Introduction

Good job today. To reiterate, basically I think your research question should be something like:

"Research shows that districts that consist of more homogeneous groups of voters achieve better representation on several dimensions. Meanwhile, many statutes require that districts be"compact", a term with many interpretations. It is not clear, however, that more compact districting plans (however compactness is measured) result in more homogeneous districts. Are compactness and homogeneity fundamentally conflicting goals? Are some measures of compactness more consistent with homogeneity than others?"

My contribution

- i develop the human compactness metric
- i use MCMC to measure spatial diversity
- first to measure relationship between compactness and normative outcomes?
- at the very least, first to measure relationship between compactness and spatial diversity

Theoretical stu

Why compactness

- states mandate it
- good check against gerrymandering

Why spatial diversity

> spatial diversity remained a statistically significant predictor of roll-o rate. With these variables held constant at their means, a House district's shift from the tenth to the ninetieth percentile in spatial diversity was associated with an increase in roll-o rate of about six percentage points

The final political gerrymandering issue that I investigate is how spatial diversity relates to common dis- trict plan metrics such as partisan bias and electoral responsiveness. As noted earlier, partisan bias refers to the divergence in the share of seats that each party would win given the same share of the statewide vote.254 For example, if Democrats would win forty-eight percent of the seats with fifty percent of the vote (in which case Republicans would win fifty-two percent of the seats), then a district plan would have a pro-Republican bias of two percent. Electoral responsiveness refers to the rate at which a party gains or

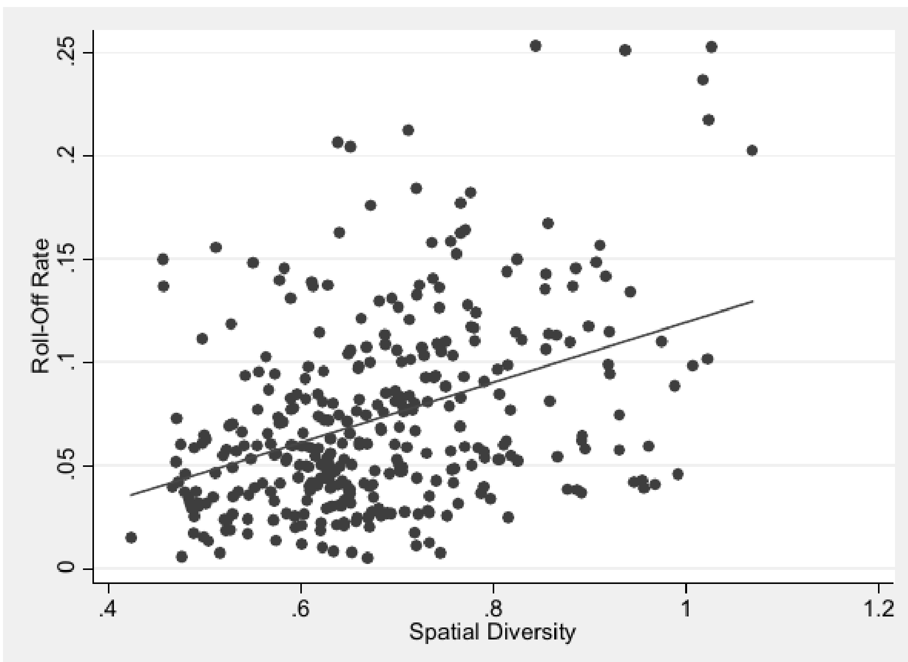


Figure 1: E ect of spatial diversity on electoral rollo

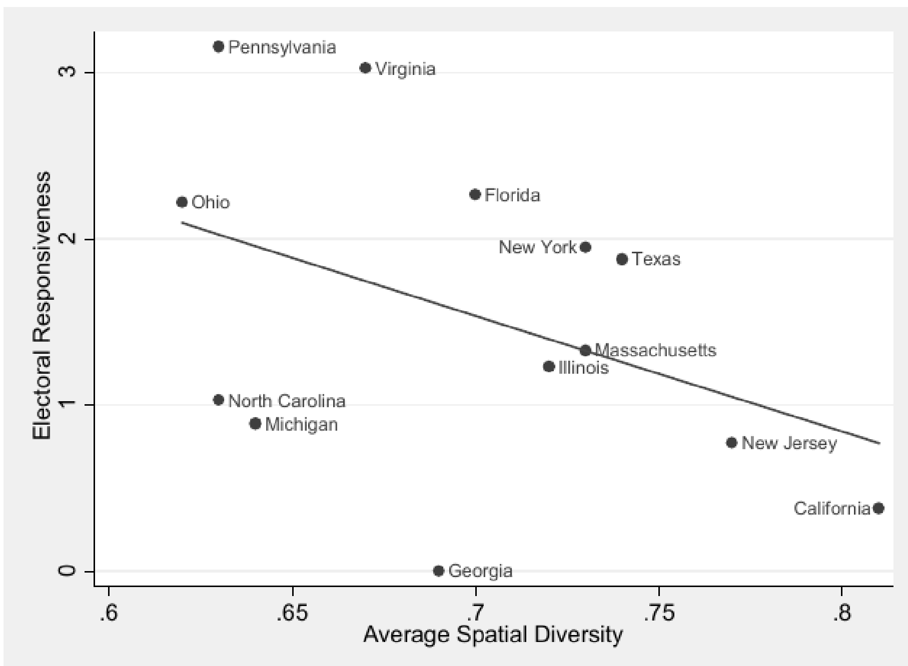


Figure 2: Effect of electoral responsiveness on spatial diversity

loses seats given changes in its statewide vote share. For instance, if Democrats would win ten percent more seats if they received five percent more of the vote, then a plan would have a responsiveness of 2.0. 255 As a general matter, the lower a plan's bias, and the higher its responsiveness, the better the plan is thought to be. 256

255 See Gelman & King, supra note 11, at 544–45 (defining bias and responsiveness). 256 Reducing bias all the way to zero is unproblematic. However, very high rates of responsiveness are undesirable because they result in large changes in seat shares despite only small shifts in vote shares. Fortunately, the responsiveness scores reported here are not high enough to raise such concerns.

Figures 12 and 13 show how states' spatial diversity averages were related to partisan bias and electoral responsiveness in the 2006, 2008, and 2010 elections.260 I include only states with at least ten congres- sional districts (because bias and responsiveness are not very meaning- ful for states with small numbers of seats), 261 and I use the absolute value of bias (because I am interested in the metric's magnitude rather than its orientation). As is evident from the first chart, spatial diversi- ty has a curvilinear relationship with bias. At lower levels of spatial diversity, that is, bias tends to decrease as spatial diversity increases; but at higher levels of spatial diversity, bias and spatial diversity tend to move in tandem. The curve as a whole is clearly U-shaped. 262 This result suggests that states seeking to treat the major parties as equitably as possible should not minimize the average spatial diversity of their districts. Consistent with the relevant literature, high levels of geographic variation are associated with high bias; 263 they both imply

The story with responsiveness is more straightforward. As the se- cond chart illustrates, responsiveness simply tends to decrease as aver- age spatial diversity increases. The states whose districts are most homogeneous, on average, are also the states whose elections are most responsive to changes in public opinion. In contrast, the states whose districts are most heterogeneous are also the ones in which even large swings in voter sentiment have little impact on the parties' seat shares. This finding indicates that while high spatial diversity is not a prereq- uisite for a partisan gerrymander (low spatial diversity can also do the trick), it is indeed an e ective way to protect incumbents of both par- ties from shifting political tides. Advocates of responsive elections, then, may push without hesitation for spatially homogeneous districts to be drawn, since it is these districts that seem most likely (in the ag- gregate) to reflect the public's evolving preferences. 265

Stephanopoulos [2012]

How compactness might affect spatial diversity

Previous work

- people have done how compactness affects competitiveness (schutzman 2020)
- people have used MCMC approach to look at how districts are more or less competitive (daryl's work)

One factor ignored in this analysis, which is critical to theprocess of drawing districts, isrespecting communities-of-interest.Even defining and locating geographically such communities is avery difficult problem, let alone the determination of whether ornot to preserve that group in a single district. We therefore pro-pose our analysis as a framework for discussion about trade-offs inredistricting rather than as a policy recommendation.

In this work, we have demonstrated with a simple model thatdemanding districts be drawn to be as compact as possible anddemanding that they satisfy a notion of partisan symmetry areincompatible, but to different degrees depending on the particularfeatures of the geographic region in question. Since existing propos-als and methodologies for automated and algorithmic redistrictinginvolve finding an approximate solution to an optimization problem,it is important to understand how changing the objective functionof these procedures can affect the outcome. As more jurisdictionsconsider redistricting reforms, they should be cautious about abdi-cating the line drawing process to algorithms which encode valuesdifferent from those of the voters who use the districts to elect theirrepresentatives.

Methods

Key research questions:

1. Do more compact districts have better, equal, or worse spatial diversity scores?
2. Is there an inherent trade-off between compactness and homogeneity?
3. Does spatial diversity give us a normative basis to select one compactness metric over another?

The research procedure:

1. Generate a large and representative subset of plausible districting plans
2. Evaluate compactness and spatial diversity scores on those plans
3.

TODO write about the two most important bits in my methods: choosing a *compactness measure and choosing a MCMC algorithm to generate the large and* representative subset.

Overview of compactness measures

To empirically evaluate a trade-off between compactness and homogeneity, we must first define the metrics by which we evaluate a proposed districting plan over each of these dimensions. Here, I introduce many different compactness measures. I give a brief overview of the different types of measures and explain the pros and cons of each. I present a compactness measure that I develop and explain my decision to use an ensemble of four compactness measures to increase the robustness of my results. [1]

Over a hundred compactness measures have been proposed in the literature. Here, I focus on two main families: *geometric* compactness metrics and *point-wise distance* metrics.

Geometric compactness metrics

Geometric compactness metrics are by far the largest class of compactness measures. They look at some geometric properties of proposed districts. These properties are most often shapes, area or perimeter—although more esoteric measures do exist. Here, I explain the three most popular compactness measures, although other popular compactness measures e.g. Schwartzberg are qualitatively similar.

Polsby-Popper

The Polsby-Popper measure is by far the most popular measure used in the literature. It is the ratio of the area of the district to the area of a circle whose circumference is equal to the perimeter of the district [Polsby and Popper, 1991]. A perfect circle has a Polsby-Popper score of 1.

$$4\pi \times \frac{A}{P^2}$$

Reock

The Reock score is a measure of the ratio of the district to the area of the minimum bounding circle that encloses the district's geometry [Reock, 1961].

$$\frac{Area}{AreaOfMinimumBoundingCircle}$$

[1] I use the phrases "compactness metric" and "compactness measure" interchangably.

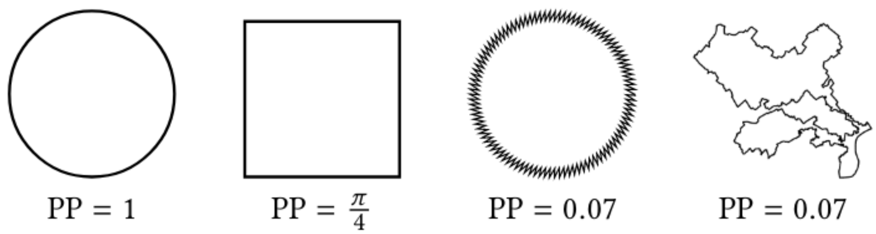


Figure 3: Polsby-Popper scores of four example regions: a perfect circle, a square, a circle with a ragged boundary, an an example district from a Pennsylvania plan. Taken from Schutzman [2020].

