From Interpolation to Imputation: Ballot Completion in a Ranked Choice Election

---

A Thesis

Presented to

The Division of Mathematics and Natural Sciences

Reed College

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Arts

Jay Lee

May 2019

Approved for the Division
(Mathematics - Statistics)

_____

Heather Kitada Smalley

Acknowledgements

To Heather, for consistently keeping me on track to actually graduate and being a wonderful thesis advisor;

To Andrew, Paul G., Kristin, and Paul M., for giving me a series of good jobs and unofficially advising me over the last few years;

To the Reed Collegium Musicum and the Anna Men, for giving me a few hours a week to take my mind off of this and make some music;

To Mia, Monika, Jade, and Scottie's Gang, for each giving me a semblance of a social life and getting me off campus;

To my family, for not getting too mad at me for not calling as much as I should;

Thank you.

List of Abbreviations

ACS: American Community Survey
BIC: Bayesian Information Criterion
GIS: Geographic information system
IIA: Independence of irrelevant alternatives
IPW: Inverse probability weighting
IRV: Instant-runoff voting
MAR: Missing at random
MCAR: Missing completely at random
MI: Multiple imputation
MLE: Maximum likehood estimation
MNAR / NI: Missing not at random / Non-ignorable
NP: Non-deterministic polynomial time
PAC: Political action committee
Q-Q: Quantile-quantile
RCV: Ranked choice voting
RHD: Random hot deck
SE: Standard error
SF: San Francisco
STV: Single transferable vote
VAP: Voting-age population
VEP: Voting-eligible population

Table of Contents

List of Tables

List of Figures

Abstract

Ranked choice voting (RCV) is an alternative voting system where voters rank multiple candidates (instead of simply selecting their highest preference), then these multiple preferences are taken into account during the vote tabulation. One of the arguments against implementing RCV is that it is harder for voters to participate in. Two of the reasons for this are the more complicated ballot design and the extra effort that goes into forming an ordered preference of candidates. To evaluate this claim, we examine rates of ballot errors and undervoting (ranking fewer than the allowed number of candidates) in some American elections conducted with RCV. Results are somewhat inconclusive, but indicate that the variables which are significant in predicting RCV ballot completion are similar to those which are significant in predicting voting rates in general.

We further investigate the impact of overvotes and undervotes on an RCV election. By simulating vote choices for those not present in the ballot data (*imputing* the missing data), we obtain a number of simulated datasets. By re-tabulating the election results with this new data, we create a set of simulated elections where overvoting and undervoting never occur. These are used as a comparison against the true election to see what impact these undervotes and overvotes have on the electoral outcome. Methods used include hot deck imputation and multinomial logistic regression. Our results here show little effect on who actually gets elected, but indicate that the margin of victory might have been smaller without the presence of overvotes and undervotes. This result shows promise for further research on the topic, particularly using different methods and considering different elections.

Introduction

This thesis investigates issues in ranked choice voting (RCV), using a 2018 mayoral election in San Francisco as our observation. It is split into two major parts, each attempting to answer a separate research question. The motivation for these research questions comes out of previous work on RCV with Matthew Yancheff and Mia Leung, advised by Professors Andrew Bray (statistics) and Paul Gronke (political science) at Reed College. Multiple times while presenting this research into RCV, we were asked what the effect would be "if everybody voted". To these now-anonymous[1] questioners, I am grateful.

This thesis was written using the `thesisdown` R Markdown template for Reed College (Ismay & Solomon, n.d.). The computation was performed in the R programming language (R Core Team, 2018), with significant use of the following packages:

- `rcv`, for tabulating the results of our ranked choice elections (Lee & Yancheff, 2019).

- `sf`, for spatial analysis and areal interpolation (Pebesma et al., n.d.).

- `simputation`, for multiple imputation of our ballot data (Vanderloo, 2017).

- The `tidyverse`, for data wrangling and visualization ("Tidyverse," n.d.).

All of the data used, R code, and source documents can be found at my GitHub page: https://github.com/jayleetx/thesis.

Part 1

Chapters 1-3 investigate the demographic trends associated with the phenomena of overvoting and undervoting in RCV. Chapter 1 reviews the literature around RCV in general, its implementation in San Francisco, and comparisons between RCV and the more common *plurality* voting system. Chapter 2 describes the methods used to conduct this investigation, particularly the data transformations that were required to obtain usable data. Chapter 3 details the results of the investigation, validations of the methods used, and some implications for further research.

---

[1]Two years ago, I did not think to write down their names for this purpose.

Part 2

Chapters 4-6 use missing data imputation methods to simulate a situation where there were no ballot errors, particularly overvoting and undervoting. Chapter 4 covers the theory around missing data and a comparison of multiple possible methods to deal with this phenomenon. Chapter 5 explains the imputation methods used in this study, as well as problems encountered along the way which had to be dealt with. Chapter 6 presents the conclusions from the analysis.

Chapter 1

RCV Background

1.1 What is ranked choice voting?

Ranked choice voting (RCV), also known as the *alternative vote* or *instant-runoff voting* (IRV) is an alternative voting method to the *first-past-the-post* or *plurality* election system more familiar to American voters, where the candidate with the most votes wins. Each voter, instead of choosing their highest preference among a set of candidates for an office, ranks some subset of the candidates in order of preference. This system (or a close variant) is used in Australia, Ireland, and some American jurisdictions: the state of Maine; San Francisco, CA; Minneapolis, MN; and Cambridge, MA; among others.

The single-winner RCV tabulation algorithm generally proceeds as follows[1]:

1. Identify each voter's most preferred candidate that has not yet been eliminated. Count up these preferences by candidate.

2. If one candidate has a majority (50% + 1) of the unexhausted votes, they are declared the winner and counting stops.

3. The candidate with the lowest number of votes is eliminated.

4. The ballots counted for that candidate are each transferred to the voter's next choice if one exists, or if one does not exist the ballot is "exhausted" and removed from counting for further rounds.

5. Return to 1.

Most jurisdictions that use RCV have slightly different rules for edge cases and ballot errors, but this algorithm is what distinguishes RCV from other ranked voting systems[2]. A close variant of RCV is the single transferable vote (STV) method[3], which can be used to elect multiple candidates, e.g. for a school board, instead of just one. In

---

[1]Modified from Arrow, Sen, & Suzumura (2002).

[2]E.g. Borda ("Borda Count," n.d.), Condorcet ("Condorcet Method," 2008), Contingent ("Electoral Systems," 2013), etc.

[3]More accurately, RCV is the single-winner implementation of the STV algorithm.

the US, this is used in Cambridge, MA and Minneapolis, MN to elect multi-member offices (Douglas, 2013).

1.2   Frequently Used Terms

Below are some definitions for frequently used terms later on. These are not all ubiquitous (for example, "undervote" has another meaning in other voting research), but we define them here for clarity.

- *Overvote*: when a voter ranks multiple candidates in the same slot. This slot is typically thrown out entirely in counting, because it's often not possible to determine which candidate was preferred.

- *Undervote*: when a voter does not rank candidates in all of the slots available to them. This is not a problem in counting, and is explicitly allowed in the laws of most jurisdictions. A plurality election analog would be voting in high-profile races (e.g. presidential), but not down-ticket decisions (e.g. local water board). This is different than other definitions of "undervote", which refer to a voter participating in one election on a ballot but not another one.

- *Skipped vote*: when a voter ranks no candidate at slot x, but ranks a candidate at slot $y > x$. This is typically not a problem in counting, but different jurisdictions have different rules about whether a voter's ballot is exhausted at this point or continues on to their next ranked choice. Plurality voting has no analog to this, because each race only has one "ranking"[4].

- *Duplicated vote*: when a voter ranks the same candidate for distinct slots x and y. This is typically not a problem for counting, and the first ranking for the candidate is used. Similar to a skipped vote, plurality voting has no analog to this. When we fit models to discuss undervoting, this category will be included as well under the heading of "undervoting" because it is also a form of not ranking as many unique candidates as offered.

- *Ballot exhaustion*: as ballot counting progresses, some ballots will become "exhausted" when all the candidates selected are eliminated. Suppose the final count in an election is between candidates B and D, and a voter ranked candidates C-A-E. Their ballot would not be counted in this final round, as they expressed no preference for either candidate B or D. An analogous situation in a plurality election might be voting in the general election but not a runoff, i.e. only having a say in part of the election.

Overvotes, skipped votes, and duplicated votes are really only interpretable as "ballot mistakes": for example, even if a voter truly prefers two candidates equally, the ballot instructions (should) make it clear that ranking them at the same slot is not allowed.

---

[4]First!

1.3 Claims about RCV

There are plenty of arguments for and against implementing RCV in place of plurality in different jurisdictions (see Part 1.5), both mathematical and practical, but in this study we'll focus on evidence surrounding one major argument against it - RCV is harder for voters to participate in than a plurality system. There are two major reasons cited for this:

- The physical design of an RCV ballot is usually more complicated than a plurality ballot, because there has to be a system to encode a more full preference among the candidates than just selecting one candidate

- The process of forming a multi-candidate preference inherently takes more mental effort than just choosing a favorite candidate

The first facet of this argument should be reflected in ballot errors made by voters. Compared to plurality voting, we expect more errors in an RCV ballot just because the ballot is more complicated. There are also more potentials for error in the RCV system generally. The only "errors" in a plurality ballot are incompletely marking a candidate (think incorrect Scantron bubbling, or hanging chads) or overvoting, both of which are potential pitfalls for a ranked choice ballot as well. On top of these, there are the potential errors of duplicated and skipped votes unique to ranked ballots[5].

The second facet should be reflected in incomplete ballots filled out by voters. Given that they understand how to encode their preferences on the ballot, there is still the non-trivial task of forming such a preference. Structurally, some of the factors that should affect this incompleteness are:

- The number of candidates running for a position

- The number of candidates voters can rank

- The number of seats elected in a given race

This first variable is determined for each election, the second for each jurisdiction, and the third a mix of both. For a clear example of these differences, consider a 2016 San Francisco Board of Supervisors race (District 3) versus a 2017 Cambridge City Council race (Table 1.1).

1.4 History of RCV in the US

In the United States, there have been two major periods of RCV implementation in various jurisdictions. Between 1915 and 1950, 24 American cities chose to institute

---

[5]These types of errors are not uniform, and some jurisdictions are more forgiving than others about rules for counting these errors. While it may be apparent that a voter who listed the same candidate 3 times (A-A-A) prefers that candidate, a candidate ranking of A-B-A is harder to extract a clear preference from. Skipped votes are where we see the most variance in jurisdiction counting rules: if a voter marks the ballot A-__-B, skipping the second slot, some jurisdictions will ignore the skip and treat B as the voter's second choice, while others will stop counting after A is eliminated (ignoring their vote for B), and others yet will throw out the ballot entirely.

Table 1.1: Comparison of RCV rules by jurisdiction

	San Francisco 2016	Cambridge 2017
Candidates running	2	27
Candidates rankable	2 (Generally, up to 3)	27 (Generally, all)
Seats elected	1	9

RCV as a form of local election. By 1965, however, all of these except for Cambridge, MA had eliminated the policy change. Then, in the 2000s, there was a resurgence of uptake in a different set of American cities[6], including Minneapolis and a handful in the San Francisco Bay Area. While Cambridge has consistently used the multi-winner (STV) method to elect City Council and School Board seats, the modern resurgence of RCV almost universally deals with single-winner elections. Research argues that RCV appears in jurisdictions where there is strong multi-party support for the reform - the RCV method itself gives individual parties less power in the election process, so powerful single parties usually don't have reason to support it (Santucci, 2017).

1.5   Why, or why not, implement RCV?

1.5.1   Pros

Mathematical benefits

One of the primary mathematical benefits of RCV is that it ensures a majority winner in the final round, not just a plurality winner. In jurisdictions without rules for 50% minimums, a common phenomenon is a candidate winning an election with less than 50% of the vote (a plurality, rather than a majority). The major conceptual issue with this is that more people preferred a candidate other than the one who was elected[7]. RCV requires that a winning candidate receive at least 50% of the votes remaining[8], ensuring that a majority of voters prefer the elected candidate to other candidates. This is particularly important in jurisdictions with strong third-party support and more than two viable candidates, like Maine. Former Governor Paul LePage, a Republican, won his first election in 2010 with 38.1% of the vote, compared to Independent Eliot Cutler's 36.7%[9]. This was one of the reasons cited for Maine's recent move to RCV for statewide elections.

RCV is also less vulnerable to strategic voting than plurality. Strategic voting[10] is a scenario where a voter does not reflect their true preferences on their cast ballot

---

[6]Mostly in the American West: there are 9 cities west of the Mississippi River currently using RCV ("Where is Ranked Choice Voting Used?" n.d.) and only 4 east of it.

[7]The "ideal" for electoral systems is the Condorcet condition: the candidate elected should beat all other candidates in one-on-one contests.

[8]See below for issues with this "remaining" concept.

[9]A margin of about 7,500 votes. Democrat Libby Mitchell received 19%.

[10]Also known as "tactical" or "insincere" voting.

in order to affect the outcome of the election. For example, a third party supporter may cast their vote for one of the two major party candidates because otherwise they feel like it won't count. While not impossible, under RCV it is NP-complete[11] for a given voter to determine if voting strategically (misrepresenting their true preferences) can help elect a more-preferred candidate (Bartholdi & Orlin, 1991). This is true even if you know the full preferences of every other voter, which is often a practical impossibility on its own.

No secondary elections

There are two major types of "secondary elections" used in American voting: primary elections and runoff elections. Primaries are used by political parties to select their nominee for a general election, so the voters of any one party aren't split between different candidates. Runoffs are most often used when no candidate in the general election surpasses 50% of the vote total. Typically the top two candidates from the general election[12] advance to a later runoff. These secondary elections face two main challenges:

- Low turnout. Secondary elections as a whole face low turnout (Ranney, 1972; Wright, 1989). Research shows that people don't actually like voting that much - the more frequently elections are held, the lower turnout will be for all of them generally. Secondary elections increase the number of elections in a period, so this is one possible reason why they generally have low turnout. Further research indicates that holding elections concurrently with a presidential election "increase[s] the likelihood that citizens will vote" (Boyd, 1986). This is seen in off-year Congressional elections, where turnout drops from presidential years. Typically general elections are held concurrently with presidential elections, so secondary elections cannot be held at the same time as a presidential election and they should thus suffer in turnout. This low turnout has consequences for representation in the system. The same research (Ranney, 1972) finds that while primary voters are not ideologically unrepresentative of general election voters, they are both demographically unrepresentative and unrepresentative on some major issues. Traditional knowledge holds that primary voters are more committed partisans than general election voters, leading the eventual candidates in a general election to be polarized away from the "center" of political ideas (Hill, 2007).

- High costs. The higher costs associated with secondary elections are a little more intuitive than turnout issues - it takes money to hold elections. Pollworkers have to be paid, facilities have to be reserved, and candidates have to do more campaigning. A 2011 City Council runoff in Plano, TX cost the city an extra

---

[11]A computer cannot find the solution "quickly", particularly as the number of voters increases. This makes such an event difficult to voluntarily induce.

[12]Or primary election - seven Southern states require primary winners to obtain 50% of the vote to get on the general election ballot, and some other states have a requirement of 40% (Wilson, 2014).

$73,000 (Rush, 2015). A 2012 Alabama runoff for multiple seats cost the state about $3 million (The Associated Press, 2014).

Since RCV eliminates the need for primary and runoff elections while still ensuring majority rule (which is the main reason for these elections), it should avoid the problems of lower turnout and higher costs associated with secondary elections[13]. Summed up, RCV decreases costs and and increases turnout by eliminating these low-turnout elections (Morales, 2018).

Reduces strategic voting - spoiler candidates and third parties

The "spoiler effect" is when a third party candidate draws votes away from the ideologically closest major party candidate, thus contributing to the election of the other major party candidate. The most famous large-scale accusation of this was in the 2000 US presidential election. Green Party candidate Ralph Nader drew about 3% of the national vote, more than the margin of victory for George W. Bush over Al Gore[14]. In the especially consequential state of Florida, Nader took 1.6% of the vote, almost 200 times greater than the margin between the two major party candidates of less than .01 percentage points[15] ("2000 Presidential General Election Results," 2001). Many believed that Nader, generally seen as more liberal than the Democrat Gore, drew votes from the Democratic base that would have helped Gore win the election otherwise. While research into third-party voters casts some doubt on this theory's applicability in 2000[16], public opinion still rests on the idea that Nader cost Gore the presidency[17] (Herron & Lewis, 2007). In fact, a pro-Republican PAC aired campaign ads promoting Nader in Democratic states in an attempt to pull votes from Gore (Meckler, 2000).

One of the reasons for major party voters to support RCV is that it avoids this spoiler issue. Under a plurality system, voters are discouraged from voting for third-party candidates because it could help elect their least-preferred of the two major candidates - "the greater of two evils", so to speak. Under RCV, however, voters can be heard throughout multiple rounds across separate candidates. Third-party voters can thus vote for their most preferred candidate, then still have their vote count for a preferred major-party candidate if their first option is eliminated.

Conversely, one of the reasons for third-party voters to support RCV is that it helps third party candidates get elected. Voters can ignore this aforementioned facet of strategic voting and select their truly preferred candidate. Third-party supporters

---

[13]Or, at least part of these problems.

[14]Admittedly, the margins are less clear-cut than this at the state level, where the margins actually matter for the Electoral College.

[15]Not to point fingers at Nader alone in this case - while he was the most popular third party candidate by far, all 8 official third-party candidates received more votes than the major-candidate margin of only 537 votes.

[16]This doubt, summarized - while Nader's Florida voters potentially would have broken enough for Gore to put him over the top, this was more a factor of the unusually close margin between the two major candidates than anything that Nader aided in particular.

[17]One of the sections of Nader's Wikipedia page is entitled "Spoiler controversy" in regards to this election.

who were worried about the spoiler effect, then, can vote for their true preference of a third party and not inadvertently help a less-preferred major candidate get elected. As people abandon this strategy, third parties will receive more votes from people no longer worried about the spoiler effect, and this could get third party candidates elected.

Disincentivizes negative campaigning

Ranked choice voting should incentivize candidates to avoid negative campaigning. In a plurality election, while a negative campaign might ostracize some of their opponents' supporters (though candidates don't care about voters who are committed to their competitors) Candidate X might improve their position by bringing more swing voters to their side. Under RCV, however, alienating Candidate Y's voters could backfire in the event that Candidate Y is eliminated and these voters decide to support Candidate Z (who didn't insult them) in the next round, leading to Candidate X's defeat. Research supports this theory - a 2016 study showed that voters in cities using RCV are more satisfied with campaigns than in cities who use plurality methods, and consider RCV campaigns to have a less negative tone overall (Donovan, Tolbert, & Gracey, 2016). In San Francisco's first RCV election, there were joint fundraisers between candidates, and one district even saw regular "Candidates Collaborative" public meetings between many candidates to discuss issues affecting the district, where "the setting [was] decidedly congenial" (Murphy, 2004).

An interesting case study of this phenomena is in the 2018 San Francisco mayoral election (Kukura, 2018). There were three frontrunners heading into election day, all incumbent members of the city's Board of Supervisors: London Breed, Jane Kim, and Mark Leno. As polls showed Breed ahead about a month before the election, Kim and Leno held a joint press conference to endorse the other as voters' second choices. By drawing second-choice votes from the other candidate, the remaining candidate hoped to overcome the gap between them and Breed. In the actual election, the standing when it came time to proceed to the final round of counting was 102,767 for Breed, 68,707 for Leno, and 66,043 for Kim. While a significant proportion of Kim's voters transferred to Leno after her elimination, in the final round Breed surpassed Leno by about 2,000 votes[18] (Figure 1.1).

---

[18]Data from ("Ranked Choice Voting Results Table," 2018); Figure 1.1 and analysis with Lee & Yancheff (2019).
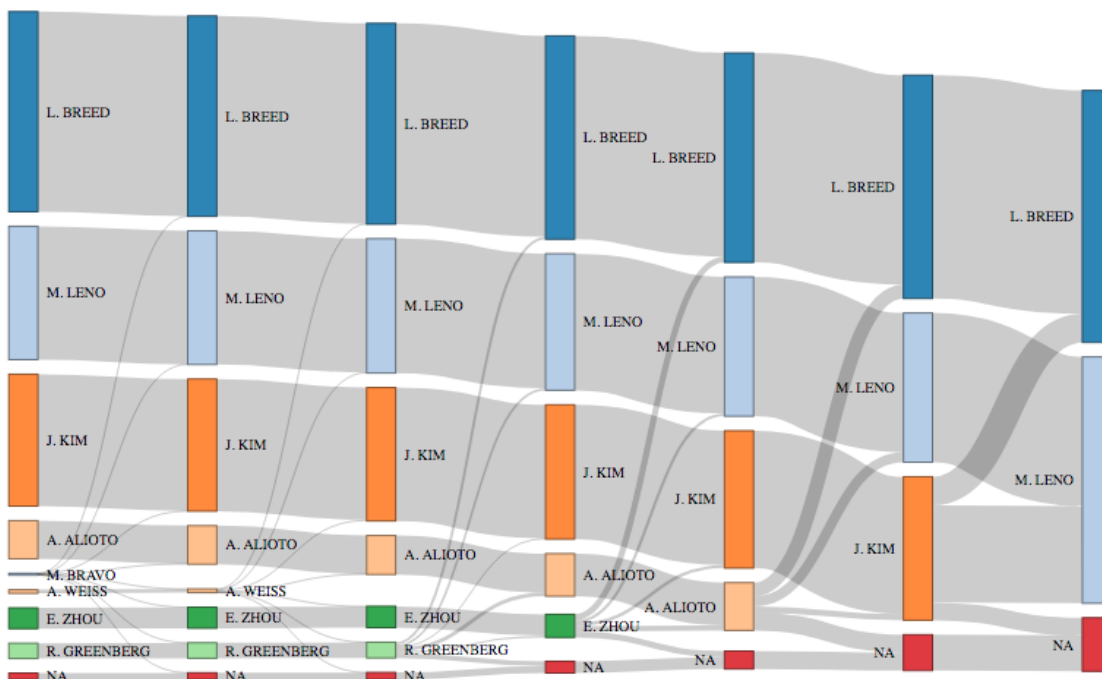
Figure 1.1: Sankey diagram of vote tabulation

Figure 1.1 is a *Sankey diagram* for this election, a type of data visualization used to display flows between states. Each "column" is a round of tabulation, with bar heights proportional to the number of votes counted for each candidate in that round. At the end of each round, the candidate with the lowest amount of votes is eliminated, and their votes get transferred to the voters' second choices. We see from this plot that Breed, Leno, and Kim were the clear front-runners throughout the election, and that a high proportion of Kim's voters transferred to Leno in the final round of tabulation.

Though it's outside the scope of this research to tell if this cross-endorsement was effective[19], there is some evidence in favor of this theory. Leno received almost 70% of the votes previously counted for Kim compared to Breed's 20%, bringing Breed's final margin of victory down to only 1 percentage point. In previous rounds of the election, no single candidate ever received more than 35% of the transferred votes from an eliminated candidate[20], so this is an unusual observation at least.

Minority representation

While much of the American literature on minority representation under voting systems has focused on gerrymandering and single- versus multi-member districts, there is some study of representation under RCV independent of the number of seats up for election. John, Smith, & Zack (2018) find an increase under RCV in the number of

---

[19]Other confounding factors counld exist: maybe Kim and Leno had similar enough positions that this scenario would have happened without the endorsement, maybe this number is only significant because in the final rounds there were only 2 candidates for second choice votes to flow to, etc.

