THE IMMEDIATE AND SUBSEQUENT EFFECTS OF RANKED-CHOICE VOTING
ON VOTER TURNOUT

Sarah Kornfeld

June 2022

A thesis submitted to the department of
Mathematical Methods in the Social Sciences (MMSS)

Advised by Professor John Bullock

Northwestern University
# Abstract

2

# Acknowledgments

3

# Introduction

4
- Ranked-Choice Voting
  - Ranked-Choice Voting and Voter Turnout

5

# Literature Review

7

# Data

10
- Data Collection Methods
11
- Advantages of Dataset
11
- Drawbacks of Dataset
11

# Methods

15
- Hypotheses and Predictions
15
- Models
  - Synthetic Control
  - Linear Regression
16

# Results and Analysis

19
- Synthetic Control: Hennepin County
  - Hennepin County: Presidential Elections
  - Hennepin County: State Elections
19
- Synthetic Control: Ramsey County
  - Ramsey County: Presidential Elections
  - Ramsey County: State Elections
23

# Conclusion

28

# Bibliography

31

# Appendix A: R Code for Scatterplot & Linear Modeling

34

# Appendix B: R Code for Synthetic Control Method

37
Abstract

Ranked-Choice Voting (RCV) is a voting system where voters are tasked to rank the candidates in an election. A candidate wins if they have received over 50% of the first-choice vote. If not, the candidate with the least first-choice votes is eliminated, and thus voters' rankings are accordingly adjusted until a winner emerges. There are contradicting opinions about how this voting system affects voter turnout. This paper highlights the effect of RCV on voter turnout in both state elections and presidential elections in Minnesota. The data shows that RCV tends to increase voter turnout and that the increase is more pronounced in subsequent elections after the implementation of RCV.
Acknowledgments

I would first like to thank my parents and siblings for supporting me throughout my academics; I could not have done it without your help. To my housemates, thank you for celebrating every little win with me all year. To my friends and classmates, late-night Main, last-minute problem sets, and every class in between would not have been half as fun without all of you. To Ola, thank you for helping me sort through the data and analysis. I could not have completed this thesis without the hours spent on Zoom and your countless ideas on how to solve each data issue that arose. And, to my advisor, Professor John Bullock, thank you for guiding me through this process and supporting me no matter how far behind I was.
Introduction

The strength of the United States democracy depends on citizens' participation in politics and, thus, exercising their right to vote. Specifically, a high voter turnout is crucial because it increases the likelihood that those in office represent the general public's views (Root & Kennedy, 2018, p. 1). While many citizens cannot vote due to voter suppression, ex-offender laws, and other barriers, others do not vote because they feel disconnected from the current political system and believe their vote will have little impact (Root & Kennedy, 2018, p. 1). 65% of registered voters ages 18-25 who did not vote in 2016 did not vote because they “didn’t like the candidates/issues” (Circle, 2018). This statistic may result from the fact that despite having a two-party system, in 2013, 42% of voters identified as independent, which stems from isolation from both parties (Jones, 2014). Thus, our election system does not incorporate the views of a large proportion of our country. Because of the two-party system, many liberal ideas are conflated with the Democratic Party and many conservative ideas with the Republican Party. This leads voters to believe that they cannot express their political preferences through voting without supporting a two-party candidate. To ease this alienation, a proposed method is to change the voting system to ranked-choice voting from plurality voting. One way ranked-choice voting may increase voter turnout is by providing space for candidates that do not fit into current Democratic and Republican molds. Therefore, more third-party candidates will be motivated to run for office, thereby increasing voter turnout by mobilizing voters who may agree with these non-establishment candidates’ views (FairVote, n.d.)
**Ranked-Choice Voting**

Ranked-choice voting is a system in which voters rank their preferences for candidates rather than indicating only their top choice. Voters receive a ballot where they can display the order of their candidate preference. A candidate must receive over 50% of the first-choice vote to win. If no candidate gets the majority of the first-choice votes, the candidate with the lowest percentage of first-choice votes is eliminated. For ballots where the eliminated candidate was the first-choice, the ballot’s votes are reordered to account for the eliminated candidate, bumping the voter’s second-choice vote to the first-choice spot. Next, the votes are recounted. This process continues until one candidate wins the election by receiving over 50% of the first-choice votes (Ballotpedia, n.d.).

**Ranked-Choice Voting and Voter Turnout**

Ranked-choice voting elections can impact election participation because it allows voters to vote for the candidate that most aligns with their beliefs without the fear of “wasting” their vote if their chosen candidate is not the winner of the Democratic or Republican primary election. Since 1920, only four third-party presidential candidates have won at least one electoral vote (Klein, 2018). Yet, despite the almost zero probability that they will win a presidential election, the existence of third-party candidates in presidential elections has impacted American politics. In the 2000 presidential election, the winner of the presidency needed to win Florida. Ultimately, George Bush won Florida by fewer than 600 votes, giving him the victory (The Editors of Encyclopaedia Britannica, 2021). Ralph Nader, the Green Party candidate, however, received over 100,000 votes in Florida (Klein, 2018). As the Green Party tends to fall on the left of the two-dimensional political spectrum, it is assumed that many of these 100,000 voters would
have ranked Al Gore above George Bush if given the opportunity to indicate their preferences, which would have enabled Gore to become President (Yue, 2015). While some critics categorize these participants as wasting their votes, others believe it is an integral part of the democratic process to indicate their dissatisfaction with the two mainstream candidates. Therefore, RCV uniquely allows one to vote for the candidate that resonates with them while casting a vote that will contribute to one’s preferred mainstream candidate’s campaign.
Literature Review