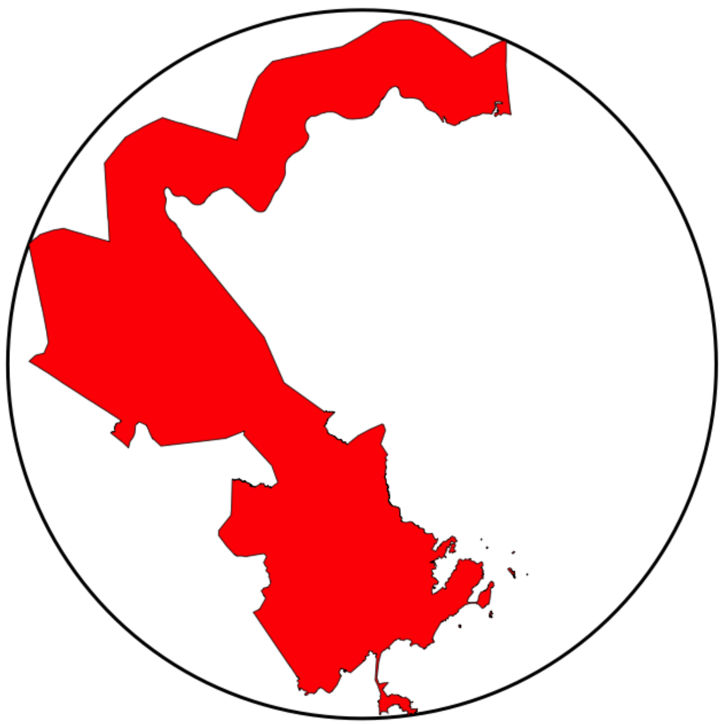


Figure 4: A visualisation of the Reock metric. Taken from https://fisherzachary.github.io/public/r-output.html

Convex Hull

The Convex Hull metric is a ratio of the area of the district to the area of the minimum convex polygon that can enclose the district's geometry. A circle, square, or any other convex polygon has the maximum Convex Hull score of 1.

$$\frac{Area}{AreaOfMinimumConvexPolygon}$$

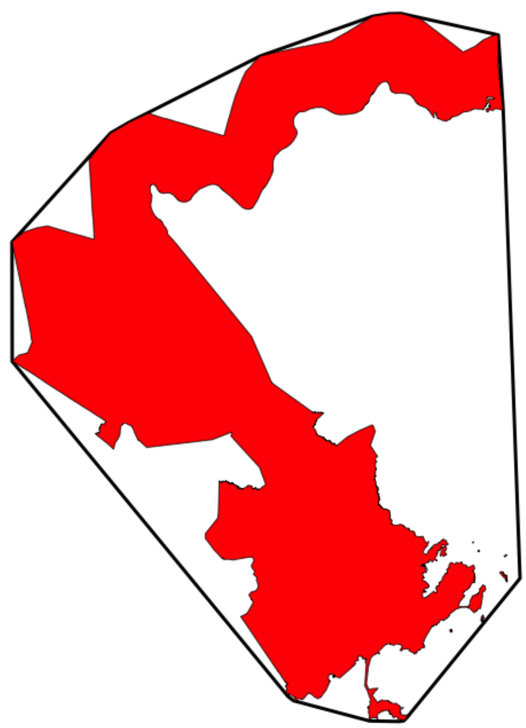


Figure 5: A visualisation of the Convex Hull metric. Taken from https://fisherzachary.github.io/public/r-output.html

Choosing a compactness metric

Which compactness measure should we choose? All three compactness measures are well-cited in the literature and enjoy widespread use. They have been cited in

U.S. Supreme Court cases, *amici* briefs, and redistricting commissions [Moncrief, 2011].

Despite their widespread acceptance, however, the problems with compactness measures are many, and well-covered in the literature. As an example, the most popular compactness measure in the literature—Polsby-Popper—is sensitive to small perturbations in data resolution (the coastline problem).[2] The same is true for other geometric compactness measures: no single metric is perfect. It is therefore important to use an *ensemble* of compactness measures to make sure that one's data and conclusions are robust.

But even this is not enough. Because all three of these compactness measures are purely geometric, they are all vulnerable to a specific family of geographic perturbations. Indeed, Barnes and Solomon [Forthcoming] show that minimally tweaking the geometric features of states is enough for the four most popular compactness measures (Polsby-Popper, Convex Hull, Reock, Schwartzberg) to give very different conclusions on nominally identical data.

Thus, it is important to include a non-geometric compactness measure in the ensemble to guard against the possibility that the results are driven by a specific quirk in geography. Many such measures have been proposed. For instance, Dube and Clark [2016] bring in a discipline of mathematics—graph theory—to formulate a new metric of compactness.

Point-wise distance compactness measures

However, one particular class of metrics I term *point-wise distance* compactness stands out for its ease of understanding (critical if it is to be persuasive to Supreme Court judges), theoretical attractiveness, and academic consensus. Roughly speaking, this class of compactness metrics tries to measure the distance between voters in a district, and assigns higher scores the lower that distance is.

This class of metrics enjoys strong theoretical grounding. Paramount to the idea of single-member districts is that there is some value in voters who live in the same area being put into the same district. Eubank and Rodden [2019]:

> "Voters in the same area are likely to share political interests; voters in the same area are better able to communicate and coordinate with one another; politicians can better maintain connections with voters in the same area; voters in the same area are especially likely to belong to the same social communities — all suggest the importance of voters being located in districts with their geographic peers."

[2] The Polsby-Popper metric measures the ratio of the area of the district to the area of a circle whose circumference is equal to the perimeter of the district. But depending on the resolution of the map, the perimeter can be effectively infinite. Barnes and Solomon find that the choice of resolution has "a substantial impact on compactness scores, with the Polsby-Popper score especially affected."

In contrast, districts that put people with unrelated, faraway others carve voters out of their natural communities and are thus to be avoided. We care about whether co-districtors live in the same area and belong to the same communities of interest, not just the compactness of their electoral district. And point-wise distance metrics deliver exactly that.

Therefore, point-wise distance metrics are more intuitive to laymen and possess a normative bent that more abstract mathematical compactness measures lack. It has therefore been an active area of development in the literature. Chambers and Miller [2010] present a measure of "bizarreness", which is the "expected relative difficulty in traveling between two points within the district". And Fryer Jr and Holden [2011] measures "the distance between voters within the same district relative to the minimum distance achievable".

Despite the relative merits of point-wise distance metrics, there are two areas of improvement—one theoretical, the other empirical. Firstly, all point-wise distance metrics suggested in the literature use Euclidean distances. But many have rightly suggested that we should consider travel times/driving durations instead. For instance, while Fryer Jr and Holden [2011] used Euclidean distance in their metric, they point out its shortcomings:

> Suppose there is a city on a hill. On the West side is [a] mild, long incline toward the rest of the city, which is in a plane. On the East side is a steep cliff, either impassable or with just a narrow, winding road that very few people use. While the next residential center to the East is much closer to the hilltop on a horizontal plane, it is much further on all sorts of distances that we think might matter: transportation time, intensity of social interactions, sets of shared local public goods and common interests, etc. Thus, for all practical purposes, one probably wants to include the hilltop in a Western district rather than an Eastern one. More general notions of distance can handle this.

The "impassable" region on the East would have a short Euclidean distance, and any districting plan that put the hilltop with the Eastern district would be unfairly penalised by these point-wise distance metrics. On the other hand, the impassable region would have a long driving duration, accurately reflecting the political geography. In this and many other cases like it (e.g. large bodies of water), driving durations better reflect a state's unique political geographies.

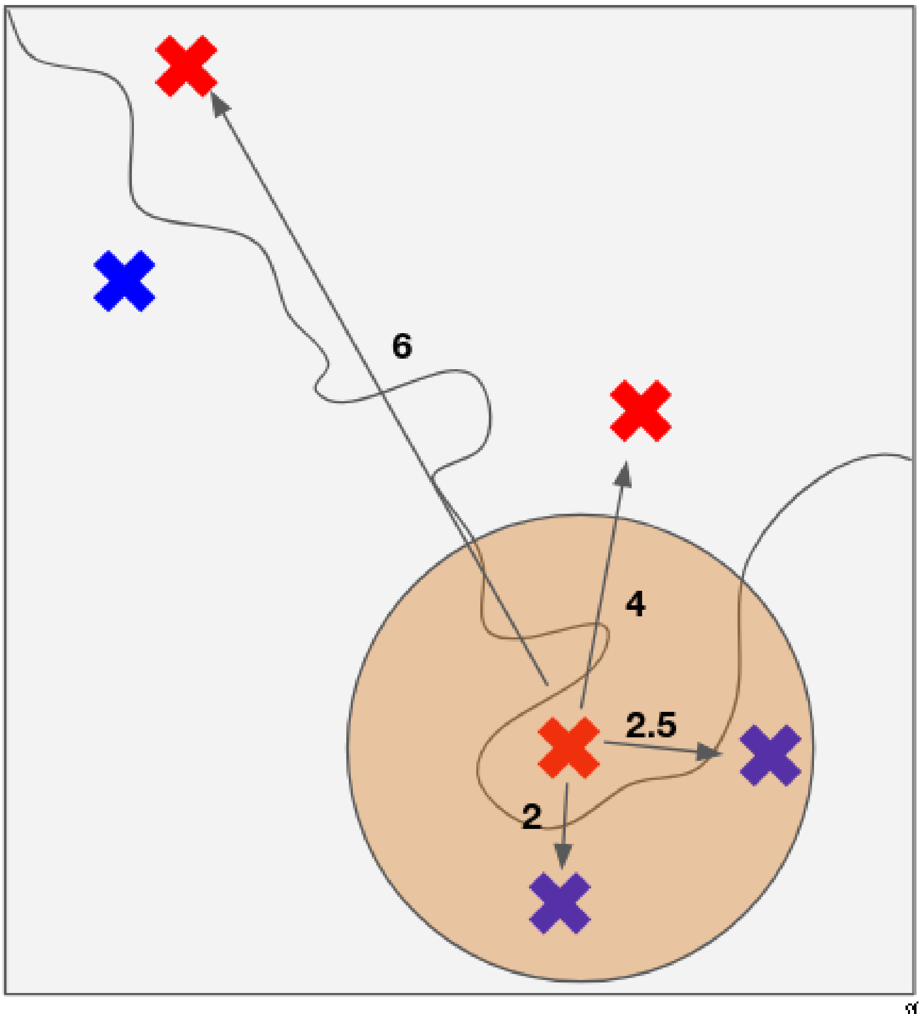
After acknowledging the shortcomings of Euclidean distance, Fryer Jr and Holden specifically suggest using driving durations to improve their metric: "one can extend much of [our analysis] by using driving distance or what legal scholars refer to as 'communities of interest' ".

There are thus strong theoretical grounds for using driving durations in point-wise distance metrics. Why then have scholars not adopted it, seeing as they agree on its superiority? This brings me to my empirical criticism: the point-wise distance metrics scholars have proposed are either far too computationally complex to compute at scale, or have restrictions that make using travel times difficult, if not impossible. For instance, the metric that Fryer Jr and Holden [2011] propose requires solving an *NP-complete* problem. A term used in computer science, an NP-complete problem scales exponentially with the size of the input. This makes it prohibitively expensive on larger states. And while they have an approximation that runs much quicker, they provide no bounds on the correctness of this approximation.

Similarly, Olson has a metric that minimises the average distance from each voter to the center of their district. He says of travel times "that it might be the right kind of thing to measure, but it would take too long. . . The large amount of map data and extra computer time to calculate all those travel times would slow the process down horribly. It would then require a room filling supercomputer to get an answer in a reasonable amount of time." [Olson, 2010]. And finally, Chambers and Miller's measure cannot feasibly be improved with driving durations due to the difficulty of finding point-to-point travel distances without passing through another district. This is because most routing engines allow you only to specify a route between two (or more) points. They do not further allow you to specify regions through which the route cannot pass.