[20]Except for round 2, where all 3 votes for the same write-in candidate transferred to Breed.

racial and ethnic minority candidates who run in the Bay Area, and an increase in the chance that women and minority women win their election. Their theory for this result is that RCV lowers the barrier to entry in an election, making it more feasible for minority candidates to campaign, and that women (particularly women of color) are better suited in general to the less negative campaigning and coalition-building that RCV promotes in candidates. Another study reports that because minority voter turnout in secondary elections is significantly worse than White turnout (moreso than in a general election), the elimination of these secondary elections should increase the relative say of minority voters in elections overall (Callaghan, 2017).

Turnout improvements

Voting system reform advocates claim that these improvements to the voting process will boost turnout by generally increasing public trust in the effectiveness of elections. While there are many different strategies employed by campaigns and advocacy groups to boost turnout, these methods don't have as much effect as changing methods to ranked voting. Phone and direct mail get-out-the-vote campaigns "typically [yield] less than 1 percentage point boosts in voter turnout", and while in-person efforts are better they are less efficient at contacting voters (Morales, 2018).

Another argument is a little less intuitive - "[IRV] boosts turnout via elimination of low-turnout elections". While we typically think about turnout in terms of the number of people voting across similar elections (e.g. change in turnout from 2010 to 2014 in Congressional races), RCV can improve turnout by increasing the number of people who get to participate in an election overall compared to the number of people who participate in typically low-turnout primary and general elections (Morales, 2018).

1.5.2 Cons

Mathematical failings

Arrow (1963) proved a fundamental result about voting theory. Any voting system where voters rank candidates, including RCV, cannot satisfy all of the following conditions[21]:

1. *Unrestricted scope*: a voter must be allowed to individually rank any ordering of candidates that they choose.

2. *Monotonicity*: if a voter changes their vote to rank candidate A higher on their ballot, candidate A cannot perform worse in the election[22].

3. *Independence of irrelevant alternatives* (IIA): if voters change their orderings but all maintain their respective preference between candidate A and candidate B, then A and B must stay in the same relative position in the final outcome. That is, the final outcome of A and B only depends on each voter's preference between A and B, not any of the other candidates.

---

[21]Assuming at least 2 voters and at least 3 candidates.
[22]This condition is important in discouraging strategic voting.

4. *Non-imposition*: every possible ordering of candidates is a potential election result[23].

5. *Non-dictatorship*: there is no single voter whose preferences unilaterally dictate the election results. A dictator would be a voter such that if the dictator preferred A to B, the election would prefer A to B regardless of all other voters' choices.

Lijphart & Grofman (1984) shows that there are also cases when RCV does not produce a Condorcet winner. One ideal for a voting system is that it satisfies the *Condorcet condition*: if there is a candidate that beats every other candidate in a one-on-one contest[24], that candidate should be elected (they are the Condorcet winner). Additionally, RCV fails Arrow's monotonicity condition: there are cases[25] where an increased ranking for a candidate may hurt them overall.

What is a majority, anyway?

Under RCV, the problem of ballot elimination can result in "majorities" that actually aren't as such. Since it's not always possible for voters to rank every candidate[26], there will almost certainly be some number of voters who did not list a ranking for any of the candidates still remaining in the final round of the election. Thus the majority that the winner has collected is only a majority of the unexhausted ballots, which may make it less than 50% of the total counted ballots (Petrangelo, 2013). This does introduce a less expansive form of the spoiler problem - while voters aren't "punished" for putting a less electable candidate for their first choice, they are punished for filling all three available slots with minor candidates. In a 2015 study of four RCV elections, Burnett and Kogan found that all four had enough eliminated ballots to only give the winner an overall plurality, not the majority sought after. This is a problem that requires a substantial fix, as "even individuals who mark three distinct choices often face the prospect of exhaustion, so education alone will not fix the problem" (Burnett & Kogan, 2015).

A research question that might address this issue of majorities is whether the pluralities generated by RCV are generally "bigger" (closer to the ideal of a true majority) in some way than under plurality, and how this is affected by technical rules like the number of candidates that voters are allowed to rank. Some research shows that RCV does perform better than a two-round runoff in this case - as a percentage of the voters in the first round, RCV consistently has more voters counted in the final round compared to top-two runoffs (Richie & Brown, 2017).

---

[23]Conditions 2 and 4 were later combined into the condition of *Pareto efficiency*: if every voter prefers candidate A to candidate B, the final outcome must also prefer A to B.

[24]This does not always happen. Sometimes a cycle can occur: rock beats scissors, scissors beats paper, paper beats rock.

[25]Theoretical, if not necessarily likely in a real election. As mentioned earlier, it is NP-complete to determine how to vote strategically in a given RCV election.

[26]This may be disallowed, as SF currently only allows 3 rankings (due to technical limitations), or it may just be infeasible - Cambridge often has 20+ candidates on the ballot.

Legal challenges

In addition to the political hurdle of passing RCV legislation[27], there are legal challenges as well. In the aftermath of Maine's recent ballot initiative instituting RCV, legal questions were brought to the state Supreme Court by the State Senate, due to conflict between the initiated bill's language about a "majority" versus the state constitution's language requiring a "plurality". There was a ruling from the court that resulted in an amended law instituting RCV for only primary elections, but a "People's Veto" later occured through initiative that removed the law, followed by a later suit by the state Republican Party to block RCV. Maine has so far used RCV in 2018 for both the primary and general elections to elect certain offices ("Ranked Choice Voting in Maine," 2019).

This anecdote examplifies the issue - while public support may be behind RCV in jurisdictions that enact it, legal challenges abound. A (failed) legal challenge was brought against RCV in San Francisco by a defeated candidate in 2011 (Hawkins, 2011). Another challenge (also failed) was brought by Bruce Poliquin against Maine after his loss to Jared Golden in their 2018 Senate race (Mistler, 2018). In Pierce County, WA (home of Tacoma), RCV was repealed by voters after the US Supreme Court court sided with a challenge to reinstate Washington's top two primary-runoff system. With the re-implementation of the top two primary, there was no need to have an RCV election (between two candidates) and the measure was repealed (Eberhard, 2017).

More complicated tabulation

By its nature, the tabulation of RCV elections is more complicated than counting results in plurality races. There are multiple rounds, so counting takes longer. Typically ballot counts are done at the precinct or county level, and then numbers are sent to a state elections office for full tabulation and verification of the results. Under RCV, however, the process typically requires all ballots to be sent (physically or electronically) to the state office for the multiple rounds of counting, further increasing the time needed to count. Additionally, if the jurisdiction is unable to obtain software to count ballots, hand counting of the ballots is necessary. This increased time in counting is an issue because "the elapsed time between Election Day and providing some results is one of the critical factors to maintaining voter engagement and trust in the process" ("Comments on Proposed Rules for RCV," n.d.).

In jurisdictions with some form of electronic voting, the cost of transitioning to RCV can be a disincentive as well. Not all jurisdictions have hardware capable of conducting an RCV election, and in those that do, the vendor[28] may not currently have software that can tabulate the results of the election. Getting around these problems is significant, as it requires either a technology upgrade[29] or a switch to

---

[27]Neither major party is particularly interested in pushing the issue, and Republicans in particular are often opposed to it (Woodard, 2018).

[28]Typically jurisdictions contract to election system vendors.

[29]This is not a problem unique to RCV - election technology across the country is aging and in need of replacement ("Funding Elections Technology," 2019).

some combination of paper ballots and hand-counting (Morales & Eberhard, 2017).

Less intuitive rules

The plurality idea, while clearly not the only voting rule out there, is one of the most simple. One of the problems with RCV is that it doesn't always satisfy this idea simply - RCV sometimes disagrees with the plurality method on choosing a winner. In the Maine Senate race between Bruce Poliquin and Jared Golden, while Poliquin was on top at the end of the first round of counting, neither had 50% and Golden took the lead (and the election) after other candidates were eliminated and their votes transferred to voters' second choices (Miller & Thistle, 2018). The more complicated process and this seemingly unfair (at first glance) rejection of plurality means that voters in general have a harder time understanding and trusting the results of an RCV election. This complication leads to an increase in overvotes (Kimball & Anthony, 2016), which causes more ballots to be fully or partially rejected in the counting process. RCV also sees increased costs in voter education, particularly for the first election conducted - In Minneapolis' first RCV election, 30% of the estimated $365,000 additional cost[30] of RCV was spent on voter education (Kimball, 2010). Neely & Cook (2008) identify three types of "voter errors" - fatigue (intentionally opting out, or accidentally skipping a portion of the ballot), confusion, and lack of information to form a preference. All of these lead to overvotes and undervotes, and are ways in which voters could be aided by a more intuitive process.

Lack of true adoption by voters

While RCV provides the opportunity for voters to rank multiple candidates, not all voters will do so[31]. This is entirely reasonable - forming a complex preference between multiple candidates is naturally more difficult than selecting a favorite from the same set. This phenomenon of undervoting, however, means that some ballots become exhausted in the course of tabulating the election and some voters don't have a say in the final round of the election. If a voter only lists their first choice and skips the second and third choices, they potentially lose out on two extra rounds of tabulating where their vote still counts towards the decision.

1.6   Research into San Francisco

The San Francisco Bay Area in particular has been an RCV hotspot in the modern American resurgence, so some good literature exists specifically studying the impact of RCV there.

---

[30]About a third of this total was one-time costs for implementing the new system.

[31]Which is fully legal in US jurisdictions - nobody is forced to vote.

1.6.1 Results supporting RCV

San Francisco appears to exhibit many of the positive characteristics purported by RCV proponents. The city is facing lower costs, experiencing more minority candidates and officeholders, and high turnout (C. Hill et al., 2018). S. Hill & Hernandez (2018) note that the exhausted ballots in a recent election had minimal impact on the results. Jerdonek (2006) finds that a 2005 election had notably high turnout, particularly in certain poor and racially diverse neighborhoods. Henry (n.d.) finds a decrease in runoff elections as a result of RCV, and no significant impact on voter turnout.

1.6.2 Mixed results

Dennis (2004) finds higher rates of overvotes in SF's first RCV election (November 2004) than were expected, somewhat positively correlated with the number of candidates on the ballot in each district. Almost a third of the ballots contained an undervote. Looking at racial and ethnic groups, while Hispanic voters were more likely to find the RCV method easy to use than Asian voters, the former were also more likely to have completed their ballot incorrectly. One possible explanation of this phenomenon is that advocacy groups "succeeded in heightening awareness about the new voting system amongst the Asian community" prior to the election, increasing their wariness of it as well.

Neely & Cook (2008) extend this basic analysis of ballot errors to explore some theories related to racial and ethnic divisions (and others). There was a significant decrease in undervoting among racial and ethnic minority groups across 4 years of elections, and a significant increase in undervotes among women as well. Districts with more candidates[32] and more campaign spending were less likely to have undervotes. The likelihood of ranking all 3 available candidates was affected by racial and ethnic categories, but not as much as it was affected by factors such as prior exposure to RCV and available election information (campaign spending, number of candidates, etc.).

There is evidence that minority communities see higher rates of overvotes, particularly in precincts that are heavily Black, Latino, foreign-born, elderly, or low-income. This pattern appears to be no more severe than under plurality voting, however, so the problem there is likely not with RCV in particular (Neely & McDaniel, 2015).

1.6.3 Results criticizing RCV

McDaniel (2016) reports a worsening of the racial voting gap in RCV compared to previous plurality elections in SF. Additionally, he finds lower turnout more generally and an increase in ballot-marking errors that cause a ballot to be disqualified. Cook (2011) notes a vitriolic final month of campaigning in a 2011 election, contrary to the good feelings reported by Donovan et al. (2016) and others. There was low voter turnout as well, and combined with the RCV process this made it hard to find an "electoral mandate" with which to rightfully govern. Mcdaniel (2016) argues that

---

[32]These also had more overvotes, in keeping with Dennis' findings.

RCV "obscures racial group interests for voters", and reports a decrease in turnout under RCV for Black and White voters. In 2018, he additionally finds higher levels of racially polarized voting between White and Asian voters (McDaniel, 2018).

Chapter 2

Data structure and demographic analysis methods

An ideal data format for this study would include both individual-level voting behavior and individual-level demographic information. Such data would lend itself well to a logistic regression, where the demographics of individual voters are predictors and undervoting (or overvoting) is the "success" response. This is an infeasible goal, however, as official voting records are completely anonymized (as they should be). As such, we have to aggregate upwards to obtain our data.

2.1 Data Structure and Source

We use two major types of data for this study: election records, from the San Francisco Department of Elections, and demographic information, from the US Census.

2.1.1 Election records

From the San Francisco Department of Elections, we have a cast ballot record of the June 2018 mayoral election. This is presented by the city as two text files:

- The ballot image ("Ranked-Choice Voting," 2018a), a 45-character fixed-width file with fields corresponding to election, candidate, unique voter number (anonymized and disconnected from any voter registration ID), ranking, precinct, and other information. Each of these is encoded numerically, so each line of the file appears as a 45-digit number (see Figure 2.1).

- The lookup table ("Ranked-Choice Voting," 2018b), an 83-character fixed-width file that defines the encodings used to create the numerical values in the ballot image.

We also obtain a precinct boundary shapefile from the Department of Elections ("Maps," n.d.).

```
000000900000066030000000100200004060010000128000
000000900000066030000001002000040600200000001
000000900000066030000001002000040600300000001
000000900000066040000001002000040600100012300
000000900000066040000001002000040600200012500
000000900000066040000001002000040600300012100
000000900000066050000001002000040600100012500
000000900000066050000001002000040600200012400
000000900000066050000001002000040600300012600
000000900000066060000001002000040600100012300
000000900000066060000001002000040600200012500
000000900000066060000001002000040600300012800
000000900000066070000001002000040600100012600
000000900000066070000001002000040600200012900
000000900000066070000001002000040600300013000
000000900000066080000001002000040600100012300
000000900000066080000001002000040600200000001
000000900000066080000001002000040600300000001
```

Figure 2.1: Original ballot image data

In the ballot file, each voter is spread across three rows of data (one for each ranking). We use the `clean_ballot()` function from our `rcv` R package (Lee & Yancheff, 2019) to read in the data and combine the ballot image with the lookup table. From this data we obtain the full ranked preference set of candidates (up to 3) from each voter, displayed in Figure 2.2.

	pref_voter_id	contest	precinct	1	2	3
1	000006095	Board of Supervisors, District 11	Pct 1107	KIM ALVARENGA	AHSHA SAFAI	MAGDALENA DE GUZMAN
2	000006096	Board of Supervisors, District 11	Pct 1107	AHSHA SAFAI	BERTA HERNANDEZ	FRANCISCO HERRERA
3	000006097	Board of Supervisors, District 11	Pct 1107	AHSHA SAFAI	NA	NA
4	000006098	Board of Supervisors, District 11	Pct 1107	AHSHA SAFAI	NA	NA
5	000006099	Board of Supervisors, District 11	Pct 1107	BERTA HERNANDEZ	BERTA HERNANDEZ	BERTA HERNANDEZ
6	000006100	Board of Supervisors, District 11	Pct 1107	AHSHA SAFAI	NA	NA
7	000006101	Board of Supervisors, District 11	Pct 1107	AHSHA SAFAI	KIM ALVARENGA	NA
8	000006102	Board of Supervisors, District 11	Pct 1107	AHSHA SAFAI	MAGDALENA DE GUZMAN	BERTA HERNANDEZ
9	000006103	Board of Supervisors, District 11	Pct 1107	FRANCISCO HERRERA	FRANCISCO HERRERA	KIM ALVARENGA
10	000006104	Board of Supervisors, District 11	Pct 1107	KIM ALVARENGA	KIM ALVARENGA	KIM ALVARENGA
11	000006105	Board of Supervisors, District 11	Pct 1107	AHSHA SAFAI	KIM ALVARENGA	MAGDALENA DE GUZMAN
12	000006106	Board of Supervisors, District 11	Pct 1107	KIM ALVARENGA	KIM ALVARENGA	AHSHA SAFAI
13	000006107	Board of Supervisors, District 11	Pct 1107	AHSHA SAFAI	BERTA HERNANDEZ	FRANCISCO HERRERA
14	000006108	Board of Supervisors, District 11	Pct 1107	MAGDALENA DE GUZMAN	FRANCISCO HERRERA	AHSHA SAFAI
15	000006109	Board of Supervisors, District 11	Pct 1107	KIM ALVARENGA	NA	NA
16	000006110	Board of Supervisors, District 11	Pct 1107	KIM ALVARENGA	BERTA HERNANDEZ	FRANCISCO HERRERA
17	000006111	Board of Supervisors, District 11	Pct 1107	FRANCISCO HERRERA	BERTA HERNANDEZ	MAGDALENA DE GUZMAN
18	000006112	Board of Supervisors, District 11	Pct 1107	KIM ALVARENGA	FRANCISCO HERRERA	AHSHA SAFAI

Figure 2.2: Cleaned RCV ballot

Since the voter-level data is fully anonymized, we have no demographic information at the individual level. For any given voter, because the most identifying piece of information we have about them is their precinct, it is impossible to determine their gender or ethnicity. To gain insight into these demographic trends, we instead

aggregate up to the precinct level.

At the precinct level, we can now only study the rates of these ballot phenomena (as opposed to an individual's response). For example, in Precinct 1101, we see that 23.7% of voters undervote. Rather than building a classification model (error / no error), we can now instead build regression models with a numerical dependent variable - the rate of ballot error. Moving up this level of abstraction does remove some granularity from the model (inference and prediction at the precinct level is less specific than at the individual level), but this is the least we can do that still allows us to to access demographic information.

2.1.2 Demographic information

Now that we have precinct-level ballot completion rates, we need to obtain precinct-level demographic information in order to build a model. Our demographic data comes from the 2012-2016 American Community Surveys (ACS) as part of the 2018 Census Planning Database (US Census Bureau, 2018). This data was chosen[1] because it contained demographic variables which we thought would be informative to our question: age, race and ethnicity, education, and poverty levels.

One consideration with this is that the voting population is not always representative of the general population. It might be more informative to obtain demographics about the voting population specifically, rather than the entire population of each precinct. However, the only information directly available about voters is their age (from the county voter registration file), and we consider the discontinuity introduced by using two different data sets for demographic information to be more of an issue than using a less accurate (but universal) Census data set.

We also obtain a block group boundary shapefile from the Census Bureau ("2010 TIGER/Line Shapefiles," n.d.).

2.2 Data processing methods

2.2.1 Areal Interpolation

We now encounter a problem. Since census regions (block groups, tracts, etc.) are set by the federal government, and election precincts are set by San Francisco County[2], the regions don't line up cleanly[3] (see Figure 2.3). Given this mismatch, how do we obtain demographic estimates for our precincts?

Stated more generally: if we have a division of a geographic region and some set of properties on the divisions, how do we estimate measures of these properties for other possible divisions?

---

[1]As opposed to the 2010 Census, which is also included in the Planning Database. While the Census has "zero error" since it's not a survey, it is also more out of date this late in the decade and didn't contain the same useful variables we were interested in.

[2]The city and county government are unified in this case, because the county comprises entirely of the city of San Francisco.

[3]There's no inherent reason that they should, it's just unfortunate for this study.

Precincts   Block Groups

Figure 2.3: Boundary mismatch between data sources

One field of GIS (geographic information system) research that looks into this is called *areal interpolation*. The goal of areal interpolation is to take a variable distributed over a set of "source zones" and estimate its distribution over a set of "target zones" (Schroeder, 2007). These two zoning systems must overlap at least partially to get any estimate for the target zones (e.g. information about Oregon doesn't directly tell us anything about information in Washington, under any zoning system), but are incompatible in some way[4]. When compatibility is not an option, areal interpolation can help get parameter estimates for the target zones.

Some common types of areal interpolation are:

- Areal weighting, using the assumption that populations are distributed uniformly on a region.

- Modified areal weighting, which creates a continuous map from region to region to better reflect changes in population density.

- Target density weighting, which uses extra information about the target zones (typically population density) to increase accuracy (Schroeder, 2007).

- Dasymetric mapping, which uses auxiliary data (also called *ancillary data*) to produce a continuous estimate of population density (Sleeter and Gould, 2008). Examples of ancillary data include land cover information, parcel classifications, and street locations.

---

[4]A set of "compatible" zoning systems would be something like the US Census' hierarchical systems: multiple blocks are combined to make a block group, multiple block groups are combined to make a census tract, etc. Since the areas overlap neatly, directly adding together certain numbers (like population) from block groups can get a very accurate estimate for the census tract.

In this work we will use areal weighting for our interpolation. The other methods above, while typically more accurate, are infeasible under the constraints of the research[5]. The assumption of uniformity is not necessarily correct, but in an urban area such as San Francisco it is more accurate than an area like the entire state of California, with a large urban/rural spread.

Areal weighting

Areal weighting first makes the assumption that populations are distributed uniformly over a space. If Region X has 100 people and we split X in half spatially, then we assume that 50 people are in each half. This also applies to sub-populations: if Region X has 20 non-White Hispanic people, then we assume each half has 10 non-White Hispanic people. This type of data, counts that can be divided into sub-regions, is called *spatially extensive* data. Extensive data is data that applies to an entire region, but not any given sub-region.

Conversely, data that applies to any given sub-region of a region is called *spatially intensive.* Properties like population density are spatially intensive under the uniformity assumption, because the ratio of population to area does not change upon examining a sub-region. Percentages are also spatially intensive - considering the sub-population as above, the percentage of non-White Hispanic people in Region X (20 out of 100, 20%) does not change when we look at one of the sub-regions (10 out of 50, 20%).

We will use the following example case to illustrate the areal interpolation process, illustrated in Figure 2.4. Suppose regions A, B, C, and D are the source regions (each taking up a quadrant of the square), and regions X, Y, and Z are the target regions (each taking up a third of the square vertically). Further suppose that the boundaries of the source regions and target regions are fully coincident[6], that is $A \cup B \cup C \cup D = X \cup Y \cup Z$.

---

[5]Mostly time limitations; see Conclusion section for further research ideas.

