The Measurement of Partisan Sorting for 180 Million Voters

Jacob R. Brown[1*] and Ryan D. Enos[2*]

[1] Institute for Quantitative Social Science and Department of Government, Harvard University, Cambridge, Massachusetts, USA

[2] Institute for Quantitative Social Science and Department of Government, Harvard University, Cambridge, Massachusetts, USA

[*] Corresponding authors: Jacob R. Brown (jrbrown@g.harvard.edu); ORCID 0000-0001-8037-5360; Ryan D. Enos (renos@gov.harvard.edu); ORCID 0000-0002-5855-7464.

Abstract

Segregation across social groups is an enduring feature of nearly all human societies and is associated with inter-group conflict, prejudice, inefficient resource allocation, and poor democratic governance. In many countries, reports of growing geographic political polarization have also raised concerns about the stability of democratic governance. To explore political polarization, we create the largest individual-level mapping of human social-group sorting on record. Using advances in spatial data computation, we implement the first reported large-scale measures of individual segregation by calculating the local residential partisan segregation of every registered voter in the United States, creating a spatially-weighted measure for over 180 million individuals. With these data, we present new evidence of extensive partisan segregation in the United States. A large portion of voters live with virtually no exposure to voters from the other party in their residential environment. Such high-levels of partisan isolation can be found across a range of places and densities and are distinct from, and sometimes in tension with, racial segregation. Moreover, Democrats and Republicans living in the same city — or even the same neighborhood — are segregated by political party.

Introduction

Segregation between human social groups is associated with a range of profoundly negative outcomes, including inter-group conflict, prejudice, inefficient resource allocation, poor democratic governance, and other socially deleterious effects (1; 2; 3; 4). Perhaps this is because interpersonal contact is an important determinant of intergroup relations (5; 6) and, more generally, interpersonal connections within social networks have been shown to

have significant influences on behavior (7; 8; 9; 10; 11). Drawing on these associations and using aggregate data, popular and scholarly accounts of politics in the United States and, increasingly, other Western democracies, describe stark partisan segregation, with members of different political parties living separate lives, resulting in partisan rancor and threatening the functions of the democracy (12; 13; 14; 15; 16). Yet, despite the association between segregation and important outcomes and the claims of increasing partisan segregation, the measure of segregation among partisans, as with the measurement of segregation for most social groups, is severely limited: researchers must usually rely on data aggregations that do not include the actual locations of individuals and, thus, measurements are limited to summaries across large geographies and the experience of individual exposure across groups is masked.

Using data on the exact residential address of every registered voter in the United States and harnessing advances in spatial data computation, we measure the local partisan segregation for each of these voters, creating a spatially-weighted cross-partisan exposure measure for over 180 million individuals. These data constitute the first measure of individual spatial segregation at this scale for any society and yield new evidence of the extent of partisan segregation in the United States, allowing us to examine the degree to which individuals are sorted by partisanship with respect to individual neighbors and within small geographic units, such as cities or neighborhoods.

A large portion of US voters live in partisan bubbles with virtually no exposure to the other party. In particular, the modal Democrat in the US lives in extreme political isolation, where more than 95% of their encounters in their residential environment are expected to be with other Democrats, implying little exposure to competing political ideas from neighbors. Similarly high-levels of partisan isolation are also present for Republicans living in certain areas and, for voters of both parties, high levels of segregation can be found across a range of places and densities and are distinct from, and sometimes in tension with, racial segregation.

Moreover, even when Democrats and Republicans live in the same city — or even the same neighborhood — they are residentially sorted by political party.

In the United States, political party affiliation is considered a social identity, analogous to race or religion (17), and is a powerful predictor of a host of attitudes and behaviors (18), including behaviors outside of the explicitly political realm (19; 20).

Because partisanship is correlated with political ideology and other attitudes and behaviors, the extent of a voter's partisan isolation is likely to affect their exposure to individuals different from themselves and to competing sociopolitical viewpoints, thus affecting a range of important outcomes. Cross-group exposure can be consequential for the shaping of intergroup attitudes and behaviors (6), including the prejudicial attitudes that are leveled across parties in the United States (21). Isolated partisan environments may also affect behavior through channels other than (a lack of) interpersonal contact: indeed, human behavior can be shaped by low-level environmental cues (22; 23), such as the norms displayed by neighbors, and randomized-controlled-trials have shown that political messaging from neighbors, such as the posting of yard-signs, has a persuasive effect on voting behavior (24).

Isolation may also contribute to the increasing ideological extremity on both the mass and elite levels in the United States (25): in the marketplace for political ideas, exposure to out-partisans may allow for the transmission of competing views (26) that can reduce extremism (27). Furthermore, the extremity of political views is correlated with political participation (28) and participation is correlated with influence (29), raising the potential that the most isolated partisans are exerting the most influence over politics.

Of course, even with residential segregation, cross-group exposure may happen in other environments (e.g., (30; 31)). But residential segregation has been shown to shape attitudes and behavior even when accounting or interpersonal cross-group contact (4) and residential segregation may also affect campaign and other elite behavior by allowing politicians to narrow the ideological appeal of their message (32; 33) thereby lessening the potential for

voters to be exposed to cross-cutting appeals and allowing politicians to avoid moderating their messages.

Segregation has also been shown to reduce cooperation for shared benefit across groups (4) and the division of partisans in geographic space is associated with levels of trust in government and anti-system attitudes (34). Once in place, the contribution of segregation to these behaviors can become self-reinforcing if partisans avoid living in areas where they would be a minority, thus even initially small levels of clustering could drive extreme segregation (35), further separating partisans and reinforcing behavioral and attitudinal separation (1).

Results

This research was approved by the Harvard University Committee on the Use of Human Subjects. To calculate partisan segregation, we rely on data containing information on every one of the 180,735,645 registered voters in the United States as of June 2018. Until very recently when some states introduced automatic registration of citizens, voter registration was voluntary in nearly every state. As such, our data include about 80% of the voting eligible population, which is about 92% of the approximately 250 million persons in the United States over 18 years old, and does not include non-citizens and people not allowed to vote in certain states because of felony convictions and other reasons. With these data, we have the social group membership and exact address of almost 75% of the adult population of the United States, making ours the largest-scale mapping of contemporary individual-level local social environments ever published ((36) use historical US Census data to determine the next door neighbors of every person reached by the 1940 US Census).

When a person registers to vote, they provide a home address and, in most states declare affiliation with a political party. We use these data to construct measures of segregation, leveraging advances in geographic data science: using Geohash techniques that store latitude and longitude coordinates as strings rather than locations, we can efficiently measure the

spatial relationships between large numbers of individuals (37). For each of $n = 180,660,202$ geocoded individuals, we measure the distance to their $k = 1,000$ nearest neighbors as defined by the closest geodesic distances from the registered voters' residences, creating a distance measure for $n \times k$ (over 180 billion) dyadic relationships. Thus, for every voter, we identify how close they live to each of their $1,000$ nearest neighbors and combine this information with data on their neighbors' partisanship to construct individual-level measures of partisan exposure and isolation. We decided on $k = 1,000$ by testing a random sample of 1,000 voters with $k = 50,000$ and found little additional information after $1,000th$ k (see Supporting Information).

With these data, we are able to be the first to implement spatially-weighted measures of segregation (38) at a large scale. Measures of segregation typically available to researchers are not spatially-weighted and do not use individual data, and therefore are based on the composition of groups in a chosen geographic unit. These measures usually must make the unrealistic assumption of common geographic context for all individuals within the chosen unit and are sensitive to common problems of aggregate measurement, including the Modifiable Areal Unit Problem (MAUP) and problems of scale, which mean that measures of segregation can be extremely sensitive to researcher choices of geographic unit.