Given the difficulties of adapting existing point-based distance metrics to use driving durations, I develop a new measure called *human compactness*. This metric incorporates driving durations at the very outset, and builds in optimisations to run quickly.

The human compactness metric measures the ratio of driving durations between one's nearest neighbours and one's fellow districtors. This ratio ranges from 0 to 1. The higher this ratio is, the more compact the district.



$$\frac{2 + 2.5}{4 + 6}$$

TODO put a figure here about human compactness

Intuitively, it encourages drawing districts that put one's next-door neighbours together in the same district.

First, I use driving durations rather than Euclidean (as-the-crow-flies) distances between voters. As mentioned, many previous scholars have suggested exactly this, giving it strong theoretical support. It keeps the metric robust to quirks in political geography like mountains and lakes, and better represents the notion of natural communities. The use of driving durations also enjoys empirical support—the use of driving durations seems strictly superior in many cases involving human-scale distances. Working with Nicholas Eubank and Jonathan Rodden, I update their gerrymandering-detection metric to use driving durations instead [Eubank and Rodden, 2019]. We find a consistently different picture of the social context of American suburban voters, raising the possibility of false positives under the Euclidean distance measure [Eubank, Lieu, and Rodden, Forthcoming]. Given that there are strong theoretical and empirical reasons to adopt driving durations, this is a large improvement.

The second improvement I make is algorithmic and computational. My metric improves upon the algorithmic complexity of Fryer Jr and Holden's algorithm from an NP-hard problem to one with a $O(n^2)$ polynomial runtime. This is an exponential decrease in algorithmic complexity, which means the disparity between my metric and Fryer Jr and Holden's increases as the input size grows. I also use programming techniques like precomputation and memoisation to decrease the time taken to compute the metric greatly. My implementation is competitive with geometry-based compactness measures like Reock: on my machine, both metrics took roughly the same amount of time (~0.20s per step). Further details on these algorithmic optimisations can be found in Appendix A.

Given these considerations, I settle on an ensemble of four different compactness measures: Polsby-Popper, Reock, Convex Hull, and Human Compactness. I exclude the Schwartzberg metric as the Schwartzberg and Polsby-Popper measure are largely mathematically equivalent. Finally, I include my point-wise distance metric. This maximises the robustness and validity of my results.

Generating plans with automated districting algorithms

In order to find out whether compactness measures track spatial diversity, we have to generate many counterfactual plausible plans that span the entirety of possible districting plans and measure the correlation between compactness and spatial diversity. This requires using a computer to draw a large number of plans according to some criteria.

The idea of drawing a large number of districting plans with a computer has a long and storied history, starting in the 60s and 70s. The approach has almost always been used to identify gerrymandering; for instance Cirincione et al. [2000] build an algorithm to "quantitatively [assess] whether the [1990 South Carolina] plan is a racial gerrymander". More recently, Chen et al. [2013] "generat[e] a large number of hypothetical alternative districting plans that are blind as to party and race, relying only on criteria of geographic contiguity and compactness." They do this using a Markov Chain simulation algorithm, a procedure that makes iterative changes for a large number of steps until a unique districting plan emerges. At each step of Cirincione et al.'s algorithm, they randomly select a Census Block Group to serve as a "seed" of the district, then randomly add its neighbouring block groups to it until a district with the desired population is formed. Similarly, Chen et al. begin by initialising all precincts as an individual, separate district, then randomly agglomerating neighbouring precincts until the desired number of districts is reached.

While this standard iterative algorithm enjoys a certain degree of success, it has one crippling weakness. The way in which this class of algorithms operates necessarily explores only a tiny subset of all possible districting plans. Subsequent work pointed out this flaw: Magleby and Mosesson wrote that automated processes "may take a biased sample of all possible legislative maps... and fail to e□ciently produce a meaningful distribution of all alternative maps". And Fifield et al. contend that "[standard Monte Carlo algorithms] are unlikely to yield a representative sample of redistricting plans for a target population." [3] This poses a huge issue for the validity of any statistical analysis, because any correlation that we discover on a biased subset of plans may be spurious when measured over the actual distribution of plans. [4]

Markov Chain algorithms

Thankfully, scholars have developed an improvement over the standard algorithm with stronger theoretical guarantees. This second class of algorithms reframe the districting problem as a *graph partition* problem (borrowing insights from graph theory and computer science), and use a *Markov Chain Monte Carlo* (MCMC) approach to sample possible districting plans. This approach is best laid out in Fifield et al. [Working Paper]. Broadly speaking, the approach initialises a specific graph partition as a step in the Markov Chain, then *flips* a random node of the graph to get another valid partition. This process is repeated until the Markov Chain approaches its steady state distribution: when this happens, the Markov chain is called "well-mixed".

This class of algorithms inherit desirable well-known theoretical guarantees of the Markov Chain.[5] They are therefore much less likely (both theoretically and empirically) to generate a biased subset of plans. Conducting a small-scale validation study on a 25-precinct set, Fifield et al. compare the distribution of plans generated by their algorithm to those generated by the standard redistricting algorithm. They prove that their algorithm produces plans that hew much more closely to the *actual* distribution of all possible districting plans.

[3] See Fifield et al. [Working Paper], pg. 16, for a technical explanation of why these algorithms don't produce uniform redistricting plans: "For example . . . , the creation of earlier districts may make it impossible to yield contiguous districts. These algorithms rely on rejection sampling to incorporate constraints, which is an ine□cient strategy. More importantly, the algorithms come with no theoretical result and are not even designed to uniformly sample redistricting plans."

[4] Generating a biased sample is not necessarily a problem if all you want to do is *optimise*, e.g. draw the most compact plan possible. Recent work builds upon this standard algorithm, using Voronoi diagrams or iterative flood fill procedures rather than random chance, to assign the precincts to be agglomerated. See Levin and Friedler [2019] for a technical overview.

[5] See DeFord et al. [2019b] for a technical overview.

Due to the many advantages of the MCMC approach, I use it in all my analyses. However, there are many ways to conduct an MCMC analysis. The key question is how one should sample from the near-infinite pool of possible plans. State-of-the-art literature in this space use one of three main approaches, all of which have their pros and cons.

The first is to get a sense for the properties of extremely compact plans under each compactness measure by using a local optimization technique, starting at a whole bunch of different initial seeds using the single node `Flip` proposal. This approach gives us the most compact plans, and is often used to find the "maximal" or "best" districting plans. However, it will—by design—only explore a very tiny subset of all plausible districting plans. Also, because the `Flip` proposal is very state-dependent, the initial state can affect the results greatly.

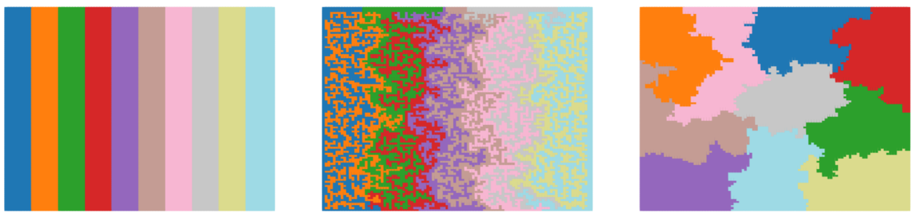
The second is a middle-of-the-road approach, using a global proposal distribution and a Metropolis-Hasting acceptance function to sample from a distribution over plans that is proportional to $e^{(} - \beta \times Compactness)$. This will give us a distribution of plans that is biased towards compact ones, but also contains some noncompact plans. One can get different distributions of plans depending on the specific acceptance (score) function. For instance DeFord and Duchin [2019] prioritises plans that have fewer locality splits and/or sustain a Black majority-minority district, and Herschlag et al. [2018] use a complicated score function that takes into account county splitting, population deviation, compactness and minority representation. Using this approach, I would define four different score functions corresponding to the different compactness measures, and compare the resulting distributions that result from each measure.

Finally, one can sample from a distribution that doesn't incorporate any compactness score at all and extract the plans that achieve a good score under each metric. This approach is used in DeFord et al. [2019a], where they generate a large neutral ensemble of districting plans and then subsequently filter the plans according to increasingly strict vote-band constraints. The advantage of this approach is that it casts the widest net: all plausible districts (subject to the equal population bound) are explored. The disadvantage is that the odds of sampling an 'optimal' district are incredibly low, which makes it suboptimal for algorithms that aim to build the "best" plan.

Choosing the best MCMC approach

I use the third approach in my analysis.

Local optimisation approaches like the `Flip` proposal have one key problem. The "mixing time" of the Markov Chain under the `Flip` proposal—that is, the number of steps it takes for the Markov Chain to be "close enough" to the stationary distribution—is very large. What that means is that the `Flip` proposal tends to generate very uncompact, snakelike districts in the beginning, as can be seen in Figure 6. It will take millions of steps for plans under the `Flip` proposal to reach a satisfactory districting plan. As such, I prefer the Recombination (Recom) distribution by DeFord et al., which uses a spanning tree method to bipartition pairs of adjacent districts at each step [DeFord et al., 2019a]. This proposal distribution improves upon



(a) Initial Partition (b) 1,000,000 Flip steps (c) 100 ReCom steps

Figure 6: The `Recom` proposal generates more realistic plans in much fewer steps. Taken from DeFord et al. [2019b].

Mixing time aside, the extreme compactness of the first approach is in fact something that I want to avoid. I aim to find out if mandating compactness in state constitutions can inadvertently adversely affect democratic representation. But restricting one's analysis to extremely compact plans deals a huge blow to the external validity of any such finding. Redistrictors care about a lot of other considerations apart from compactness, and therefore most definitely do not optimise solely over compactness. State constitutions demand that plans be "reasonably compact", not "maximally compact": it's vanishingly unlikely that those extremely compact plans would resemble the types of plans that would be drawn in real life. As such, *even* if I found that optimally compact plans had greater spatial diversity, this would have very little bearing on redistricting policy. It's far more instructive to see whether the relationship holds in the plans that legislators could be expected to draw.

Given that legislators care a lot about many different considerations, might it be better to try and include these considerations into the score function? This is what the second approach does. While this approach holds strong theoretical merit, I find that this approach introduces too many degrees of freedom. The choice of what factors to include in the score function is contentious: Herschlag et al. use population deviation, Polsby-Popper score, county boundaries and minority deviation. But they could just as easily have included factors such as number of cut edges (proposed in Dube and Clark [2016]), for instance. Even if there is a strong justification for including exactly those factors, there is still significant researcher freedom to operationalise the scores. For instance, Herschlag et al. and DeFord and Duchin both include a population deviation score, but operationalise the metric differently. Furthermore, any score function has to be assigned specific weights— but this assignment is somewhat arbitrary and open to argument. For instance, Herschlag et al. "chose a VRA score function which awards lower scores to districting plans which had one district close to 44.48% African-Americans and a second district close to 36.20% African-Americans", on the basis that the 2016 districting plan which was accepted by the Court had districts with those proportions. But this is incredibly arbitrary. Obviously, just because a particular district was accepted by the Court with those proportions of African-Americans doesn't imply that those exact proportions of African-Americans are optimal.

To be clear, these problems are not insurmountable. If there is a strong theoretical basis for one particular operationalisation over another, then the criticism of researcher fiat largely loses its bite. Furthermore, the results obtained are robust to a variety of perturbations. Herschlag et al. [2018] change the weights and threshold values as a robustness check and find qualitatively similar results. Nonetheless, different results can occur. And if two different operationalisations or factor weights yield qualitatively different results, how would we adjudicate between them? For these reasons, I choose not to use the second approach.

