Causal Inference Summative

Analyzing the Impact of Election Swings on Democratic Vote Percentage

April 26, 2024

Z0195806
bjsn39@durham.ac.uk

Abstract

This thesis explores the effects of electoral swings on voter turnout within U.S. House elections from 1932 to 2008. Employing Difference in Differences (DiD) and Propensity Score Matching (PSM), the study analyzes how shifts exceeding five percentage points influence the Democratic vote percentage (DemPct). Results indicate that both positive and negative swings significantly affect voter turnout. The analysis reveals a nuanced response from voters to different types of electoral changes, contributing new insights into voter behavior and electoral dynamics. These findings have practical implications for political strategists and underscore the complexity of electoral influence on voter participation. The rigorous methodological approach strengthens the study's contributions to political science, particularly in understanding causal relationships where experimental methods are impractical.

Contents

1. Introduction

1.1. Study Overview

Negative Election Swings are pivotal events that can reshape the political landscape, influencing both the strategic decisions of parties and the participation rates of voters. This thesis examines the impact of negative election swings—defined as significant shifts away from the incumbent or previously dominant party—on voter turnout in U.S. House elections from 1932 to 2008[1]. Utilizing a Difference in Differences (DiD) approach complemented by Propensity Score Matching (PSM), this research aims to robustly isolate the effects of such electoral dynamics from other variables affecting voter engagement.

1.2. Theoretical Framework

This research is grounded in the potential outcomes framework, which allows for a structured exploration of causal relationships by hypothesizing different potential realities—one where the intervention (negative election swing) occurred and one where it did not. This approach is critical for estimating the specific effect of negative campaigning on voter turnout by comparing observed outcomes with estimated counterfactuals.

1.3. Research Questions and Hypotheses

"How does negative election swing affect voter turnout in U.S. House elections?" Based on the dataset, the hypothesis posits that negative election swing has a statistically significant impact on voter turnout, considering the potential for voter disillusionment and fatigue.

1.4. Study Quantification and Protenial Result Framework

- **Definition of Negative Swing**: Defining negative swings quantitatively with threshold instead of a float number.
- **DiD Application**: Implementing Difference in Differences to assess changes in voter turnout between districts experiencing negative swings and those without such swings, across the front and back election cycles.
- **PSM Utilization**: Using Propensity Score Matching to control for confounding variables that could affect turnout, such as demographic shifts, economic changes, etc., ensuring that the treatment and control groups are well-matched.

2. Theory

2.1. Election Swing

An electoral swing analysis (or swing) shows the extent of change in voter support, typically from one election to another, expressed as a positive or negative percentage[2]. In this study, I use One Party Swing, the math formula is as follows:

$$\mathrm{OnePartySwing} = \mathrm{CurrentVotePercentage} - \mathrm{PreviousVotePercentage}$$

2.2. Voter Turnout and Political Swing Research

Here is some literature review about the relationship between voter turnout and political swing. According to the correlates of voter turnout, Richard W. delved into different metrics for calculating voter turnout such as the Voting Eligible Population (VEP) and Citizen Voting Age Population (CVAP)[3]. Another research from the MIT Election Lab examines voter turnout over time, providing a comprehensive historical perspective[4].

Thomas Cao's study (2023) conducted during the 2022 midterm elections investigated how information about bipartisan oversight of the electoral process affects voter confidence and turnout[5]. Duquette discusses the disproportionate influence of swing states due to the winner-take-all nature of the U.S. Electoral College[6].

Political researchers also put there attention on the election analysis with causal inference. In his work, "The Statistics of Causal Inference: A View from Political Methodology," Luke Keele provides a comprehensive overview of the statistics used in causal inference within political science. He discusses the necessary assumptions for attributing a causal interpretation to statistical estimates and emphasizes the importance of identification assumptions in statistical analyses concerning causal effects[7]. Robert in his work using RDD to prove there is no causal effect of Incumbency Advantage in the U.S. House Elections[8].

2.3. Donation in this Study

In my research, I delve into the impact of negative electoral swings on voter turnout, building on existing research on electoral swings and voting behavior. At the same time, reference was made to the variables used by predecessors. Most previous studies have focused on the impact of electoral swings on party policies and election results, but few studies have analyzed in detail how negative swings affect voters' voting participation from the perspective of causal inference. By combining difference-in-differences (DiD) methods and propensity score matching (PSM), my study not only fills this gap but also improves the rigor of the study, allowing us to more accurately estimate the impact of this swing on voter behavior practical impact.

3. Data

The dataset[1] used in this study comprises a selection of variables derived from the U.S. House Elections spanning from 1932 to 2008. The data was sourced from a comprehensive electoral database, ensuring a robust foundation for the analysis. Here is a brief description of each variable included in the dataset:

1. **Main Variable for Causal Inference**
   - DWinNxt (Nominal Scale): Indicates whether the Democrat wins the next election (Yes = 1, No = 0). It measures the subsequent success of the Democratic candidate, providing insights into election trends and the persistence of political support.
   - DemWin (Nominal Scale): Indicates whether the Democrat won the current election (Yes = 1, No = 0). This variable is critical for understanding immediate electoral outcomes.
   - DPctNxt (Ratio Scale): Represents the percentage of the vote the Democratic candidate receives in the next election, providing a quantifiable measure of electoral support over successive election cycles.
   - DemPct (Ratio Scale): Represents the percentage of the vote the Democratic candidate receives in the current election, offering insights into the immediate electoral landscape.

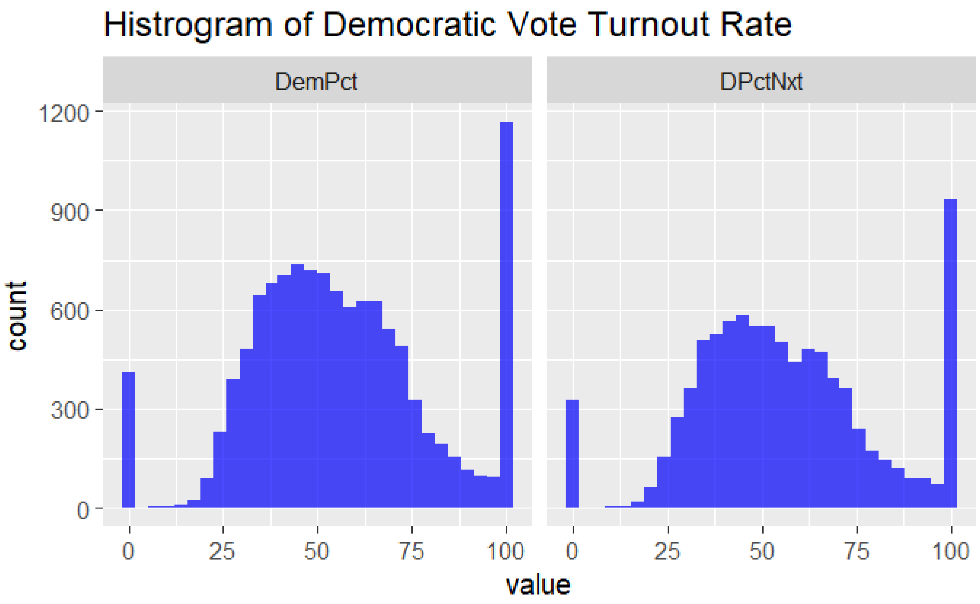


Figure 1: Histogram of the Turnout Rate in current year and next year

Based on these four variables, the tables of Before-treatment and Post-treatment can be assembled. TThis can help us build the **DiD** model.

