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).

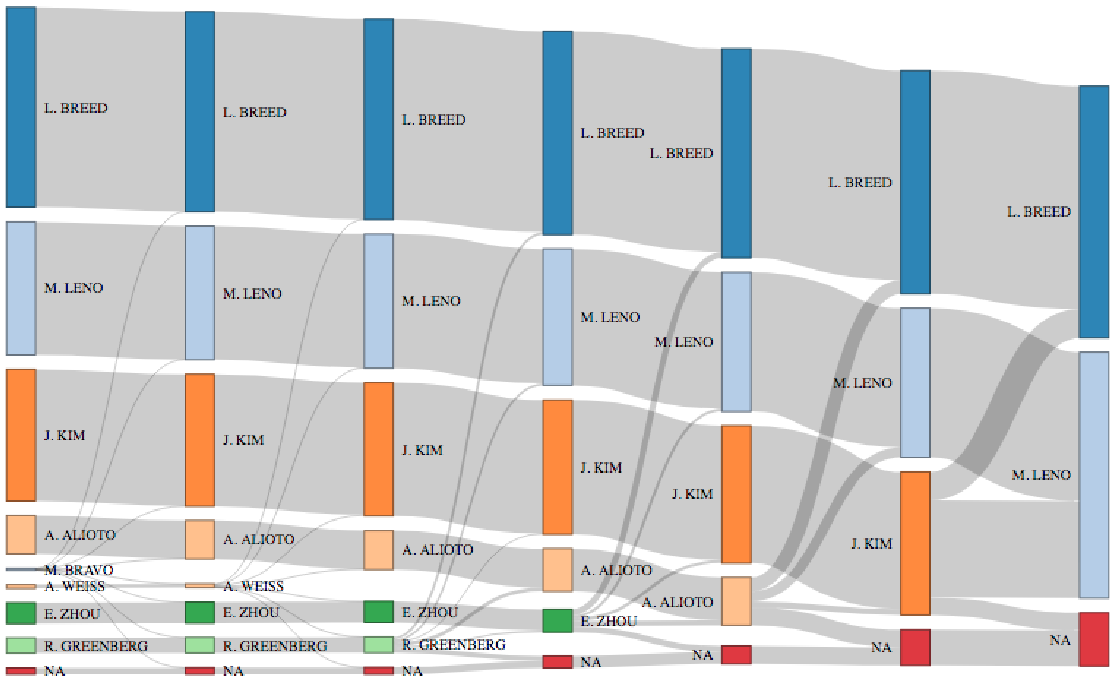


Figure 1.1: Sankey diagram of vote tabulation

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.

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).

