Disaggregation of Unsorted Votes

MGGG Redistricting Lab

Contents

Contributors

Amy Becker contributed to the study design and analysis in this report.

Version 0.0 | July 27, 2021

1 Introduction

There are many aspects of the redistricting process that require strong data-disaggregation methods. In general, the need for these methods arise when specific sources of data are not available for the desired geographic units. For example, redistricting practitioners usually build districting plans out of precincts or small Census units (e.g. Blocks or Block Groups), and they need to have accurate voting data on these building-block units in order to assess redistricting criteria such as partisan fairness and Voting Rights Act compliance.

However, some modes of voting, including absentee and early voting, are often only reported at the county level. We call these votes *unsorted* as they have not been sorted into respective precincts. Unfortunately, counties are generally too coarse to use as district building blocks[1], so we need methods to accurately distribute those county-level votes down to more suitable building blocks such as precincts. This is particularly true when using the 2020 election data because such a large fraction of voters used these alternative voting modes. In this report we compare several different methods for county-to-precinct vote disaggregation and discuss various methodological considerations for practitioners.

The general problem of translating data from one type of spatial unit to another is a common data challenge that faces redistricting practitioners, and many methods have been developed to address this need. One of the simplest such methods is areal-weighting,[2] which allocates data proportionally to the area of the spatial units. Areal-weighting is often poorly suited for distributing demographic and voting data, as we will see later in this report, because it allocates more people and votes to larger-area units, which actually tend to be more rural and less-populated. Still this method is commonly used, including in many automated GIS tools such as the ArcGIS Spatial Join Tool[3]. When working with demographic and voting data, practitioners should take care to understand how their data is being joined and allocated when relying on such tools.

In the redistricting context, more accurate data disaggregation techniques often involve allocating data proportionally to well-correlated proxy value(s) (for example, allocating votes proportionally to population). The MGGG Redistricting Lab's MAUP library[4] provides a flexible set of general-purpose tools that are suitable for most standard geographic data aggregation and disaggregation tasks.

For the specific task of county-to-precinct disaggregation of unsorted votes, we consider some of these more common strategies in addition to some more-tailored methods.

[1] Some plans are built out of whole-county building blocks, most notably Iowa's Congressional maps. But in general counties are too few and/or too populous to keep whole *and* provide enough flexibility to achieve other redistricting criteria such as population balance.

[2] See Amos, McDonald, and Watkins, *When Boundaries Collide: Constructing a National Database of Demographic and Voting Statistics*, `https://doi.org/10.1093/poq/nfx001`

[3] `https://pro.arcgis.com/en/pro-app/latest/tool-reference/analysis/spatial-join.htm`

[4] `https://github.com/mggg/maup`

2 Methodology

We compare several different methods to address the following common scenario. A state reports its in-person, election-day votes at the precinct level but reports absentee votes and early votes only at the county level. In fact many states do ultimately report votes at the precinct level, but they neither break these precinct vote totals down by constituent voting modes nor transparently describe their disaggregation methods.

Because precincts nest into counties and unsorted votes are reported at the county level, we can reduce the disaggregation problem to one of defining a weighting scheme over the precincts in a county. The county vote totals for each candidate and voting mode can then be allotted according to these weights in order to form precinct level vote estimates. Because we assume to have county totals for each candidate and by each mode, these weighting schemes can be determined independently for each.

Though there is little to be found in the literature on actual implementation or analytical comparisons of methods for this specific problem, there is a brief high level discussion of ways to "account for non-polling place votes" in Michael McDonald's *Presidential Vote within State Legislative Districts*[5]. In this report we describe various ways to actually implement some of those high level ideas, and we point the reader to Appendix A as well as our code base[6] for further details.

2.1 County-to-precinct Disaggregation Methods

To define the disaggregation methods, we introduce the following notation. For an arbitrary Precinct P in County C, we show how each method calculates the weight w_P for Precinct P. The weighting scheme for the county would be formed by calculating such a w_{P_i} for each Precinct P_i in County C.

1. **Uniformly**
   - **Definition:** Distribute votes evenly to every precinct in the county.

$$w_P = \frac{1}{\#\, precincts(C)}$$

   - **Notes:** This method gives the same weighting scheme across the county for each candidate and voting mode. It places equal votes in each of a county's precincts.

2. **By Area**
   - **Definition:** Distribute votes proportionally to the precinct's share of the county's area.

$$w_P = \frac{area(P)}{area(C)}$$

   - **Notes:** This method gives the same weighting scheme across the county for each candidate and voting mode. It places more votes in a county's larger-area precincts.

[5] https://www.jstor.org/stable/pdf/24711104.pdf?refreqid=excelsior%3Ad141b71d186bce89c88fa3de4742d560

[6] https://github.com/mggg/county-vote-disaggregation

3. **By Total Population**

   - **Definition:** Distribute votes proportionally to the precinct's share of the county's total population.
   $$w_P = \frac{population(P)}{population(C)}$$
   - **Notes:** This method gives the same weighting scheme across the county for each candidate and voting mode. It places more votes in the county's most populated precincts.

4. **By Voting-Aged Population (VAP)**

   - **Definition:** Distribute votes proportionally to the precinct's share of the county's voting-aged population.
   $$w_P = \frac{VAP(P)}{VAP(C)}$$
   - **Notes:** This method gives the same weighting scheme across the county for each candidate and voting mode.

5. **By Citizen Voting-Aged Population (CVAP)**

   - **Definition:** Distribute votes proportionally to the precinct's share of the county's citizen voting-aged population.
   $$w_P = \frac{CVAP(P)}{CVAP(C)}$$
   - **Notes:** This method gives the same weighting scheme across the county for each candidate and voting mode.

6. **By Election-Day Votes**

   - **Definition:** Distribute each candidate's votes proportionally to the precinct's share of the county's in-person, election-day votes for that candidate.
   $$\text{(for Candidate 1)} \quad w_P = \frac{\text{Election-Day votes for Candidate 1 in Precinct } P}{\text{Election-Day votes for Candidate 1 in County } C}$$
   $$\text{(for Candidate 2)} \quad w_P = \frac{\text{Election-Day votes for Candidate 2 in Precinct } P}{\text{Election-Day votes for Candidate 2 in County } C}$$
   $$\vdots$$
   - **Notes:** This method gives the same weighting scheme across the county for each voting mode, but different weighting schemes for each candidate.

7. **By Voterfile Total Voters**

   - **Definition:** Distribute votes proportionally to the precinct's share of the county's total voters according to the voterfile.
   $$w_P = \frac{Voted(P)}{Voted(C)}$$
   - **Notes:** This method gives the same weighting scheme across the county for each candidate and voting mode.

8. **By Voterfile Total Voters Minus Election-Day Voters**

   - **Definition:** Distribute votes proportionally to the precinct's share of the county's total voters according to the voterfile minus the number of election-day voters.

$$w_P = \frac{\text{Voterfile Voters in Precinct } P - \text{Election-Day Voters in Precinct } P}{\text{Voterfile Voters in County } C - \text{Election-Day Voters in County } C}$$

   - **Notes:** This method gives the same weighting scheme across the county for each candidate and voting mode.

9. **By Voterfile Mode**

   - **Definition:** Distribute votes for each voting mode proportionally to the precinct's share of the county's voters *by that mode* according to the voterfile.

$$\text{(for Absentee Voters)} \quad w_P = \frac{\text{Voterfile Absentee Voters in Precinct } P}{\text{Voterfile Absentee Voters in County } C}$$

$$\text{(for Early Voters)} \quad w_P = \frac{\text{Voterfile Early Voters in Precinct } P}{\text{Voterfile Early Voters in County } C}$$

   - **Notes:** This method gives the same weighting scheme across the county for each candidate but different weighting schemes for each voting mode.

10. **By Voterfile Party**

   - **Definition:** Distribute votes for each candidate proportionally to the precinct's share of the county's voters *registered to the candidate's party* according to the voterfile.

$$\text{(for Candidate 1)} \quad w_P = \frac{\text{Voterfile Voters Registered to Candidate 1's Party in Precinct } P}{\text{Voterfile Voters Registered to Candidate 1's Party in County } C}$$

$$\text{(for Candidate 2)} \quad w_P = \frac{\text{Voterfile Voters Registered to Candidate 2's Party in Precinct } P}{\text{Voterfile Voters Registered to Candidate 2's Party in County } C}$$

$$\vdots$$

   - **Notes:** This method gives the same weighting scheme across the county for each voting mode but different weighting schemes for each candidate, though different candidates from the same *party* will have the same weighting scheme.

11. **By Voterfile Mode and Party**

   - **Definition:** Distribute votes for each candidate and voting mode proportionally to the precinct's share of the county's voters *that are registered to the candidate's party and voted via that mode* according to the voterfile.

$$\text{(for Candidate 1 Absentee Voters)} \quad w_P = \frac{\text{Voterfile Absentee Voters Registered to Candidate 1's Party in Precinct } P}{\text{Voterfile Absentee Voters Registered to Candidate 1's Party in County } C}$$

$$\text{(for Candidate 1 Early Voters)} \quad w_P = \frac{\text{Voterfile Early Voters Registered to Candidate 1's Party in Precinct } P}{\text{Voterfile Early Voters Registered to Candidate 1's Party in County } C}$$

$$\text{(for Candidate 2 Absentee Voters)} \quad w_P = \frac{\text{Voterfile Absentee Voters Registered to Candidate 2's Party in Precinct } P}{\text{Voterfile Absentee Voters Registered to Candidate 2's Party in County } C}$$

$$\text{(for Candidate 2 Early Voters)} \quad w_P = \frac{\text{Voterfile Early Voters Registered to Candidate 2's Party in Precinct } P}{\text{Voterfile Early Voters Registered to Candidate 2's Party in County } C}$$

$$\vdots$$

- **Notes:** This method gives different weighting schemes for each candidate and each mode, though candidates from the same party will have the same weighting scheme.

For each candidate and voting mode, once these weights have been calculated for each precinct in a county, they are multiplied by the county's vote totals for that candidate and by that mode. Specifically, if we let $V_{(C,X,M)}$ be the total number of unsorted votes for Candidate X via mode M in County C, then $Est_{(P,X,M)}$ are the number of votes for Candidate X via mode M that are estimated for precinct P by the given disaggregation method, where:

$$Est_{(P,X,M)} = V_{(C,X,M)} \cdot w_P$$

Precinct-level vote estimates by-mode are formed for the whole state by performing these calculations separately on each county and then grouping all of the resulting precinct estimates together. See Appendix A for detailed examples of these calculations.

2.2 Other Methods

This list of disaggregation methods is not intended to be comprehensive, but rather to demonstrate a sample of options that practitioners might consider trying. There are many other plausible ways to distribute votes, including more complicated methods (e.g. incorporating other demographic data such as age or race or even other sources of data such as polls). In fact, since each method uses only within-county data, practitioners can even disaggregate votes from different counties using different methods.