During the summer of 2020, ranked-choice voting (RCV) was a popular topic in political news due to the re-adoption of the voting system for New York State’s gubernatorial election (Montanaro, 2021). Many opinions on RCV flooded the headlines, both in favor and against the voting system, citing the ability to vote for one’s favorite candidate rather than strategizing who may win and citing the failure to truly elect the most popular candidate (Vanden Heuvel, 2021; Jacoby 2020). Additionally, there are many retrospective hypotheses claiming different election outcomes with the implementation of RCV, both in favor of liberal and conservative agendas. Due to the Libertarian Party being the third most, albeit distant, popular party, researchers believed that Trump would have gained more electoral votes under RCV in 2020, possibly supporting a Trump victory (Cervas & Grofman, 2021, p. 4). Others believe that Hillary Clinton could have won the 2016 presidential election using RCV (Matthews, 2016). Thus, because RCV can potentially change political outcomes and many jurisdictions across the US have adopted it, ranked-choice voting is a timely topic to explore.

The case for ranked-choice voting is supported by many due to RCV’s ability to positively impact the political climate. Namely, ranked-choice voting motivates minorities to vote, encourages civil campaigning and political discourse, and improves public views of campaigns (Brockington et al., 1998; Donovan et al., 2016, p. 162). When comparing minority mobilization in a cumulative voting system—another name for ranked-choice voting—to a limited voting system, researchers found that more minorities voted (Brockington et al., 1998). While the direct cause of this shift in voter demographics is unknown, researchers hypothesize that the increased opportunity of minority candidates to receive the nomination in RCV motivates more minorities to engage in the election (Brockington et al., 1998). Thus, RCV can
increase minority representation in the electorate and in elected positions. In addition to changing the makeup of people engaged in the elections, RCV also positively shifts the public’s opinion about elections. In a study comparing RCV to a plurality election, voters found that RCV elections were less harmful because candidates criticized each other less (Donovan et al., 2016, p. 162). Further research also indicated that through the increased positive perception of politics, young voter turnout increased in RCV elections (Juelich & Coll, 2021, p. 329). Therefore, the restructuring of the voting method mobilizes young voters, counteracting their overwhelmingly negative view of politics. In addition to increasing voter turnout in regular elections, RCV increases voter participation in runoff elections because it does not require another ballot to be cast (Kimball & Anthony, 2016, p. 21). Therefore, it further allows the public view to be represented in their elected officials because, historically, there is a high drop-off from the general election to the runoff election, and RCV avoids the need for multiple elections.

The advantages of ranked-choice voting can be diminished due to the difficulty of the process. Ranked-choice voting requires knowledge about one’s individual preferences, all the candidates, and the rules of ranked-choice voting. There is a concern that these difficulties will disproportionately affect minorities as systemic barriers give them less access to tools to increase this knowledge. Precisely, RCV is thought to be difficult for voters with low levels of education (McDaniel, 2016, p. 402). As low educational attainment tends to be correlated with race due to underfunding and systemic barriers, RCV has the potential to discourage minority voter turnout (Darling-Hammond, 1998). To explore this concern, researchers asked a sample of Democratic voters to engage in ranked-choice voting among candidates in the Democratic primary election. This sample demonstrated no significant difference in perceived difficulty between race or gender (Coll, 2021, p. 303). Instead, the researchers found that older, less politically interested,
and more ideologically conservative participants self-reported a higher difficulty with RCV (Coll, 2021, p. 294). Thus, people who do not know much about the Democratic primary candidates and are older struggle disproportionately. This finding is hopeful since it is not a systemic issue but rather somewhat related to voter knowledge. Thus, greater education from more explicit instructions or better preparation (i.e., a nonpartisan voter guide) can help alleviate this difficulty and allow RCV to increase voter turnout.

Yet, in contrast to the other findings on the impact of RCV on voter turnout, in San Francisco’s 2009 mayoral election, both Black and white voter turnout decreased compared to other racial groups (McDaniel, 2016, p. 388). However, as most of the research on RCV focuses on natural studies, including the research in this paper, the research lacks external validity. Therefore, due to the discrepancy in findings across different samples, more research needs to be conducted about how RCV affects voter turnout. Is the effect constant throughout time, or does turnout for subsequent RCV elections differ from the first election?
Data

Minnesota provides a compelling case study to evaluate the effects of RCV because of the variation of election types; some districts use ranked-choice voting while others continue to use plurality voting. Therefore, all of the analyzed data describes Minnesotan counties. Specifically, the data contains samples from Ramsey County and Hennepin County to represent treated counties as they both transitioned from plurality voting to RCV in 2011 and 2009, respectively, and samples from 80 other Minnesotan counties that continue to use plurality voting. The data spans from 2000 to 2019, capturing both the pre-treatment period and the post-treatment period of the implementation of RCV. Moreover, because the data spans 2019, we can analyze both the immediate effect of the shift to RCV and the longer-term impact.