2. **Intervention**
   - ElcSwing (Interval Scale): Measures the change in the percentage of votes between current and previous elections for the Democratic candidate. It quantifies the magnitude and direction of electoral shifts, providing a key indicator of political change.

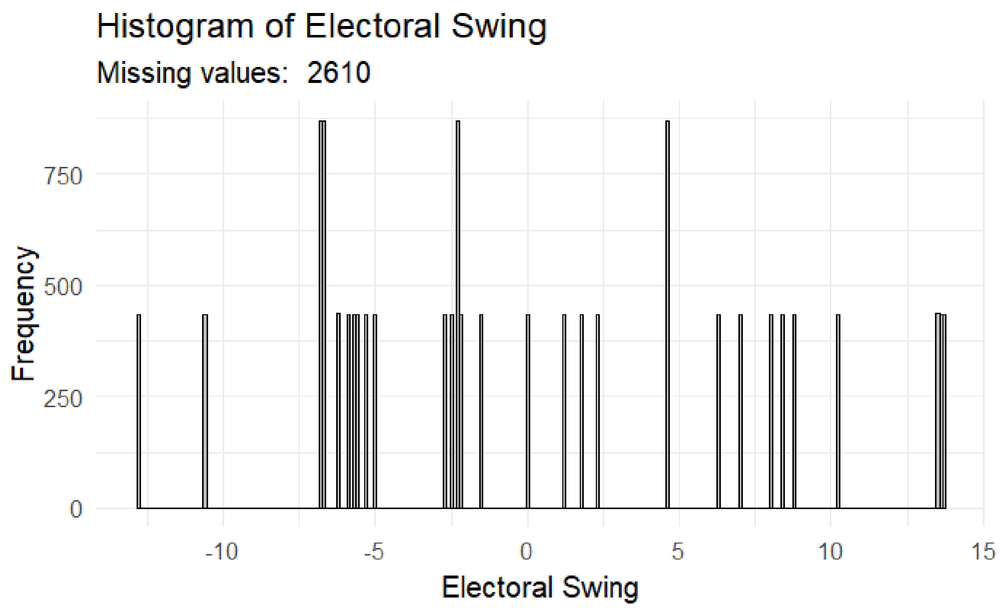


Figure 2: Electoral Swings Before Conversion

By observing the distribution of electoral swing, manually set the value less than −5 to **TreatmentNeg=1** and the value greater than 5 to **TreatmentPos=1**.

3. **Other covarites inspried by the previous research**
   - **Social Data**

- UrbanPct (Ratio Scale): The percentage of the population living in urban areas within the state. This demographic factor is essential for analyzing voter behavior across different urbanization levels.
- GovWkPct (Ratio Scale): The percentage of the state's workforce employed in government jobs. This variable can indicate the economic structure of a region and its potential influence on voting behavior.

- **Economic Data**
  - DSpndPct (Ratio Scale): Reflects the percentage of total campaign expenditure attributed to the Democratic candidate. This financial metric helps gauge the intensity and scale of campaign efforts.
  - DDonaPct (Ratio Scale): Measures the percentage of total donations received by the Democratic candidate, offering insight into financial backing and supporter enthusiasm.
- **Political Data**
  - YearElec (Interval Scale): The year in which the election took place. This temporal variable allows for the analysis of trends and patterns over time.
  - StAlphCd (Nominal Scale): The state's alphabetical code, which is crucial for geographic comparisons and understanding regional political dynamics.

These covariates are used to refit the model when DiD fails.

Data Set Build

The data set originally did not contain time, but by extracting the two data from last year and this year, and then marking them with postTrt, the time factor can be cleverly extracted to facilitate DiD analysis.

This is current year data Before-treatment

StAlphCd	DemWin	DemPct	Treatment	Other...	PostTrt
...	...	...	...	...	0

This is next year data Post-treatment

StAlphCd	DWinNxt	DPctNxt	Treatment	Other...	PostTrt
...	...	...	...	...	1

Other exploratory data analyzes and their codes on these variables are placed in **Appendix 2**.

3.1. Methodology: Difference in Differences (DiD)

Why DiD? DiD is advantageous because it helps control for unobserved confounders that are constant over time, as well as common shocks that affect all units within the dataset. By comparing the changes in outcomes over time between a treatment group and a control group, DiD isolates the "difference" that is attributable to the intervention (election swing), assuming that these groups would have followed **parallel paths** in the absence of the treatment.

3.1.1. Mathematical Representation of DiD

$$Y_{\mathrm{ij}} = \beta_0 + \beta_1 T_i + \beta_2 P_t + \beta_3 (T_i \times P_t) + X_{\mathrm{it}}\beta + \varepsilon_{\mathrm{it}}$$

Where:

- Y_{ij}: DemPct for district i in year j.

- T_i: High negative election swing as the Treatment.
- P_t: A binary variable indicating whether the year is a post-election swing year.
- X_{it}: A vector of control variables.
- ε_{it}: The error term.

3.1.2. Treatment Definition

Positive/Negative Value of Election Swing: Districts are separately categorized for positive swings (greater than +5 points) and negative swings (less than −5 points) to examine differing impacts.

3.1.3. First Differences

- First, calculate the average Democratic percentage (DemPct) by time and treatment group to establish a baseline and post-treatment outcome measurement.

Post-treatment	**Treatment**	**DemPct**
0	0	51.30630
0	1	49.51980
1	0	51.08934
1	1	53.36237

- This step simplifies the model by focusing on aggregate changes, reducing the influence of outliers or extreme values.

3.1.4. Average Treatment Effect (ATT)

The ATT is computed by taking the difference in outcomes between pre-treatment and post-treatment periods across the treatment and control groups. This figure represents the net effect of the election swing.

$$\text{ATE}_{\text{DiD}} = \left(\hat{Y}_{\text{T,post}} - \hat{Y}_{\text{T,pre}}\right) - \left(\hat{Y}_{\text{C,post}} - \hat{Y}_{\text{C,pre}}\right)$$

In our case, the ATT represents the change in Democratic vote percentage is 4.059529.

3.1.5. Counterfactual Outcome and Test the Parallel Trend Assumption

Constructs a counterfactual scenario for what the outcome would have been in the absence of the treatment, enhancing the robustness of causal inference by providing a baseline against which actual outcomes can be compared.