Only one of our proposed methods, disaggregation 'By Voterfile Mode and Party', actually gives different weighting schemes for each candidate *and* mode. Ideally, disaggregation methods would have the flexibility to reflect the differences between these votes distributions, as the (unknown) ground truth distributions are unlikely to be identical for different candidates and modes. However, in practice it is not so straightforward to design methods that account for both of these aspects at once in a reasonable way.

Finally, it may be tempting to use precinct-level vote returns from previous elections to dictate disaggregation methods. The idea here is that if a precinct's number of voters (or proportion of the county votes) is fairly stable between elections, then we can base our estimates on prior data (i.e from elections when there were more in-person voters and thus less uncertainty). Still, those prior elections are also unlikely to have accurately and/or transparently sorted any alternative-modality votes into precincts, and for places with a non-trivial number of such votes, these methods may simply reinforce error in how the data is already reported, rather than accurately sort the votes. Furthermore, voting behavior in 2020 was markedly different than prior elections. Still, there may be some places where using previous returns to inform vote disaggregation can help improve estimates.

2.3 Data Requirements

The methods listed above require practitioners to have access to various sources of data:

- **Election Returns:** Most crucially, the votes totals themselves are required. All of the methods require knowing the number of unsorted votes at the county level and the list of each county's precincts used in a given election. The 'By Election-Day Votes' and 'By Voterfile Total Voters Minus Election-Day Voters' disaggregation additionally require tallies of the in-person, election-day votes in each precinct in a county.
- **Shapefiles:** Although not technically required to calculate the proportions, shapefiles are likely needed to determine various precinct-level data fields. In particular the 'By Area', 'By Total Population', 'By VAP', and 'By CVAP' methods all require having those data fields at the precinct level. Demographic data is usually attached to precincts by aggregating from Block or Block Group level units. Determining the correct relationship between these units generally requires having shapefiles for each and then using tools like `maup` to help translate data on one set of units to data on another set of units. Block and Block Group shapefiles are readily available from the Census. Precinct-level shapefiles often pose a challenge. They are not always provided by the state and even when they are, they are often outdated as precinct lines fluctuate election-to-election. Poor precinct stability and ambiguous precinct matching pose problems in general with disaggregation methods.
- **Voterfiles:** Several of the described methods require voterfiles. Specifically, voterfile history by precinct, voting mode, and party registration. The availability of this data varies widely state-to-state. Some states offer these files freely to anyone, while others charge money or provide them only to residents or candidates. Additionally, voterfile data fields vary state-to-state. Some states require party registration, whereas others do not include or require it. If precincts are not included or do not match well onto the precincts from election returns, practitioners can geocode voter addresses and use precinct shapefiles to help assign voters to precincts. Additionally there are commercially available voter data sets for each state, though it is often not clear what sources and methods were used to generate these.

2.4 An Important Note on Distinguishing Distributions

There are some important subtleties when thinking about vote distributions. In many contexts, we are more familiar with working with distributions of vote *shares*. Figure 1, for example shows the two-way partisan share of 2020 Presidential votes for each precinct in Mecklenburg County, North Carolina (which contains Charlotte, the state's largest city) for each voting mode. The redder precincts indicate a higher share of Trump votes and the bluer precincts indicate a higher share of Biden votes.

We include these maps to emphasize the distinction between two often conflated questions: 'How does the number of Trump Early votes in Precinct A compare to the number of Biden Early votes in Precinct A?" and "How does the number of Trump Early votes in Precinct A compare to the number of Trump Early votes in Precinct B?" The former may ultimately be of more interest, but it tells us very little about *where* to place actual votes (a precinct that leans 70% Biden could have 7 Biden votes or 7000 Biden votes!). We focus instead on estimating the latter, noting that a strong disaggregation method would ideally preserve precinct-level partisan shares as well as vote estimates (as we will see in the following section).

Figure 2 shows instead the type of distributions we focus on in our disaggregation methods. The top row shows four *target vote distributions* (Early and Absentee vote distributions, for each of the two major candidates) for the 2020 Presidential Election. These maps show how each of these types of votes were actually distributed across the county. That is the different color shades show how, for example, the number of Trump Early votes in one precinct compares to the number of Trump Early votes in a neighboring precinct. Note that these four distributions are not terribly different from each other, even though they are showing votes for opposing candidates, which may seem counter-intuitive! In fact the precincts with the most voters are likely to have relatively higher numbers of *both* Trump votes *and* Biden votes compared to precincts with fewer voters. This does not tell us *anything* about how the number of Trump votes compares to the number of Biden votes in any precinct, only how each type of vote itself varies across the county.

These top-row target distributions are the distributions that we are trying to approximate with our weighting schemes. The other maps in Figure 2 (below the top row) show the various weighting schemes generated by our methods described in Section 2.1. Put simply, the main goal of this work is to find which weighting-scheme map (from below the top row) best approximates each of the top-row target vote-distribution maps.

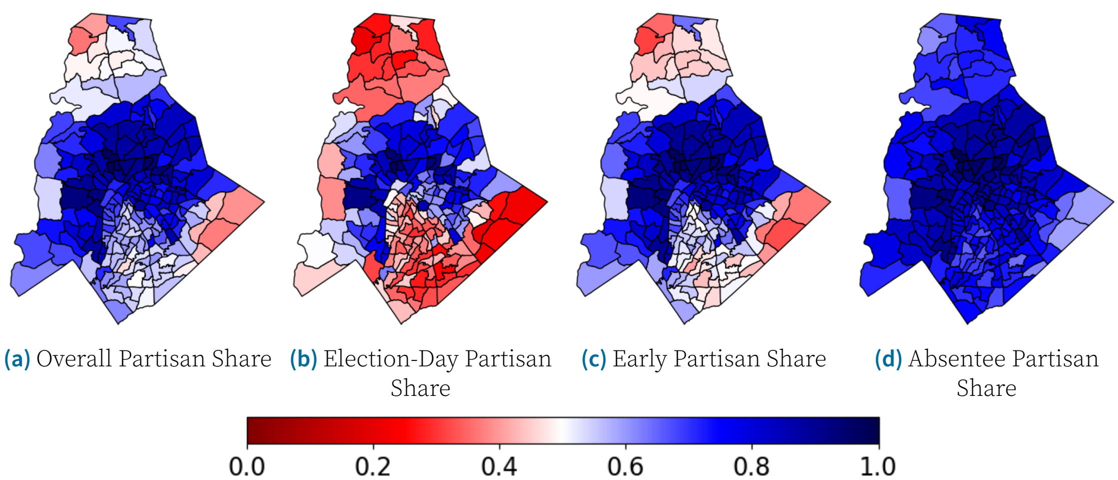
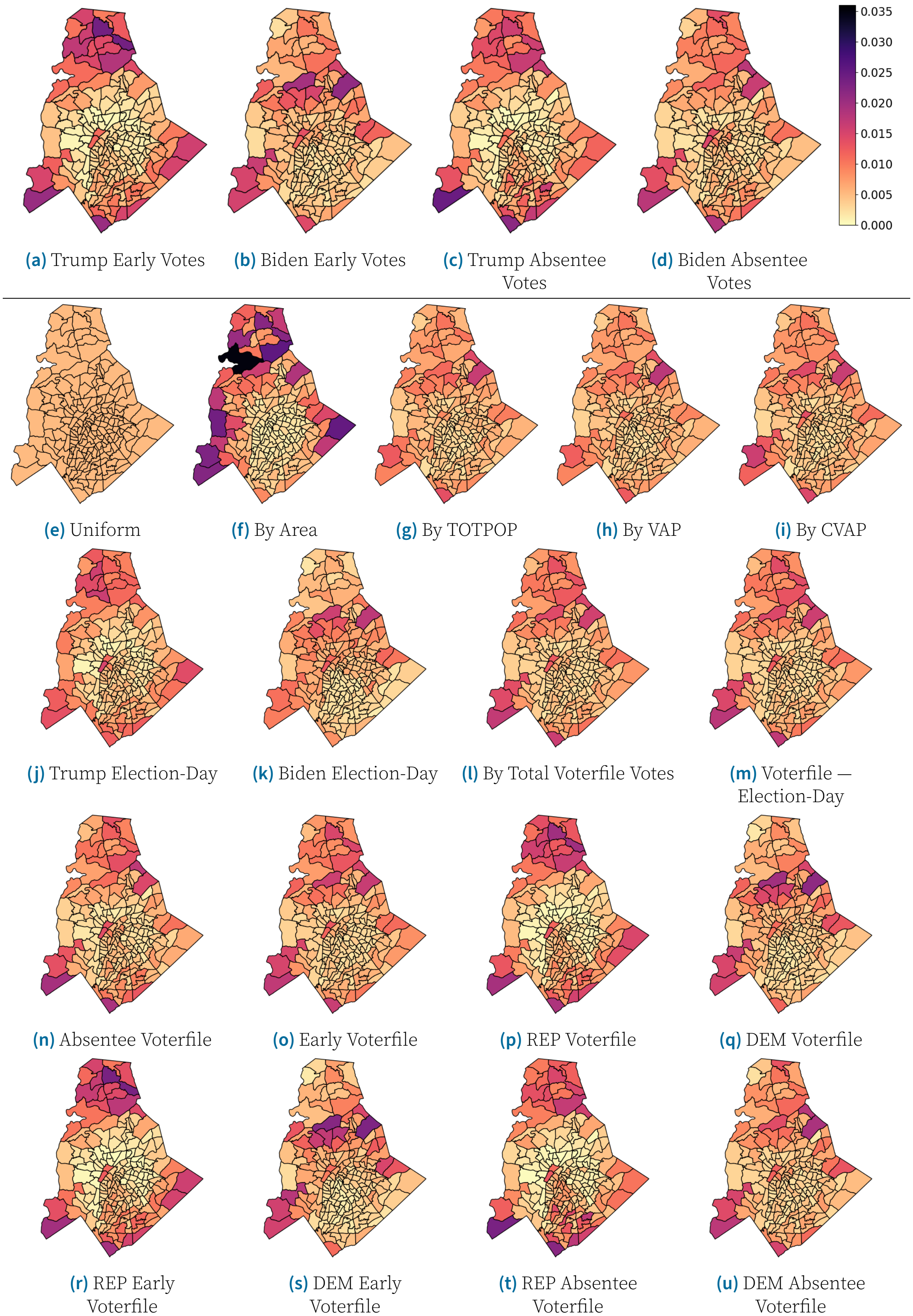


Figure 1. Precinct-level partisan share by voting mode for 2020 Presidential Votes in Mecklenburg County, North Carolina.



(a) Trump Early Votes (b) Biden Early Votes (c) Trump Absentee Votes (d) Biden Absentee Votes

(e) Uniform (f) By Area (g) By TOTPOP (h) By VAP (i) By CVAP

(j) Trump Election-Day (k) Biden Election-Day (l) By Total Voterfile Votes (m) Voterfile — Election-Day

(n) Absentee Voterfile (o) Early Voterfile (p) REP Voterfile (q) DEM Voterfile

(r) REP Early Voterfile (s) DEM Early Voterfile (t) REP Absentee Voterfile (u) DEM Absentee Voterfile

Figure 2. Target vote distributions compared to different weighting-scheme options for 2020 Presidential Votes in Mecklenburg County, North Carolina.

