



# Partisan spatial sorting in the United States: A theoretical and empirical overview <sup>☆,☆☆</sup>



Ethan Kaplan <sup>a,\*</sup>, Jörg L. Spenkuch <sup>b,\*</sup>, Rebecca Sullivan <sup>c</sup>

<sup>a</sup> University of Maryland, United States

<sup>b</sup> Northwestern University, United States

<sup>c</sup> Federal Housing Finance Agency, United States

## ARTICLE INFO

### Article history:

Received 29 December 2020

Revised 31 March 2022

Accepted 14 April 2022

Available online 25 May 2022

### Keywords:

Sorting  
Polarization  
Voting  
Elections  
Ideology  
Partisanship

## ABSTRACT

We develop a variance-like index of heterogeneity in partisanship and use it to measure spatial sorting. We prove that our index is the only one (up to a linear transformation) that satisfies seven theoretical properties, all of which are intuitively desirable. Based on this index we document the long-run evolution of geographic sorting along partisan lines in the American electorate. We provide evidence that spatial cleavages have increased dramatically since the mid-twentieth century. At no point since the Civil War have partisans been as clustered *within* the boundaries of individual states as today. Nonetheless, even when geographic sorting is measured at the precinct level, differences across communities tend to be significantly smaller than differences within. In this sense, the American electorate continues to be more diverse within than across areas.

© 2022 Elsevier B.V. All rights reserved.

## 1. Introduction

According to popular perception, ordinary Americans are not only divided in their allegiance to one of the two major parties, but partisan divisions also manifest themselves across space. Republican supporters live in “red,” rural states, while Democrats reside in “blue,” urban areas along the coasts. Some even argue that partisans have become so clustered in like-minded communities that the resulting spatial fissures are tearing American society apart (see, e.g., Bishop, 2008). But how sorted are Americans really? And has the degree of sorting changed much over time?

In order to measure sorting and thereby answer these questions, we introduce a variance-like index of heterogeneity in ideology or

partisanship. We show that our index is the only one that satisfies a set of seven intuitively desirable criteria. Chiefly among them, the variance index allows us to exactly decompose overall heterogeneity in partisanship into differences across and within communities. As a result, we can gauge the degree of sorting along ideological lines by comparing partisan heterogeneity across areas to the heterogeneity within the respective communities.

Although we are primarily interested in partisan sorting, the usefulness of our index is not limited to this particular context. Other applications might include (1) measuring the degree of sorting by race across and within schools and further into classrooms, (2) comparing the degree of educational sorting across and within firms, or (3) measuring the degree of sorting by comparing heterogeneity in income at various levels of aggregation.

After analytically developing our index, we use it to measure partisan geographic sorting dating back to 1856—the first presidential election with both a Democratic and a Republican candidate. Such a long horizon is useful for putting recent trends into perspective. We compare our measure of sorting to other common measures, most of which display overall similar trends. Our measure, however, has the advantage of being decomposable into constituent parts. This is important for comparing the degree of sorting across different levels of aggregation, such as states, counties, and precincts.

Our paper contributes both to the theory and to the empirical measurement of sorting. Although our analysis speaks to divisions

\* We thank Jorge Perilla-Garcia and Daniel Kolliner for excellent research assistance. We have also benefited from helpful comments by Morris Fiorina, Jim Gimpel, Marc Hetherington, Greg Martin, Pablo Montagnes, Cody Tuttle, Haishan Yuan, and audience members at the 2018 APSA Annual Meeting. Any mistakes, however, are solely our own.

\*\* Rebecca Sullivan contributed to this research in her personal capacity and not as part of her official duties at the Federal Housing Finance Agency. The analysis and conclusions are those of the authors alone and should not be represented or interpreted as conveying an official Federal Housing Finance Agency position, policy, analysis, opinion, or endorsement. Any errors or omissions are the sole responsibility of the authors.

\* Corresponding authors.

E-mail addresses: [edkaplan@umd.edu](mailto:edkaplan@umd.edu) (E. Kaplan), [j-spenkuch@kellogg.northwestern.edu](mailto:j-spenkuch@kellogg.northwestern.edu) (J.L. Spenkuch), [rebecca.sullivan@fhfa.gov](mailto:rebecca.sullivan@fhfa.gov) (R. Sullivan).

within the American electorate by documenting trends in partisan sorting over long periods of time, we do not directly contribute to the debate on the causes and consequences of sorting (see, e.g., Bishop, 2008; Gimpel and Hui, 2015; McDonald, 2011; Mummolo and Nall, 2017; Martin and Webster, 2020). We merely assess the claim that partisans are increasingly clustered across space. To date, the literature has focused on potential downstream effects of sorting. Much less work has been done to directly assess whether partisans are, in fact, more geographically sorted than in the past. Moreover, the little evidence that does exist is, for the most part, based on state-level differences, without a principled way of drawing comparisons over time (e.g., Glaeser and Ward, 2006; Hopkins, 2017). Glaeser and Ward's (2006) assertion that partisan segregation is one of the big myths of American electoral geography is, therefore, speculative.

Two recent studies based on detailed voter-registration records do present evidence of partisan clustering. Brown and Enos (2018) use a snapshot from 2017 to show that, today, Democrats and Republicans are nearly as segregated as racial minorities. Sussell (2013) relies on data from California spanning the period from 1992 to 2010. His results suggest that partisan segregation increased noticeably during this time.<sup>1</sup>

Our analysis unearths a rich set of previously unknown facts. Specifically, we find that, *within* states, partisan sorting has increased approximately five fold between its nadir in 1976 and the most recent presidential election in 2016. Surprisingly, since the 1970s, our measure of within-state geographic sorting is nearly perfectly correlated with Poole and Rosenthal's (Poole and Rosenthal, 1997) well-known index of polarization in the U.S. House ( $\rho = .95$ ). Regardless of whether we measure within-state sorting at the county or precinct level, the data reveal a dramatic increase in spatial differences—especially over the last five election cycles. Geographic sorting within states is currently at a historic high.<sup>2</sup> Although we do find a rise in sorting across states, the red-blue state divide is significantly lower than it was in the period surrounding the Civil War or even in the mid-1890s through the mid-1920s. Moreover, the rise in state-level partisan sorting is not nearly as sharp as the increase in sorting across counties with the same state.

Finally, though we document a dramatic rise in spatial sorting over the last few decades, all of our results imply that differences between individuals within counties or precincts are, on average, many times greater than differences across space. At the same time, we emphasize that it is difficult to say how much geographic sorting is “too much.” Current levels are high by historical standards, and there simply does not exist enough evidence on the causal effects of geographic cleavages on democratic outcomes to speculate about potential consequences.

By developing a theoretically grounded measure of geographic sorting and by documenting the recent increase therein, our anal-

<sup>1</sup> Specifically, Sussell (2013) computes isolation and segregation indices using partisan registration rates as well as presidential election returns. Eleven of his twelve measures of partisan segregation increased during this period, with rates of growth ranging from 2.1% to 23.1%. (Brown and Enos (2018) compute isolation and exposure indices but only at one point in time.

<sup>2</sup> In independent, simultaneous work, Darmofal and Strickler (2019) also present time-series evidence on long-run sorting. An important difference between their approach and ours is that they rely on Bishop's (2008) concept of “landslide counties” to measure geographic sorting. As previously pointed out by Abrams and Fiorina (2012) and Klinkner (2004a, 2004b), using landslide counties to assess changes in sorting is theoretically problematic because the results can be highly dependent on arbitrary cutoffs. As a consequence, some of Darmofal and Strickler's substantive conclusions differ greatly from ours. While they find that “the percentage of the voting public living in heavily or landslide partisan counties in the twenty-first century is well within a normal historical range” (p. 83), we show that, when properly measured, current levels of voter partisan sorting are very high by historical standards.

ysis paves the way for research on a number of important questions related to political sorting. For instance, are spatial divisions in the electoral landscape a cause or a consequence of elite polarization? Does the clustering of like-minded partisans lead to better or worse representation? Does it cause legislative dysfunction? Does partisan sorting create ideological echo chambers—as asserted by Bishop (2008)—or is it irrelevant for the evolution of voters' views and preferences? On theoretical grounds the answers to these questions are inherently ambiguous. What our analysis establishes is that, today, partisans are more geographically clustered than at any time in recent memory.

## 2. Measuring geographic sorting: theory

Before discussing our findings on historical patterns of geographical sorting by partisanship, we first ask how sorting on ideology should be measured in the first place. The literature has heretofore been eclectic in its measurement of political sorting. Bishop (2008), for instance, calculates the share of voters living in “landslide counties,” i.e., counties in which one party achieved a victory margin of at least 20%. Abrams and Fiorina (2012) criticize this measure for being arbitrary and vague, and Klinkner (2004a, 2004b) shows that small definitional changes lead to as much as a 25% reduction in the number of voters in such counties.

In what follows, we propose seven properties that any good measure of ideological heterogeneity ought to possess. The first six of them are self-evidently desirable, while the last property is tailored towards comparing heterogeneity across and within regions. Being able to compare heterogeneity across and within geographic areas is important because if partisans sort across space, then we would expect much of the extant heterogeneity to be captured by differences across rather than within communities.

We prove that there exists one and only one (up to a constant positive multiple) index that satisfies all of our theoretical desiderata. To be clear, there is a vast literature that axiomatically derives different indices.<sup>3</sup> Our contribution is to recognize that the mathematical structure of quantifying the degree of heterogeneity in ideology or partisanship is very similar to that of measuring inequality. We can, therefore, build upon prior work, particularly Bosmans and Cowell (2010), and bring some of its insights to bear on our question. In our discussion below, we motivate our axioms with reference to the measurement of partisan sorting; however, as we have noted, there are many other potential applications.<sup>4</sup>

*Mathematical Preliminaries.*— We first assume that the researcher observes a valid proxy for voters' ideology or partisanship.<sup>5</sup> Formally, let there be  $n$  individuals, whose preferences are characterized by  $\mathbf{x} = (x_1, \dots, x_n)$ . We use  $\bar{x} = (1/n)\sum_{i=1}^n x_i$  to denote the mean of  $\mathbf{x}$ , while  $\bar{\mathbf{x}}$  is an  $n \times 1$  vector with  $\bar{x}$  in every position.

**Definition.** An index of ideological heterogeneity is a function  $P$  that assigns a real number to any vector of preferences  $\mathbf{x}$ , i.e.,  $P: \mathbb{R}^n \rightarrow \mathbb{R}$ .

*Desirable Properties.*— Any measure of ideological heterogeneity ought to have a well-defined and easily interpretable baseline. Our first axiom, therefore, states that measured heterogeneity should be equal to zero when all voters have identical preferences.