A standard measure of segregation is *Exposure*, which captures the extent that members of one group encounter members of another group (or their own group, in the case of *Isolation*) in their local environments (39). This and other standard measures of segregation are aspatial, making the stringent assumption that where individuals live in relation to each other within the geographic unit has no bearing on exposure. This assumption creates a *checkerboard problem*, wherein the measures are unable to distinguish between different spatial distributions of individuals even if the likely exposure was very different across the distributions (40). For example, in Figure 1, using the standard measure of exposure the levels of segregation on the top left and top right would be the same, despite the starkly

different spatial relationships across groups. When measuring partisan segregation, because partisans are known to cluster in certain types of places, for example Democrats in the densest parts of cities (41), not accounting for the spatial relationship of voters may lead to highly misleading inferences (Figure 1).

By measuring the distance between individuals, our measures are not subject to these issues. Our primary measure is a weighted average of exposure (38) where the proportion of each party in an individual's 1,000 nearest neighbors is weighted by the inverse of the distance in meters from each neighbor ($w_k = \frac{1}{d+1}$). Weighting by distance gives greater emphasis to an individual's closest neighbors, so the partisanship of a next door neighbor is more important when describing partisan exposure than the neighbor further away (Figure 1). These individual-level measures of *Spatial Exposure* also allow us to flexibly explore segregation at any level, from segregation from one's most immediate neighbors up to any arbitrarily large level of geography (Figure 2).

In measuring partisan segregation, we confront the challenge of how to account for voters who cannot or do not explicitly declare membership in a party. Partisanship is recorded at the time voters register in 31 states and the District of Columbia. If we were to measure segregation only in these states, we would miss large sections of the US. Furthermore, in all states, some voters choose not to register with one of the two major political parties, even if they have the option of doing so. If we were to measure segregation only among voters officially registering as Democrats or Republicans, we would clearly misrepresent the levels of isolation or exposure to voters with similar or different political ideologies: extensive evidence shows that all but a small proportion of officially independent and minor-party voters have stable preferences for one of the major parties and ideological orientations indistinguishable from those of major party members (42; 43; 44; 45).

Thus, before constructing measures of segregation, we impute partisanship for voters not registered as Democrats or Republicans. Such imputation techniques, relying on similar

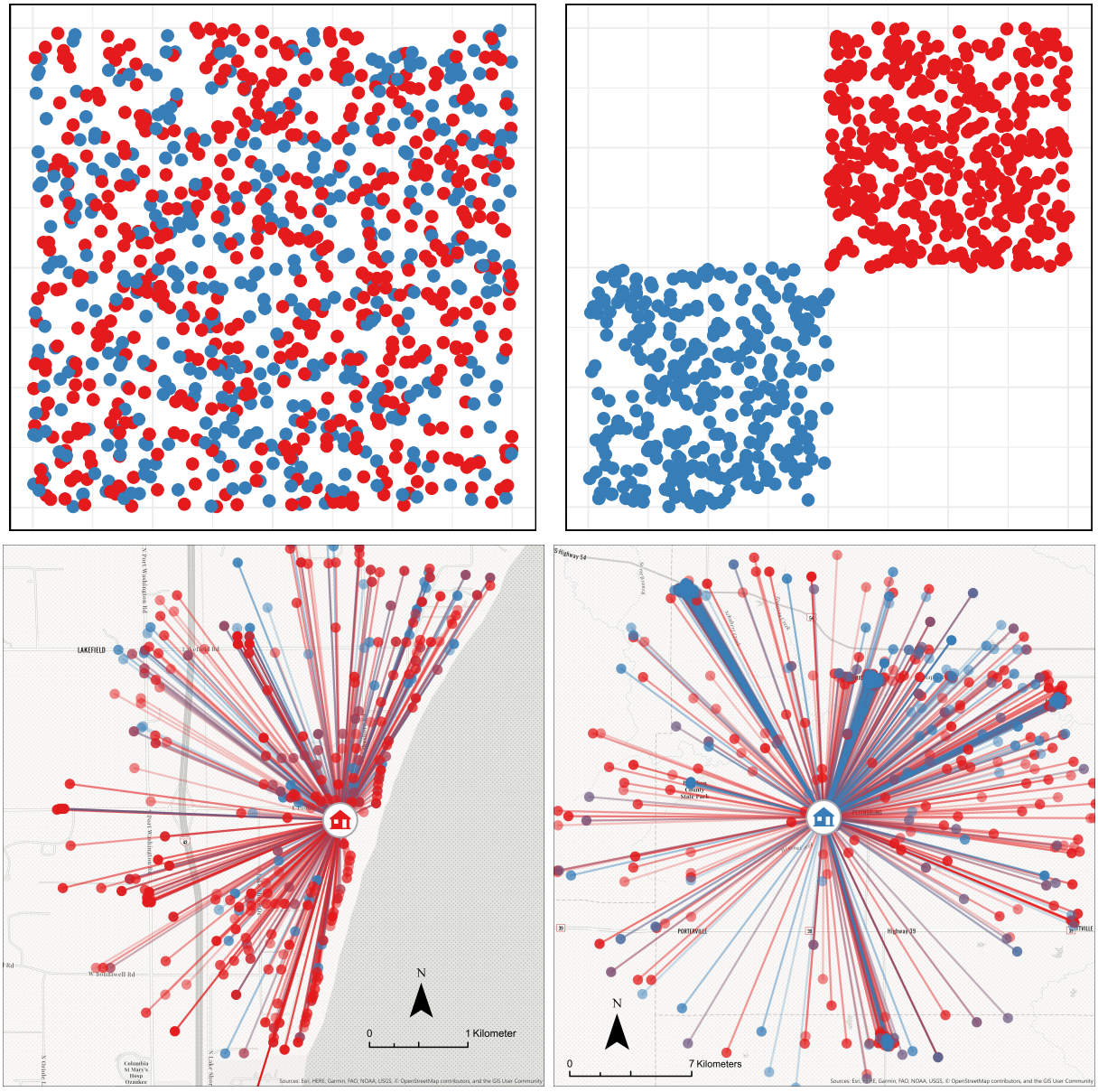

Figure 1: Using aspatial measures of segregation, the average exposure of red and blue individuals in the top left and and top right is the same. Spatial weighting captures the apparent much higher segregation in the top left. Drawing on real data, the Republican, designated by the red house, in a suburb of Milwaukee, WI (bottom left) has an unweighted Democratic exposure of .36 but because other Republicans are also clustered along the lake shore, many less than 1km away, the Spatial Exposure to Democrats falls to .15. The Democrat in rural southeastern Kansas (bottom right) has an unweighted Republican exposure of .64 but because other Democrats in the area are clustered into the centers of the small towns more than 7km away, the voter is isolated from other Democrats and Spatial Exposure to Republicans is .98.