The primary dataset with county-level voter turnout was provided by Eamon McGinn, a Ph.D. candidate at the University of Technology Sydney, researching the effects of instant run-off voting on participation and civility. I merged McGinn’s dataset with multiple datasets from Data USA that described Minnesota county-level racial and ethnic demographic data from 2013 to 2019. But, I could not use this merged dataset to analyze the combined effect of RCV and voter demographics on turnout because it lacks demographic data from the pre-treated period. Additionally, the treated counties differed from the untreated counties on more variables than the implementation of RCV, failing to follow the requirement of parallel trends for a difference-in-difference analysis.

The merged dataset contains two main categories of variables: voting variables and demographic variables. The voting variables describe the type of election, the voter turnout, the number of pre-registered and daily-registered voters, and information about the frequency of election format. The demographic variables represent the proportion of residents who identify
with a racial identity category and if they identify as Hispanic or Latino, with each county’s proportions summing to one.

**Data Collection Methods**

The large data set was obtained by cleaning and merging the dataset from Eamon McGinn and 82 datasets describing county racial and ethnic demographics. I received the data from McGinn over email as it is not publically available, and he has not completed the paper that uses this data. I downloaded the demographic data from Data USA from datausa.io because it is publicly available.¹

**Advantages of Dataset**

The dataset covers voter turnout from 2000 to 2019 in 82 of the 87 counties in Minnesota. As Hennepin County and Ramsey County adopted ranked-choice voting in 2009 and 2011, the period the data spans provide an ideal range of dates, providing data from before the change, data from right after the change, and data from a few years after the change. This range allows the dataset to be helpful in evaluating the immediate and subsequent effects of RCV on voter turnout.

**Drawbacks of Dataset**

I approached this data with the intent to use heterogeneity to identify how the interactions of election type with race and ethnicity impacted voter turnout. This proposed analysis will help determine if there are race-based barriers or equalizers of ranked-choice voting. For example, suppose the data identified race-based barriers to RCV displayed through a lower turnout in

¹ [https://datausa.io/profile/geo/minnesota](https://datausa.io/profile/geo/minnesota)
neighborhoods with a lower proportion of white individuals than those neighborhoods with a predominantly white population. In this case, future research could focus on how to alleviate these disparities. If the barriers were irreconcilable, the results could help motivate the termination of ranked-choice voting because it exacerbates unequal voting opportunities. On the other hand, if the data demonstrated that RCV increased voter turnout among minority voters, the analysis could motivate more voting districts to implement RCV voting in future elections. If a median result arose—for example, minority voter turnout only increased after the second election that used RCV—districts could implement measures to improve voter knowledge and experience with RCV before an upcoming election. Therefore, these results have the potential to impact future election formats.

Yet, because ranked-choice voting has gotten the most traction in urban counties with diverse constituents, this data fails to satisfy the heterogeneity requirements necessary for most analysis. As shown in Figure 1, there is a high correlation between higher than average Black and African American counties and counties that adopted ranked-choice voting in Minnesota. Therefore, because the treated counties all have a higher proportion of Black constituents than average, we cannot see the effect of implementing RCV because it is unclear if the difference in voter turnout results from RCV or voter demographics.

Figure 1 is a scatterplot, where each data point is a unique Minnesotan county in a year ranging from 2000 to 2019. The X-axis displays the proportion of the county’s population who self-identify as white. The Y-axis indicates the proportion of the county’s population who identify as Black and/or African American. Overall, this scatterplot illustrates that all Minnesotan counties in the last 20 years have had a high white population and a low Black population. But, the cluster towards the top of the figure displays a group of counties in recent
years where the Black population represents over 10% of the county’s population. The color of a point, blue or coral, respectively indicates whether or not that county used RCV in the specified year. The cluster of blue points demonstrates the lack of heterogeneity. Almost all counties with Black populations exceeding 10% are treated, and all treated counties have a Black population above 10%. Thus, this figure exemplifies how the treatment—usage of ranked-choice voting—and racial diversity are highly correlated, preventing the analysis of ranked-choice voting by voter demographics. This trend is apparent in all racial groups.

Figure 1: Racial Demographics and RCV in Minnesota

Figure 1 displays the proportion of residents who identify as Black or African American against the proportion of residents who identify as white in Minnesota. Each data point is one county, in one year from 2000 to 2019. The blue points are counties and years that used RCV, while the coral points are counties and years that used plurality voting.
Additionally, the demographic data only spans from 2013 to 2019. Therefore, the pre-treatment period predates the available demographic data. While other demographic data exists, the available datasets fail to segment it by county or only exist in 10-year increments. Thus, the interaction of racial and ethnic demographics and turnout cannot be compared in both pre and post-treatment periods.

Lastly, the novelty of ranked-choice voting limits the potential for data analysis. While many cities adopted RCV in the early 20th century, by the 1950s, most states shifted back to a plurality voting system (Zoch, 2020, p. 1). Just recently did ranked-choice voting gain traction again. Counties in Minnesota only started using RCV again in 2009, and even then, only a few of them adopted the method (FairVote, n.d.). Therefore, we lack data to make confident conclusions about the effects of ranked-choice voting because of the few counties and the few years where these elections exist and the lack of comparable voting districts.
Methods

As noted, initially, I planned to use regressions and difference-in-difference models to identify the effect of the interaction of race and RCV on voter turnout to drive policy changes. However, because of the homogeneity of the treated counties, this was no longer possible. Therefore, I regarded the two treated counties, Hennepin and Ramsey, as case studies and used the synthetic control method. This method allows speculation about how the treated counties would have progressed if they had not been treated, driving an estimate indicating the effect of the treatment.

