Supporting Information for The Measurement of Partisan Sorting for 180 Million Voters

Jacob R. Brown*and Ryan D. Enos†

1 Data Processing

The analysis in this manuscript relies on the spatial processing technique of Geohashing. Geohash is a public domain encoding system that stores spatial location data in strings of letters and numbers. This technique is an application of z-order curves, transforming multi-dimensional data to a single dimension to allow for more efficient processing of such data Niemeyer (2008). These techniques were applied to the latitude and longitude coordinates of the residence of each registered voter in the United States, so that k-means analysis can identify the nearest neighbors and measure distances between neighbors for each voter in the nationwide voterfile. This data processing plan was developed in collaboration with the Harvard University's Center for Geographic Analysis (CGA), and CGA implemented the nearest neighbors analysis and distance calculations. The processing was implemented on Amazon Web Services (AWS) across 10 Postgres instances with PostGIS add-ons. Processing of the entire file in this setup took approximately 4 weeks in continuous computation time. The output was dataset of 1000 x 180,735,645 rows listing the $1,000$ nearest neighbors of

*Institute for Quantitative Social Science and Department of Government, Harvard University, jrbrown@g.harvard.edu

†Institute for Quantitative Social Science and Department of Government, Harvard University, renos@gov.harvard.edu

each voter, with columns for the neighbor's partisanship and the distance they live from the voter.

With these data in hand, we calculated weighted averages of Spatial Partisan exposure and Isolation and other partisan segregation, racial segregation, and general summary statistics for each voter in the file. This stage of the analysis was also conducted on AWS, using an instance calibrated for parallel processing in `R`. Total computation time for this stage was over 200 hours, completed in multiple installments.

2 Spatial versus Aspatial Measures

Here, we present comparison statistics of weighted versus unweighted partisan exposure measures. In our analyses, our primary measure of partisan exposure and isolation are averages of exposure to neighbors of each party in a voter's $1,000$ nearest neighbors, weighted by the inverse distance the voter lives from each neighbor. An exposure measure that does not weight for distance is making the strong assumption that distance does not matter, so that the spatial distribution of partisan within the $1,000$ nearest neighbors is uninformative as to partisan context. In Figure S1 we show the spatially weighted and unweighted distributions and in Figure S2 we show the nationwide distribution of the change in exposure for each individual voter when we weight by distance. In Figure S3 we show the absolute differences between spatial and aspatial measures. For many voters, the change is small, likely reflecting homogeneity within their 1,000 nearest neighbors. But for a large portion of voters, we see that, for many voters, not accounting for the spatial relationships between them and their neighbors significantly distorts the measurement of partisan exposure and isolation. For both parties, partisan isolation appears lower, and exposure appears higher, when distance is not incorporated, with fewer Democrats living in extreme isolation (> 0.95). The Republican distributions spreads with the incorporation of distance.

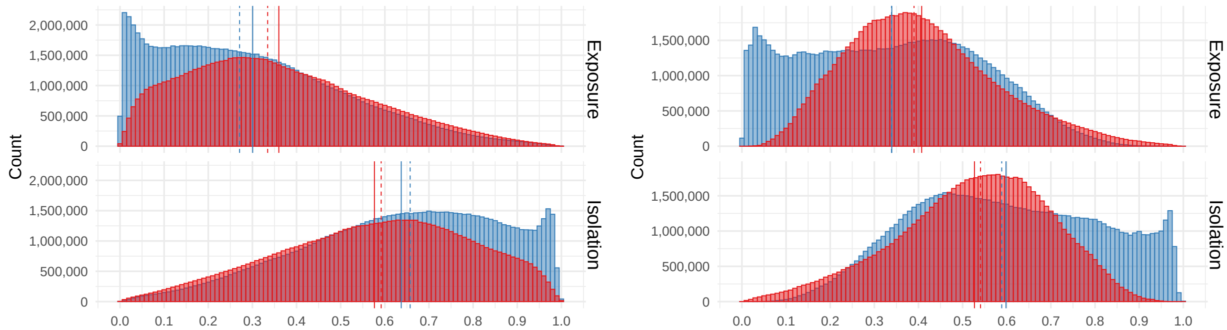


Figure S1: Spatial versus Aspatial Exposure/Isolation

Nationwide distribution of spatial (left) and aspatial (right) partisan isolation and exposure separately for Democrats (blue) and Republicans (red). Solid vertical lines represent mean values and dashed lines represent median values. The distributions are weighted by the posterior partisan probabilities.

(a) Percentage Point Change

(b) Percent Change

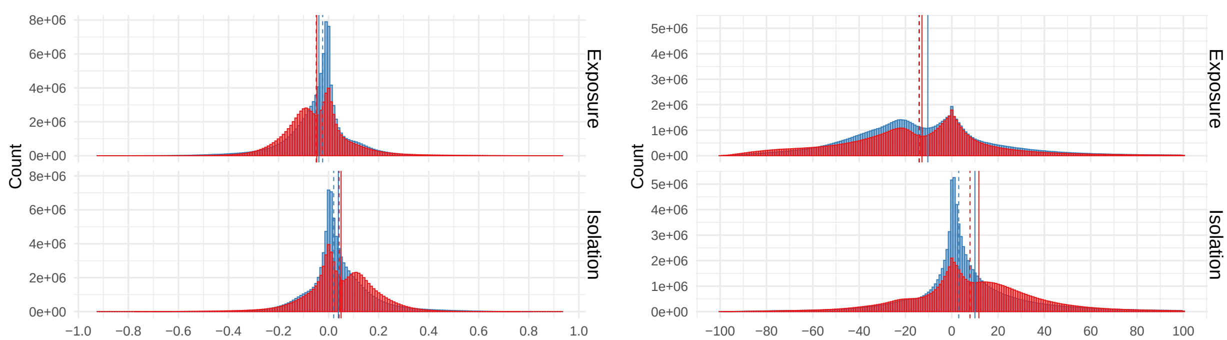


Figure S2: Individual Differences in Spatial versus Aspatial Exposure/Isolation

Nationwide distribution of individual-level changes in partisan Exposure and Isolation separately for Democrats (blue) and Republicans (red). The histograms on the left show the percentage point difference in spatial and aspatial exposure, while the histograms on the right show the percent change. Solid vertical lines represent mean values and dashed lines represent median values. The distributions are weighted by the posterior partisan probabilities.

(a) Percentage Point Change (b) Percent Change

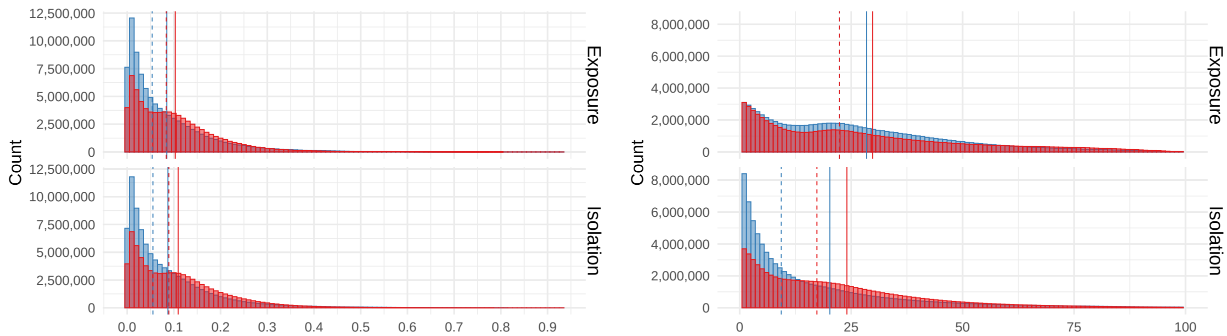


Figure S3: Individual Absolute Differences in Spatial versus Aspatial Exposure/Isolation

Nationwide distribution of individual-level absolute changes in partisan Exposure and Isolation separately for Democrats (blue) and Republicans (red). The histograms on the left show the percentage point absolute difference in spatial and aspatial exposure, while the histograms on the right show the absolute percent change. Solid vertical lines represent mean values and dashed lines represent median values. The distributions are weighted by the posterior partisan probabilities.

3 Nearest Neighbor Analysis up to 50,000 Neighbors

In our analysis we measure partisan residential context by determining the $1,000$ nearest neighbors to each voter. Prior to conducting this analysis, we took a random sample of $10,000$ voters and located their $50,000$ nearest neighbors, so as to determine the number of nearest neighbors at which adding more neighbors the the analysis is not informative as to a voer's Spatial Exposure and Isolation. For these $10,000$ voters, we calculated their Spatial Exposure and Isolation at varying levels of neighbors from 5 up to $50,000$. Figure S4 plots the distribution percentiles of Spatial Exposure and Isolation for this sample across number of neighbors. We see that the major changes in Spatial Exposure and Isolation that result from adding more neighbors to the analysis levels off after 1,000 neighbors. Above that, more neighbors add little information to our proximity-weighted measures.

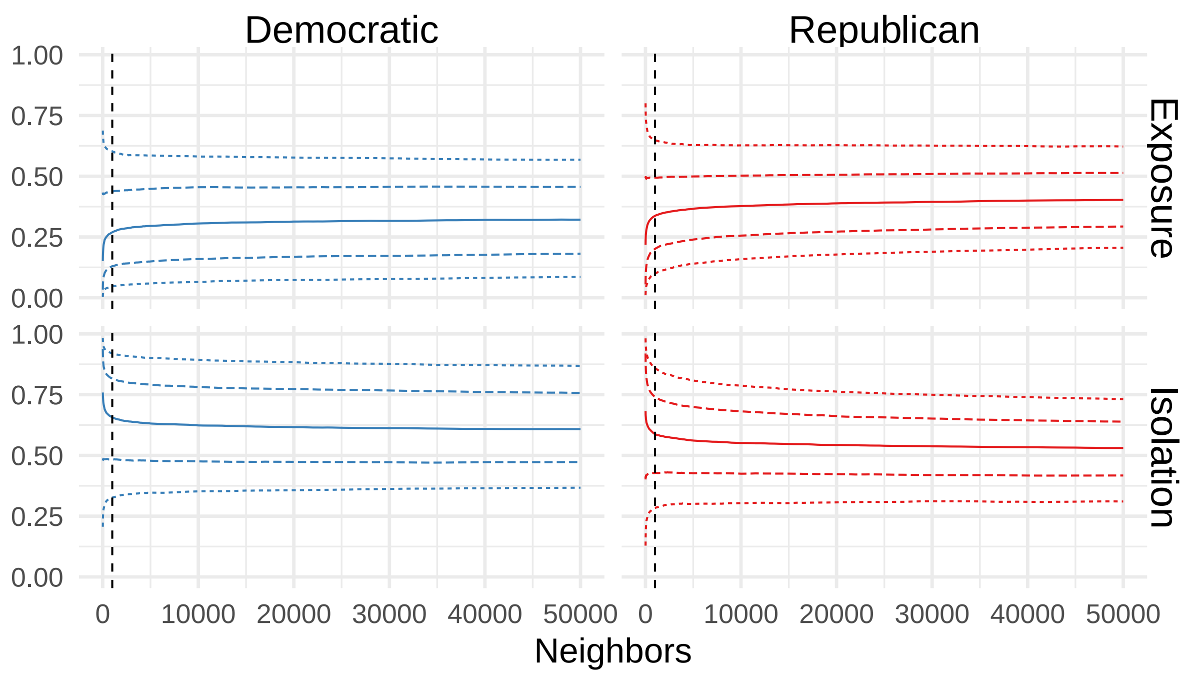


Figure S4: Change in Spatial Exposures by Number of Neighbors

Lines display the percentiles of of Democratic and Republican Spatial Exposure/Isolation at different levels of neighbors for the sample of 10,000 voters where we identified their 50,000 nearest neighbors. The solid lines represent the median or 50th percentile, the dashed line represent the 25th and 7th percentiles, and the dotted lines the 10th and 90th percentiles. The horizontal dashed black line represents the 1,000 neighbor mark. We see that the major changes in Spatial Exposure and Isolation that result from adding more neighbors to the analysis levels off after 1,000 neighbors. Above that, more neighbors add little information to our proximity-weighted measures.