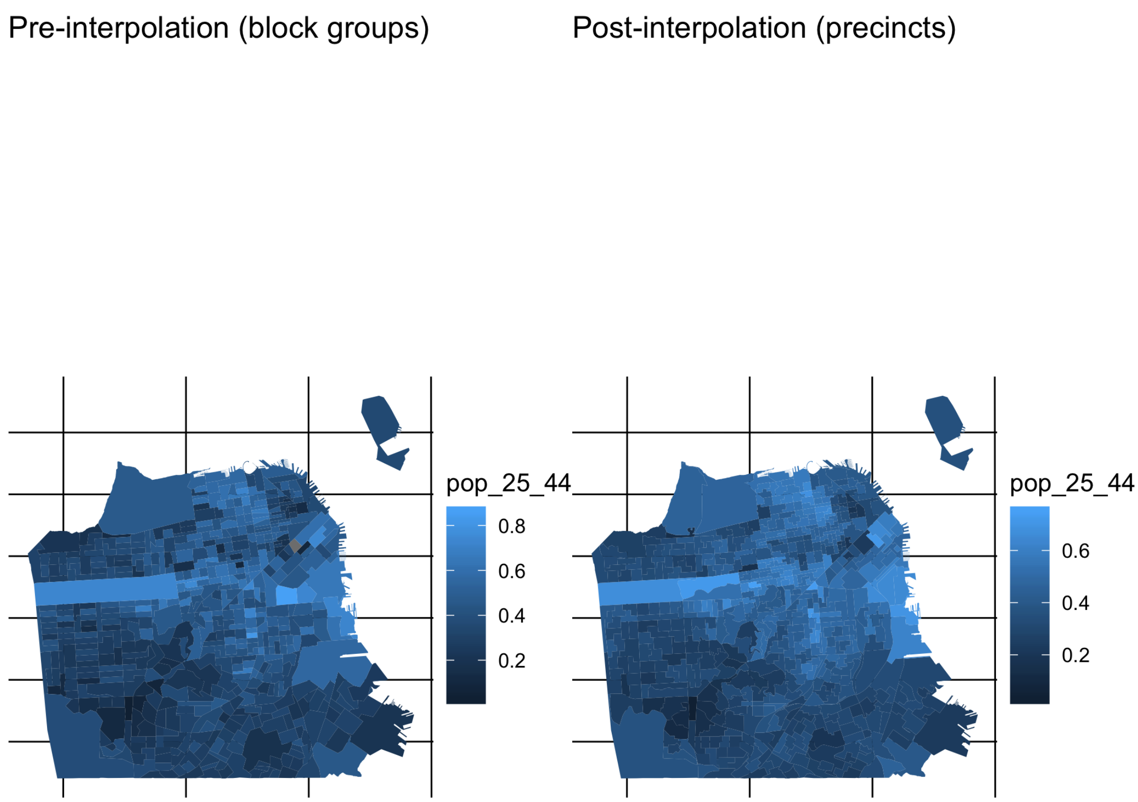


Figure 2.5: Proportion of residents aged 25-44

### 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).

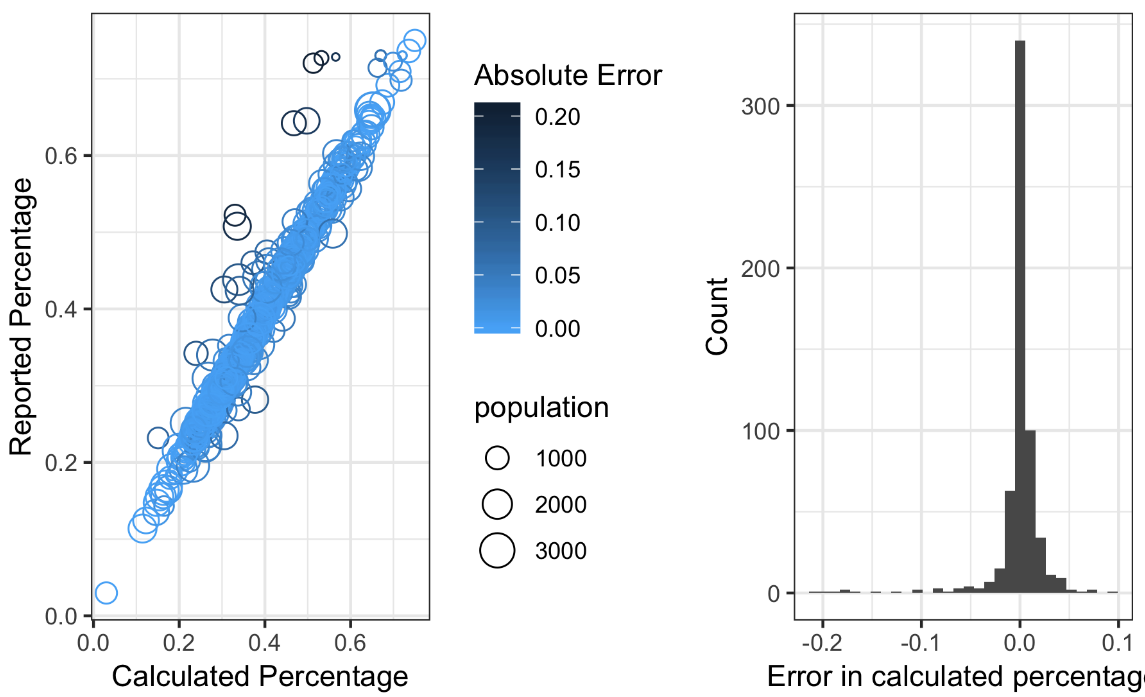


Figure 2.6: Sample error in intensive interpolation - Percentage aged 25-44

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

# 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.

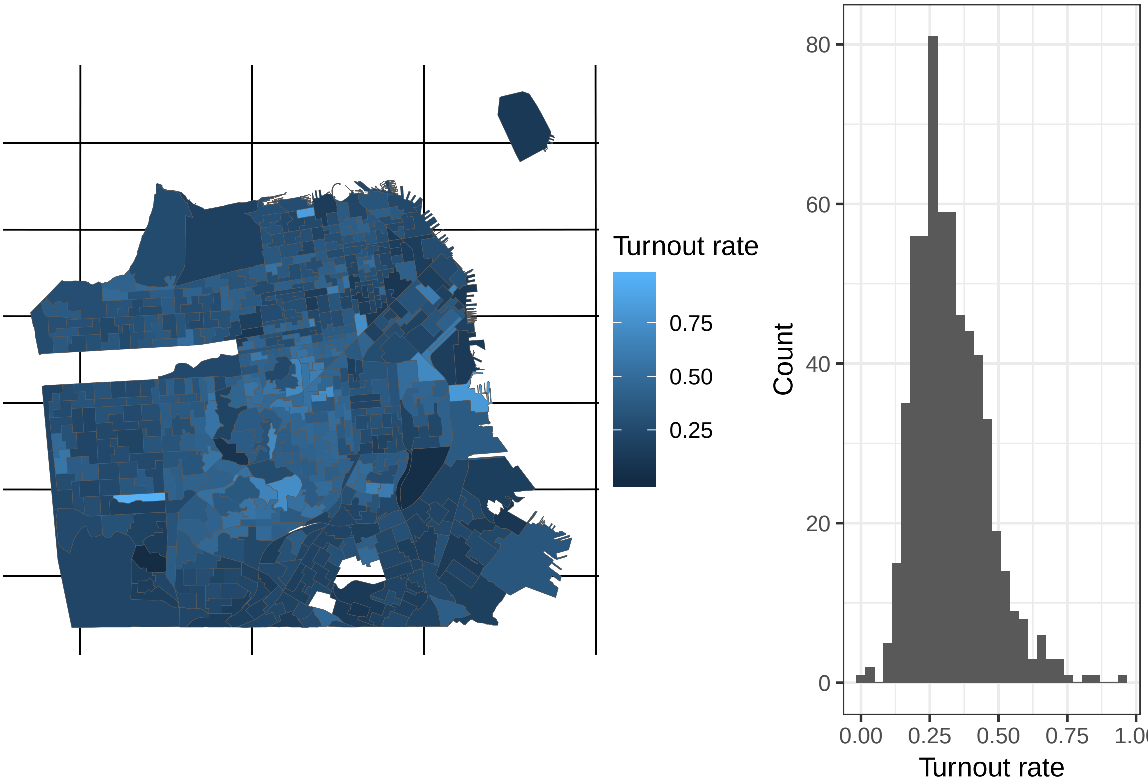