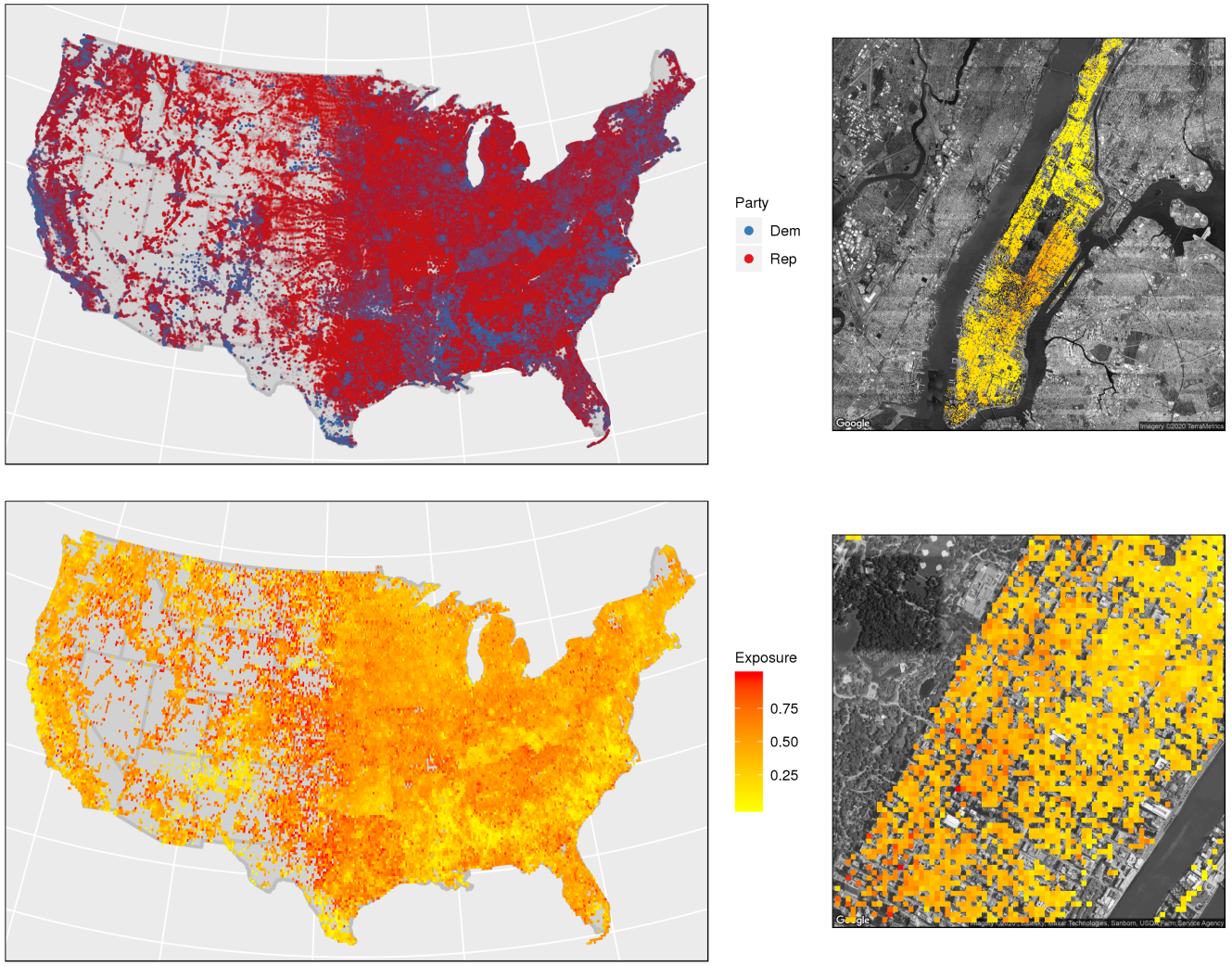

Figure 2: Measuring Spatial Exposure across increasingly small geographies. The exact residential location of every Democrat and Republican in the United States (top left) can be used to measure each Democrats' Spatial Exposure to Republicans and this can be averaged across arbitrarily small grid cells for display purposes (bottom left). Exposure can be examined across any resolution: in Manhattan, NY (top right), the drastically different residential exposure to Republicans can be seen, with Democrats on the northern and southern extremes of the island having virtually no residential exposure to Republicans, while Democrats on the Upper East Side have exposure as high as 50% due to the clustering of Republicans in this area. Zoomed into the Upper East Side of Manhattan (bottom right), the clustering of Republicans along Central Park and, thus, Democrats' decreasing exposure to Republicans moving northeasterly, can also be seen.

information, are commonly used by for political campaigns. We impute through a three-step process where we first code a voter as a Democrat or Republican based on the last partisan primary in which they cast a ballot, e.g., if a voter votes in a primary election to select Democratic candidates for office, we impute that voter as a Democrat. Next, we classify voters registered to a third-party with a clear left/right ideological lean as Democrats (left) or Republicans (right) (see Supporting Information). With these updated counts, we then impute partisanship for the remaining independents through a Bayesian process using priors constructed from 2016 precinct-level presidential vote share and individual-level demographic characteristics. This is a process similar to imputation methods for race successfully used in academic research (46; 47) and by political campaigns (33). Through this process, we classify 89% of voters not registered to a major party as leaning toward either Democrats or Republicans, which is very close to the proportion of non-partisan voters who consistently vote either Democrat or Republican according to previous studies (43) (see Supporting Information for details of imputation process, summary statistics, and the sensitivity of our results to this imputation). Then using contact information from the voterfile, we validate these imputations in a survey of 12,221 voters, finding that we impute voters' partisanship with an accuracy of 77%, which is near the maximum accuracy that may be possible given the instability observed in previous surveys of self-reported partisanship (i.e., (17)), indicating a high degree of accuracy for our method of imputation. We also show that voters with imputed partisanship are nearly ideologically identical to non-imputed voters, both in states with partisan registration and those without, indicating that our imputation of those not registered as partisans accurately models the exposure to competing political ideas experienced when sharing residential environments with registered partisans.

These imputations yield a probabilistic score of being a partisan. When calculating isolation for voter i, we weight the contribution of each neighbor k by these probabilities, which will, on average, recover the probability of encountering members of the other party in their

local environment. When making aggregate summaries of Democratic and Republican Exposure (Isolation), we also weight central tendencies and distributions by posterior partisan probabilities. We can also use these scores to assign each voter a discrete party affiliation based on the party with the highest probability for each neighbor. In this case, exposure is unweighted by party. In the Supporting Information, we present results with party classified in this way and with no imputation of partisanship of any kind.

Spatially weighted measures of Partisan Isolation and Exposure are defined as:

$$\text{Spatial Isolation}_i = \frac{\sum_{k=1}^{1000} \frac{1}{(d_k+c)^a} \mathbb{P}(p_k = p_i)}{\sum_{k=1}^{1000} \frac{1}{(d_k+c)^a}}$$

$$\text{Spatial Exposure}_i = \frac{\sum_{k=1}^{1000} \frac{1}{(d_k+c)^a} \mathbb{P}(p_k = q_i)}{\sum_{k=1}^{1000} \frac{1}{(d_k+c)^a}}$$

where p_i is the partisan identification of voter i, q_i is the opposite partisan identification (if Democrat, Republican, if Republican, Democrat) of voter i, p_k is the partisan identification of neighbor k, $\mathbb{P}(p_k = p_i)$ is the posterior probability that neighbor k has the same partisanship as neighbor i, d_k is the distance in meters neighbor k lives from voter i, c is a constant adjustment made so that when $d_k = 0$ the expression is not undefined, and a is an exponent to which we raise the denominator of the distance weight — to control how much weight is given to proximity in the measure. In the main analysis we set $c = 1$ and $a = 1$. Setting c higher would decrease the weight given to the smallest distances, and setting a higher would give greater weight to distance in general, placing even more emphasis on the closest neighbors and increasing the intensity of segregation (in the Supporting Information, we show results with other weighting schemes).

Spatial Exposure and Isolation represents a person's residential partisan experience. Of course, there can be partisan exposure, outside of the residential context, for which we cannot account and, within a voter's residential context, other variables, such as the density

of non-voters, can influence the likelihood of interaction with partisan neighbors. Spatial Exposure and Isolation represent the interactions that come from the spatial arrangement of partisans: the likelihood, all else being equal, of encountering a neighbor of a given party in one's residential life. Exposure ranges from 0 to 1, with 0 being no exposure to the other party and 1 being only exposure to the other party. Isolation of 1 is perfect isolation, encountering only one's own party, and 0 is encountering only the other party. Exposure of .01 would mean that we expect, all else equal, only 1 out of 100 interactions in a voter's residential context to be with a person from the other party. Exposure to the out-party is not necessarily the inverse of isolation because a small portion of voters are true independents, neither registering with nor otherwise aligning with either party. Thus, voters not perfectly isolated from their own party might still not have any exposure to voters from the other party.