3 Analysis

To compare various disaggregation methods we used North Carolina and Oklahoma as case studies. For each state we analyzed the major-party votes in the 2016 and 2020 Presidential elections. We differentiated between three voting modes: Election-Day, Early, and Absentee. Table 1 shows a summary of the statewide vote breakdown by party and mode.

State	Year	Party	ALL	Absentee	Early	Election-Day
North Carolina	2016	**TOTAL**	4,551,947	190,127 4.2%	2,836,372 62.3%	1,525,448 33.5% %
		Republican (Trump)	2,362,631 51.9%	98,147 2.2%	1,376,149 30.2%	888,335 19.5%
		Democrat (Clinton)	2,189,316 48.1%	91,980 2.0%	1,460,223 32.1%	637,113 14.0%
	2020	**TOTAL**	5,443,067	977,636 18.0%	3,576,323 65.7%	889,108 16.3%
		Republican (Trump)	2,758,775 50.7%	277,862 5.1%	1,890,765 34.7%	590,148 10.8%
		Democrat (Biden)	2,684,292 49.3%	699,774 12.9%	1,685,558 31.0%	298,960 5.5%
Oklahoma	2016	**TOTAL**	1,369,511	95,350 7.0%	145,869 10.7%	1,128,292 82.4%
		Republican (Trump)	949,136 69.3%	60,975 4.5%	95,767 7.0%	792,394 57.9%
		Democrat (Clinton)	420,375 30.7%	34,375 2.5%	50,102 3.7%	335,898 24.5%
	2020	**TOTAL**	1,524,170	274,217 18.0%	164,994 10.8%	1,084,959 71.2%
		Republican (Trump)	1,020,280 66.9%	111,171 7.3%	109,186 7.2%	799,923 52.5%
		Democrat (Biden)	503,890 33.1%	163,046 10.7%	55,808 3.7%	285,036 18.7%

Table 1. Vote breakdown by party and voting mode for 2016 and 2020 Presidential elections in North Carolina and Oklahoma.

3.1 Data

Both states report precinct level votes broken down by voting mode,[7] which allow us to compare the vote disaggregation estimates to the 'ground truth' values given by the state.[8] Both states also provide freely available voterfiles with voter history by mode and party affiliation.

North Carolina election returns, precinct shapefiles, and voter history were downloaded from

[7]Oklahoma election results are broken down this way for all counties except for Tulsa County and Oklahoma County, which report Election-Day results at the precinct level but report alternative-modality vote sums for the county.

[8]Technically, these states rely on (undisclosed) disaggregation methods themselves to sort ballots when the precincts are ambiguous, but these states provide the closest data to 'ground truth' available.

the North Carolina State Board of Elections.[9] Oklahoma election returns[10] and voter list[11] were provided by the Oklahoma State Election Board. We used the Oklahoma shapefile prepared by MGGG.[12] For both states, VAP and total population came from the 2010 Census and CVAP came from the 2015-2019 5-year ACS.

Even though these states provided the best available opportunity to test various disaggregation methods against an underlying ground truth, data challenges still emerged. Most notably, perfect precinct matching is not always available. In North Carolina, for example, even though a large majority of the votes are sorted into actual geographic precincts, there are still some "precincts" that do not correspond to polling-place geographies (e.g. 'ONE STOP WALNUT COVE' in STOKES County, North Carolina). Additionally, for North Carolina, there were discrepancies between the "Precinct Sorted Results"[13] (which we used as our ground truth) and the official state-level election results.[14]. These discrepancies were relatively minor for the 2020 elections but noticeably larger for the 2016 elections.

3.2 Methodolgy

For each state (North Carolina and Oklahoma), year (2016 and 2020), and alternative voting mode (Early and Absentee), we used each of the eleven disaggregation methods described in Section 2.1 to generate precinct-level vote estimates. Though the election results for these two states are available already sorted into precincts by voting mode, we only used county-aggregated totals of Early and Absentee votes to form our estimates. That is, we used only the information that would be available in states that *do not* report their sorted votes in this way. The precinct-sorted vote data was then used as the "ground truth" against which to compare the quality of our estimates.

3.3 Results

One important question that practitioners may face is: which of these disaggregation methods is best? Answering this question is not so straightforward and ultimately depends on the intended use of the disaggregated data. We explore several alternative ways to compare the quality of methods.

3.3.1 Precinct-Level Performance

First, we assess how well these methods perform at the precinct level. To do so, we must decide what attribute(s) of the estimated vote disaggregation to evaluate. To illustrate why this decision matters, we use 2020 Trump Early votes in North Carolina to compare the performance of the 'By Election-Day' and 'By Voterfile Mode and Party' disaggregation methods.

Figure 3 shows a counts-based evaluation: how close is the disaggregation-estimated *number* of Trump Early votes in each precinct to the actual number of Trump Early votes observed in each

[9] https://www.ncsbe.gov/results-data

[10] https://oklahoma.gov/elections/election-info/election-results.html

[11] https://oklahoma.gov/elections/candidate-info/voter-list.html

[12] https://github.com/mggg-states/OK-shapefiles

[13] https://www.ncsbe.gov/results-data/election-results/historical-election-results-data

[14] https://er.ncsbe.gov

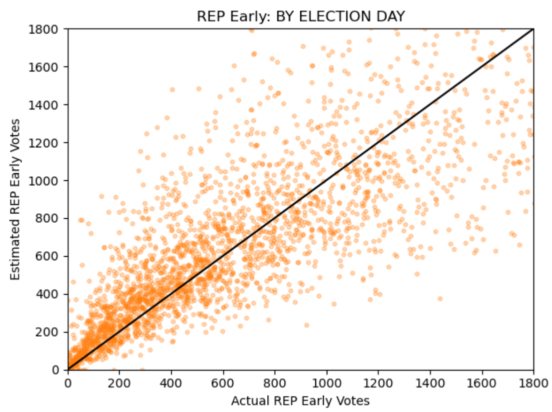
precinct, for each method. The figure shows scatter plots and residual plots for the 'By Election-Day' method (on the left) and the 'By Voterfile Mode and Party' method (on the right). We see that the latter falls much closer to the line $y = x$. That is, the predicted number of votes is very close to the true number of votes for almost all of the precincts. This is supported by the residuals plot, which shows a thin spike near zero.

One concern with counts, however, is that small relative error is disproportionately penalized in larger counties. Suppose a county has 100,000 Early Biden votes to distribute and a disaggregation method estimates that Precinct A had 5100 of those votes, but in reality the precinct only had 5000 of those votes. This will give a residual error of 100 votes, whereas in a county with only 1000 Early Biden votes, similar relative error would only lead to a residual error of 1 vote. Because state county populations, and therefore vote totals, vary drastically (e.g. North Carolina county populations range from 4000 to 1.1 million), counts-based evaluation may be misleading[15].

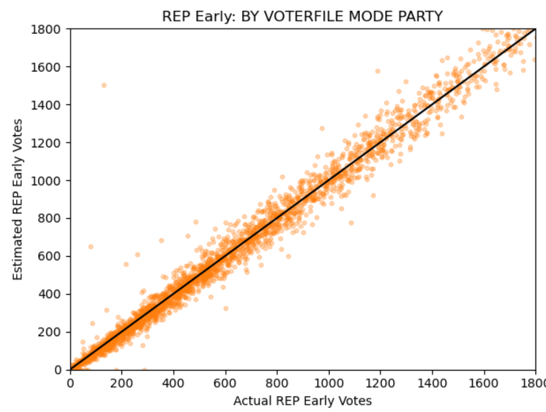
Figure 4 shows instead a percentage-based evaluation that measures error as a percentage of the county-level vote aggregates. Again we see that 'By Voterfile Mode and Party' estimates outperform 'By Election-Day' estimates. Though this evaluation normalizes precincts across the state, it can create the reciprocal problem: small absolute error is disproportionately penalized for small counties. A precinct vote estimate that is only off by a few votes seems a lot worse in small counties than in big counties according to this measure.

A third evaluation is partisanship of the precincts: How well does the disaggregation method preserve the partisan balance of each precinct? Figure 5 shows the analogous plots for this measure. Here the advantage of the 'By Voterfile Mode and Party' method is much less pronounced.

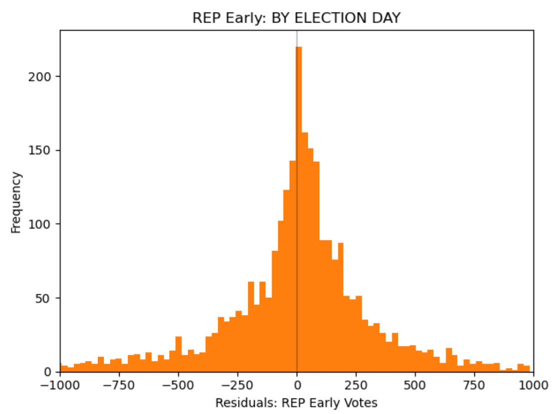
[15]This is especially true when ground truth is near zero for some precincts in large counties. Having data points clustered along the y-axis drives down measures of correlation.



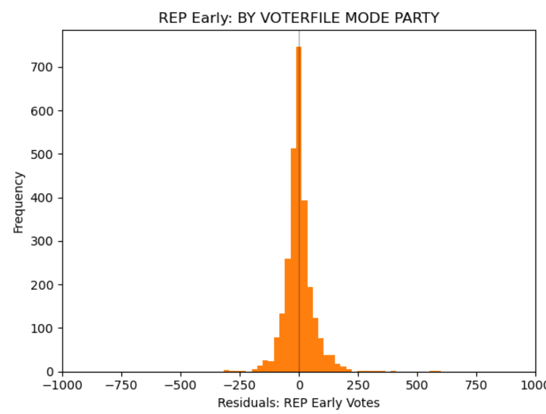
(a) Counts Scatter: By Election-Day



(b) Counts Scatter: By Voterfile Mode and Party

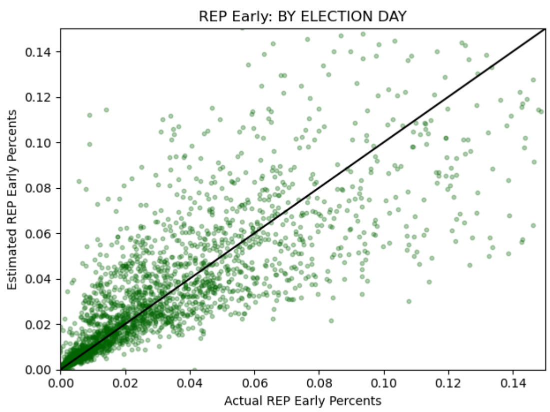


(c) Counts Residuals: By Election-Day

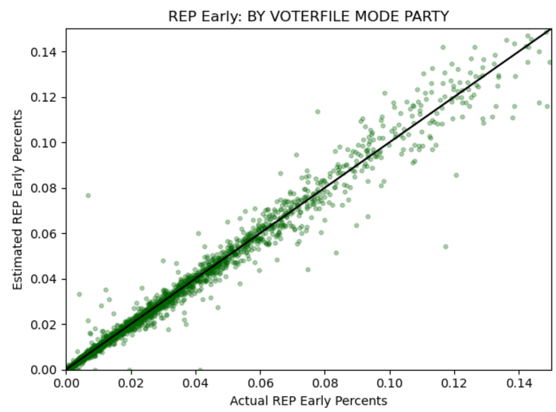


(d) Counts Residuals: By Voterfile Mode and Party

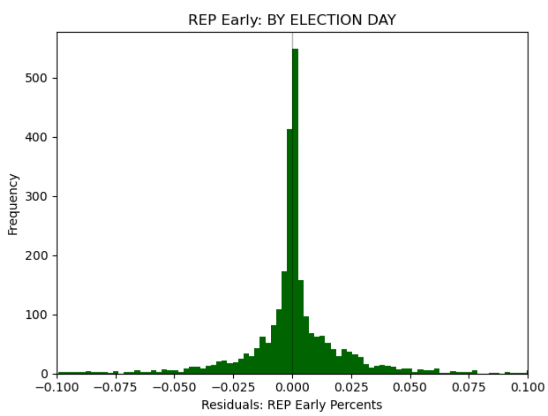
Figure 3. Comparison of 'By Election-Day' and 'By Voterfile Mode and Party' disaggregation methods on estimating the precinct-level number of Republican Early votes cast in the 2020 Presidential Election in North Carolina.