I will use linear regression to estimate the long-term effects of ranked-choice voting on voter turnout. This method will provide a future template for interpreting how voter turnout differs from subsequent RCV elections, displaying the predicted learning curve.

Hypotheses and Predictions

I predict that voter turnout will decrease in the first RCV election but increase in subsequent elections. Past research found that older and less politically interested participants reported more difficulty with ranked-choice voting (Coll, 2021, p. 294). Therefore, with uncertainty around the method, less politically engaged members of the electorate may be discouraged from voting in the first RCV election. However, I believe this hesitation will be short-lived as familiarity with the method will increase engagement.

Additionally, RCV will ultimately increase voter turnout because it allows for a more diverse ballot that encourages minority voters to vote. I predict that this trend will be present in both districts because their similar demographic makeup made them incomparable to one
another. Thus, I will compare both districts to similar linear combinations of other non-treated counties.

Models

Synthetic Control

Synthetic Control allows for causal inference of ranked-choice voting elections on voter turnout in two counties. To do so, a treated county is compared to a weighted sum of all the untreated counties, the controls. This process is completed for both Hennepin County and Ramsey County. Therefore, the treatment effect size is calculated with the following equation (Swarup, 2021):

\[ \tau_t = Y^I_{jt} - Y^N_{jt} \]

where \( \tau_t \) is the effect size at a given time period after treatment. \( Y^I_{jt} \) represents the outcome value of the treated county and \( Y^N_{jt} \) represents the outcome value of the linear combination of the control group. In this paper, \( Y^I_{jt} \) represents the voter turnout at time \( t \) of the indicated treated county. \( Y^N_{jt} \) is calculated with the following equation:

\[ \hat{Y}^N_{jt} = \sum_{j=2}^{J+1} \omega_j Y^i_{jt} \]

where \( j=2 \) through \( J+1 \) are the indexes for the untreated counties, \( \omega_j \) is the weight given to a specified \( j \) county and \( Y^i_{jt} \) is the turnout in each time period \( t \) for each untreated county. The specified weight for each untreated county is determined through finding the linear combination that makes the control group most similar to the treated country before RCV was implemented.
Mathematically, this is calculated by selecting the weights of each control unit to minimize the following equation:

$$
T_0 \sum_{t=1}^{T_0} (Y_{1t} - \sum_{j=2}^{J} \omega_j(V)Y_{jt})^2
$$

where $T_0$ is the last period before the treatment, $J$ is the total number of control counties, $Y$ is the outcome variables, and $\omega(V)$ is the function determining weight. This equation will be minimized when the synthetic control is the closest to the case study group in the pre-treatment period, allowing for a more accurate estimate of treatment effect size (Cunningham, 2021).

**Linear Regression**

I will also use a linear regression model to evaluate the prediction that ranked-choice voting will initially decrease then ultimately increase voter turnout. This prediction is based on the idea that the introduction of RCV will cause voter confusion and difficulty that will eventually be eased once the system becomes mainstream. Specifically, I regress the voter turnout on the first, second, third, and fourth ranked-choice voting elections in a county with the following equation:

$$
Turnout = \beta_0 + \beta_1 * firstRCVelection + \beta_2 * secondRCVelection
\quad + \beta_3 * thirdRCVelection + \beta_4 * fourthRCVelection
$$

where firstRCVelection, secondRCVelection, thirdRCVelection, and fourthRCVelection are binary variables taking on the value of 1 if the election corresponds the noted number of RCV elections since implementation and 0 otherwise. $\beta_0$ is the baseline, representing the average voter turnout for a non-RCV election in Hennepin County or Ramsey County. The coefficients on each variable describe the expected change in turnout for the corresponding election compared to the
turnout for the baseline, a non-RCV election. Because of the limited treated counties, the analysis is conducted at the voting district level rather than the county level.
Results and Analysis

Next, I will discuss the synthetic control method findings and the preliminary linear regression analysis. I ran four separate synthetic control models—two for each treated county—analyzing voter turnout. Additionally, I ran a linear regression, regressing turnout on elections since RCV implementation.

Synthetic Control: Hennepin County

Hennepin County has three different types of elections: presidential, state, and mayoral elections. I segmented the data by election type and analyzed them separately because of the assumption that voter turnout would differ as a function of election type. Therefore, I looked at the impact of RCV on voter turnout by type of election rather than across all elections in a county as a whole. However, the mayoral data was excluded because it fails to fit the data qualifications for synthetic control. Specifically, there are not enough control counties that hold mayoral elections in the same years as Hennepin County. Thus, for Hennepin county, I analyzed the effect of RCV on voter turnout for state elections and then again for presidential elections.