With the spatially weighted measure of exposure in hand, we can see how the probability of interaction with an out-party member can be dramatically different when not accounting for the spatial distribution of neighbors (Figure 1): for half of voters, not accounting for distance distorts exposure by 22% or more; for 25% electorate, the distortion is 40% or more; and, in some extreme cases, the distortion exceeds 100% (see Supporting Information for full distribution of standard versus Spatial Exposure).

Average Exposure

The average Democrat's exposure to the out-party is 0.30 and the average Republican's exposure is 0.36. Given that the nationwide proportion of Democrats and Republicans are 51% and 43% respectively, this represents substantially lower cross-party exposure than would be expected if partisans were not sorting into different residential environments. But, moreover, the national distribution (Figure 3) shows a large portion of voters living in extreme isolation: nationwide, the modal Democrat lives with virtually no exposure to out-partisan

neighbors (Exposure < 0.05) and a majority of Democrats and Republicans live with isolation levels well below the threshold of 0.60 set by (1) for high isolation in the context of race in municipal areas. At these levels, for the modal Democrat in the US less than 1 in 10 of their interactions in their residential environment will, on average, be with a Republican.

Democrats and Republicans are Segregated in Most Types of Places

Partisan segregation is not distributed evenly across the US and, thus, the level of residential interaction across party lines will vary significantly depending on where a voter lives. The flexibility of individually measured segregation allows us to describe segregation across places defined in any way and, thus, we can see how partisans are likely to interact across the suburbs and central cities of both large and small urban areas, as well as rural locations.

In major urban areas (Core Based Statistical Areas (CBSAs) with over 1 million residents, the largest of which is the New York-Newark-Jersey City area, the smallest of which is the Tuscon, Arizona area) Democratic exposure to Republicans is extremely low, especially in the dense urban cores (Figure 4). Notably, a large plurality of Democratic voters live in these areas and the very-low levels of exposure extend even to the medium density suburbs of these major areas and to minor urban areas (those with less than 1 million residents, the largest of which is Honolulu).

It is only when reaching areas where relatively few voters are located — the very low density places on the fringe of or entirely outside of urban areas — that Republican exposure to Democrats becomes lower than Democratic exposure to Republicans. In these areas, many Republicans live with extremely low levels of exposure to Democrats and the typical Republican begins to experience the sharp isolation characterizing the experience of the typical Democrat in more dense locations. But isolation in rural areas is not only present for Republicans: the modal Democrat in low density places outside of urban areas, often rural Hispanics, African Americans, and American Indians, also has levels of exposure less than

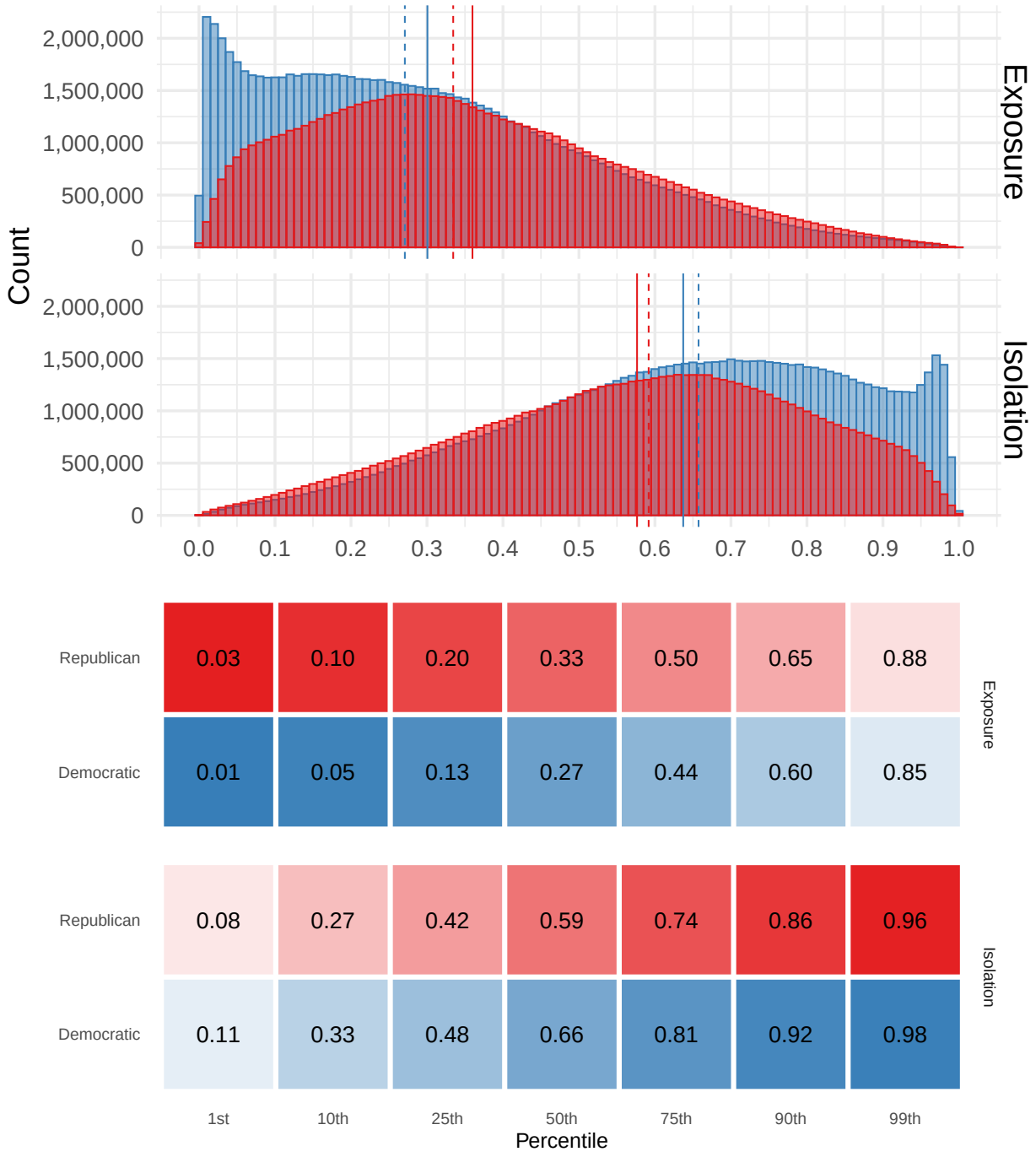

Figure 3: Nationwide distribution of partisan Spatial Isolation and Exposure separately for Democrats (blue) and Republicans (red). Solid vertical lines represent mean values and dashed lines represent median values. Colored cells present spatially weighted proportion of out-party (Exposure) or in-party (Isolation) neighbors across percentiles. The distributions are weighted by the posterior partisan probabilities.