4 Imputation of Partisanship

Our analyses rely on imputations of partisanship for voters not registered to a political party in the voterfile. We impute partisanship for voters who do not have partisanship explicitly recorded through a three-step process. First, we code remaining non-partisans as Democrat or Republican based on the last partisan primary in which they voted: most states that do not record partisanship — and some that do record partisanship — have open primaries, so voters can cast a ballot in either parties' primary, providing an indication of partisan attachment (Table S2). Next, we code individuals registered to third parties with clear ideological leans into the Democrat or Republican classification. Table S1 shows our classifications of each third party. Lastly, we impute partisanship for the remaining non-partisans through a Bayesian process combining the individual-level probability of different combinations of voter demographics conditional on partisanship based on a nationwide sample of voterfile validated respondents from the Cooperative Congressional Election study (CCES), with a geographic prior based on the 2016 vote share of the precinct in which the voter lives, net the Republican and Democratic counts in that precinct.

Precinct level vote shares were constructed from data provided by the MIT Election Lab and augmented with data collected from individual states. Using a series of spatial and tabular joins, we successfully merge 98.66% of voters with their precinct-level returns. There is some variation in the success of merging across states, ranging from 0% non-merged in several states to 13.01% in South Dakota, which is a state in which some counties do not have spatially defined precinct, but in which voters can vote in whatever voting precinct they choose. For voters for whom we could not put in a precinct, we instead constructed a geographic prior from county-level returns (Table S3).

To construct the individual level probability, we use the voterfile validated sample of CCES respondents (that is voters who, prior to anonymization, are matched to individuals on the voterfile so that aspects of their identity are confirmed against official records) and for four variables, race, age, gender, and turnout in the 2016 election, we calculate the probability of every unique combination, conditional on self-reported partisanship (Democrat, Republican, or independent, where "lean" Democrat or Republican were incorporated into Democrat or Republican). We coarsen age into quantiles (18-34, 35-50, 51-62, > 63) for this analysis, and group CCES respondents into strata based on these four variables. We then calculate the proportion of the three partisan groups in the CCES sample (lean Democrat, Republican, or true independent) who fit into each strata. We then ascribe, to each unclassified non-partisan in the nationwide file, the probabilities of their demographic makeup conditional on each partisan group.[1,2]

$$Pr(X_i = x|D_i) = \frac{\sum_{j=1}^{n_D} \mathbb{I}(X_j = x)}{n_D}$$

$$Pr(X_i = x|R_i) = \frac{\sum_{j=1}^{n_R} \mathbb{I}(X_j = x)}{n_R}$$

$$Pr(X_i = x|I_i) = \frac{\sum_{j=1}^{n_R} \mathbb{I}(X_j = x)}{n_I}$$

where X_i is the demographic makeup of voter i in the nationwide voterfile, X_j is the demographic makeup of CCES respondent j, n_D is the number of self-reported Democrats in the CCES sample, n_R is the number of self-reported Republicans in the CCES sample, and n_I is the number of self-reported independents in the CCES sample.

The geographic prior is constructed by taking the precinct-level (or county for small

[1] 810,364 voters in the L2 voterfile do not have a gender listed. For these voters, we imputed Male or Female gender based on their age and first name, using the `R` package `gender`, which compares name and age to census data on frequency of male and female names across years. With this we successfully impute gender for all but 60,407 of the voters with missingness in the gender variable.

[2] 3,017,665 voters in the L2 file do not have their age recorded on the voter list. For these voters, we impute their age by taking the median age of other voters in the file with the same first name and same gender. We do this for all voters with missingness save those in Wisconsin, where, relying on Yougov survey data indicating that voters with no age in the voterfile are overwhelmingly in the 18-34 age demographic, we categorize their age in the youngest stratum. In this way, we were able to successfully impute age for all but 381,926 of the voters with missingness in the age variable.

portion of voters, see above) probability that a non-Democrat or Republican cast a vote for the Republican or Democratic candidate in 2016. This is done by taking the number of votes for the Republican candidate in the 2016 general election – Donald Trump – and the Democratic candidate– Hillary Clinton – and subtracting the number of Republicans (registered, third party lean, or by primary voting) from the vote count for the Republican candidate and subtracting the number of Democrats from the vote count for the Democratic candidate. The remaining votes for the Republican (Democrat), after accounting for the registered Republicans (Democrats) over the remaining total votes cast after accounting for the registered Republicans (Democrats) defines the geographic prior of being a Republican (Democrat). The probability of being an independent is 1 minus the probability of being a Democrat or Republican. This is:

$$Pr(\mathrm{R}_p) = \frac{\text{Trump Votes}_p - \text{Republicans}_p}{\text{Total Votes}_p - \text{Republicans}_p - \text{Democrats}_p}$$

$$Pr(\mathrm{D}_p) = \frac{\text{Clinton Votes}_p - \text{Democrats}_p}{\text{Total Votes}_p - \text{Republicans}_p - \text{Democrats}_p}$$

$$Pr(\mathrm{I}_p) = 1 - Pr(\mathrm{D}_p) - Pr(\mathrm{R}_p)$$

With these probabilities, we use Bayes formula to construct the posterior probability of being a Democrat, Republican or Independent:

$$Pr(\mathrm{R}_i|X_i) = \frac{Pr(X_i|\mathrm{R}_i)Pr(\mathrm{R}_c)}{Pr(X_i|\mathrm{R}_i)Pr(\mathrm{R}_c) + Pr(X_i|\mathrm{D}_i)Pr(\mathrm{D}_c) + Pr(X_i|\mathrm{I}_i)Pr(\mathrm{I}_c)}$$

$$Pr(\mathrm{D}_i|X_i) = \frac{Pr(X_i|\mathrm{D}_i)Pr(\mathrm{R}_c)}{Pr(X_i|\mathrm{R}_i)Pr(\mathrm{R}_c) + Pr(X_i|\mathrm{D}_i)Pr(\mathrm{D}_c) + Pr(X_i|\mathrm{I}_i)Pr(\mathrm{I}_c)}$$

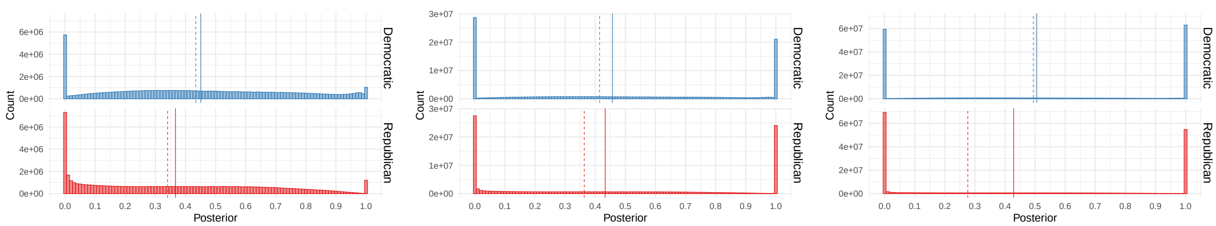
$$Pr(\mathrm{I}_i|X_i) = \frac{Pr(X_i|\mathrm{I}_i)Pr(\mathrm{R}_c)}{Pr(X_i|\mathrm{R}_i)Pr(\mathrm{R}_c) + Pr(X_i|\mathrm{D}_i)Pr(\mathrm{D}_c) + Pr(X_i|\mathrm{I}_i)Pr(\mathrm{I}_c)}$$

With this process, we classify 89% of voters not registered to a major party as lean

Democrats or lean Republicans.[3] Table S4 shows the percent of the electorate classified as Democrats, Republicans, and Independents at each step of the imputation process prior to the Bayesian imputation, then shows the weighted averages of the posterior probabilities that result from our multi-step imputation.

Figure S5 shows the distribution of posterior partisanship probabilities separately for Democratic and Republican partisanship across (a) all voters where we impute partisanship using the Bayesian imputation, (b) all voters who are not registered to the Democratic or Republican party in the L2 files, so for whom we code partisanship through their primary voting, third party affiliation, or Bayesian imputation, and (c) all voters in the L2 file, registered to a major party or otherwise. Voters who are registered to a major party, vote in a partisan primary, or are registered to a party with a clear ideological lean will have a posterior partisan probability of 1 for the appropriate party and 0 for the out-party. This is reflected in the distribution of all voters, where the most common posterior values are overwhelmingly 0's and 1's. We further observe this pattern in the distribution across unaffiliated voters – voters not registered as Democrats or Republicans – for whom some imputation is required. In the distribution of voters for whom we were unable to code based on primary voting or third party affiliation, and thus relied on Bayesian imputation, we see some posteriors between 0 and 1.

[3] 59.89% of the electorate is not registered to the Democratic or Republican party in the L2 voterfile. We can compare this to the partisan breakdown after our three-step imputation, where the weighted (weighted by posterior partisan probability) proportion of the electorate that is Independent is just 6.58%. 100% − 6.58%/59.89% = 89.01%.



(a) Bayesian Imputation (b) Unaffiliated Voters (c) All Voters

Figure S5: Posterior Partisanship

Histograms show the distribution of posterior partisanship probabilities separately for Democratic and Republican partisanship across (a) all voters where we impute partisanship using the Bayesian imputation, (b) all voters who are not registered to the Democratic or Republican party in the L2 files, so for whom we code partisanship through their primary voting, third party affiliation, or Bayesian imputation, and (c) all voters in the L2 file, registered to a major party or otherwise. Note that voters who are registered to a major party, vote in a partisan primary, or are registered to a party with a clear ideological lean will have a posterior partisan probability of 1 for the appropriate party and 0 for the out-party. Solid vertical lines plot the mean and dashed vertical lines plot the median of each distribution.

Table S1: Parties Coded as Lean Democrat/Lean Republican

	Party	Code
1	Democratic	Lean Democrat
2	Republican	Lean Republican
3	Non-Partisan	Unknown
4	Registered Independent	Unknown
5	American Independent	Lean Republican
6	Other	Unknown
7	Libertarian	Lean Republican
8	Independence	Unknown
9	Unknown	Unknown
10	Green	Lean Democrat
11	Declined to State	Unknown
12	Conservative	Lean Republican
13	Peace and Freedom	Lean Democrat
14	Working Family Party	Lean Democrat
15	Constitution	Lean Republican
16	Reform	Unknown
17	Constitutional	Lean Republican
18	Natural Law	Unknown
19	Women's Equality Party	Lean Democrat
20	Moderate	Unknown
21	Progressive	Lean Democrat
22	American	Unknown
23	Mountain	Unknown
24	Liberal	Lean Democrat
25	Green Libertarian	Lean Democrat
26	Socialist	Lean Democrat
27	Independent Democrat	Lean Democrat
28	Patriot	Unknown
29	Independent Republican	Lean Republican
30	Socialist Labor	Lean Democrat
31	Christian	Unknown
32	Harold Washington Democrat	Lean Democrat
33	Communist	Lean Democrat
34	Taxpayers	Unknown
35	Social Democrat	Lean Democrat
36	Consumer	Unknown
37	Right to Life	Lean Republican
38	Citizens	Unknown
39	Whig	Unknown
40	Rainbow	Lean Democrat
41	Freedom	Unknown
42	Anarchist	Unknown
43	Bull Moose	Unknown
44	Populist	Unknown
45	Tea	Lean Republican
46	Prohibition	Unknown
47	Free Choice	Unknown
48	Federalist	Unknown
49	Worker's Party	Lean Democrat
50	Labor	Lean Democrat
51	Harold Washington Republican	Lean Republican
52	Harold Washington	Lean Democrat
53	Individualist	Unknown
54	Alliance	Unknown
55	Citizens Republican	Lean Republican
56	Natural Party	Unknown
57	Grass Roots	Unknown
58	Tax	Unknown
59	Solidarity	Unknown
60	Peoples	Unknown