[6]This example ignores the case where the covered regions are not coincident, which is a possibility in general. In this research we enforce coincidence in our data, however.
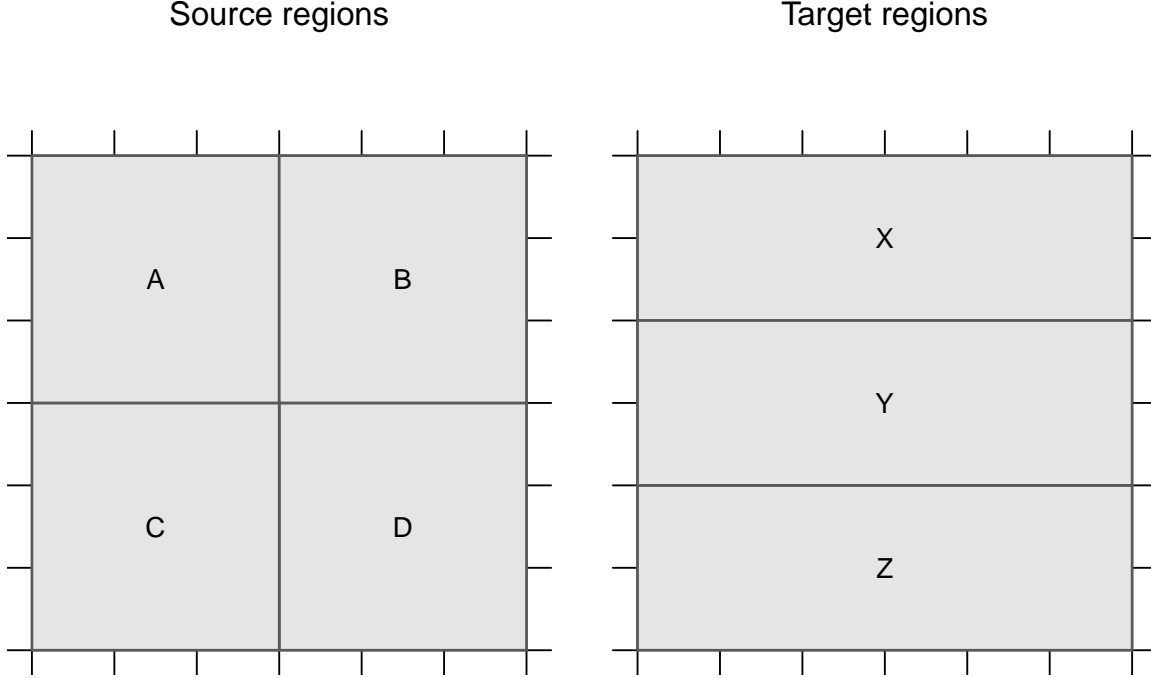
Source regions   Target regions

Figure 2.4: Example regions for interpolation

In general, denote a source region by S_i (over index set I) and a target region by T_j (over index set J). For any region R, denote the area of R by $Area(R)$, the measure of a given extensive property of R by x_R, and the measure of a given intensive property of R by y_R. These measures are known in the source regions, but unknown in the target regions (hence, the interpolation). Denote an estimate of a quantity with a caret, e.g. $\widehat{y_R}$.

Say we want to estimate x_{T_j}, the measure of the extensive property in region T_j. For all $j \in J$, we can estimate this property with

$$\widehat{x_{T_j}} = \sum_{i \in I} \frac{Area(S_i \cap T_j)}{Area(S_i)} \cdot x_{S_i}$$

For each source region, calculate the proportion of the region that lies inside the target region. These proportions are the weights that go into a weighted sum of the source regions' properties. In the example case, consider target region X:

$$\widehat{x_X} = \frac{Area(A \cap X)}{Area(A)} \cdot x_A + \frac{Area(B \cap X)}{Area(B)} \cdot x_B + 0 + 0 = \frac{2}{3}(x_A + x_B)$$

Since A and B each have $2/3$ of their area inside target region X, we estimate that $2/3$ of the extensive quantity x_A is inside region X (similarly for x_B). Adding these together gives us an estimate $\widehat{x_X}$.

For intensive properties, the process is slightly different. For all $j \in J$, we can estimate an intensive property y_{T_j} with

$$\widehat{y_{T_j}} = \sum_{i \in I} \frac{Area(S_i \cap T_j)}{Area(T_j)} \cdot x_{S_i}$$

Since the intensive property isn't "divisible" within a region, we instead take a weighted mean of the component source regions of the target region, where each component is weighted by the amount of the target region it takes up. Again considering target region X:

$$\widehat{y_X} = \frac{Area(A \cap X)}{Area(X)} \cdot y_A + \frac{Area(B \cap X)}{Area(X)} \cdot y_B + 0 + 0 = \frac{1}{2}(y_A + y_B)$$

Since A and B each take up half of target region X, each of them gets weighted by half before being added to get the estimate of y_X.

To validate the application of this method to the data at hand, we can visually compare the spatial distribution of a variable before the interpolation to its distribution after the interpolation (Figure 2.5).

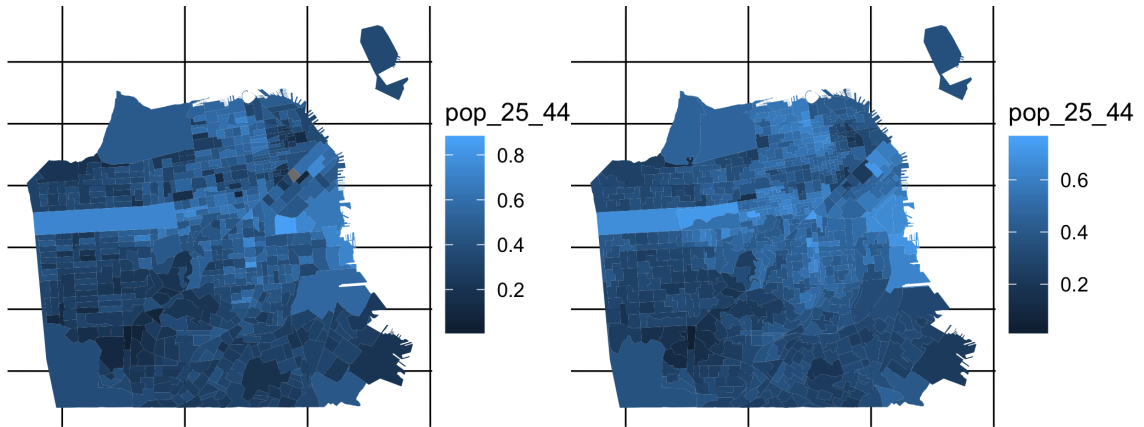Pre-interpolation (block groups)   Post-interpolation (precincts)

Figure 2.5: Proportion of residents aged 25-44

Interpolation - counts vs. percentages

One consideration in the data preparation stage was whether to use counts or percentages as input for the regression. The census data contains total population in a region, as well as a count and percentage for a given variable (say, people between the ages of 25 and 44). The percentage can equivalently be calculated by dividing the count by the population. After performing the areal interpolation on the data, we ran this percentage calculation again to double check that it lined up with the reported percentages (post-interpolation). Our intuition was that these steps should be commutative - calculating a percentage and then interpolating should have the same result as interpolating and then calculating the percentage. This was not the case, however - in the variable for population between 25 and 44, the error between these two methods ranged from -26 to 12 percentage points (Figure 2.6).
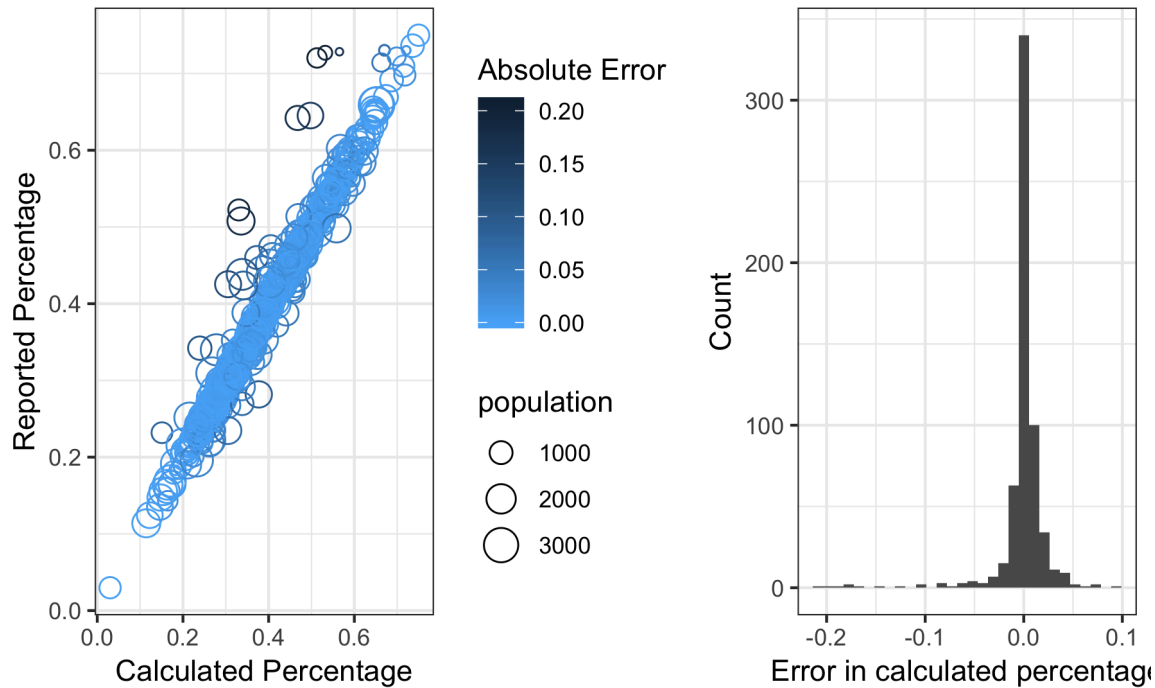
Figure 2.6: Sample error in intensive interpolation - Percentage aged 25-44

As it turns out, these steps are not commutative in this way, and the observed error is mostly a function of the weighting between different steps in the process. For example, consider a simple case: suppose source regions A and B are fully contained in target region X, and split X in half. Let the number of people between 25 and 44 in A be 4 (out of 10 total) and in B be 3 (out of 5 total). Taking the percentages first gives us a proportion of 0.4 in A and 0.6 in B. Using the intensive interpolation method on these proportions, both are weighted by the amount of X that the region takes up (half, in each case) and added, so the average weighted by area is 0.5. Conversely, using the extensive interpolation method on the count and popuation, we see that X

has 7 people between 25 and 44 (out of 15 total). Taking the percentage, we see that the proportion of this variable in X is ~0.47.

In short, the "calculating proportion" and "areal interpolation" steps are NOT commutative, because of the differences in weights when using intensive vs. extensive interpolation methods. Both are calculating a weighted mean of sorts, but intensive interpolation is weighting by **area**, while extensive interpolation is weighting by **population**. In this case we see that the latter is more accurate - source regions with more people should have greater impact on the estimated measures in target regions, because the variables we are dealing with are in terms of people rather than space. As such, for this study we will interpolate only the count data (population, numbers of people for each measured variable) and then calculate proportions in the target regions after the interpolation. These proportions are better suited to the regression to ensure that variables are of the same scale and we can compare coefficient estimates.

Boundary mismatch

A further issue appears - just like the precinct and census boundaries don't line up because they come from different sources, the outside boundaries of the precinct and census files don't line up. The census boundaries have a lower resolution on the whole than then precinct boundaries. This leads to issues when performing the areal interpolation, because parts of a census boundary that are outside of a precinct will be dropped and cause an underestimate of the true measure of the variables. We address this by bounding both files to only consider the space that is contained in both[7]. Since we are dealing with a fundamental unit of people instead of space, this ensures that every person is counted, and changes some of the spatial weights calculated. The assumption here is that any region which is only contained in one of the files (precincts without census, or census without precincts) has zero population, because every person should be contained in both a precinct and a census region.

---

[7]No census tracts were fully removed in this process, but this causes one precinct, Precinct 9900, to be cut off. However, this precinct is a semi-exclave of the county on Alameda Island, across the San Francisco Bay. Since this land, an undeveloped former naval air base, is uninhabited (Levi, 2018) its removal does not impact our results.
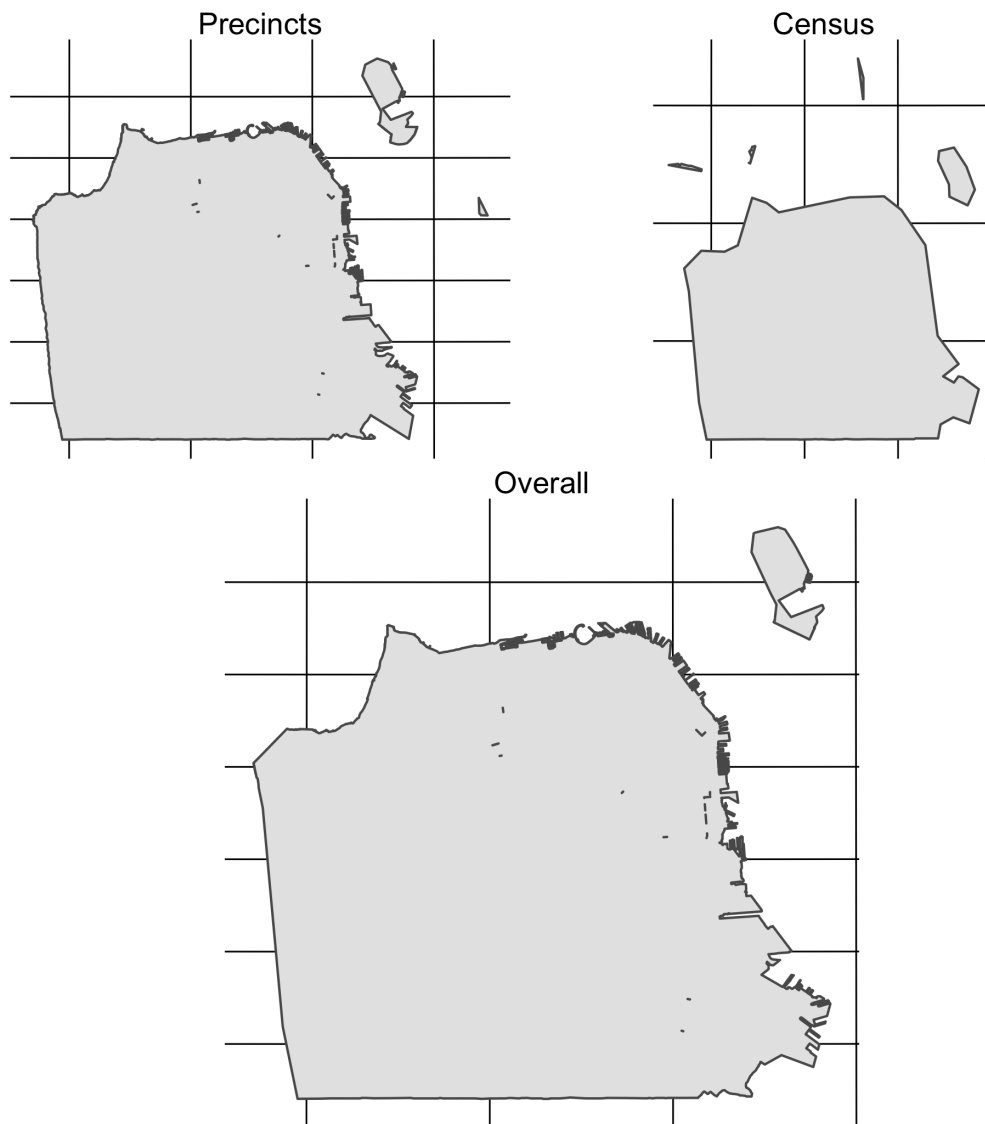
Figure 2.7: Boundary intersection between shapefiles

Figure 2.7[8] displays the result of this intersection between the two regions. The impact on the census tracts is quite visible - notably the eastern coastline is much more defined, and the gap in Treasure Island (the large island off the northeastern shore) appears. The impact on the precinct boundaries is less apparent, but still there - for example, in the southeastern corner it is visible how the angles in the precinct boundary have softened to the more rounded final shape.

Figure 2.8 is a comparison of the original shapefiles[9] (including all divisions) to

---

[8]The plots were clipped to display the same area. There are some extra islands out of range on the precinct and census maps that are not shown.

[9]The original census regions contained the Farallon Islands, which have been removed from this plot because they are uninhabited, 30 miles into the Pacific Ocean, and messed up the scale of the graph.

the bounded versions.

Figure 2.8: Original versus bounded shapefiles

Precinct Consolidation

One peculiarity in San Francisco is the combination of certain precincts during elections. California state law allows for counties to "consolidate" precincts with low numbers of registered voters during an election. This eases administration by not requiring counties to set up and staff a full polling place in a precinct with few voters, while still giving voters a physical polling location in their approximate area[10]. While San Francisco does this less often than other counties, there are still some precincts that are consolidated each election.

The areal interpolation process produces demographic estimates for areas in the map, however, which are non-consolidated. This causes a mismatch: we have demographic data for individual precincts, but election data for the consolidated precincts. To address this issue, we have split the election data (a count of overvotes

---

[10]This method of precinct consolidation may change with the 2018 California Voter's Choice Act, which lets counties (except for Los Angeles County) move fully to vote-by-mail in combination with voting centers (Allen & Hertzberg, 2016).

Table 2.1: Consolidated Precincts - Original Data

Election Pct.	Pct. Number	Over	Not over	Under	Not under
Pct 1104/1105	1104	1	455	133	323
Pct 1104/1105	1105	1	455	133	323
Pct 7509/7511	7511	0	491	151	340
Pct 7509/7511	7509	0	491	151	340
Pct 7527/7528	7527	2	392	119	275
Pct 7527/7528	7528	2	392	119	275

Table 2.2: Consolidated Precincts - Weighted Data

Election Pct.	Pct. Number	Weight	Over	Not over	Under	Not under
Pct 1104/1105	1104	0.4060452	0	185	54	131
Pct 1104/1105	1105	0.5939548	1	270	79	192
Pct 7509/7511	7511	0.7438036	0	365	112	253
Pct 7509/7511	7509	0.2561964	0	126	39	87
Pct 7527/7528	7527	0.0129859	0	5	2	4
Pct 7527/7528	7528	0.9870141	2	387	117	271

and undervotes) in the consolidated precincts into their two component precincts. This split is weighted by the population of each precinct in 2010[11]. For example, suppose Precinct X/Y had 100 undervotes, the population of Precinct X in 2010 was 325, and the population of Precinct Y in 2010 was 175. Then the "population split" between X and Y is 65%-35%, and we adjust the undervotes accordingly: Precinct X should have 65 of the 100 undervotes, and Precinct Y should have the remaining 35.

Examples of this in the data are presented here. Table 2.1 is the initial data, with full consolidated counts doubled in the precincts (and thus doubled across rows), and Table 2.2 is the data with proper weights applied. Note that this is 6 (combined into 3) out of the 12 total precincts (combined into 6) that experienced this phenomenon.

2.3 Calculating overvote and undervote rates

The dependent variables considered here are rates of overvoting and undervoting in each precinct. In the raw ballot image data, two fields are flags indicating whether a voter had either an undervote or an overvote. We redefined the category of "undervote" to include duplicated votes as well[12], but drew our overvote metric directly from these

---

[11]The most simple weight would be 50-50 for each split, but this is inaccurate for many of these precincts: see the gap in the weights for Pct 7527/7528. The flaw in this method was discovered after calculating turnout of over 1000% for Precinct 7527...

[12]E.g., voting for Mark Leno three times is functionally equivalent to voting for him once and no other candidates.

Table 2.3: Processed Ballot Image

Contest	Voter ID	1	2	3	Precinct	Overvote	Undervote
Mayor	000188699	JANE KIM	MARK LENO	LONDON BREED	Pct 7916	FALSE	FALSE
Mayor	000208276	LONDON BREED	MARK LENO	NA	Pct 7876	FALSE	TRUE
Mayor	000052986	LONDON BREED	MARK LENO	JANE KIM	Pct 7204	FALSE	FALSE
Mayor	000165119	LONDON BREED	MARK LENO	ANGELA ALIOTO	Pct 7003	FALSE	FALSE
Mayor	000085554	MARK LENO	LONDON BREED	ANGELA ALIOTO	Pct 1135	FALSE	FALSE
Mayor	000041954	RICHIE GREENBERG	RICHIE GREENBERG	RICHIE GREENBERG	Pct 1132	FALSE	TRUE
Mayor	000141209	JANE KIM	MARK LENO	NA	Pct 7301	FALSE	TRUE
Mayor	000156890	LONDON BREED	NA	NA	Pct 7012	FALSE	TRUE
Mayor	000234948	LONDON BREED	ANGELA ALIOTO	ELLEN LEE ZHOU	Pct 9446	FALSE	FALSE
Mayor	000088384	MARK LENO	RICHIE GREENBERG	ANGELA ALIOTO	Pct 9127	FALSE	FALSE

flags. In this view, an overvote in a given ranking appears as `NA` in that ranking, so the only way to observe an overvote from this data is with the given flag. Since we recalculate the undervote variable, the new definition of it will always include any overvote (since these are `NA` in the data). This does not lead to a particular imprecision in our results, but should be made note of conceptually. In this election, 0.305% of mayoral voters overvoted, and 31.1% of mayoral voters undervoted[13].

2.4 Regressions

To evaluate what demographic factors impact these ballot phenomena, we use two forms of regression:

- Linear regression. Here the dependent variables are the rate (between 0 and 1) of overvoting and undervoting in each precinct. This limited range may violate some of the assumptions of linear regression (normally distributed residuals, etc.), but we will address this for each model in particular in Chapter 3.

- Logistic regression. Here the dependent variables to train the model are individuals (who either overvoted or not / undervoted or not). This study is a good fit for logistic regression because of this inherent binary nature in overvoting and undervoting, but the demographic data issues are a limitation. Since the demographic data is at the precinct level, every voter in a given precinct is given the same demographic information as predictors, which is an interesting situation. The real advantage to using the logistic method is that it cannot predict an overvote or undervote rate (for any set of demographic values) outside of $[0, 1]$.

To choose which variables went into the final models produced, we used bootstrapping and best subset selection. We bootstrapped the data multiple times, each time performing best subset selection on the data and choosing the model with the lowest BIC[14]. Then, across all of the bootstrapped models, we identified variables that were

---

[13]Further research could examine how this compares to other RCV elections, in SF or otherwise.

[14]The Bayesian Information Criterion is a metric that can be used to select a "best" model specification from a finite set of options. It is chosen here for consistency, because one of the packages we used for analysis internally uses this criterion.

included most often[15] to put into our final model.

In addition to modelling overvote and undervote rate, we also model turnout rate in general. This gives us an opportunity to see if the demographic factors associated with RCV ballot issues are different from those associated with voting more generally.

---

[15]Arbitrarily, any variable that appeared in more than half of the sample models was included.

Chapter 3

Demographic analysis results

For the following models, I use variables[1] taken from the US Census 2018 Planning Database[2], estimated for the election precincts through the areal interpolation method[3]. All data is self-reported through the Census process[4]. A description of these variables is in the Appendix.

3.1 Voter turnout

The turnout across the city is displayed in Figure 3.1. This election apparently saw low turnout (generally sub-50% in precincts), and has a right-skewed distribution - that is, a handful of precincts had notably high turnout rates. There is no discernable spatial pattern to the turnout rate, other than maybe some slightly higher turnout in the central and east central parts of the city.

---

[1]These variables do experience some collinearity, but were included anyway in an attempt to "cover all our bases" in regards to the demographic variables. Further research could better refine the choice of variables.

[2]All descriptions taken from the Planning Database as well.

[3]We removed from the dataset one outlier with a turnout rate of 124%, Precinct 7024. We believe this is a particularly egregious inaccuracy in the interpolation process of calculating population. The gaps in the maps in this chapter are Golden Gate Park (to the northwest), Crocker-Amazon Playground and John McLaren Park (to the south), and our removed precinct 7024 (to the southeast).