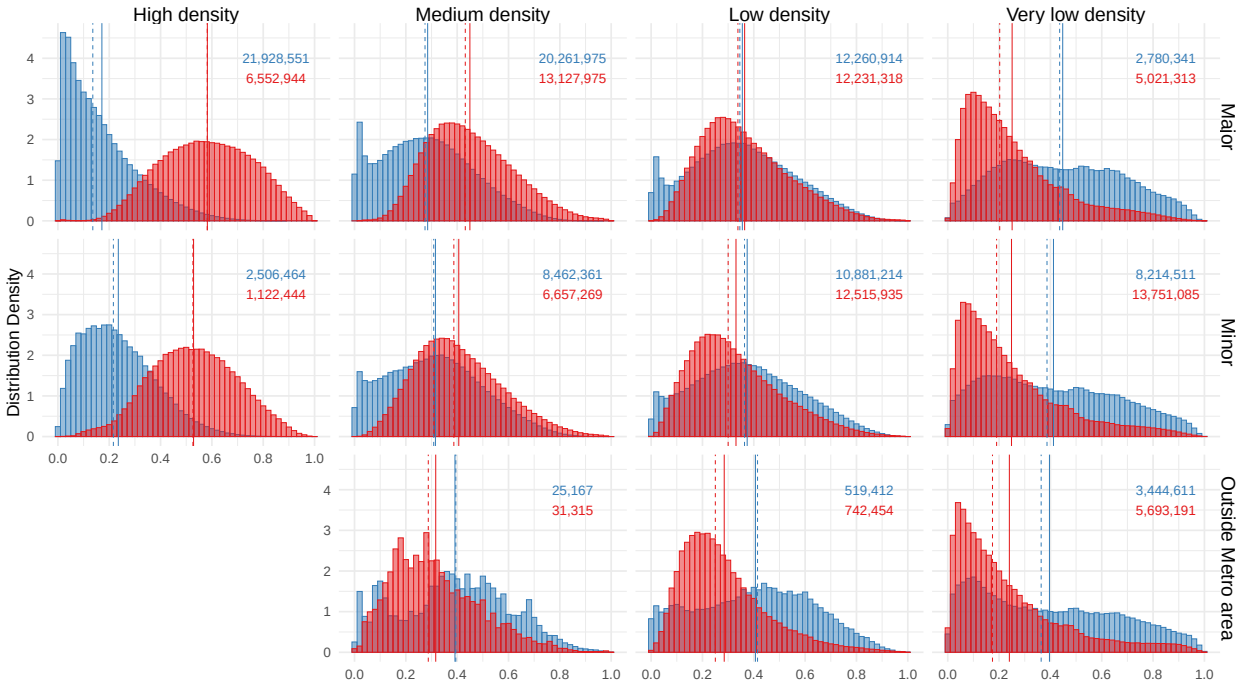

Figure 4: Isolation by Density and Urban Area. Histograms shows individual-level Spatial Partisan Exposure subset by the population density of the voter's census tract and the type of urban area in which they live, separately for Democrats (blue) and Republicans (red). Solid vertical lines represent mean values and dashed lines represent median values. Distributions are weighted by posterior partisan probabilities. Numbers in blue and red represent estimates of number of Democrats and Republicans, respectively, in the urban area-density subset.

0.10.

Even in the low density outer suburbs of the major, and some minor, urban areas, a significant portion of Democrats have virtually no exposure to Republicans in their residential environment and the modal Republican also has low levels (≈ 0.20) of exposure to Democrats. These low levels of exposure, for both Republicans and Democrats, in the low density suburbs of major urban areas suggests that segregation is not only a product of the sorting of Democrats into central cities and Republicans into suburbs — rather, even within the suburban fringes of urban areas, Democrats and Republicans are separated.

Partisans Sort Even Within Neighborhoods

The severity of residential segregation demonstrates significant separation of partisans across the United States, but the extent of sorting can be better understood by comparing cross-party exposure of Republicans and Democrats living within the same places. A portion of the high isolation experienced by Democrats in the dense areas of large cities, for example, is to be expected because many more Democrats live within these large cities than do Republicans. But to what extent do Democrats and Republicans living in the same areas still have different levels of partisan exposure?

We compare the exposure of partisans to that of out-partisans living in the same geographic area, and find that even conditional on the choice to live in a larger geography, say within a particular city or even within the same neighborhood, Democrats and Republicans still cluster with voters from their own party. We demonstrate this by constructing an index of Relative Exposure, defined as the difference between the average Spatial Exposure of partisans within a geographic unit to the average Spatial Isolation of out-partisans also living in that geographic unit. So, the Relative Exposure for Democrats in a geographic unit would be the difference between average Democratic exposure to Republicans minus Republican exposure to Republicans within that unit. This captures the extent to which members of one

party experience different partisan environments than members of the other party living in the same geography. If, after moving into said geography, no further sorting were occurring, we would expect this difference to be 0. So, for example, a city where Republicans have a relative exposure of -0.20 would be a city where out of a voter's 100 nearest neighbors, Republicans on average have 20 more nearest neighbors who are also Republicans than the average Democrat has Republican neighbors. Notably, such comparisons across groups in the same geographic area are not possible when using traditional aspatial measures of exposure.

We calculate Relative Exposure across a range of geographies. Because we measure the exposure of individuals, we can measure Relative Exposure in any arbitrarily defined unit and therefore avoid problems such as MAUP. But to facilitate comparison with previous research, we use commonly used geographies and compare Democrats and Republicans living in the same state, urban area (CBSA), county, city/town, ZIP Code, and Census Tract. At every level of geography, both Republicans and Democrats have Relative Exposure far lower than 0, indicating that substantial sorting does occur (Figure 5). Large scale sorting can be seen at the state level, likely driven by the clustering of Democrats into urban areas, but even within cities Democrats and Republicans sort into different places, again indicating that partisan segregation is not merely a result of large-scale geographic trends, such as a urban-rural divide. Moreover, even within Census Tracts, a small level of geography often used in social science research to represent a neighborhood, both Democrats and Republicans cluster more with their co-partisans than out-partisans. Even at these small geographic levels, the difference between Democrats' and Republicans' partisan environments is 11 percentage points greater than we would expect if there was no partisan sorting within the geography. This disparity indicates that, even after the Democrats and Republicans make the choice to live in similar neighborhoods, they still live with noticeably different levels of partisan exposure (in the Supporting Information, we demonstrate the robustness of these results to dropping neighbors who live in the same household as the voter).