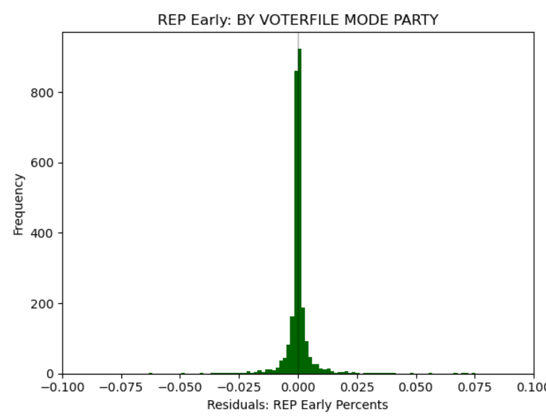
(a) Percent Scatter: By Election-Day



(b) Percent Scatter: By Voterfile Mode and Party

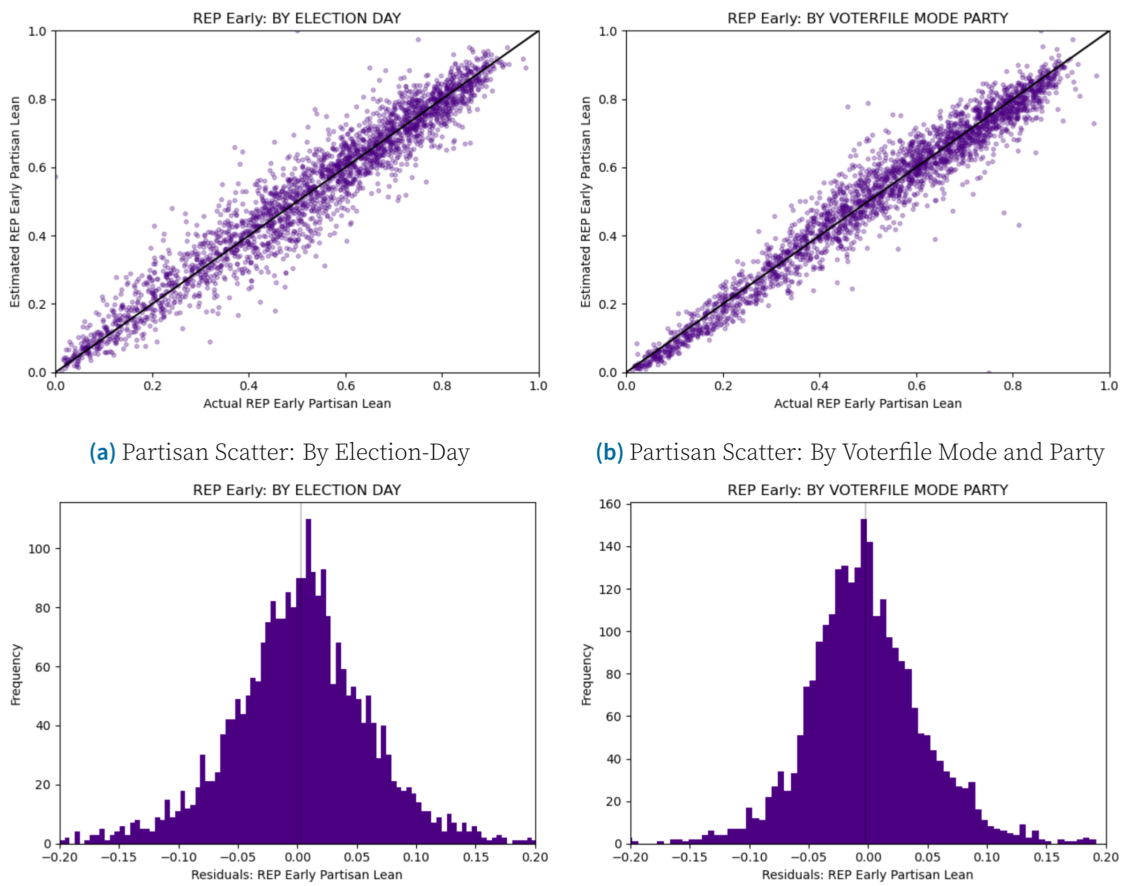


(c) Percent Residuals: By Election-Day



(d) Percent Residuals: By Voterfile Mode and Party

Figure 4. Comparison of 'By Election-Day' and 'By Voterfile Mode and Party' disaggregation methods on estimating the precinct-level percentages of the county totals for Republican Early votes cast in the 2020 Presidential Election in North Carolina.



(a) Partisan Scatter: By Election-Day

(b) Partisan Scatter: By Voterfile Mode and Party

(c) Partisan Residuals: By Election-Day

(d) Partisan Residuals: By Voterfile Mode and Party

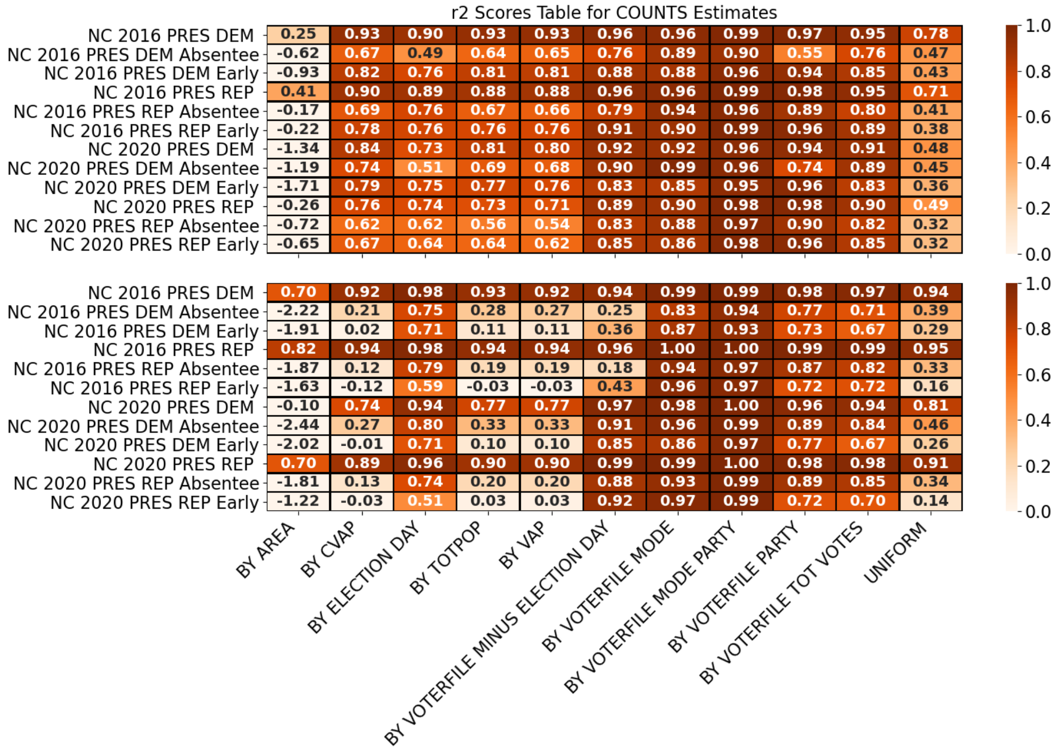
Figure 5. Comparison of 'By Election-Day' and 'By Voterfile Mode and Party' disaggregation methods on estimating the precinct-level Republican partisan share of Early votes cast in the 2020 Presidential Election in North Carolina.

Ideally, a strong disaggregation method would be robust across all of these precinct-level evaluations (and others). Figures 6, 7, and 8, summarize these three evaluation measures (counts-based, percent-based, and partisan) across all methods and instances. For each of the two states, years, and major parties, we consider how well the methods estimated the party's Absentee vote distribution, the party's Early vote distribution, and the party's overall vote distribution across the state (which includes the Election-Day votes). It is these overall party vote distributions that are likely to be of more importance in most redistricting contexts.

For each figure we provide heatmaps for the R^2 values, where values closer to 1 indicate stronger estimates, and root mean square error (rmse), where values closer to 0 indicate smaller error. In each table the darker shades correspond to better estimates of the ground truth.

Comparing trends across all tables gives a fuller picture of which methods tend to perform well in general. Row-to-row differences are prominent but are not the focus of this analysis. Instead, we ask *within* each row which column(s) perform well and then zoom out to visualize how the relative performance extends to larger trends.

We see that the 'By Election-Day' and the various Voterfile methods tend to perform well whereas 'Uniform' and 'By Area' perform poorly. For partisan evaluation in particular, 'By Election-Day', 'By Voterfile Mode and Party', and 'By Voterfile Party' outperform the others.