Figure 3.1: Observed turnout rate by precinct

### 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.

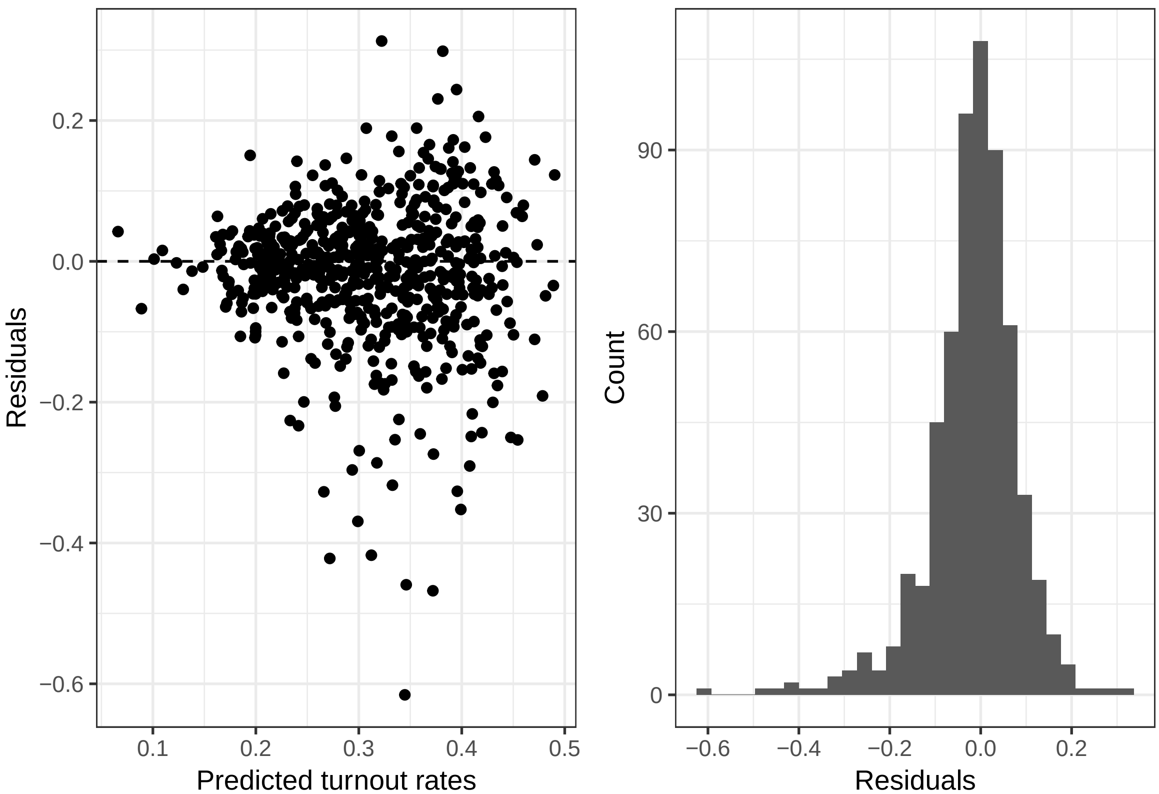


Figure 3.3: Logistic turnout model validation

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 |

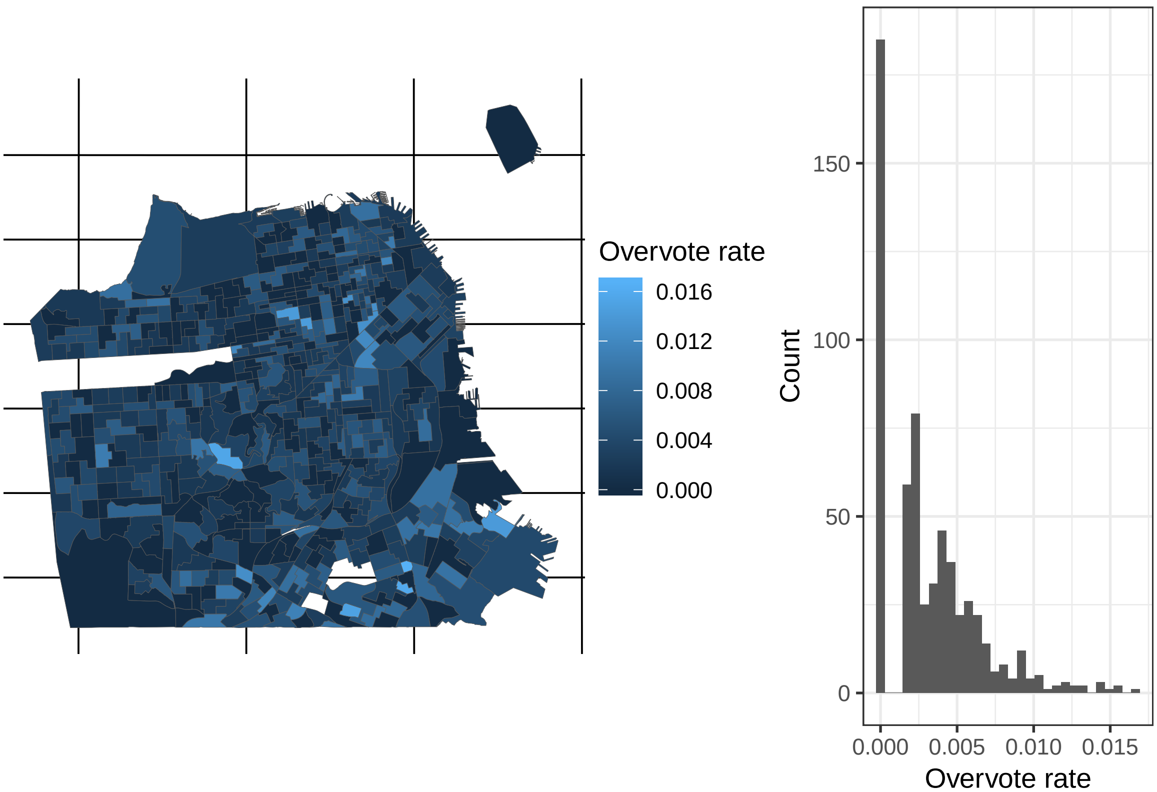


Figure 3.4: Observed overvote rate by precinct