<sup>3</sup> See, e.g., Esteban and Ray (1994) for a well-known index of polarization.

<sup>4</sup> Our index is particularly useful when attempting to quantify nested sorting. Some examples besides the one presented in this paper would include sorting by race into school districts, into schools within districts and into classrooms within schools; or sorting of workers by education level across industries, across firms within industries and across plants within firms.

<sup>5</sup> In our empirical application, we use electoral returns to proxy for the partisanship of voters.

**Axiom 1** (Normalization).  $P(\mathbf{x}) = 0$  whenever  $x_i = x_j$  for all  $i, j$ .

In addition, an index of heterogeneity in ideology or partisanship should not change if voters become *uniformly* more liberal or conservative. As commonly understood, heterogeneity refers to differences in preferences rather than the extremity of their mean. Hence, **Axiom 2** requires that uniform changes in voters' preferences have no effect on  $P$ .

**Axiom 2** (Translational Invariance).  $P(\mathbf{x} + \mathbf{c}) = P(\mathbf{x})$  for any  $\mathbf{c} = (c, \dots, c) \in \mathbb{R}^n$ .

Since we are concerned with voters rather than political elites, we also think it desirable that all individuals receive equal weight. That is, conditional on the *distribution* of preferences, measured heterogeneity should not depend on who holds which views (**Axiom 3**).

**Axiom 3** (Anonymity).  $P(\mathbf{y}) = P(\mathbf{x})$  whenever  $\mathbf{y}$  is simply a permutation of  $\mathbf{x}$ .

Nor should it matter how many individuals there are (**Axiom 4**). In particular, an exact doubling of the population maintaining the distribution of preferences should not impact the index.

**Axiom 4** (Population Independence).  $P(\mathbf{x}, \mathbf{x}) = P(\mathbf{x})$ .

Independence of population size is important for directly comparing differently-sized groups of voters. By imposing **Axiom 4**, we ensure that our conclusions about the evolution of partisan sorting across space and time are solely due to changes in the distribution of voters' preferences rather than differences in population size.

Our next axiom requires that small changes in preferences lead only to small changes in measured heterogeneity.

**Axiom 5** (Continuity).  $P$  is continuous in every element of  $\mathbf{x}$ .

Continuity fails for all indices that rely on cutoff values to classify states, counties, or any other group of voters. Threshold-based indices are problematic because substantively minor differences between voters across space or time may give the (false) impression of large differences or changes. [Ansolabehere et al. \(2006\)](#), for instance, argue that categorizing states as either "red" or "blue" obscures the fact that most of America is actually "purple." [Klinkner \(2004a,b\)](#) makes a similar point when he criticizes Bishop's ([Bishop, 2008](#)) measure of "landslide counties." He even demonstrates that small changes to the cutoff used to define "landslides" have a big effect on the results. By contrast, a continuous measure of voter heterogeneity is immune to such problems.

An important additional requirement is that as voters' preferences diverge, measured heterogeneity increases.

**Axiom 6** (Spread Responsiveness). If  $\mathbf{x} = (x_1, x_2)$  with  $x_1 \leq x_2$  and  $\mathbf{x}' = (x_1 - c, x_2 + c)$  for some  $c > 0$ , then  $P(\mathbf{x}') > P(\mathbf{x})$ .

In words, **Axiom 6** deals with the minimal case of an electorate of only two individuals. If the ideological distance between the two increases (without changing the mean), then measured heterogeneity must go up. As an example, if there are no political differences among people, then our index should be zero; as differences emerge, our measure of heterogeneity should rise. Any index that does *not* satisfy this property is an inherently flawed measure of heterogeneity across voters.<sup>6</sup>

<sup>6</sup> We define **Axiom 6** in terms of two voters so that it is straightforward to say whether heterogeneity should be increasing or decreasing. With three or more individuals, it is possible for an increased spread between one pair of individuals to coincide with a decline between other pairs, in which case it is *a priori unclear whether heterogeneity should go up or down*.

In our view, **Axioms 1–6** are not controversial. They are desirable for any measure of voter heterogeneity. Our last axiom is the least trivial one. Yet, it is crucial for assessing the importance of geographic divisions.

As illustrated in [Fig. 1](#), even absent any macro-level differences in the overall composition of the electorate, voters today might be living in more homogeneous communities than just a few decades ago. That is, they might be better *sorted*. Conversely, the American electorate as a whole might have become more polarized without any widening of spatial cleavages. Hence, assessing claims of spatial sorting involves a comparison of differences across and within communities. Put differently, we need to be able to disentangle communities becoming more or less alike from changes in how internally differentiated the respective groups of voters are.

In addition, absent a commonly accepted definition of "community," we need to be able to consistently do so at different levels of spatial aggregation. Suppose, for instance, that, according to  $P$ , voters within every single electoral precinct in some state have become more extreme over time, without any narrowing in the differences across precincts. Then if we use  $P$  to assess heterogeneity in the state as a whole, it should also indicate rising heterogeneity at the state level. **Axiom 7** ensures that this is the case.

**Axiom 7** (Decomposability). There exists a nonnegative weighting function  $\omega$  such that (i)  $P(\mathbf{x}, \mathbf{y}) = \omega(\bar{x}, \bar{y}, n_x, n_y)P(\mathbf{x}) + \omega(\bar{y}, \bar{x}, n_y, n_x)P(\mathbf{y}) + P(\bar{x}, \bar{y})$  for all  $\mathbf{x} \in \mathbb{R}^{n_x}$  and  $\mathbf{y} \in \mathbb{R}^{n_y}$ ; and (ii)  $\omega(\bar{x}, \bar{y}, n_x, n_y) + \omega(\bar{y}, \bar{x}, n_y, n_x) = 1$ .

Intuitively, the axiom stipulates that a useful measure of heterogeneity ought to be decomposable into across- and within-group components.<sup>7</sup> In our application, the former measures geographic sorting, i.e., the extent of mean differences across, say, states, counties, towns, or neighborhoods, etc. The latter is a weighted average of the heterogeneity within each group of voters.

**Axiom 7** guarantees that these decompositions are exact and, when carried out at different levels of aggregation, mutually consistent. After all, it should be irrelevant whether we first decompose national differences to the state level and then to the county and precinct level, or if we directly work with the latter. In the remainder of this paper, we rely heavily on decompositions at various levels of aggregation in order to assess whether Democratic and Republican supporters are more geographically clustered today than in decades past. It is, therefore, important for  $P$  to ensure that our findings for different levels of aggregation are mutually consistent. Moreover, decomposability allows us to determine the relative importance of changes in sorting at different levels of aggregation, such as states, counties, or precincts.

As a technical matter, we restrict  $\omega$  to be an arbitrary function of mean preferences as well as groups' sizes. We further require that all weights be non-negative and sum up to one. This last condition ensures that, if there are no mean differences across communities, then society as a whole shall not be deemed more (less) polarized than its most (least) polarized subgroup.

*A Unique Index.*— We view each of the properties in **Axioms 1–7** as desirable for an index that is being used to document geographic divisions over time. Given these axioms, we can formally prove that there exists a uniquely good measure.

<sup>7</sup> We also note that we could alternatively look at voter heterogeneity across and between non-spatially defined groups. For example, we could use our index to look at heterogeneity within and across income or educational groups at different levels of aggregation.

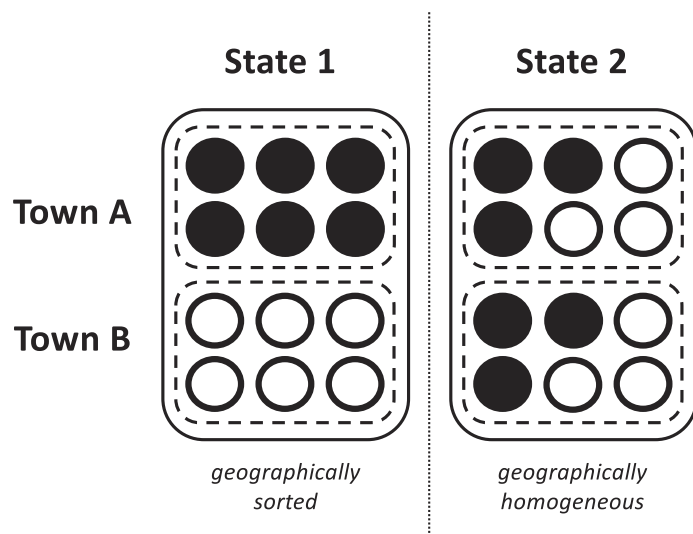


Fig. 1. Heterogeneity Across vs. Within Communities.

**Proposition 1.** An index satisfies Axioms 1–7 if and only if it is a positive scalar multiple of  $P(\mathbf{x}) = (1/n) \sum_{i=1}^n (x_i - \bar{x})^2$ . Since  $P$  corresponds to the population variance, we refer to this index as the variance index.

In words, the proposition establishes that the variance index is the only measure of voter heterogeneity that has all of the desired properties. Any other index violates at least one of our desiderata.<sup>8</sup>

As a corollary to Proposition 1, the weights needed to disaggregate the variance index across different groups of voters are simply the groups’ population shares.

**Corollary.** Suppose that  $P(\mathbf{x})$  satisfies Axioms 1–7, then  $\omega(\bar{x}, \bar{y}, n_x, n_y) = \frac{n_x}{n_x + n_y}$ .

While Proposition 1 holds given any unidimensional representation of individuals’ preferences or actions, it is silent on how to best gauge ideology or partisanship. As a result, comparisons between different groups of voters may well depend on the underlying measure of preferences. We, therefore, advocate that the variance index be used with the understanding that any conclusion is inextricably tied to the representation of preference on which it is based. That is, the variance index measures sorting in whatever facet of voters’ preferences or actions is captured by  $x$ .

### 3. Methods and data

#### 3.1. Mapping theory into data

Since we are interested in assessing the extent of spatial cleavages over long periods of time, most of our empirical application focuses on geographical sorting in partisanship as captured by election results. Survey data would provide an alternative measure of partisan sentiment from which we could potentially compute partisan sorting. Unfortunately, survey data are scant before 1930 and rarely allow for systemic valid inferences below the state level even today. If we want to contrast geographic sorting today with the divisions that existed during the New Deal or during

<sup>8</sup> See Massy and Denton (1988) for a useful discussion of the properties of different measures of segregation, many of which may seem *prima facie* useful for measuring geographical sorting.