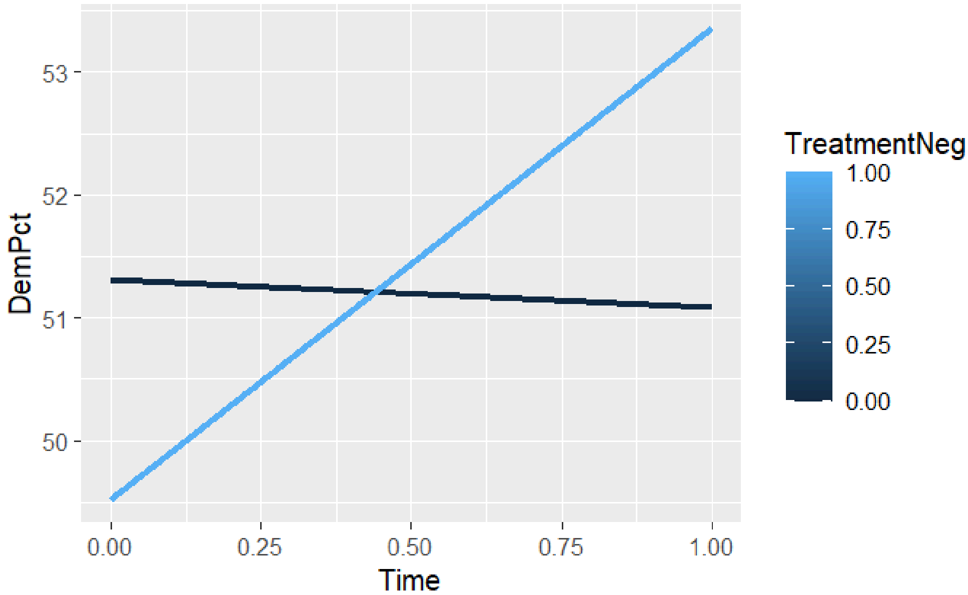


Figure 3: Graphical Representation of the Average Treatment Effect

As can be seen from the image of ATE, due to the existence of covariates, Treatment does not meet the parallel hypothesis, so next we will add covariates to the linear model and PSM for analysis.

3.2. Methodology: Propensity Score Matching (PSM)

Why Use PSM? PSM helps in balancing the observed covariates between the treated and control groups, reducing bias due to confounding variables. By ensuring that the treatment and control groups are comparable on all observed covariates, PSM makes it more plausible that differences in outcomes (like voter turnout) are due to the treatment (negative swing) rather than other factors. In observational studies where the treatment is not randomly assigned, PSM is a powerful tool to mimic randomization, thereby strengthening the validity of causal conclusions.

3.2.1. Model Building

I begin by estimating the propensity scores using a logistic regression model (glm) where the probability of being exposed to a negative swing (TreatmentNeg) is modeled as a function of several covariates (TreatmentPos, DemWin, DWinPrv, DSpndPct, DDonaPct, GovWkPct, UrbanPct). These covariates are chosen based on their potential to influence both the likelihood of experiencing a negative swing and the outcome (voter turnout). **Appendix 5.1**

3.2.2. Propensity Score Calculation

The propensity scores are calculated using the fitted values from the logistic regression, predicting the likelihood that each unit in the dataset receives the treatment based on observed characteristics. **Appendix 5.2**

3.2.3. Matching

With the MatchIt package, you use these propensity scores to match units in the treatment group (those experiencing negative swings) with similar units in the control group (those not experiencing negative swings). The nearest neighbor matching method is used, which pairs units with the closest propensity scores in a one-to-one fashion without replacement. **Appendix 5.3**

3.3. Linear Model with Covariates

After matching, I fit a linear model that includes the covariates used in the propensity score model to estimate the effect of negative swings on voter turnout while controlling for potential confounders.

This model allows us to assess the impact of negative swings on voter turnout more accurately by accounting for the influence of other variables. **Appendix 3.3**

After doing the robustness check, we can get the final best model.

DemPct ~ Time:TreatmentNeg + TreatmentPos + DemWin + DWinPrv + DSpndPct + DDonaPct + GovWkPct + UrbanPct

4. Fingdings

4.1. DiD Results

Since the parallelism assumption is not satisfied, DiD continues to the part where ATE is calculated. Then turn more energy to finding covariates.

4.2. Analysis based on the summary of PSM (Appendix 5)

Main Obervation Items

1. Std. Mean Diff.: This is the standardized mean difference for each covariate between the treatment and control groups. Values closer to zero indicate a good balance.
2. Var. Ratio: The variance ratio between the treated and control groups. A ratio close to 1 indicates similar variance.
3. eCDF Mean and Max: These are the empirical cumulative distribution function mean and maximum differences, respectively. Smaller values suggest that the distribution of covariates between groups is more similar.

From initial model

- Variables such as DSpndPct, DDonaPct, and GovWkPct show reasonable balance in terms of mean differences and variance ratios, suggesting initial comparability.

However, significant differences are noted in distance, and the treatment presence (TreatmentPos) is significantly different, indicating a need for matching to better balance these characteristics.

Post-Matching Balance

- The distance metric's standardized mean difference significantly decreases, highlighting the effectiveness of the matching in aligning the propensity scores between groups.
- DemWin, DWinPrv, and demographic variables like GovWkPct and UrbanPct show improved balance, although some, like GovWkPct, still exhibit relatively larger differences, which may need further investigation or additional covariate adjustment.

Analysis and Interpretation

- The improvements in balance metrics post-matching suggest that the matching process effectively reduces the bias due to observed confounding variables. This enhances the credibility of causal inferences about the effect of negative election swings.
- The matched dataset can now be used to estimate the causal impact of negative swings more reliably, as it minimizes the influence of confounding variables that could otherwise skew the analysis.

4.3. Linear Model Results

- This model effectively quantifies how both positive and negative election swings, along with other political and demographic factors, impact Democratic vote percentages. The significance and effect size of the interaction term (Time:TreatmentNeg) highlight the particular influence of temporal dynamics combined with negative swings, offering insights into how electoral contexts evolve and affect voter behavior over time.

- These findings not only enhance our understanding of electoral dynamics but also provide actionable insights for political strategists and policymakers looking to understand or influence electoral outcomes.

5. Discussion and Conclusion

5.1. Summary of Findings

- This study utilized Difference in Differences (DiD) and Propensity Score Matching (PSM) to explore the impact of negative and positive election swings on voter turnout. Key findings include:
- The presence of negative election swings is associated with a significant, albeit smaller, increase in DemPct over time, suggesting a complex response by voters to shifts in electoral sentiment.
- Other factors such as prior electoral wins, campaign spending, and donations also play substantial roles in influencing DemPct.

5.2. Theoretical Implications

The results affirm the theory that both positive and negative electoral dynamics can mobilize voters, possibly due to heightened political engagement or increased campaign efforts in swing districts. These findings enrich the understanding of electoral behavior, suggesting that voters respond not just to the direction of the swing but to the magnitude and context of the change.

5.3. Practical Implications

For political strategists and campaigners, recognizing the nuanced effects of electoral swings can inform more targeted and effective campaign strategies. For policymakers, understanding these dynamics could guide efforts to enhance voter engagement and turnout, particularly in swing regions.

5.4. Limitations and Future Research

One limitation of this study is its reliance on historical electoral data, which may not fully capture the contemporary political landscape's nuances. Additionally, while PSM improves the robustness of causal inference, it cannot account for unobserved confounders that might still affect the outcomes.

Future research could extend these findings by incorporating more granular data on voter demographics or by examining the role of digital media in influencing voter behavior in swing districts. Longitudinal studies could also assess the long-term effects of repeated swings in the same districts to understand cumulative impacts on voter turnout.

Bibliography

[1] R. S. Erikson and K. Rader, "Replication Data for: Much Ado About Nothing: RDD and the Incumbency Advantage." [Online]. Available: https://doi.org/10.7910/DVN/567RS6

[2] Wikipedia, "Swing (politics) — Wikipedia, The Free Encyclopedia." [Online]. Available: https://en.wikipedia.org/wiki/Swing_(politics)

[3] R. Frank and F. Martínez i Coma, "Correlates of Voter Turnout," *Political Behavior*, vol. 45, pp. 607–633, 2023, doi: 10.1007/s11109-021-09720-y.