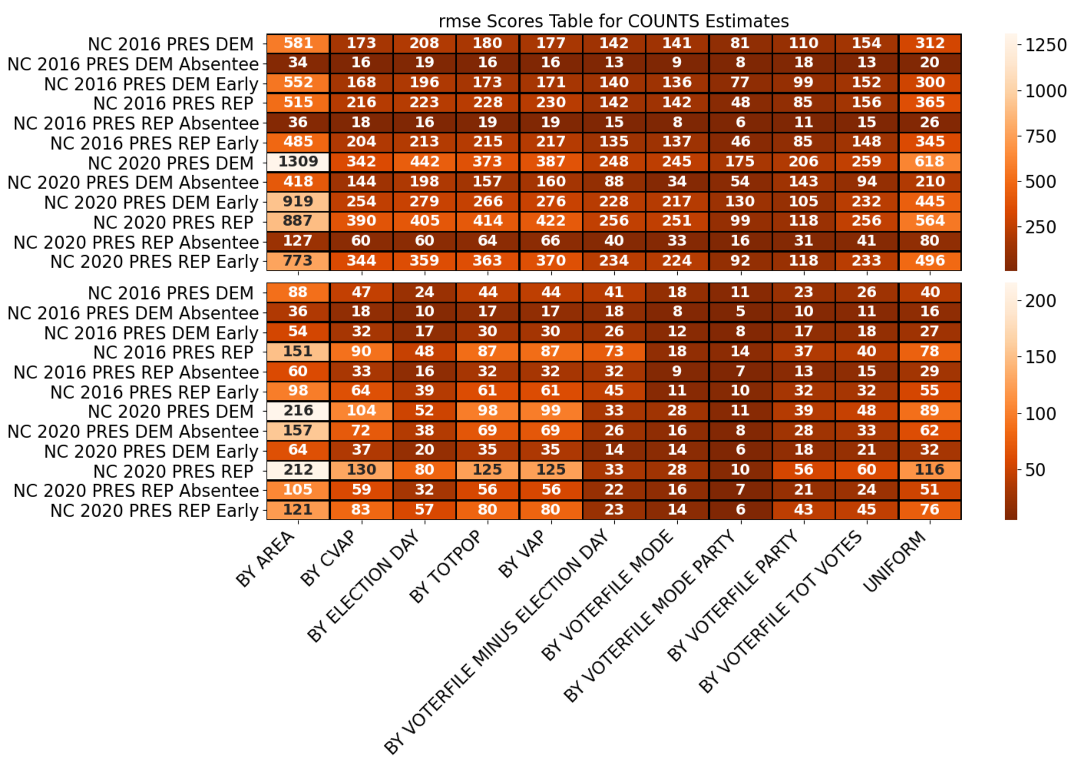


Figure 6. Counts-based precinct-level evaluations

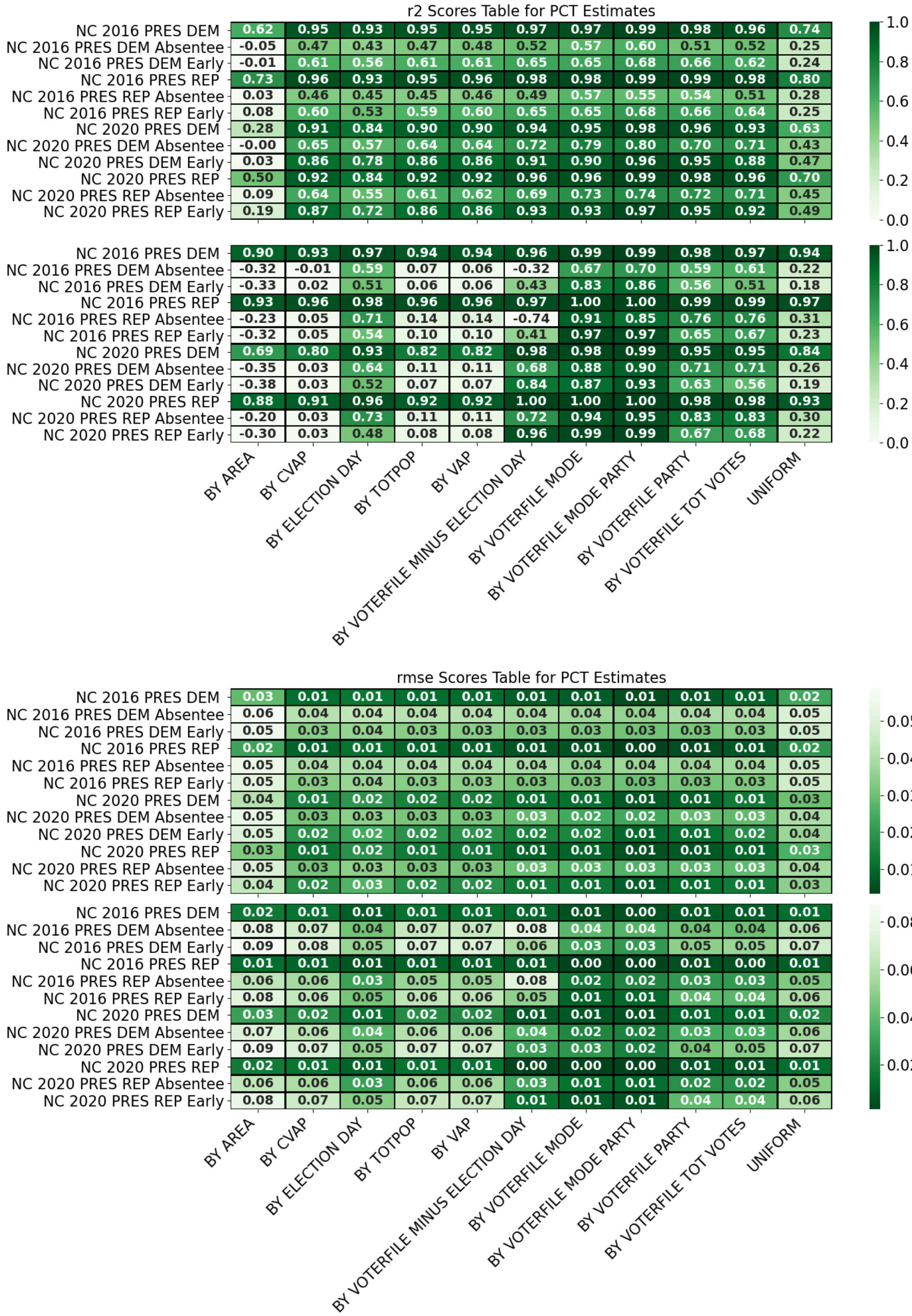


Figure 7. Percents-based precinct-level evaluations

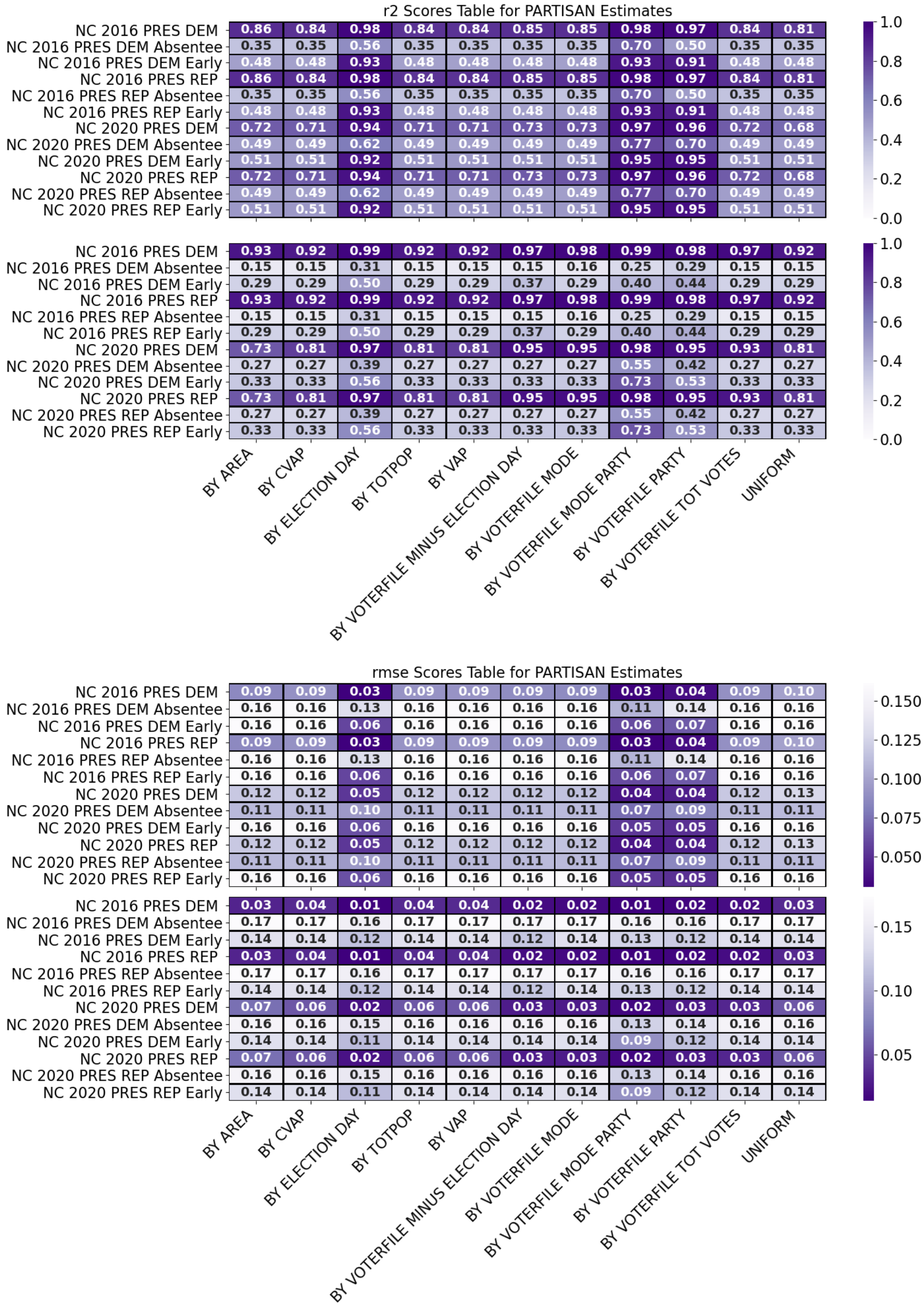


Figure 8. Precinct-level partisan preservation evaluations

3.3.2 Aggregate Performance

Again, an ideal disaggregation method would accurately estimate votes at the precinct level, but in most redistricting contexts, the more important question is how well these methods perform at the district level. Properly assessing redistricting criteria such as partisan fairness or VRA compliance for a newly proposed plan requires having accurate district-level vote totals. Some disaggregation methods may have relatively noisy precinct-level estimates but strong district-level estimates if the error of neighboring precincts balances out well. Alternatively, even small amounts of error at the precinct level can be amplified at the district level, depending on the geographic distribution of the estimation error across a county.

Notably, when re-aggregated to the county level, vote estimates should be preserved *exactly*, since these totals are already known for the county. Districts that are wholly comprised of counties will not be affected by this disaggregation error, but plans with many county splits or states with very large counties that need to be divided into many districts will have at least some degree of disaggregation error.

Figure 9 compares the estimated district-level Biden two-way share of 2020 Presidential votes across disaggregation methods for 30 of North Carolina's 120 State House districts. The leftmost column shows the ground truth target percentages for these districts. The other columns are estimates derived by different disaggregation methods. Here we can ask which column best approximates the leftmost column?

Several of the rows (e.g. Districts 1, 5, 6, and 13) are identical across the columns. These are districts that are comprised of whole counties. A few rows (e.g. Districts 3 and 16) are not identical but are fairly consistent across methods. The other districts were chosen for inclusion because they show how much variation can occur across methods.

Figure 10 shows the partisan share residuals from Figure 9. That is, the predicted two-way Biden share minus the ground truth two-way Biden share for each district. Redder cells indicate underestimates of the true two-way Biden share and greener cells indicate overestimates of the true two-way Biden share.