Table S2: State Registration Rules and Primary Types

	State	State Recorded PID	Democratic primary type	Republican primary type
1	Alabama	No	Closed	Open
2	Alaska	Yes	Open	Closed
3	Arizona	Yes	Semi-closed	Semi-closed
4	Arkansas	Yes (optional)	Open	Open
5	California	Yes	Top-two	Top-two
6	Colorado	Yes	Semi-closed	Semi-closed
7	Connecticut	Yes	Closed	Closed
8	Delaware	Yes	Closed	Closed
9	District of Columbia	Yes	Closed	Closed
10	Florida	Yes	Closed	Closed
11	Georgia	No	Open	Open
12	Hawaii	No	Open	Open
13	Idaho	Yes	Semi-closed	Semi-closed
14	Illinois	No	Open	Open
15	Indiana	No	Open	Open
16	Iowa	Yes	Open	Open
17	Kansas	Yes	Semi-closed	Semi-closed
18	Kentucky	Yes	Closed	Closed
19	Louisiana	Yes	Non-partisan	Non-partisan
20	Maine	Yes	Closed	Closed
21	Maryland	Yes	Closed	Closed
22	Massachusetts	Yes	Semi-closed	Semi-closed
23	Michigan	No	Open	Open
24	Minnesota	No	Open	Open
25	Mississippi	No	Open	Open
26	Missouri	No	Open	Open
27	Montana	No	Open	Open
28	Nebraska	Yes	Semi-closed	Semi-closed
29	Nevada	Yes	Closed	Closed
30	New Hampshire	Yes	Semi-closed	Semi-closed
31	New Jersey	Yes	Semi-closed	Semi-closed
32	New Mexico	Yes	Closed	Closed
33	New York	Yes	Closed	Closed
34	North Carolina	Yes	Semi-closed	Semi-closed
35	North Dakota	No	Open	Open
36	Ohio	No	Open	Open
37	Oklahoma	Yes	Semi-closed	Closed
38	Oregon	Yes	Closed	Closed
39	Pennsylvania	Yes	Closed	Closed
40	Rhode Island	Yes	Semi-closed	Semi-closed
41	South Carolina	No	Open	Open
42	South Dakota	Yes	Semi-closed	Semi-closed
43	Tennessee	No	Open	Open
44	Texas	No	Open	Open
45	Utah	Yes	Semi-closed	Semi-closed
46	Vermont	No	Open	Open
47	Virginia	No	Open	Open
48	Washington	No	Non-partisan	Non-partisan
49	West Virginia	Yes	Semi-closed	Semi-closed
50	Wisconsin	No	Open	Open
51	Wyoming	Yes	Open	Open

Table S3: State-level voter to precinct matches

	State	Precinct Matches	County Matches	Total Voters	Percent Precinct	Percent County
1	Alabama	2725805	216864	2942669	92.63%	7.37%
2	Alaska	440960	5891	446851	98.68%	1.32%
3	Arizona	3355012	70571	3425583	97.94%	2.06%
4	Arkansas	1349036	65582	1414618	95.36%	4.64%
5	California	18196337	3859	18200196	99.98%	0.02%
6	Colorado	3015783	123998	3139781	96.05%	3.95%
7	Connecticut	2119260	3380	2122640	99.84%	0.16%
8	DC	396394	48	396442	99.99%	0.01%
9	Delaware	619808	311	620119	99.95%	0.05%
10	Florida	12338852	6432	12345284	99.95%	0.05%
11	Georgia	5465155	3775	5468930	99.93%	0.07%
12	Hawaii	636251	134	636385	99.98%	0.02%
13	Idaho	681658	19089	700747	97.28%	2.72%
14	Illinois	7705542	36662	7742204	99.53%	0.47%
15	Indiana	3716361	40913	3757274	98.91%	1.09%
16	Iowa	1885254	2172	1887426	99.88%	0.12%
17	Kansas	1456847	94437	1551284	93.91%	6.09%
18	Kentucky	2856862	40351	2897213	98.61%	1.39%
19	Louisiana	2763402	10876	2774278	99.61%	0.39%
20	Maine	849285	54260	903545	93.99%	6.01%
21	Maryland	3798705	311	3799016	99.99%	0.01%
22	Massachusetts	4020865	36088	4056953	99.11%	0.89%
23	Michigan	6643980	244	6644224	100.00%	0.00%
24	Minnesota	3144404	515	3144919	99.98%	0.02%
25	Mississippi	1704121	70162	1774283	96.05%	3.95%
26	Missouri	3307305	267974	3575279	92.50%	7.50%
27	Montana	556203	17874	574077	96.89%	3.11%
28	Nebraska	975054	75018	1050072	92.86%	7.14%
29	Nevada	1401254	41825	1443079	97.10%	2.90%
30	New Hampshire	808929	88	809017	99.99%	0.01%
31	New Jersey	5330146	13094	5343240	99.75%	0.25%
32	New Mexico	1077260	89	1077349	99.99%	0.01%
33	New York	10820724	447366	11268090	96.03%	3.97%
34	North Carolina	5897797	17601	5915398	99.70%	0.30%
35	North Dakota	327978	6117	334095	98.17%	1.83%
36	Ohio	7189406	4410	7193816	99.94%	0.06%
37	Oklahoma	1551914	128206	1680120	92.37%	7.63%
38	Oregon	2593732	204739	2798471	92.68%	7.32%
39	Pennsylvania	7643777	501	7644278	99.99%	0.01%
40	Rhode Island	690427	803	691230	99.88%	0.12%
41	South Carolina	2811852	43177	2855029	98.49%	1.51%
42	South Dakota	416486	62290	478776	86.99%	13.01%
43	Tennessee	3245682	0	3245682	100.000%	0.00%
44	Texas	13223300	5243	13228543	99.96%	0.04%
45	Utah	1332678	44331	1377009	96.78%	3.22%
46	Vermont	408720	37	408757	99.99%	0.01%
47	Virginia	4922433	55586	4978019	98.88%	1.12%
48	Washington	4136953	86870	4223823	97.94%	2.06%
49	West Virginia	1070862	4154	1075016	99.61%	0.39%
50	Wisconsin	4445976	659	4446635	99.99%	0.01%
51	Wyoming	231091	1298	232389	99.44%	0.56%
52	Total	178303878	2431767	180735645	98.655%	1.35%

Table S4: Partisan Breakdowns at Each Imputation Stage

	Party	L2 %	After Primary Coding %	After Third Party Coding %	After Full Imputation %
1	Democrat	23.16%	34.70%	34.88%	50.54%
2	Independent	59.89%	36.38%	35.60%	6.58%
3	Republican	16.95%	28.92%	29.52%	42.88%

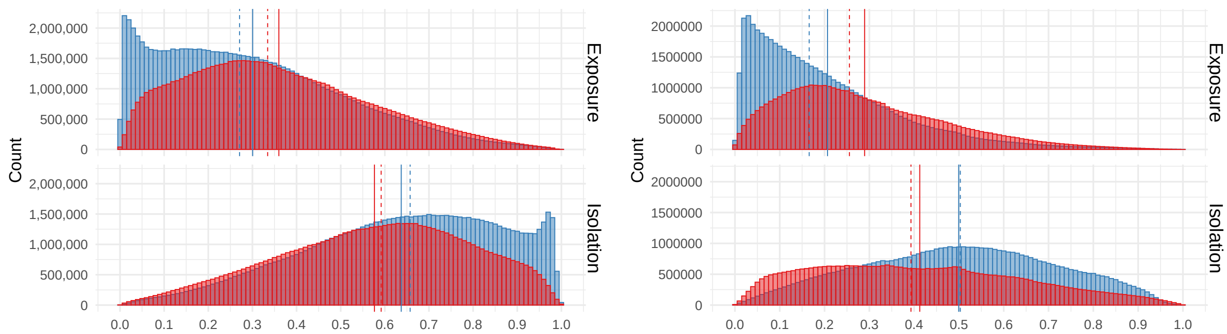


Figure S6: Exposure and Isolation with Imputation (left) and without (right).

5 Results without Imputation

Figure S6 displays the results with imputation next to the results without imputation. We see that imputing for partisanship added many Democrats living in high isolation to the Democratic distribution, while the Republican isolation distribution shifted slightly towards higher isolation. Exposure increases as there are now very few independents left in the sample.

6 Imputation Accuracy

In order to assess the accuracy of our imputation, we surveyed 12,221 voters, randomly sampled after stratification by state and whether partisanship was visible on the voterfile with an over-sample of non-partisans. Respondents were contacted by email from email addresses linked to the voterfile by the vendor L2. In the survey we validate L2's linking, finding that, conditional on getting a response, 86.1% of respondents report to being the person to whom the email address was matched in the voter list. We limit our analysis to these voters ($n = 10,519$).

To conduct the survey, we sent emails containing the invitation to participate in our online Qualtrics survey to 1,753,493 unique voters. Of these emails, 47.2% bounced, indicating that the email was invalid or that our email was rejected by a server, perhaps for spam protection. Thus, 925,339 unique voters received an invitation to participate in the survey, and we received 12,221 responses, a response rate of 1.3%, which is similar to the single-digit response rates expected for modern phone or email surveys. In our analysis of the survey, we construct survey weights to account for non-response and report results with and without these weights below.

Using these data, we assess validity and accuracy of our imputation in two ways: first by comparing our imputed partisanship to self-reported partisanship (again including "lean" partisans in the parties) and also by comparing the ideology of imputed and non-imputed voters.

Figure S7 plots the imputed posterior partisanship probabilities for our survey respondents who are not registered as Democrats or Republicans against their rates of self-reported partisanship. A perfect correlation would follow the 45 degree lines in these figures. We see that our partisan predictions are strongly correlated with self-reported partisanship, and approach the levels of accuracy that we might expect given the levels of partisan instability in survey response (see main text). We are most accurate when our imputation is most confident, where much of the support of our imputation lies, in the survey sample and in the unaffiliated voter population, and at the high ends, we approach the levels of accuracy possible with the proportions of partisans in our sample.

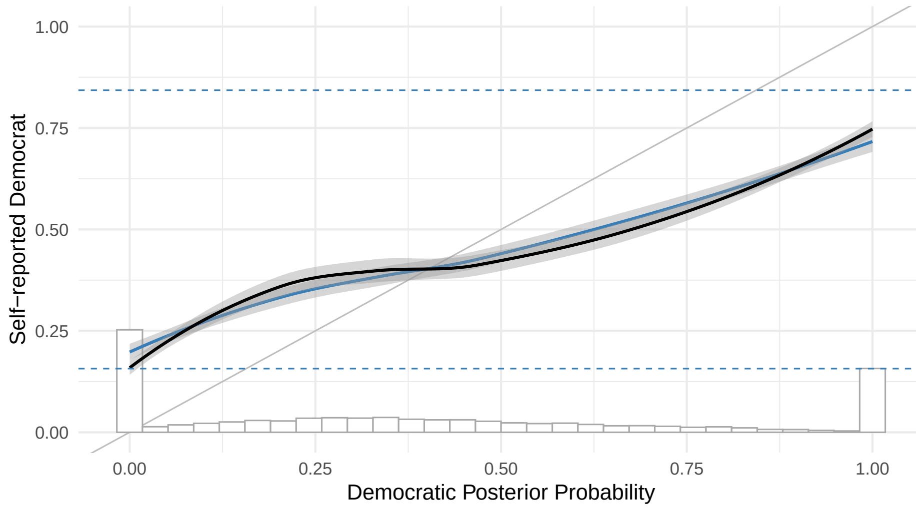
To more systematically test the accuracy of the imputation, we compute Brier Scores detailing the mean squared error, or the average squared deviation of each survey respondent's posterior partisan probability from their actual partisanship. A Brier score is designed to assess the magnitude of deviations for a probabilistic forecast, and produces a statistic on a 0 to 1 scale, with 0 being zero deviations, or perfect accuracy, and 1 being complete deviations, or zero accuracy.[4] Brier Scores can be inverted (1-score) and interpreted similar to the rate of accuracy of the forecast. Thus, when we observe a Brier score of 0.23 (0.24 unweighted) for Democratic partisanship and 0.23 (.25 unweighted) for Republican partisanship, this indicates our forecast is accurately predicting partisanship at rates of approximately 77% for both parties. Figure S8 further illustrates the accuracy of the forecast, plotting the histogram of raw squared deviations across units, as well as the average and median squared deviations. We see that most of the units have very small deviations, with the medians very close to 0. Figures S9 and S10 demonstrate the consistency of these patterns for survey respondents living in different states, different types of urban areas, and different densities.