The last approach is one that makes the fewest assumptions. It generates a neutral ensemble and does not favour one plan over another (except for some minimal compactness and population deviation requirements). This approach gives us the largest space of plausible plans, which has a key advantage: it allows the results to be applicable even for districting algorithms that do not use an MCMC approach. This includes not only the regular low-tech way of drawing districts, but also other automated districting algorithms like Magleby and Mosesson [2018] and Levin and Friedler [2019].

Therefore, I elect to use the last, "neutral walk" approach. I use a global `Recom` proposal to generate the states, but accept every proposal subject to minimal population deviation requirements. This gives me a neutral ensemble of 10,000 plans for every state.

Method

1. Generate a neutral ensemble of 10,000 districting plans for every state
2. Calculate spatial diversity and compactness scores for each districting plan
3. Perform data analysis (OLS regressions, difference-in-means test)
4. Generate 100,000 districting plans

I download Census Tract level. These can be downloaded from the United States Census Bureau website.

I use the open-source software library GerryChain to generate the ensembles. Replication code and data are included in the Supplementary Information. I obtain the ReCom Markov chain procedure from one of the co-authors of the DeFord et al. [2019b] paper, and generate 10,000 districting plans for 10 states (Connecticut, Georgia, Idaho, Louisiana, Maine, Maryland, New Hampshire, Rhode Island, Utah, and Wisconsin) for a total of 100,000 plans.

TODO talk about why I chose those states

2. Calculate spatial diversity and compactness scores for each of the 100,000 districting plans

I obtain data on spatial diversity from Professor Nicholas Stephanopoulos. The dataset gives *factor scores* for each Census Tract in the country. A district's spatial diversity score is calculated by the sum of the standard deviation of each factor score, normalised by the proportion of the variance each factor score explains. As an example, consider a district made up of three Census Tracts (A, B, C) and let each Tract have three factor scores (1, 2, 3). Let the proportion of the variance explained by each factor score be 50%, 30% and 20% respectively. Then the total spatial diversity score would be

$$\sigma(A_1, B_1, C_1) \times 0.5 + \sigma(A_2, B_2, C_2) \times 0.3 + \sigma(A_3, B_3, C_3) \times 0.2$$

Because

I obtain voter data from Eubank and Rodden,

In order to do this, I have to

3. Analyse

As a robustness check, I rerun all the analyses

by taking the sum of square roots rather than the arithmetic mean.

This penalises districting plans that have a large di�erence between districts e.g. one very good district and one very bad one.

Results

My key results are as follows:

1. Political geography largely pins down the spatial diversity of each individual district.
   - Small urban districts have high SD, large rural ones have low SD.
   - This is largely a function of the state that district belongs to.
2. What "good" plans look like can be quite di�erent (qualitatively) depending on the choice of compactness measure.
3. Di�erent compactness measures are correlated with one another.
   - The geometric compactness measures agree most with one another, the human compactness measure not as much.
4. OLS regressions and di�erence-in-means tests suggest that only the human compactness measure is negatively correlated with spatial diversity: geometric/dispersion based measures have either no or a positive (bad) e�ect on spatial diversity.

5. A difference-in-means test suggests that the most compact districting plans are indeed less spatially diverse than average.
    - Not sure how to reconcile this with result (3) — will discuss this with you later today.
6. A difference-in-means test suggests that the most compact plans under human compactness are less spatially diverse than the most compact plans under geometric/dispersion based measures.

Overall, the evidence suggests that optimising over compactness will give you less spatially diverse districts, and human compactness will do the best job of it.

Initial analysis

Firstly, what do the generated districts (and plans) look like,

What is the distribution of spatial diversity in each state?

After having obtained all the plans and their corresponding scores, I plot the plans with the best and worst spatial diversity and compactness scores to get an understanding for the types of plans that each metric encourages. This will give us valuable intuition for understanding the subsequent results.

For ease of exposition I show states with only two districts, but the analysis extends to states with any number of districts. (Plots of the other eight states are available in the Supplementary Information). I also use Polsby-Popper to represent the other two dispersion-based compactness metrics as my explanations are similarly applicable to those metrics.

Figure 7 plots the best and worst plans according to several metrics. Let us begin with the middle row (Polsby-Popper), as its interpretation is the most straightforward. The Polsby-Popper (and other dispersion-based) metric penalises districts that are very "snakelike" and prefers districts that have regular shapes like squares or circles. This is clearly reflected in the plot. The best plan has a district with a very regular shape, and the worst plan has a snakelike district that contorts through half the state.

On the top row is human compactness. A good plan under human compactness minimises the total travel times between every member of the district. This encourages small, compact districts that avoid splitting urban centers.

We can see that the top plan under human compactness corresponds well to the actual population density of New Hampshire as seen in Figure 8. The top plan puts the two most populous and urban counties in New Hampshire—Rockingham and Hillsborough—together in the same district. The worst plan under human compactness splits the counties in such a way that one's co-districtors are far away, and one's nearest neighbours are in a separate district.

As expected, the top plan under spatial diversity (bottom row) closely resembles the top plan under human compactness. In relatively homogeneous New Hamp-

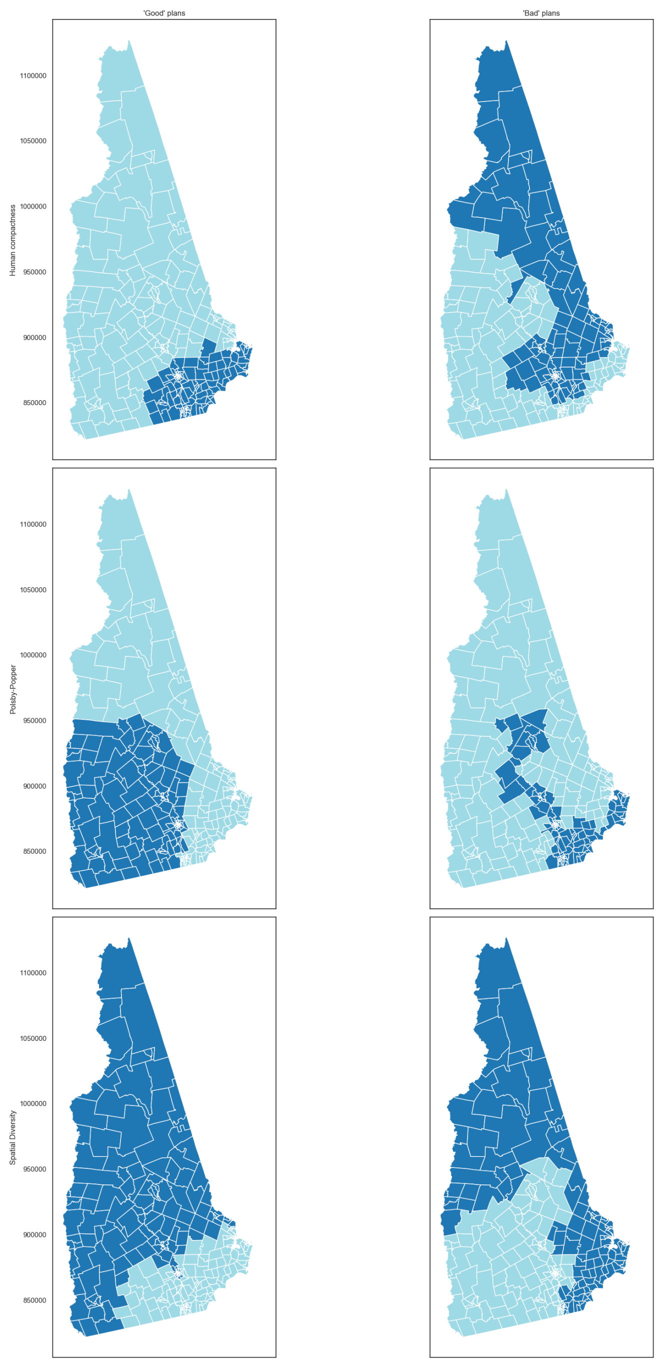


Figure 7: Best and worst districting plans of New Hampshire under different metrics

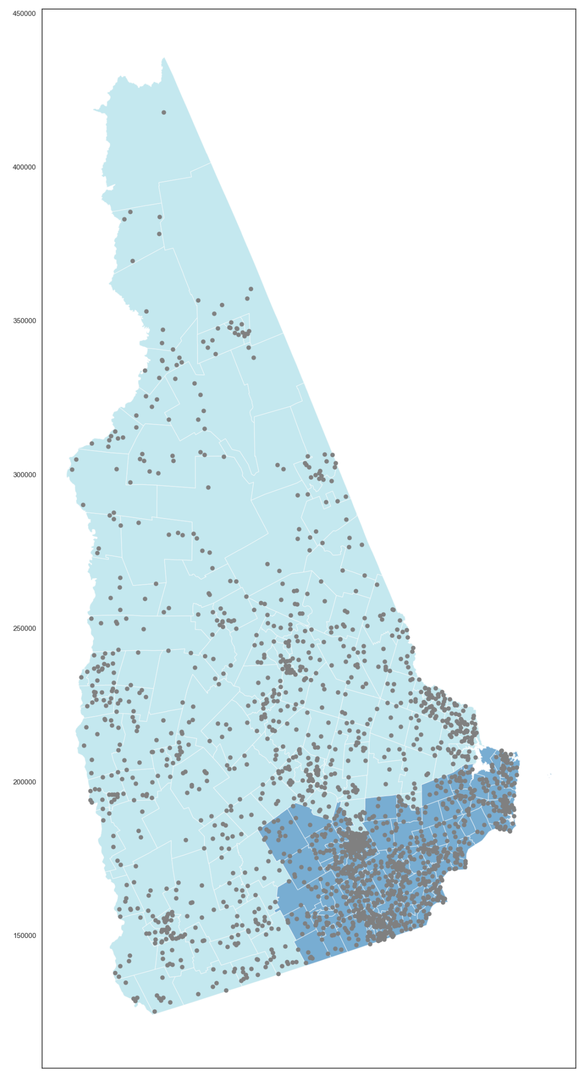


Figure 8: Population density plot of New Hampshire. Each dot represents roughly 600 people.

shire, the main source of spatial diversity is the urban-rural divide. A plan that keeps urbanites together in one district is favoured under spatial diversity.

And while the worst plan under spatial diversity looks different from that under human compactness at first glance, they are actually quite similar. Both plans split up the two populous urban counties, having a "fish-hook" shaped district that starts from the rural north of the state and swoops down to the south to carve out a large part of the counties.

This case study shows that dispersion-based measures may not always reflect existing communities of interest. This seems to fuel criticism of dispersion-based measures on exactly that basis ("it makes no sense to combine areas that have nothing in common except that they fit neatly into a square" [Wolf, 2015]). In this example, human compactness and spatial diversity agree neatly on what the best districting plans should look like.

While human compactness generally tracks spatial diversity better than other compactness metrics (I provide evidence for this later), it does not always do so. Figure 9 gives the population of Idaho. We can see that a large proportion of the population is concentrated in a U-shaped "belt" spanning the southern half of the state. A good plan under spatial diversity will attempt to put this relatively urban "belt" in the same district, and this is indeed what we observe in Figure 10. But due to its great distance and jagged perimeter, such a plan is penalised under both human compactness and dispersion-based measures, both of which prefer a relatively compact square-shaped district.

As we can see, compactness measures need not always agree with spatial diversity. . . .

Political geography largely pins down the spatial diversity of each individual district

Here, I present some observations. Firstly, small urban districts have high spatial diversity, while large rural ones have low spatial diversity. Sec While spatial diversity varies enormously between districts, this is to a large extent dependent on the state's political geography. If

This suggests that while districting we should not expect it to largely affect . . .