### 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.

## 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.

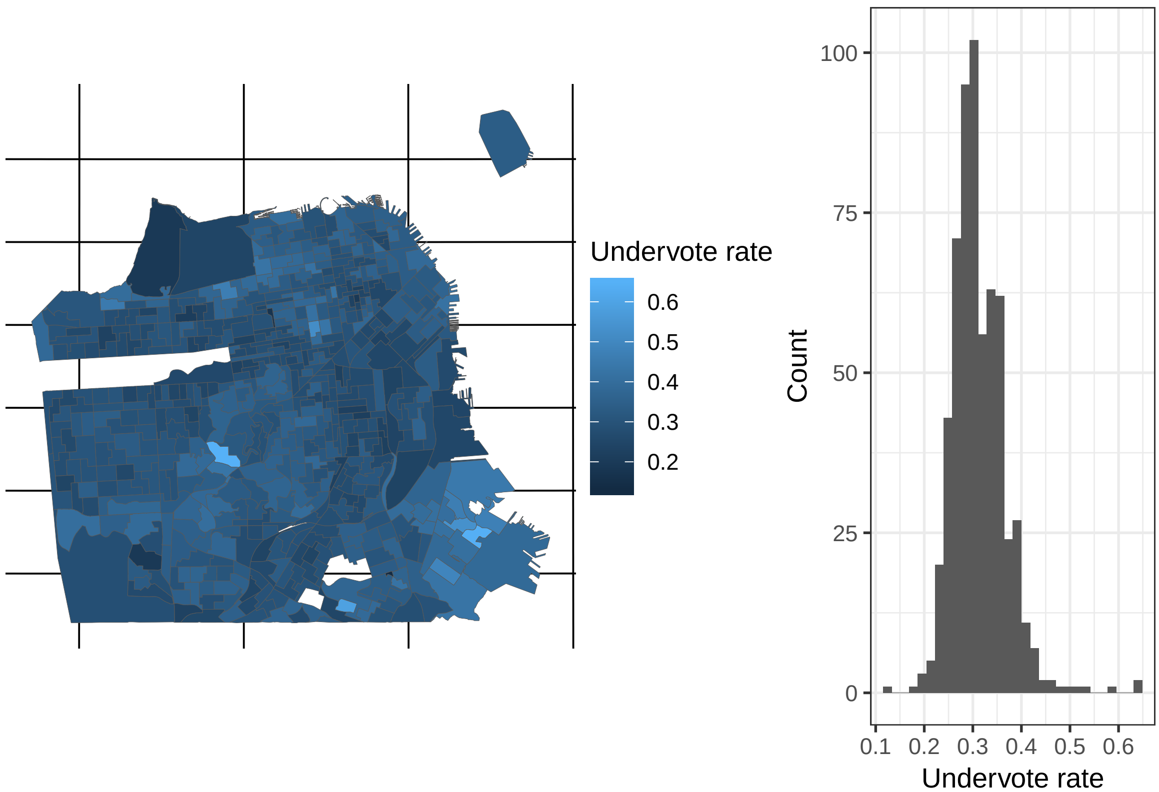


Figure 3.7: Observed undervote rate by precinct

### 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 |

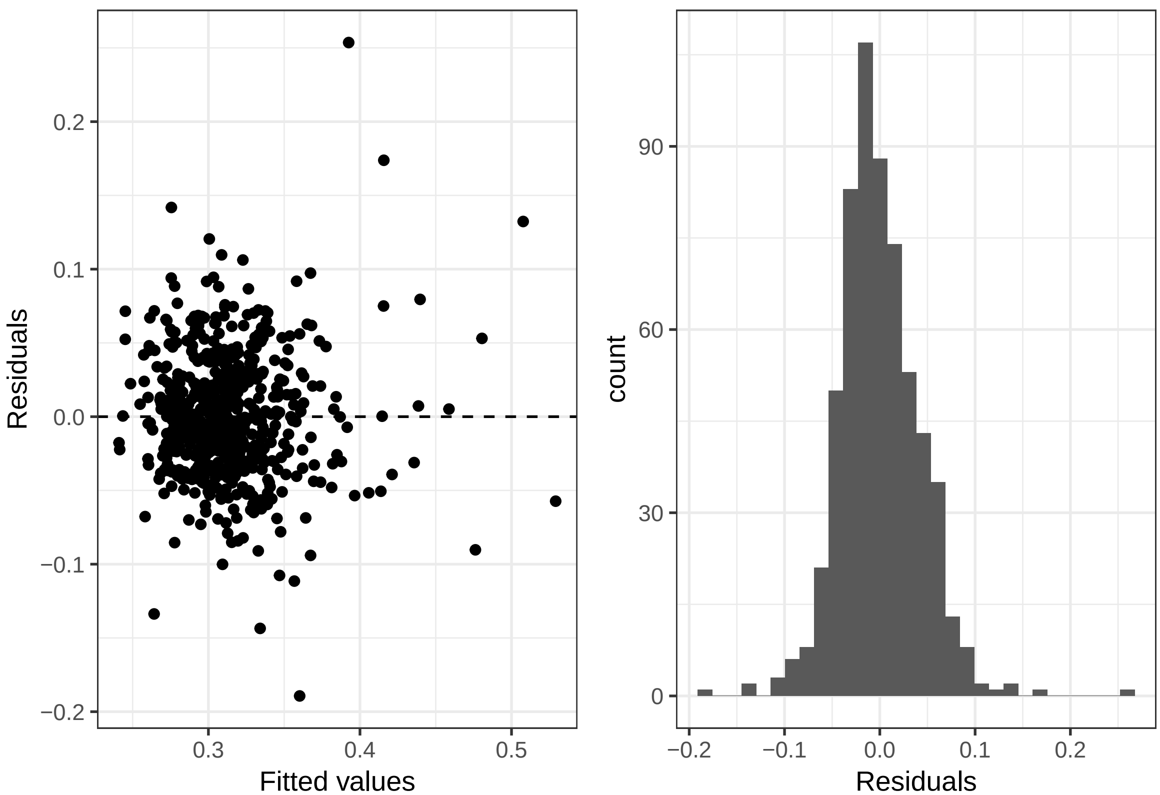


Figure 3.8: Linear undervote model validation

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.