In comparing our imputed partisanship to self-reported partisanship, we present unweighted survey results and results that incorporate survey weights created to make the survey population more comparable to the population of voters for whom we imputed partisanship, that is voters not explicitly registered as Democrats or Republicans in the L2 voter

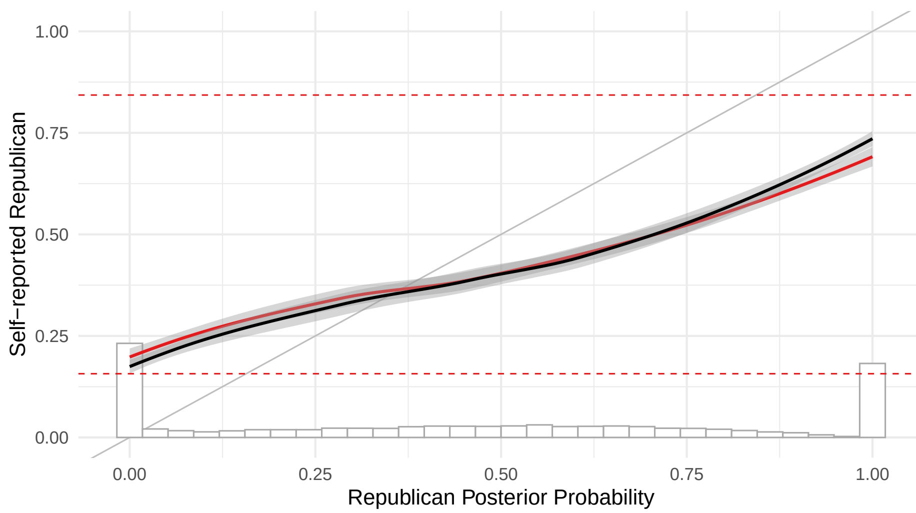
[4] This is,

$$\text{Brier Score}_p = \frac{1}{n}\sum_{i=1}^{n}[P(p)_i - \mathbb{I}(p_i = p)]^2$$

where the Brier score for predicting p partisanship is the summation of the squared deviations of $P(p)_i$ (the posterior partisan probability of p partisanship for respondent i) from $\mathbb{I}(p_i = p)$ (an indicator variable equaling 1 if the survey respondent reports as being p partisanship).



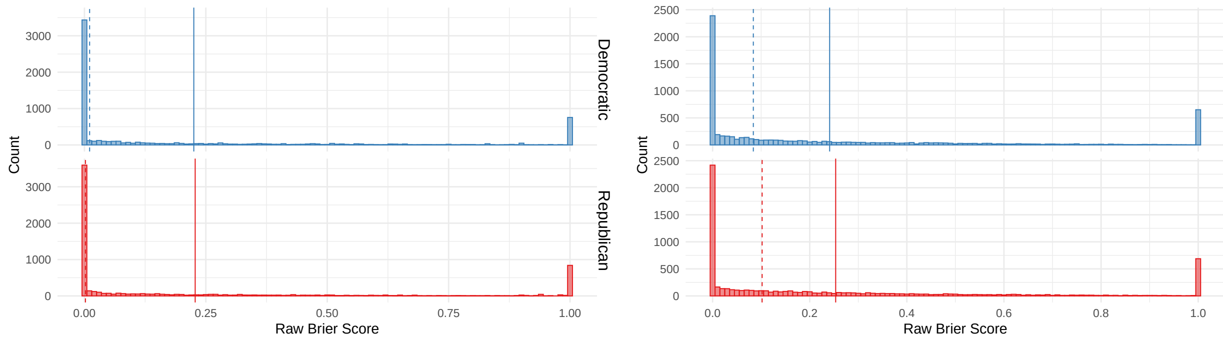
(a) Democratic



(b) Republican

Figure S7: Percent self-report Partisan Category by Posterior Partisan Probability

Figures show the LOESS lines plotting the relationship between posterior partisan probability (Democratic on left, Republican on right) and the rates of survey respondents reporting as the corresponding partisanship. The correlation is limited to the subset of survey respondents ($n = 7,087$) who are not registered with a major political party. Black lines plot the LOESS curve with survey weights incorporated, red/blue lines without survey weights. The 45-degree grey line plots a perfect 1-to-1 relationship between posterior partisan probability and self-reported partisanship. The horizontal dotted lines show the rates at which survey respondents who are registered Democrats/Republicans self-report partisanship in agreement (or disagreement for the lower lines) with their actual partisan registration. That is, the upper blue (red) dotted line represents the proportion of survey respondents we know are registered Democrats (Republicans) who self report as Democrats (Republicans), and the lower dotted line represents the proportion who do not self report as Democrats (Republicans). These lines represent lower and upper bounds on how accurate we can expect our forecast to appear when measured against survey data. The histogram on the bottom plots the frequency distribution of posterior partisan probabilities across the unaffiliated subset.



(a) Survey Weights

(b) Unweighted

Figure S8: Brier Score Distributions

Histograms shows Brier score distribution for survey respondents not registered to the Democratic or Republican party ($n = 7,087$). The left panel shows the distribution weighted by survey weights, and the right panel shows unweighted distribution. Dashed vertical lines show the median and solid vertical lines show the mean for the distribution.

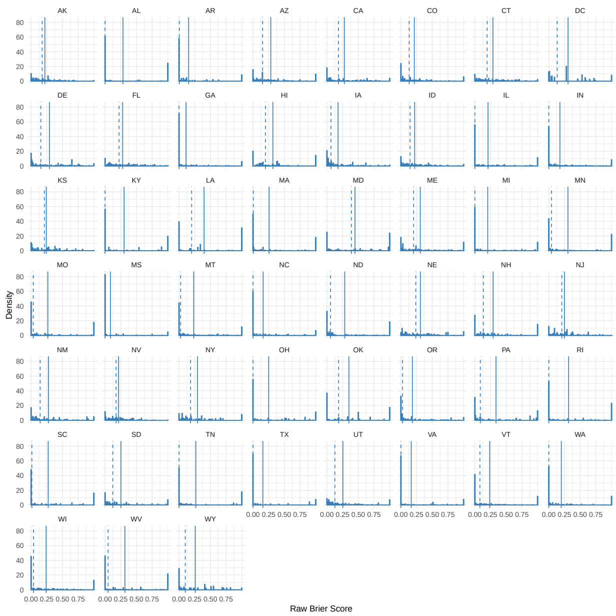
lists. These weights incorporate observable information on race, age, gender, vote history, and the type of urban area (major, minor, or outside) and population density in which the respondent lives. We also model response bias within the sample of people emailed surveys along the same observable variables. The survey weights are the combination of these models.

To construct the weights, we first model the likelihood of having an email address attached to voter records for each unaffiliated voter in the nationwide voter list.

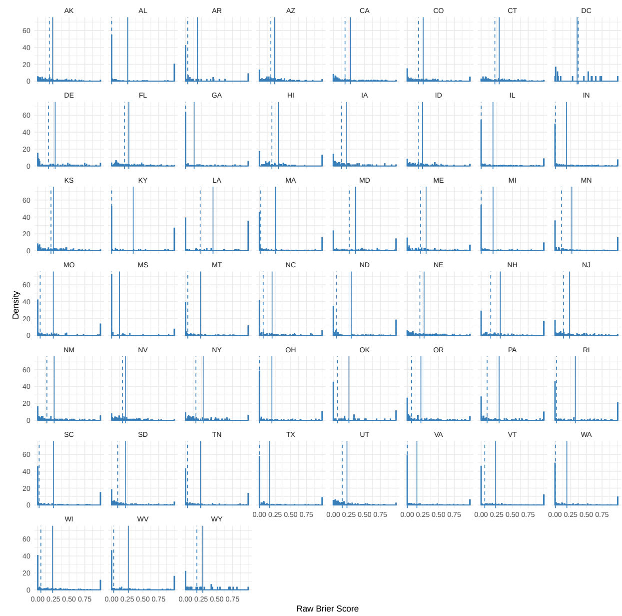
$$P(\text{Email}_i|\mathbf{X}_i) = g^{-1}(\beta\mathbf{X}_i)$$

where $g^{-1}(\cdot)$ is the inverse-logit function, $\mathbf{X}_i$ is a vector of covariate values for voter i's state of residence, pre-imputation partisan identification, race, age, gender, whether or not they voted in 2016, the category (High, Medium, Low, Very low) of population density of the tract in which they live, and the type of urban area (Major, Minor, Outside Metro area) in which they live.

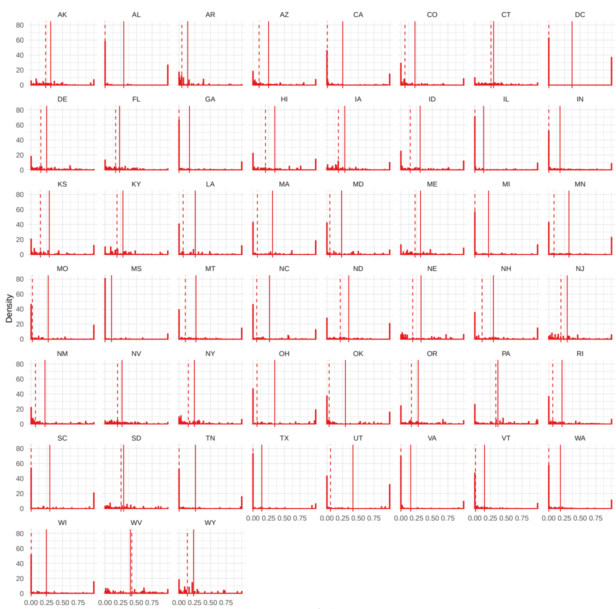
Next, we model the likelihood of a voter with an email being contacted for our survey.



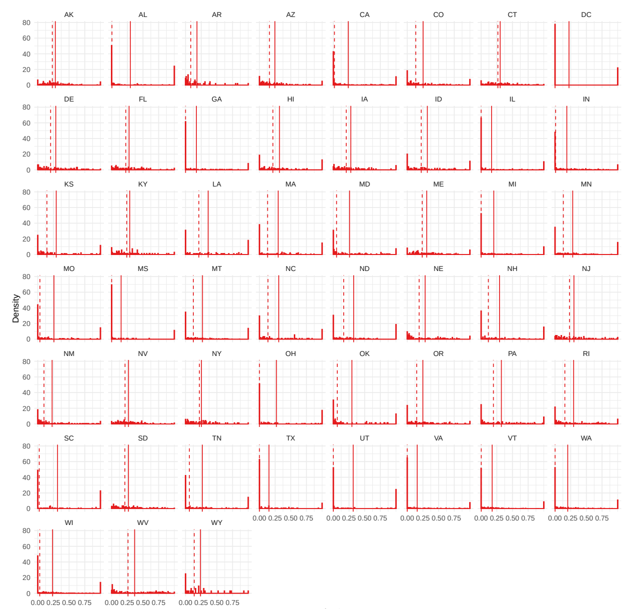
(a) Democratic: Survey Weights



(b) Democratic: Unweighted



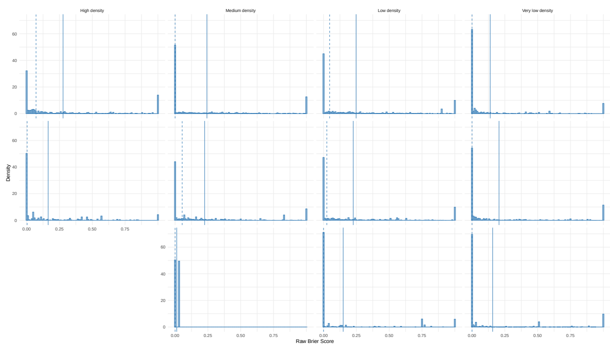
(c) Republican: Survey Weights



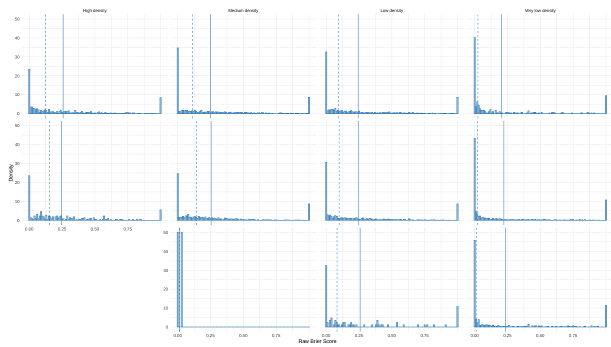
(d) Republican: Unweighted

Figure S9: Brier Score Distributions by State

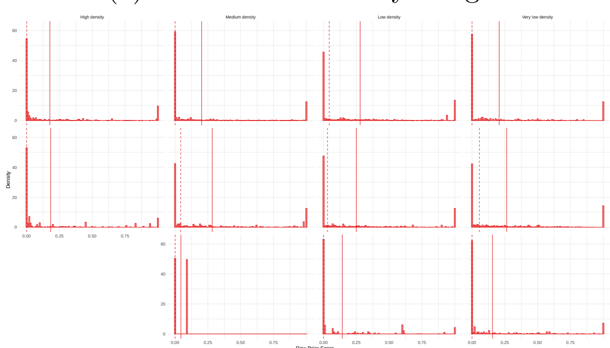
Histograms shows Brier score distribution for survey respondents not registered to the Democratic or Republican party, subset by each state. The left panel shows the distribution weighted by survey weights, and the right panel shows unweighted distribution. Dashed vertical lines show the median and solid vertical lines show the mean for the distribution.