Consistently, the 'By Election-Day', 'By Voterfile Party', and 'By Voterfile Mode and Party' methods most closely align with the ground truth across districts than the other disaggregation methods. These are the three disaggregation methods that account for differences in estimated vote distributions across parties.

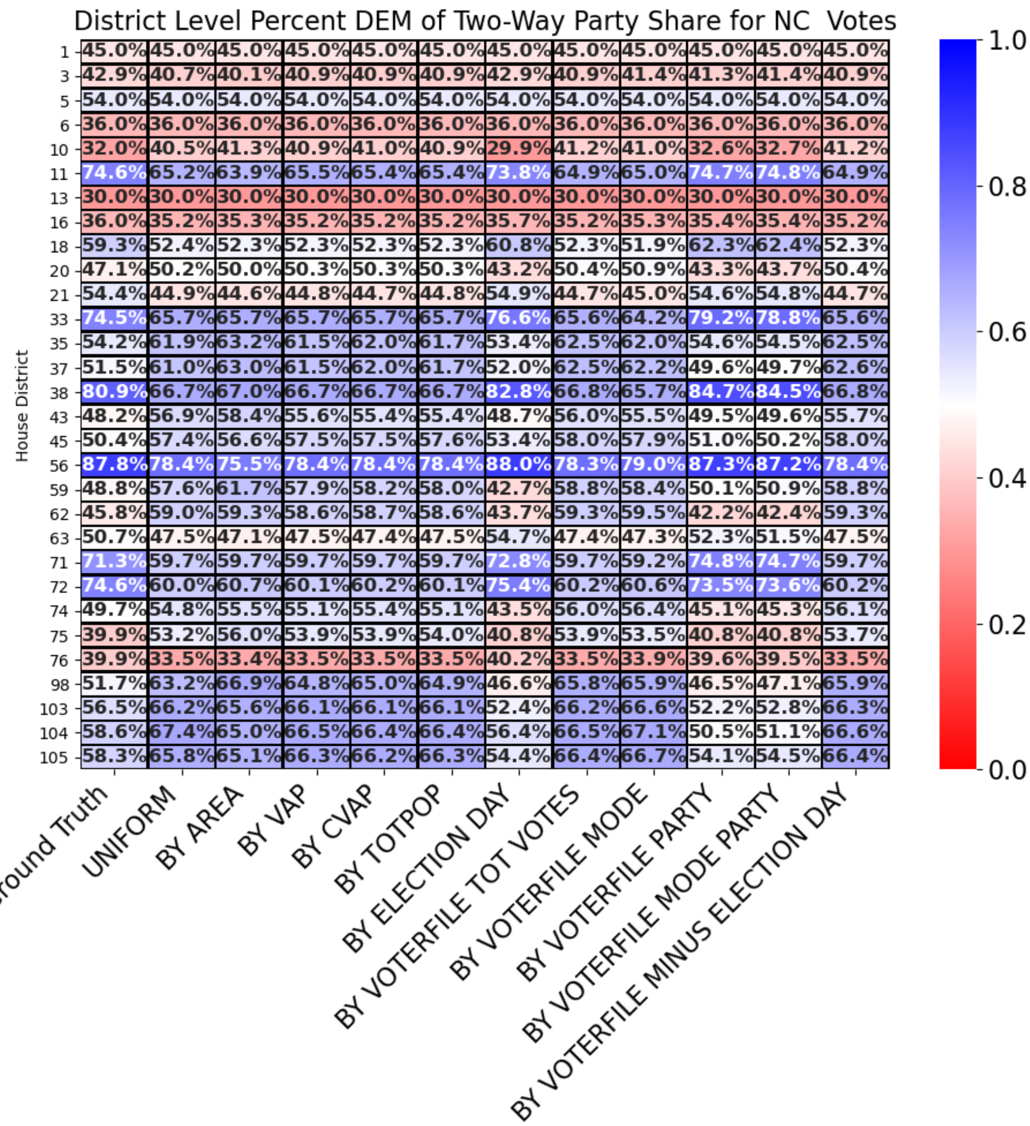


Figure 9. Biden share of major party votes in 2020 Election for 30 different North Carolina State House Districts. The left column shows the observed (ground truth) percentages in these districts and the other columns show disaggregation estimates.

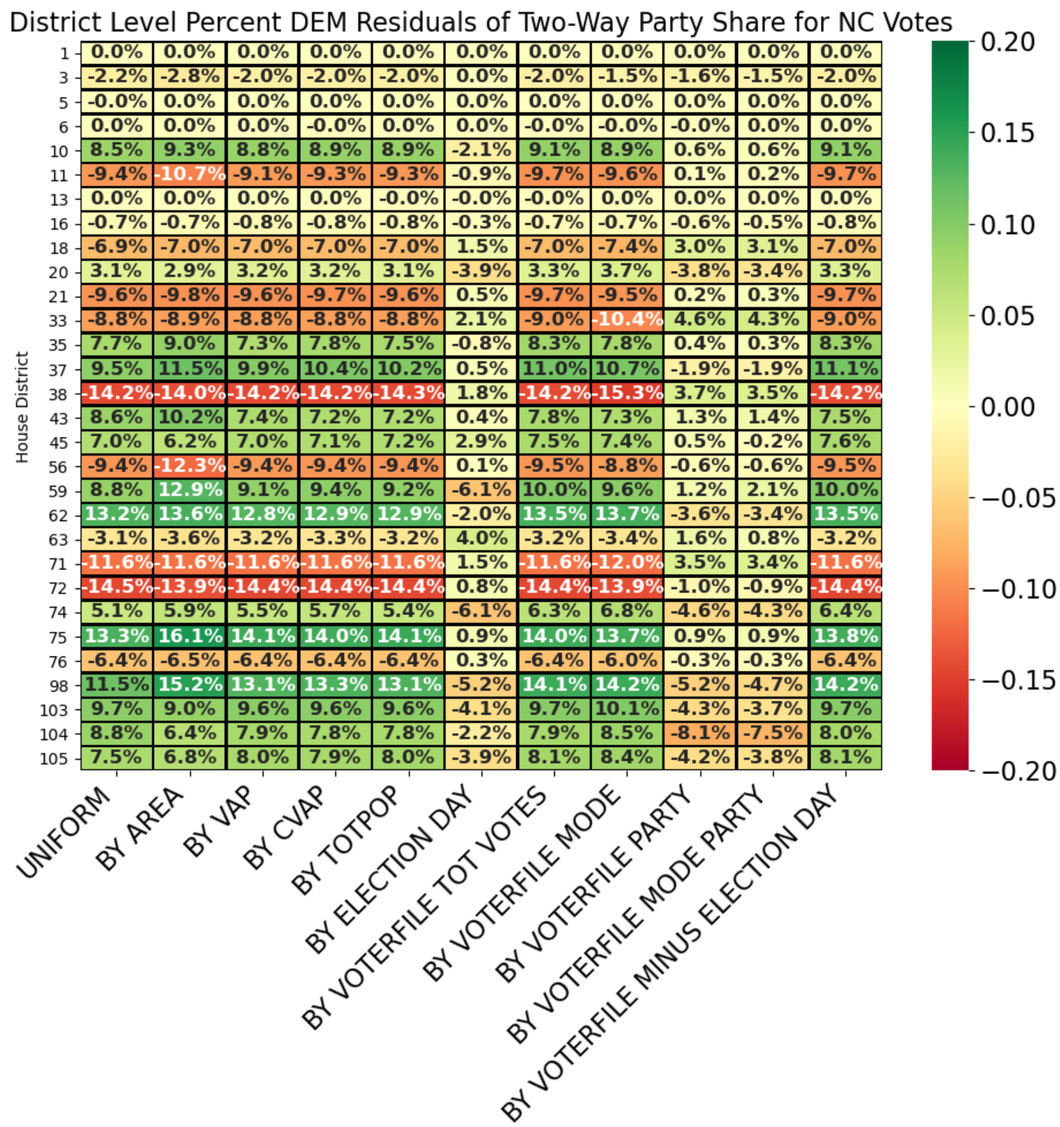


Figure 10. Residuals from Figure 9 measured as: predicted partisan share minus ground truth for each method. Each column shows a different disaggregation method. Redder cells indicate districts for which the Biden share of major party votes in the 2020 Election was *underestimated* by its column's method and greener cells indicate districts for which the Biden share was *overestimated* by its column's method.

4 Conclusion

Redistricting analysis requires accurate voting data at the district building-block level. However, some votes are only reported in aggregate at the county level (e.g. Absentee Mail-In and Early votes), so practitioners must use county-to-precinct disaggregation methods to estimate how these votes should be allocated to precincts across each county.

We compared 11 different disaggregation methods using the 2016 and 2020 Presidential elections in North Carolina and Oklahoma—two states whose election data and voterfile availability facilitate the comparison of disaggregation estimates to ground truth values.

Across the board, disaggregating by a weighting scheme based on the voterfile's joint distribution of voters *by voting mode and party registration* consistently outperforms the other methods. Disaggregating based on voterfile mode alone provides strong estimates for the distribution of vote totals and percentages across a county, and disaggregating based on voterfile party alone preserves partisan balance well. In the absence of a (good quality) voterfile, disaggregating based on the distribution of election-day votes outperforms other voterfile-less disaggregation techniques, especially when estimating partisan balance.

A Method Details and Examples

In this Appendix, we walk through a small toy example to show how the disaggregation metho- calculations are carried out.

Our example consists of a single county, C, with three precincts, P_1, P_2, and P_3 (shown in Figure 11) for a single two-party election. Weighting schemes for an entire state are calculated independently for each county, year, and election, so it suffices to show these calculations for a single county and election.

The data required to calculate the weighting schemes for each disaggregation method are given in three tables: Table 2 shows the tabular data attached to the precinct shapefile, Table 3 shows the election returns for County C, and Table 4 shows the voterfile data for County C.

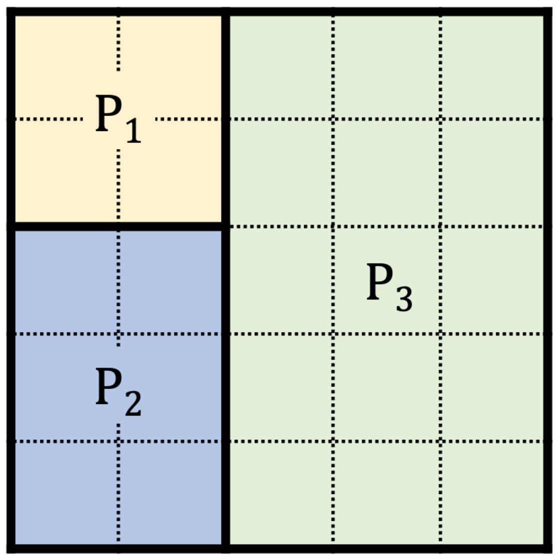


Figure 11. Geography of the three precincts of County C

County	Precinct	TOTPOP	VAP	CVAP	AREA
C	P1	15000	13000	11000	4
C	P2	9000	8000	6000	6
C	P3	6000	4000	3000	15
TOTALS:		30000	25000	20000	25

Table 2. Shapefile data for County C

For each disaggregation method, we show how to generate the weighting scheme(s) and how that would be applied to determine precinct-level vote estimates by mode and party. We use w_1, w_2, and w_3 to denote the weights of Precincts P_1, P_2, and P_3, respectively, in each weighting scheme.