[4] MIT Election Data and Science Lab, "Voter Turnout." [Online]. Available: https://electionlab.mit.edu/research/voter-turnout

[5] MIT Election Data and Science Lab, "Voter Confidence and Electoral Participation." [Online]. Available: https://electionlab.mit.edu/articles/voter-confidence-and-electoral-participation

[6] C. Duquette, F. Mixon, and R. Cebula, "Swing States, the Winner-Take-all Electoral College, and Fiscal Federalism," *Atlantic Economic Journal*, vol. 45, p. , 2017, doi: 10.1007/s11293-016-9526-2.

[7] L. Keele, "The Statistics of Causal Inference: A View from Political Methodology," *Political Analysis*, vol. 23, no. 3, pp. 313–335, 2015, doi: 10.1093/pan/mpv007.

[8] R. S. Erikson and K. Rader, "Much Ado About Nothing: RDD and the Incumbency Advantage," *Political Analysis*, vol. 25, no. 2, pp. 269–275, 2017, doi: 10.1017/pan.2017.1.

Appendix for Causal Inference Summative

Z0195806

2024-04-26

Appendix 0 Research Question

"The impact on Final Turnout Rate by Election Swing"

Appendix 1 Read the .dta file

Data Resource: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/567RS6

1.1 Using haven to load the "RD__IP.dta" file

```r
# change to current directory
setwd(getwd())
# install.packages('haven')
library(haven)
data = read_dta("./data/RD_IP.dta")
```

1.2 Variable Explanation

1. **StAlphCd** - State Alpha Code: Abbreviation of the state.
2. **CDNumAtL** - Congressional District Number at Large.
3. **YearElec** - Year of the Election.
4. **Ob** - Observation number or identifier.
5. **Use** - Indicator for whether the data row is usable for analysis.
6. **CngrsNum** - Congress Number, indicating which session of Congress.
7. **StICPSR** - State ICPSR Code, a unique identifier for states in political databases.
8. **StPOAbrv** - State Postal Abbreviation.
9. **DifDPct** - Difference in Democratic Percentage points between elections.
10. **DemPct** - Percentage of vote that went to the Democratic candidate.
11. **DemWin** - Indicator of whether the Democratic candidate won.
12. **DWinPrv** - Indicator of whether the Democrat won the previous election.
13. **DWinNxt** - Indicator of whether the Democrat won the next election.
14. **DifPVDec** - Difference in Presidential Vote share by the decade.
15. **OpenSeat** - Indicates if the election was for an open seat (no incumbent).
16. **IncStatus** - Incumbency status.
17. **PrvTrmsD** - Number of previous terms served by the Democratic candidate.
18. **PrvTrmsO** - Number of previous terms served by the opposing candidate.

19. **VoteTotW** - Total votes for the winning candidate.
20. **VoteTotL** - Total votes for the losing candidate.
21. **VoteCast** - Total votes cast in the election.
22. **SoSDem** - Secretary of State is a Democrat (indicator).
23. **PrvElcOb** - Observation number of the previous election.
24. **NxtElcOb** - Observation number of the next election.
25. **DemInc** - Indicates if the incumbent is a Democrat.
26. **DemOpen** - Indicates if the open seat was previously held by a Democrat.
27. **NonDInc** - Indicates if the incumbent is not a Democrat.
28. **NonDOpen** - Indicates if the open seat was previously held by a non-Democrat.
29. **DExpAdv** - Democratic Experience Advantage.
30. **RExpAdv** - Republican Experience Advantage.
31. **ForgnPct** - Percentage of foreign-born population in the district.
32. **GovWkPct** - Percentage of government workers in the district.
33. **BlackPct** - Percentage of African American population in the district.
34. **UrbanPct** - Percentage of urban population in the district.
35. **VtTotPct** - Voter turnout percentage.
36. **DPctNxt** - Democratic vote percentage in the next election.
37. **DifDPNxt** - Difference in Democratic Percentage in the next election.
38. **DPctPrv** - Democratic vote percentage in the previous election.
39. **DifDPPrv** - Difference in Democratic Percentage in the previous election.
40. **ElcSwing** - Electoral swing.
41. **GovDem** - Indicates if the governor is a Democrat.
42. **IncDWNOM1** - NOMINATE score of the incumbent Democrat.
43. **DSpndPct** - Percentage of campaign spending by the Democrat.
44. **DDonaPct** - Percentage of donations received by the Democrat.
45. **CQRating3** - Congressional Quarterly rating for the district.
46. **DifIPPct** - Difference in Incumbent Party Percentage points.
47. **IPwin** - Incumbent Party win indicator.
48. **close05** - Close election indicator based on a margin of 0.5%.
49. **IPPctNxt** - Incumbent Party percentage in the next election.
50. **IPPctPrv** - Incumbent Party percentage in the previous election.
51. **DifIPPPrv** - Difference in Incumbent Party Percentage in the previous election.
52. **IPInc** - Incumbent Party Incumbency.
53. **PrvTrmsIP** - Previous Terms of Incumbent Party.
54. **IPExpAdv** - Incumbent Party Experience Advantage.
55. **IPSwing** - Incumbent Party Swing.
56. **CQRatingIP** - Congressional Quarterly rating based on Incumbent Party.
57. **IPSpndPct** - Incumbent Party Spending Percentage.
58. **IPDonaPct** - Incumbent Party Donation Percentage.
59. **SoSIP** - Secretary of State from Incumbent Party.
60. **GovIP** - Indicates if the governor is from the incumbent party.
61. **DifPVIP** - Difference in Presidential Vote share attributed to the incumbent party, averaged over the decade.

```
# write.csv(data, "./data/RD_IP.csv", row.names = FALSE)
```

