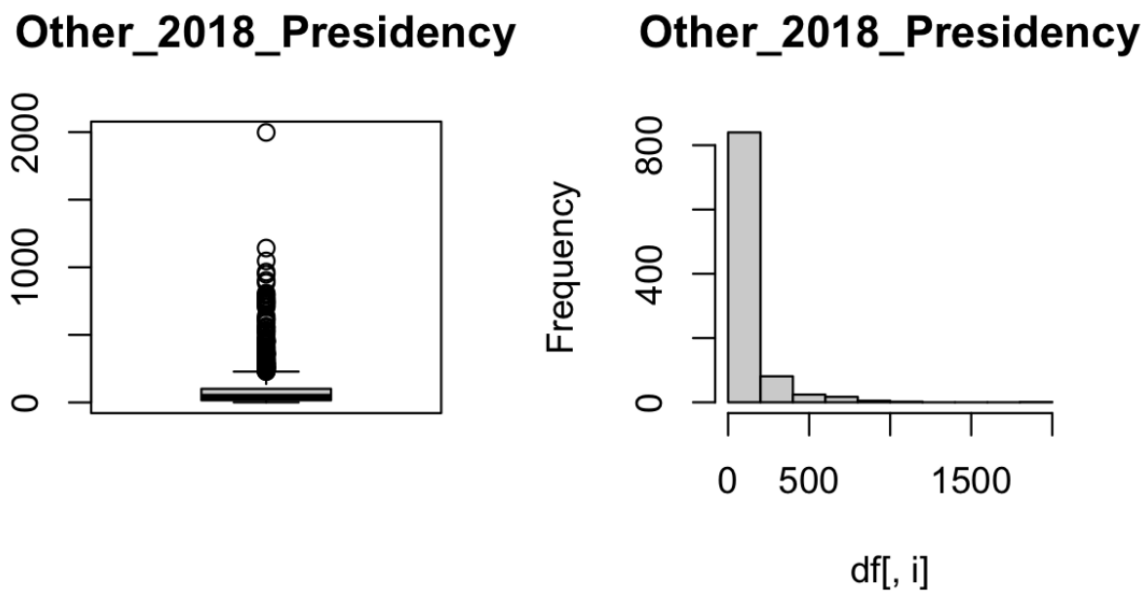

Fig. 18. Boxplot and Histogram

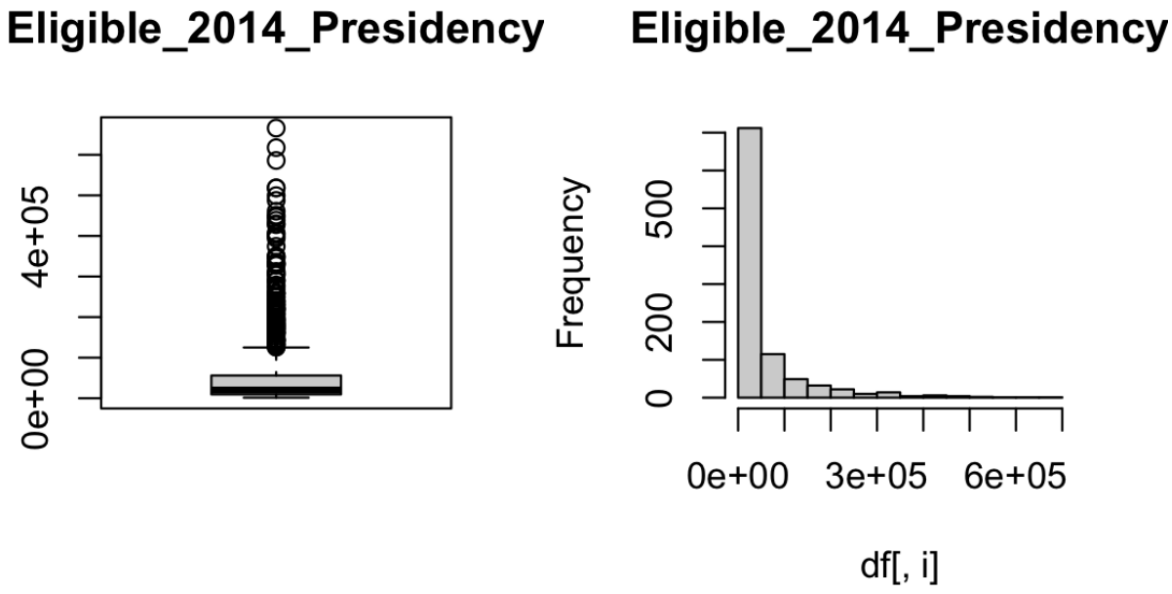

Fig. 19. Boxplot and Histogram

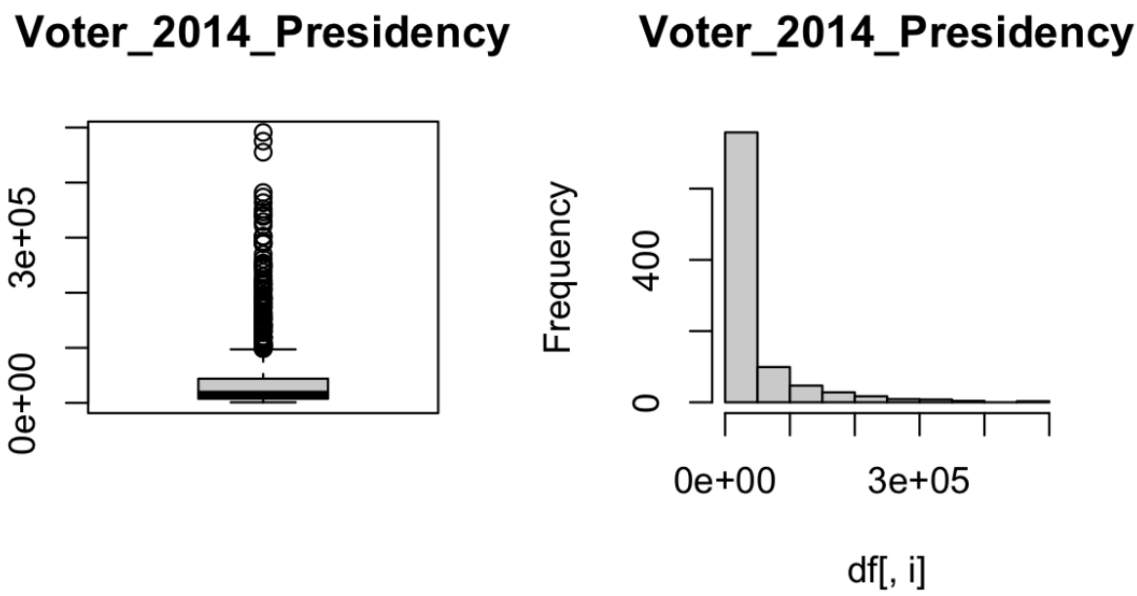

Fig. 20. Boxplot and Histogram

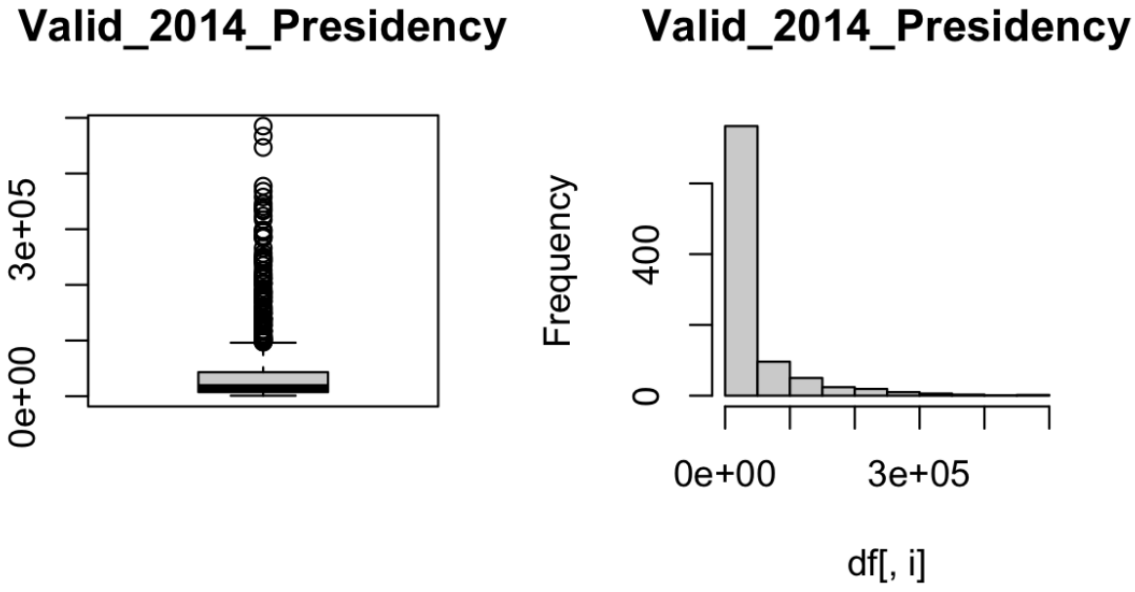

Fig. 21. Boxplot and Histogram

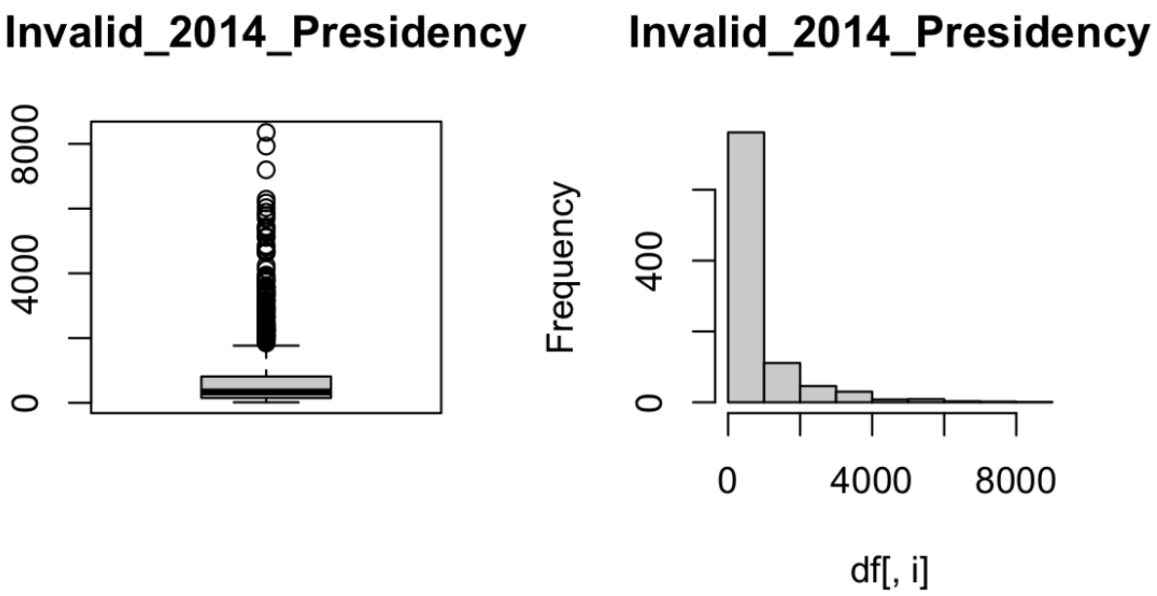

Fig. 22. Boxplot and Histogram

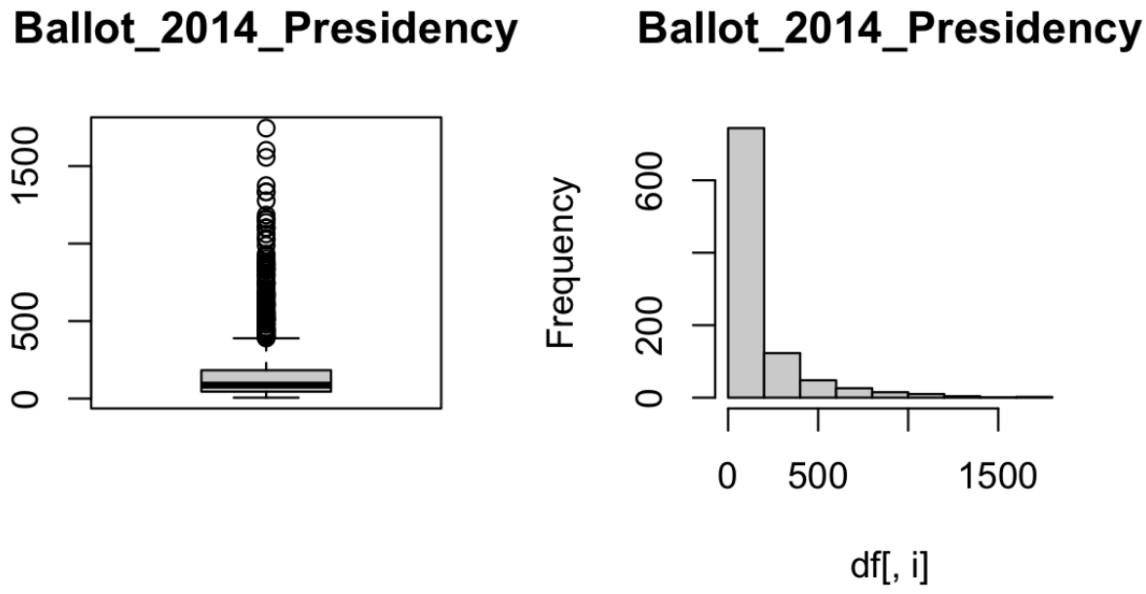