Reconstruction, we are forced to rely on electoral returns as a proxy for voters’ partisan preferences.<sup>9</sup>

To operationalize the theoretical insights in the previous section, consider a presidential election in which the Democratic candidate received  $n^D$  votes, while the Republican one garnered  $n^R$ . We represent the choices of voters in this election by letting  $x$  be a vector with  $n^D$  ones and  $n^R$  zeros. Denoting the Democratic two-party vote share as  $v$ , the variance index simplifies to

$$P(x) = \frac{1}{n} [n^D(1 - v)^2 + n^R v^2] = v(1 - v), \tag{1}$$

where  $n = n^D + n^R$ . Based on this expression,  $P$  is minimized when all voters back the same candidate, i.e.,  $v = 1$  or  $v = 0$  and  $P(x) = 0$ ; it is maximized when the electorate is equally split, i.e., when  $v = .5$  so that  $P(x) = .25$ .<sup>10</sup> When interpreting the magnitude of the variance index and its components, it is often useful to do so with these numbers in mind.

In this context, it is important to distinguish ideological extremity from spatial sorting along partisan lines. An area with a very high or very low Democratic vote share is likely an ideologically extreme one. For instance, in 2016 three counties had a Democratic two-party vote share below 5%: King County (TX), Roberts County (TX) and Garfield County (MT); while another sixty counties saw Democratic vote shares below 10%. On the Democratic side, in addition to the eight wards of Washington D.C., four counties had two-party Democratic vote shares in excess of 90%: Prince George’s County (MD), Oglala Lakota County (SD), Bronx County (NY), and San Francisco County (CA). The aforementioned places are likely some of the most partisan within the United States. They contribute substantially to partisan sorting across counties; internally, however, they are very homogeneous.

By contrast, the Democratic and Republican two-party vote shares fell within 0.2 percentage points of parity in eight counties: Clark County (WA), Lorain County (OH), Winnebago County (IL),

<sup>9</sup> We note that, in general elections in the U.S., there is rarely reason to cast strategic ballots. It is, therefore, reasonable to assume that votes proxy for partisan preferences.

<sup>10</sup> We ignore votes for third-party candidates. The number of votes for these candidates is small in most, though not all, elections that we study. One advantage of our approach is that it easily accommodates third party candidates as long as we have a good measure of the “partisan distance” between the independent candidate and each of the other parties or candidates in the election.



Kent County (RI), Panola County (MS), Kendall County (IL), Nash County (NC), and Teton County (ID). These counties contributed the least to our measure of partisan sorting precisely because they had the greatest extent of internal heterogeneity.

As matter of notation, when we measure differences across states, we let  $\bar{x}_s$  be an  $|n_s| \times 1$  vector with the Democratic vote share in state  $s$  while, for each state,  $x_s$  is an  $n_s \times 1$  vector with  $n_s^D$  ones and  $n_s^R$  zeros, one for each individual within the states. We then calculate the extent of sorting across states as

$$\begin{aligned}
 \underbrace{P(\bar{x}_1, \dots, \bar{x}_S)}_{\text{heterogeneity across states}} &= \underbrace{P(x)}_{\text{overall heterogeneity}} - \underbrace{\sum_{s=1}^S \frac{n_s}{n} P(x_s)}_{\text{heterogeneity within states}} \\
 &= \sum_{s=1}^S \frac{n_s}{n} (v_s - v)^2. \tag{2}
 \end{aligned}$$

In some of what follows, we report cross-state sorting relative to the overall level of partisan heterogeneity, i.e.,  $P(\bar{x}_1, \dots, \bar{x}_S)/P(x)$ . We refer to this ratio as the across-state share and interpret it as the fraction of heterogeneity in partisanship that is attributable to systematic differences across states.

To assess the relative importance of geographic cleavages at different levels of aggregation, we repeatedly apply our decomposition to geographic units that are nested. For instance, since counties are nested within states, we can further decompose Eq. (2) into

$$\begin{aligned}
 \underbrace{P(x)}_{\text{overall heterogeneity}} &= \underbrace{P(\bar{x}_1, \dots, \bar{x}_S)}_{\text{heterogeneity across states}} + \underbrace{\sum_{s=1}^S \frac{n_s}{n} P(\bar{x}_{1,s}, \dots, \bar{x}_{C_s,s})}_{\text{within-state heterogeneity across}} \\
 &+ \underbrace{\sum_{s=1}^S \sum_{c=1}^{C_s} \frac{n_{c,s}}{n} P(x_{c,s})}_{\text{heterogeneity within counties}}, \tag{3}
 \end{aligned}$$

where  $C_s$  denotes the number of counties in state  $s$ , and  $x_{c,s}$  represents the preference profile of voters in county  $c$  in the same state. Intuitively, the first term on the right-hand side in Eq. (3) measures the importance of differences in voters' mean preferences across states. The second term tells us how geographically divided voters are, on average, across counties within the same state. The last term measures the degree of partisan heterogeneity within individual counties. Thus, our decomposition can be thought of as disentangling differences between individuals within the same county from differences in the average across counties within the same state, as well as mean differences across states. Below, we demonstrate that the importance of these components varies considerably over the long arc of American history.

For the most recent period, we also assess partisan sorting at the precinct level. Precinct-level data allow us to document trends in geographic differences at a much finer scale, but only for a shorter time frame and subject to the caveat that precinct boundaries are not temporally stable. Calculating precinct- rather than county-level heterogeneity in partisanship requires nothing more than an appropriate change of indices in the equation above.<sup>11</sup>

### 3.2. Data sources

We obtained county-level presidential election returns for the years 1972 through 2016 from the *CQ Voting and Elections Collection* and the remainder from *Inter-university Consortium for Political and Social Research* (1999). Our county-level time series

<sup>11</sup> Note, our main results would remain qualitatively unchanged if we scaled votes in general elections by the respective candidates' idealpoints. This is because for any two-candidate election, scaling votes corresponds to a linear transformation of  $x$ , which simply yields a scalar multiple of the variance index. In races with three or more candidates, this equivalence need not hold. Candidates' relative positions may affect both levels and shares of geographic heterogeneity among voters.

starts in 1856, the first year in which both Democratic and Republican candidates competed in a presidential election. Precinct-level electoral returns come primarily from the Harvard Election Data Archive. We collect electoral returns both for presidential elections as well as elections for the House of Representatives. The precinct-level presidential election data is available from 2000 to 2016 whereas the precinct-level data on house elections ends in 2012. Unfortunately, coverage of the Harvard Election Data Archive varies significantly over time. Thus, whenever necessary, we supplement the precinct-level data with information from *David Leip's Atlas of U.S. Elections* and with information that we collected directly from different Secretaries of State. The latter are additionally used to correct a number of anomalies in the raw data (see [Appendix B](#) for details).

## 4. Partisan sorting over time: evidence

### 4.1. National time series

We now present our first decomposition of the variance index. We begin by calculating the degree of partisan sorting across states because it is the highest interesting level of spatial aggregation and because "red states" and "blue states" have received substantial attention in both the academic and popular discourse.

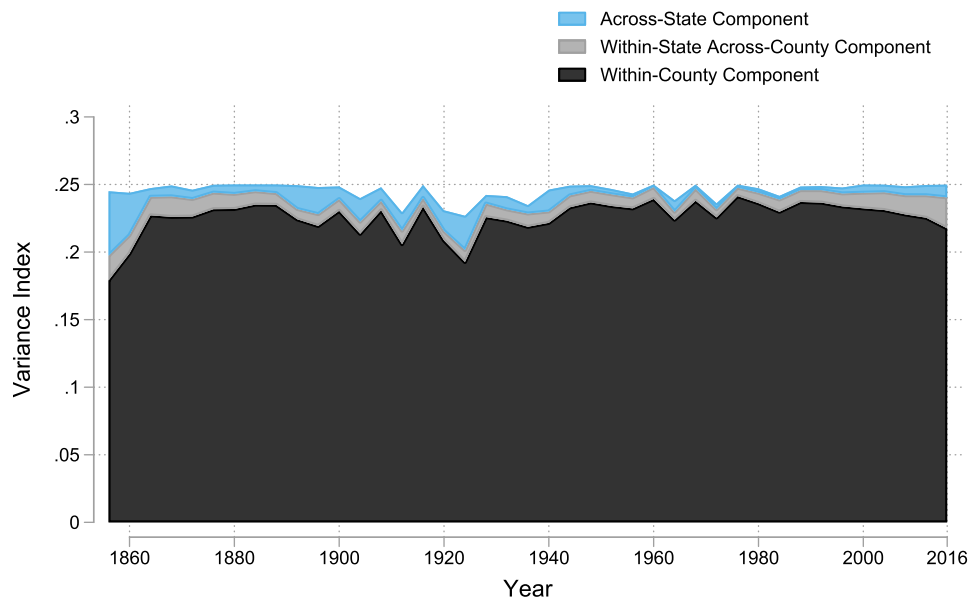
Relying on the expression in Eq. (3), [Fig. 2](#) computes the total variance in two-party votes for president and its decomposition into (1) an across-state component (shaded in blue), (2) a within-state across-county component (shaded in gray), and (3) a within-county component (shaded in black). We present this decomposition for every presidential election from 1856 through 2016.<sup>12</sup> The blue area then corresponds to the first term on the right side of Eq. (3), the gray area to the second term, and the black area to the third term.

[Fig. 3](#) complements the previous figure by presenting the across-state and within-state across-county components as a fraction of the total variance index for the respective election. Doing so highlights the time trends in both measures of sorting. It also makes it more straightforward to interpret their magnitudes and thus the change of relative partisan cleavages across space over time.