County	Precinct	All Republican Votes	All Democrat Votes	Republican, Election-Day Votes	Democrat, Election-Day Votes	Republican, Absentee Votes	Democrat, Absentee Votes	Republican, Early Votes	Democrat, Early Votes
C	P1	?	?	900	600	?	?	?	?
C	P2	?	?	900	100	?	?	?	?
C	P3	?	?	1200	300	?	?	?	?
TOTALS:		4500	5500	3000	1000	500	1500	1000	3000

Table 3. Election return data for County C

County	Precinct	VOTED	VOTED: Registered Republican	VOTED: Registered Democrat	VOTED: Election-Day	VOTED: Absentee	VOTED: Early	VOTED: Election-Day, Registered Republican	VOTED: Election-Day, Registered Democrat	VOTED: Absentee, Registered Republican	VOTED: Absentee, Registered Democrat	VOTED: Early, Registered Republican	VOTED: Early, Registered Democrat
C	P1	4700	2000	2700	1500	1200	2000	500	1000	500	700	1000	1000
C	P2	2900	1500	1400	1000	400	1500	500	500	300	100	700	800
C	P3	2400	1500	900	1500	400	500	1000	500	200	200	300	200
TOTALS:		10000	5000	5000	4000	2000	4000	2000	2000	1000	1000	2000	2000

Table 4. Voterfile data for County C

A.1 Sample Disaggregation Calculations

1. **Uniformly**

- **Definition:** Distribute votes evenly to every precinct in the county.

$$w_P = \frac{1}{\# \, precincts(C)}$$

- **Example:**

	P_1	P_2	P_3	C **Totals**
Weights	$w_1 = 1/3$ = **0.33**	$w_2 = 1/3$ = **0.33**	$w_3 = 1/3$ = **0.33**	1
Absentee Republican Vote Estimates	$500 \times w_1 = 500 \times 0.33$ = **166.67**	$500 \times w_2 = 500 \times 0.33$ = **166.67**	$500 \times w_3 = 500 \times 0.33$ = **166.67**	500
Absentee Democrat Vote Estimates	$1500 \times w_1 = 1500 \times 0.33$ = **500**	$1500 \times w_2 = 1500 \times 0.33$ = **500**	$1500 \times w_3 = 1500 \times 0.33$ = **500**	1500
Early Republican Vote Estimates	$1000 \times w_1 = 1000 \times 0.33$ = **333.33**	$1000 \times w_2 = 1000 \times 0.33$ = **333.33**	$1000 \times w_3 = 1000 \times 0.33$ = **333.33**	1000
Early Democrat Vote Estimates	$3000 \times w_1 = 3000 \times 0.33$ = **1000**	$3000 \times w_2 = 3000 \times 0.33$ = **1000**	$3000 \times w_3 = 3000 \times 0.33$ = **1000**	3000

- **Data Used:** Number of Precincts in County C and vote TOTALS from Table 3
- **Notes:** This method gives the same weighting scheme across the county for each candidate and voting mode. It places equal votes in each of a county's precincts.

2. **By Area**

- **Definition:** Distribute votes proportionally to the precinct's share of the county's area.

$$w_P = \frac{area(P)}{area(C)}$$

- **Example:**

	P_1	P_2	P_3	C Totals
Weights	w_1 = 4/25 = **0.16**	w_2 = 6/25 = **0.24**	w_3 = 15/25 = **0.6**	1
Absentee Republican Vote Estimates	500 × w_1 = 500 × 0.16 = **80**	500 × w_2 = 500 × 0.24 = **120**	500 × w_3 = 500 × 0.6 = **300**	500
Absentee Democrat Vote Estimates	1500 × w_1 = 1500 × 0.16 = **240**	1500 × w_2 = 1500 × 0.24 = **360**	1500 × w_3 = 1500 × 0.6 = **900**	1500
Early Republican Vote Estimates	1000 × w_1 = 1000 × 0.16 = **160**	1000 × w_2 = 1000 × 0.24 = **240**	1000 × w_3 = 1000 × 0.6 = **600**	1000
Early Democrat Vote Estimates	3000 × w_1 = 3000 × 0.16 = **480**	3000 × w_2 = 3000 × 0.24 = **720**	3000 × w_3 = 3000 × 0.6 = **1800**	3000

- **Data Used:** AREA column of Table 2 and vote TOTALS from Table 3
- **Notes:** This method gives the same weighting scheme across the county for each candidate and voting mode. It places more votes in a county's larger-area precincts.

3. **By Total Population**

- **Definition:** Distribute votes proportionally to the precinct's share of the county's total population.

$$w_P = \frac{population(P)}{population(C)}$$

- **Example:**

	P_1	P_2	P_3	C Totals
Weights	w_1 = 15000/30000 = **0.5**	w_2 = 9000/30000 = **0.3**	w_3 = 6000/30000 = **0.2**	1
Absentee Republican Vote Estimates	500 × w_1 = 500 × 0.5 = **250**	500 × w_2 = 500 × 0.3 = **150**	500 × w_3 = 500 × 0.2 = **100**	500
Absentee Democrat Vote Estimates	1500 × w_1 = 1500 × 0.5 = **750**	1500 × w_2 = 1500 × 0.3 = **450**	1500 × w_3 = 1500 × 0.2 = **300**	1500
Early Republican Vote Estimates	1000 × w_1 = 1000 × 0.5 = **500**	1000 × w_2 = 1000 × 0.3 = **300**	1000 × w_3 = 1000 × 0.2 = **200**	1000
Early Democrat Vote Estimates	3000 × w_1 = 3000 × 0.5 = **1500**	3000 × w_2 = 3000 × 0.3 = **900**	3000 × w_3 = 3000 × 0.2 = **600**	3000

- **Data Used:** TOTPOP column of Table 2 and vote TOTALS from Table 3
- **Notes:** This method gives the same weighting scheme across the county for each candidate and voting mode. It places more votes in the county's most populated precincts.

4. **By Voting-Aged Population (VAP)**

- **Definition:** Distribute votes proportionally to the precinct's share of the county's voting-aged population.

$$w_P = \frac{VAP(P)}{VAP(C)}$$

- **Example:**

	P_1	P_2	P_3	C **Totals**
Weights	w_1 = 13000/25000 = **0.52**	w_2 = 8000/25000 = **0.32**	w_3 = 4000/25000 = **0.16**	1
Absentee Republican Vote Estimates	500 × w_1 = 500 × 0.52 = **260**	500 × w_2 = 500 × 0.32 = **160**	500 × w_3 = 500 × 0.16 = **80**	500
Absentee Democrat Vote Estimates	1500 × w_1 = 1500 × 0.52 = **780**	1500 × w_2 = 1500 × 0.32 = **480**	1500 × w_3 = 1500 × 0.16 = **240**	1500
Early Republican Vote Estimates	1000 × w_1 = 1000 × 0.52 = **520**	1000 × w_2 = 1000 × 0.32 = **320**	1000 × w_3 = 1000 × 0.16 = **160**	1000
Early Democrat Vote Estimates	3000 × w_1 = 3000 × 0.52 = **1560**	3000 × w_2 = 3000 × 0.32 = **960**	3000 × w_3 = 3000 × 0.16 = **480**	3000

- **Data Used:** VAP column of Table 2 and vote TOTALS from Table 3
- **Notes:** This method gives the same weighting scheme across the county for each candidate and voting mode.

5. **By Citizen Voting-Aged Population (CVAP)**

- **Definition:** Distribute votes proportionally to the precinct's share of the county's citizen voting-aged population.

$$w_P = \frac{CVAP(P)}{CVAP(C)}$$

- **Example:**

	P_1	P_2	P_3	C **Totals**
Weights	w_1 = 11000/20000 = **0.55**	w_2 = 6000/20000 = **0.3**	w_3 = 3000/20000 = **0.15**	1
Absentee Republican Vote Estimates	500 × w_1 = 500 × 0.55 = **275**	500 × w_2 = 500 × 0.3 = **150**	500 × w_3 = 500 × 0.15 = **75**	500
Absentee Democrat Vote Estimates	1500 × w_1 = 1500 × 0.55 = **825**	1500 × w_2 = 1500 × 0.3 = **450**	1500 × w_3 = 1500 × 0.15 = **225**	1500
Early Republican Vote Estimates	1000 × w_1 = 1000 × 0.55 = **550**	1000 × w_2 = 1000 × 0.3 = **300**	1000 × w_3 = 1000 × 0.15 = **150**	1000
Early Democrat Vote Estimates	3000 × w_1 = 3000 × 0.55 = **1650**	3000 × w_2 = 3000 × 0.3 = **900**	3000 × w_3 = 3000 × 0.15 = **450**	3000

- **Data Used:** CVAP column of Table 2 and vote TOTALS from Table 3
- **Notes:** This method gives the same weighting scheme across the county for each candidate and voting mode.

6. **By Election-Day Votes**

- **Definition:** Distribute each candidate's votes proportionally to the precinct's share of the county's in-person, election-day votes for that candidate.

$$\text{(for Candidate 1)} \quad w_P = \frac{\text{Election-Day votes for Candidate 1 in Precinct } P}{\text{Election-Day votes for Candidate 1 in County } C}$$

$$\text{(for Candidate 2)} \quad w_P = \frac{\text{Election-Day votes for Candidate 2 in Precinct } P}{\text{Election-Day votes for Candidate 2 in County } C}$$

$$\vdots$$

- **Example:**

	P_1	P_2	P_3	C **Totals**
Weights (Republican)	w_1 = 900/3000 = **0.3**	w_2 = 900/3000 = **0.3**	w_3 = 1200/3000 = **0.4**	1
Absentee Republican Vote Estimates	500 × w_1 = 500 × 0.3 = **150**	500 × w_2 = 500 × 0.3 = **150**	500 × w_3 = 500 × 0.4 = **200**	500
Early Republican Vote Estimates	1000 × w_1 = 1000 × 0.3 = **300**	1000 × w_2 = 1000 × 0.3 = **300**	1000 × w_3 = 1000 × 0.4 = **400**	1000
Weights (Democrat)	w_1 = 600/1000 = **0.6**	w_2 = 100/1000 = **0.1**	w_3 = 300/1000 = **0.3**	1
Absentee Democrat Vote Estimates	1500 × w_1 = 1500 × 0.6 = **900**	1500 × w_2 = 1500 × 0.1 = **150**	1500 × w_3 = 1500 × 0.3 = **450**	1500
Early Democrat Vote Estimates	3000 × w_1 = 3000 × 0.6 = **1800**	3000 × w_2 = 3000 × 0.1 = **300**	3000 × w_3 = 3000 × 0.3 = **900**	3000

- **Data Used:** Republican Election-Day Votes column from Table 3 for weights for Republican weights and estimates, Democrat Election-Day Votes column from Table 3 for weights for Democrat weights and estimates, and vote TOTALS from Table 3
- **Notes:** This method gives the same weighting scheme across the county for each voting mode, but different weighting schemes for each candidate.