[4]In regards to the `female` variable: the Census specifically asks about binary sex; there are currently no questions about gender identity. We thus refer to this as "percentage female" later on to avoid undue conclusions as a result of this distinction.
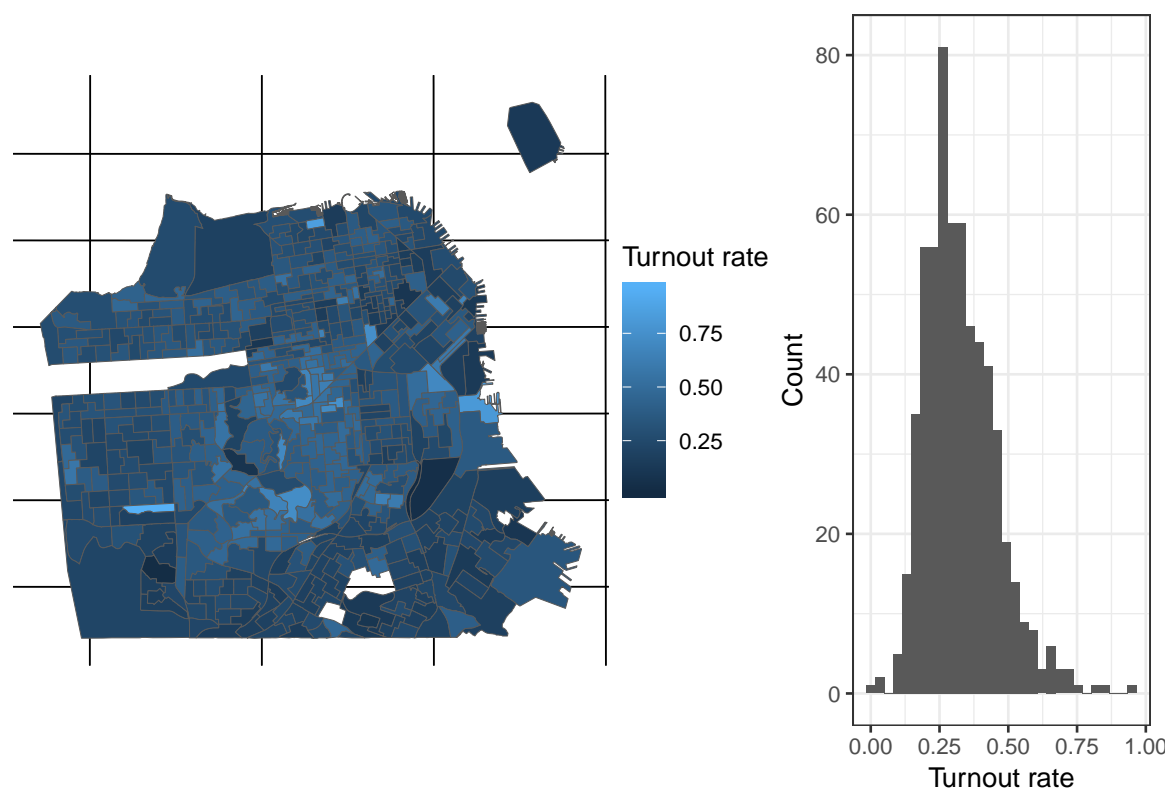
Figure 3.1: Observed turnout rate by precinct

3.1.1   Linear

With a BIC value of 420.5, the best linear model (Table 3.1) we found for predicting turnout[5] included variables for age, education, and race. Having more young people (ages 18-24) and people without high school degrees was negatively correlated with turnout, while having more middle-aged, college-educated, and White people was positively correlated with turnout. This is consistent with general literature on voter turnout - racial and ethnic minorities vote less, while the older and better educated vote more.

---

[5]This was actually set to predict the natural log of turnout because the raw turnout saw heteroskedasticity in residuals, which indicates that a linear model was not an accurate descriptor of this phenomenon.

Table 3.1: Linear model for RCV voter turnout (log)

term	estimate	std.error	statistic	p.value
Intercept	-1.6641254	0.1423196	-11.692877	0.0000000
pop_18_24	-1.7780018	0.2265656	-7.847626	0.0000000
pop_45_64	0.9452890	0.2411011	3.920717	0.0000985
no_hs	-0.7112645	0.2721442	-2.613558	0.0091868
college	0.4605347	0.1931316	2.384565	0.0174111
white	0.3787332	0.1462976	2.588786	0.0098666
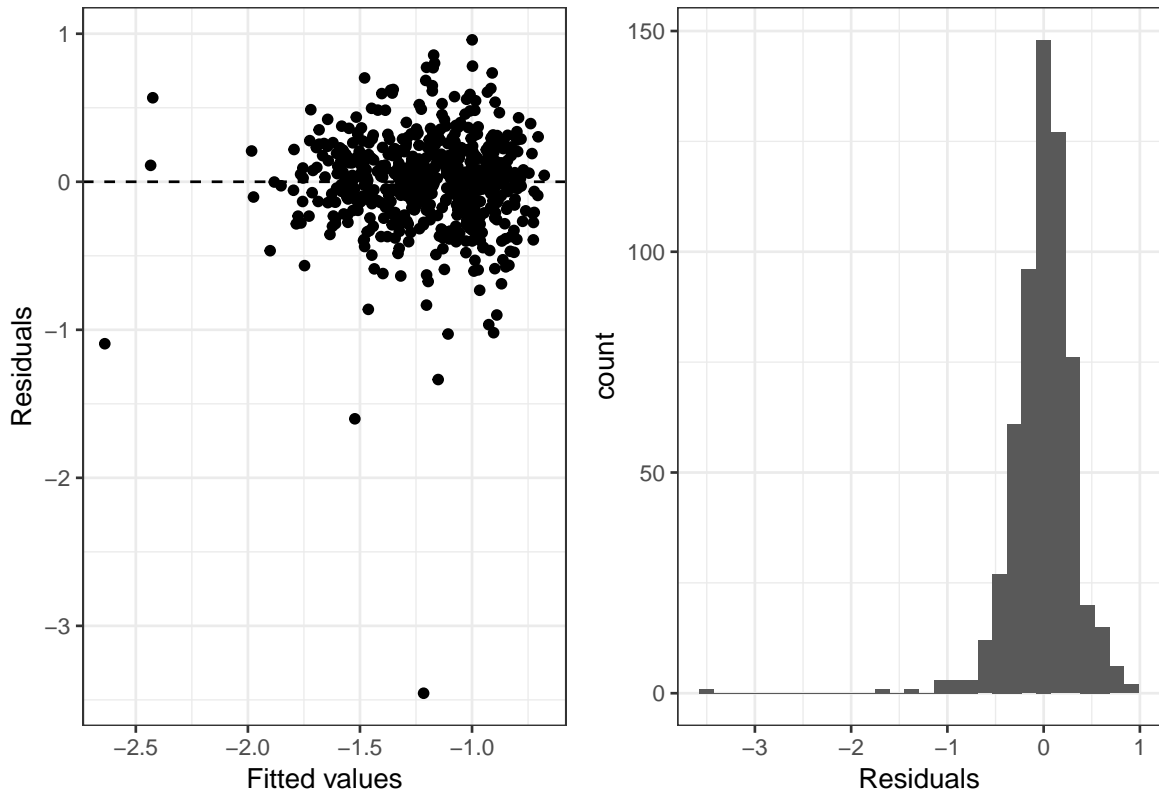
Figure 3.2: Linear turnout model validation

Figure 3.2 details the residuals from the model. These are normally distributed, with no systematic change in variance. There is one particularly low outlier, but on the whole the residuals are distributed evenly. Unfortunately, the residuals here are quite large - since we're predicting turnout rates, a "valid" residual[6] should only range between -1 and 1. This indicates that this model is not particularly well fitted to the data.

---

[6]That is, a residual where the predicted value is indeed a rate between 0 and 1.

Table 3.2: Logistic model for RCV voter turnout

term	estimate	std.error	statistic	p.value
Intercept	-0.8735670	0.1123734	-7.7737871	0.0000000
college	0.5432950	0.0372575	14.5821495	0.0000000
no_hs	-0.7027148	0.0620006	-11.3340093	0.0000000
pop_18_24	-1.3239340	0.0665276	-19.9005207	0.0000000
pop_45_64	1.9863400	0.0761654	26.0792823	0.0000000
poverty	-0.6150431	0.0481346	-12.7775777	0.0000000
asian	-1.2430027	0.1070423	-11.6122624	0.0000000
pac_islander	-0.3304069	0.2998461	-1.1019217	0.2704957
pop_65_plus	1.4978564	0.0667685	22.4335837	0.0000000
female	-0.1874513	0.0532668	-3.5191037	0.0004330
native	-6.1301592	0.6277355	-9.7655131	0.0000000
no_english	-0.1342568	0.0535481	-2.5072196	0.0121685
hispanic	-0.8818015	0.1061429	-8.3076859	0.0000000
white	-0.4570681	0.1069476	-4.2737557	0.0000192
black	-0.9048225	0.1158023	-7.8135098	0.0000000
pop_25_44	0.3555263	0.0530427	6.7026379	0.0000000
other_race	-0.0807165	0.2791823	-0.2891177	0.7724913

3.1.2   Logistic

With a BIC value of 3.3019×10^4, the best logistic model (Table 3.2) we found for predicting turnout[7] included all of our variables. Education and age were positively correlated with turnout, while poverty, percentage female, and non-English-speaking population were negatively correlated. All of the racial groups were correlated negatively with turnout, and there is no distinct pattern among minority groups versus White voters.

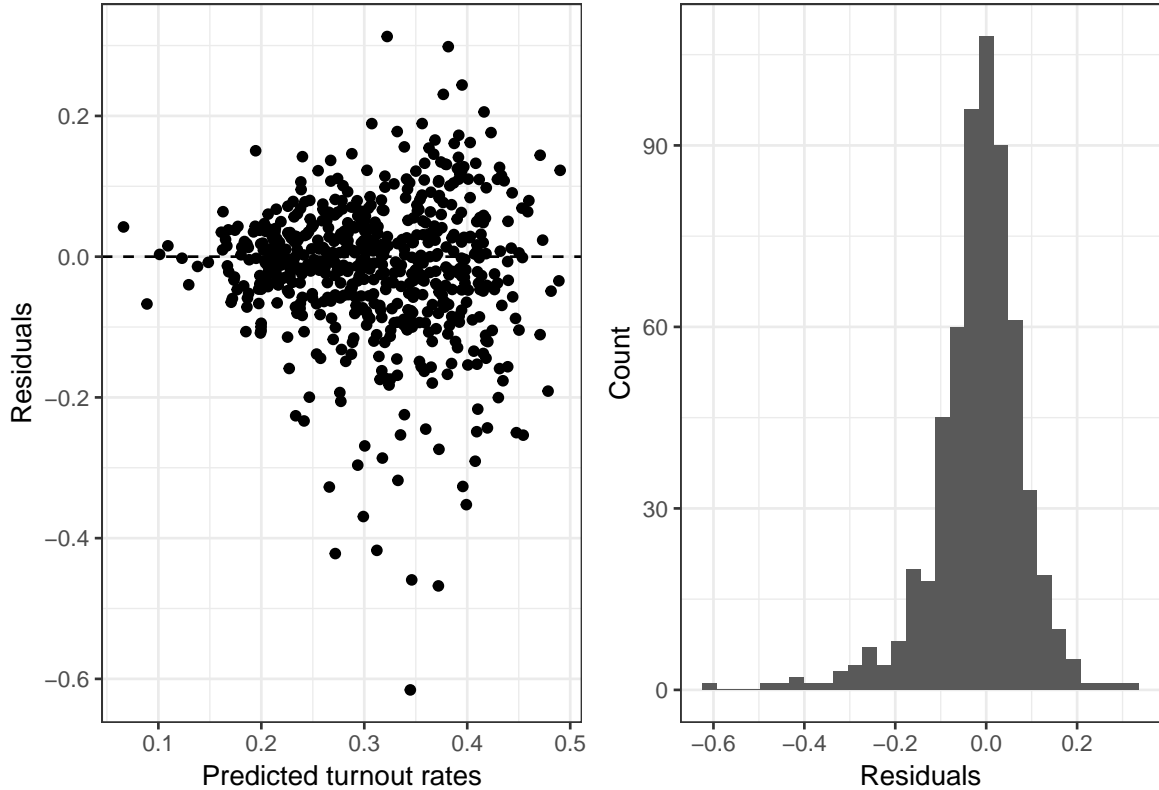[7]Now, we are predicting the raw, un-logged turnout scores again.

Figure 3.3: Logistic turnout model validation

Figure 3.3 details the residuals of our predictions[8]. These are normally distributed (with a slight left skew), and notably are much smaller in general than the linear method[9]. The middle 95% of residuals for the linear model lie between -0.62 and 0.599, significantly wider than the -0.254 and 0.15 from the logistic method. We see some heteroskedasticity, as the residuals are much smaller at the high and low ends of the predicted values. While a better version of this model may exist, the smaller residuals alone indicate that this is a much better explanation of the turnout method than the linear model.

3.2   Overvoting

The rate of overvoting across the city is displayed in Figure 3.4. Overvote rates are usually very small fractions, and here we see that no precinct had an overvote rates higher than 2%. Many, in fact, had no overvotes at all. There is no discernable spatial pattern to the overvote rate.

---

[8]This is somewhat an abuse of the linear model validation plots. Since we're using the logit link function to predict a turnout rate (rather than an individual's binary turnout measure), this was the best visualization I could come up with to express the "accuracy" of the logistic method. There are no equivalent assumptions that these plots address, like in the linear case.

[9]They are bound to $[-1, 1]$ because the logistic method outputs a probability, but are still much smaller aside from this limitation.

Table 3.3: Linear model for overvote rates

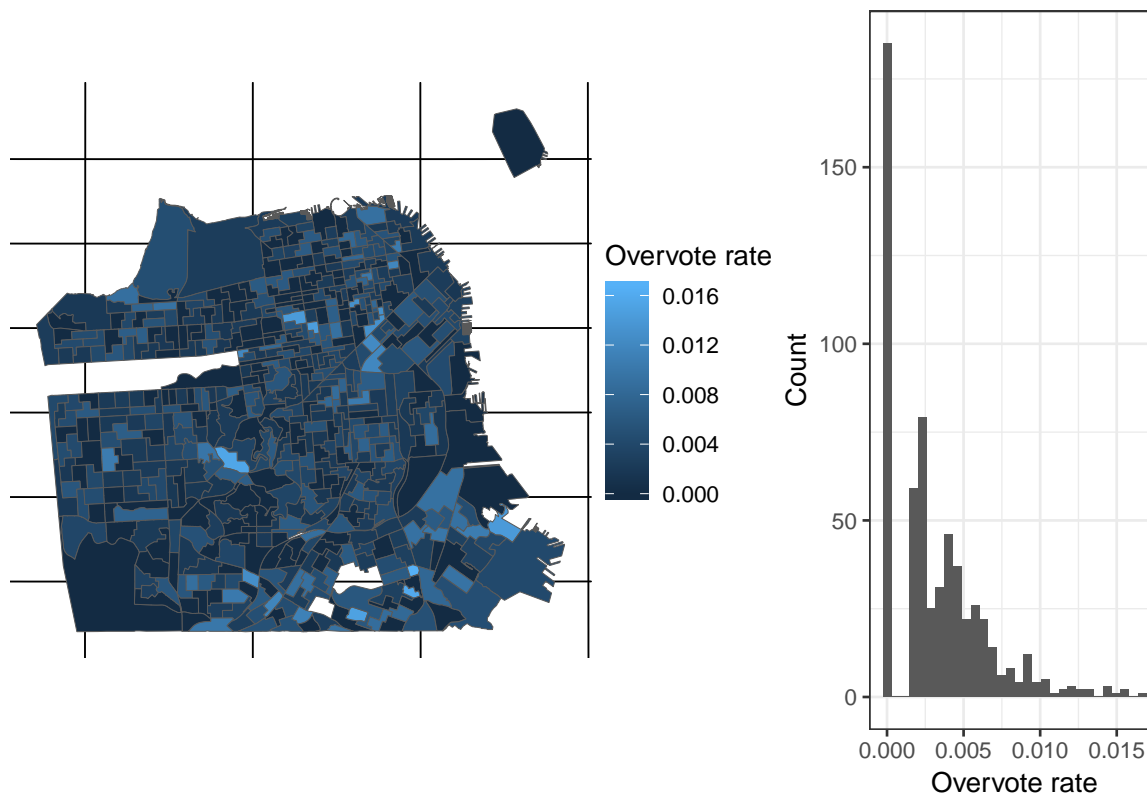| term	estimate	std.error	statistic	p.value
Intercept	0.0025553	0.0001451	17.615646	0
black	0.0117459	0.0015530	7.563142	0

Figure 3.4: Observed overvote rate by precinct

3.2.1   Linear

With a BIC value of -5267.6, the best linear model (Table 3.3) we found for predicting overvote rates included only the variable `black`. Having more African-Americans in a precinct was positively correlated with overvoting.
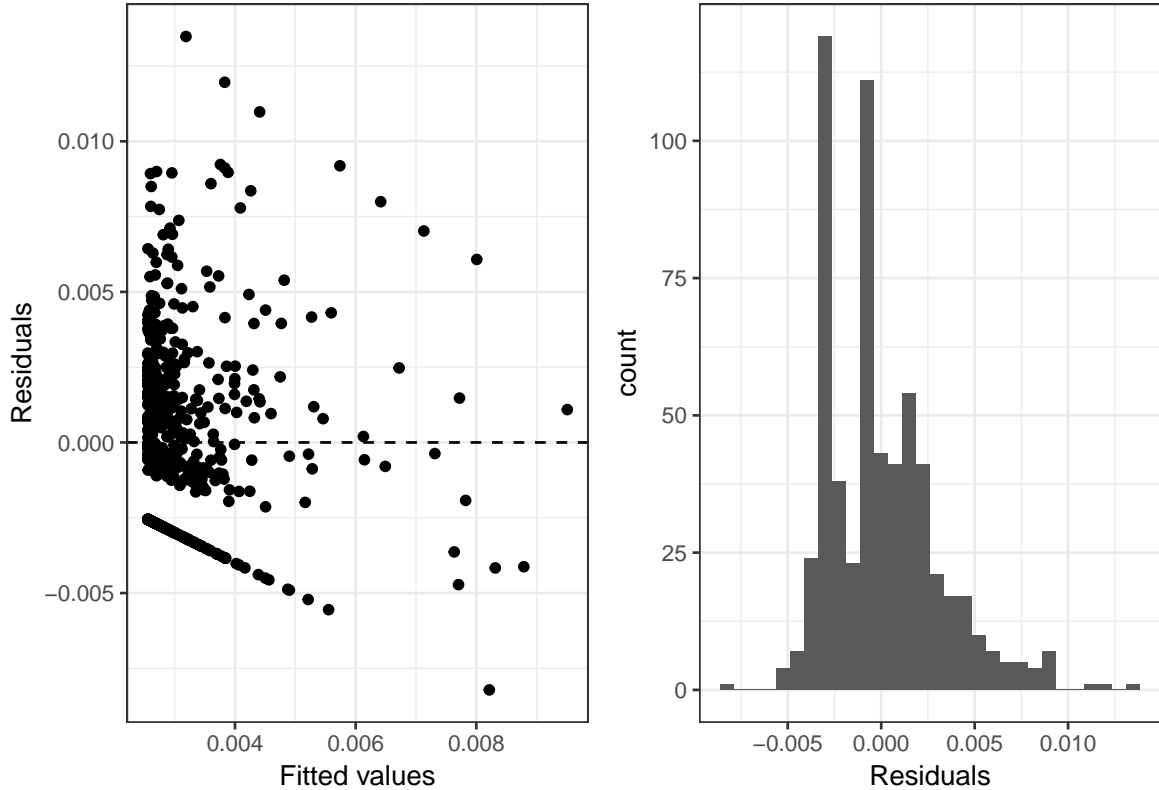
Figure 3.5: Linear overvote model validation

Figure 3.5 details the residuals from the model. Here we see that the linear model is not a great fit for this data. The combination of low overvoting rates and a lower bound at 0 mean that we see significant heteroskedasticity in the residuals. They are skewed in a strange way due to the prevalence of so many 0s, with the "normal mass" of residuals showing a right skew but the 0 residuals causing large spikes on the negative end. Interestingly, there also seems to be a line of residuals in the upper section of the scatter plot, indicating that there are a set of precincts with roughly the same high overvote rate that were predicted to have much lower rates. All this indicates that this model is poorly fitted to the data.

3.2.2 Logistic

With a BIC value of 1733.4, the best logistic model (Table 3.4) we found for predicting overvote rates also included only the variable `black` (again positively correlated with overvoting).

Table 3.4: Logistic model for overvote rates

term	estimate	std.error	statistic	p.value
Intercept	-5.925391	0.0425953	-139.109149	0
black	2.550279	0.3543068	7.197942	0
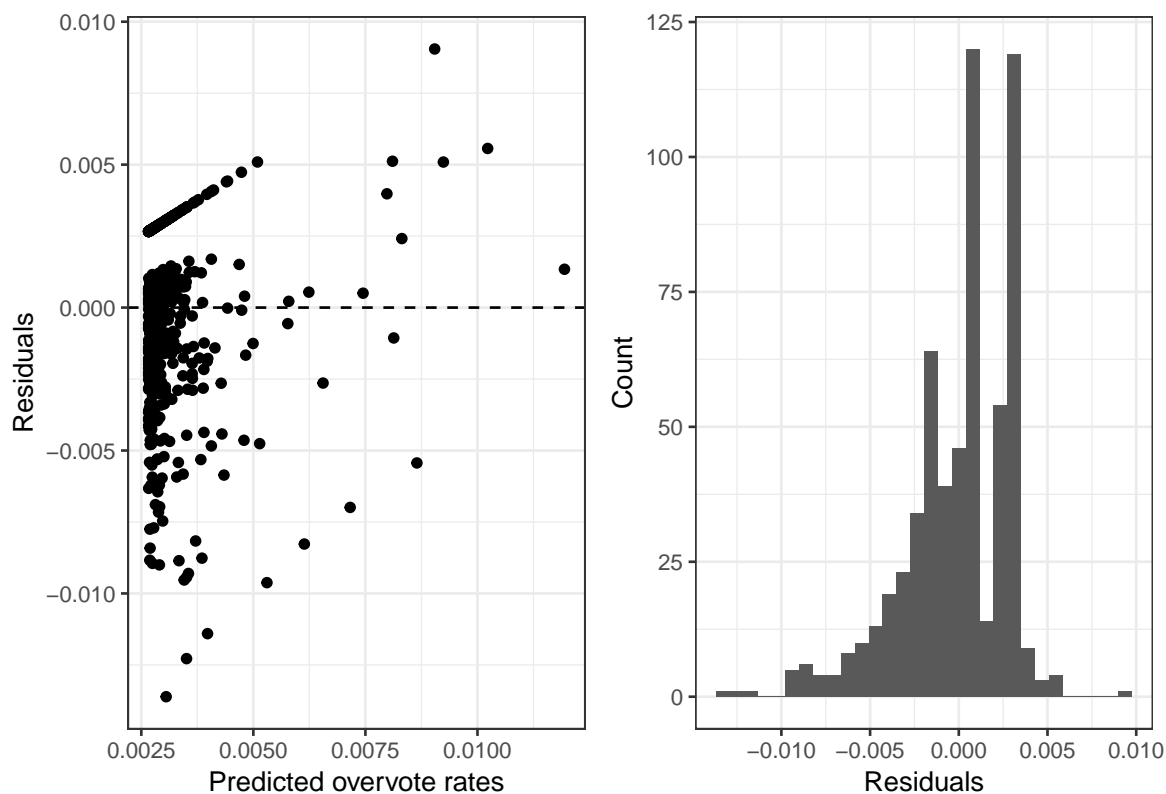
Figure 3.6: Logistic overvote model validation

Figure 3.6 details the residuals of our predictions. We see a similarly strange pattern in the residual scatterplot as in the linear overvote model, but this time the lower line is curved and the upper line is straight. There is no clear modal pattern in the residuals, except for an increase moving from the negative end that looks roughly like a bell curve. The middle 95% of the data is equally spread between the two methods ([-0.004,0.008] for the linear model and [-0.008,0.004] for the logistic), but they have different signs - the logistic model has more negative values, the linear more positive. Neither the linear nor the logistic model appears to be a good fit for overvote rates here. Perhaps a zero-inflated method would perform better with this unusual distribution.