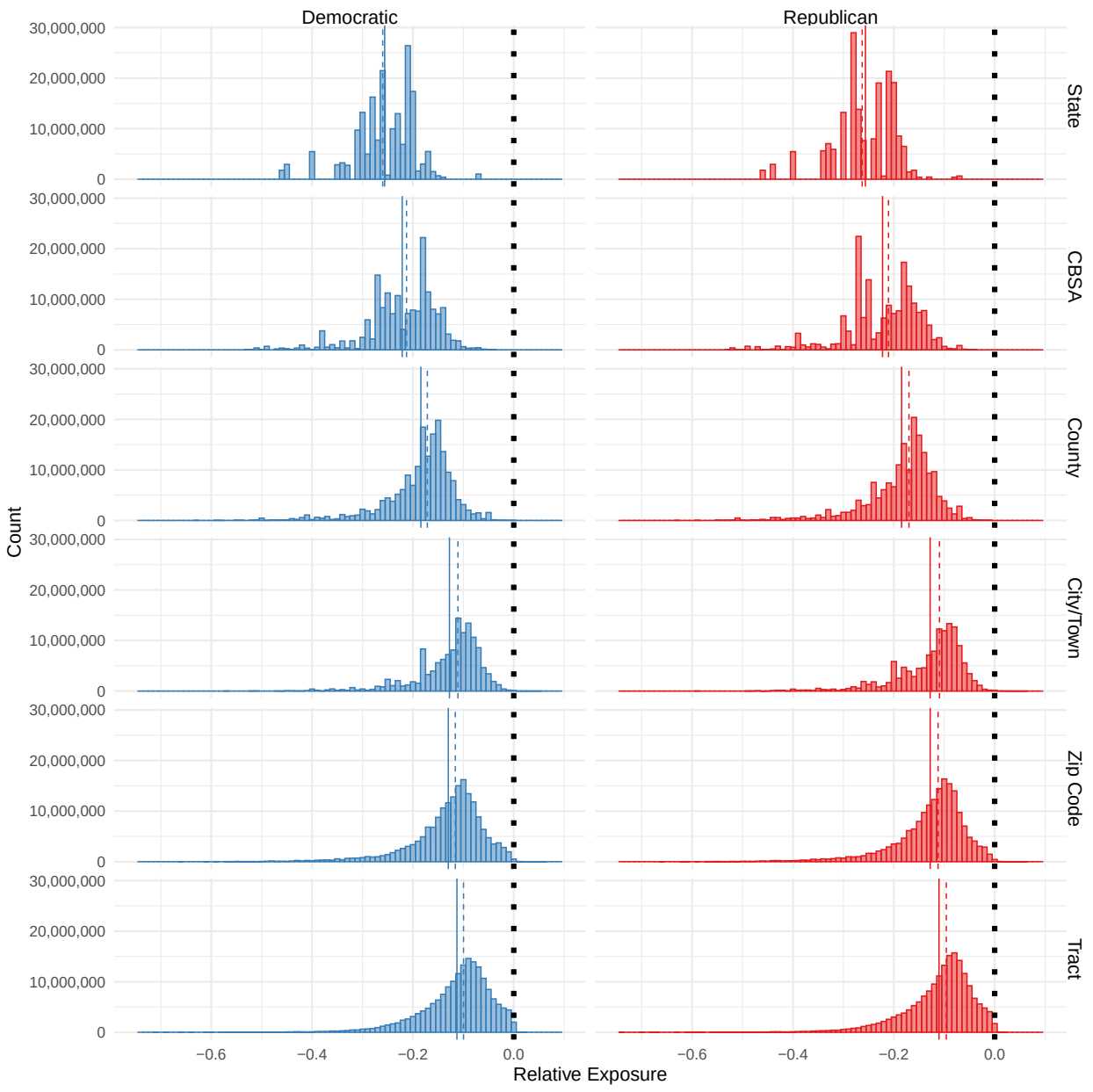

Figure 5: Relative Exposure by Geography. Histograms show the weighted nationwide distribution of relative exposure across geographic units for Democrats (blue) and Republicans (red). Distributions are weighted by population and the y-axis represents the number of individual voters. Solid vertical lines represent mean values and dashed lines represent median values. Geographies are ordered from bottom to top in decreasing average size.

Partisan and Racial Segregation are Distinct

In the US, race is highly correlated with partisanship such that Democrats are much more likely than Republicans to be non-white. As such, some partisan segregation will likely be a function of the significant racial segregation in the US. Because we can observe the race of individuals on the voterfile, we can examine how much partisan sorting is merely a function of racial sorting.

We compare partisan exposure measured among all neighbors to the same measures but with exposure only measured among neighbors of the same race. Among white voters, the distribution of the difference between exposure calculated among all voters and only among their white neighbors are narrowly centered around 0, indicating that, on average, partisan exposure within race for white voters mirrors general partisan segregation and, thus, partisan segregation is not only a function of racial segregation (see Supporting Information for full distributions).

In fact, for white Democrats, racial segregation may actually reduce the levels of partisan segregation that would be expected if more white Democrats lived near non-whites. Partisan exposure, subsetted to whites and non-whites by party (Figure 6) reveals high partisan isolation of non-white Democrats and, among the relatively small numbers of non-white Republicans the highest exposure of any group. This is likely attributable to the forces of racial segregation clustering non-whites together, regardless of partisanship. On the other hand, levels of Spatial Exposure for white Democrats and white Republicans are largely similar and higher than for non-white Democrats. This discrepancy between white and non-white voters suggests that white voters sorting away from non-whites (e.g., (48)) increases their exposure to white members of the outparty, thus increasing the levels of partisan exposure over what would be present if partisan segregation were merely a function of racial segregation.

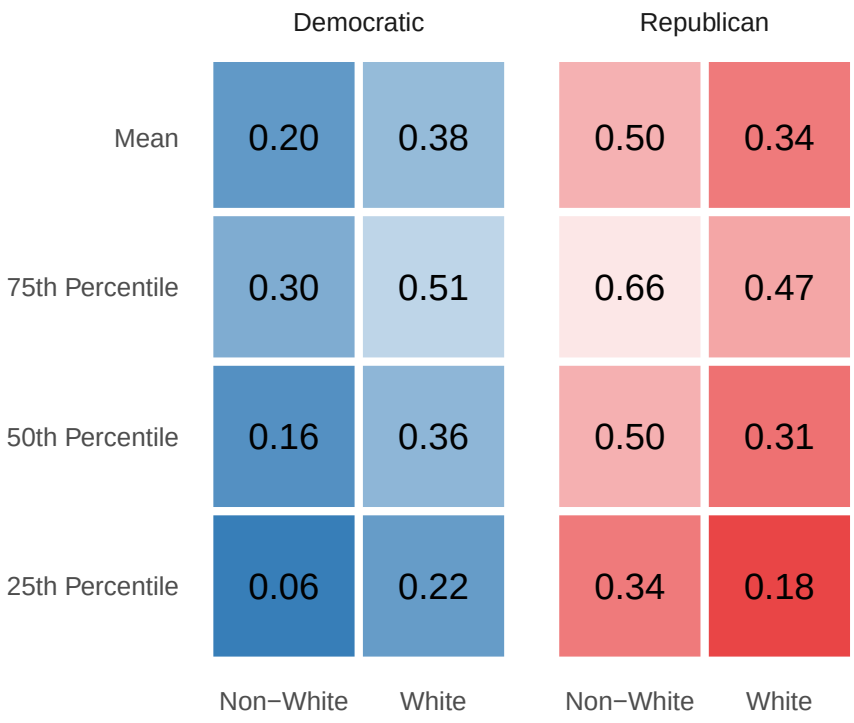

Figure 6: Nationwide Percentiles of Partisan Exposure/Isolation by Party and Race. Colored cells show Spatial Exposure to out-party for white and non-white Democrats (blue) and Republicans (red). Central tendencies and percentiles are from distributions weighted by posterior partisanship probabilities.

Discussion

The isolation of voters by political party is a much talked about feature of contemporary American politics. Uneven distributions of votes threatens equitable representation (14), and isolation from opposing political viewpoints may influence the development of partisan affect, policy preferences, and patterns of political behavior and shape the stragegies of campaigns and elected officials. Despite claims of a starkly segregated America, partisan segregation to date has been imprecisely measured, so the extent of partisan sorting was only understood across very large geographies.

By using geo-located records to develop a spatially-weighted measure of exposure to neighbors of both parties for every voter in the United States, we move beyond conventional measures of segregation by measuring exposure at the individual-level, circumventing common problems of aggregate measurement present in most measures of segregation, and by incorporating the distance between neighbors into our measurement. This yields the first large-scale accounting of individual spatial segregation of any social groups anywhere in the world.

Our results show high partisan segregation across the country, with most voters of both political parties living in partisan bubbles with little exposure to the other party. These high levels of isolation exist across regions and densities. Democratic exposure to Republicans is on average lower than Republican exposure to Democrats, markedly lower in high and medium densities, and higher than is Republican exposure to Democrats in low density areas. Republican exposure is lowest in very low density areas. We also demonstrate that partisan segregation is distinct from racial segregation, and, for white voters, racial sorting likely reduces partisan isolation. Comparing Democrats and Republicans who live in the same city, or even in a neighborhood, we find substantially large differences in partisan environments, evidence that partisan sorting is driven by forces beyond the decision to live

in a specific city or neighborhood.