This easily extends to the level of the state. Of course, a state's political geography

Figure 11 is a kernel density estimation (KDE) plot of the distribution of spatial diversity in all districts. As in Stephanopoulos's results, the distribution appears log-normal, with a noticeable tail on the right that contains a number of especially heterogeneous districts.

A tempting conclusion to draw from the data is that these districts are equally distributed over the different states. In reality, though, the districts of a state

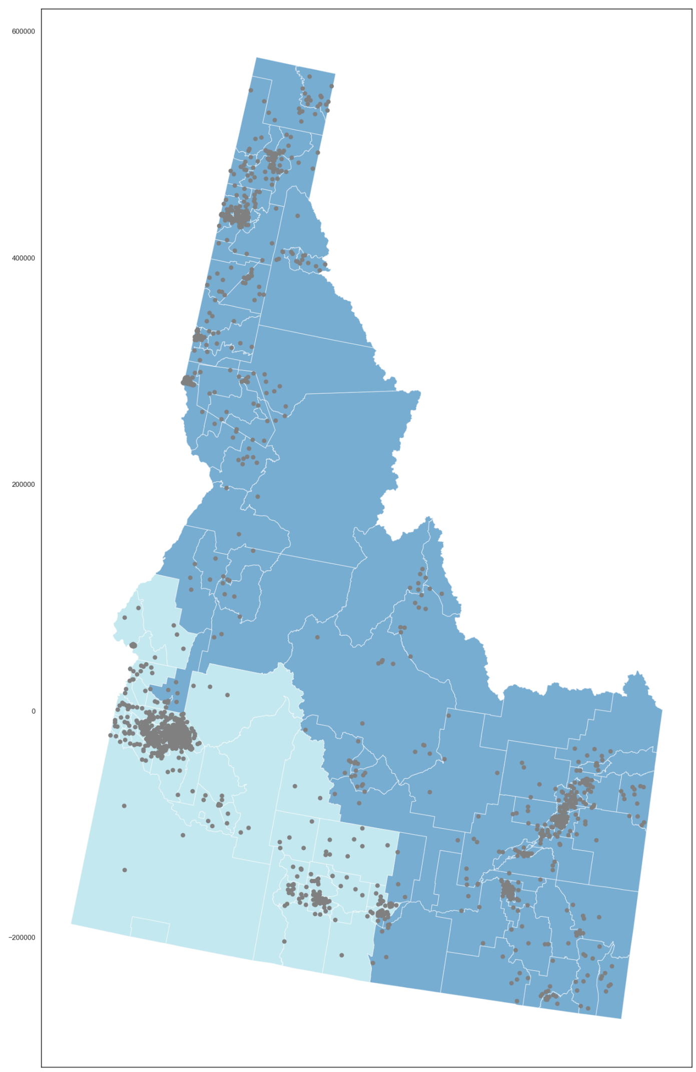


Figure 9: Population density plot of Idaho. Each point represents ~700 people.

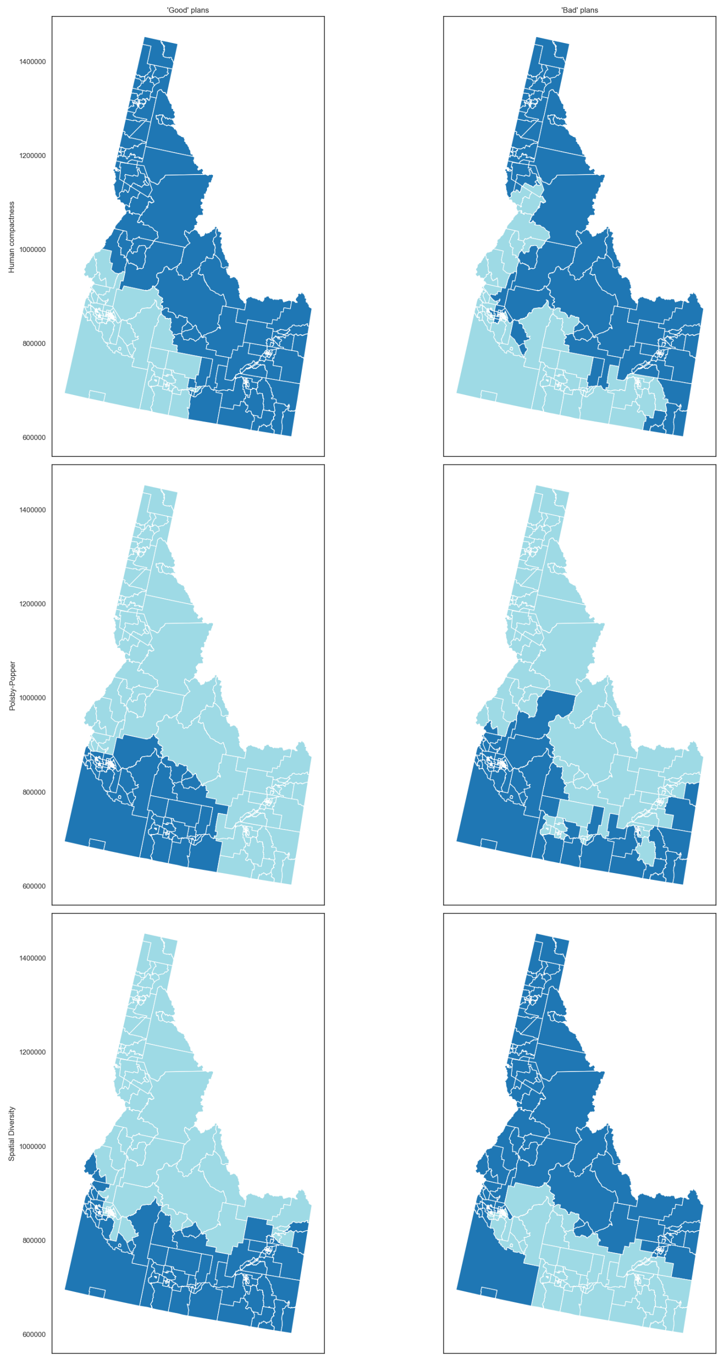


Figure 10: Best and worst districting plans of Idaho under different metrics

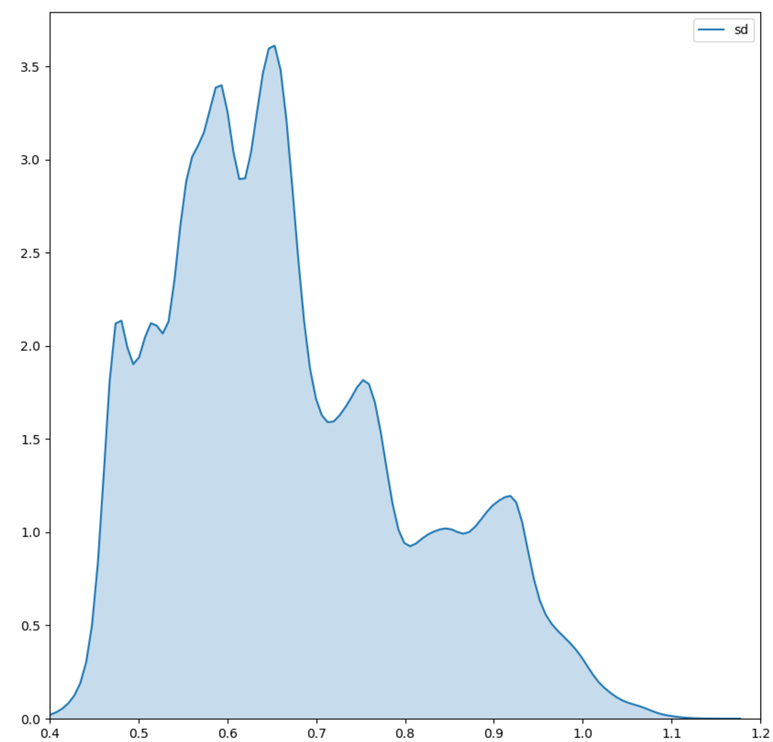


Figure 11: Spatial diversity of all districts

can only take on a small range of values no matter how a districting plan is drawn. Figure 12 demonstrates. The peaks imply a multimodal distribution where individual districts are clustered around certain values and not others. This is most starkly displayed in the states with only two districts. Despite the fact that the redistricting algorithm is continuous, there is a sharp bimodal distribution present in the states of Idaho and Maine, and to a lesser degree Utah and New Hampshire.

This finding is somewhat surprising. It implies that even though the MCMC algorithm explores the entire set of feasible districting plans, any district in any feasible plan will take on a specific form. In other words—no matter how one draws the plan, each district's spatial diversity is largely pinned down by its state's political geography. Some states have very spatially diverse districts, some states have very homogeneous ones, and this is a function of their geography and not the way the districts are drawn.

Given that each district's spatial diversity is largely exogenous, we should expect each state's overall spatial diversity not to vary much as well. Indeed, we see in Figure 13 that each state occupies a narrow band in the range of possible spatial diversity scores. While the range of spatial diversity scores ranges from 0.50 to 0.80, the range of a state's spatial diversity score is only 0.05. While this range is small, it is not insignificant. Figure 2 shows that an increase in a state's spatial diversity by 0.05 is correlated with a decrease in electoral responsiveness by 0.3, about 10% of the variance.

Districts that are small and urban are usually more spatially diverse

Maryland provides the clearest example, although the same pattern repeats in all other states...

This seems rather intuitive, as

Look at Stephanopoulos's factors and the weightages.

Compactness measures largely agree with one another, but human compactness less so

This is the aggregated pairwise Pearson's correlation coe□cient on the aggregated dataset. We can see that PP, Reock and CH largely agree with one another (pairwise correlation ranges from 0.5 to 0.6), but the human compactness metric disagrees.

```
            sd        hc        pp     reock        ch
sd     1.000000 -0.357926 -0.094629 -0.007729 -0.455142
hc    -0.357926  1.000000 -0.409995 -0.111860  0.110439
pp    -0.094629 -0.409995  1.000000  0.480940  0.606623
```