Appendix 2 Exploratory Data Analysis

2.1 Load Required Libraries

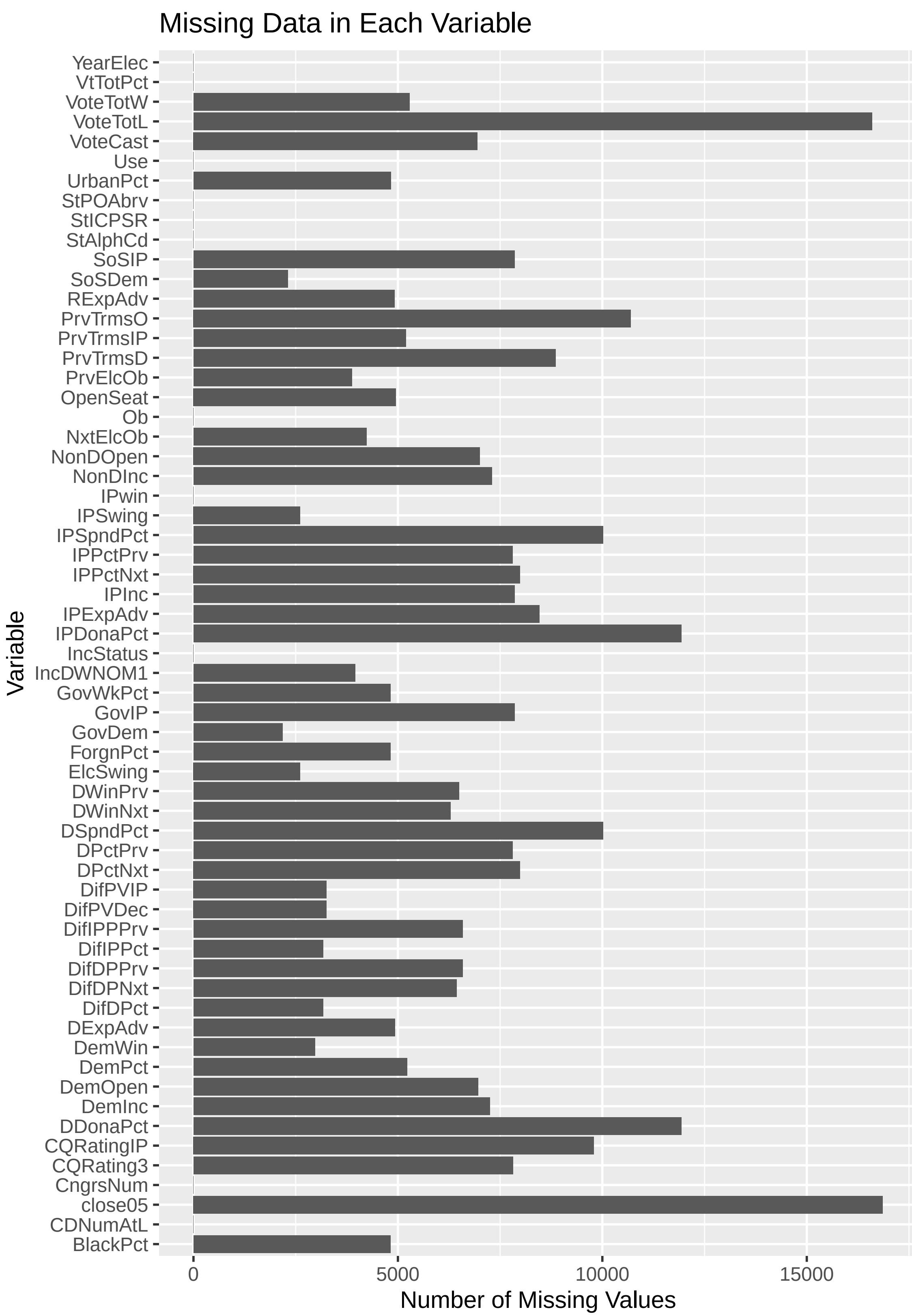
```r
# Load necessary libraries from tidyverse individually
# For creating visualizations
# For data manipulation
library(dplyr)
# # For reading CSV data
# library(readr)
# For data tidying
library(tidyr)
```

2.2 Inspect the Dataset

```r
# View the first few rows of the dataset
# head(data)
# Get a summary of the dataset
# summary(data)
# Check the structure of the dataset
# str(data)
```

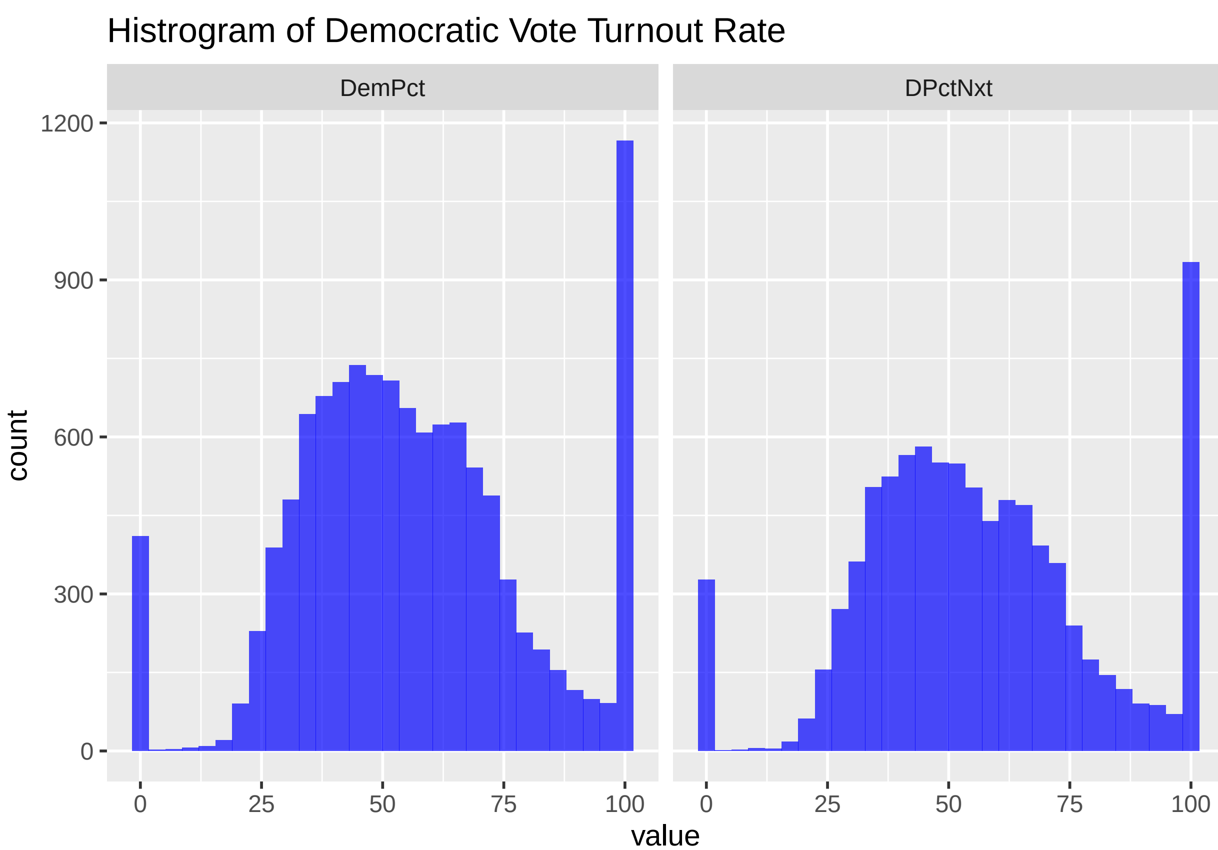
Check the missing values:

```r
# Checking for missing values visually
library(ggplot2)
data %>%
  summarise_all(funs(sum(is.na(.)))) %>%
  gather(key = "variable", value = "n_missing") %>%
  ggplot(aes(x = variable, y = n_missing)) +
  geom_col() +
  coord_flip() +
  labs(x = "Variable",
       y = "Number of Missing Values",
       title = "Missing Data in Each Variable")
```