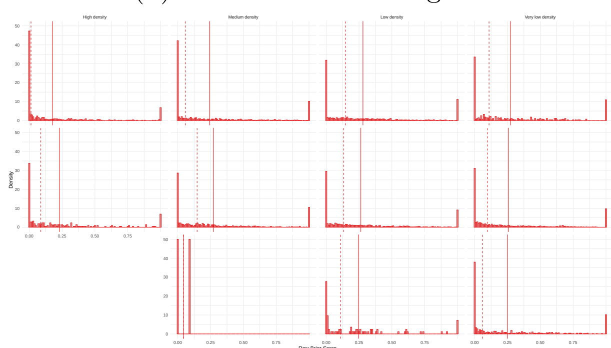
(a) Democratic: Survey Weights



(b) Democratic: Unweighted



(c) Republican: Survey Weights



(d) Republican: Unweighted

Figure S10: Brier Score Distributions by Urban Area and Density

Histograms shows Brier score distribution for survey respondents not registered to the Democratic or Republican party, subset by urban area type and density. The left panel shows the distribution weighted by survey weights, and the right panel shows unweighted distribution. Dashed vertical lines show the median and solid vertical lines show the mean for the distribution.

Potential survey respondents were randomly sampled after stratification by state and whether partisanship was visible on the voterfile with an over-sample of non-partisans. We then multiply the probability of having an email with the probability of contact to produce a design probability (probability of having an email and receiving a survey invite) for each voter in our survey contact list.

$$P(\text{Contact}_i|\text{State}_i, \text{Partisan Registration}) = g^{-1}(\beta_0 + \beta_1\text{State}_i + \beta_2\text{Partisan Registration})$$

$$P(\text{Email}_i \cap \text{Contact}_i|\mathbf{X}_i) = P(\text{Email}_i|\mathbf{X}_i) \times P(\text{Sampled}_i|\text{State}_i, \text{Partisan Registration})$$

After receiving responses, we model the likelihood of receiving a response within our contacted sample. We then multiple response probability with our design probability to get the overall probability of having an email, being contact, and getting a response. We re-scale these weights as is common practice so that their sum reflects the total number of respondents.

$$P(\text{Response}_i|\mathbf{X}_i) = g^{-1}(\beta\mathbf{X}_i)$$

$$\text{Survey Weight}_i = \frac{1/(P(\text{Email}_i \cap \text{Contact}_i \cap \text{Response}_i|\mathbf{X}_i))}{\sum_{i=1}^{n} 1/(P(\text{Email}_i \cap \text{Contact}_i \cap \text{Response}_i|\mathbf{X}_i))} \times n$$

where n is the number of survey respondents.

Figure S11 and Table S5 detail the distribution of weights across the entire survey sample and the subset that are not registered as Republicans or Democrats. Tables S6 through S10 and Figure S12 demonstrate how the survey sample (weighted and unweighted) compares

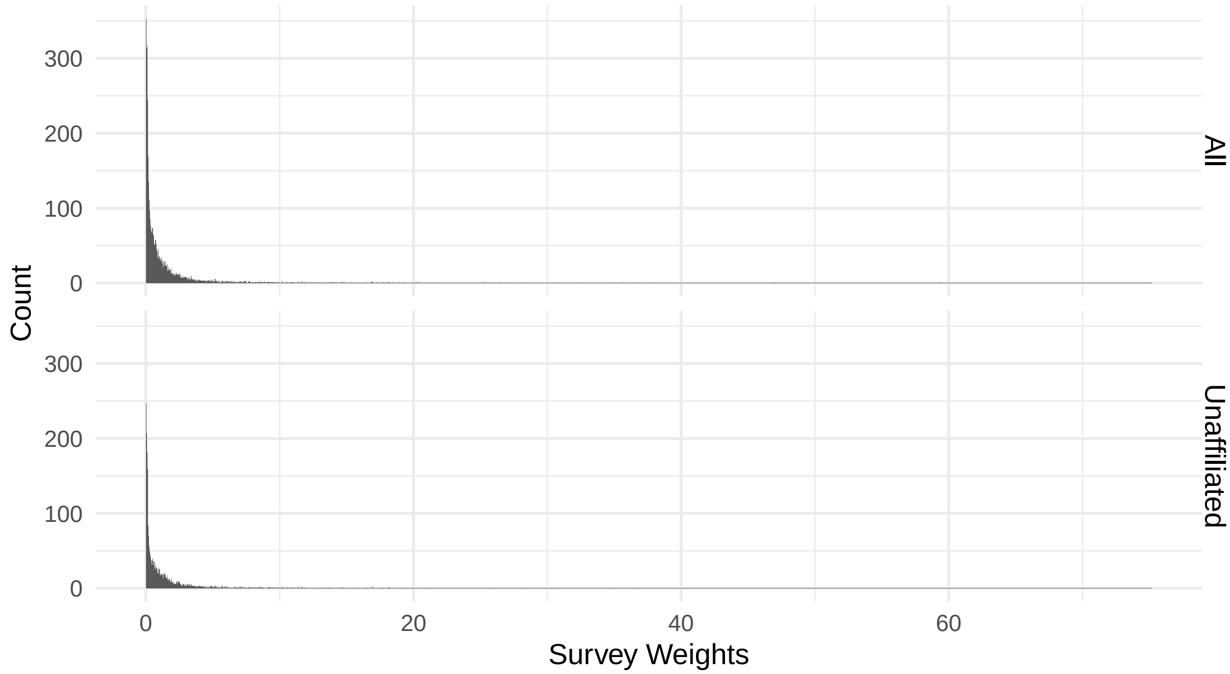


Figure S11: Survey Weights Histogram

Histogram displays distribution of survey weights across the entire survey sample ($n = 12,221$) in the top panel and the subset of the survey sample comprised of voters not registered to major political party ($n = 7,087$) in the bottom panel. Weights are scaled so that the sum equals the total number of respondents in the sample.

across key observable variables (race, gender, population density, urban area, 2016 turnout, and age) to the entire population of registered voters in the US, the entire population of US voters not registered to a major party, the subset of voters for whom L2 has an email, and the sample of voters whom we sent an invitation to participate in the survey. Across variables, the survey sample is comparable to the populations from which it was drawn, but is older, has higher proportions of men, Whites, people living in lower population densities and outside of major urban areas, and higher levels of 2016 turnout than the comparison populations. Once survey weights are incorporated these levels look very similar across variables.

We can also validate the premise and accuracy of our imputation by examining the political ideology of imputed to non-imputed voters. If the voters we impute have, on average,

Table S5: Survey Weights Percentiles

	Sample	0%	1%	10%	25%	50%	75%	90%	99%	100%
1	All	0.03	0.06	0.13	0.41	1.16	2.36	8.34	0.03	0.06
2	Unaffiliated	0.01	0.02	0.05	0.11	0.41	1.20	2.39	8.46	36.57

Table displays percentiles of survey weights across the entire survey sample ($n = 12,221$) in the top row and the subset of the survey sample comprised of voters not registered to major political party ($n = 7,087$) in the bottom row. Weights are scaled so that the sum equals the total number of respondents in the sample.

Table S6: Race Survey Comparison

	Sample	White	Black	Hispanic	Asian	Other
1	Survey	80.24%	3.28%	3.94%	1.78%	10.76%
2	Survey Weighted	66.06%	11.15%	9.35%	3.24%	10.19%
3	Unaffiliated Voters	65.30%	10.32%	10.12%	3.10%	11.15%
4	All Voters	64.06%	10.65%	11.20%	2.99%	11.10%
5	L2 Emails	65.32%	10.17%	10.36%	2.90%	11.25%
6	Survey Invitation	70.98%	8.72%	6.57%	2.55%	11.18%

Table S7: Gender Survey Comparison

	Sample	Male	Female
1	Survey	50.89%	48.93%
2	Survey Weighted	46.16%	53.84%
3	Unaffiliated Voters	47.74%	52.04%
4	All Voters	46.78%	53.01%
5	L2 Emails	46.14%	53.70%
6	Survey Invitation	47.01%	52.80%

Table S8: Population Density Survey Comparison

	Sample	High density	Medium density	Low density	Very low density
1	Survey	11.37%	27.76%	33.04%	27.79%
2	Survey Weighted	17.72%	29.20%	29.46%	23.62%
3	Unaffiliated Voters	14.89%	29.40%	30.90%	24.77%
4	All Voters	18.97%	28.90%	29.15%	22.92%
5	L2 Emails	19.50%	30.29%	29.32%	20.86%
6	Survey Invitation	12.96%	27.46%	30.76%	28.77%

Table S9: Urban Area Survey Comparison

	Sample	Major	Minor	Outside Metro area
1	Survey	38.78%	53.21%	8.01%
2	Survey Weighted	55.11%	38.03%	6.86%
3	Unaffiliated Voters	54.36%	38.88%	6.76%
4	All Voters	55.70%	38.18%	6.13%
5	L2 Emails	58.65%	36.38%	4.96%
6	Survey Invitation	40.75%	50.21%	9.03%