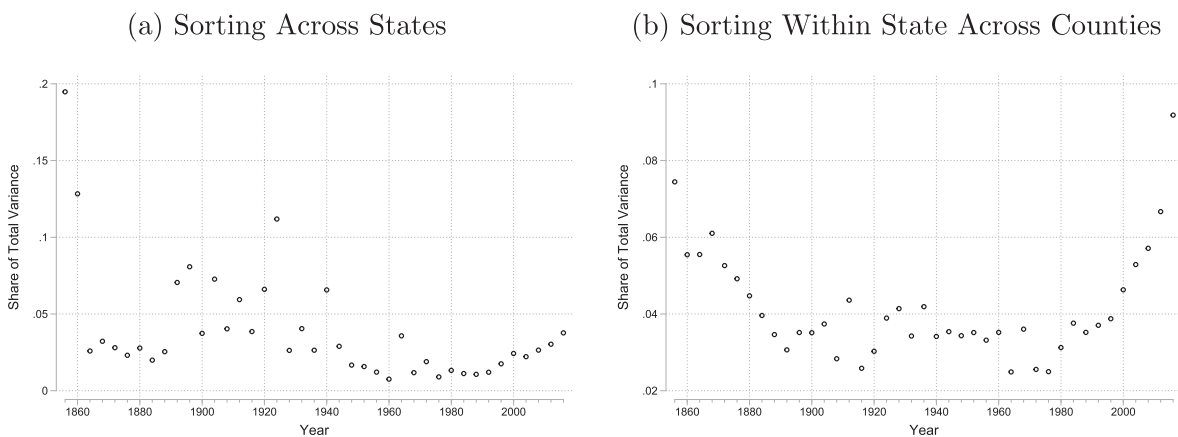
When expressed as a fraction of the total variance index, the across-state and within-state across-county components can be interpreted akin to the (partial)  $R^2$  in a standard linear regression model. The across-state component, for instance, corresponds to  $R^2$  in an individual-level regression of a partisan indicator on state fixed effects (and no constant). In other words, the across-state component in [Fig. 3](#) corresponds to the share of the national variation in partisanship that can be explained by state of residence. The within-state across-county component tells us what fraction of the overall variation in partisanship can be explained by voters from the same states choosing different counties of residence.<sup>13</sup>

<sup>12</sup> Note, four parties received electoral votes in the 1860 election: the Republican Party, the Constitutional Union Party, the Northern Democratic Party and the Southern Democratic Party. The Constitutional Union Party and the Northern Democratic Party received a total of 51 electoral votes, while the Southern Democratic Party garnered 72 votes. The Republican Party won 180 votes in the Electoral College. Our results for 1860 are based on votes cast for the Northern Democratic Party.

<sup>13</sup> For a different way to think about the size of the numbers in [Fig. 3](#), suppose that 15% of the total variance in partisanship is across counties. Now, imagine that people re-sorted so that the fraction of Democrats and Republicans was the same across all counties but the average within-county variance in partisanship remained the same. In such a case, the overall variance in partisanship would decrease by 15%. Alternatively, if we could spatially re-arrange voters so that every county was completely homogeneous but the variance across counties remained the same, then the overall variance in partisanship would decrease by 85%.



**Fig. 2.** Decomposition of the Variance Index, Presidential Elections 1856–2016. *Notes:* Figure shows a decomposition of the variance index for each presidential election from 1856 to 2016. As explained in the text, the decomposition is based on the expression in Eq. (3).



**Fig. 3.** Measuring Sorting Across States and Counties, Presidential Elections 1856–2016. *Notes:* Figure measures partisan sorting across states (panel A) and across counties within the same states (panel B) for each presidential election from 1856 to 2016. All numbers correspond to the respective terms in Eq. (3) as a fraction of the overall variance index for the respective election.

Taken together, Figs. 2 and 3 document five interesting facts. First, for most elections the level of overall variance in partisanship is close to its mechanical maximum at .25.<sup>14</sup> Second, although partisan differences across states have been increasing in recent decades, across-state sorting was higher in the U.S. at many previous points in time—regardless of whether we measure the across-state component of the variance index in levels or as a fraction of the overall heterogeneity at the respective point in time. Third, spatial differences across counties were highest and partisan heterogeneity within counties lowest around the time of the Civil War and in recent years. Fourth, across-county sorting has been steadily increasing over the past four decades. Fifth, partisan heterogeneity within counties is many times higher than differences across counties.

Of course, the choice to measure sorting based on the two-party vote share is somewhat consequential for our results on across-state sorting. The years in which the across-state component of the variance index is the highest are the ones where independents

performed well at the national level. Parties other than the Democratic and Republican parties managed to garner more than 10% of the vote in six elections within our sample: 1912 (27% Bull Moose, 6% Socialist), 1860 (18% Constitutional Democratic, 12% Constitutional Union), 1856 (22% American “Know Nothing”), 1992 (18% Reform), 1924 (17% Progressive), and 1968 (14% American Independent). Ignoring these years would remove some of the years with the most across-state sorting. Nonetheless, there are still many years between 1892 and 1940 with larger mean differences across states than in the recent period. As a result, though contemporaneous partisan sorting across states has risen in recent years, it is not historically high.<sup>15</sup>

<sup>15</sup> We note that our variance index allows us to compute across-state sorting across multiple parties as well as with the two-party vote share. The problem is determining positions in a partisan or ideological spectrum of third party candidates. In a previous draft, we showed how to do this when looking at across-county sorting on ideology using primary elections. However, we do not know of a reasonable way to measure ideological or partisan differences between independent candidates going back further in time.

<sup>14</sup> This is because the popular vote share of the two main parties tends to be close in most presidential elections.

Interestingly, partisan sorting across states remained largely the same directly after the disenfranchisement of African-Americans following the withdrawal of the Northern Army from the South (1877), as well as after their re-enfranchisement due to the passage of the Voting Rights Act (1965). The lack of visible impact of the latter may, in part, be because the loyalties of both African-American and white Southerners lay with the Democratic party at the time. Hence, the expansion of the franchise did little to change the spatial distribution of partisan allegiances.

A particularly stark picture emerges when we turn from sorting *across states* to geographic sorting across counties *within* the same states. Besides assessing the importance of geographic divisions at a lower level of aggregation, a benefit of measuring sorting within rather than across states is that we hold fixed the competitiveness of the race as well as other electoral circumstances that might affect voters.

The combined across-state and within-state across-county evidence we present shows that while the red-state/blue-state divide has been increasing, this aspect of spatial sorting which has been covered so heavily in the press is actually neither the largest component of the increase in sorting nor is its current levels close to historical heights. Relative to 2016, sorting across states was far higher leading up to the Civil War and substantially higher for most of the time from 1892 to 1924. Outside of these two periods, the 2016 presidential election was the third-most polarized across states—right after 1932 and 1940. In contrast, political sorting across counties within states has seen a more steep increase and is at historic heights.

Within individual states, the most geographically homogeneous presidential elections were in the 1960s and early- to mid-1970s. It was precisely during this time period, following the passage of the Civil Rights Act, that Southern Democrats started to realign with the Republican Party. The realignment of the South temporarily reduced differences across space relative to the widening divisions within the electorate of Southern counties. The 1964 and the 1972 presidential elections were also two of the six least competitive presidential races—right after the election of Harding in 1920, Wilson in 1912, Coolidge in 1924, and FDR in 1936.

The highest degree of within-state across-county sorting in our data is recorded in 2016, followed closely by the election of 1856, the 2012 election, and then the contentious 1860 presidential election that spawned the Civil War.<sup>16</sup> We note that, different from the results on across-state sorting, elections with large third-party vote shares are not outliers. This is likely due to a combination of factors. In particular, looking within states controls for many dimensions of the electoral environment: state competitiveness in the presidential race, other state-specific elections which impact turnout, qualification for the ballot by third-party or independent candidates, and even state-specific campaigning strategies by presidential campaigns.

Despite the re-emergence of geographic sorting along partisan lines, the divisions across states and across counties within states displayed in Fig. 2 are small in comparison to the within-county component of the variance index. The numbers in Fig. 3 make clear that except for three elections (i.e., 1856, 1860, and 1924), geographic sorting across states and across counties within the same state jointly accounts for less than 15% of the overall variance index (i.e.,  $P(x)$ ), and in only one year does it amount to more than 20% (1856). Even in light of rampant disenfranchisement in the South throughout much of this period, the evidence implies that, since at least the mid-19th century, the partisanship of American voters has always varied much more within rather than across areas.

<sup>16</sup> The results for 1856 and 1860 should be interpreted with caution, as the parties that competed in these elections were not truly national. With minor exceptions, this is not an issue for the remainder of our time series.

More formally, the within-county component of the variance index (i.e., the third term in Eq. (3)) has always exceeded all other components by at least a factor of three.

To benchmark the size of our estimates, we have simulated the decomposition in Fig. 2 under the assumption that voters are not spatially sorted. Our simulation holds constant the number of voters in each county and randomly assigns each voter to either the Democratic or Republican party according to the national two-party vote share in the election year. Unsurprisingly, our simulations show that, if voters are randomly sorted, then in every election, the within-county component of the variance in partisanship equals nearly 100% of the total variance. As a result, the findings above simultaneously imply two things: (1) There is much more heterogeneity in partisanship within than across counties, while (2) the mean differences that we do observe across counties are substantially greater than would be expected if there was no sorting along partisan lines.

One potential explanation for the high within-component of the variance index is that we are measuring within-state partisan sorting at the county level. As explained above, we rely on county-level electoral returns for the simple reason that the data go as far back as the existence of the Republican party. However, the mean county size is more than 100,000 residents. Due to the Modifiable Areal Unit Problem our county-level analysis may miss a significant amount of sorting that is only occurring at lower levels of aggregation. It is, for instance, conceivable that most sorting occurs across towns or neighborhoods *within* a county. For this reason, we also conduct decompositions at the precinct level. With slightly more than 1,000 registered voters on average, precincts are substantially smaller than counties, which should enable us to detect even very localized sorting.

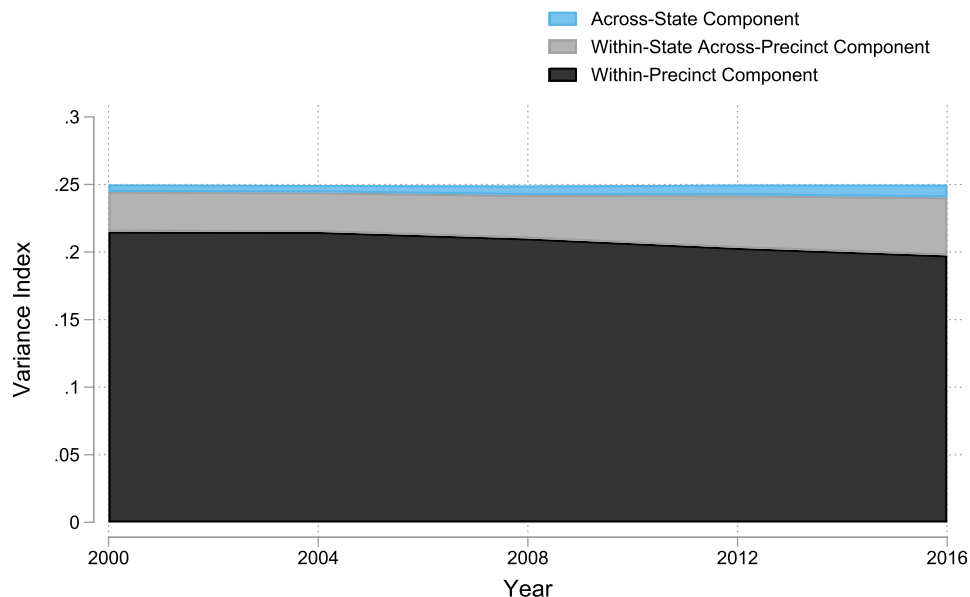
Unfortunately, precinct-level data are only available as of the 2000 election, and only for a subset of states. In particular, we have data for 22 states (including Washington, D.C.) in the 2000 general election, for 33 states in 2004, 39 in 2008, 40 in 2012, and 49 in 2016. In Fig. 4, we depict the across-state variance in blue, the within-state across-precinct variance in gray, and the within-precinct variance in black. Fig. 4 clearly shows that, in recent years, the U.S. is becoming more politically sorted at the precinct level. Moreover, the heterogeneity in partisanship is, on average, much higher within than across precincts. Our precinct-level results thus mirror the coarser (but longer) county-level analysis.