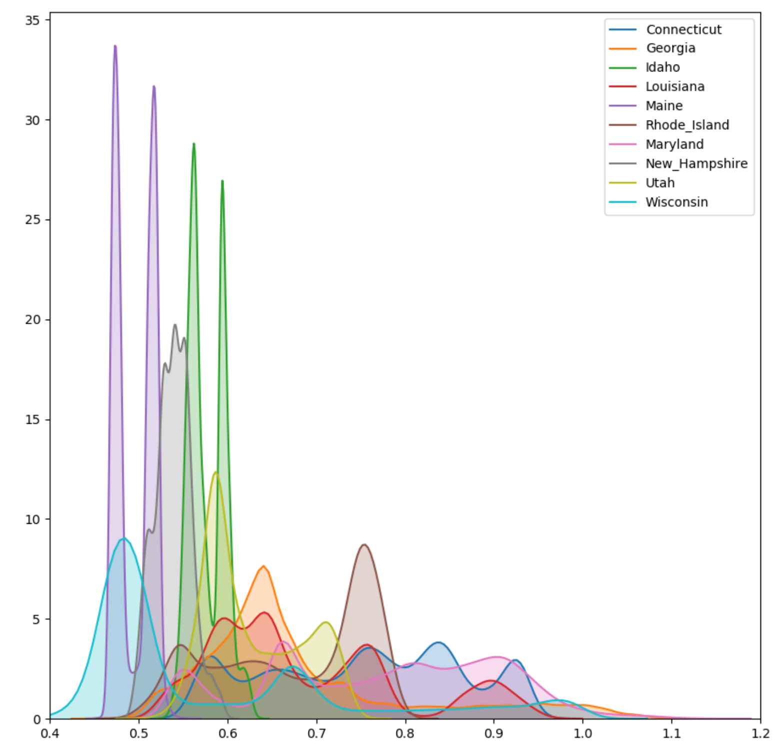


Figure 12: Spatial diversity of districts binned by state

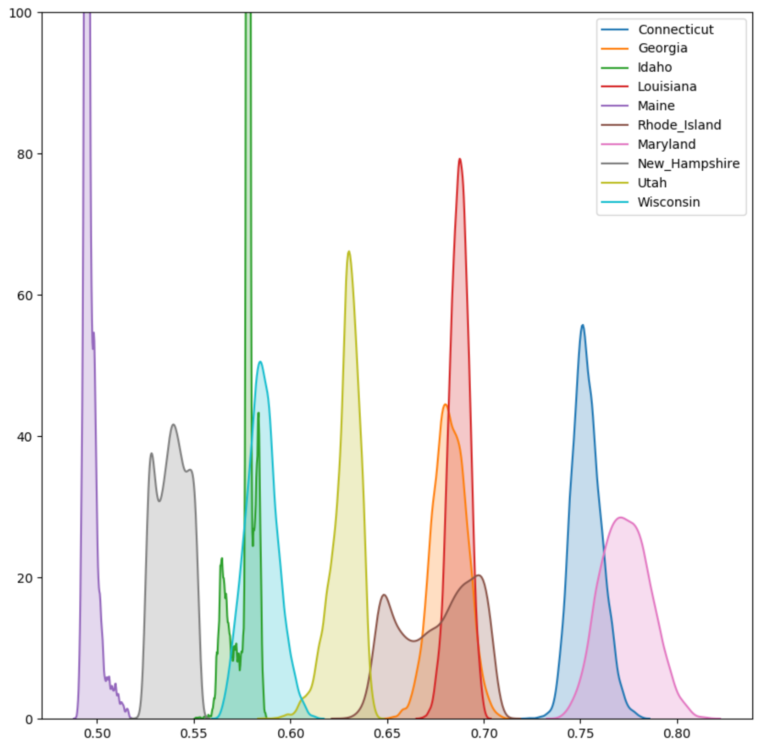


Figure 13: Overall spatial diversity of districting plans by state

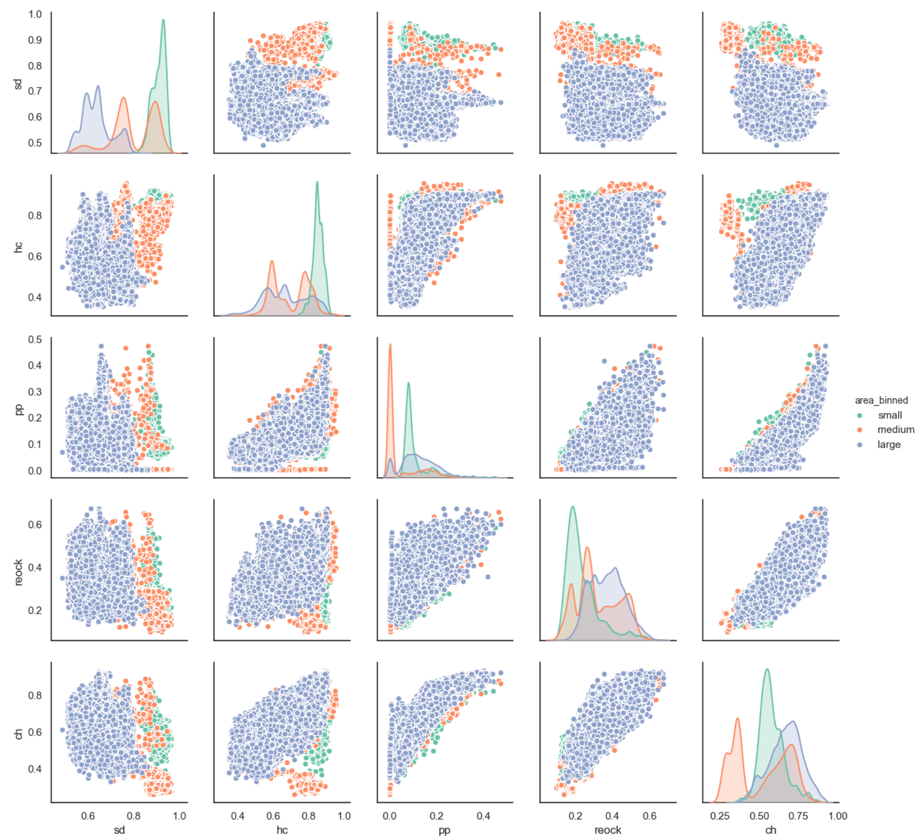


Figure 14: Pairwise plot of Maryland's districts: urban districts have highest SD

```
reock -0.007729 -0.111860  0.480940  1.000000  0.629797
ch    -0.455142  0.110439  0.606623  0.629797  1.000000
```

This is the aggregate analysis though. We should look at the individual states, and the results look better. (Appendix C?)

The human compactness metric

One way to visualise is the

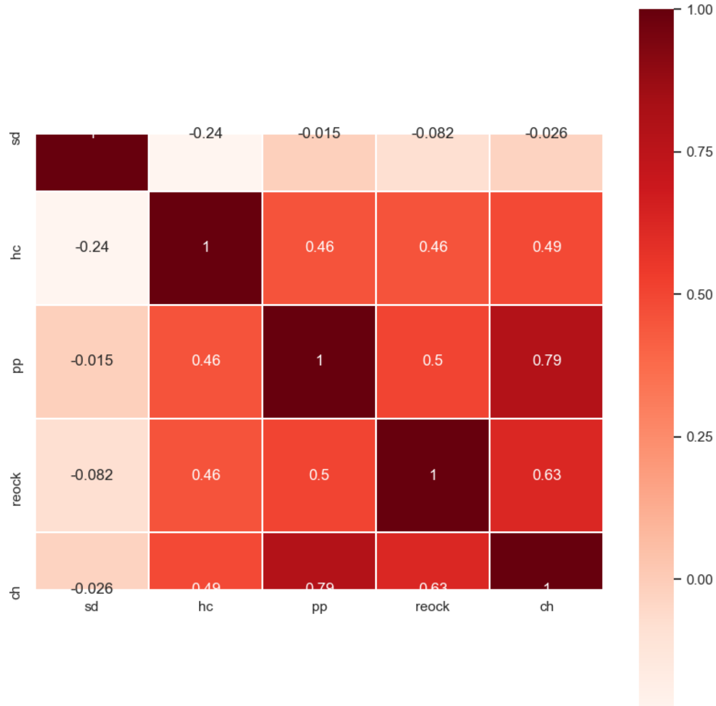


Figure 15: Correlation heatmap of Connecticut

A similar way to visualise the findings...

Figure 17 is a correlation plot between spatial diversity and the various compactness metrics for the state of Connecticut.

This means that the use of an ensemble is quite warranted? How to justify...

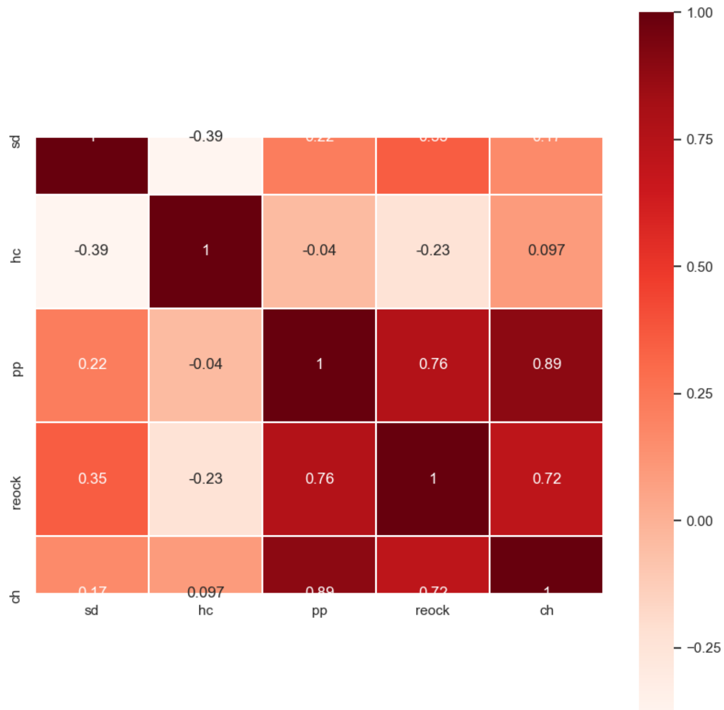


Figure 16: Correlation heatmap of Connecticut

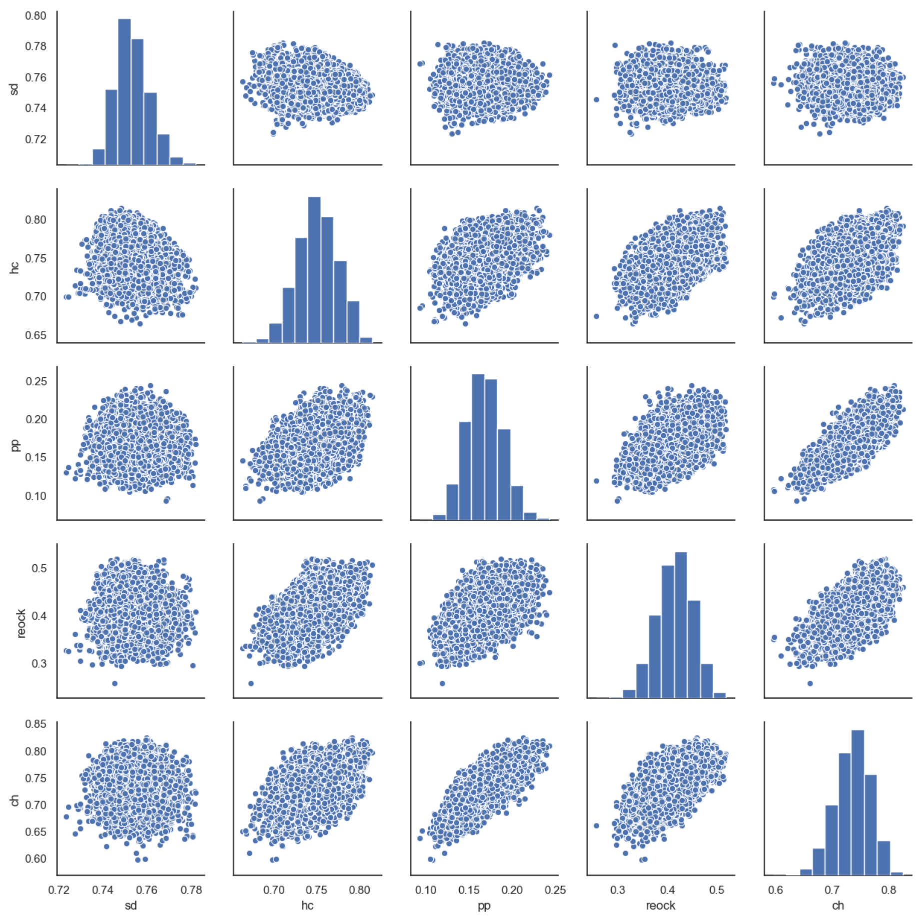


Figure 17: Correlation plot of Connecticut

The overall effect of compactness on spatial diversity is equivocal. Only human compactness has a significant negative effect on spatial diversity.