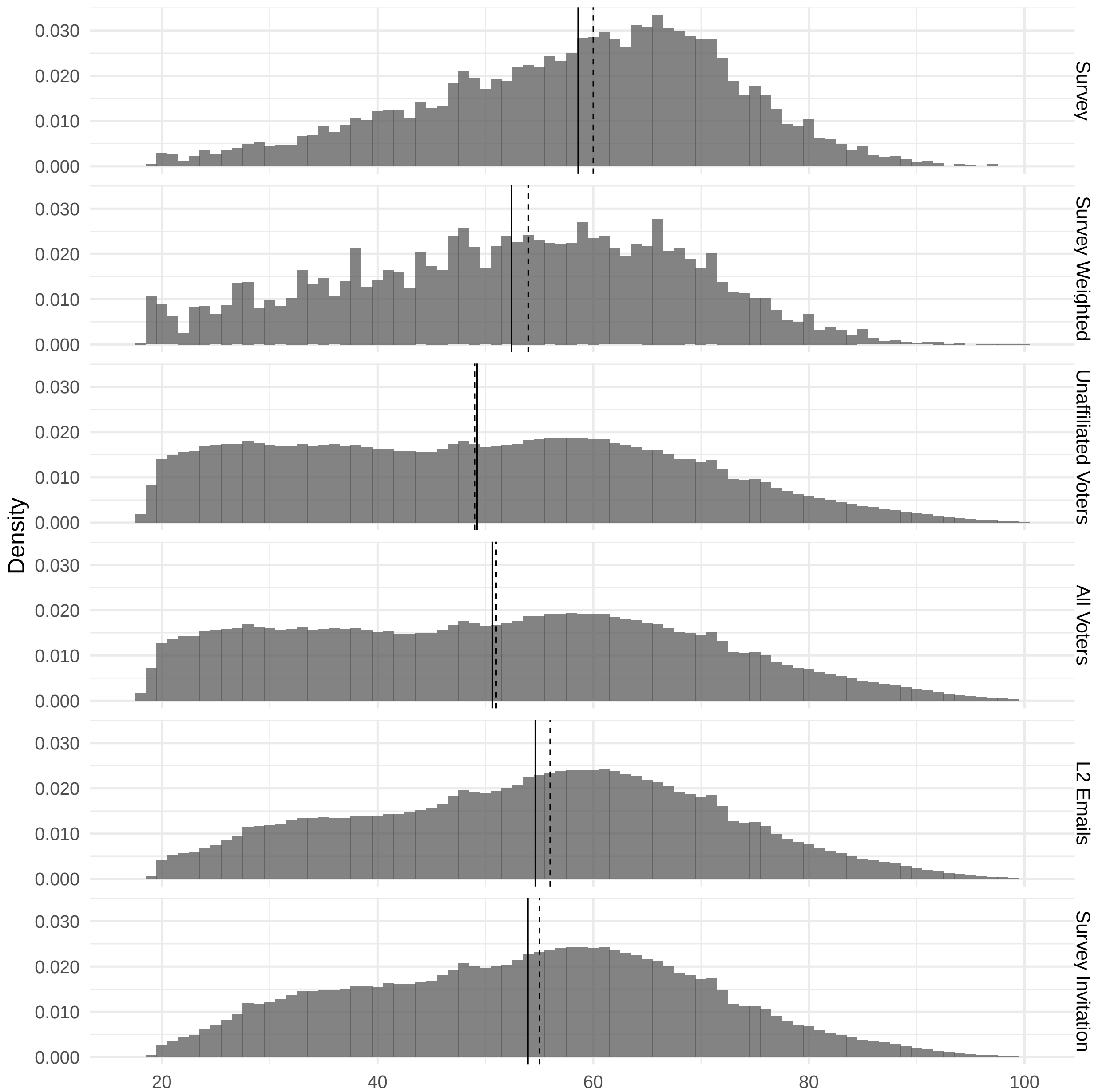


Figure S12: Age Survey Comparison

Histograms plot the distribution of age for the survey sample, the weighted survey sample, the entire population of registered voters in the US, the entire population of US voters not registered to a major party, the subset of voters for whom L2 has an email, and the sample of voters whom we sent an invitation to participate in the survey. Vertical solid lines plot the mean and vertical dashed lines plot the median of the distributions.

Table S10: 2016 Turnout Survey Comparison

	Sample	Vote 2016
1	Survey	89.89%
2	Survey Weighted	70.41%
3	Unaffiliated Voters	66.22%
4	All Voters	70.02%
5	L2 Emails	79.19%
6	Survey Invitation	74.48%

very different political ideology than those not imputed, then imputing for the purposes of measuring partisan exposure could be misleading because it would artificially inflate the levels of exposure to ideologically (dis)similar voters. On the other hand, if imputed and non-imputed voters have similar ideologies, it demonstrates that not imputing would be a mistake because it would cause us to understate levels of exposure to these ideologies.

Across a number of tests, we find strong consistency between imputed and non-imputed voters in our survey data. In Figure S13, we compare responses on a 7-point scale of ideology from "Extremely Liberal" to "Extremely Conservative." This ideology scale is standard on large-scale political science surveys, such as the American National Election Study. Comparing imputed and non-imputed voters yields similar distributions within party, as defined by self-reported responses to a three item question about their partisanship. For Democrats ($N = 2,914$) a Kolmogorov-Smirnov test for a difference in distributions between imputed and non-imputed voters yields $D = 0.022, p = 0.893$, not allowing us to reject the null hypothesis of no difference in distributions at $p < .05$. Kolmogorov-Smirnov for Republicans ($N = 3,067$) also shows similar ideology across imputed and non-imputed voters ($D = 0.029, p = 0.564$). In Figure S14 we also compare ideology across imputed and non-imputed partisans within party but further subset the data by states in which party registration is possible (see Table S2). Within party, the ideology across imputed and non-imputed individuals and different types of states is very similar.

We can also use ideology to validate that our imputations reflect ideological variation so

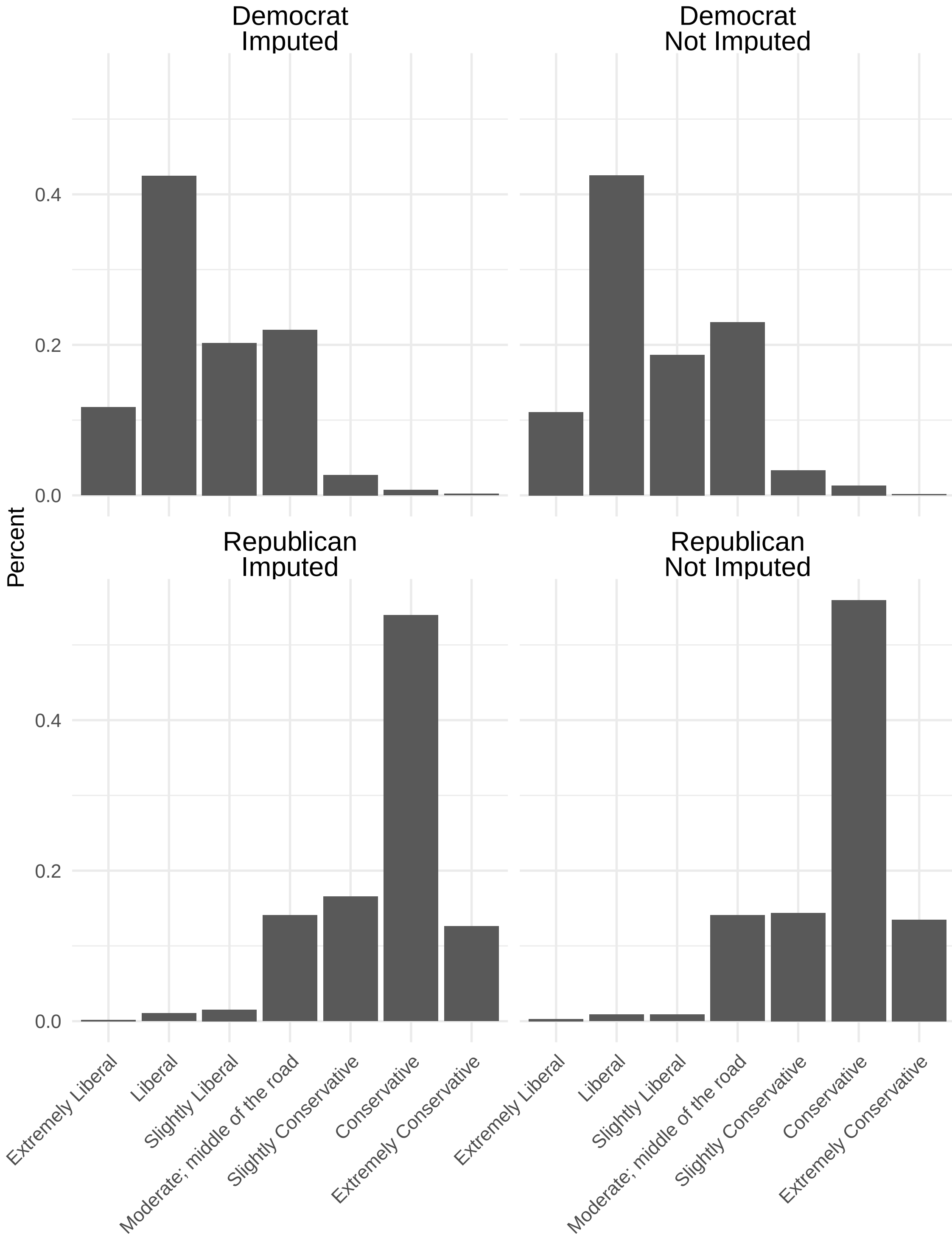


Figure S13: Distribution of self-reported ideology for self-described Republicans and Democrats among voters for whom partisanship was or was not imputed.

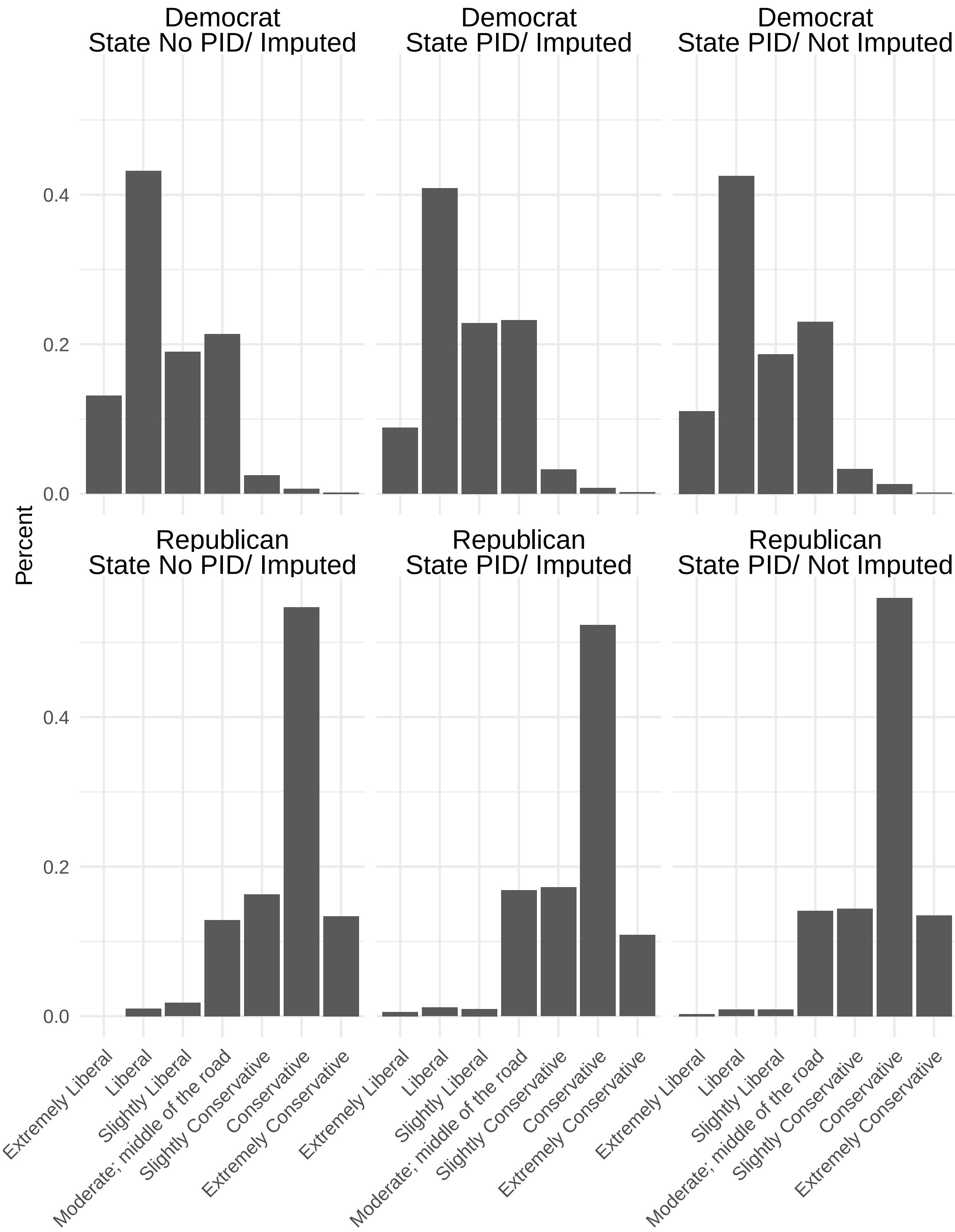


Figure S14: Distribution of self-reported ideology for self-described Republicans and Democrats among voters for whom partisanship was or was not imputed and by whether state records party identifcation.

that the probability of partisanship is correlated with ideology, that is those we impute as more likely to be Republican are more conservative than those we impute as more likely to be Democrats. We do this in Figure S15 where we show that our posterior probability of being a Democrat ($Pr(D)$) is correlated with two measures of ideology and that the variation is consistent across imputed and non-imputed voters and across states with and without party registration. The first is on the self-reported scale of ideology, where 1 means "Extremely Liberal" and 7 means "Extremely Conservative". In the top figure, as $Pr(D)$ increases (divided into five quantiles on the x-axis), the average self-reported ideology also becomes more liberal. In the lower figure, we scale ideology using survey respondents' statement of being "for" or "against" eight issues before Congress. These were (issues were presented in random order):

1. Repeal Affordable Care Act: Would repeal the Affordable Care Act of 2009 (also known as Obamacare).
2. American Health Care: Would repeal the tax penalties on individuals for not maintaining health coverage and on employers for not offering coverage. Would end subsidies to help people purchase insurance and would end funding for states that expanded Medicaid.
3. Financial CHOICE Act: Allows banks to not be subject to the heightened regulatory requirements of Dodd-Frank by maintaining enough reserve funds to withstand a financial downturn. Grants the president the power to fire the head of the Consumer Financial Protection Bureau and the Federal Housing Finance Agency at any time and without cause. Repeals a rule which prevents commercial banks from making speculative investments for their own profits.
4. Kate's Law: Increases criminal penalties for individuals in the country illegally who are convicted of certain crimes, deported, and then re-enter the U.S. illegally.
5. Countering America's Adversaries Through Sanctions Act: Places sanctions on Iran, North Korea, and Russia. Sets into law sanctions imposed by the Obama administration for Russia's interference in Ukraine, Syria, and the 2016 presidential election. Requires the president to get congressional approval before easing or lifting sanctions on Russia.