3.3 Undervoting

The rate of undervoting across the city is displayed in Figure 3.7. These are somewhat normally distributed, and show no clear spatial pattern. Most precincts had between 20% and 40% of voters undervote.
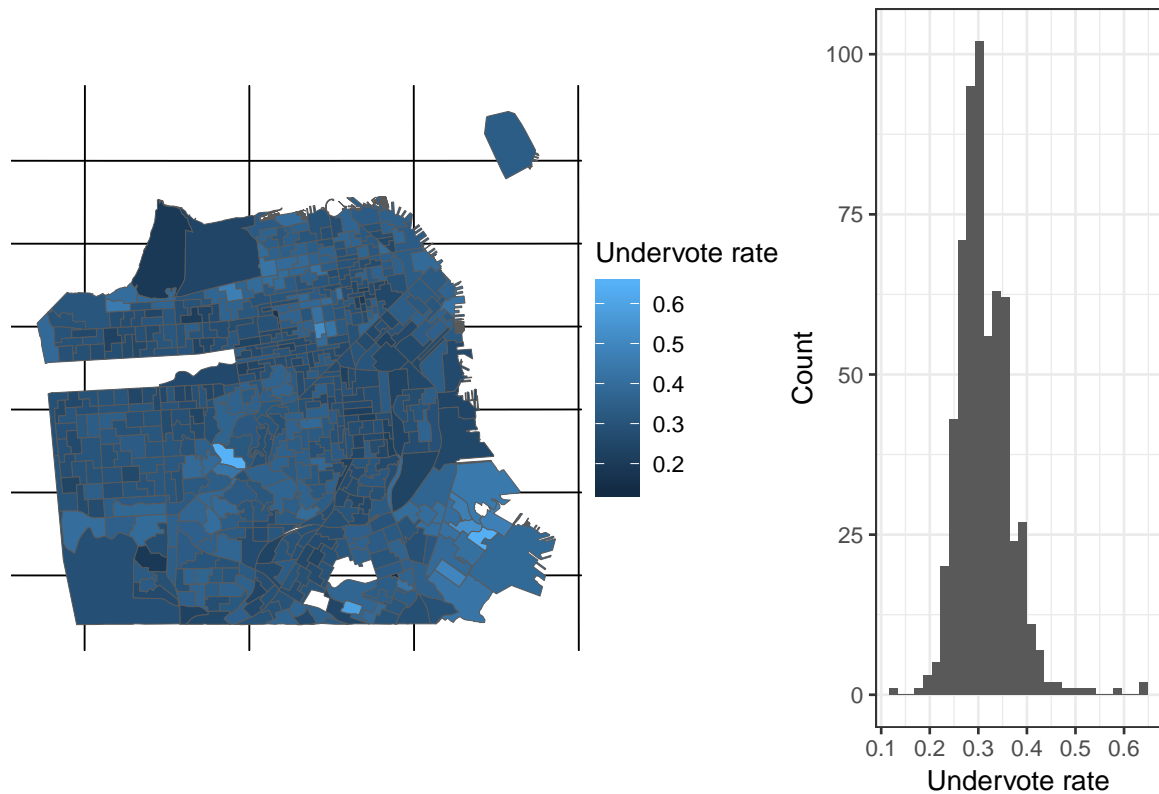
Figure 3.7: Observed undervote rate by precinct

3.3.1 Linear

With a BIC value of -2039.4, the best linear model (Table 3.5) we found for predicting undervote rates included predictors about age, education, and race. The three youngest age categories were negatively correlated with undervoting, both education variables were positively correlated, and three ethnic groups (`black`, `pac_islander`, `white`) were positively correlated as well. While this doesn't agree with previous literature on voting habits, neither does it provide a clear pattern with which to base a disagreement on. Particularly, we would not expect the percentage of people with a college degree *and* the percentage of people with no high school diploma to trend in the same direction.

Table 3.5: Linear model for undervote rates

term	estimate	std.error	statistic	p.value
Intercept	0.4063591	0.0243818	16.666513	0.0000000
black	0.2855476	0.0269697	10.587734	0.0000000
pop_18_24	-0.2706175	0.0323329	-8.369733	0.0000000
pop_25_44	-0.3173180	0.0232191	-13.666272	0.0000000
pop_45_64	-0.2452776	0.0405118	-6.054474	0.0000000
college	0.1048400	0.0258765	4.051554	0.0000576
pac_islander	0.6842866	0.1690560	4.047691	0.0000586
white	0.0529474	0.0189889	2.788329	0.0054681
no_hs	0.1262080	0.0361406	3.492142	0.0005148

Figure 3.8: Linear undervote model validation
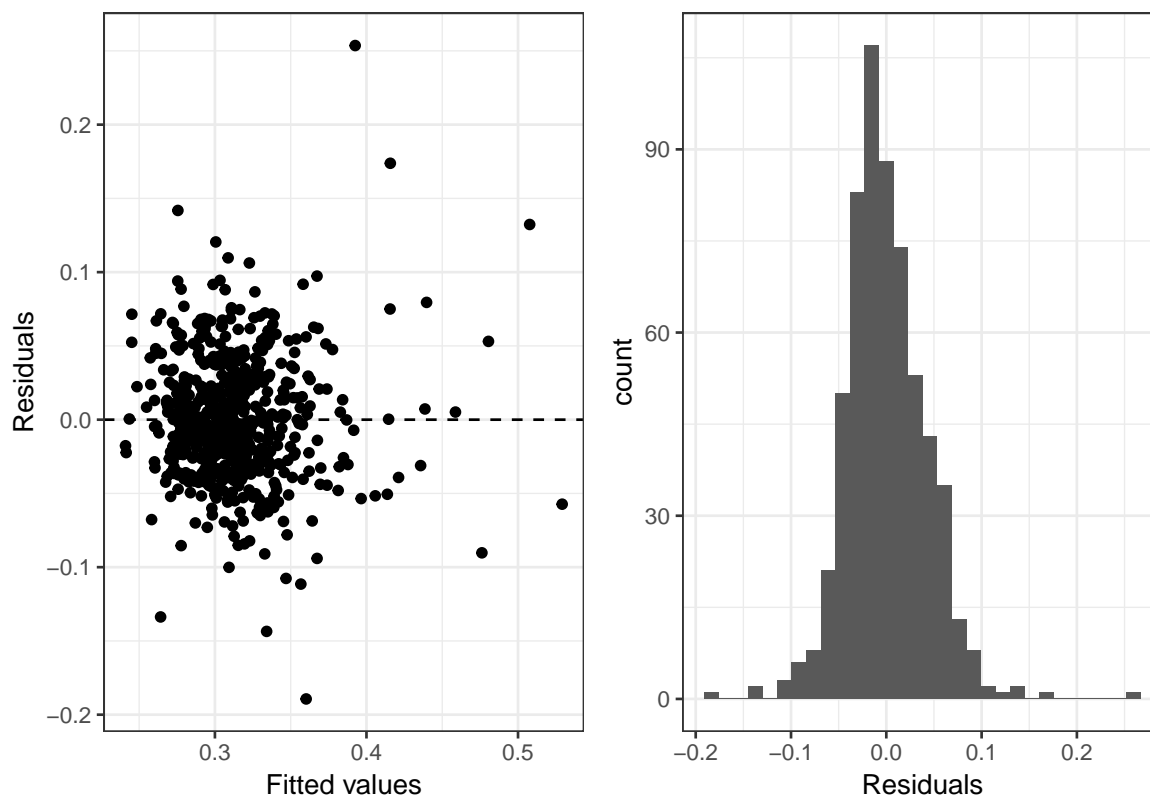
Figure 3.5 details the residuals from the model. The linear model seems to be a decent fit for this data, with relatively normal residuals and no clear variance pattern. The scale of the residuals could be improved, but on the whole this is not a poorly fitting model.

Table 3.6: Logistic model for undervote rates

term	estimate	std.error	statistic	p.value
Intercept	-1.4927657	0.1762335	-8.470387	0.0e+00
black	2.4368837	0.1921335	12.683285	0.0e+00
college	0.5488777	0.0667635	8.221218	0.0e+00
pop__18__24	-1.0290278	0.0950082	-10.830940	0.0e+00
pop__25__44	-1.3470105	0.0604195	-22.294293	0.0e+00
pop__45__64	-0.9796601	0.1029356	-9.517211	0.0e+00
white	1.2514086	0.1738961	7.196299	0.0e+00
asian	1.0982593	0.1729419	6.350453	0.0e+00
hispanic	1.0818430	0.1769331	6.114420	0.0e+00
no__hs	0.4545196	0.0946414	4.802546	1.6e-06
pac_islander	3.5887730	0.5142211	6.979047	0.0e+00

3.3.2 Logistic

With a BIC value of 5493.9, the best logistic model (Table 3.6) we found for predicting undervote rates also included variables in age, education, and race. We see this strange pattern in variable coefficients here as well - both educational groups and all of these ethnic groups are positively associated with undervoting (no clear minority vs. White distinction), and the three younger age categories are negatively associated with it.
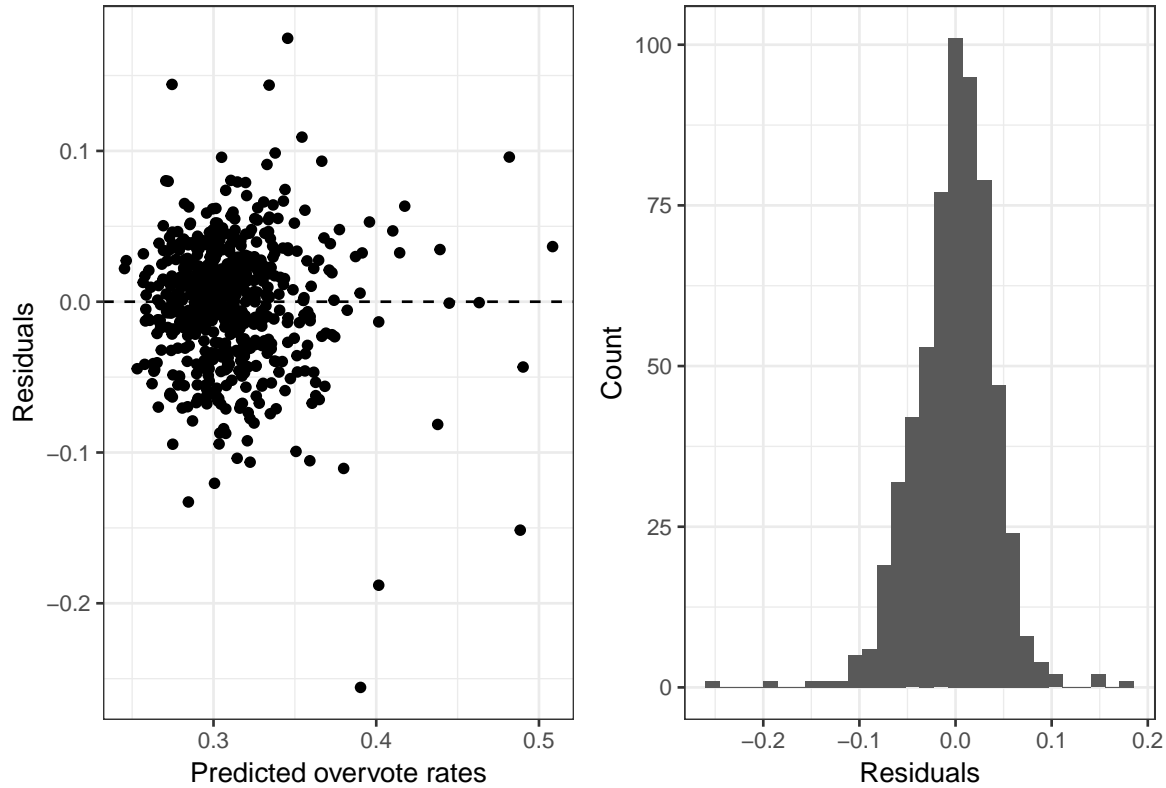
Figure 3.9: Logistic undervote model validation

Figure 3.9 details the residuals of our predictions. These look very similar to the linear model: approximately normally distributed, with no clear heteroskedasticy, and even the same cluster shape in the residual scatter plot. They even have similar spreads in the residuals ([-0.073,0.079] for the middle 95% of the linear model's residuals and [-0.084,0.074] for the logistic). Under all these metrics, the logistic model performs well. Despite these models seeming to fit appropriately, it's hard to draw a useful conclusion from them due to the aforementioned issues with variable meaning.

3.4   Summary

Table 3.7: Comparison of models fit

Metric	Model	BIC	Residual spread (95%)	Residual shape
Turnout	Linear	420.537241	1.219	Normal
	Logistic	3.301895×10^4	0.404	Normal
Overvoting	Linear	-5267.5585476	0.012	Non-normal
	Logistic	1733.3913291	0.011	Non-normal
Undervoting	Linear	-2039.4170766	0.152	Normal
	Logistic	5493.8649095	0.158	Normal

In Table 3.7 we compare the models fit earlier. While the BIC is not entirely comparable between the linear and logistic models, we can still draw some insights here. A lower BIC is better, so on the whole the models for overvoting and undervoting were much better fits than the models for turnout. The spread of residuals also supports the above. We did see, however, that

1. The overvoting models had some significantly non-normal residual patterns, and

2. The undervoting models showed strange relationships between the coefficients.

Overall, however, this analysis shows that the rates of overvoting and undervoting are affected by some of the same demographic considerations as turnout rate in general is affected by: age, race, and education.

Chapter 4

Missing data literature review

4.1 Missing data

Many statistical methods require complete data sets, that is data where every possible observation has a proper value, to work correctly (univariate means, regression, etc.). In many fields of research, however, data collection often results in missing data. Respondents skipping questions (nonresponse), respondents not completing all phases of a multi-stage survey (attrition), government agencies not reporting certain data (redaction), and errors in data collection are all examples of reasons that data can be missing.

Suppose we have a matrix $A = (a_{ij})$ with some missing data, which has m rows and n columns. Let $\tilde{A}$ be the complete version of dataset A, where every missing data point in A is now observed. Let M be an $m \times n$ matrix such that

$$M = (m_{ij}), \text{ where } m_{ij} = \begin{cases} 1, & \text{when } a_{ij} \text{ is missing} \\ 0, & \text{when } a_{ij} \text{ is not missing} \end{cases}$$

Let $\tilde{A}_{obs}$ be the observed data in $\tilde{A}$, that is the entries in $\tilde{A}$ corresponding to the entries in M that equal 0, and $\tilde{A}_{mis}$ be the unobserved data in $\tilde{A}$, that is the entries in $\tilde{A}$ corresponding to the entries in M that equal 1. Let ϕ denote some unknown parameters.

The literature on missing data has identified three distinct subtypes of missing data (Little & Rubin, 2014).

- Missing completely at random (MCAR). The distribution of missing values does not depend on any data in Y, observed or missing.

$$f(M|\tilde{A}, \phi) = f(M|\phi) \ \forall \ Y, \phi$$

  An example of this is a survey where the respondent mislabeled a question answer by accident.

- Missing at random (MAR). The distribution of missing values depends on the observed data, but not any missing data.

$$f(M|\tilde{A}, \phi) = f(M|Y_{obs}, \phi) \ \forall \ \tilde{A}_{mis}, \phi$$

An example of this is an election poll where members of one party are less likely to report their true vote choice than members of another party. The distribution of missing vote choice depends on an observed variable, party identification.

- Non-ignorable (NI), or missing not at random (MNAR). The distribution of missing values depends on the missing values themselves. An example of this is high-income respondents leaving the income field blank in a survey to obscure their true earnings.

4.2   Methods of dealing with missing data

As with any data problem, a common task for a dataset with missing data is to obtain a useful estimator. Let $\hat{x}$ be an estimator of some population parameter x. Some of the qualities that are good in an estimator are:

1. Lack of bias. $\hat{x}$ is *unbiased* if $E(\hat{x}) - x = 0$, that is we expect $\hat{x}$ to be an accurate guess at x (Nikulin, n.d.).

2. Efficiency. An estimator $\hat{x}_1$ is more efficient than $\hat{x}_2$ if the sampling variance of the estimator $\hat{x}_1$ is less than the variance of the estimator $\hat{x}_2$, that is we expect $\hat{x}_1$ to be a more precise guess at x. An unbiased estimator can be called "efficient" if its variance is equal to the Cramer-Rao Lower Bound: the inverse of the Fisher Information of the unknown parameter (Fowler, n.d.).

3. Provides good estimates of uncertainty. Metrics such as confidence intervals created by the estimator should be accurate (Allison, 2011).

Listed below are some of the methods for dealing with missing data, and the qualities of the estimators they produce.

4.2.1   Listwise deletion

Listwise deletion (or *complete-case analysis*) is the most common method of dealing with missing data, as many statistical packages perform listwise deletion on datasets before running most analyses because they need complete data. In listwise deletion, any row with a missing value is fully removed from the dataset. This can lead to biased parameter estimates when the data is not MCAR. Consider the high-income censoring case as above, with average income being the parameter of interest. If high earners are censoring their income and they are dropped from the dataset, our estimate of average income will be negatively biased. Additionally, listwise deletion will reduce the number of overall observations going into any analysis, which reduces the statistical power of methods and increases the size of equivalent confidence intervals (Little & Rubin, 1987).

In terms of the three goals of an estimator above, listwise deletion performs moderately. It provides good estimates of uncertainty, and can be unbiased under certain conditions (MCAR, and some MAR cases). It is quite inefficient, however,

because it tosses away so much available data. Variance increases with a reduction in the number of data points, so more data points are needed to obtain an estimator with a desired variance (Allison, 2011).

4.2.2 Maximum likelihood estimation

Maximum likelihood estimation (MLE) methods generally ignore the missing data. The process is similar to any maximum-likelihood-based parameter estimation with full data: all of the available data is used to build a likelihood model for the parameter(s) of interest, and then the likelihood equation is maximized to obtain an estimate. MLE inferences which ignore the missing-data mechanism like this only require MAR to be valid (Little & Rubin, 1987).

MLE satisfies all of the estimator criteria listed above well, but does have some drawbacks. It takes some computationally intensive software to calculate, and involves fitting some parameters on the data that may or may not apply (Allison, 2011). In our case, because we're dealing with a very complex estimator (election winner as determined by the RCV algorithm) that truly depends on every single data point, any MLE method would be an intensely computational black box that produces a likely uninterpretable result.

4.2.3 Inverse probability weighting

Inverse probability weighting (IPW) attempts to correct for the bias of listwise deletion by weighting each observed data point by its probability of being missing. A probability model is specified that determines if a case is complete or not: $P(\text{complete} \mid \text{observed data})$[1]. Then a weight is assigned to each case, which is the inverse of its probability of being missing based on the observed data. The desired analysis is then performed with the newly-weighted data (Seaman & White, 2013).

IPW can outperform listwise deletion in terms of bias, particularly in the MNAR case where the overweighting of similar-to-missing data helps to correct for the bias in listwise deletion. IPW is less efficient than multiple imputation, because it uses only complete cases to build its weights (as opposed to MI, which uses all available data) (Seaman & White, 2013).

4.2.4 Imputation - single and multiple

Imputation, unlike the previous three methods, is a way of filling in the missing data points directly[2]. In *single imputation*, a single value is estimated for each missing cell in the data. Some variants of single imputation are:

- Mean imputation, where the mean of the observed data column is taken. This can lead to biased estimates if the data is not MCAR (the observed data is not

---

[1]The reader may note that this is similar to the definition of data being missing at random.
[2]Rather than going straight to an adjusted conclusion/parameter.

representative of the missing data), and also negatively impacts measures of variance.

- Hot deck imputation, where a value from an otherwise similar row in the data is taken. This comes in two subtypes:

  - Random hot deck, where a random case is drawn to fill in the missing values.
  - Sequential hot deck, where the most immediately previous matching case is drawn to fill in the missing values.

- Regression imputation, a model-based approach where the missing value is predicted using a regression equation that is trained using the observed data.

- Multinomial imputation, a different model-based approach which is tailored for a categorical response variable.

- Last value imputation, where the missing value is copied from the immediately preceding observed value[3]. This is particularly common in time series data, where this is a natural ordering of the observations.

Multiple imputation (MI) is a method of running single imputation multiple times to create extra variability among estimates. A dataset is imputed multiple times (with some variability between imputations), then for each new dataset the parameter of interest is calculated. This added variability moderates the appearance of "certainty" that comes with single imputation (Little & Rubin, 1987).

Hot deck imputation is unbiased under MCAR and MAR (Little & Rubin, 1987). Imputation typically underestimates variance in parameter estimates, because (after a random draw to fill in the data) it uses a deterministic method to calculate the estimate. Multiple imputation can increase this variance through multiple random draws, but still often underestimates variance. Depending on the parameter of interest, often five imputations is enough to obtain nearly-efficient estimates with MI (Allison, 2011).

4.3   What if everybody votes (fully)?

The question of "what if everybody voted" (or "what if we had 100% turnout) is one well-studied in literature, and particularly under alternative voting methods including RCV. Consider Table 4.1, which shows the twenty most common combinations of votes in the 2018 SF mayoral election. The most common vote choice was London Breed and nobody else[4], which shows the potential for many more people to fill in 2nd

---

[3]This is similar to sequential hot deck, but last value imputation doesn't have the same concerns about similarity between the cases; it just repeats the immediate last data point measured.

[4]Breed herself may have increased the rate of appearance of this combination. When asked at a candidate forum who her second and third choices for mayor would be, she replied "My No. 2 choice is London Breed, and my No. 3 choice is London Breed" (Berman, 2018). An interesting study would

Table 4.1: Examination of voter counts

1	2	3	Count	Proportion
LONDON BREED	NA	NA	20430	0.0812656
JANE KIM	MARK LENO	AMY FARAH WEISS	15654	0.0622678
LONDON BREED	MARK LENO	JANE KIM	11937	0.0474825
MARK LENO	JANE KIM	LONDON BREED	11074	0.0440497
LONDON BREED	JANE KIM	MARK LENO	10848	0.0431507
JANE KIM	MARK LENO	LONDON BREED	10484	0.0417028
MARK LENO	LONDON BREED	JANE KIM	7705	0.0306486
JANE KIM	MARK LENO	NA	7633	0.0303622
MARK LENO	JANE KIM	NA	7525	0.0299326
LONDON BREED	MARK LENO	NA	6595	0.0262333
LONDON BREED	MARK LENO	ANGELA ALIOTO	6188	0.0246144
MARK LENO	JANE KIM	ANGELA ALIOTO	5811	0.0231147
JANE KIM	LONDON BREED	MARK LENO	5761	0.0229159
MARK LENO	NA	NA	5725	0.0227727
JANE KIM	MARK LENO	ANGELA ALIOTO	4649	0.0184926
MARK LENO	JANE KIM	AMY FARAH WEISS	4514	0.0179556
LONDON BREED	JANE KIM	ANGELA ALIOTO	4238	0.0168577
LONDON BREED	ANGELA ALIOTO	MARK LENO	4231	0.0168299
LONDON BREED	ANGELA ALIOTO	NA	3548	0.0141131
ANGELA ALIOTO	NA	NA	3524	0.0140176

and 3rd vote choices[5]. While this (and most other undervotes in the first few rows here) would not have changed the election result because of how they voted for final candidates, consider row 20: people who only voted for Angela Alioto, the eventual fourth-place finisher. There were 3,524 people who voted for Alioto alone, which is larger than the final margin of 2,268 between London Breed and Mark Leno. The presence of undervotes like this has the potential to change the results of an election.