To ensure that the patterns above are not an artifact of the varying panel structure, we have replicated the results for a balanced panel of 12 states that we observe continuously from 2000 onwards. We present the resulting estimates numerically in Table 1. For ease of interpretation and comparison, we normalize all numbers by the overall variance index in the respective year. The numbers in a particular column thus correspond to the share of the variation in voters' partisanship that can be explained by mean differences across the respective geographic units.

Reassuringly, the degree of spatial sorting evident in the unbalanced panel is similar to that in the balanced one. This suggests that the observed trends are not due to the entry of highly heterogeneous states into our sample, but are instead reflective of the fact that partisan sorting is a national phenomenon.

Table 1 also replicates our main analysis at the precinct level. Unsurprisingly, the degree of partisan sorting is markedly greater at the precinct level than at the county level. More important, partisan sorting shows similar rates of growth irrespective of the level of aggregation. In other words, the trend towards greater geographic sorting is also borne out on the more-localized precinct level.

The first column in Table 1 presents the within-state across-county share, while the second column shows the national



**Fig. 4.** Precinct-Level Decomposition, Presidential Elections 2000–2016. *Notes:* Figure shows a decomposition of the variance index for each presidential election from 2000 to 2016. As explained in the text, the decomposition is based on the expression in Eq. (3) in conjunction with precinct-level election data.

**Table 1**  
Geographic Partisan Sorting, 2000–2016

Year	Within-State Across-County Share	Nationwide Across-County Share	Within-State Across-Precinct Share	Nationwide Across-Precinct Share	Within-State Across-Precinct Share
2000	.047	.071	.120	.139	.128
2004	.054	.075	.120	.139	.126
2008	.059	.084	.132	.157	.146
2012	.069	.097	.160	.187	.174
2016	.095	.130	.178	.209	.194
Panel:	Balanced	Balanced	Unbalanced	Unbalanced	Balanced

*Notes:* Table shows the share of the variance index due to geographic sorting at the county and precinct levels, either within states or nationwide. There are 22 states (incl. Washington D.C.) in our precinct-level data for 2000, 33 states in 2004, 39 in 2008, 40 in 2012, and 49 in 2016. The balanced precinct-level panel includes 12 states.

across-county share of the variance index based on our precinct-level data. The third and fourth columns show the within-state and national across-precinct shares, respectively. The precinct-level shares are 2–3 times the size of the county-level shares. Also, the across-county share nearly doubles between 2000 and 2016 whereas the across-precinct share increases by about 50%. Since the number of states for which we have data grows dramatically over time, we show that our results are not due to compositional changes in the states in our sample (cf. Columns 1, 2 and 5). Regardless of whether we consider a balanced or unbalanced sample of states, the results are qualitatively similar.

To benchmark the estimates in Table 1, we have computed a second set of placebo decompositions under the assumption of no sorting along partisan lines. Even at the substantially smaller precinct level, we find that random sorting leads to nearly 100% of the variance being within rather than across precincts.

In sum, all pieces of evidence tell a similar story. Geographic sorting has been increasing over the last half-century and strongly so in recent years. Yet, spatial cleavages in partisanship pale relative to differences within even the smallest of geographical units.

In Appendix Fig. A.1, we compare our findings to results based on four other commonly used measures of spatial segregation and polarization. Specifically, we compare the across-county

component of our variance index to the dissimilarity index, the Gini coefficient, the isolation index, and the polarization index of Esteban and Ray (1994). To make these measures more comparable, we compute all of them at the county level and normalize each to 100 in 1976<sup>17</sup>. Recall, the dissimilarity index measures the average gap in Democratic and Republican vote shares across counties. The Gini coefficient computes the fraction of partisans that would have to be redistributed in order to eliminate sorting across counties (i.e., to equalize the two-party vote share for the Democrats across counties). The isolation index computes the probability that people of opposite political parties randomly meet within their counties. Finally, the polarization index is similar to the Gini coefficient except that it raises population weights to a power greater than unity. Although our index is more sensitive and thus varies more over time than do the other ones, all of the measures we consider, with the exception of the isolation index, show a similar qualitative U-shaped pattern over the past century and a half. The variance index, however, is the only one which satisfies all of the theoretical desiderata in Axioms 1–7.

<sup>17</sup> We choose to normalize the indices to 100 in 1976 because that is the year in which our variance index achieves its minimum



One possible explanation for increasing geographic polarization among voters is that turnout patterns have changed over time. If this were the case, then our results might reflect a change in voting patterns rather than a change in residential partisan clustering. For such a theory to explain our findings it would need to be the case that turnout by Democrats increased in Democratic party strongholds, while that of Republicans increased in predominantly Republican areas. If correct, then our results would only document geographic sorting among actual voters, but not among citizens more generally. Although the former would be interesting in and of itself, we address this potential concern in two complementary ways.

First, we regress the change (from one election to the next) in the within-state across-county component of the variance index on the percentage change in turnout in the same state. Over the entire period from 1856 to 2016, there is only a weak relationship between the two variables. Though the estimated coefficient is statistically significant, it is small and negative ( $\beta = -.003, p = .006$ ). For the more recent period of rising partisan sorting (i.e., from 1976 onwards), we do obtain a larger but *negative* and statistically significant estimate ( $\beta = -.009, p = .000$ ).<sup>18</sup> While lower turnout is associated with more geographic sorting, the resulting coefficient is at least one order of magnitude too small to explain the rise in the within-state across-county share. For example, the first percentile of negative percentage changes in turnout is  $-0.788$ . A decline of this magnitude would lead to a .007 increase in our measure, while the actual rise in within-state across-county sorting is greater than .07 during the post-1976 period. Moreover, during the time period of most rapid increase in partisan sorting, we have seen an increase rather than a decrease in turnout. We thus conclude that changes in overall turnout do not appear nearly large enough to explain our findings.

It could, of course, be the case that turnout changed differentially for Democratic and Republican supporters, and that these changes offset each other. In order to assess the plausibility of this explanation, we follow [Sussell \(2013\)](#), and turn to voter registration data from the state of California. Specifically, we collected these data for 2006, 2008, 2011 and 2017, and, for each year, compute within-state across-precinct sorting among registered Democrats and Republicans. The results are presented in [Appendix Table A.2](#). Reassuringly, we observe an increase in partisan sorting of about 25%.<sup>19</sup>

#### 4.2. Partisan sorting state-by-state

The times series evidence presented above implicitly averages across states. In principle, it is possible that some states experience very high levels of geographic sorting, whereas others see almost none. In order to investigate differences across states, we return to our county-level election data and document within-state across-county sorting separately for each state. To conserve on space, [Fig. 5](#) reports results for four different elections: 1860, 1972, 2012 and 2016. Since the total variance in partisanship varies by state, we calculate the across-county component of the variance index separately by state and normalize by the total variance in partisanship in the state.

The 1860 presidential election is a definite outlier. Virginia (47.36%), New Jersey (27.14%) and Missouri (23.77%) accounted

for almost all of the variance in electoral returns.<sup>20</sup> All other states except for Maryland (10.13%) had within-state across-county shares below 10%.<sup>21</sup> By contrast, in 1976, mean differences across counties accounted for more than 10% of within-state heterogeneity only in Washington, D.C. (15.32%). Out of all other states, only New York (5.22%) had an across-county heterogeneity share greater than 5%. By 2012, within-state across-county sorting had risen to more than 5% in 32 states and Washington, D.C., of which nine saw shares greater than 10%. The states with the highest across-county shares were Maryland (14.70%), Georgia (14.36%), Mississippi (11.83%) and Louisiana (10.61%)—all of which are Southern. Four years later, spatial divisions within states had increased yet further, on average. Fifteen states had across shares greater than 10%. The five geographically most-polarized states in 2016 were Maryland (18.27%), Georgia (17.42%), Missouri (15.29%), New York (13.47%), and Illinois (13.32%).

Looking at the evidence, three clear patterns emerge. (1) The rise in partisan sorting across counties reflects a broad-based phenomenon that is apparent in all states. (2) At any given point in time, there are considerable differences across states, and (3) higher levels of geographic sorting are present in the South.

What explains these differences? Although the main purpose of our paper is to document rather than explain patterns of geographical sorting along partisan lines over time, we now briefly explore potential reasons for the observed changes. To do so we turn to the 2010 Decennial Census and calculate, for each state, the across-county variance in median household income, educational attainment (as measured by college graduation), the share of whites and blacks, and the share of urban households. We then regress these variables on our measure of geographic sorting within each state in 2016. [Appendix Table A.3](#) presents the results. Remarkably, all five factors together explain nearly 60% of the variation in our data ( $R^2 = .599$ ). Looking at the explanatory power of each variable in isolation, we see that states with a higher level of geographic sorting on partisanship are, first and foremost, states with high degrees of racial clustering, with more sorting on education, and (to a lesser degree) sorting on income. In fact, within-state clustering of African-Americans and whites respectively explains 43.0% and 39.2% of the between-state differences respectively in partisan sorting. Surprisingly, urban-rural differences appear to be the least important predictor—though we stress again that the correlations in [Appendix Table A.3](#) should not be interpreted as causal.

#### 5. Concluding remarks

The analysis in this paper documents that the long arc of legislative polarization has carried with it a similar arc in partisan sorting across space. Our results are based on a novel approach to measuring geographic sorting of voters. Specifically, we introduce the variance index, which is the only measure of voter heterogeneity that satisfies six intuitively desirable criteria *and* that can be perfectly decomposed into across- and within-group components.