## 6.2 London Breed's final vote share

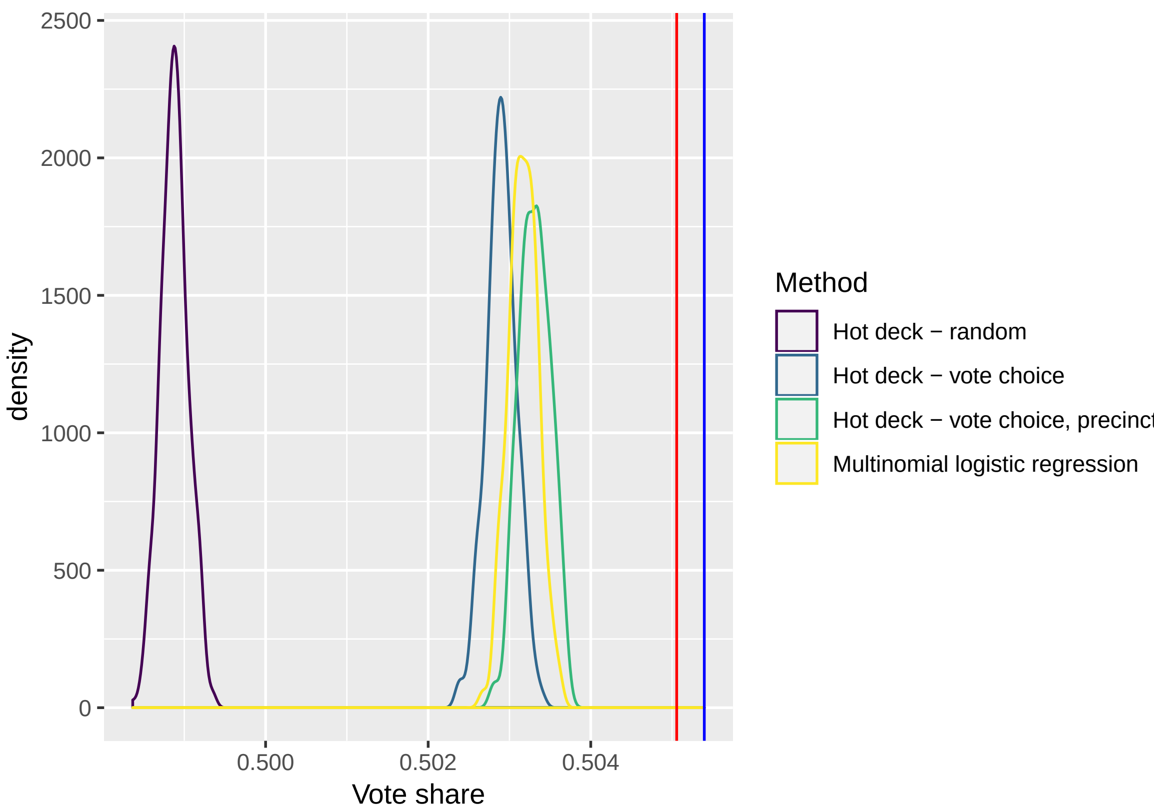


Figure 6.1: London Breed's simulated vote share - unadjusted variances

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 |

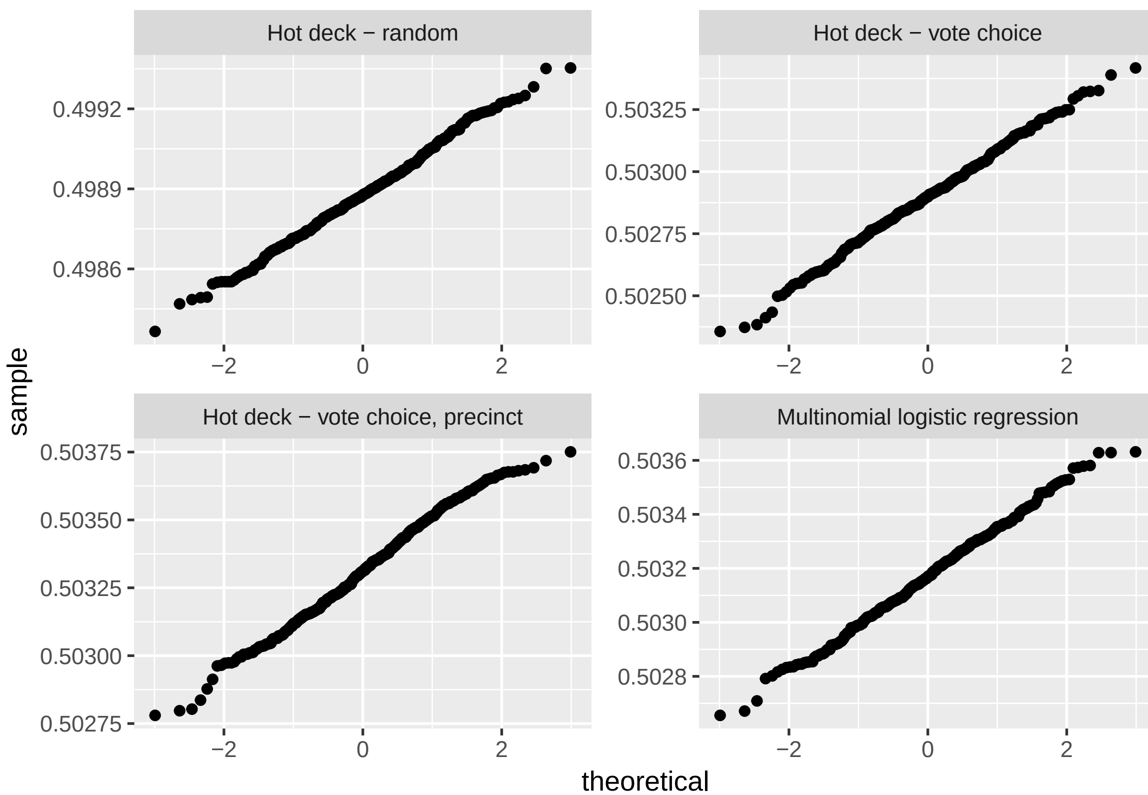


Figure 6.2: Normality check, Q-Q plot of Breed's vote share

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.