7. **By Voterfile Total Voters**
   - **Definition:** Distribute votes proportionally to the precinct's share of the county's total voters according to the voterfile.

$$w_P = \frac{Voted(P)}{Voted(C)}$$

   - **Example:**
   - **Data Used:** VOTED column of Table 4 and vote TOTALS from Table 3
   - **Notes:** This method gives the same weighting scheme across the county for each candidate and voting mode.

8. **By Voterfile Total Voters Minus Election-Day Voters**
   - **Definition:** Distribute votes proportionally to the precinct's share of the county's total voters according to the voterfile minus the number of election-day voters.

$$w_P = \frac{\text{Voterfile Voters in Precinct } P - \text{Election-Day Voters in Precinct } P}{\text{Voterfile Voters in County } C - \text{Election-Day Voters in County } C}$$

	P_1	P_2	P_3	C **Totals**
Weights	w_1 = 4700/10000 = **0.47**	w_2 = 2900/10000 = **0.29**	w_3 = 2400/10000 = **0.24**	1
Absentee Republican Vote Estimates	500 × w_1 = 500 × 0.47 = **235**	500 × w_2 = 500 × 0.29 = **145**	500 × w_3 = 500 × 0.24 = **120**	500
Absentee Democrat Vote Estimates	1500 × w_1 = 1500 × 0.47 = **705**	1500 × w_2 = 1500 × 0.29 = **435**	1500 × w_3 = 1500 × 0.24 = **360**	1500
Early Republican Vote Estimates	1000 × w_1 = 1000 × 0.47 = **470**	1000 × w_2 = 1000 × 0.29 = **290**	1000 × w_3 = 1000 × 0.24 = **240**	1000
Early Democrat Vote Estimates	3000 × w_1 = 3000 × 0.47 = **1410**	3000 × w_2 = 3000 × 0.29 = **870**	3000 × w_3 = 3000 × 0.24 = **720**	3000

- **Example:**

	P_1	P_2	P_3	C **Totals**
Weights	w_1 = (4700-1500)/(10000-4000) = **0.533**	w_2 = (2900-1000)/(10000-4000) = **0.317**	w_3 = (2400-1500)/(10000-4000) = **0.15**	1
Absentee Republican Vote Estimates	500 × w_1 = 500 × 0.533 = **266.67**	500 × w_2 = 500 × 0.317 = **158.33**	500 × w_3 = 500 × 0.15 = **75**	500
Absentee Democrat Vote Estimates	1500 × w_1 = 1500 × 0.533 = **800**	1500 × w_2 = 1500 × 0.317 = **475**	1500 × w_3 = 1500 × 0.15 = **225**	1500
Early Republican Vote Estimates	1000 × w_1 = 1000 × 0.533 = **533.33**	1000 × w_2 = 1000 × 0.317 = **316.67**	1000 × w_3 = 1000 × 0.15 = **150**	1000
Early Democrat Vote Estimates	3000 × w_1 = 3000 × 0.533 = **1600**	3000 × w_2 = 3000 × 0.317 = **950**	3000 × w_3 = 3000 × 0.15 = **450**	3000

- **Data Used:** VOTED column of Table 4, Republican Election-Day Votes and Democrat Election-Day Votes columns (summed) from Table 3, and vote TOTALS from Table 3
- **Notes:** This method gives the same weighting scheme across the county for each candidate and voting mode.

9. **By Voterfile Mode**

- **Definition:** Distribute votes for each voting mode proportionally to the precinct's share of the county's voters *by that mode* according to the voterfile.

$$\text{(for Absentee Voters)} \quad w_P = \frac{\text{Voterfile Absentee Voters in Precinct } P}{\text{Voterfile Absentee Voters in County } C}$$

$$\text{(for Early Voters)} \quad w_P = \frac{\text{Voterfile Early Voters in Precinct } P}{\text{Voterfile Early Voters in County } C}$$

- **Example:**
- **Data Used:** VOTED: Absentee and VOTED: Early columns of Table 4 and vote TOTALS from Table 3
- **Notes:** This method gives the same weighting scheme across the county for each candidate but different weighting schemes for each voting mode.

10. **By Voterfile Party**

	P_1	P_2	P_3	C Totals
Weights (Absentee)	w_1 = 1200/2000 = **0.6**	w_2 = 400/2000 = **0.2**	w_3 = 400/2000 = **0.2**	1
Absentee Republican Vote Estimates	500 × w_1 = 500 × 0.6 = **300**	500 × w_2 = 500 × 0.2 = **100**	500 × w_3 = 500 × 0.2 = **100**	500
Absentee Democrat Vote Estimates	1500 × w_1 = 1500 × 0.6 = **900**	1500 × w_2 = 1500 × 0.2 = **300**	1500 × w_3 = 1500 × 0.2 = **300**	1500
Weights (Early)	w_1 = 2000/4000 = **0.5**	w_2 = 1500/4000 = **0.375**	w_3 = 500/4000 = **0.125**	1
Early Republican Vote Estimates	1000 × w_1 = 1000 × 0.5 = **500**	1000 × w_2 = 1000 × 0.375 = **375**	1000 × w_3 = 1000 × 0.125 = **125**	1000
Early Democrat Vote Estimates	3000 × w_1 = 3000 × 0.5 = **1500**	3000 × w_2 = 3000 × 0.375 = **1125**	3000 × w_3 = 3000 × 0.125 = **375**	3000

- **Definition:** Distribute votes for each candidate proportionally to the precinct's share of the county's voters *registered to the candidate's party* according to the voterfile.

$$\text{(for Candidate 1)} \quad w_P = \frac{\text{Voterfile Voters Registered to Candidate 1's Party in Precinct } P}{\text{Voterfile Voters Registered to Candidate 1's Party in County } C}$$

$$\text{(for Candidate 2)} \quad w_P = \frac{\text{Voterfile Voters Registered to Candidate 2's Party in Precinct } P}{\text{Voterfile Voters Registered to Candidate 2's Party in County } C}$$

$$\vdots$$

- **Example:**

	P_1	P_2	P_3	C Totals
Weights (Republican)	w_1 = 2000/5000 = **0.4**	w_2 = 1500/5000 = **0.3**	w_3 = 1500/5000 = **0.3**	1
Absentee Republican Vote Estimates	500 × w_1 = 500 × 0.4 = **200**	500 × w_2 = 500 × 0.3 = **150**	500 × w_3 = 500 × 0.3 = **150**	500
Early Republican Vote Estimates	1000 × w_1 = 1000 × 0.4 = **400**	1000 × w_2 = 1000 × 0.3 = **300**	1000 × w_3 = 1000 × 0.3 = **300**	1000
Weights (Democrat)	w_1 = 2700/5000 = **0.54**	w_2 = 1400/5000 = **0.28**	w_3 = 900/5000 = **0.18**	1
Absentee Democrat Vote Estimates	1500 × w_1 = 1500 × 0.54 = **810**	1500 × w_2 = 1500 × 0.28 = **420**	1500 × w_3 = 1500 × 0.18 = **270**	1500
Early Democrat Vote Estimates	3000 × w_1 = 3000 × 0.54 = **1620**	3000 × w_2 = 3000 × 0.28 = **840**	3000 × w_3 = 3000 × 0.18 = **540**	3000

- **Data Used:** VOTED: Registered Republican column of Table 4 for Republican weights and estimates, VOTED: Registered Democrat column of Table 4 for Democrat weights and estimates and vote TOTALS from Table 3
- **Notes:** This method gives the same weighting scheme across the county for each voting mode but different weighting schemes for each candidate, though different candidates from the same *party* will have the same weighting scheme.

11. **By Voterfile Mode and Party**

- **Definition:** Distribute votes for each candidate and voting mode proportionally to the precinct's share of the county's voters *that are registered to the candidate's party and voted via that mode* according to the voterfile.

$$\text{(for Candidate 1 Absentee Voters)} \quad w_P = \frac{\text{Voterfile Absentee Voters Registered to Candidate 1's Party in Precinct } P}{\text{Voterfile Absentee Voters Registered to Candidate 1's Party in County } C}$$

$$\text{(for Candidate 1 Early Voters)} \quad w_P = \frac{\text{Voterfile Early Voters Registered to Candidate 1's Party in Precinct } P}{\text{Voterfile Early Voters Registered to Candidate 1's Party in County } C}$$

$$\text{(for Candidate 2 Absentee Voters)} \quad w_P = \frac{\text{Voterfile Absentee Voters Registered to Candidate 2's Party in Precinct } P}{\text{Voterfile Absentee Voters Registered to Candidate 2's Party in County } C}$$

$$\text{(for Candidate 2 Early Voters)} \quad w_P = \frac{\text{Voterfile Early Voters Registered to Candidate 2's Party in Precinct } P}{\text{Voterfile Early Voters Registered to Candidate 2's Party in County } C}$$

$$\vdots$$

- **Example:**

	P_1	P_2	P_3	C **Totals**
Weights (Absentee Republican)	w_1 = 500/1000 = **0.5**	w_2 = 300/1000 = **0.3**	w_3 = 200/1000 = **0.2**	1
Absentee Republican Vote Estimates	500 × w_1 = 500 × 0.5 = **250**	500 × w_2 = 500 × 0.3 = **150**	500 × w_3 = 500 × 0.2 = **100**	500
Weights (Early Republican)	w_1 = 1000/2000 = **0.5**	w_2 = 700/2000 = **0.35**	w_3 = 300/2000 = **0.15**	1
Early Republican Vote Estimates	1000 × w_1 = 1000 × 0.5 = **500**	1000 × w_2 = 1000 × 0.35 = **350**	1000 × w_3 = 1000 × 0.15 = **150**	1000
Weights (Absentee Democrat)	w_1 = 700/1000 = **0.7**	w_2 = 100/1000 = **0.1**	w_3 = 200/1000 = **0.2**	1
Absentee Democrat Vote Estimates	1500 × w_1 = 1500 × 0.7 = **1050**	1500 × w_2 = 1500 × 0.1 = **150**	1500 × w_3 = 1500 × 0.2 = **300**	1500
Weights (Early Democrat)	w_1 = 1000/2000 = **0.5**	w_2 = 800/2000 = **0.4**	w_3 = 200/2000 = **0.1**	1
Early Democrat Vote Estimates	3000 × w_1 = 3000 × 0.5 = **1500**	3000 × w_2 = 3000 × 0.4 = **1200**	3000 × w_3 = 3000 × 0.1 = **300**	3000

- **Data Used:** VOTED: Absentee, Registered Republican column of Table 4 for Absentee Republican weights and estimates, VOTED: Early, Registered Republican column of Table 4 for Early Republican weights and estimates, VOTED: Absentee, Registered Democrat column of Table 4 for Absentee Democrat weights and estimates, VOTED: Early, Registered Democrat column of Table 4 for Early Democrat weights and estimates, and vote TOTALS from Table 3
- **Notes:** This method gives different weighting schemes for each candidate and each mode, though candidates from the same party will have the same weighting scheme.