State elections elect state officials during non-presidential election years. They occur every four years and start in 2002 in this dataset. Presidential elections also happen every four years and begin in 2000 in this dataset. As the synthetic control method compares the treated county to a linear combination of untreated counties, we assume that the difference in voter turnout in the post-treatment period results from introducing ranked-choice voting for each election type.
Hennepin County: Presidential Elections


Figure 2 displays the synthetic control model for Hennepin County’s presidential elections. The graph indicates that while presidential election voter turnout was consistently lower in Hennepin County than in the synthetic control group, the difference decreased with the implementation of ranked-choice voting. Specifically, in the pre-treatment period, the difference in voter turnout is on average 8%. In 2012, the difference decreased to 1.5%, and in 2016 to 3%, showing a positive effect. Additionally, Hennepin County turnout averages at 72% in the pre-treatment period. In comparison, in 2012, the voter turnout was 81.5%, and in 2016, 79.5%. Thus, there is also an increase in voter turnout, displaying a positive effect in terms of the total turnout. Therefore, the results follow the overarching prediction that RCV will increase voter turnout. But, these results contradict the learning curve hypothesis as the difference between 2012 and 2016 increased and voter turnout decreased. While these results may display that excitement over RCV translated to voter turnout immediately peaking and then dropping, the results may also indicate another story, such as excitement in candidates or voter suppression.
The synthetic control method creates a control group that is most similar to the pre-period of the treatment group based exclusively on voter turnout. Voter turnout is the sole similarity measure because I could not obtain demographic data for the pre-period. Therefore, there is no guarantee that the treated counties and control counties are similar; the treated counties are significantly more diverse. Thus, there is a possibility that the change is due to other efforts to mobilize minority voter bases.

**Hennepin County: State Elections**

In the dataset, Hennepin County had five years of state elections: 2002, 2006, 2010, 2014, and 2018. As Hennepin County introduced RCV in 2009, only two years fall in the pre-treatment period and three in the post. While state elections also occur in years with presidential elections, those years are not included in this analysis because presidential
candidates tend to drive more significant voter turnout. Thus, only data points from midterm elections and state general elections are included here.

Figure 3 illustrates the synthetic control output for Hennepin County’s state elections. In the pre-treatment period, voter turnout for both the control county and Hennepin County fluctuated around 65%, with turnout higher in the control in 2002 and turnout higher in Hennepin county in 2006. In 2010, the first state election using RCV, voter turnout in Hennepin County dropped to 57%, while turnout in the control increased above 2006 levels. Thus, initially, there was a negative effect in terms of both the difference in voter turnout and the mere voter turnout. Then, in 2014, the second state election using RCV, voter turnout in Hennepin County dropped to 67.5%, the lowest in the entire period. In addition, the turnout difference between Hennepin County and the synthetic control reached its largest at 7.5%. Thus, on both accounts, RCV again harmed voter turnout. Then, in 2018, voter turnout in Hennepin County reached 77%, surpassing the synthetic control turnout at 75%. Thus, mere turnout dramatically increased in the second RCV state election. The voter turnout in Hennepin County exceeded the voter turnout in the synthetic control county, displaying an extremely positive effect. These results align with the overarching and specific predictions as voter turnout ultimately increased, but only after an initial decrease. Thus, these results back the learning curve hypothesis.
Figure 3: Synthetic Control on Hennepin County State Elections

**Synthetic Control: Ramsey County**

I repeated the same synthetic control process for Ramsey County as I did for Hennepin County. In contrast, Ramsey County has four election types: presidential, state, mayoral, and municipal elections. Similarly, I segmented the data by election type and found that the mayoral election data and the municipal election date fail to fit the data qualifications for synthetic control. Just as with Hennepin County, there were not enough untreated mayoral and municipal elections in the same year as Ramsey County’s mayoral and municipal elections. Because voter turnout is the sole measure of similarity, it was unclear if merging close years—e.g., grouping 2009 and 2010 mayoral elections—was a valid approach. Thus, the synthetic control method was only used for presidential and state elections. State elections are elections that elect state officials during non-presidential election years. They occur every four years and, in this dataset, start in 2002. Presidential elections also happen every four years and begin in 2000.
Ramsey County: Presidential Elections

In the dataset, there are four years where presidential elections occurred: 2004, 2008, 2012, and 2016. As Ramsey County began ranked-choice voting in 2011, two of the year fall into the pre-treatment period and two into the post.

Figure 4 displays the synthetic control output for Ramsey County presidential elections. Similar to Hennepin County, the voter turnout in Ramsey County is lower than the synthetic control in all years. In the pre-treatment period, the synthetic control county’s voter turnout, on average, is 8.5% higher than Ramsey County. Then in 2012, the first RCV presidential election, voter turnout in Ramsey County increased to over 80%, and the synthetic control voter turnout was around 82.5%. Thus, the difference between the two decreased, showing a positive effect. Then, in 2016, the difference between the two counties increased, and voter turnout in Ramsey County decreased. Yet, the difference remained smaller than pre-treatment differences, and the voter turnout remained above the pre-treatment voter turnout. Thus, in 2016, RCV still had a positive effect but a more negligible positive effect compared to 2012. These results support the hypothesis that voter turnout will ultimately increase due to RCV elections but contradict the learning curve hypothesis. Just as for Hennepin County, one explanation for this discrepancy is the difference in candidates in 2012 and 2016, voter suppression, and the difference in the political climate.
Ramsey County: State Elections

In the dataset, Ramsey County has five years with state elections: 2002, 2006, 2010, 2014, and 2018. As Ramsey County introduced RCV in 2011, three election years fall in the pre-treatment period and two in the post.