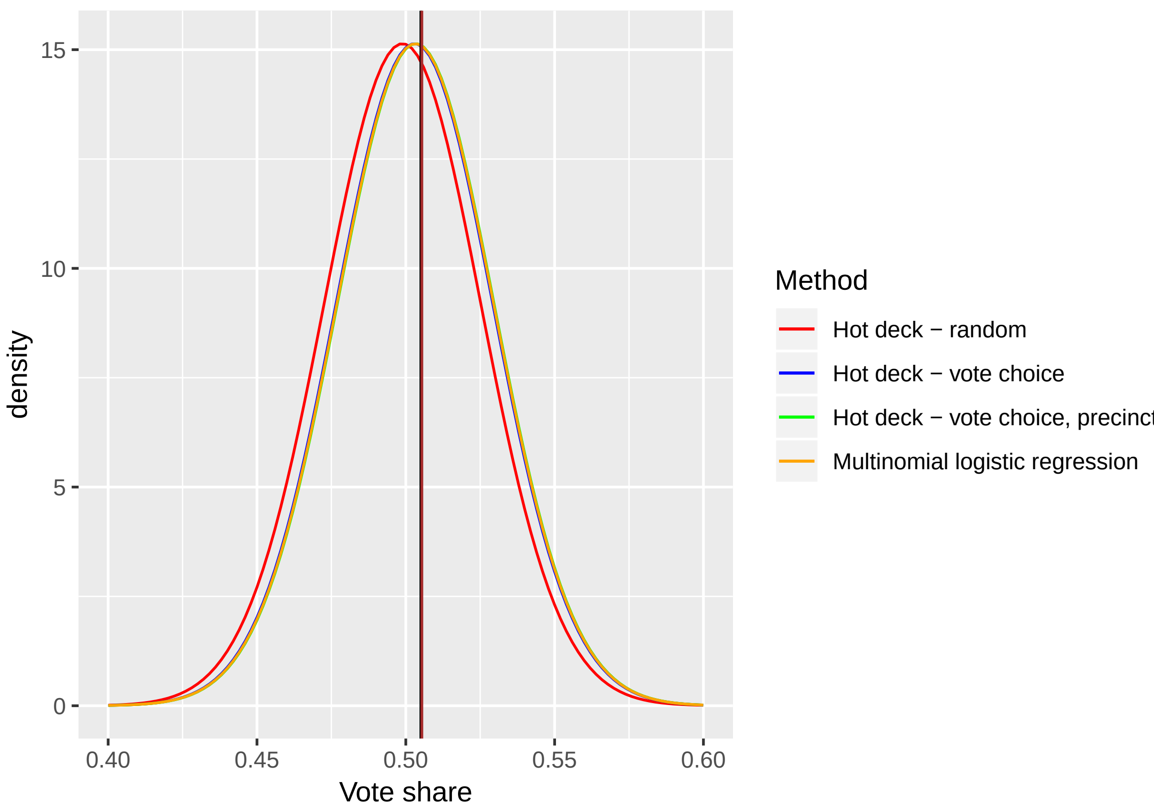


Figure 6.3: London Breed's simulated vote share - adjusted variances

## 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.

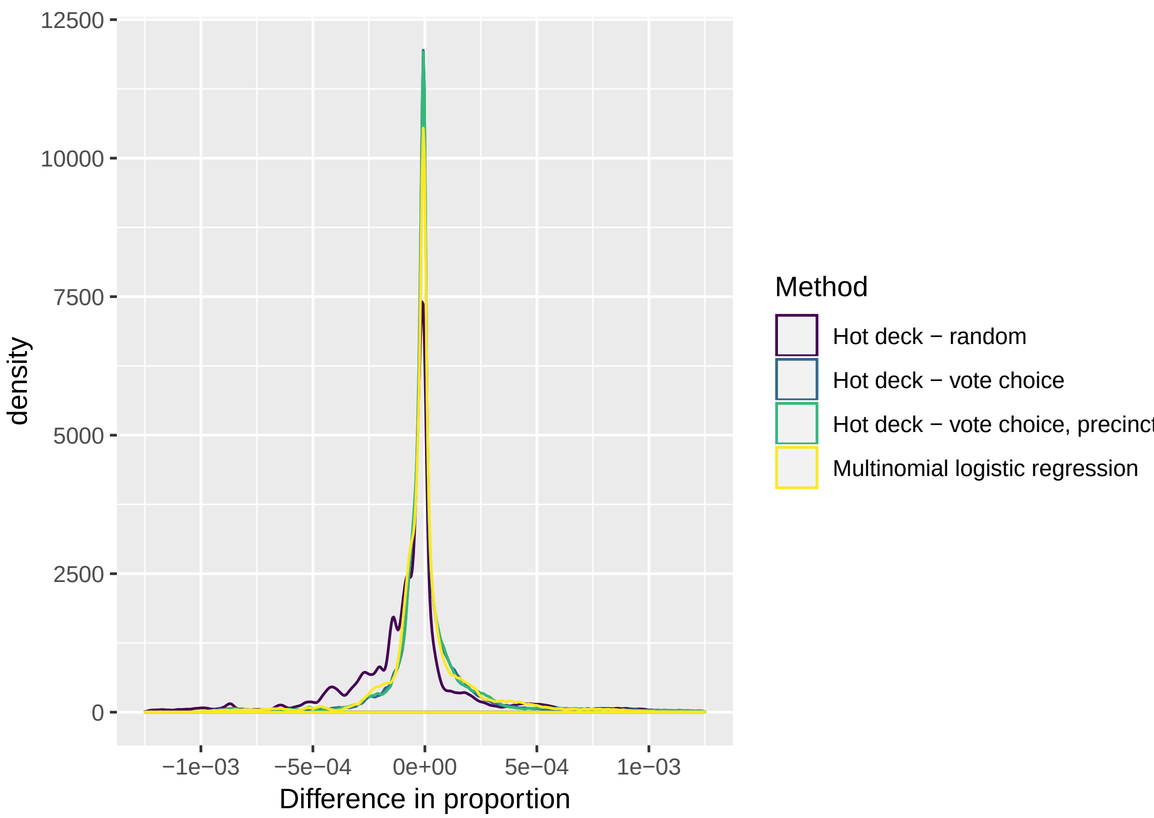


Figure 6.4: Difference in vote combination rates