Multivariate regression with country dummies

We cannot simply run a regression aggregating every single district as each state has a unique distribution of spatial diversity and compactness. Consider the following. Within each state, increasing compactness decreases spatial diversity. But on the aggregate, states with high spatial diversity also have low compactness. In this case, regressing spatial diversity on the aggregate level would give an inflated estimate of the actual effect, falling afoul of the *ecological fallacy*. I illustrate this in figures 18 and 19. In Figure 18, I plot a graph of human compactness on the x-axis and spatial diversity on the y-axis. The overall trend seems to be slightly negative: in most of the groups, there is a slight negative correlation between human compactness and spatial diversity. However, we would obtain erroneous results if we aggregated the different states and ran a singular regression. This is depicted in Figure 19: due to the *between-group* correlation of compactness and spatial diversity, the estimate of the effect is biased. We must therefore control for state when running the regression. Thus, I run a multivariate regression with the functional form

$$SpatialDiversity = \beta_0 + \beta_1 Compactness + \beta_2 State$$

where $State$ is a dummy variable, taking care to avoid the dummy variable trap.

Table 1 shows the results for human compactness. I run the same regression for each compactness metric and obtain the following:

```
HC: -0.0404, t-value -40.632
PP: +0.0251, t-value 29.841
Reock: +0.0209, t-value 27.645
CHull: -0.0016, t-value -1.801
```

I find that only human compactness has a statistically significant negative coefficient on spatial diversity, while Polsby-Popper and Reock have a significant positive effect on spatial diversity. This initial result suggests two things: firstly, and rather disappointingly, that optimising over the two most popular compactness measures may have adverse effects on electoral competitiveness and responsiveness. More encouragingly, though, these effects can be mitigated by the judicious choice of compactness measure. The results show that optimising over Convex Hull does not come at the cost of diversity, and that increasing human compactness actually decreases spatial diversity.

The most compact plans have lower spatial diversity than average

While the results of the overall regression are discouraging, this may not be the last word. The neutral ensemble approach means that the generated plans run

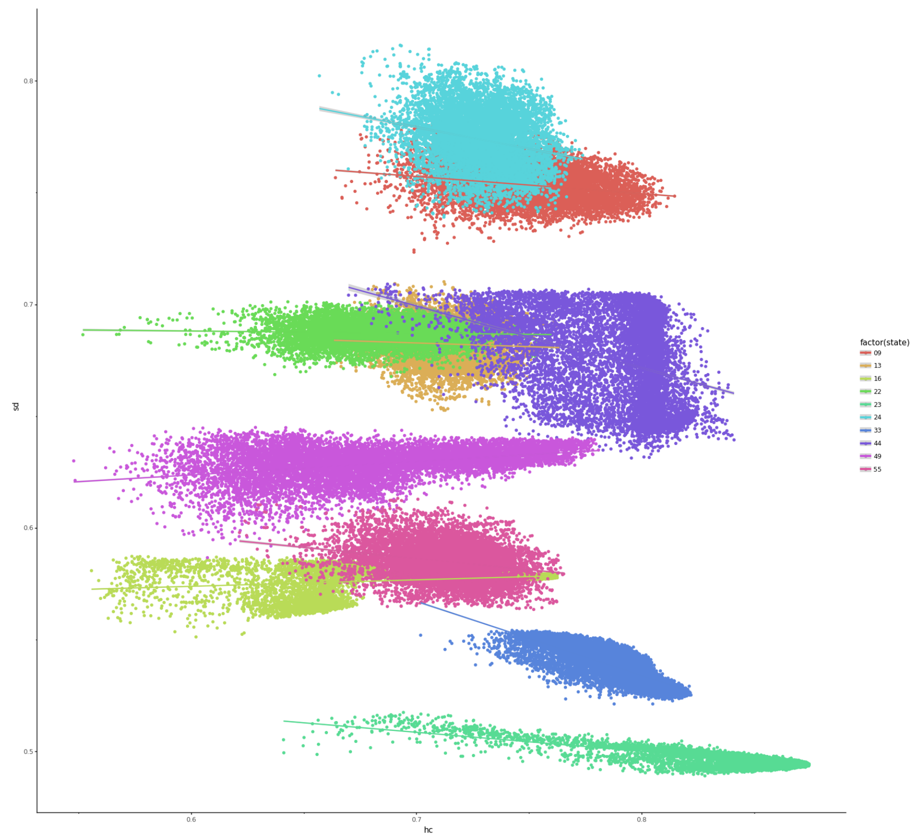


Figure 18: The individual-level regressions show a weak downward trend between human compactness and spatial diversity

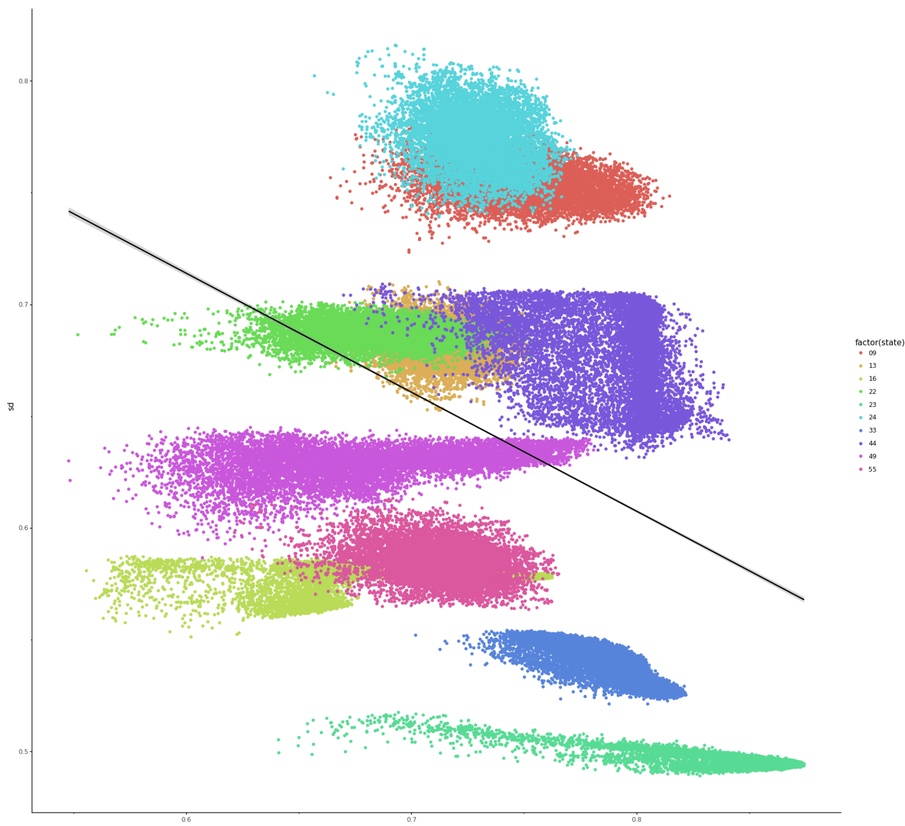


Figure 19: Aggregating the individual states gives an inflated estimate of the effect of compactness and commits the ecological fallacy

Table 1: OLS Regression of Spatial Diversity on Human Compactness with Country Dummies

Dep. Variable:	sd	R-squared:	0.988
Model:	OLS	Adj. R-squared:	0.988
Method:	Least Squares	F-statistic:	8.188e+05
Date:	Wed, 11 Mar 2020	Prob (F-statistic):	0.00
Time:	20:23:45	Log-Likelihood:	3.2365e+05
No. Observations:	100000	AIC:	-6.473e+05
Df Residuals:	99989	BIC:	-6.472e+05
Df Model:	10		

	coef	std err	t	P>|t|	[0.025	0.975]
C(state)[09]	0.7837	0.001	1042.069	0.000	0.782	0.785
C(state)[13]	0.7111	0.001	993.725	0.000	0.710	0.713
C(state)[16]	0.6054	0.001	856.490	0.000	0.604	0.607
C(state)[22]	0.7149	0.001	1039.373	0.000	0.714	0.716
C(state)[23]	0.5303	0.001	626.929	0.000	0.529	0.532
C(state)[24]	0.8030	0.001	1097.735	0.000	0.802	0.804
C(state)[33]	0.5705	0.001	725.232	0.000	0.569	0.572
C(state)[44]	0.7073	0.001	899.177	0.000	0.706	0.709
C(state)[49]	0.6561	0.001	959.927	0.000	0.655	0.657
C(state)[55]	0.6138	0.001	858.803	0.000	0.612	0.615
hc	-0.0404	0.001	-40.632	0.000	-0.042	-0.038

Omnibus:	3979.140	Durbin-Watson:	1.171
Prob(Omnibus):	0.000	Jarque-Bera (JB):	9332.569
Skew:	-0.236	Prob(JB):	0.00
Kurtosis:	4.420	Cond. No.	67.9

the whole gamut of compactness scores, including both highly compact plans and highly noncompact ones in the sample of 100,000. In reality, though, legislators will try to optimise for compactness to some degree. A plan proposed in real life—while not being optimally compact—would be reasonably so. Rather than regressing over the entire sample, then, we should specifically check the spatial diversity of plans which exceed the threshold of “reasonable compactness”.

But what is the threshold of “reasonable compactness”? The choice of the threshold cannot be determined *a priori*. One would have to know the distribution of compactness in a sample of plans generated in real life. Of course, as real-life expert districtors do not produce a distribution of plans, this is also a tall order. I therefore run the same OLS regression for different thresholds of “reasonable compactness”, ranging from the top 90% (excluding the bottom 10%) of plans to the top 10% of plans.[6] The results are as follows:

[6] The results are similar when we take the top 5% or 2% of plans, but the small sample

10th percentile:

	coef	std err	t	P>|t|	[0.025	0.975]
hc	-0.0256	0.001	-19.660	0.000	-0.028	-0.023
pp	0.0339	0.001	34.757	0.000	0.032	0.036
reock	0.0287	0.001	34.723	0.000	0.027	0.030
ch	-0.0001	0.001	-0.143	0.886	-0.002	0.002

25th percentile:

	coef	std err	t	P>|t|	[0.025	0.975]
hc	-0.0274	0.002	-16.018	0.000	-0.031	-0.024
pp	0.0357	0.001	30.861	0.000	0.033	0.038
reock	0.0307	0.001	33.635	0.000	0.029	0.032
ch	-0.0024	0.001	-2.153	0.031	-0.005	-0.000

50th percentile:

	coef	std err	t	P>|t|	[0.025	0.975]
hc	-0.0514	0.003	-15.722	0.000	-0.058	-0.045
pp	0.0154	0.001	10.269	0.000	0.012	0.018
reock	0.0267	0.001	24.400	0.000	0.025	0.029
ch	-0.0119	0.001	-8.627	0.000	-0.015	-0.009

75th percentile:

	coef	std err	t	P>|t|	[0.025	0.975]
hc	-0.1154	0.007	-17.023	0.000	-0.129	-0.102
pp	-0.0338	0.002	-14.814	0.000	-0.038	-0.029
reock	0.0185	0.002	12.136	0.000	0.015	0.021
ch	-0.0589	0.002	-30.961	0.000	-0.063	-0.055

90th percentile:

	coef	std err	t	P>|t|	[0.025	0.975]
hc	-0.0396	0.013	-3.098	0.002	-0.065	-0.015
pp	0.0412	0.005	7.756	0.000	0.031	0.052
reock	0.0074	0.003	2.285	0.022	0.001	0.014
ch	-0.0274	0.004	-6.862	0.000	-0.035	-0.020

The results vary somewhat depending on our choice of threshold, but are on the whole remarkably consistent. The Reock measure performs poorly in all thresholds. The Polsby-Popper metric is not much better either. Only when the threshold is set to the top 25% of plans does the coe□cient go below 0, and the e□ect reverses when we look at the top 10% of plans. I am inclined to believe that is an outlier. The Convex Hull metric is the best of the dispersion-based metrics. It consistently has a negative coe□cient, although the negative coe□cients are very small—particularly when the threshold is low. Finally, the human compactness metric performs well on all subsample. The coe□cient on human compactness is larger than all the other metrics on all the thresholds—a

sizes of those thresholds mean that it is di□cult to get statistical significance.

strong indication that it is the metric that best minimises spatial diversity.