Citrin, Schickler, & Sides (2003) use auxiliary data to simulate vote choice of non-voters in US Senate elections. They raise an interesting consideration: while the overarching metric of electoral change is an election flipping, the numerical "amount" of change in an election may be large but not enough to overcome a yet larger margin of victory. For example, if full voter turnout improved Democratic vote shares by 5% across the country, this would only flip the seats where a) a Democrat lost, and b) the margin of victory for the other candidate was small enough that a 5% vote share would overturn it. We should thus consider alternate methods of assessing the impact of no undervotes or overvotes, aside from looking at who wins the election.

Liu (2014) investigates imputation methods for vote choice in a Taiwanese election

---

be to see if "London Breed" repeated 3 times shows up more often than we would expect it to, given how often other candidates appear 3 times.

[5]This data has been cleaned to consolidate repeated vote choices, so there is no difference between listing only "London Breed" as first or listing her in all three slots.

(conducted with plurality). They find that "the extent to which MI corrects the distribution [of vote share] is very limited, although the direction of adjustment is correct". Bernhagen & Marsh (2010) use multiple imputation to simulate choices in place of undervotes in an Irish general election (conducted with RCV). They find that the techniques applied were normally "not large enough to affect the counterfactual estimation of election results under universal turnout". These studies offer the idea that MI may not shift the observed data enough to observe a change, even if one would truly be present under full voting. Further work could be done into how well MI mirrors the correct amount of variance under different conditions (MCAR, MAR, etc.).

In a FairVote report, S. Hill & Hernandez (2018) find a similar result in the 2018 San Francisco mayoral election: the exhausted ballots in this election did not have the potential to swing the election away from Breed. They also find that the majority of exhausted ballots were "voluntarily exhausted" meaning that the voter undervoted, rather then being exhausted because they ran out of available slots to list. Additionally, the involuntarily exhausted ballots were likely to pull for London Breed over Leno. In addition to the small impacts of MI mentioned above, this study indicates that the positioning of the undervotes themselves in this specific election may be a factor that contributes to a lack of change in our results. Taking this into concern in combination with the MI result, we may expect to not see much variance among methods.

4.4   Peculiarities with this study

4.4.1   Violated assumptions

Earlier we made the implicit assumption that every "missing" data point (unobserved value) has some true value. In the case of ranked choice voting, this is not necessarily true. It is a fully rational choice for a voter to only list one candidate when offered three options, if they truly only have one preference. While this would seem to make this study just an exercise in missing data methods, this is no more the case here than in other studies of imputing vote choice.

Kroh (2006) has a good discussion of methods for working with imputations when "Don't know" is a valid response, or there are valid reasons for the inapplicability of a survey question. While Kroh's study does not apply to our particular research question and data, it provides a nuanced take on the assumptions inherent in missing data imputation.

Given these issues, we believe that our data is MAR and imputation will be a good method to use. Part 1 of the thesis showed that undervoting is dependent on the demographic qualities we're considering, so it's not MCAR. While it's quite difficult to tell if the data is MNAR[6], we believe that people's hypothetical 2nd and 3rd preferences themselves are not what is causing these vote choices to be missing (with the exception of truly not having such a preference).

---

[6]Since the missing data itself is what affects missing-ness, we have no way of knowing what this mechanism is caused by from looking at the observed data alone.

4.4.2 Monotonicity

In particular, our data is *monotone* missing data - that is, for a certain order of our variables (Y_i), an observation having a missing value in column Y_j implies that it has a missing value in all columns Y_J where $J > j$ (Little & Rubin, 2014). In our case, what this looks like is that a voter cannot have a third choice for mayor without having a second choice present[7]. There are additional methods that can be used with monotone data, on which we do not go into greater detail here. These are described at more length in Rubin (1974).

4.4.3 Not all undervotes are created equal

In the imputation process, we fill in all undervotes[8] as if they are treated equally. This is not the case, however, as the undervote of somebody who listed Mark Leno first is inherently not the same as that of somebody who listed Jane Kim first, because she was eliminated in the tabulation process. We must treat these equally, however, because under a given imputation we do not know which candidates will eventually be eliminated. This also maintains the spirit of our research question, "what if there were no undervotes".

4.4.4 100% turnout

Due to data limitations, in this study we did not truly ask "what if everybody voted", but instead "what if there were no overvotes or undervotes". This is a different question, both conceptually and in practice: the latter question does not involve inferring the preferences of total non-voters, for example. This is an intriguing question for further study, but in this particular research we were examining the phenomena of overvoting and undervoting, rather than nonvoting. Some of the literature above is particular to the idea of 100% turnout, but still has important insight into our research question.

---

[7]Or rather, we have enforced this condition on our data. We treat the presence of a third choice and the absence of a second choice in a given voter's ballot as an error, and re-align our data so that the unique choices provided by each voter are filled in "minimally" to satisfy this monotone condition. We consider such an enforcement to be reasonable, given our access to voter's true preferences being limited to their reporting as such on the ballot.

[8]And weight them, in quantitative analyses.

Chapter 5

Missing data methods

5.1 Imputation

In dealing with our missing data, we chose here to use multiple imputation for a handful of reasons. First, it had the best conceptal analog to "what if there were no overvotes or undervotes", and is thus an appropriate method for measuring the effect of said ballot phenomena. Second, it does not have the same distributional assumptions that come with methods like maximum likelihood estimation. There were six methods of dealing with the missing data that we used:

- Original cases (or, not dealing with it at all). This original election data is a baseline case to compare all methods applied to.

- Listwise deletion. This is one form of analog to "what if there were no overvotes or undervotes", but instead of using all of the available information on voters it removes incomplete information and leaves us with a less accurate measure. This does, however, give us a "full vote choice distribution" from the original ballots that we can use to compare to other methods. This full vote choice distribution lets us compare these methods at a more granular level than the simple parameter of "who won". For example, we can see what proportion of voters under our multinomial method below voted for the `Kim - Leno - Weiss` combination, compared to the 9.03% of full voters that had this combination in the original ballot.

- Random hot deck imputation (RHD) based on nothing: that is, each voter is assigned a truly random candidate (pulled from the existing distribution of that ranking slot) that is unique from their earlier vote choices. This ignores some of the conditional distribution information available in the data. For example, given the cross-endorsement between Jane Kim and Mark Leno, we might expect their voters to be more likely to support the other candidate as their second choice than an average voter. Randomly pulling a second choice to impute for a Kim voter would then underestimate the chance that Kim would have been chosen.

Table 5.1: Examination of full voter counts

1	2	3	Count	Proportion
JANE KIM	MARK LENO	AMY FARAH WEISS	15654	0.0903336
LONDON BREED	MARK LENO	JANE KIM	11937	0.0688841
MARK LENO	JANE KIM	LONDON BREED	11074	0.0639041
LONDON BREED	JANE KIM	MARK LENO	10848	0.0625999
JANE KIM	MARK LENO	LONDON BREED	10484	0.0604994
MARK LENO	LONDON BREED	JANE KIM	7705	0.0444628
LONDON BREED	MARK LENO	ANGELA ALIOTO	6188	0.0357087
MARK LENO	JANE KIM	ANGELA ALIOTO	5811	0.0335332
JANE KIM	LONDON BREED	MARK LENO	5761	0.0332447
JANE KIM	MARK LENO	ANGELA ALIOTO	4649	0.0268277

- RHD based on vote choice alone. Missing data points are matched to potential donors by their corresponding 1st vote choice when imputing the 2nd vote choice, and by their first two vote choices when imputing the 3rd vote choice. This solves some of the conditional information problems presented by the fully random imputation.

- RHD based on vote choice and precinct. This is similar to the imputation vote choice alone, but now donors also must be from the same precinct as the recipient. This method controls for which precinct the voter is from, which attempts to get at differences between precincts in candidate support.

- Multinomial logistic regression based on vote choice and demographic data. The demographic information given to each voter was the rates of demographic groups in their precinct[1], as described in Chapter 2. This is similar to a multinomial regression based on vote choice and precinct alone, but also includes more information about certain precincts being more demographically similar than others.

5.1.1 Issues

Underestimate of variance

As expressed in the previous chapter, single imputations underestimate variance because they place an unwarranted certainty upon the newly imputed data, which is then fed into a deterministic process to calculate an estimator. This is overcome through the use of multiple imputation.

---

[1]This leaves us somewhat open to the genetic fallacy of assuming that every person in a precinct is identical to the precinct as a whole. However, since we are not interpreting the coefficients of this model, this precinct demographic information can just be treated as more information to make an accurate predictive model. It is a way of quantifying how similar different precincts are to each other, rather than simply treating each one as a wholly independent entity.

There are a set of methods to calculate the mean and variance of estimators from multiple imputation, known as "Rubin's Rules" (Rubin, 1987). For a set of estimates $\{\hat{\theta}_i\}$ of a parameter θ and m the number of imputations, the pooled mean can be calculated

$$\bar{\theta} = \frac{1}{m} \sum_{i=1}^{m} \hat{\theta}_i,$$

that is the mean of the estimates.

For variance, we must consider two quantities. Within each imputation, any estimate will have a standard error $SE_i{}^2$. We can then calculate the overall variance *within* imputations,

$$V_W = \frac{1}{m} \sum_{i=1}^{m} SE_i^2,$$

the average squared standard error within an imputation. There is also variance *between* each dataset, which comes from uncertainty about what the true values are that we are attempting to impute. This is expressed as

$$V_B = \frac{\sum_{i=1}^{m}(\theta_i - \bar{\theta})^2}{m - 1},$$

that is the sample variance of the estimators. Then, the *total* variance of our estimate is

$$V_T = V_W + V_B \left(1 + \frac{1}{m}\right),$$

a combination of these two facets of variance[3]. This is the way in which multiple imputation adjusts for the decreased variance that comes with single imputation.

While our RCV algorithm does not produce typical sample parameter estimates, these rules are still useful for obtaining an accurate sense of variability in our quantitative estimates.

Imputation on an imputation

Since we impute the 3rd choice based on the first and second choices, this means we're imputing some rows based on the imputed (unobserved) 2nd choice. This can lead to further issues of certainty, because the further imputations treat the imputed 2nd choice as known. We also believe this is overcome through the use of multiple imputations.

Incomplete hot deck matching

Some combinations of vote choices (typically including unpopular candidates), or precinct and vote choices, are unique or not present in the data. This means that the hot deck method cannot match any donor to impute a value. When this is the

---

[2]If our estimate is a mean, this is the standard deviation of the vector divided by the square root of the sample size. If the estimate is a proportion $\hat{p}$, the standard error is $\sqrt{\hat{p}(1 - \hat{p})/n}$.

[3]The $\left(1 + \frac{1}{m}\right)$ term is a bias correction for the estimate of variance.

Table 5.2: Proportion of unfilled 3rd vote choices through imputation
steps

Imputation step	Proportion
Original	0.3106906
1st, 2nd choice & precinct	0.0057996
1st, 2nd choice	0.0000040
1st choice	0.0000000

case, we then fill in with the next-best method, that is: if there is no combination
of 1st and 2nd that matches the case we need to impute 3rd for, we fall back to just
matching 1st and imputing the value. Table 5.2 is an illustration of how often this
phenomenon happens. Even in the worst case scenario out of all the imputations (3rd
vote choice, imputed on 1st choice, 2nd choice, and precinct), only 0.580% of 3rd vote
choices are missing. Of the choices that were missing originally, even this method
is able to match 98.1% of them. Since this is a small problem, we feel it does not
negatively impact our results.

Choice of which imputations to perform

There are any number of alternative imputation methods we could have used to get
simulated elections. The ones here were chosen for a mix of applicability (not all
methods apply to cetegorical data as well as numerical data), feasibility, and our
subjective assessment of what would make an interesting comparison[4].

For all of these methods, we had to ensure that no duplicate votes were created in
the process[5]. For example, a multinomial model considering demographic information
alone (no vote choices) was considered, but did not work because it created duplicated
votes for a number of the voters.

5.2   Data preparation

The election data was transformed first, to remove any duplicated votes (e.g., a row
listing London Breed three times was reduced to its functional equivalent: listing
London Breed once) and "left-align" all ballots so that ballots with one missing value
always had it in slot 3, and ballots with two missing values always had them in slots 2
and 3.

The demographic data used is the same that was used to answer the previous
research question: precinct-level demographic rates, obtained through areal interpola-
tion, and joined to each voter by their precinct. This is the most granular method we
have to establish demographics for each voter, because the only identifying information

---

[4]E.g., the missing data methods progress from least information used (listwise deletion) to most
information used (all available vote choices and precinct demographic information).

[5]As this would represent an "undervote", not using as many unique candidates as allowed.

we have about any given voter is their precinct. As such, the multinomial model used is not considering a voter's personal demographics, but rather the demographics of their precinct at large. As mentioned earlier, this gives us a way to quantify certain precincts as being more or less "distant" demographically to each other, as opposed to treating them all fully independently.

5.3 Metrics for comparing methods

At the end of the day, the ultimate metric of any election-related experiment like this is the winner of the election. While we are foremostly concerned with this, other metrics should not be ignored for consideration as well. We also examine:

- London Breed's[6] vote share in the final round of counting. This metric gives us a quantitative assessment of how close an election was to changing its eventual winner. Here we ignore exhausted ballots and just consider the binary comparison between the final two candidates.

- Intermediate switches in rankings. The smaller absolute margins between candidates in earlier rounds of counting provide a potentially easier opportunity to change the ultimate result (who is elected). These have a cascading effect on later rounds of the election, so even a seemingly insignificant swap in an earlier round of the tabulation can impact the eventual winner.

- Changes in proportions of vote combinations. This makes use of the calculated proportions of each combination of vote choices from the listwise deletion mentioned earlier. This is an attempt at a quantitative assessment of these intermediate ranking swaps, to see how big the impact of the imputation was. If a method has a strangely high or low proportion of a certain vote choice combination compared to the original data (here equivalent to the listwise deletion), then that method may have had some serious impact on the election results.

- Number of exhausted voters by the final round. This is secondary in terms of electoral impact, but does illustrate the impact of undervotes on representation. One of the complaints against RCV is that it has large numbers of exhausted ballots by the final round of the election (depending on how many candidates run and how many candidates voters are allowed to rank). By removing the undervotes and overvotes, we can see how much of the ballot exhaustion phenomenon is attributable to the limit on voter rankings.

---

[6]Breed was the winner of the real-life election.

Chapter 6

Data imputation results

For context, Table 6.1 displays the original election tabulation[1].

6.1 Final winner

We see in Table 6.2 that London Breed wins consistently in all methods except the fully random hot deck imputation, in which Mark Leno was consistently elected. As this is the least "realistic" of any of the methods (since it uses the least available information to predict vote choice), we can consider it a less accurate measure of potential change. It does potentially indicate that Leno had more overall support than Breed (in terms of being listed anywhere on the ballot), just positioned in the wrong place for it to push him to victory. Perhaps under a different voting system such as Borda count, which awards points to candidates based on any ranking they appeared in[2], Leno would have won.

---

[1]This differs slightly from the official results in ("Ranked Choice Voting Results Table," 2018); specifically, our vote counts are consistently an undercount of the official results. We believe this differece is due to the omission of some ballots from the published data, particularly ballots with a scanning error that have to be hand-counted.

[2]Under Borda Count, a candidate would be given 3 points for a 1st ranking from a voter, 2 points from a 2nd ranking, and 1 point from a 3rd ranking ("Borda Count," n.d.).

Table 6.1: Original election results

Candidate	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7	Round 8
BREED	91918	91921	92039	92233	93363	96028	101335	113247
LENO	61276	61276	61392	61734	62675	63837	67902	110979
KIM	60644	60644	60735	61320	61727	63059	65549	NA
ALIOTO	17447	17447	17628	17810	19463	21609	NA	NA
ZHOU	9521	9521	9631	9758	10526	NA	NA	NA
GREENBERG	7016	7016	7093	7189	NA	NA	NA	NA
WEISS	1661	1661	1720	NA	NA	NA	NA	NA
BRAVO	890	890	NA	NA	NA	NA	NA	NA
ROGERS	3	NA	NA	NA	NA	NA	NA	NA
NA	3640	3640	3778	3972	6262	9483	19230	29790

Table 6.2: Election results count by method

Method	Candidate	First place finishes
Original ballot	BREED	1
Listwise deletion	BREED	1
Hot deck - random	LENO	360
Hot deck - vote choice	BREED	360
Hot deck - vote choice, precinct	BREED	360
Multinomial logistic regression	BREED	360

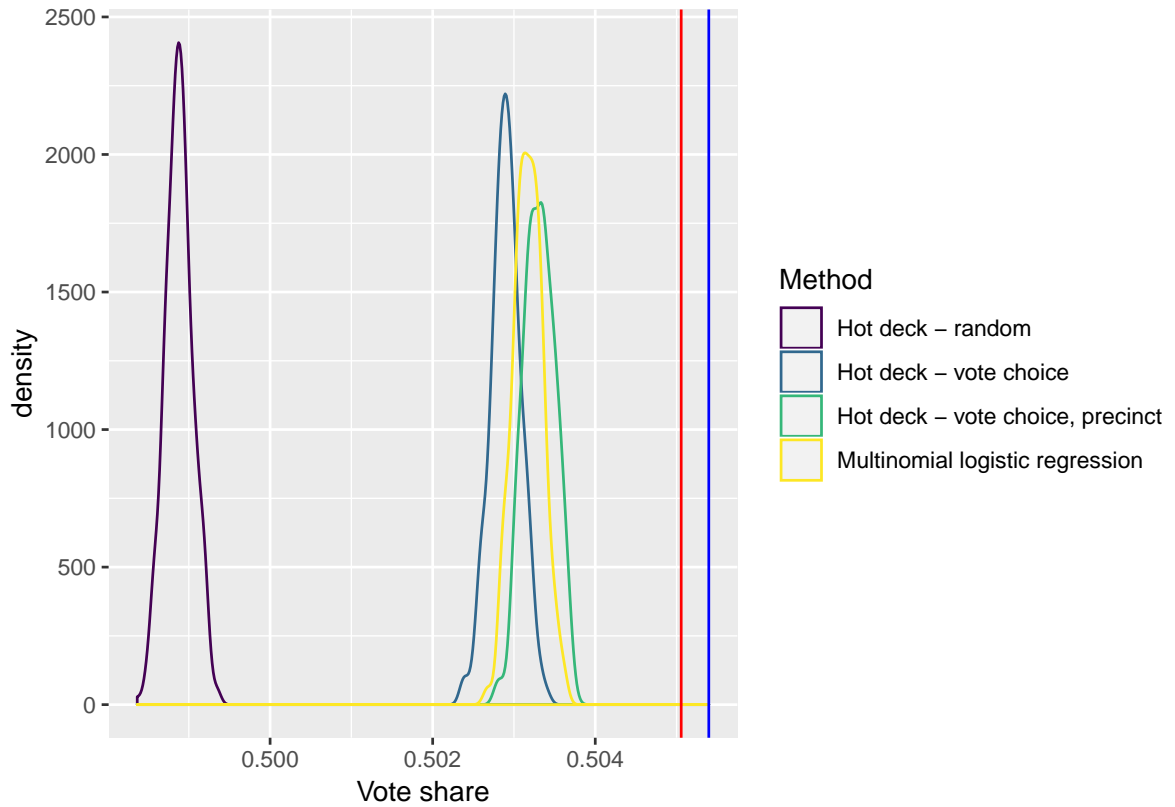## 6.2 London Breed's final vote share

Figure 6.1: London Breed's simulated vote share - unadjusted variances

Figure 6.1 is a density plot representing London Breed's estimated final vote share across the four imputation methods. The red vertical line is the original vote share, and the blue line is her share under listwise deletion. The imputation methods that use more information[3] indicate that Breed would likely still win if there were no undervotes or overvotes, but her margin of victory would be smaller than it was in the actual election. This could potentially offer some support for the claim that "the extent to which MI corrects the distribution [of vote share] is very limited, although the direction of adjustment is correct" (Liu, 2014).

For each imputation method, we have calculated the estimated mean and variance (using Rubin's rules) of London Breed's calculated final vote share[4]. One of the assumptions for using these rules is that the parameter estimates be normally distributed, which they appear to be from Figure 6.1. We assess that as follows:

---

[3]Everything except the fully random hot deck.

[4]Here a more typically specified question might be: out of all the voters who listed either London Breed or the other calculated finisher in the top two spots (Mark Leno or Jane Kim), what proportion preferred Breed to the other candidate?

Table 6.3: London Breed's simulated vote share - adjusted variances

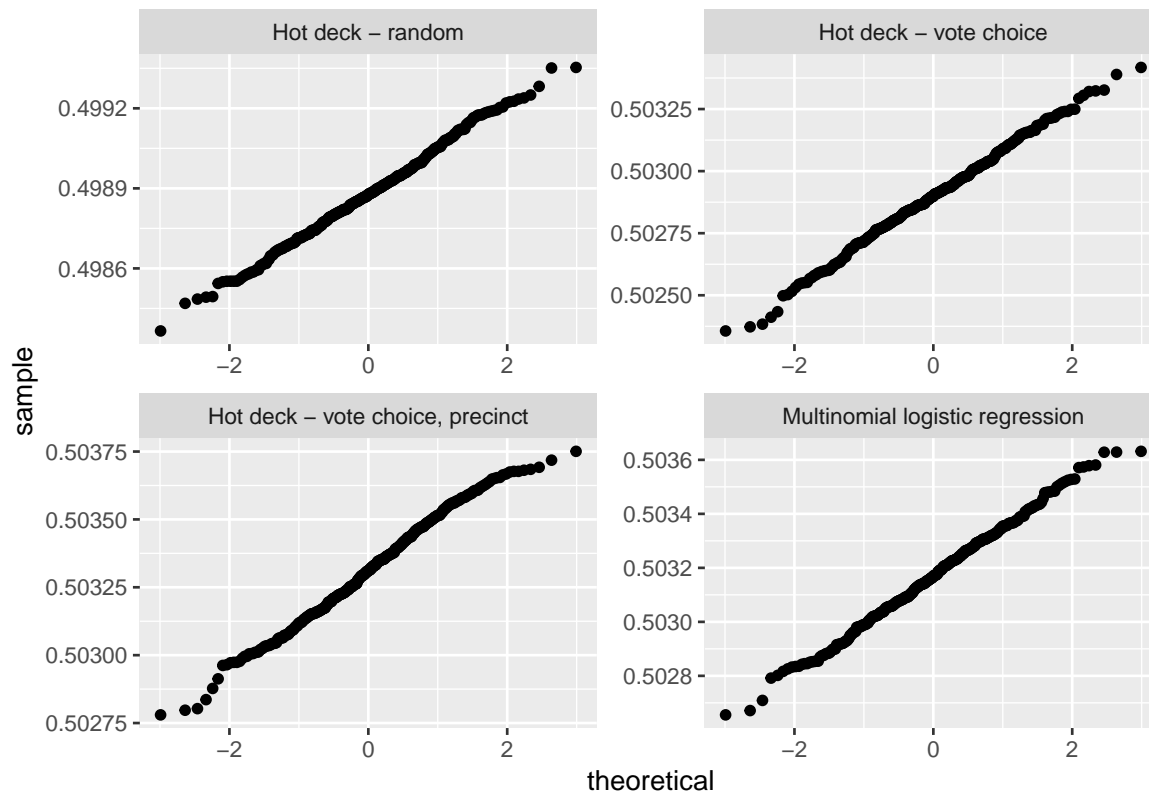| Method	$\hat{p}$	V_W	V_B	V_T
Original ballot	0.5050574	0.2499744	0	0.0000000
Listwise deletion	0.5053969	0.2499709	0	0.0000000
Hot deck - random	0.4988793	0.0006944	0	0.0006945
Hot deck - vote choice	0.5028990	0.0006944	0	0.0006945
Hot deck - vote choice, precinct	0.5033102	0.0006944	0	0.0006945
Multinomial logistic regression	0.5031699	0.0006944	0	0.0006944

Figure 6.2: Normality check, Q-Q plot of Breed's vote share

The Q-Q plots all follow a sufficiently linear path to assume normality in the data, so we meet the assumption to use Rubin's rules here. When we apply Rubin's rules to correct our underestimate of variance, the situation changes wildly[5]. Looking at Figure 6.3, the overlap between the various methods is much more palpable. Breaking this down in Table 6.3, it is clear that the majority of the variance in these estimated distributions comes from the variance *within* each imputation[6], as opposed to the

---

[5]Here the black and brown lines are the vote share for the original and listwise methods (respectively), which overlap at this scale.