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

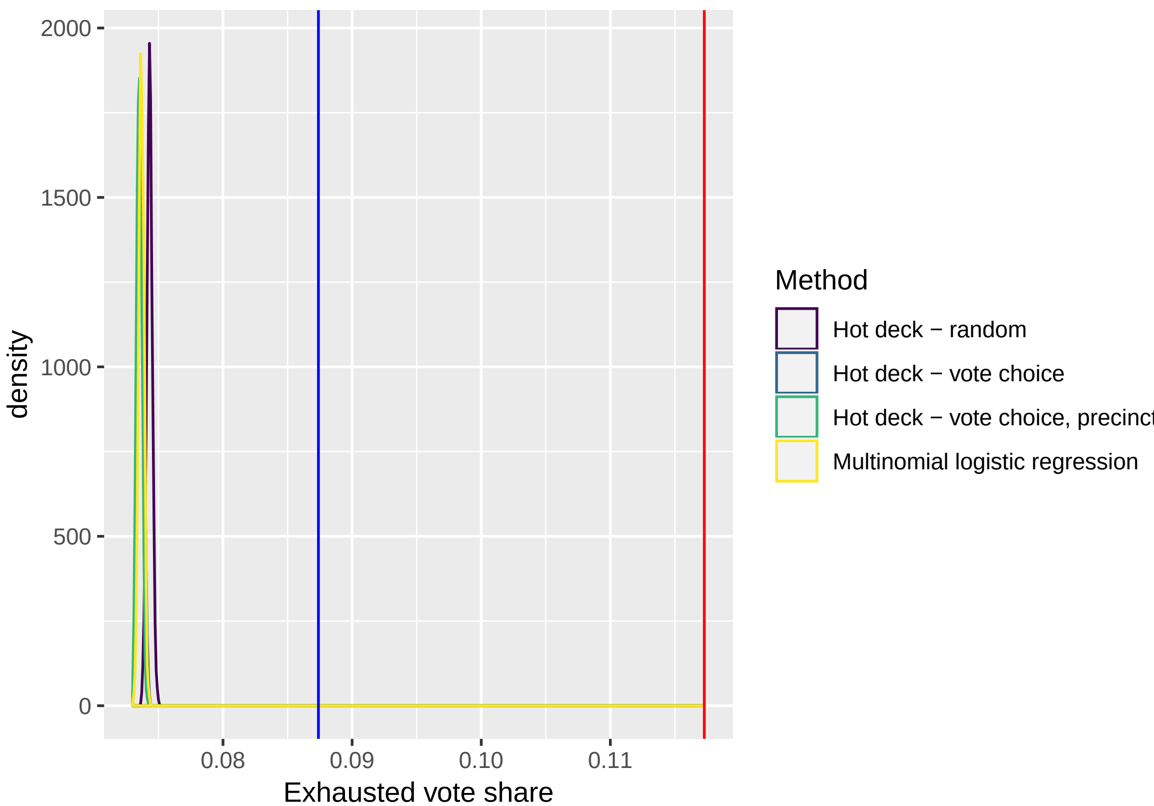


Figure 6.5: Simulated exhausted vote share - unadjusted variances

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 |