Given that the high-levels of partisan segregation cannot be explained entirely by urban/rural or racial sorting, what accounts for the partisan segregation, in some cases extreme, which can be found across many types of geographies and even within neighborhoods? This is a topic that deserves further research, but we note that a fruitful line of inquiry may be the influence of micro-level behaviors on these large-scale patterns. While the best available evidence shows that most voters consider the partisan composition of an area to be low on their list of priorities when choosing neighborhoods (49), it is still possible partisan differences in income and lifestyle preferences, such as transportation and type of housing, may drive some voters to select different cities, neighborhoods, and, in some cases, streets or houses within neighborhoods, even if partisanship is not an explicit criteria for selection. As partisanship becomes more correlated with lifestyle differences (50), such sorting may be further exacerbated.

Furthermore, there is mounting evidence for party-based affective attitudes among Americans that are, by some measures, stronger than affect based on race (21). Given that individual attitudes were responsible for some — although certainly not all — of the large-scale racial sorting that occurred across neighborhoods in the United States (35; 1), it is possible that some voters, especially if they have already selected a city or neighborhood in which to live, will make decisions based on the partisanship of their neighbors that will drive some of the clustering we observe. Moreover, if voters will make decisions about where to live based-on the characteristics of their potential neighbors that are correlated with partisanship, such as their race and income, this can also drive the sorting we observe, even at small levels like the neighborhood.

Methods

Voterfile

Our voterfile is obtained from L2, a commercial data vendor working with both major political parties in the United States. The file contains 180,735,64, reflecting the count of registered voters in the United States as of June 2018. We removed 75,443 that were could not be successfully geocoded leaving 180,660,202. L2 and other commercial vendors obtain this data from state governments who collect it for the purposes of administering elections and helping incumbent politicians in their campaigns (33). The vendors then sell this data to political campaigns for voter targeting. Voter data provided by the states usually includes a large number of records that are invalid because the registrant has died or moved and re-registered. L2 and other vendors attempt to remove these records before selling the data. The commercial vendors also attempt to link people across time, preserving parts of their records as they change addresses and/or states delete older data.

Voters' race is available on voter files either because it is recorded at the time of registration in some states or from imputation methods similar to those we implemented for party in this study (see (46)).

Partisanship Imputation Accuracy

We calculate the accuracy of our imputation of partisanship using a Brier score method to measure deviations of the imputed posterior partisan probabilities from the self-reported ideology of the survey respondents. This yields an accuracy 77%. The error rate in this context does not mean that we incorrectly predict 23% of voters' partisan exposure. Since our measures of Spatial Partisan Exposure and Isolation is a weighted average of the probability of Republican and Democratic affiliation across all neighbors, when our probabilistic forecast underestimates the extent to which a voter prefers the Republican to the Democratic party,

or vice versa, we are directly accounting for this uncertainty. See Supporting Information.

Looking at other data from our survey and previous panel surveys give a sense of the expected upper-bound of accuracy in surveying partisan identity, which our 77% accuracy approaches. Our survey included registered voters and the match between registered party and self-reported partisan identity from the survey is 84% (84.31% for Republicans and 83.37% for Democrats). This is similar to what we might expect based on overtime instability of partisan identification in panel surveys: (17) report stability of self-reported partisanship in panel surveys of .80 to .85 over two years, .82 to .84 over one year, and .83 to .90 over less than a year. In most cases with our data, the registration happened many years before the survey, so overtime stability of 84% slightly exceeds what might be expected from the two-year findings of (17). Note to that, assuming that partisanship, as a social identity, will change for only a very small number of people in less than a year, the less than a year stability of .83 to .90 in a period of less than a year gives an estimate of surveys of instability that comes from stochastic fluctuations or survey error and, therefore, the absolute maximum percent match between a survey response and another measure of partisanship.

Nationwide Proportion of Democrats and Republicans

The nationwide proportion of Democrats and Republicans of 51% and 43%, respectively, is calculated by taking the average of posterior partisan probabilities across all US voters.

Density Classifications

We classify high, medium, low, and very low density by the population per square mileage of the Census Tract in which a voter lives, using the classifications developed by David H. Montgomery of CityLab (Table 1).

Classification	Households Per Square Mile	% of Voters
Very low density	< 102	22.92%
Low density	$[102, 800)$	28.90%
Medium density	$[800, 2213)$	29.15%
High density	$>= 2213$	18.97%

Table 1: Density Classifications by Households Per Square Mile

Geographic Definitions and Sources

We use geographic units of state, urban area, county, city/town, ZIP Code, Census Tract, and precincts. Precincts are used to create the imputations of partisanship. All geographic units, except precincts, come from United States Census definitions from 2010 or later. Urban areas are defined by Core-Based Statistical Ares and city/town are defined by Census Places, which includes incorporated places and Census Designated Places that are not incorporated.

Voters' state, county, ZIP Code, and Census Tract were taken directly from the voter file. Urban area is defined by county location extracted from the voterfile and city/town is defined by a spatial merge of voters individual locations' with Census Bureau provided shapefiles. Precincts were collected from the MIT Election Lab (`https://electionlab.mit.edu/`) and the authors' own searches of state election administrations.

The distribution of registered voters in each geography is listed in Table 2 and the percent of voters with non-missing cases for each of these variables is listed in Table S1 in the Supporting Information.

Geography	0.1%	1%	10%	25%	50%	75%	90%	99%	99.9%	Mean
State	237,298	283,212	478,367	976,729	2,798,419	4,334,925	7,643,908	15,710,329	17,948,081	3,542,357
CBSA	6,377	7,637	14,757	22,027	43,011	106,995	338,680	2,885,697	6,720,367	184,977
County	179	628	2,794	6,173	14,734	38,272	120,097	763,377	1,892,687	57,480
City/Town	2	10	68	183	622	2,456	8,595	55,440	295,080	4,542
Zip Code	1	2	94	353	1,460	7,114	16,900	33,084	45,909	5,272
Tracts	4	248	1,082	1,580	2,275	3,149	4,120	6,660	11,578	2,489

Table 2: Quantiles of Registered Voters by Geography

Relative Exposure

Relative Exposure is the difference between the average Spatial Exposure of partisans within a geographic unit to the average Spatial Isolation of out-partisans also living in that geographic unit. This is:

$$\text{Relative Exposure}_{g,p} = \frac{\sum_{i \in \{g\}} \text{P}(p)_i \ \text{Exposure}_{i,p}}{\sum_{i \in \{g\}} \text{P}(p)_i} - \frac{\sum_{i \in \{g\}} \text{P}(q)_i \ \text{Isolation}_{i,q}}{\sum_{i \in \{g\}} \text{P}(q)_i}$$

where Relative Exposure for party p within geographic unit g is the weighted average (weighted by, $P(p)_i$, the posterior probability of voter i being in party p), of party p's exposure to out-party q, across all voters i in geographic unit g, minus the weighted average (weighted by $P(q)_i$) of party q's Isolation, or exposure to itself.

We compare Democrats and Republicans living in the same state, urban area, county, city/town, ZIP Code, and Census Tract. At each level of geography, for both Democrats and Republicans a population-weighted T-test for a mean different than zero yields $p <$ .00000001. See Supporting Information for further details.

Data Availability

All replication data and code are available in the Harvard University Dataverse: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/A40X5L.

References

[1] Massey, D. S. & Denton, N. A. *American Apartheid: Segregation and the Making of the Underclass* (Havard University Press, Cambridge, MA, 1993).

[2] Alesina, A., Baqir, R. & Easterly, W. Public goods and ethnic divisions. *The Quarterly Journal of Economics* **114**, 1243–1284 (1999).

[3] Trounstine, J. Segregation and inequality in public goods. *American Journal of Political Science* (2015).

[4] Enos, R. D. *The Space Between Us: Social Geography and Politics* (Cambridge University Press, New York, 2017).

[5] Allport, G. W. *The Nature of Prejudice* (Addison-Wesley, Cambridge, MA, 1954).

[6] Pettigrew, T. F. & Tropp, L. A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology* **90**, 751–783 (2006).