Fig. 23. Boxplot and Histogram

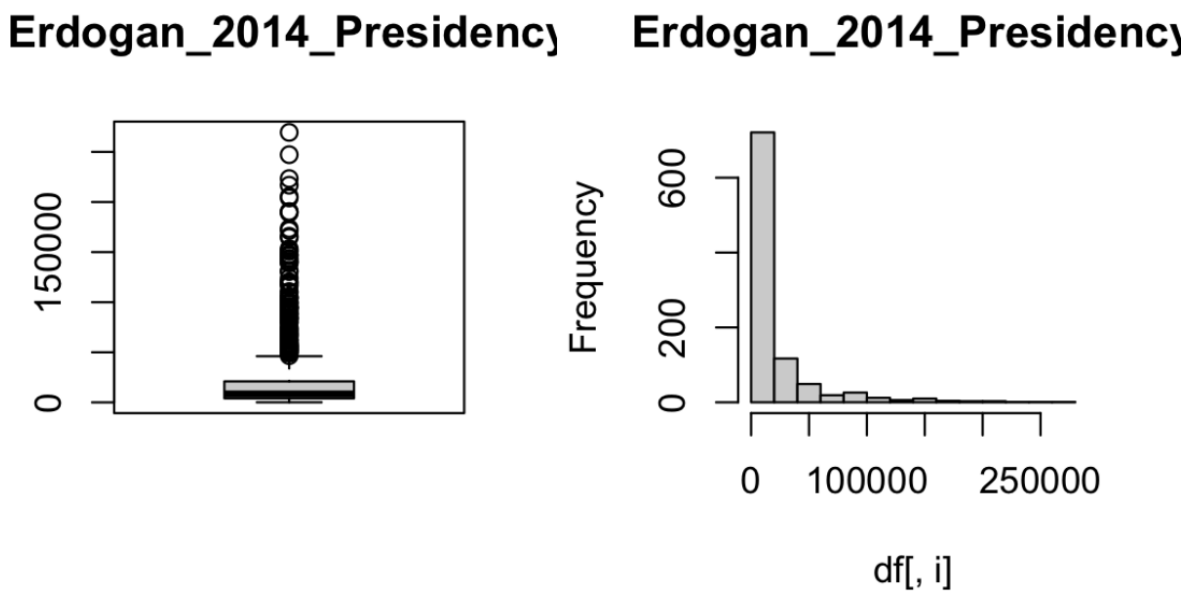

Fig. 24. Boxplot and Histogram

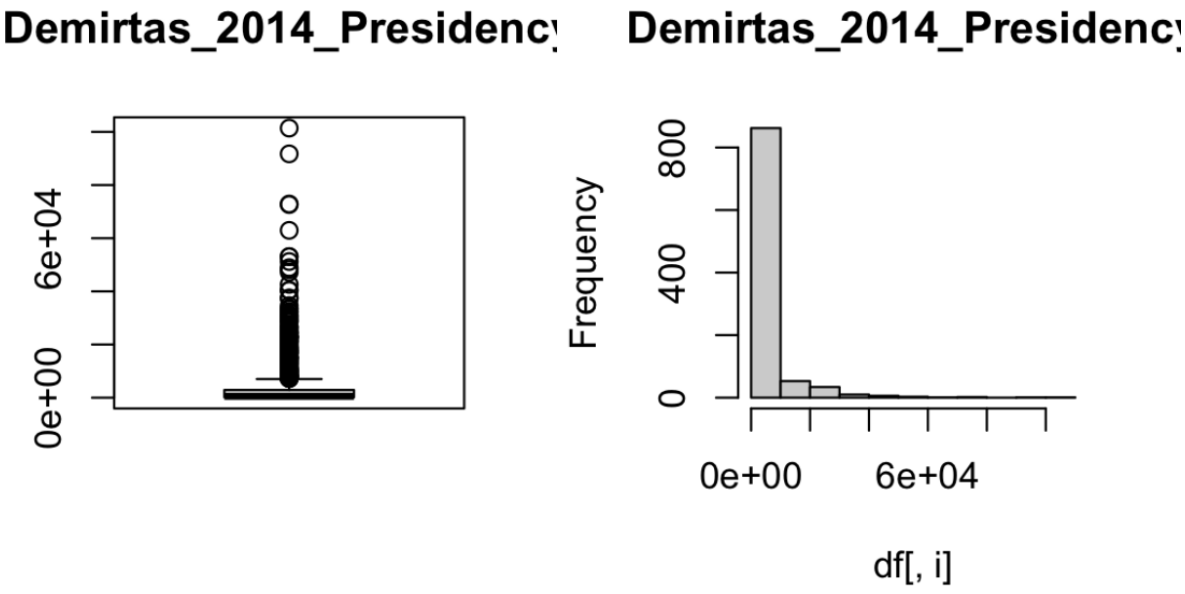

Fig. 25. Boxplot and Histogram

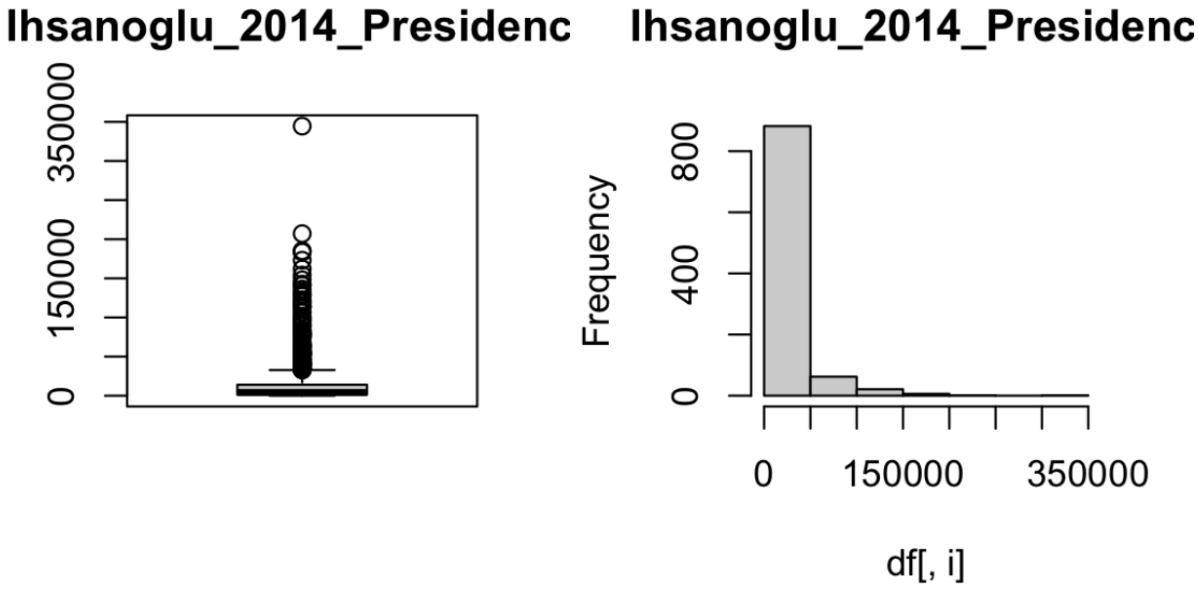

Fig. 26. Boxplot and Histogram

Bibliography

[1] F. Benford. The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78(4):551–572, 1938. Accessed 6 Apr. 2024.