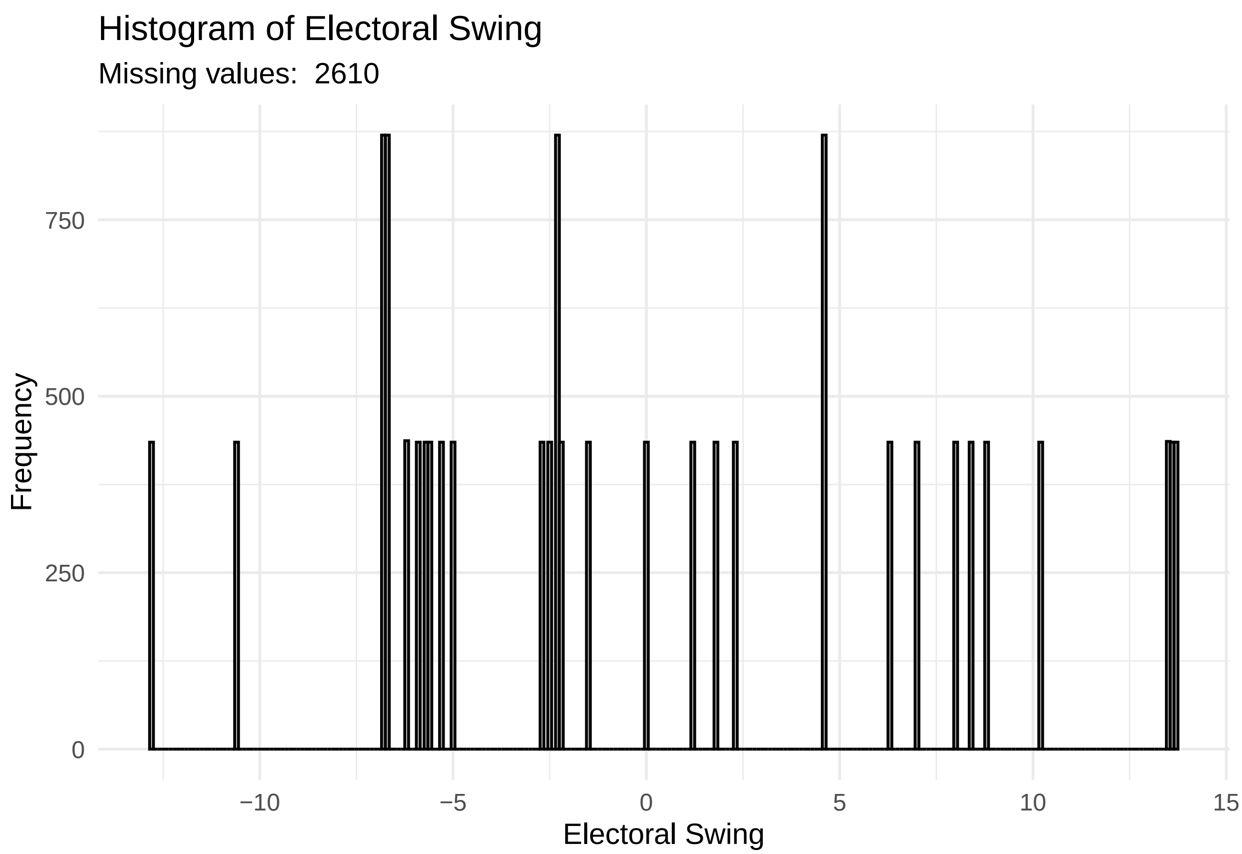

Missing Data in Each Variable
Variable
YearElec
VtTotPct
VoteTotW
VoteTotL
VoteCast
Use
UrbanPct
StPOAbrv
StICPSR
StAlphCd
SoSIP
SoSDem
RExpAdv
PrvTrmsO
PrvTrmsIP
PrvTrmsD
PrvElcOb
OpenSeat
Ob
NxtElcOb
NonDOpen
NonDInc
IPwin
IPSwing
IPSpndPct
IPPctPrv
IPPctNxt
IPInc
IPExpAdv
IPDonaPct
IncStatus
IncDWNOM1
GovWkPct
GovIP
GovDem
ForgnPct
ElcSwing
DWinPrv
DWinNxt
DSpndPct
DPctPrv
DPctNxt
DifPVIP
DifPVDec
DifIPPPrv
DifIPPct
DifDPPrv
DifDPNxt
DifDPct
DExpAdv
DemWin
DemPct
DemOpen
DemInc
DDonaPct
CQRatingIP
CQRating3
CngrsNum
close05
CDNumAtL
BlackPct
0
5000
10000
15000
Number of Missing Values

2.3 Initial Exploratory Data Analysis

```
# Plotting histograms for continuous variables
# such as vote percentages and differences
library(ggplot2)
data %>%
  select(DemPct, DPctNxt) %>%
  gather(key = "variable", value = "value") %>%
  ggplot(aes(x = value)) +
  geom_histogram(bins = 30, fill = "blue", alpha = 0.7) +
  facet_wrap(~variable, scales = "free_x") +
  labs(title = "Histrogram of Democratic Vote Turnout Rate")
```




```
ggplot(data, aes(x = ElcSwing)) +
  geom_histogram(binwidth = 0.1,
                 fill = "grey",
                 color = "black") +
  theme_minimal() +
  labs(title = "Histogram of Electoral Swing",
       subtitle = paste("Missing values: ",
                        sum(is.na(data$ElcSwing))),
       x = "Electoral Swing",
       y = "Frequency")
```




```
data.did.current = data %>%
  select(DemWin, DWinPrv, DemPct, DSpndPct,
         DDonaPct, YearElec, StAlphCd,
         GovWkPct, UrbanPct, ElcSwing, StPOAbrv
  ) %>%
  drop_na() %>%
  mutate(Time = 0)

data.did.next = data %>%
  select(DWinNxt, DemWin, DPctNxt, DSpndPct,
         DDonaPct, YearElec, StAlphCd,
         GovWkPct, UrbanPct, ElcSwing, StPOAbrv
  ) %>%
  drop_na() %>%
  mutate(Time = 1)

data.did.next = data.did.next %>%
  rename(DemWin = DWinNxt,
         DWinPrv = DemWin,
         DemPct = DPctNxt)
data.did.all = bind_rows(data.did.current, data.did.next)
```

Appendix 3 DiD

3.1 The Dataset

Absolute Value of Election Swing

```
data.did.all.abs = data.did.all %>%
  mutate(Treatment = ifelse(abs(ElcSwing) > 5, 1, 0))
```

Pos/Neg Value of Election Swing

```
data.did.all.posneg = data.did.all %>%
  mutate(TreatmentPos = ifelse(ElcSwing > 5, 1, 0)) %>%
  mutate(TreatmentNeg = ifelse(ElcSwing < -5, 1, 0))
```

3.2 First Differences

```
differences <- data.did.all.posneg %>%
  group_by(Time, TreatmentNeg) %>%
  summarise(DemPct = mean(DemPct, na.rm = TRUE))
differences
```

```
## # A tibble: 4 x 3
## # Groups:   Time [2]
##    Time TreatmentNeg DemPct
##   <dbl>        <dbl>  <dbl>
## 1     0            0   51.3
## 2     0            1   49.5
## 3     1            0   51.1
## 4     1            1   53.4
```

3.2.1 The Average Treatment Effect (ATT)

```
T0Trt0 = differences[1, 3]
T0Trt1 = differences[2, 3]
T1Trt0 = differences[3, 3]
T1Trt1 = differences[4, 3]
(T1Trt1-T1Trt0)-(T0Trt1-T0Trt0)
```

```
##     DemPct
## 1 4.059529
```