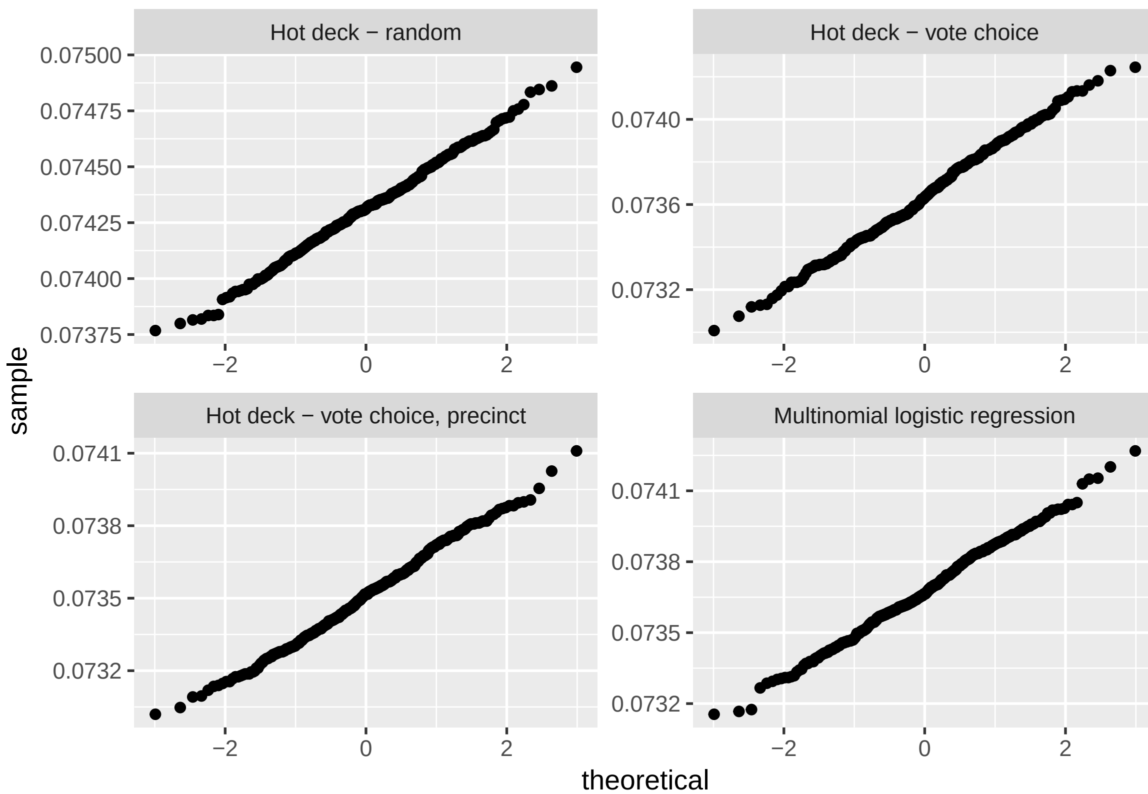


Figure 6.6: Normality check, Q-Q plot of exhausted vote share

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.

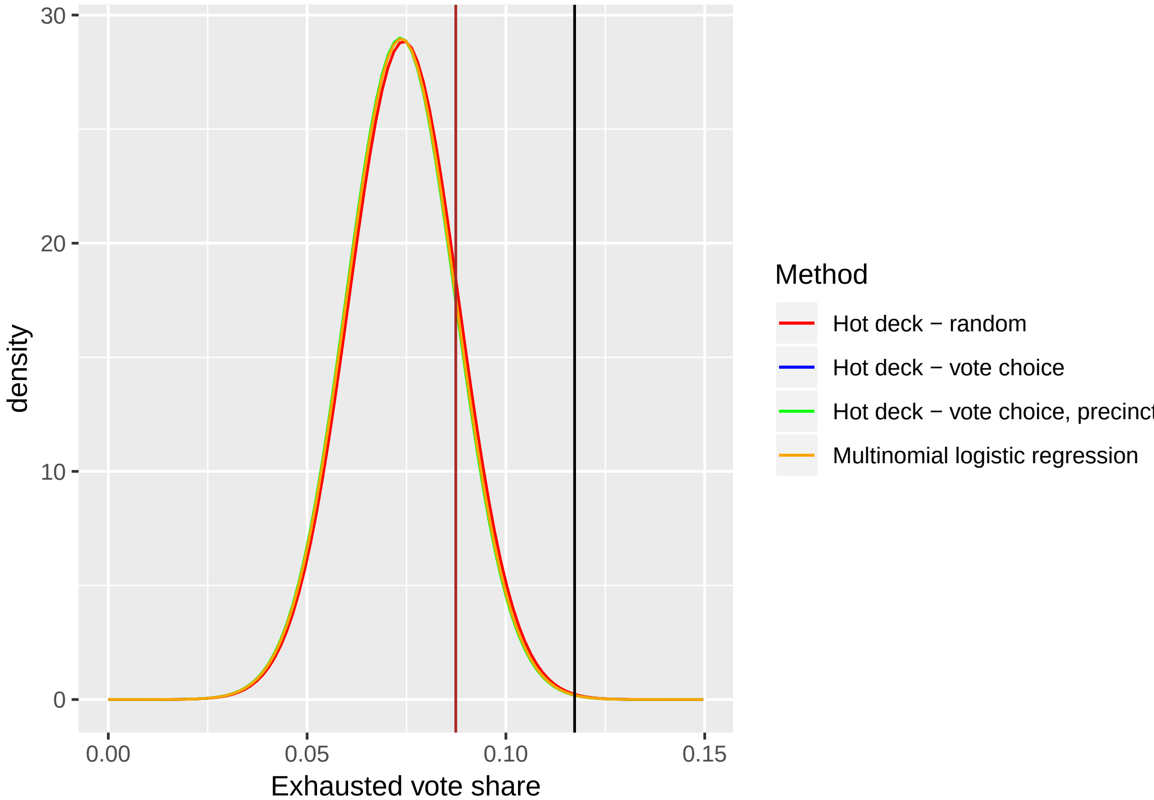


Figure 6.7: Simulated exhausted vote share - adjusted variances