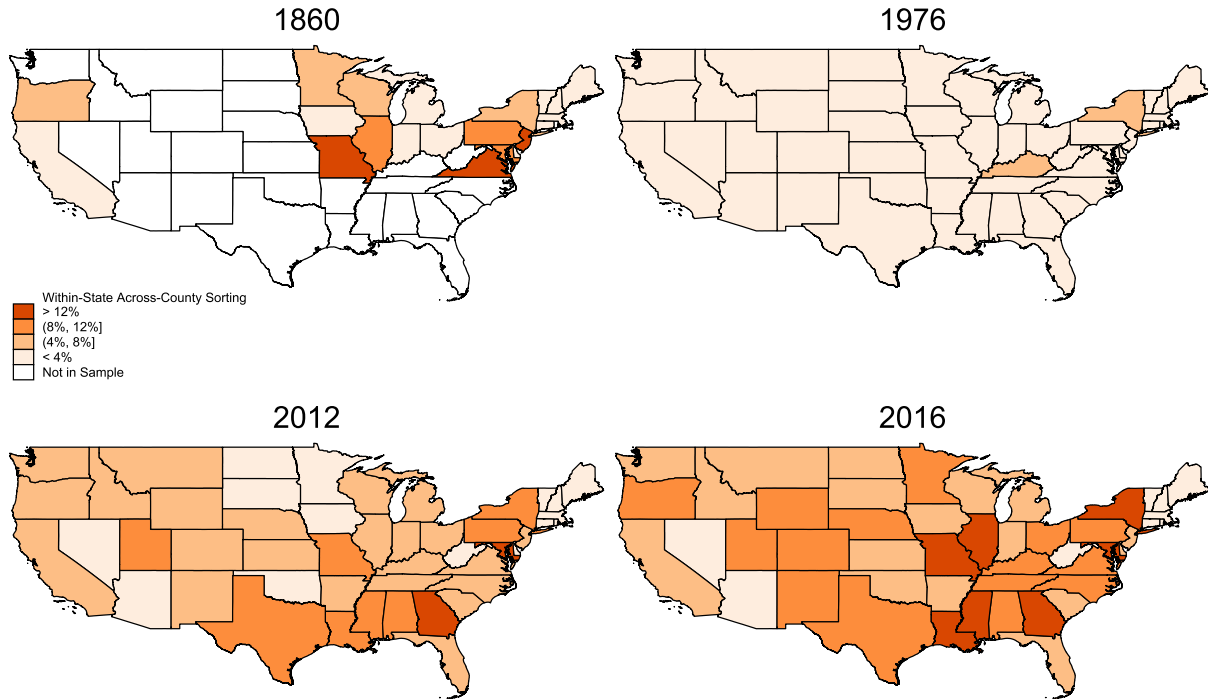
Relying on this decomposition, we show that there has been a steady rise in geographic partisan sorting since the early 1970s. This trend has accelerated after 2000. Current partisan cleavages across states are as high as at any time in the last fifty years, and geographic partisan sorting within states is at an all-time high in the post-Civil War era. In particular, recent times of legislative polarization have also been times of high geographic clustering of partisans.

<sup>20</sup> Of course, in the mid-19th century, a much smaller number of people lived in the U.S., and an even smaller number were eligible to vote—only white males in most states.

<sup>21</sup> A number of southern states are absent from our data because the Republican party was not on the ballot.

<sup>18</sup> If Republican supporters are more likely to be marginal voters in areas where the Democrats are dominant, while Democrats are more likely to be marginal voters in counties where the Republican party is dominant, then such a negative relationship may be expected. Exactly this is shown in [Fujiwara et al. \(2016\)](#).

<sup>19</sup> This analysis disregards a significant number of independent and third-party voters. As a robustness check, we have replicated our analysis assuming that these individuals' partisan preferences are located at the midpoint between Democrats and Republicans and obtained qualitatively equivalent results.



**Fig. 5.** Within-State Across-County Sorting, by State and Presidential Election. *Notes:* Figure shows our measure of partisan sorting across counties within the same state separately for each state in each of the 1860, 1976, 2012, and 2016 presidential elections. Since the total variance in partisanship varies by state, we normalize the across-county component of the variance index for that state by the total variance in the state in that year.

Yet partisan heterogeneity within counties or precincts is many times greater than divisions across space. For instance, mean differences across precincts account for only about 1/3th of the overall variance index. By this measure, the American electorate continues to be much more diverse within than across communities, even when the latter are narrowly defined.

Nonetheless, we stress that we do not know how much geographic sorting is “too much.” By historical standards, spatial cleavages are very large, and even small amounts of sorting may lead to legislative dysfunction and conflict, especially in a winner-take-all electoral system (see, e.g., Hopkins, 2017). Further, we do not know whether geographic sorting is a cause or a consequence of polarization in Congress. Does partisan sorting have an effect on the evolution of voter views and voter preferences? Does it affect the quality of representation or the provision of local public goods? All of these questions are substantively important. Given that current levels of geographic sorting have not been seen in generations, we hope that our findings and methods pave the way for future research along these lines.

**Appendix A. Proofs**

**Lemma 1** Bosmans and Cowell, 2010. An index  $P$  satisfies Axioms 1–5, the strict Pigou-Dalton principle, and admits an aggregation function  $A$ , which is continuous and strictly increasing in its first two arguments, with  $P(x, y) = A(P(x), P(y), \bar{x}, \bar{y}, n_x, n_y) \forall x \in \mathbb{R}^{n_x}, y \in \mathbb{R}^{n_y}$ , if and only if there exists some  $\kappa \in \mathbb{R}$  and a continuous, strictly increasing function  $f : \mathbb{R} \rightarrow \mathbb{R}$ , with  $f(0) = 0$ , such that, for all  $z \in \mathbb{R}^n$ ,

$$f(P(z)) = \begin{cases} \frac{1}{n} \sum_{i=1}^n \{ \exp(\kappa[z_i - \bar{z}]) - 1 \} & \text{if } \kappa \neq 0 \\ \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2 & \text{if } \kappa = 0 \end{cases}$$

**Proof.** See Bosmans and Cowell (2010). □

**Lemma 2.** Axioms 3, 6 and 7 together imply the strict Pigou-Dalton principle.

**Proof.** The strict Pigou-Dalton principle requires that  $P(z) < P(z')$  whenever  $z = (z_1, \dots, z_i, \dots, z_j, \dots, z_n)$  with  $z_i \leq z_j$  and  $z' = (z_1, \dots, z_i - c, \dots, z_j + c, \dots, z_n)$  for some  $c > 0$ . Let  $y \in \mathbb{R}^n, x = (x_1, x_2)$  with  $x_1 \leq x_2$ , and  $x' = (x_1 - c, x_2 + c)$ . By Axiom 7,

$$\begin{aligned} P(x, y) &\geq P(x', y) \\ \omega(\bar{x}, \bar{y}, 2, n)P(x) + \omega(\bar{y}, \bar{x}, n, 2)P(y) + P(\bar{x}, \bar{y}) &\geq \omega(\bar{x}, \bar{y}, 2, n)P(x') + \omega(\bar{y}, \bar{x}, n, 2)P(y) + P(\bar{x}, \bar{y}) \\ P(x) &\geq P(x') \end{aligned}$$

Since  $P(x) < P(x')$  by Axiom 6, it follows that  $P(x, y) < P(x', y)$ , as desired. Anonymity, i.e., Axiom 3, further ensures that the Pigou-Dalton principle is satisfied for mean-preserving spreads in arbitrary positions. □

**Lemma 3.** Suppose  $P(x) = \frac{1}{q} \sum_{i=1}^n \{ \exp(\kappa[x_i - \bar{x}]) - 1 \}$  for some  $q \in \mathbb{R}$  and  $\kappa \neq 0$ . Then there exists no weighting function  $\omega$  that satisfies Axiom 7.

**Proof.** Our proof is in two parts. First, we show that a weighting function satisfies condition (i) of the axiom if and only if  $\omega(\bar{x}, \bar{y}, n_x, n_y) = \frac{n_x}{n_x + n_y} \frac{\exp(\kappa \bar{x})}{\exp\left(\kappa \left(\frac{n_x}{n_x + n_y} \bar{x} + \frac{n_y}{n_x + n_y} \bar{y}\right)\right)}$ . We then prove that this implies  $\omega(\bar{x}, \bar{y}, n_x, n_y) + \omega(\bar{y}, \bar{x}, n_y, n_x) \neq 1$  whenever  $\kappa \neq 0$ , which violates condition (ii).

It is easy to verify that  $\omega(\bar{x}, \bar{y}, n_x, n_y) = \frac{n_x}{n_x+n_y} \frac{\exp(\kappa\bar{x})}{\exp\left(\kappa\left[\frac{n_x}{n_x+n_y}\bar{x} + \frac{n_y}{n_x+n_y}\bar{y}\right]\right)}$  satisfies condition (i) if  $P(x) = \frac{1}{q} \frac{1}{n} \sum_{i=1}^n \{\exp(\kappa[x_i - \bar{x}]) - 1\}$ . To prove that it is the only weighting function that does so, let  $y = \bar{y}$ . In this particular case,  $P(y) = 0$  and condition (i) reduces to  $P(x, y) = \omega(\bar{x}, \bar{y}, n_x, n_y)P(x) + P(\bar{x}, \bar{y})$ . Letting  $\bar{z} = \frac{n_x}{n_x+n_y}\bar{x} - \frac{n_y}{n_x+n_y}\bar{y}$  and substituting for  $P$  gives:

$$\frac{1}{q} \frac{1}{n_x+n_y} \left( \sum_{i=1}^{n_x} \{\exp(\kappa[x_i - \bar{z}]) - 1\} + \sum_{i=1}^{n_y} \{\exp(\kappa[\bar{y} - \bar{z}]) - 1\} \right) = \omega(\bar{x}, \bar{y}, n_x, n_y) \frac{1}{q} \frac{1}{n_x} \sum_{i=1}^{n_x} \{\exp(\kappa[x_i - \bar{x}]) - 1\} + \frac{1}{q} \frac{1}{n_x+n_y} \left( \sum_{i=1}^{n_x} \{\exp(\kappa[\bar{x} - \bar{z}]) - 1\} + \sum_{i=1}^{n_y} \{\exp(\kappa[\bar{y} - \bar{z}]) - 1\} \right)$$

Solving this expression for  $\omega(\bar{x}, \bar{y}, n_x, n_y)$  yields

$$\begin{aligned} \omega(\bar{x}, \bar{y}, n_x, n_y) &= \frac{n_x}{n_x+n_y} \frac{\sum_{i=1}^{n_x} \exp(\kappa[x_i - \bar{z}]) - \sum_{i=1}^{n_x} \exp(\kappa[\bar{x} - \bar{z}])}{\sum_{i=1}^{n_x} \{\exp(\kappa[x_i - \bar{x}]) - 1\}} \frac{\exp(\kappa\bar{z})}{\exp(\kappa\bar{x})} \frac{\exp(\kappa\bar{x})}{\exp(\kappa\bar{z})} \\ &= \frac{n_x}{n_x+n_y} \frac{\sum_{i=1}^{n_x} \exp(\kappa x_i) - \sum_{i=1}^{n_x} \exp(\kappa\bar{x})}{\sum_{i=1}^{n_x} \{\exp(\kappa x_i) - \exp(\kappa\bar{x})\}} \frac{\exp(\kappa\bar{x})}{\exp(\kappa\bar{z})} \\ &= \frac{n_x}{n_x+n_y} \frac{\exp(\kappa\bar{x})}{\exp(\kappa\bar{z})} \end{aligned}$$

This shows that  $\omega(\bar{x}, \bar{y}, n_x, n_y) = \frac{n_x}{n_x+n_y} \frac{\exp(\kappa\bar{x})}{\exp\left(\kappa\left[\frac{n_x}{n_x+n_y}\bar{x} + \frac{n_y}{n_x+n_y}\bar{y}\right]\right)}$  is the only weighting function that satisfies condition (i) when  $y = \bar{y}$ . Hence, there cannot exist a different weighting function that satisfies the same condition for all  $x \in \mathbb{R}_x^n$  and  $y \in \mathbb{R}_y^n$ , which completes the first part of the proof.