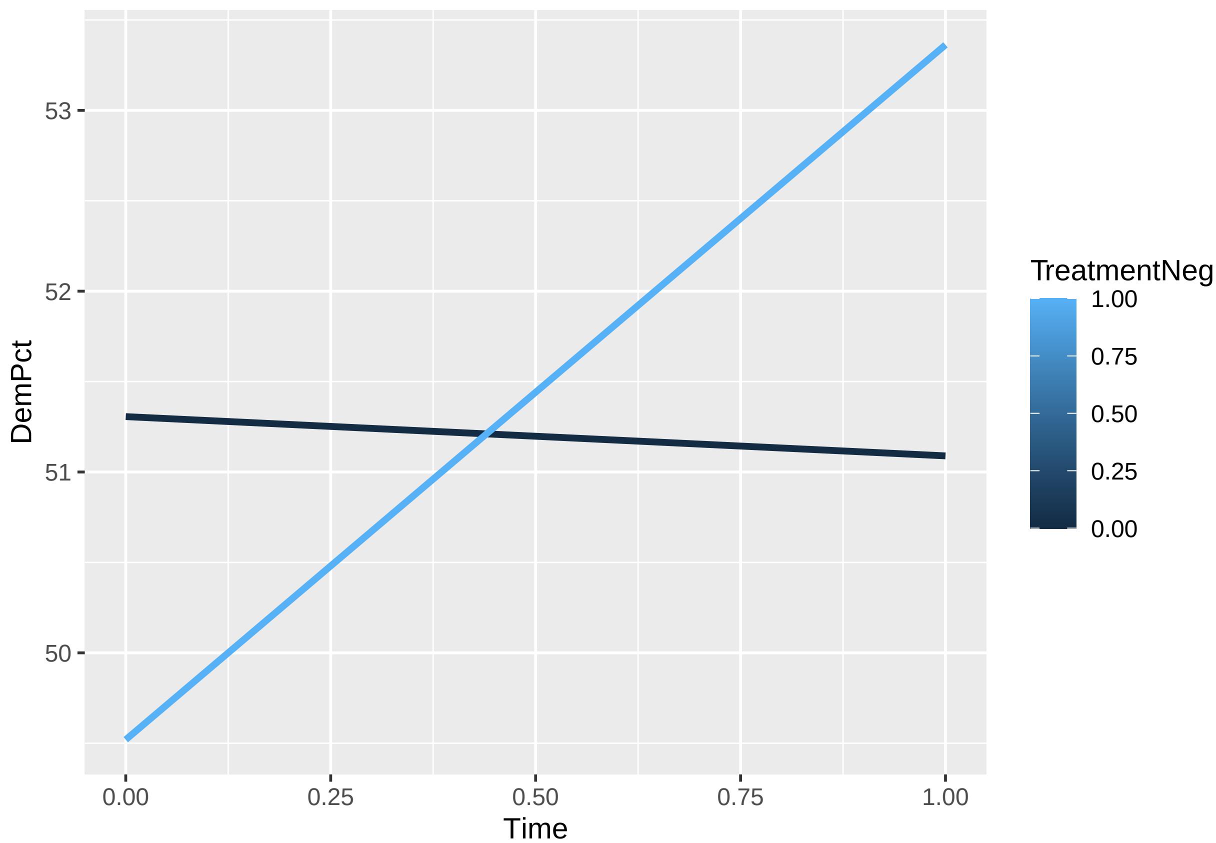
3.2.2 Counterfactual Outcome

```r
# Calculate counterfactual outcome
Trt1_counterfactual = tibble(
  Time = c(0, 1),
  TreatmentNeg = c(0, 1),
  DemPct = as.numeric(c(T0Trt1, T0Trt1-(T0Trt0-T1Trt0)))
)
# combine them together
did_plotdata = bind_rows(differences, Trt1_counterfactual)
did_plotdata
```

```
## # A tibble: 6 x 3
## # Groups:   Time [2]
##    Time TreatmentNeg DemPct
##   <dbl>        <dbl>  <dbl>
## 1     0            0   51.3
## 2     0            1   49.5
## 3     1            0   51.1
## 4     1            1   53.4
## 5     0            0   49.5
## 6     1            1   49.3
```

```r
differences %>%
  ggplot(aes(x=Time, y=DemPct, group=TreatmentNeg)) +
  geom_line(aes(color=TreatmentNeg), size=1.2)
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



3.3 Calculating the DiD Estimator via lm

```
model.1 = lm(
  DemPct ~ Time * Treatment + DemWin + DWinPrv +
    DSpndPct + DDonaPct + GovWkPct + UrbanPct,
  data = data.did.all.abs
)
```

```
summary(model.1)
```

```
##
## Call:
## lm(formula = DemPct ~ Time * Treatment + DemWin + DWinPrv + DSpndPct +
##     DDonaPct + GovWkPct + UrbanPct, data = data.did.all.abs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -51.028  -5.717  -0.361   5.565  47.016
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      25.073779   0.829307  30.235  < 2e-16 ***
## Time             -0.167660   0.497685  -0.337 0.736223
```

```
## Treatment      -0.478259   0.461497  -1.036 0.300109
## DemWin         19.187167   0.582551  32.936  < 2e-16 ***
## DWinPrv         1.946120   0.685397   2.839 0.004541 **
## DSpndPct        0.064185   0.017382   3.693 0.000225 ***
## DDonaPct        0.138305   0.017021   8.125 5.76e-16 ***
## GovWkPct        0.222157   0.079258   2.803 0.005086 **
## UrbanPct        0.031759   0.006258   5.075 4.04e-07 ***
## Time:Treatment  2.188857   0.650196   3.366 0.000768 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.5 on 4359 degrees of freedom
## Multiple R-squared:  0.7199, Adjusted R-squared:  0.7193
## F-statistic:  1245 on 9 and 4359 DF,  p-value: < 2.2e-16
```

```r
model.2 = lm(
  DemPct ~ Time * TreatmentPos + Time * TreatmentNeg +
    DemWin + DWinPrv + DSpndPct + DDonaPct +
    GovWkPct + UrbanPct,
  data = data.did.all.posneg
)
```

```r
summary(model.2)
```

```
##
## Call:
## lm(formula = DemPct ~ Time * TreatmentPos + Time * TreatmentNeg +
##     DemWin + DWinPrv + DSpndPct + DDonaPct + GovWkPct + UrbanPct,
##     data = data.did.all.posneg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -51.790  -5.647  -0.297   5.541  46.730
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       25.081412   0.827153  30.323  < 2e-16 ***
## Time              -0.172451   0.496254  -0.348 0.728228
## TreatmentPos       1.543819   0.653081   2.364 0.018127 *
## TreatmentNeg      -1.308521   0.496844  -2.634 0.008477 **
## DemWin            18.967111   0.582877  32.541  < 2e-16 ***
## DWinPrv            2.337877   0.687568   3.400 0.000679 ***
## DSpndPct           0.063938   0.017341   3.687 0.000230 ***
## DDonaPct           0.136645   0.017000   8.038 1.17e-15 ***
## GovWkPct           0.217617   0.079110   2.751 0.005969 **
## UrbanPct           0.032313   0.006258   5.164 2.53e-07 ***
## Time:TreatmentPos -1.128221   0.918746  -1.228 0.219513
## Time:TreatmentNeg  3.554766   0.701508   5.067 4.20e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.47 on 4357 degrees of freedom
## Multiple R-squared:  0.7216, Adjusted R-squared:  0.7209
## F-statistic:  1027 on 11 and 4357 DF,  p-value: < 2.2e-16
```