The average spatial diversity of top plans under human compactness is significantly lower than the average spatial diversity of top plans under other compactness metrics

The OLS regressions we run give the relationship between compactness and spatial diversity. But perhaps one is not concerned about the marginal effect of compactness on diversity. One might ask a more basic question: if we mandate that plans are "reasonably compact"—whatever that means—and force legislators to propose only plans that cross a threshold of reasonable compactness, will that adversely affect spatial diversity?

If there is indeed a fundamental trade-off between compactness and spatial diversity, then we should observe the average spatial diversity of highly compact plans to be higher than the spatial diversity across all plans. I therefore compare the mean spatial diversity of top 500 plans under each compactness metric to the mean spatial diversity of all plans. As a robustness check, I look at different proportions (top 10%/5%/2%) and obtain almost-identical results. The results are as follows:

```
Mean SD of plans with highest Human Compactness scores: 0.635558
Mean SD of plans with highest Polsby-Popper scores: 0.640954
Mean SD of plans with highest Reock scores: 0.639897
Mean SD of plans with highest Convex Hull scores: 0.639985
Mean SD of all plans: 0.639758
```

Only the mean spatial diversity of human compactness is significantly lower than the mean spatial diversity of all plans. In order to check the significance of this result, I run a differences-in-means test using Welch's t-test. I use Welch's t-test as Student's t-test relies on a homogeneity in variances assumption. When the assumption of equal variances is not met, Student's t-test yields unreliable results, while Welch's t-test controls Type 1 error rates as expected [Delacre et al., 2017]. In this case, since the top plans come from different distributions, it is unlikely that the variances are homogeneous. The results are as follows:

```
Welch's t-tests for the top 5% of plans

HC vs All: statistic=[-3.36526759]), pvalue=[0.00076992]
Reock vs All: statistic=[0.97597048]), pvalue=[0.32912173]
PP vs All: statistic=[0.11228718]), pvalue=[0.91059979]
CHull vs All: statistic=[0.18211076]), pvalue=[0.85550249]
```

Only human compactness had a statistically significant difference in mean spatial diversity. For completeness, I ran pairwise differences-in-means tests between all four metrics, for a total of 6 tests. The results are as follows:

```
Welch's t-tests for the top 5% of plans (significant results only)
```

```
HC vs PP: statistic=[-3.16361084]), pvalue=[0.00156292]
HC vs Reock: statistic=([-2.53127357]), pvalue=[0.01138011]
HC vs CHull: statistic=[-2.57101923]), pvalue=array([0.0101543]))
```

As expected, there were no significant di□erences in means between any of the geometric compactness metrics, but there was a significant di□erence in the means between human compactness and the other compactness metrics. Similar results obtain when I rerun the tests for the top 10% and top 2% of plans under each compactness metric. The results show that the top plans under human compactness have significantly lower spatial diversity than the top plans under other compactness metrics.

While this analysis is suggestive, there are two rejoinders to this. Firstly, one could argue that the di□erence in means is quite small: only 1.5% of the total variance in spatial diversity. Secondly, one might think that looking only at the aggregated results could be misleading. A di□erence in means in the aggregate could be due to one or a few outlier states driving the results.

To address these two criticisms, I run Welch's t-tests for each metric for all ten states (giving a total of 40 t-tests). The full list of t-tests is available in Appendix B.

Once again, human compactness performs the best. The top plans under the Reock metric have statistically significant negative di□erences in spatial diversity means in 3 out of 10 states. Polsby-Popper and Convex Hull do a little better with 4 out of 10 states. Human Compactness has a whopping seven states. If we look at *meaningful* di□erences—not just statistically significant ones (instances where the mean is lower by more than 5% of the total variance)—then human compactness outperforms by a wide margin. Human compactness has a statistically significant and meaningfully lower spatial diversity in six of the states. Reock does in two states, and Convex Hull and Polsby-Popper only in one. Finally, in two cases (both under the human compactness metric), the di□erence is so meaningful that it makes up 25% and 35% of the total variance. Concretely, the spatial diversity of all 10,000 New Hampshire plans lie within a range of 0.03. The top 1,000 plans under human compactness have a spatial diversity that is 0.01 lower than the mean — a very meaningful e□ect that spans one-third of the total range. Far from being a small e□ect, it seems that the choice of compactness metric to optimise over can have very meaningful impacts.

In order to understand

Table 2 shows what percentile the top 10 percent of plans under each metric would occupy in the distribution of 10,000 plans (lower is better). If there is no relationship between a compactness metric and spatial diversity, then we should expect the mean percentile to lie around 50 percent. If, however, the top plans under a metric are significantly less spatially diverse, then we should see a low percentile for many of the states. In the table, I have **bolded** the best-performing metric in each row, subject to it being less than the median

Table 2: What percentile the top 10 percent of plans under each metric occupy

	hc	pp	reock	ch
0 Connecticut	**34.31**	54.02	55.61	48.25
1 Georgia	48.29	**44.24**	48.34	47.62
2 Idaho	59.92	48.62	**20.90**	26.88
3 Louisiana	**39.03**	39.12	42.45	41.24
4 Maine	26.22	92.48	78.12	**23.56**
5 Rhode Island	**23.32**	56.46	53.71	52.70
6 Maryland	36.99	**33.00**	33.00	48.68
7 New Hampshire	**8.25**	58.08	40.30	65.73
8 Utah	77.05	61.72	58.57	59.92
9 Wisconsin	**34.09**	42.14	47.26	43.07
Mean percentile	**38.75**	52.99	47.83	45.77

(<50th percentile). As before, I run robustness checks and get qualitatively similar results for various threshold cut-offs.

As we can see, the human compactness metric consistently outperforms the other metrics, forestalling the criticism that the results may be driven by outliers. While human compactness does particularly well in New Hampshire and Rhode Island, it still performs best overall even if we remove those two states from consideration.

Discussion and further work

Is there a fundamental trade-off between compactness and communities of interest? ...

In hindsight, it was certainly the right call to use an ensemble of compactness metrics, because of the possibility (and existence!) of outliers. (Elaborate...)

Cast doubt upon work that only uses a singular compactness metric: external validity is very low, vulnerable to outliers (not trying to single out Schutzman's work or anything...)

The low correlation between human compactness and the other compactness measures suggests that human compactness is qualitatively different from geometric/dispersion-based compactness measures.

human compactness better tracks spatial diversity, which in turn correlates with democratic outcomes like responsiveness and competitiveness.

TODO explain why human compactness performs better...

...

The results suggest that optimising for human compactness will tend to give plans

In my work, I analyse 10 out of the 50 states. Restricting analysis to a subset of states is common in other redistricting work, due to the onerous computational burdens of the procedure. DeFord et al. [2019a] measure the effect of competitiveness on partisanship for five states, and Schutzman [2020] looks at the trade-off between compactness and partisan symmetry for only two states. My analysis covers more states than much of the literature, but further work should nonetheless extend the analysis to cover more states—especially large states like Texas, Florida and California.

Acknowledgements

Big thanks to Daryl Deford and Filip

and Stephanopoulos

Am I allowed to thank Andy or name him as my supervisor? Ask Tak Huen about this

Appendix A: Explanation of human compactness metric

Appendix B: Results of difference-in-means tests for individual states

Here I compare the average spatial diversity of all 10,000 plans per state to the average spatial diversity of the 500 most compact plans per state.

I present the results for each state and each metric in the ensemble, using Welch's t-test.

	state	metric	mean_diff	variance	pct_variance	t-stat	p-value
0	0	hc	-0.003460	0.058642	-5.900607	-17.425785	1.366961e-61
1	0	pp	0.000069	0.058642	0.118009	0.288166	7.732681e-01
2	0	reock	0.000381	0.058642	0.650317	1.624462	1.045297e-01
3	0	ch	-0.001042	0.058642	-1.776135	-5.014481	6.033771e-07
4	1	hc	-0.000513	0.057499	-0.892208	-1.868482	6.193359e-02
5	1	pp	-0.001423	0.057499	-2.475505	-4.986335	7.054193e-07
6	1	reock	-0.000498	0.057499	-0.865298	-1.692770	9.076060e-02
7	1	ch	-0.000678	0.057499	-1.178930	-2.231754	2.581874e-02
8	2	hc	0.001489	0.036047	4.131827	26.809567	2.038788e-153
9	2	pp	0.001104	0.036047	3.062205	10.321991	2.820313e-24
10	2	reock	-0.000188	0.036047	-0.520417	-0.859779	3.900941e-01
11	2	ch	0.000383	0.036047	1.063637	2.841225	4.560090e-03
12	3	hc	-0.001257	0.033457	-3.756204	-9.240446	9.523388e-20
13	3	pp	-0.001245	0.033457	-3.720159	-7.632057	4.670461e-14
14	3	reock	-0.000776	0.033457	-2.318159	-5.132091	3.320205e-07
15	3	ch	-0.000927	0.033457	-2.770633	-7.108140	1.896994e-12
16	4	hc	-0.001902	0.028376	-6.704063	-49.155427	0.000000e+00
17	4	pp	0.005131	0.028376	18.081320	38.153281	1.090249e-206
18	4	reock	0.001304	0.028376	4.596054	20.334160	7.714653e-84
19	4	ch	-0.002035	0.028376	-7.171113	-50.341694	0.000000e+00
20	5	hc	-0.019707	0.077819	-25.324736	-43.785027	7.817121e-271
21	5	pp	0.007385	0.077819	9.490310	14.029691	8.033314e-42
22	5	reock	0.005601	0.077819	7.197869	10.059549	5.666063e-23
23	5	ch	0.004848	0.077819	6.229592	8.615116	2.011837e-17
24	6	hc	-0.004913	0.076917	-6.386934	-12.541515	4.676097e-34
25	6	pp	-0.006333	0.076917	-8.233653	-16.445177	3.655560e-55
26	6	reock	-0.006334	0.076917	-8.235342	-17.317992	1.527167e-60
27	6	ch	-0.000795	0.076917	-1.033852	-1.809545	7.061978e-02
28	7	hc	-0.011556	0.032940	-35.083239	-120.004988	0.000000e+00
29	7	pp	0.002150	0.032940	6.527335	9.218455	1.208411e-19
30	7	reock	-0.002165	0.032940	-6.573630	-11.615082	6.541658e-30
31	7	ch	0.004050	0.032940	12.294876	17.193270	1.023553e-59
32	8	hc	0.005538	0.058276	9.503582	42.778404	4.401068e-291
33	8	pp	0.002962	0.058276	5.082034	18.477814	2.578165e-69
34	8	reock	0.002492	0.058276	4.275665	14.864941	8.132217e-47
35	8	ch	0.002689	0.058276	4.613737	16.787183	1.984654e-58
36	9	hc	-0.003290	0.049699	-6.619743	-13.092609	8.687410e-37
37	9	pp	-0.001645	0.049699	-3.309711	-6.053633	1.889349e-09
38	9	reock	-0.000677	0.049699	-1.361577	-2.476624	1.340008e-02
39	9	ch	-0.001482	0.049699	-2.982079	-5.561983	3.278783e-08