[7] Centola, D. The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010). URL `https://science.sciencemag.org/content/329/5996/1194`. `https://science.sciencemag.org/content/329/5996/1194.full.pdf`.

[8] Fowler, J. H. & Christakis, N. A. Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences* **107**, 5334–5338 (2010). URL `https://www.pnas.org/content/107/12/5334`. `https://www.pnas.org/content/107/12/5334.full.pdf`.

[9] Rand, D. G., Arbesman, S. & Christakis, N. A. Dynamic social networks promote cooperation in experiments with humans. *Proceedings of the National Academy of*

Sciences **108**, 19193–19198 (2011). URL https://www.pnas.org/content/108/48/19193. https://www.pnas.org/content/108/48/19193.full.pdf.

[10] Bond, R. M. *et al.* A 61-million-person experiment in social influence and political mobilization. *Nature* **489**, 295–298 (2012). URL https://doi.org/10.1038/nature11421.

[11] Becker, J., Porter, E. & Centola, D. The wisdom of partisan crowds. *Proceedings of the National Academy of Sciences* **116**, 10717–10722 (2019). URL https://www.pnas.org/content/116/22/10717. https://www.pnas.org/content/116/22/10717.full.pdf.

[12] Bishop, B. *The Big Sort* (Houghton Mifflin Harcourt, New York, 2009).

[13] Abrams, S. J. & Fiorina, M. P. The big sort that wasn't: A skeptical reexamination. *PS Political Science and Politics* **45**, 203 (2012).

[14] Chen, J. & Rodden, J. Unintentional gerrymandering: Political geography and electoral bias in legislatures. *Quarterly Journal of Political Science* (2013).

[15] Nall, C. The political consequences of spatial policies: How interstate highways facilitated geographic polarization. *The Journal of Politics* **77**, 394–406 (2015).

[16] Martin, G. J. & Webster, S. W. Does residential sorting explain geographic polarization? *Political Science Research and Methods* 1–17 (2018).

[17] Green, D. P., Palmquist, B. & Schickler, E. *Partisan Hearts and Minds* (Yale University Press, New Haven, CT, 2004).

[18] Achen, C. H. & Bartels, L. M. *Democracy for realists: Why elections do not produce responsive government* (Princeton University Press, 2017).

[19] Hersh, E. D. & Goldenberg, M. N. Democratic and republican physicians provide different care on politicized health issues. *Proceedings of the National Academy of Sciences* **113**, 11811–11816 (2016).

[20] Chen, M. K. & Rohla, R. The effect of partisanship and political advertising on close family ties. *Science* **360**, 1020–1024 (2018).

[21] Iyengar, S. & Westwood, S. J. Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science* **59**, 690–707 (2015).

[22] Aarts, H. & Dijksterhuis, A. The silence of the library: environment, situational norm, and social behavior. *Journal of personality and social psychology* **84**, 18 (2003).

[23] Cialdini, R. B., Reno, R. R. & Kallgren, C. A. A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology* **58**, 1015 (1990).

[24] Green, D. P. *et al.* The effects of lawn signs on vote outcomes: Results from four randomized field experiments. *Electoral Studies* **41**, 143–150 (2016).

[25] Bonica, A. Mapping the ideological marketplace. *American Journal of Political Science* **58**, 367–386 (2014).

[26] Huckfeldt, R. & Sprague, J. Networks in context: The social flow of political information. *American Political Science Review* **81**, 1197–1216 (1987).

[27] Zaller, J. R. *The Nature and Origins of Mass Opinion* (Cambridge University Press, New York, 1992).

[28] Converse, P. E. The nature of belief systems in mass publics. In Apter, D. E. (ed.) *Ideology and Discontent* (Free Press, New York, 1964).

[29] Verba, S., Schlozman, K. L. & Brady, H. E. *Voice and Equality: Civic Voluntarism in American Politics* (Harvard University Press, Cambridge, MA, 1995).

[30] Athey, S., Ferguson, B., Gentzkow, M. & Schmidt, T. Experienced segregation. *Working Paper, Stanford University* (2019).

[31] Phillips, N. E., Levy, B. L., Sampson, R. J., Small, M. L. & Wang, R. Q. The social integration of american cities: Network measures of connectedness based on everyday mobility across neighborhoods. *Sociological Methods & Research* 0049124119852386 (2019).

[32] Enos, R. D. & Hersh, E. D. Party activists as campaign advertisers: The ground campaign as a principal-agent problem. *American Political Science Review* **109**, 252–278 (2015).

[33] Hersh, E. *Hacking the Electorate: How Campaigns Perceive Voters* (Cambridge University Press, Cambridge, UK, 2015).

[34] Cramer, K. J. *The Politics of Resentment: Rural Consciousness in Wisconsin and the Rise of Scott Walker* (University of Chicago Press, 2016).

[35] Schelling, T. C. *Micromotives and Macrobehavior* (Norton, New York, 1971).

[36] Logan, T. D. & Parman, J. M. The national rise in residential segregation. *The Journal of Economic History* **77**, 127–170 (2017).

[37] Niemeyer, G. Geohash.org (2008). URL `http://geohash.org/site/tips.html`.

[38] Reardon, S. F. & O'Sullivan, D. Measures of spatial segregation. *Sociological Methodology* **34**, 121–162 (2004).

[39] Massey, D. S. & Denton, N. A. The dimensions of residential segregation. *Social Forces* **67**, 281–315 (1988).

[40] White, M. J. The measurement of spatial segregation. *American journal of sociology* **88**, 1008–1018 (1983).

[41] Rodden, J. A. *Why Cities Lose: The Deep Roots of the Urban-Rural Political Divide* (Basic Books, 2019).

[42] Keith, B. E., Magleby, D. B., Nelson, C. J., Orr, E. A. & Westlye, M. C. *The myth of the independent voter* (Univ of California Press, 1992).

[43] Magleby, D., Nelson, C. & Westlye, M. *The myth of the independent voter revisited*, 238–263 (2011).

[44] Hawkins, C. B. & Nosek, B. A. Motivated independence? implicit party identity predicts political judgments among self-proclaimed independents. *Personality and Social Psychology Bulletin* **38**, 1437–1452 (2012).

[45] Klar, S. & Krupnikov, Y. *Independent politics* (Cambridge University Press, 2016).

[46] Enos, R. D. What the demolition of public housing teaches us about the impact of racial threat on political behavior. *American Journal of Political Science* **60**, 123–142 (2016).

[47] Imai, K. & Khanna, K. Improving ecological inference by predicting individual ethnicity from voter registration records. *Political Analysis* **24**, 263–272 (2016).

[48] Farley, R., Steeh, C., Krysan, M., Jackson, T. & Reeves, K. Stereotypes and segregation: Neighborhoods in the detroit area. *The American Journal of Sociology* **100**, 750–780 (1994).

[49] Mummolo, J. & Nall, C. Why partisans do not sort: The constraints on political segregation. *The Journal of Politics* **79**, 45–59 (2017). URL `https://doi.org/10.1086/687569`. `https://doi.org/10.1086/687569`.

[50] Hetherington, M. & Weiler, J. *Prius Or Pickup?: How the Answers to Four Simple Questions Explain America's Great Divide* (Houghton Mifflin, 2018).

Acknowledgements

The authors thank Ben Lewis and Devika Kakkar at the Harvard Center for Geographic Analysis and Angelo Dagonel for research assistance. We thanks seminar participants at the University of Pittsburgh, New York University, Northeastern University, Princeton University, Brown University, University of Massachusetts – Amherst, and Harvard University.

Author Contributions

JB and RE both contributed to the conception, design, analysis, data-collection, and writing.

Competing Interests

The authors have no financial or non-financial competing interests.