```r
model.3 = lm(
  DemPct ~ Time:TreatmentNeg + TreatmentPos +
    DemWin + DWinPrv + DSpndPct + DDonaPct +
    GovWkPct + UrbanPct,
  data = data.did.all.posneg
)
```

```r
summary(model.3)
```

```
##
## Call:
## lm(formula = DemPct ~ Time:TreatmentNeg + TreatmentPos + DemWin +
##     DWinPrv + DSpndPct + DDonaPct + GovWkPct + UrbanPct, data = data.did.all.posneg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -51.517  -5.705  -0.296   5.637  46.579
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       24.331908   0.755064  32.225  < 2e-16 ***
## TreatmentPos       1.402897   0.441039   3.181 0.001479 **
## DemWin            19.162608   0.579352  33.076  < 2e-16 ***
## DWinPrv            2.038605   0.681004   2.994 0.002773 **
## DSpndPct           0.063045   0.017354   3.633 0.000283 ***
## DDonaPct           0.138678   0.017003   8.156 4.48e-16 ***
## GovWkPct           0.242600   0.078685   3.083 0.002061 **
## UrbanPct           0.033216   0.006255   5.310 1.15e-07 ***
## Time:TreatmentNeg  2.573221   0.404649   6.359 2.24e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.48 on 4360 degrees of freedom
## Multiple R-squared:  0.7209, Adjusted R-squared:  0.7204
## F-statistic:  1408 on 8 and 4360 DF,  p-value: < 2.2e-16
```

Appendix 4 Synthetic Control

```r
# install.packages("Synth")
```

```r
data.synth.pre = data %>%
  select(DemWin, DWinPrv, DemPct, DSpndPct,
         DDonaPct, YearElec, StAlphCd, StPOAbrv,
         GovWkPct, UrbanPct, ElcSwing, StAlphCd, YearElec
  ) %>%
  drop_na() %>%
  mutate(Time = 0)

data.synth.post = data %>%
  select(DWinNxt, DemWin, DPctNxt, DSpndPct,
```

```r
         DDonaPct, YearElec, StAlphCd, StPOAbrv,
         GovWkPct, UrbanPct, ElcSwing, StAlphCd, YearElec
  ) %>%
  drop_na() %>%
  mutate(Time = 1)

data.synth.post = data.synth.post %>%
  rename(DemWin = DWinNxt,
         DWinPrv = DemWin,
         DemPct = DPctNxt)

data.synth.all <- bind_rows(data.synth.pre, data.synth.post)
```

```r
# Define the treated unit - example: State with code 'CA'
treated_unit <- "CA"
```

```r
# Variables used as predictors
predictor_vars <-
  c("DemPct", "DSpndPct", "DDonaPct",
    "GovWkPct", "UrbanPct")
```

```r
library(Synth)
```

```
## ##
## ## Synth Package: Implements Synthetic Control Methods.

## ## See https://web.stanford.edu/~jhain/synthpage.html for additional information.
```

```r
# data.prep <- dataprep(
#   foo = complete_cases_data,
#   predictors = c("DemPct", "DSpndPct", "DDonaPct", "GovWkPct", "UrbanPct", "ElcSwing"),
#   predictors.op = "mean",
#   time.predictors.prior = 2000:2008,
#   special.predictors = list(
#     list("DemWin", 2000:2008, "mean")
#   ),
#   dependent = "DemWin",
#   unit.variable = "StAlphCd",
#   unit.names.variable = "StPOAbrv",
#   time.variable = "YearElec",
#   treatment.identifier = 1,
#   controls.identifier = setdiff(unique(complete_cases_data$StAlphCd), 1),
#   time.optimize.ssr = 1930:2008,  # Ensure this period is covered by your data
#   time.plot = min(complete_cases_data$YearElec):max(complete_cases_data$YearElec)
# )
```

Appendix 5 PSM

```r
# install.packages("MatchIt")
```

5.1 Model Building

```r
data.glm = data.synth.all %>%
  mutate(TreatmentPos = ifelse(ElcSwing > 5, 1, 0)) %>%
  mutate(TreatmentNeg = ifelse(ElcSwing < -5, 1, 0))
model.psm = glm(TreatmentNeg ~ TreatmentPos +
    DemWin + DWinPrv + DSpndPct + DDonaPct +
    GovWkPct + UrbanPct, data = data.glm,
    family = "binomial")
```

5.2 Propensity Score Calculation

```r
data.glm$pscore = predict(model.psm, type = "response")
```

5.3 Matching

```r
library(MatchIt)
matchit_res <- matchit(TreatmentNeg ~ TreatmentPos +
    DemWin + DWinPrv + DSpndPct + DDonaPct +
      GovWkPct + UrbanPct,
    data = data.glm,
    method = "nearest", distance = "logit")
matched_data <- match.data(matchit_res)
```

```r
summary(matchit_res)
```

```
##
## Call:
## matchit(formula = TreatmentNeg ~ TreatmentPos + DemWin + DWinPrv +
##     DSpndPct + DDonaPct + GovWkPct + UrbanPct, data = data.glm,
##     method = "nearest", distance = "logit")
##
## Summary of Balance for All Data:
##              Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## distance            0.5261        0.3487          1.8881     0.1604    0.2247
## TreatmentPos        0.0000        0.2916         -0.8453          .    0.2916
## DemWin              0.5308        0.5431         -0.0247          .    0.0123
## DWinPrv             0.5653        0.5292          0.0729          .    0.0361
## DSpndPct           53.0047       51.5699          0.0461     0.8441    0.0337
## DDonaPct           54.7841       52.9094          0.0531     0.8725    0.0303
## GovWkPct            6.7452        7.2394         -0.2364     1.1465    0.0904
## UrbanPct           69.3892       70.4412         -0.0397     1.0717    0.0115
##              eCDF Max
```

```
## distance       0.3236
## TreatmentPos   0.2916
## DemWin         0.0123
## DWinPrv        0.0361
## DSpndPct       0.0672
## DDonaPct       0.0706
## GovWkPct       0.1360
## UrbanPct       0.0274
##
## Summary of Balance for Matched Data:
##              Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## distance            0.5261        0.4739          0.5562     0.5692    0.1086
## TreatmentPos        0.0000        0.0373         -0.1080          .    0.0373
## DemWin              0.5308        0.5556         -0.0498          .    0.0248
## DWinPrv             0.5653        0.5497          0.0316          .    0.0157
## DSpndPct           53.0047       52.7492          0.0082     0.8291    0.0374
## DDonaPct           54.7841       54.5380          0.0070     0.8519    0.0353
## GovWkPct            6.7452        7.2707         -0.2514     1.1260    0.0938
## UrbanPct           69.3892       71.4301         -0.0771     1.2347    0.0242
##              eCDF Max Std. Pair Dist.
## distance       0.1879          0.5562
## TreatmentPos   0.0373          0.1080
## DemWin         0.0248          0.9651
## DWinPrv        0.0157          0.9575
## DSpndPct       0.0686          1.2120
## DDonaPct       0.0718          1.1959
## GovWkPct       0.1485          0.7954
## UrbanPct       0.0621          0.9758
##
## Sample Sizes:
##           Control Treated
## All          2517    1852
## Matched      1852    1852
## Unmatched     665       0
## Discarded       0       0
```