As with presidential elections, Ramsey County’s voter turnout is lower than the control group’s voter turnout. In 2014, the first ranked-choice voting election in Ramsey County, this difference increased, and overall voter turnout in Ramsey County decreased, displaying a negative effect on both accounts. Then in 2018, both voter turnouts escalated above pre-treatment levels, and the difference decreased to below pre-treatment levels. This trend follows the learning curve hypothesis, as the immediate effect of RCV on voter turnout is negative, but the subsequent impact of RCV on voter turnout is positive.
To further explore the impact of the number of elections since the implementation of RCV on voter turnout, I ran a simple linear regression. Rather than subsetting the data by election type, I instead included all the election types and focussed on the number of RCV elections in the chosen county. Additionally, each unit is a voting district rather than a county with all the election types. Thus, each county has multiple observations each year.

In the regression, the baseline represents the average voter turnout in all plurality elections in Hennepin and Ramsey counted from 2000 until 2011 and 2009, respectively. The coefficient on firstRCVelection is thus the average voter turnout in the firstRCVelection in each district regardless of election type and year. For example, for Hennepin County, this is a mayoral election in 2009, while for Ramsey, this is a municipal election in 2011.
The regression displays that turnout decreases in all RCV elections compared to plurality elections but that the decrease tapers off as the number of RCV elections since implementation increases. Specifically, the results demonstrate that the most considerable reduction occurs in the first RCV election, then the second, then the third, and finally the fourth. Therefore, these results indicate there is a learning curve or adaptation curve and that as time progresses, turnout in ranked-choice voting elections increases. Additionally, while all these show a decrease in voter turnout, it fails to account for many differences. These RCV elections occurred after 2009, while the baseline is the average plurality elections from 2000 to 2009 and 2011. Thus, it is unclear if the difference is due to general declines in turnout due to mistrust of the system, lack of agreement with any candidate, voter suppression, and more, or a direct result of RCV. These results are highly preliminary, and thus further research is needed for conclusions.

Figure 6: Regression Results

```
Call:
  lm(formula = Turnout ~ relevel(as.factor(treat.year), ref = "none"),
     data = data2)

Residuals:     Min       1Q  Median       3Q      Max
     -0.61237 -0.11112  0.05355  0.17469  0.33207

Coefficients:                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                     0.612371   0.002831  216.28  < 2e-16 ***
relevel(as.factor(treat.year), ref = "none")fourth -0.236305   0.023081  -10.24  < 2e-16 ***
relevel(as.factor(treat.year), ref = "none")second -0.341828   0.015576  -21.95  < 2e-16 ***
relevel(as.factor(treat.year), ref = "none")third  -0.295055   0.015329  -19.25  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2209 on 6805 degrees of freedom
(128 observations deleted due to missingness)
Multiple R-squared:  0.1864,  Adjusted R-squared:  0.1859
F-statistic: 389.8 on 4 and 6805 DF,  p-value: < 2.2e-16
```
Conclusion

This analysis demonstrates that the transition to ranked-choice voting from plurality voting increases voter turnout across state and presidential elections. Additionally, the data illustrates that, on average, the change in voter turnout increases with subsequent RCV elections, following the learning curve hypothesis.

In the presidential elections in both Hennepin County and Ramsey County, not only did the difference between turnout in the treated group and the synthetic control decrease, showing a relative increase in voter turnout, but the pure voter turnout also increased. Thus, RCV had a positive effect on voter turnout across the board. This sizable positive increase occurred immediately after the implementation of RCV in 2012 and did not show a delayed effect. A possible explanation for this is that presidential elections tend to have more funding and more publicity, giving room for more information and education about the change in the voting system. In both counties, in 2016, the gap in voter turnout and pure turnout decreased from 2012 rates. But, both the net difference in turnout and pure turnout was above the pre-RCV rates. Thus, the subsequent positive effect was smaller than the immediate effect.

In the first state elections with RCV, the difference between the treated group and the synthetic control increased, and the overall voter turnout in treated counties decreased. Therefore, there was an immediate negative effect on voter turnout. While the negative impact may result from an adjustment period, this election was after the 2012 presidential election, where RCV was used. Therefore a possible explanation is that there was less funding for the state elections, and thus, combined with RCV, voter turnout decreased in both measures. But, in 2018, both voter turnouts increased, and the difference in voter turnout decreased, showing an
overall positive effect. In fact, in Hennepin, voter turnout in 2018 surpassed the synthetic control voter turnout, displaying a large subsequent positive impact.

While expected voter turnout decreases for the first through fourth subsequent RCV elections compared to the average plurality election in the treated counties, the magnitude decreases with each successive election. This result supports the theory that there is a learning curve and that ranked-choice voting takes time to adapt. Yet, it is essential to acknowledge the limitations of the dataset that may be driving these results.

This dataset has many limitations, which may be generating the mentioned results. First, the dataset only includes counties in Minnesota and therefore lacks external validity. Yet, even within Minnesota, only two counties have used RCV elections, and those two counties have the largest racial and ethnic minority populations in the state. Therefore, there may be different drivers of the observed effects that correlate with the racial and ethnic demographic of Ramsey and Hennepin. For example, in 2012, President Obama ran a re-election campaign. Voters are more likely to vote in elections where the candidates represent their racial group. 36% of the voters who voted in 2012 but did not vote in 2016 were Black, displaying that Black voters disproportionately went from voting to not voting (Bump, 2018). Thus, this trend may have impacted the voter turnout in the treated counties.