[6]Given the high number of imputations we made (360), as opposed to the 5 that typically suffice for MI, this makes some sense.

Table 6.4: Candidate ranking for each imputation method

Method	Count	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7	Rank 8	Rank 9
Original ballot	1	BREED	LENO	KIM	ALIOTO	ZHOU	GREENBERG	WEISS	BRAVO	ROGERS
Listwise deletion	1	BREED	KIM	LENO	ALIOTO	ZHOU	GREENBERG	WEISS	BRAVO	ROGERS
Hot deck - random	360	LENO	BREED	KIM	ALIOTO	ZHOU	GREENBERG	WEISS	BRAVO	ROGERS
Hot deck - vote choice	360	BREED	LENO	KIM	ALIOTO	ZHOU	GREENBERG	WEISS	BRAVO	ROGERS
Hot deck - vote choice, precinct	360	BREED	LENO	KIM	ALIOTO	ZHOU	GREENBERG	WEISS	BRAVO	ROGERS
Multinomial logistic regression	360	BREED	LENO	KIM	ALIOTO	ZHOU	GREENBERG	WEISS	BRAVO	ROGERS

variance *between* the imputations[7]. This explains why the unadjusted variance in Figure 6.1 is so small: because it only considers the variance between each imputation.
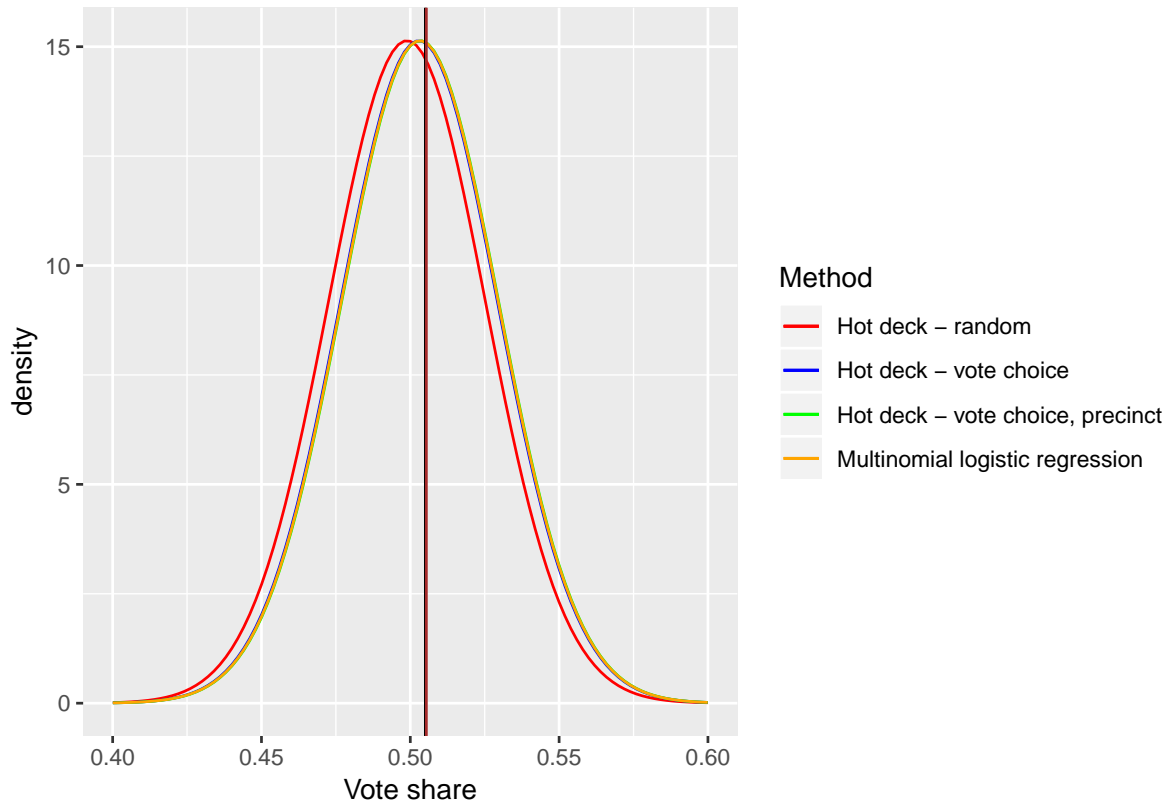

Figure 6.3: London Breed's simulated vote share - adjusted variances

6.3   Intermediate switches in rankings

Table 6.4 shows that each of the methods has the exact same candidate across all of the simulations run[8]. Under listwise deletion, we see that Jane Kim edged out Mark Leno as runner-up, but still failed to defeat Breed. Under the fully random

---

[7]This is not exactly zero across the board, as Table 6.3 would suggest, but is quite small. For the four multiple-run methods, the variance between is on the order of 10^{-8}.

[8]There were 360 simulations run for the imputation methods, and the two deterministic methods (original ballot data and listwise deletion) were each calculated once.

imputation, Leno took first place from Breed, but the remaining candidates are still arranged as they were in the original election.

We see no variability among these candidate rankings within each imputation method. While there were small changes in vote counts within methods between each simulation (see Section 6.4), this was not enough to change the eventual ordering of the candidates within any imputation method. However, between methods we do see some slight variability. As mentioned above, listwise deletion and fully random imputation are the least informed of the methods, so we consider their candidate rearrangements to be less accurate than the consistency in the more-informed methods.

6.4 Distribution of vote combinations

For each imputation method, we can compare the rate of appearance of a single vote combination to the baseline generated from the listwise deletion.
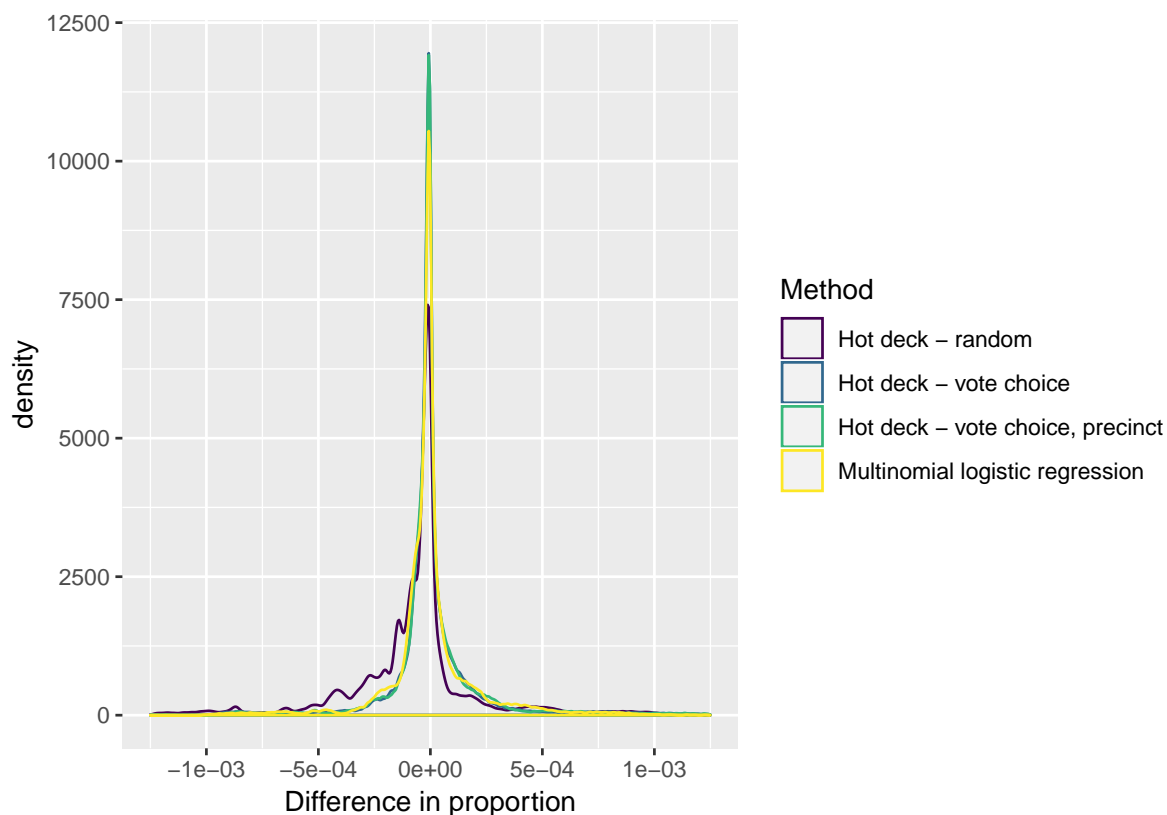

Figure 6.4: Difference in vote combination rates

Figure 6.4 is a density plot displaying the difference between the rate of a given vote combination under listwise deletion to its rate in each of the imputed methods. Overall, these changes in proportion from the original data are very small, which would seem to indicate that the imputation methods aren't actually changing the distribution of the data as a whole. This makes some sense, and lines up with our observations so far that the imputation methods (particularly the more informed

ones) don't impact the electoral results very much. Additionally, this highlights that these imputation methods suffer from the same "sampling issue" that techniques like bootstrapping have. Since we're trying to extrapolate from our data set, any inferences that we draw will be inherently centered around that observed data in some way.

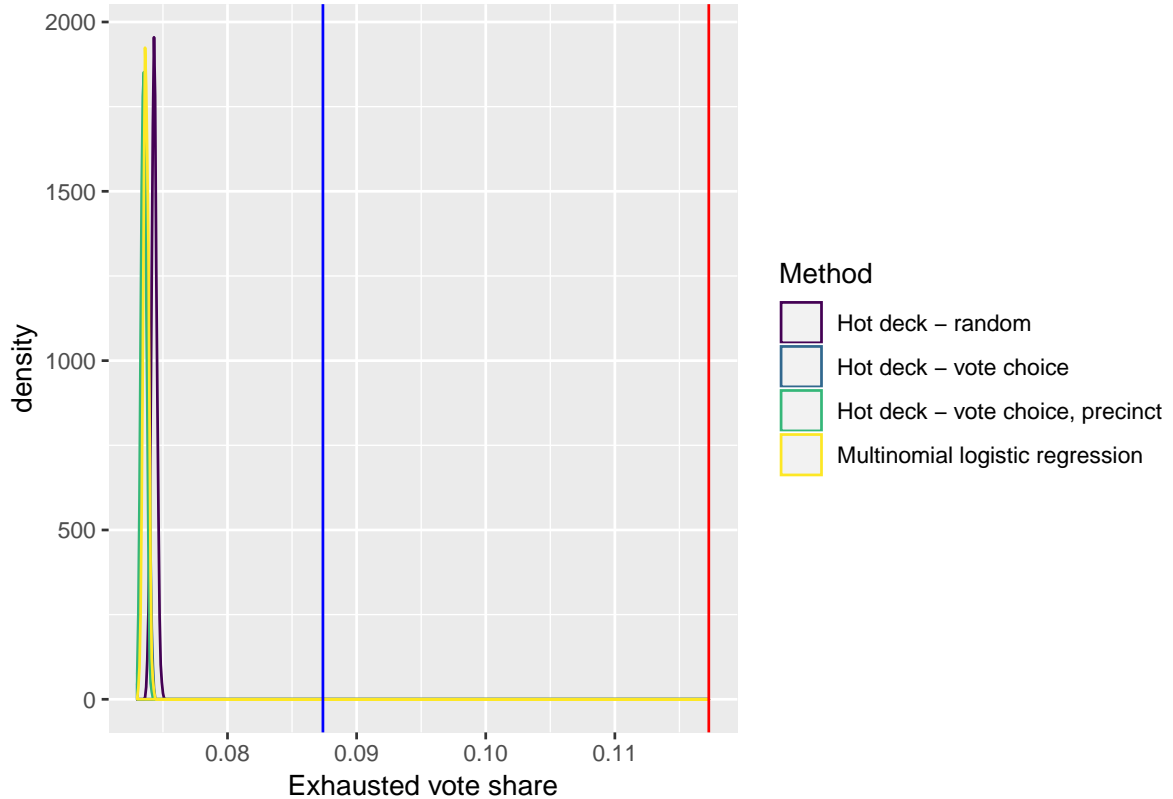## 6.5 Exhausted voters by the final round

Figure 6.5: Simulated exhausted vote share - unadjusted variances

Figure 6.5 is a density plot of the proportion of exhausted votes for each imputation method. The red line is exhausted votes in the original data, and the blue line is exhausted votes after listwise deletion. We see that the unadjusted variance for each method is small, compared to the overall decrease in exhausted votes from the original ballot data. The exhausted vote share in our imputed datasets is about 62.9% of the exhausted vote share in the original election. This indicates that about 0.4 of the exhausted data issue is "voluntary", where voters are not ranking as many candidates as they are given the option to.

We again apply Rubin's rules to calculate the estimated mean and variance of the final exhausted vote share[9]. Checking the normal assumption:

---

[9]Here a specified question might be: out of all the voters, how many did not support either of the top two finishers?

Table 6.5: Exhausted votes for each imputation method

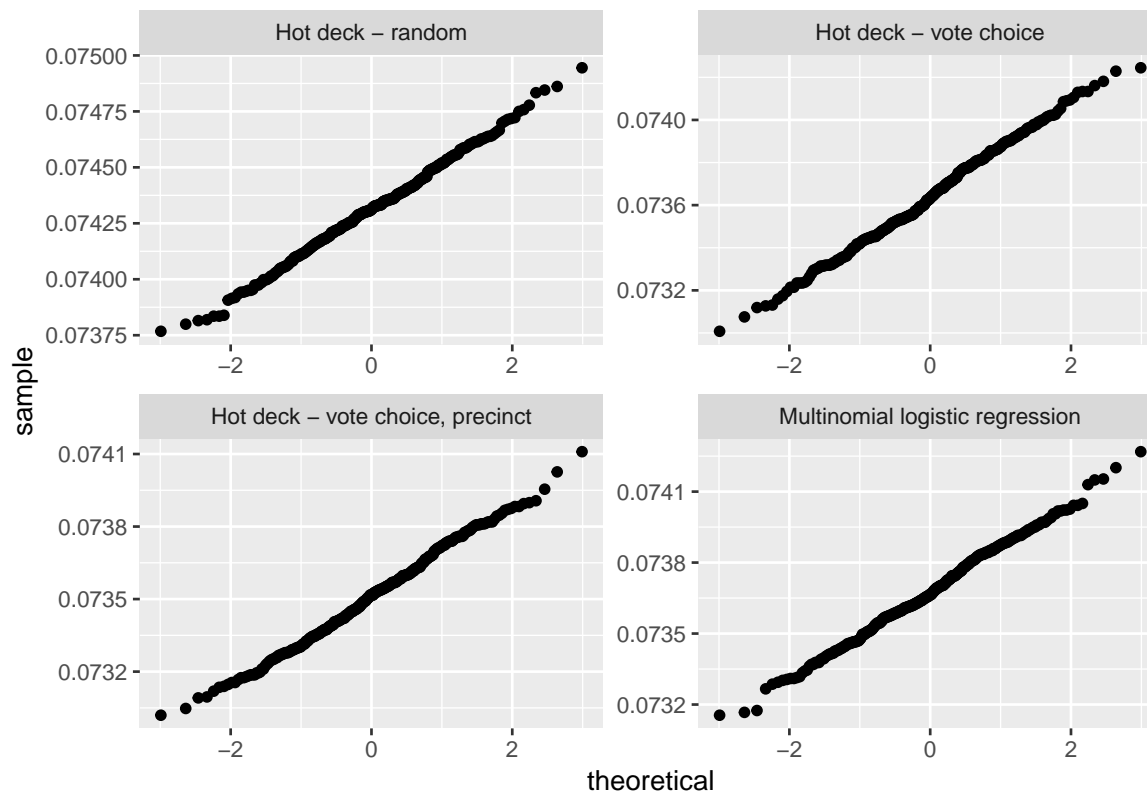| Method	$\hat{p}$	V_W	V_B	V_T
Original ballot	0.1172761	0.1035224	0e+00	0.0000000
Listwise deletion	0.0873906	0.0797535	0e+00	0.0000000
Hot deck - random	0.0743117	0.0001911	0e+00	0.0001911
Hot deck - vote choice	0.0736412	0.0001895	1e-07	0.0001895
Hot deck - vote choice, precinct	0.0735096	0.0001892	0e+00	0.0001892
Multinomial logistic regression	0.0736782	0.0001896	0e+00	0.0001896

Figure 6.6: Normality check, Q-Q plot of exhausted vote share

Since the Q-Q plots are sufficiently linear, we can assume that our parameter estimates for share of votes exhausted by the final round are normally distributed and proceed with Rubin's rules. When we apply Rubin's rules to correct our underestimate of variance, the situation changes drastically[10]. Looking at Figure 6.7, the overlap between the various methods is much more palpable. Similar to Breed's vote share, Table 6.5 shows that the majority of the variance in these estimated distributions of exhausted vote share comes from the variance *within* each imputation, as opposed to the variance *between* the imputations. In this case, however, the improvement in

---

[10]Here the black and brown lines are the exhausted vote share for the original and listwise methods (respectively).

exhausted vote share compared to the original data is so high that there is a much clearer impact from the imputation.
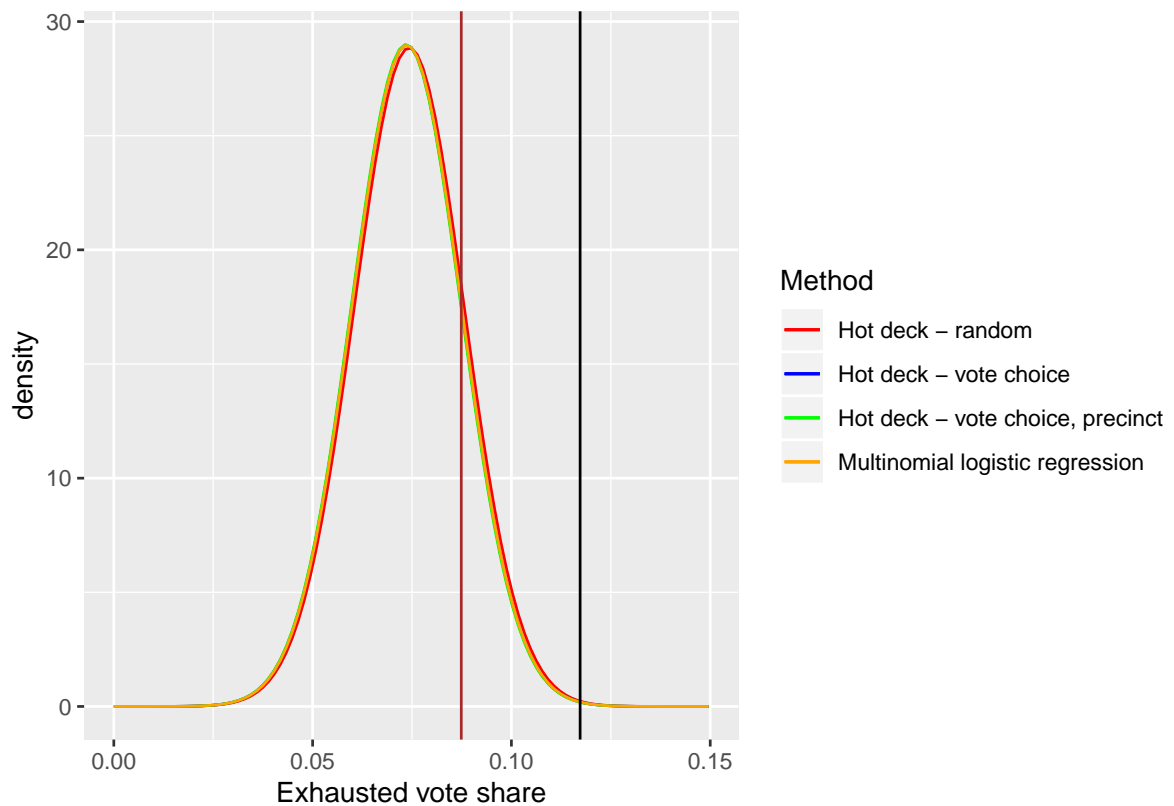

Figure 6.7: Simulated exhausted vote share - adjusted variances

Conclusion

6.6   Summary of results

6.6.1   Part 1

Part 1 left us with some interesting results. While not all of our methods gave conclusive interpretations, we saw that (on the whole) overvoting and undervoting were related to some similar variables as voter turnout is in general. The models for turnout and undervoting had relevant variables dealing with age, education and race, which are variables commonly linked to voting habits more generally in other research. Our turnout and undervoting models seemed to agree on the direction of these correlates as well. The coefficients provided by the overvoting model were clear that precincts with high number of African-American residents were more likely to overvote (and thus lose a number of votes that were attempted to be cast).

There were some weaknesses in our models, however. The poor fit and violated assumptions of the overvoting models indicate that these results may not be entirely accurate. Our undervoting model produced some unexpected relationships between demographic variables, which we could not interpret. A better tool than these regressions may be needed to parse further into this question.

We hoped that a more clear result here would be able to better inform voter education efforts, but do not feel confident in making any recommendation as such from this study. Given that the level of overvoting is so low across the board and only spikes in a few precincts, a geographically targeted education approach might be the best solution to this issue. Another course of action may be to continue voter education efforts in line with previous research on voter turnout more generally, as these same correlates are important to rates of overvoting and undervoting.