To show that  $\omega(\bar{x}, \bar{y}, n_x, n_y) + \omega(\bar{y}, \bar{x}, n_y, n_x) \neq 1$  we proceed by way of contradiction. Suppose that  $\kappa \neq 0$  and  $\omega(\bar{x}, \bar{y}, n_x, n_y) + \omega(\bar{y}, \bar{x}, n_y, n_x) = 1$ . Plugging in our candidate solution for  $\omega$  and rearranging yields the condition:

$$\frac{n_x}{n_x+n_y} \exp(\kappa\bar{x}) + \frac{n_y}{n_x+n_y} \exp(\kappa\bar{y}) = \exp\left(\kappa\left[\frac{n_x}{n_x+n_y}\bar{x} + \frac{n_y}{n_x+n_y}\bar{y}\right]\right)$$

Since  $\exp(\cdot)$  is a convex function, Jensen's inequality implies that, unless  $\kappa = 0$ , the LHS of the expression above is strictly greater than the RHS, which produces the desired contradiction.  $\square$

**Proof of Proposition 1.** It is straightforward to verify that the variance index satisfies Axioms 1–7 with  $\omega(\bar{x}, \bar{x}, n_x, n_y) = \frac{n_x}{n_x+n_y}$ . We, therefore, focus on proving that it is the only index that does so (up to scalar multiplication).

Since Axioms 3, 6 and 7 imply the strict Pigou-Dalton principle (cf. Lemma 2) and since the aggregation function in Axiom 7 is a special case of that in the Lemma 1, any heterogeneity index that satisfies Axioms 1–7 must be contained in the class of indices characterized by Lemma 1. Hence, it suffices to show that, given Axiom 7,  $f$  in Lemma 1 must be an affine transformation and  $\kappa = 0$ .

Suppose that  $f$  is, indeed, an affine transformation and that  $\kappa \neq 0$ . Then,  $P(x) = \frac{1}{q} \frac{1}{n} \sum_{i=1}^n \{\exp(\kappa[x_i - \bar{x}]) - 1\}$  for some constant  $q \in \mathbb{R}$ . From Lemma 3 we know that Axiom 7 fails in this case. It, therefore, follows that if  $f$  is an affine transformation, then  $\kappa = 0$ .

To show that  $f$  must be an affine transformation let  $n_x = n_y$  and consider any  $x \in \mathbb{R}_x^n$  and  $y \in \mathbb{R}_y^n$  such that  $\bar{x} = \bar{y}$ . By Axioms 1 and 3, condition (i) in Axiom 7 reduces to  $P(x, y) = \omega P(x) + (1 - \omega)P(y)$  with (ii)  $\omega = \omega(\bar{x}, \bar{y}, n_x, n_y) = \omega(\bar{y}, \bar{x}, n_y, n_x) = \frac{1}{2}$ . Applying  $f$  to both sides of the equation, gives

$$f(P(x, y)) = f\left(\frac{P(x) + P(y)}{2}\right)$$

Now, if  $n_x = n_y$  and  $\bar{x} = \bar{y}$ , then, relying on the explicit expressions for  $f$  in Lemma 1, it is possible to show that, for any  $\kappa$ ,

$$f(P(x, y)) = \frac{f(P(x)) + f(P(y))}{2}$$

We, therefore, have that  $f\left(\frac{P(x)+P(y)}{2}\right) = \frac{f(P(x))+f(P(y))}{2}$ , which is Jensen's Equality. The solutions to this functional equation are known to be of the form  $f(x) = qx + s$  for some constants  $q, s \in \mathbb{R}$  (cf. Aczél, 1966, ch. 2, Theorem 1). Hence,  $f$  is an affine transformation, as desired.  $\square$

## Appendix B. Data appendix

### B.1. County-level election returns

We obtained county-level presidential election returns for the years 1972 through 2016 from the *CQ Voting and Elections Collection* (<http://library.cqpress.com/elections/>) and data from 1856 to 1968 from *Inter-university Consortium for Political and Social Research* (1999). In a small number of cases, the Democratic or Republican party was not listed as fielding a candidate in a particular general election. We dealt with this issue on a case-by-case basis. In many of the affected state-years, the name of the party listed in the ICPSR data was slightly different for that particular year. In some cases, however, the state party did not list the national candidate, or the candidate did not qualify for the ballot for idiosyncratic reasons. We detail these exceptions below:

**Rhode Island, 1856:** The Republicans did not field a candidate in Rhode Island. The Democrats did and the American party (“Know-Nothing party”) put up Millard Fillmore.

**Tennessee, 1856:** The Republicans did not field a candidate in Tennessee. The Democrats did and the American party (“Know-Nothing party”) put up Millard Fillmore.

**Virginia, 1856:** The Republicans did not field a candidate in Tennessee. The Democrats did and the American party (“Know-Nothing party”) put up Millard Fillmore.

**Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, and Texas, 1860:** The Republican party did not get on to the ballot in these 10 Southern states. Instead, the Constitutional Union party did. The Constitutional Union party, a pro-union party largely in the South arguing in favor of maintaining the union by ignoring slavery, ran a candidate, as did the Southern Democratic party.

**New Jersey, 1860:** NJ selected its electors before the Dيمقراطية party split into the Northern and Southern Democratic parties at the South Carolina convention. NJ was a fusion state where different electors got to choose different Democratic candidates. We count all votes as votes for the Democratic party.

**South Dakota, 1896:** William Jennings Bryan ran as a candidate both for the Populist as well as the Democratic party. We count his votes as votes for the Democratic party.

**Wyoming, 1892:** Only the Populist and the Republican party ran. Grover Cleveland and the Democrats were not on the ballot. We do not include Wyoming, 1892.

**Oregon, 1900:** The name of the Oregon Democrat party in 1900 was the People's and Democratic party. The Republican party was called the Modern Republican party.

**Nevada, 1904:** The name of the Democratic party in 1904 was the Democrat and Silver party.

**South Dakota, 1912:** The Republican party was called the Progress Republican party.

**Mississippi, 1932, 1936:** The Republican party split into two factions: Go for the Lily-White faction and the Black-and-Tan faction. The national Republican party ran with the Lily-White faction as the Lily-White Republican party.

**Alabama, 1964:** The Alabama Democratic Party did not support the national Democratic party's nominee, Lyndon B. Johnson. Thus, the state party passed a resolution unpledging their electors. We count Alabama's votes for the Democratic party as actual Democratic votes despite their being unpledged.

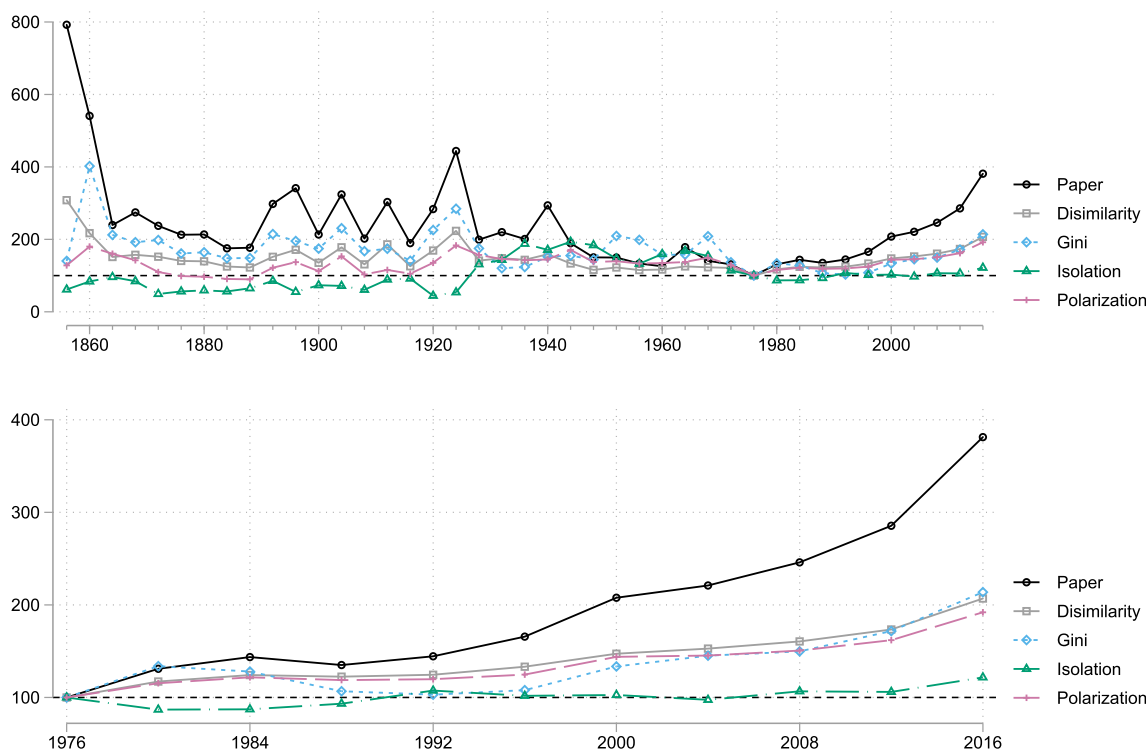
**Mississippi, 1964:** Our data for this state year come from Wikipedia ([https://en.wikipedia.org/wiki/United\\_States\\_presidential\\_election\\_in\\_Mississippi,\\_1964](https://en.wikipedia.org/wiki/United_States_presidential_election_in_Mississippi,_1964)).