Powerful data analysis tools were not used because of the small sample of treated counties. Therefore, the treated counties were treated as case studies, and the synthetic control method was used. The synthetic control was only based on previous voter turnout, not any other factors. Therefore, the similarity measure is not robust.

Because of all the limitations of this analysis, there is much room for further research. First, there will be more data with the upcoming midterm elections and the increase of districts
using RCV. Ideally, this increase in data will help satisfy heterogeneity, allowing for analyzing
the interaction between race and ranked-choice voting on turnout.

Additionally, the concern regarding the difficulty of ranking preferences extends from
ranked-choice voting and is relevant in ranking any choices. Therefore, future research about
setting up and advertising ranked-choice voting elections is essential to increase accessibility.
Overall, humans struggle to organize preferences accurately. Analysis of different restaurant
choices revealed that asking individuals to rank restaurants and deriving a ranking from
scenarios requiring binary options leads to different hierarchies (Schibrowsky & Peltier, 1995).
In other words, what we prefer and what we would choose are not equivalent. The researchers
also found that requiring justification for preferences changes preferences (Schibrowsky &
Peltier, 1995). Thus, this research hints that what we initially believed in our preferential order
might not be our actual ranking.

Lastly, the accuracy of our rankings three and below lose accuracy, indicating that only
the top two rankings are telling of preferential attitudes (Ben-Akiva et al., 1992, p. 163). Thus,
there is a disconnect between internal attitudes and stated hierarchies. Therefore, even if
ranked-choice voting is a preferential voting system to increase turnout, we may still be bad at
ranking. Thus, what can we do to improve the ability to rank candidates to make RCV
successful? The finding that RCV motivates voter turnout provides grounds to encourage
research on making the ranking process more accessible, as it proves RCV can be a sustainable
voting system.
Bibliography

References

https://ballotpedia.org/Ranked-choice_voting_(RCV)

https://doi.org/10.1016/0148-2963(92)90058-j

https://doi.org/10.2307/2647733

https://www.washingtonpost.com/news/politics/wp/2018/03/12/4-4-million-2012-obama-voters-stayed-home-in-2016-more-than-a-third-of-them-black/

https://doi.org/10.1017/S1049096521001232


https://doi.org/10.17645/pag.v9i2.3913


https://www.brookings.edu/articles/unequal-opportunity-race-and-education/


# Appendix A: R Code for Scatterplot & Linear Modeling

```r
setwd("~/Desktop/Thesis")

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──
##    ✓ ggplot2 3.3.3 ✓ purrr 0.3.4
##    ✓ tibble 3.0.6 ✓ dplyr 1.0.5
##    ✓ tidyr 1.1.2 ✓ stringr 1.4.0
##    ✓ readr 1.4.0 ✓ forcats 0.5.1

evoting_data <- read_csv("votingdata.csv")
demog_data <- read_csv("countydemographics.csv")

# Merging data sets
joined_data <- left_join(voting_data,
                          demog_data,
                          by = c("County.Name" = "county", "Year" = "Year"))

# Creating one variable for identifying as Black or AA
joined_data$'Proportion of Black and/or African American Residents' <-
              joined_data$blackAA_alone_HL + joined_data$blackAA_alone_noHL

joined_data$'Proportion of White Residents' <- joined_data$white_alone_noHL

# Converting 'treatment' to a character object
joined_data$treatment <- as.character(joined_data$Treated)
```
# Plotting Black Prop vs. white Prop

```r
ggplot(
joined_data,
aes(`Proportion of White Residents`,
    `Proportion of Black and/or African American Residents`,
    color = treatment)) +
geom_point(size = 1.5) +
xlab("Proportion of Population that Identifies as White") +
ylab("Proportion of Population that Identifies as Black and/or African American")
```

![Graph of Proportion of White Residents vs. Proportion of Black and/or African American Residents](image)

# Limiting the data set to just treated counties

treated_counties <- joined_data %>%
    subset(County.Name == "HENNEPIN" | County.Name == "RAMSEY")

# Running a linear regression

```r
lmtest <- lm(turnout ~ third + fourth + second, data = treated_counties)

summary(lmtest)
```
## Call:
`lm(formula = turnout ~ third + fourth + second, data = treated_counties)`

## Residuals:
```
  Min 1Q Median 3Q Max
-0.4002 -0.1812  0.0714  0.1650  0.2898
```

## Coefficients:
```
              Estimate Std. Error   t value Pr(>|t|)  
(Intercept)   0.57200   0.02965    19.292   <2e-16 *** 
third         -0.23894   0.15967    -1.497   0.1400   
fourth        -0.15419   0.22385    -0.689   0.4937   
second        -0.27680   0.15967    -1.734   0.0884 .
```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Residual standard error: 0.2219 on 57 degrees of freedom