6. No Sanctuary for Criminals: Withholds federal funds from states and localities that do not follow federal immigration laws.

7. Assault Weapons Ban of 2019: Makes it a crime to knowingly import, sell, manufacture, transfer, or possess a semiautomatic assault weapon or large capacity ammunition feeding device.

8. Impeaching Donald Trump, President of the United States, for high crimes and misdemeanors.

9. Federal Civilian Workforce Pay Raise Fairness Act of 2019: increases by 2.6% the rates of basic pay for federal civilian employees for 2019.

We then scale their responses to extract a measure of latent ideology for each voter using the method developed by Clinton, Jackman, and Rivers (2004). This is the same method that has been used to scale the ideology of Members of Congress and voters in previous research Tausanovitch and Warshaw (2013). The scale is arbitrary, with a mean of 0, max of 2.23, and min -1.96. More negative scores mean more liberal. The median Democrat in our data has a score of -1.07, the median Republican 0.87, and the median Independent -0.08. We examine the correlation between this measure and $Pr(D)$ and, once again, find that as $Pr(D)$ increases, their issue-scaled ideology also becomes more liberal, as indicated by lower scores on this latent dimension.

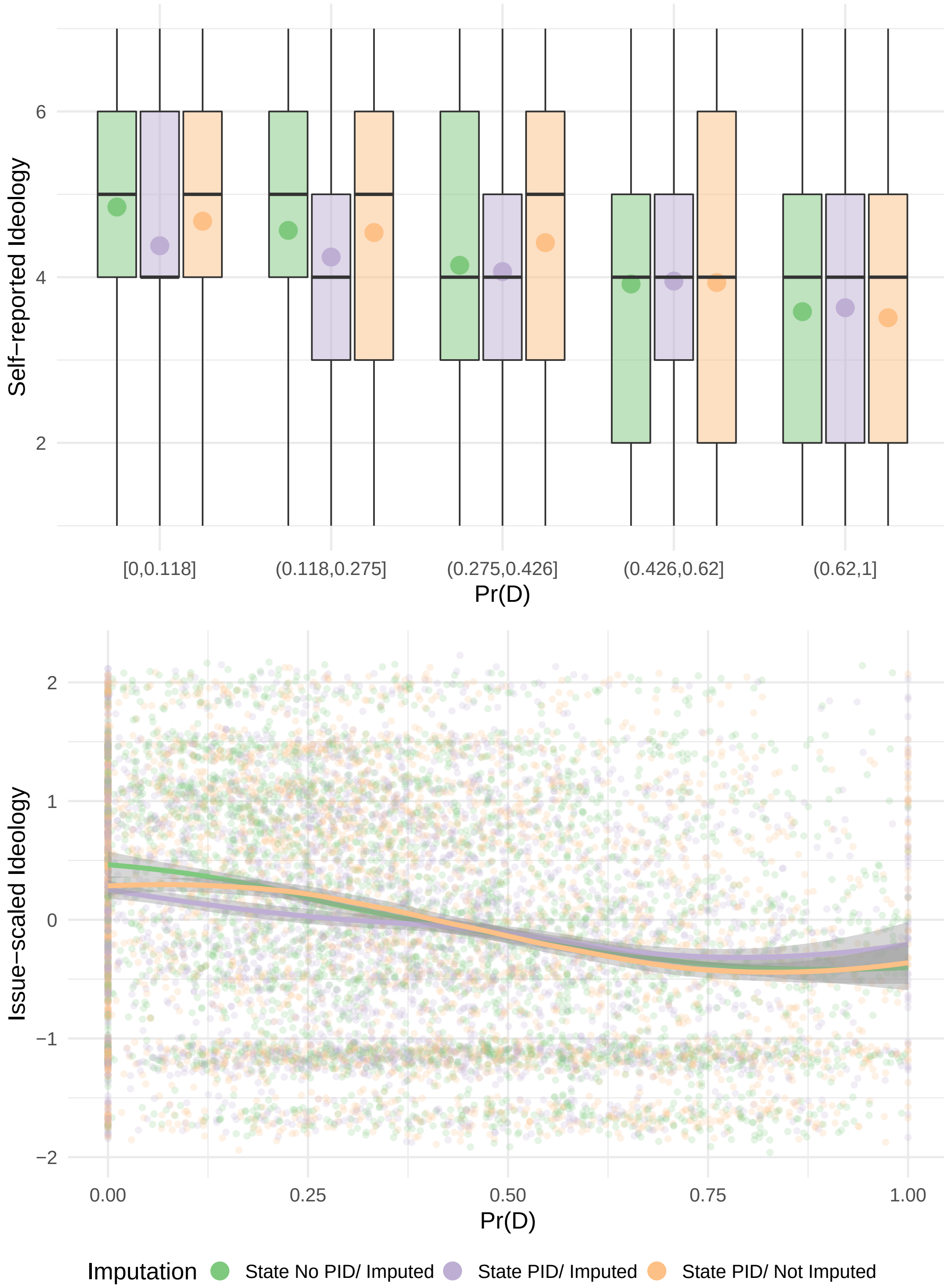


Figure S15: Relationship between $Pr(D)$ and self-reported ideology (top) and with ideology as scaled from issue support (bottom).

7 Relative Exposure Robustness to Dropping Same-household Neighbors

Here, we present the robustness of our Relative Exposure results to calculating Spatial Exposure and Isolation without including neighbors who live in the same household as the voter, leaving a measure of partisan exposure to other voters with which a voter does not live. In day to day life, a good deal of exposure to politics likely comes from people living in the same household, so including household members makes sense. However, for Relative Exposure statistics, the comparison between Democrats and Republicans living in the same geography, we want to know if they are robust to dropping same-household neighbors to demonstrate that the differences between Democrats and Republicans who live in the same town and neighborhood are not attributable only to different patterns in cohabitants, but to actual choices of where to live in relation to one's neighbors.

We identify cohabitants by finding voters registered at the exact same address. Figure S16 presents the distribution of how many neighbors (from the original 1,000) are left across voters after we drop their same-household neighbors. We see that many voters live with 0, 1 or 2 cohabitants and very few live with more than 3 registered cohabitants.[5]

After dropping same-household neighbors, we do see reductions in the differences between the partisan environments of Democrats and Republicans living in the same geographies. However, even down to the neighborhood (Census Tract) level, we still see meaningful differences between Democrats and Republicans. Figure S17 presents the distribution of Relative Exposure across different baseline geographies. We further test whether these differences are statistically significant, by estimating t-tests, weighting by the population of each unit, on the within-geography difference between Democratic and Republican partisan environments.

[5]To appear as neighbors in our analysis, these cohabitants must be registered to vote. This precludes children and other cohabitants who are unregistered by choice or for other reasons.

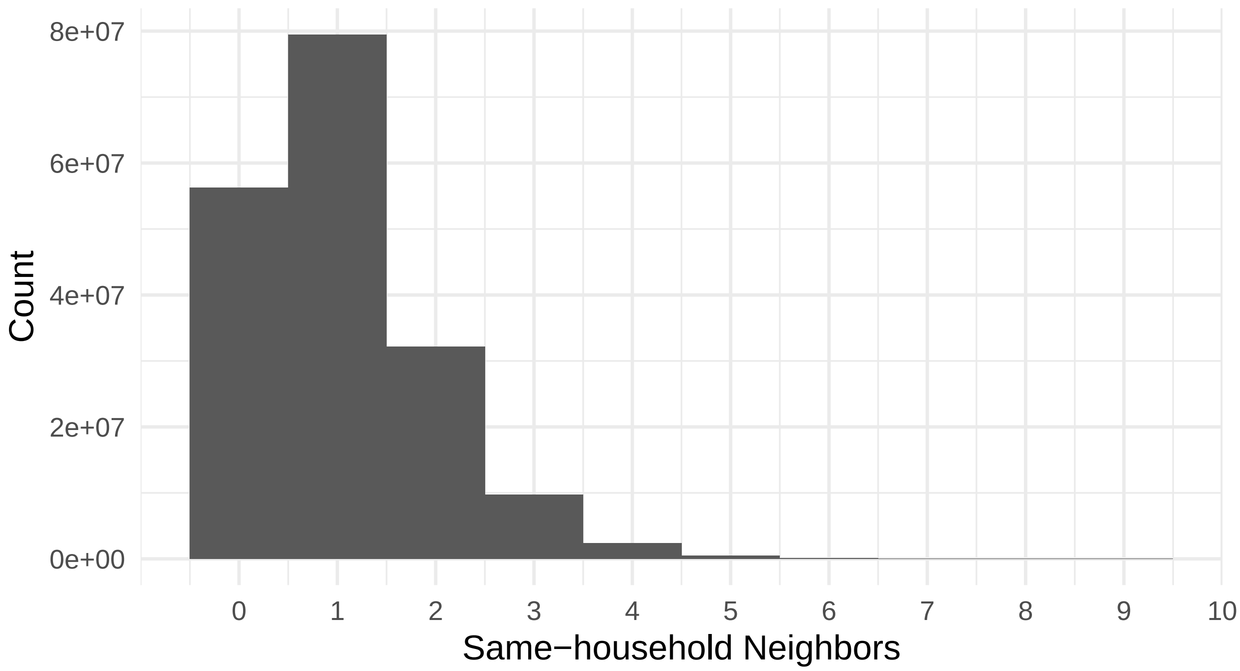


Figure S16: Neighbors after Dropping Same-Household Neighbors

Histogram displays the nationwide distribution of the number of neighbors in our analysis after dropping neighbors who live with the voter.

Tables S11 and S12 present the results of this analysis for the main results (with all 1,000 neighbors) and the results where we drop same-household neighbors. We see that results are consistent in direction and significance across baseline geographies for the main results and the results without same-household neighbors.