*B.2. Precinct-level data*

The bulk of the precinct-level voting data came from the Harvard Election Data Archive. It was downloaded in the form of text files, one for each year's worth of data for each state. These were then converted into STATA files and appended together so that each observation contains data for each precinct and year combination. Two master files were created out of the Harvard data: one for data from 2000 to 2010, and another for data from 2012. This is because the 2010 census redrew precincts and congressional districts for the 2012 election cycle. Voting data for Representatives in Congress and for President (when applicable), as well as geographically-identifying information was kept during the merge, with the rest of the data discarded.

Additional precinct-level data was also collected from various states' websites. Some of the Harvard data files were updated with corrected information from these websites, but most of the data collected from the states' websites was for state and year combinations not available in the Harvard database. Additional data was also purchased from David Leip's Election Atlas. When no precinct-level data could be found online, the Secretaries of State for each state were called to verify that no precinct-level data existed for that state. More details can be found on the exact sources of all data in the State by State section, below. Once the two master files were created, each was checked for errors and inconsistencies. A number of listed precincts in the data were discovered not to be geographical units but rather observations which reported county totals or absentee votes. These observations were discarded as a result.

Additionally, states with town-level data were double-checked to see if any more geographically specific data was available. If precinct-level data was available, it replaced the town-level data, but when no precinct-level data was available, towns were labeled as such. Some precincts seem to have voting data for one party but not the other. In most cases, this is the result of one party not putting forward a candidate for the House of Representatives in a given congressional district. The remaining cases all occur in very small precincts, usually with less than a hundred votes cast in total. We concluded that these precincts simply happened to be unanimous due to their small size (if only two votes are cast in a precinct and both happen to be for a Republican, this does not suggest that we are missing the Democratic voting data, but rather that no Democratic votes were cast in that precinct). A more detailed breakdown of precincts missing one party's voting data is available upon request, listing which districts lacked either a Democratic or Republican candidate and the sizes of the precincts imbalanced in an election where both Democratic and Republican candidates ran (see Fig. A.1).



**Fig. A.1.** Comparison with Other Indices, Presidential Elections 1856–2016. *Notes:* Figure compares across-county sorting according to our measurement approach with four other commonly used indices. As explained in the main text, we normalize each index to 100 in 1976.



When neither the Harvard Election Data Archive nor David Leip's Election Atlas contained data for a particular state-year combination or when data anomalies existed, we directly contacted the respective Secretaries of State to either obtain the data or verify that precinct-level electoral returns were not kept for the election in question. In all cases, we only include precinct data which can be aggregated to match within 5% of a state's total votes and 2% of two party vote share from a given state's election returns provided by the Federal Election Commission. In some cases, data are available for states, but the aggregated totals exclude either portions of the state, or early and absentee voting.

Appendix Table A.1 provides a complete list of state-year combination for which we have verified, trustworthy precinct-level election data.

**Table A.2**  
Geographic Sorting Among Registered Partisans in California.

Year	Within-State Across-Precinct Share of State Variance Index	Within-County Across-Precinct Share of State Variance Index	Within-State Across-County Share of State Variance Index
2006	0.126	0.058	0.068
2008	0.126	0.058	0.068
2011	0.134	0.064	0.070
2013	0.140	0.063	0.077
2017	0.162	0.066	0.096

Notes: Numbers are within-state across-precinct shares of the variance index based on voter registration data from the state of California for the years 2006, 2008, 2011, 2013, and 2017, within-county across-precinct shares, and within-state across-county shares. For the purpose of calculating and decomposing our variance index, we only consider registered Democrats and Republicans.

**Table A.1**  
Verified Presidential Precinct Data

State	2000	2004	2008	2012	2016
AK	No	Yes	Yes	Yes	Yes
AL	No	Yes	Yes	Yes	Yes
AR	Yes	Yes	Yes	Yes	Yes
AZ	No	Yes	Yes	Yes	Yes
CA	No	Yes	Yes	Yes	Yes
CO	No	Yes	Yes	No	Yes
CT	Yes	Yes	Yes	Yes	Yes
DC	No	Yes	Yes	Yes	Yes
DE	No	Yes	Yes	Yes	Yes
FL	No	No	No	No	Yes
GA	No	No	No	Yes	Yes
HI	Yes	No	Yes	Yes	Yes
IA	No	Yes	Yes	Yes	Yes
ID	Yes	Yes	Yes	Yes	Yes
IL	No	No	No	Yes	Yes
IN	No	No	No	No	Yes
KS	Yes	Yes	Yes	Yes	Yes
KY	No	No	Yes	Yes	Yes
LA	Yes	Yes	Yes	Yes	Yes
MA	No	Yes	Yes	Yes	Yes
MD	Yes	No	No	No	Yes
ME	Yes	No	Yes	Yes	Yes
MI	Yes	Yes	Yes	No	Yes
MN	Yes	Yes	Yes	Yes	Yes
MO	No	Yes	Yes	No	No
MS	No	No	Yes	Yes	Yes
MT	Yes	Yes	Yes	Yes	Yes
NC	Yes	Yes	Yes	Yes	Yes
ND	No	Yes	Yes	Yes	Yes
NE	No	No	Yes	Yes	Yes
NH	Yes	Yes	No	Yes	Yes
NJ	Yes	No	No	Yes	Yes
NM	Yes	Yes	Yes	Yes	Yes
NV	No	No	No	Yes	Yes
NY	No	No	No	No	Yes
OH	No	Yes	Yes	Yes	Yes
OK	No	Yes	No	Yes	Yes
OR	No	No	Yes	No	Yes
PA	Yes	Yes	Yes	Yes	Yes
RI	Yes	Yes	Yes	Yes	Yes
SC	No	Yes	Yes	Yes	Yes
SD	No	Yes	Yes	Yes	Yes
TN	No	Yes	Yes	Yes	Yes
TX	Yes	Yes	Yes	Yes	Yes
UT	No	No	Yes	No	Yes
VA	Yes	No	No	Yes	Yes
VT	Yes	Yes	Yes	Yes	No
WA	No	No	Yes	No	Yes
WI	Yes	Yes	Yes	No	Yes
WV	No	No	No	Yes	Yes
WY	Yes	Yes	Yes	Yes	Yes

**Table A.3**  
Predictors of Within-State Across-County Sorting in 2016.

	Within-State Across-County Share of Variance Index					
	(1)	(2)	(3)	(4)	(5)	(6)
Var. in Median HH Income	-.082 (.069)	.174** (.070)				
Var in Percent College Grad.	5.597*** (1.758)		6.228*** (1.706)			
Var in Percent White	1.005** (.494)			2.269*** (.437)		
Var. in Percent Black	1.032** (.445)				2.431*** (.349)	
Var in Urban Pop. Share	.033 (.118)					.433** (.175)
R2	.599	.201	.367	.392	.430	.112
Number of Observations	51	51	51	51	51	51

Notes: Numbers are point estimates and standard errors from regressing our measure of geographic partisan sorting in a particular state on the variables listed in the left-most column. Specifically, the dependent variable in all regressions is the within-state across-county share of heterogeneity in partisanship in the respective state, including Washington D.C. The independent variables are the county-level variance of median household income in the same state, the variance in the share of college graduates, the variance in the share of whites and blacks, as well as the county-level variance in the population percentage that is urban. Standard errors are heteroskedasticity robust and reported in parentheses. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively.

**References**

Abrams, S., Fiorina, M., 2012. The Big Sort' That Wasn't: A Skeptical Reexamination. *PS: Polit. Sci. Polit.* 45 (2), 203–210.  
 Aczél, J., 1966. *Lectures on Functional Equations and their Applications*. Academic Press, New York.  
 Ansolabehere, S., Rodden, J., Snyder, J., 2006. Purple America. *J. Econ. Perspect.* 20 (2), 97–118.  
 Bishop, B., 2008. *The Big Sort: Why the Clustering of Like-Minded America Is Tearing Us Apart*. Houghton Mifflin Harcourt, New York.  
 Bosmans, K., Cowell, F., 2010. The Class of Absolute Decomposable Inequality Measures. *Econ. Lett.* 109, 154–156.  
 Brown, J., Enos, R., 2018. The Measurement of Partisan Sorting for 180 Million Voters. *Nat. Hum. Behav.* 5 (8), 998–1008.  
 Darmofal, R., Strickler, R., 2019. *Demography, Politics, and Partisan Polarization in the United States, 1828–2016*. Springer, Cham, Switzerland.  
 Esteban, J.-M., Ray, D., 1994. On the Measurement of Polarization. *Econometrica* 62 (4), 819–851.  
 Fujiwara, T., Meng, K., Vogl, T., 2016. Habit Formation in Voting: Evidence from Rainy Elections. *Am. Econ. J. Appl. Econ.* 8 (4), 160–188.  
 Gimpel, J., Hui, I., 2015. Seeking Politically Compatible Neighbors? The Role of Neighborhood Partisan Composition in Residential Sorting. *Polit. Geogr.* 45, 130–142.  
 Glaeser, E., Ward, B., 2006. Myths and Reality of American Political Geography. *J. Econ. Perspect.* 20 (2), 119–144.

Hopkins, D., 2017. *Red Fighting Blue: How Geography and Electoral Rules Polarize American Politics*. Cambridge University Press, New York.  
 Inter-university Consortium for Political and Social Research, 1999. *United States Historical Election Returns, 1824–1968*. Inter-University Consortium for Political and Social Research [distributor], Ann Arbor, MI, <https://doi.org/10.3886/ICPSR00001.v3>.  
 Klinkner, P., 2004a. Red and Blue Scare: The Continuing Diversity of the American Electoral Landscape. *Forum* 2 (2), Article 2.  
 Klinkner, P., 2004b. Counter Response from Klinkner to Bishop and Cushing. *Forum* 2 (2), Article 9.  
 Martin, G., Webster, S., 2020. Does Residential Sorting Explain Geographic Polarization? *Polit. Sci. Res. Methods*.  
 Massy, D., Denton, N., 1988. The Dimensions of Residential Segregation. *Soc. Forces* 67 (2), 281–315.  
 McDonald, I., 2011. Migration and Sorting in the American Electorate: Evidence from the 2006 Cooperative Congressional Election Study. *Am. Polit. Res.* 39 (3), 512–533.  
 Mummolo, J., Nall, C., 2017. Why Partisans Do Not Sort: The Constraints on Political Segregation. *J. Polit.* 79 (1), 45–59.  
 Poole, K., Rosenthal, H., 1997. *Congress: A Political-Economic History of Roll Call Voting*. Oxford University Press, New York.  
 Sussell, J., 2013. New Support for the Big Sort Hypothesis: An Assessment of Partisan Geographic Sorting in California, 1992–2010. *PS: Polit. Sci. Polit.* 46(4), 768–773.