[2] C. Borghesi and J. Bouchaud. Spatial correlations in vote statistics: a diffusive field model for decision-making. *European Physical Journal B*, 75:395–404, 2010.

[3] P. Collier. *Wars, Guns, and Votes: Democracy in Dangerous Places*. Harper, New York, 2009.

[4] J. Deckert, M. Myagkov, and P. Ordeshook. Benford's law and the detection of election fraud. *Political Analysis*, 19(3):245–268, 2011.

[5] S. Gehlbach. Electoral fraud in russia: Report from the russian blogosphere, 2012. The Monkey Cage, January 27.

[6] P. Klimek, Y. Yegorov, R. Hanel, and S. Thurner. Statistical detection of systematic election irregularities. *Proceedings of the National Academy of Sciences*, 109(41):16469–16473, 2012.

[7] W. R. Mebane. Election forensics: Vote counts and benford's law. 2006. Political Methodology Society, University of California, Davis, CA.

[8] W. R. Mebane and J. S. Sekhon. Robust estimation and outlier detection for overdispersed multinomial models of count data. *American Journal of Political Science*, 48:392–411, 2004.

[9] M. Myagkov, P. C. Ordeshook, and D. Shakin. *The forensics of election fraud: Russia and Ukraine*. Cambridge University Press, 2009.

[10] J. Tucker. Enough! electoral fraud, collective action problems, and post-communist colored revolutions. *Perspectives on Politics*, 5(3):535–551, 2007.