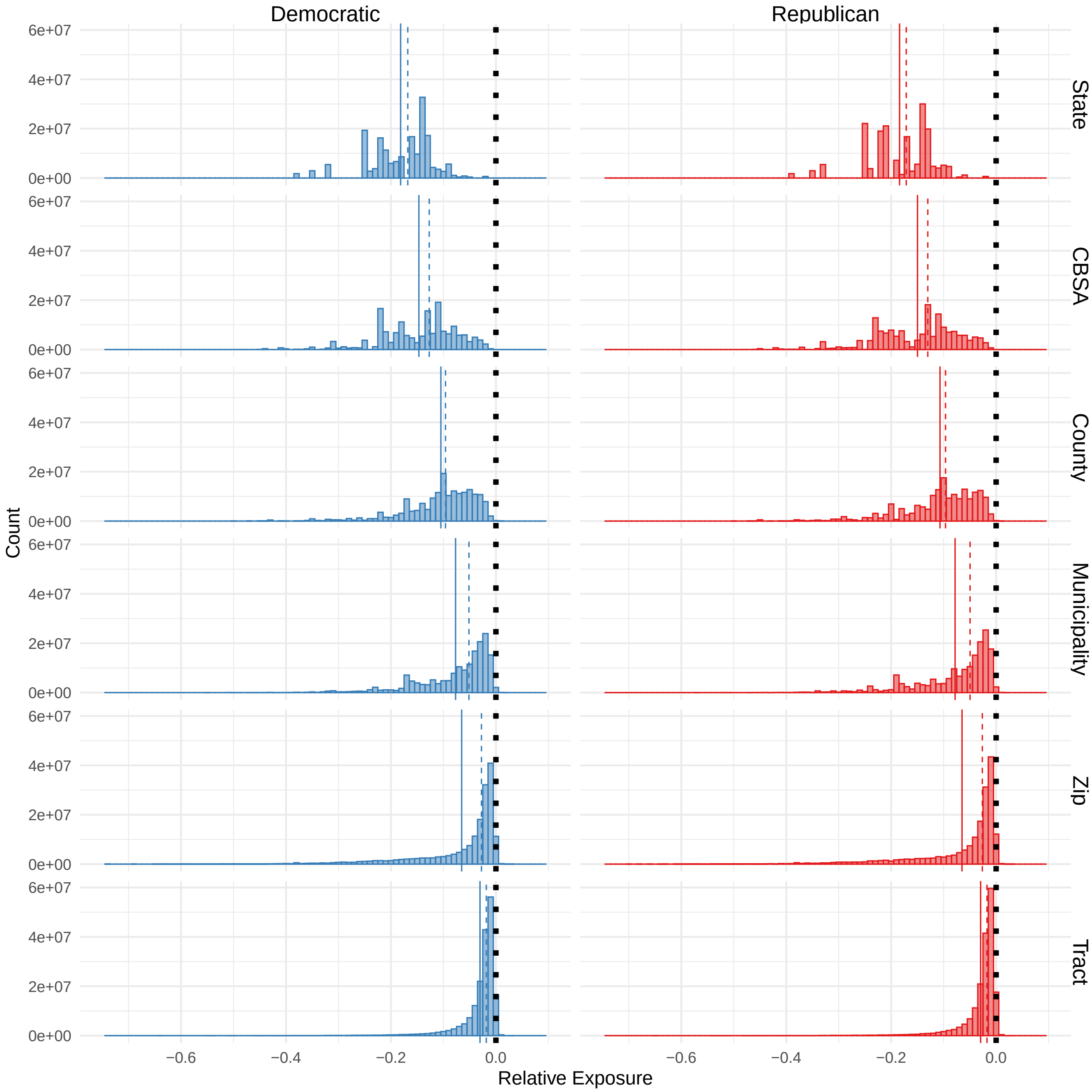


Figure S17: Relative Exposure Without Same-household Neighbors by Geography

Histograms show the weighted nationwide distribution of relative exposure without same-household neighbors across geographic units for Democrats (blue) and Republicans (red). Distributions are weighted by population and the y-axis represents the number of individual voters. Solid vertical lines represent mean values and dashed lines represent median values. Geographies are ordered from bottom to top in decreasing size.

Table S11: Relative Exposure Significance Tests

	Estimate	Std. Error	t value	Pr(>|t|)	Geography	Party
1	-0.256	0.009	-29.611	0.000	State	Democratic
2	-0.256	0.009	-29.224	0.000	State	Republican
3	-0.221	0.002	-96.106	0.000	CBSA	Democratic
4	-0.222	0.002	-92.412	0.000	CBSA	Republican
5	-0.184	0.001	-153.771	0.000	County	Democratic
6	-0.185	0.001	-148.551	0.000	County	Republican
7	-0.157	0.001	-267.322	0.000	City/Town	Democratic
8	-0.156	0.001	-257.349	0.000	City/Town	Republican
9	-0.146	0.000	-365.102	0.000	Zip Code	Democratic
10	-0.144	0.000	-356.961	0.000	Zip Code	Republican
11	-0.113	0.000	-436.321	0.000	Tract	Democratic
12	-0.110	0.000	-431.792	0.000	Tract	Republican

Table S12: Relative Exposure Significance Tests – No Household Neighbors

	Estimate	Std. Error	t value	Pr(>|t|)	Geography	Party
1	-0.182	0.009	-21.325	0.000	State	Democratic
2	-0.184	0.009	-21.143	0.000	State	Republican
3	-0.147	0.002	-59.687	0.000	CBSA	Democratic
4	-0.150	0.003	-57.891	0.000	CBSA	Republican
5	-0.105	0.001	-86.672	0.000	County	Democratic
6	-0.107	0.001	-83.176	0.000	County	Republican
7	-0.077	0.001	-134.022	0.000	City/Town	Democratic
8	-0.078	0.001	-128.570	0.000	City/Town	Republican
9	-0.065	0.000	-186.333	0.000	Zip Code	Democratic
10	-0.065	0.000	-181.528	0.000	Zip Code	Republican
11	-0.030	0.000	-206.084	0.000	Tract	Democratic
12	-0.030	0.000	-199.719	0.000	Tract	Republican

8 Within-Race Partisan Segregation

To test the extent to which partisan segregation is distinct from racial segregation, we compare our measures of partisan segregation to the same measures but with exposure and isolation only calculated among other white voters. Among white voters, the distribution of the difference between Spatial Exposure and Isolation calculated among all voters and only among their white neighbors are narrowly centered around 0, indicating that, on average, partisan isolation within race for whites mirrors general partisan segregation (Figure S18) and that high levels of partisan isolation remain, even when accounting for racial isolation. Figure S19 shows the same for non-whites, where there is more change from baseline results when looking only within group.

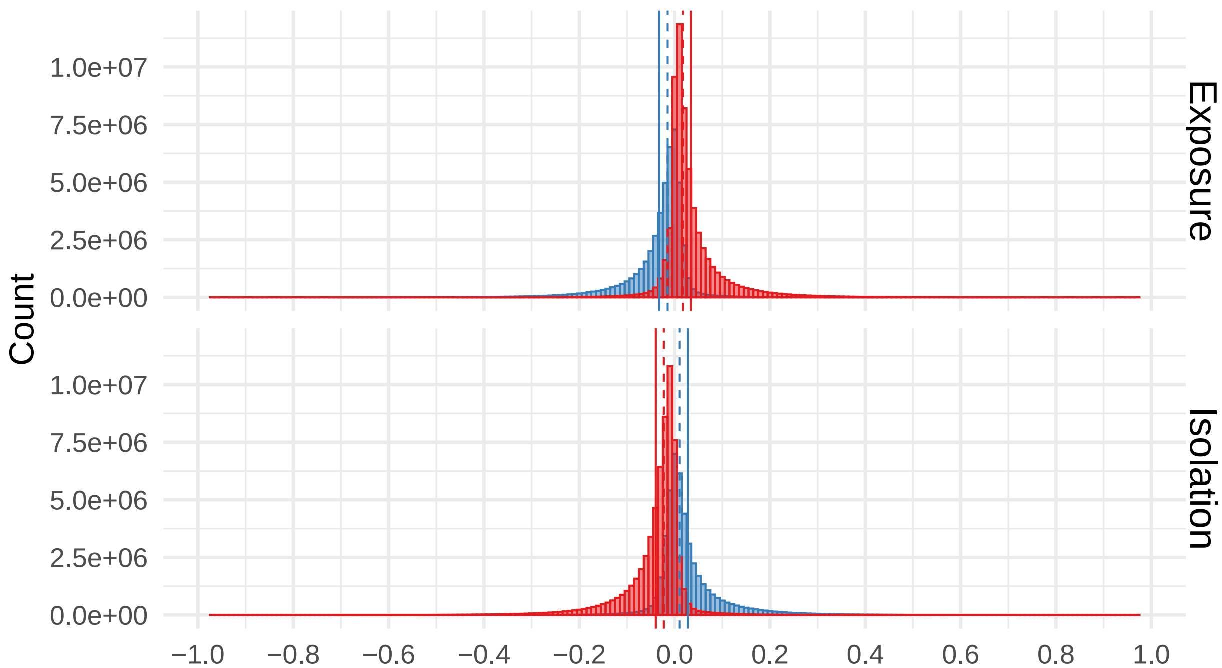


Figure S18: Partisan Segregation vs. White-only Partisan Segregation

Histograms display the distribution for white voters of differences between partisan segregation calculated from all $1,000$ nearest neighbors and partisan segregation calculated only from white neighbors. Positive Isolation values means that a voter appears less isolated by partisanship when we look only at their white neighbors. Positive Exposure values means that a voter appears to have less cross-party exposure when we only look at their white neighbors. Distributions are plotted separately for Democrats (blue) and Republicans (red). Solid lines represent mean values and dashed lines represent median values. Distributions are weighted by posterior partisan probabilities.

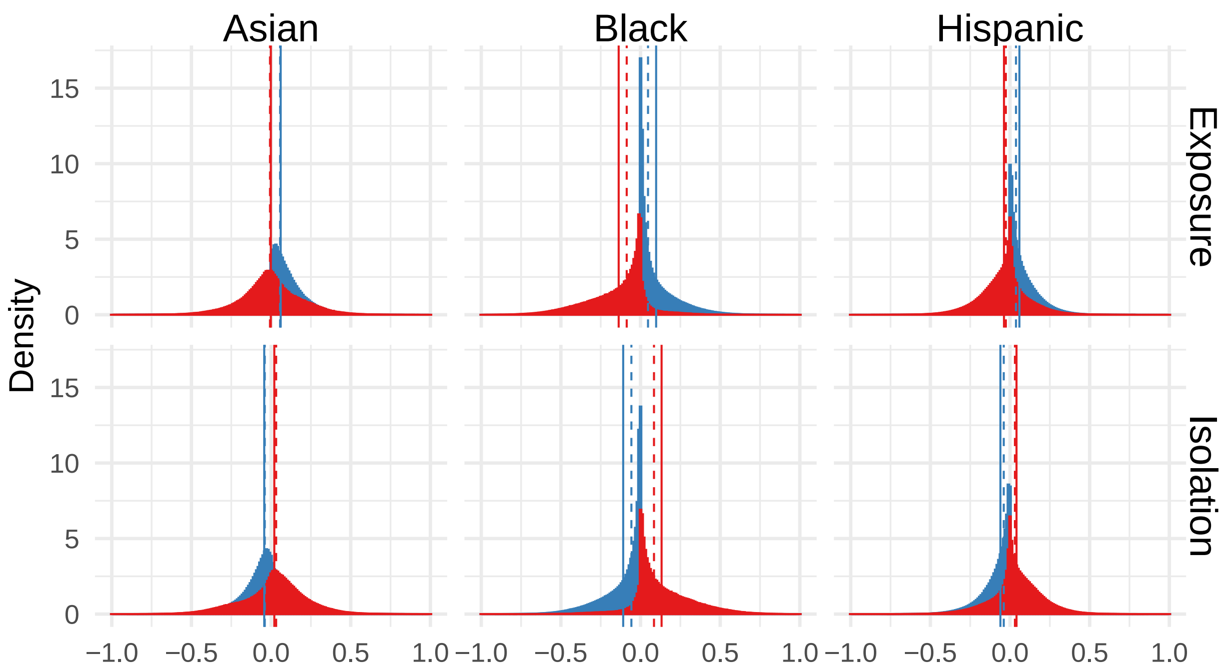


Figure S19: Partisan Segregation vs. Within-race Partisan Segregation

Histograms display the separate distributions for Black, Hispanic and Asian voters of the differences between partisan segregation calculated from all $1,000$ nearest neighbors and partisan segregation calculated only from neighbors of the same race. Positive Isolation values means that a voter appears less isolated by partisanship when we look only at their same-race neighbors. Positive Exposure values means that a voter appears to have less cross-party exposure when we only look at their within-race neighbors. Distributions are plotted separately for Democrats (blue) and Republicans (red). Solid lines represent mean values and dashed lines represent median values. Distributions are weighted by posterior partisan probabilities.

References

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. “The Statistical Analysis of Roll Call Data.” *American Political Science Review* 98(2): 355–370.

Niemeyer, Gustavo. 2008. “Geohash.org.”.

Tausanovitch, Chris, and Christopher Warshaw. 2013. “Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities.” *The Journal of Politics* 75(02): 330–342.