6.6.2   Part 2

In Part 2, we saw that the more informed MI methods did not change the election results, though they did sometimes decrease the winner's margin of victory. This is the same result obtained by Bernhagen & Marsh (2010) for a separate election. If our assumption that missing ballots are MAR is correct, this indicates that overvoting and undervoting did not impact the result of this election, though these phenomena did slightly help London Breed. When we correct for variance, our certainty that Breed always wins dramatically decreases. This is a weakness of the MI method as

applied to this data: because there is no "standard error" in vote share in an election (since it's not functioning as a survey), this adjusted variance does not necessarily reflect the actual variance in our estimate.

Given the disparity in undervote levels between different precincts (and the impact of demographic factors on these undervote levels), we expected one of the models with precinct or demographic to have more of an effect on the electoral results. We thought that if the precincts had different local preferences, then the now-higher weighting of underrepresented precincts might flip the election. One possible explanation as to why this did not occur is that this disparate precinct weighting did happen and it cancelled out between different precincts on the whole, leading to Breed's continued victory. Fairvote has produced some analysis that seems to indicate regional distinctions among candidates, which could support the claim about different local preferences ("SF Mayor's Race," n.d.-a; "SF Mayor's Race," n.d.-b; "SF Mayor's Race," n.d.-c; "SF Mayor's Race," n.d.-d). Additionally, as indicated in S. Hill & Hernandez (2018), the positioning of the undervotes themselves in this election may have limited the impact of full voting, and even under full voting Breed still would have won due to this positioning. Another answer is that MI itself may not produce enough variance to meaningfully affect these election results (even if there *were* a true change in the election winner), given the large number of data points and the inherent closed loop of using the observed data to modify itself[11]. This may be a limitation of MI, as after correcting variance with Rubin's rules there was much more potential overlap in election results (Chapter 6).

We also saw that almost half of the exhausted ballots in the final round of our original election can be explained by "voluntary" undervotes, that is voters not completing the ballot fully. This evidence complicates the claim that RCV does not produce true majorities, because the phenomenon of undervoting is not *only* a function of limited rankings imposed by the jurisdiction. Further research should be conducted into the effect of taking voluntary undervoting into account on how often we observe a non-majority in RCV elections.

6.7   Future research ideas

The most apparent next course of action in this resarch is to extend to more elections and jurisdictions that use RCV. This was not accomplished in this thesis due to issues with obtaining and merging the correct geographic boundaries to obtain precinct demographic information for more than one election cycle or more than one city. The advantage of this would be to extend how general the results are, and see if the demographic characteristics and imputation methods here have similar effects in different case studies. One consideration moving forward is that this missing data imputation method may not work well in Cambridge, where there are often far fewer available data points at later rankings in the ballot data[12] than we had in this San

---

[11]The same problem faced by methods like the bootstrap.

[12]Voters, perhaps experiencing some dizziness at the sheer number of candidates on the ballot, quite reasonably don't usually fill in choices all the way to rank 30 or so.

Francisco election.

An option to address this issue in Cambridge might be using some stochastic model (instead of hot deck imputation) to predict vote preferences. If a voter's state is the candidate they ranked in slot n, what are the "transition probabilities" of them ranking another given candidate in slot $n + 1$? This would not be a true Markov process because it would have to include some memory to ensure that no candidate appears twice for a given voter, but the Markov chain is a useful conceptual comparison.

Another desirable task would be to obtain more accurate demographic information about the election precincts. One of the limitations of this study is the inaccuracy in precinct demographic information propagated through the areal interpolation initially performed. Perhaps a more accurate interpolation could be calculated with additional ancillary data (e.g. street or zoning data, a more granular population density metric, etc.), to avoid relying on the uniform spatial population distribution assumption that we used here. Alternatively, we could obtain better demographic information about the precincts directly by examining the voter registration file for a given jurisdiction. The information carried in these files varies by state[13], but they generally include age and whether someone voted in a given election. That would give us more detailed demographic information about the people who voted in the election, as opposed to general precinct demographics. As with any data containing personally identifiable information, care must be taken to maintain privacy if such a file is used.

In future studies, a better model specification could be used to examine which variables are correlated with overvoting and undervoting. Notably, because overvoting is so rare, we see a significant number of precincts with 0 overvotes. A zero-inflated logistic model could be applied to further enhance the conclusions presented.

Finally, we could additionally expand the research question to include 100% voter turnout, along with no overvotes or undervotes. That would require getting an accurate count of the voting-eligible population (VEP) or voting-age population (VAP) within each precinct. There would then be no information for some voters about 1st candidate supported, so precinct and demographic information alone would need to be used to predict vote choices. Such a study would be less precise in conclusions than this one, but is in turn a more broadly applicable research question. We expect that this would have some precinct-level turnout jumps that could greater impact the election result; however, this hypothesis should be considered skeptically, as we theorized the same result would appear in this study.

---

[13]Race and gender are only collected in certain states, for example.

Appendix A

Description of variables

Table A.1: Census variables used in models

Variable	Description
female	The percentage of the population that is female
pop_18_24	The percentage of the population that is between 18 and 24 years old
pop_25_44	The percentage of the population that is between 25 and 44 years old
pop_45_64	The percentage of the population that is between 45 and 64 years old
pop_65_up	The percentage of the population that is 65 years old or over
hispanic	The percentage of the population that identify as "Mexican", "Puerto Rican", "Cuban", or "another Hispanic, Latino, or Spanish origin"
white	The percentage of the population that indicate no Hispanic origin and their only race as "White" or report entries such as Irish, German, Italian, Lebanese, Arab, Moroccan, or Caucasian
black	The percentage of the population that indicate no Hispanic origin and their only race as "Black, African Am., or Negro" or report entries such as African American, Kenyan, Nigerian, or Haitian

Variable	Description
native	The percentage of the population that indicate no Hispanic origin and their only race as "American Indian or Alaska Native" or report entries such as Navajo, Blackfeet, Inupiat, Yup'ik, or Central/South American Indian groups
asian	The percentage of the population that indicate no Hispanic origin and their only race as "Asian Indian", "Chinese", "Filipino", "Korean", "Japanese", "Vietnamese", or "Other Asian"
pac_islander	The percentage of the population that indicate no Hispanic origin and their only race as "Native Hawaiian", "Guamanian or Chamorro", "Samoan", or "Other Pacific Islander"
other_race	The percentage of the population that indicate no Hispanic origin and their race as other than "White", "Hispanic", "Black or African American", "American Indian or Alaska Native", "Asian", and "Native Hawaiian or Other Pacific Islander"
no_hs	The percentage of the population aged 25 years and over that are not high school graduates and have not received a diploma or the equivalent
college	The percentage of the ACS population aged 25 years and over that have a college degree or higher
poverty	The percentage of the eligible population[a] that are classified as below the poverty level given their total family or household income within the last year, family size, and family composition

[a]The "eligible population" is measured by the Census as "Number of people excluding institutionalized people, people in military group quarters, people in college dormitories, and unrelated individuals under 15 years old"

Variable	Description
no_english	The percentage of all occupied housing units[a] where no one ages 14 years and over speaks English only or speaks English "very well"

———————————

[a]This variable is at the level of housing units (not persons), and the Census measures "occupied housing units" as "Number of housing units classified as usual place of residence of the individual or group living in it"

References

2000 Presidential General Election Results. (2001, December). *Federal Election Commission*. Retrieved from https://transition.fec.gov/pubrec/2000presgeresults.htm

2010 TIGER/Line Shapefiles. (n.d.). *United States Census Bureau*. Retrieved from https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2010/&layergroup=Block+Groups

Allen, B., & Hertzberg, B. (2016, September). Elections: Vote by mail voting and mail ballot elections.

Allison, P. D. (2011). Missing Data. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE Handbook of Quantitative Methods in Psychology.*

Amunategui, M. (n.d.). Predicting Multiple Discrete Values with Multinomials, Neural Networks and the {nnet} Package. Retrieved from https://amunategui.github.io/multinomial-neuralnetworks-walkthrough/index.html

Arrow, K. J. (1963). *Social Choice and Individual Values* (2nd ed.). New Haven; London: Yale University Press.

Arrow, K. J., Sen, A. K., & Suzumura, K. (Eds.). (2002). *Handbook of social choice and welfare* (Vol. 1). Amsterdam ; Boston: Elsevier.

Bartholdi, J. J., & Orlin, J. B. (1991). Single transferable vote resists strategic voting. *Social Choice and Welfare*, *8*(4). http://doi.org/10.1007/BF00183045

Berman, R. (2018, May). Rival Candidates Try an Unusual Election Message: Vote for Both of Us. *The Atlantic*. Retrieved from https://www.theatlantic.com/politics/archive/2018/05/san-francisco-mayor-jane-kim-mark-leno-ranked-choice-voting/561053/

Bernhagen, P., & Marsh, M. (2010). Missing Voters, Missing Data: Using Multiple Imputation to Estimate the Effects of Low Turnout. *Journal of Elections, Public Opinion, and Parties.*

Borda Count. (n.d.). *Electoral Reform Society*. Retrieved from https://www.electoral-reform.org.uk/voting-systems/types-of-voting-system/borda-count/

Boyd, R. W. (1986). Election Calendars and Voter Turnout. *American Politics*

Quarterly, *14*(1-2), 89–104. http://doi.org/10.1177/1532673X8601400106

Burnett, C. M., & Kogan, V. (2015). Ballot (and voter) "exhaustion" under Instant Runoff Voting: An examination of four ranked-choice elections. *Electoral Studies*, *37*, 41–49. http://doi.org/10.1016/j.electstud.2014.11.006

Callaghan, P. (2017, October). Has ranked-choice voting lived up to its promise in the Twin Cities? *MinnPost*. Retrieved from https://www.minnpost.com/politics-policy/2017/10/has-ranked-choice-voting-lived-its-promise-twin-cities/

Citrin, J., Schickler, E., & Sides, J. (2003). What If Everyone Voted? Simulating the Impact of Increased Turnout in Senate Elections. *American Journal of Political Science*, *47*(1), 75–90. http://doi.org/10.2307/3186094

Comments on Proposed Rules for RCV. (n.d.). *League of Women Voters of Maine*. Retrieved from http://www.lwvme.org/files/Comments/_on/_Proposed/_Rules/_for/_RCV.pdf

Condorcet Method. (2008, August). *Princeton University Department of Mathematics*. Retrieved from http://web.math.princeton.edu/math/_alive/Voting/Lab1/Condorcet.html

Cook, C. (2011, December). Ranked Choice Voting. *SPUR*. Retrieved from https://www.spur.org/publications/urbanist-article/2011-12-01/ranked-choice-voting

Dennis, G. (2004). Ballot-Marking Errors in the First San Francisco Instant Runoff Election, 15.

Do, V. H., Thomas-Agnan, C., & Vanhems, A. (2014). *Spatial reallocation of areal data - another look at basic methods Réaffectation spatiale de données surfaciques - une autre approche des méthodes de base* (p. 32).

Donovan, T., Tolbert, C., & Gracey, K. (2016). Campaign civility under preferential and plurality voting. *Electoral Studies*, *42*, 157–163. http://doi.org/10.1016/j.electstud.2016.02.009

Douglas, A. (2013, November). Cambridge, Massachusetts Elections a Model for America. *FairVote*. Retrieved from https://www.fairvote.org/cambridge-massachusetts-elections-a-model-for-america

Eberhard, K. (2017, September). What Really Happened With Instant Runoff Voting in Pierce County, Washington? *Sightline Institute*.

Electoral Systems: Contingent and Alternative Voting. (2013, September). *UK Engage*.

EPSG:7132. (n.d.). *EPSG.io*. Retrieved from https://epsg.io/7132

Fowler, M. (n.d.). *Chapter 3: Cramer-Rao Lower Bound*. Binghamton University.

Funding Elections Technology. (2019, February). *National Conference of State Legis-*

latures. Retrieved from http://www.ncsl.org/research/elections-and-campaigns/funding-election-technology.aspx

Hawkins, L. D. (2011, June). Major Legal Victory for Ranked Choice Voting and Reform. *FairVote.* Retrieved from https://www.fairvote.org/major-legal-victory-for-ranked-choice-voting-and-reform

Henry, M. A. (n.d.). *THE IMPLEMENTATION AND EFFECTS OF RANKED CHOICE VOTING IN CALIFORNIA CITIES* (PhD thesis). California State University, Sacramento.

Herron, M. C., & Lewis, J. B. (2007). Did Ralph Nader spoil Al Gore's presidential bid? A ballot-level study of green and reform party voters in the 2000 presidential election*. *Quarterly Journal of Political Science*, *2*(3), 205.

Hill, C., Jerdonek, C., & Mogi, V. (2018, June). SF elections are working - and getting even better. *The San Francisco Examiner.* Retrieved from http://www.sfexaminer.com/sf-elections-working-getting-even-better/

Hill, S. (2007). Instant Runoff Voting. In *Ten Big Ideas for a New America.* New America Foundation.

Hill, S., & Hernandez, P. (2018, July). Exhausted Ballots in the 2018 San Francisco Mayoral Election. *FairVote California.*

Ismay, C., & Solomon, N. (n.d.). Thesisdown.

Jerdonek, C. (2006). *Ranked Choice Voting and Voter Turnout in San Francisco's 2005 Election* (p. 10). FairVote.

John, S., Smith, H., & Zack, E. (2018). The alternative vote: Do changes in single-member voting systems affect descriptive representation of women and minorities? *Electoral Studies*, *54*, 90–102. http://doi.org/10.1016/j.electstud.2018.05.009

Kimball, D. C., & Anthony, J. (2016). *Voter Participation with Ranked Choice Voting in the United States.*

Kimball, J. (2010, May). Estimated cost of Ranked Choice Voting in Minneapolis: $365,000. *MinnPost.* Retrieved from https://www.minnpost.com/political-agenda/2010/05/estimated-cost-ranked-choice-voting-minneapolis-365000/

Kroh, M. (2006). Taking "Don't Knows" as Valid Responses: A Multiple Complete Random Imputation of Missing Data. *Quality & Quantity*, *40*(2), 225–244. http://doi.org/10.1007/s11135-005-5360-3

Kukura, J. (2018, May). Leno and Kim Endorse Each Other for Mayor - May 10, 2018. *SF Weekly.* Retrieved from http://www.sfweekly.com/topstories/leno-and-kim-endorse-each-other-for-mayor/

Lee, J., & Yancheff, M. (2019, April). Rcv: Ranked Choice Voting.

Levi, R. (2018, November). Why Is Part of Alameda Island in San Francisco? *KQED*. Retrieved from https://www.kqed.org/news/11702058/why-is-part-of-alameda-island-in-san-francisco

Lijphart, A., & Grofman, B. (1984). *Choosing an Electoral System: Issues and Alternatives.* (G. M. Pomper, Ed.). New York, NY: Praeger Publishers.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* New York, NY: John Wiley & Sons, Inc.

Little, R. J. A., & Rubin, D. B. (2014). Introduction. In *Statistical Analysis with Missing Data* (pp. 1–23). Hoboken, NJ, USA: John Wiley & Sons, Inc. http://doi.org/10.1002/9781119013563.ch1

Liu, F. C. S. (2014). Using Multiple Imputation for Vote Choice Data: A Comparison across Multiple Imputation Tools. *Open Journal of Political Science*, *04*(02), 39–46. http://doi.org/10.4236/ojps.2014.42006

Maps. (n.d.). *San Francisco Department of Elections.* Retrieved from https://sfelections.sfgov.org/maps

McDaniel, J. (2016). Ranked Choice Voting Likely Means Lower Turnout, More Errors. *Cato Unbound*, (Should We Choose Ranked Choice Voting?).

McDaniel, J. (2018). Does More Choice Lead to Reduced Racially Polarized Voting? Assessing the Impact of Ranked-Choice Voting in Mayoral Elections. *California Journal of Politics and Policy*, *10*(2). http://doi.org/10.5070/P2CJPP10241252

Mcdaniel, J. A. (2016). Writing the Rules to Rank the Candidates: Examining the Impact of Instant-Runoff Voting on Racial Group Turnout in San Francisco Mayoral Elections. *Journal of Urban Affairs*, *38*(3), 387–408. http://doi.org/10.1111/juaf.12209

Meckler, L. (2000, October). GOP Group To Air Pro-Nader TV Ads. *Washington Post.* Retrieved from https://www.washingtonpost.com/wp-srv/aponline/20001027/aponline115918/_000.htm

Miller, K., & Thistle, S. (2018, November). Jared Golden declared winner of first ranked-choice congressional election, but challenge looms. *Portland Press-Herald.* Retrieved from https://www.pressherald.com/2018/11/15/final-ranked-choice-vote-count-slated-for-noon/

Mistler, S. (2018, November). In Tight Race, Maine Republican Sues To Block State's Ranked-Choice Voting Law. *NPR.org.* Retrieved from https://www.npr.org/2018/11/13/667435326/facing-defeat-maine-republican-sues-to-block-states-ranked-choice-voting-law

Morales, M. (2018, March). Better Voting Systems Boost Turnout. *Sightline Institute.*

Morales, M., & Eberhard, K. (2017, July). An Action Plan for Ranked-Choice-Ready

Voting Equipment. *Sightline Institute.*

Multinomial Logistic Regression | R Data Analysis Examples. (n.d.). *UCLA Institute for Digital Research & Education.* Retrieved from https://stats.idre.ucla.edu/r/dae/multinomial-logistic-regression/

Murphy, D. E. (2004). New Runoff System in San Francisco Has the Rival Candidates Cooperating. *The New York Times.*

Neely, F., & Cook, C. (2008). Whose Votes Count?: Undervotes, Overvotes, and Ranking in San Francisco's Instant-Runoff Elections. *American Politics Research, 36*(4), 530–554. http://doi.org/10.1177/1532673X08318110

Neely, F., & McDaniel, J. (2015). Overvoting and the Equality of Voice under Instant-Runoff Voting in San Francisco. *California Journal of Politics and Policy; Berkeley, 7*(4), 0_1, 0_2, 1–27. http://doi.org/http://dx.doi.org/10.5070/P2cjpp7428929

Nikulin, M. S. (n.d.). Unbiased estimator. *Encyclopedia of Mathematics.*

Pebesma, E., Bivand, R., Racine, E., Sumner, M., Cook, I., Keitt, T., . . . Pedersen, T. L. (n.d.). Sf: Simple Features for R. Retrieved from r-spatial

Petrangelo, T. (2013, October). Does ranked-choice voting guarantee a majority winner? *MinnPost.* Retrieved from https://www.minnpost.com/minnesota-blog-cabin/2013/10/does-ranked-choice-voting-guarantee-majority-winner/

Ranked-Choice Voting: Ballot Image. (2018a, June). *San Francisco Department of Elections.* Retrieved from https://sfelections.org/results/20180605/data/20180627/20180627/_ballotimage.txt

Ranked Choice Voting in Maine. (2019, January). *Maine State Legislature.* Retrieved from http://legislature.maine.gov/lawlibrary/ranked-choice-voting-in-maine/9509

Ranked-Choice Voting: Master Lookup. (2018b, June). *San Francisco Department of Elections.* Retrieved from https://sfelections.org/results/20180605/data/20180627/20180627/_masterlookup.txt

Ranked Choice Voting Results Table. (2018, June). *San Francisco Board of Elections.* Retrieved from https://sfelections.org/results/20180605/data/20180627/mayor/20180627/_mayor.html

Ranney, A. (1972). Turnout and Representation in Presidential Primary Elections. *The American Political Science Review, 66*(1), 21–37. http://doi.org/10.2307/1959276

R Core Team. (2018). R: A Language and Environment for Statistical Computing. VIenna, Austria. Retrieved from R%20Foundation%20for%20Statistical%20Computing

Richie, R., & Brown, M. (2017, December). Ranked choice voting outperforms runoffs in upholding majority rule. *FairVote.* Retrieved from https://www.fairvote.org/ranked/_choice/_voting/_outperforms/_runoffs/

_in/_upholding/_majority/_rule

Rubin, D. B. (1974). Characterizing the Estimation of Parameters in Incomplete-Data Problems. *Journal of the American Statistical Association*, *69*(346), 467–474. http://doi.org/10.2307/2285680

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons.

Rush, J. (2015, June). Cost of runoff election affects more than just candidates. *Star Local*. Retrieved from https://starlocalmedia.com/planostarcourier/news/cost-of-runoff-election-affects-more-than-just-candidates/article/_bac2a0f2-ffaa-524c-9b87-6704f60e4468.html

Santucci, J. (2017). Party Splits, Not Progressives: The Origins of Proportional Representation in American Local Government. *American Politics Research*, *45*(3), 494–526. http://doi.org/10.1177/1532673X16674774

Schroeder, J. P. (2007). Target-Density Weighting Interpolation and Uncertainty Evaluation for Temporal Analysis of Census Data. *Geographical Analysis*, *39*(3), 311–335. http://doi.org/10.1111/j.1538-4632.2007.00706.x

Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, *22*(3), 278–295. http://doi.org/10.1177/0962280210395740

SF Mayor's Race: Breed First Round %. (n.d.-a). *FairVote*. Retrieved from https://d3n8a8pro7vhmx.cloudfront.net/fairvote/pages/10829/attachments/original/1533072759/BreedFirstChoicesbyPrecinct.png?1533072759

SF Mayor's Race: Breed % of Vote in Final Round. (n.d.-b). *FairVote*. Retrieved from https://d3n8a8pro7vhmx.cloudfront.net/fairvote/pages/10829/attachments/original/1533067132/BreedandLenoFinalRoundbyPrecinct.png?1533067132

SF Mayor's Race: Kim First Round %. (n.d.-c). *FairVote*. Retrieved from https://d3n8a8pro7vhmx.cloudfront.net/fairvote/pages/10829/attachments/original/1533072757/KimFirstChoicebyPrecinct.png?1533072757

SF Mayor's Race: Leno First Round %. (n.d.-d). *FairVote*. Retrieved from https://d3n8a8pro7vhmx.cloudfront.net/fairvote/pages/10829/attachments/original/1533072758/LenoFirstChoicebyPrecinct.png?1533072758

The Associated Press. (2014, July). Runoff election Tuesday will cost Alabama $3 million. *al.com*. Retrieved from https://www.al.com/news/2014/07/runoff/_election/_tuesday/_will/_c.html

Tidyverse. (n.d.). Retrieved from https://www.tidyverse.org/

US Census Bureau, A. S. D. (2018, June). 2018 Planning Database. *United States Census Bureau*. Retrieved from https://www.census.gov/research/data/planning/

_database/2018/

Vanderloo, M. (2017, January). Simputation: Simple Imputation.

Where is Ranked Choice Voting Used? (n.d.). *FairVote*. Retrieved from https://www.fairvote.org/where/_is/_ranked/_choice/_voting/_used

Wilson, R. (2014, June). Runoff elections a relic of the Democratic South. *Washington Post*. Retrieved from https://www.washingtonpost.com/blogs/govbeat/wp/2014/06/04/runoff-elections-a-relic-of-the-democratic-south/

Woodard, C. (2018, April). How ranked-choice voting effort became a partisan flash point. *Lewiston Sun Journal*. Retrieved from https://www.sunjournal.com/2018/04/07/stark-divide-on-ranked-choice-voting/

Wright, S. G. (1989). Voter Turnout in Runoff Elections. *The Journal of Politics*, *51*(2), 385–396. http://doi.org/10.2307/2131348