## (6 observations deleted due to missingness)

## Multiple R-squared:  0.08723,  Adjusted R-squared:  0.03919

## F-statistic: 1.816 on 3 and 57 DF,  p-value: 0.1546
Appendix B: R Code for Synthetic Control Method

This code produces output for the synthetic control analysis done for the Ramsey state elections. The equivalent code is used for the other synthetic control analyses with small changes noted in each line where applicable.

```r
setwd("~/Desktop/Thesis")

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──
## ✓ ggplot2 3.3.3 ✓ purrr 0.3.4
## ✓ tibble 3.0.6 ✓ dplyr 1.0.5
## ✓ tidyr 1.1.2 ✓ stringr 1.4.0
## ✓ readr 1.4.0 ✓ forcats 0.5.1

df_synthetic <- read_csv("MergedDecisionData.csv",
                         col_types = cols(.default = col_character()))

# Clean up data

df_synthetic <- df_synthetic %>%
                   drop_na("County.Name")
                   group_by(Year, County.Name, State.General.Election, Presidential.Election,
                              Mid.Term, Municipal.Election, Mayoral.TC, Treated, STV, treat.year)

df_synthetic <- df_synthetic[-c(1:3, 5, 7:9, 13:14, 20, 25:50)]

# Clean up data by grouping at county level

df_synthetic <- df_synthetic[c(1:14)]

df_synthetic$weighted_turnout <- as.double(df_synthetic$Turnout) *
                                as.integer(df_synthetic$Total.Votes)
```
df_synthetic <- df_synthetic %>%
  group_by(
    Year, County.Name, State.General.Election, Presidential.Election,
    Mid.Term, Municipal.Election, Mayoral.TC, Treated, STV, treat.year) %>%
  summarise(
    Total.Votes = sum(as.integer(Total.Votes)),
    weighted_turnout = sum(weighted_turnout),
    Registered.at.7am = sum(as.integer(Registered.at.7am)),
    Registered.On.Day = sum(as.integer(Registered.On.Day)),
    na.rm = TRUE )

df_synthetic$turnout <- df_synthetic$weighted_turnout / df_synthetic$Total.Votes

# Remove treated non-Ramsey county data
df_synthetic <- df_synthetic %>%
  # (change for each county/election type)
  filter(County.Name != "HENNEPIN", rm.na = TRUE)

# Ensure all Ramsey districts are treated
df_synthetic <- df_synthetic %>%
  # (change for each county/election type)
  filter((County.Name == "RAMSEY" &amp; &amp; Treated == 1) | County.Name != "RAMSEY")

table(
  paste(
    df_synthetic$State.General.Election,
    df_synthetic$Presidential.Election,
    df_synthetic$Mid.Term,
    df_synthetic$Municipal.Election,
# Process the data

df_synthetic <- df_synthetic %>%
  ungroup() %>%
  mutate(
    Year = as.numeric(Year),
    election_type = case_when(
      Mayoral.TC == 1 ~ "mayoral",
      Mayoral.TC == 0 & Municipal.Election == 1 ~ "other municipal",
      Presidential.Election == 1 ~ "presidential",
      Presidential.Election == 0 & State.General.Election == 1 ~ "other state election"),
    county_election = paste(County.Name, election_type)) %>%
  drop_na() %>%
  select(County.Name, county_election, Year, Total.Votes, weighted_turnout, turnout, Registered.at.7am, Registered.On.Day) %>%

# (change for each county/election type)
filter(County.Name != "RAMSEY" | county_election == "RAMSEY other state election")

# Here is the timespan for the treatment unit

timespan <- df_synthetic %>%
  filter(county_election == "RAMSEY other state election") %>%
  pull(Year)

# We need to have a set of control counties that have all of those time periods present

df_synthetic_restricted <- df_synthetic %>%
  filter(Year %in% timespan) %>% # only take observations from matching years
  complete(Year, county_election) # fill out the missing time county_election combinations

# Find county election combinations that do not have a full timespan
incomplete_info_units <- df_synthetic_restricted %>%
  filter(is.na(turnout)) %>%
  pull(county_election)

# Remove these problematic cases
df_synthetic_restricted <- df_synthetic_restricted %>%
  filter(!(county_election %in% incomplete_info_units))

# Run the synthetic control method
library(tidysynth)
synthetic_step1 <- df_synthetic_restricted %>%
  synthetic_control(outcome = turnout,
                    unit = county_election,
                    time = Year,
                    # (change for each county/election type)
                    i_unit = "RAMSEY other state election",
                    # (change for each county/election type)
                    i_time = 2014, # first period that is treated
                    generate_placebos = T)

synthetic_step2 <- synthetic_step1 %>%
  # (change for each county/election type)
  generate_predictor(time_window = c(2002, 2006, 2010), # years in the pre-treatment period
                     turnout = mean(turnout, na.rm = T),
                     Total.Votes = mean(Total.Votes, na.rm = T),
\[ \text{Registered.at.7am} = \text{mean(Registered.at.7am, na.rm = T)}, \]
\[ \text{Registered.On.Day} = \text{mean(Registered.On.Day, na.rm = T))} \]

synthetic_step3 <- synthetic_step2 %>%
  # (change for each county/election type)
  generate_weights(optimization_window = c(2014, 2018), # years in the post-treatment period
                   margin_ipop = .02, sigf_ipop = 2, bound_ipop = 6)

synthetic_step4 <- synthetic_step3 %>%
  generate_control()

synthetic_step4 %>% plot_trends()