Outlier analysis for Pennsylvania congressional redistricting

Moon Duchin

February 2018

The Pennsylvania redistricting plan submitted by Speaker Turzai and President Pro Tem Scarnati is an *extreme outlier* among redistricting plans, according to detailed analysis and rigorous calculations of partisan skew detailed in this report.

This was assessed by a series of tests that were set up and validated independently of the Governor's counter-proposal. I have studied the Governor's proposed map using the same tests and have determined that it behaves squarely in accordance with the traditional districting principles. On the other hand, the Turzai-Scarnati plan is overwhelmingly likely to have been drawn to increase partisan advantage, since traditional districting principles alone do not explain its partisan skew. I produced over three billion maps similar to the Turzai-Scarnati proposal that are at least as compact, preserve at least as many counties, and keep population deviation to within the 1% threshold, so that a mapmaker can tune them to 1-person deviation while maintaining county preservation and compactness. The fraction of maps that were more Republican-skewed in this sample was less than one in 2 million. This means that even with conservative assumptions, there is less than a 0.1% chance that the Turzai-Scarnati plan would have its partisan skew if its authors had no partisan intent.

1 Introduction

I have been asked to use best practices from mathematics and statistics to assess whether a variety of newly proposed redistricting plans for Pennsylvania congressional districts are or are not *extreme outliers* along partisan lines. I have set up this analysis using a method that itself is symmetrical with respect to the two parties, by comparing a proposed plan to a large ensemble of alternatives produced by random changes that only take recognized and traditional districting principles into account. The principles encoded in the random walk are the ones named in the court order: respect for political boundaries, compactness, and population parity.

The method employed here is to run Markov chains to understand whether partisan scores of districting plans exhibit sensitive dependence on unstated priorities used in constructing the plan. Using modifications to two open-source packages (markovchain and redist [2, 4]), our runs characterize the Turzai-Scarnati plan (henceforth TS plan) as a partisan outlier at a very high level of statistical significance. By contrast, the Governor's counter-proposed plan (henceforth GOV plan) is not an outlier in chains initiated there.

I regard this analysis to be as robust as possible given the available data and the timeframe, and I have high confidence in the findings. I found that the TS plan is an extreme outlier under the local $\sqrt{2\epsilon}$ test from [1] at a very high level of statistical significance, while GOV showed no significant effects, i.e., there is no reason to believe that it was drawn to achieve partisan ends.

I will continue to investigate this topic and will extend this analysis with a variety of tests and approaches in the future. At the Court's request, I will gladly furnish further details and analysis.

2 Design and justification

2.1 Question to study

The basic question we are attempting to answer is:

Does a newly-proposed plan represent an extreme outlier among available alternatives?

The approach described here takes seriously the question of available alternatives; we need to control for the effects of the districting rules and for the underlying "political geography" of the state. Both of these factors may cause it to be the case that there is a systemic structural advantage for one party or the other, so it is only legitimate to compare a plan against alternative plans designed according to the same rules and with the same political geography.

2.2 Fixing a voter distribution

When assessing partisan skew, you must pick both a plan and a distribution of voters—the locations where voters live—against which to evaluate it. This is precisely the strength of the algorithmic sampling approach to studying gerrymandering: *the partisan properties of districting plans can only be understood when compared to other plans that hold constant the geography of where voters are located.*

I have studied the TS plan, the currently enacted plan, and the GOV plan with respect to many available election returns and am focusing this analysis on two races for which I believe the answers give most reliable and most easily interpretable results: Senate 2010 (R 51–D 49) and Senate 2016 (R 50.7–D 49.3). These are state-level, statewide races that have two nice features: incumbency effects do not vary across the state, and we don't have to use any interpolation techniques to model uncontested races. (These are both sources of uncertainty when using returns from U.S. Congressional races or State Legislative races.) It is common practice to prefer state-level election results over presidential races for modeling future state-level elections, because presidential races often have voter preference patterns that are quite different.[1] Sen10 has the advantage of having no incumbent in the race, but Sen16 has the advantage of being more recent. I've also considered SenW, which is defined a weighted average of those two (weighted to equalize turnout)—I consider this to be the best and most reliable snapshot of the underlying political geography in Pennsylvania right now.

In other words, we hold constant the distribution of voters—with high concentrations of Democrats in Philadelphia and Pittsburgh and all the rest of the Pennsylvania political geography—and vary only the way the state is cut up into districts. This completely controls for voter distribution effects on any partisan outcomes described in this report.

2.3 Building an ensemble of alternatives

In order to assess the qualities of a proposed plan, we consider its evolution under random transformations. This procedure is called a *Markov chain*, which moves between *states* which represent redistricting plans built out of fixed units, via *transitions* that change the district assignment of a single unit at a time. (Here, the units are voting precincts.[2]) I used two different kinds of Markov chains in conducting this study: a simple random walk and a weighted random walk. In simple random walk, the changes are made through the following process: randomly select a precinct on the boundary between two districts; check whether the new plan is contiguous and has acceptable levels of adherence to traditional districting principles, and move there. In weighted random walk, a penalty score is used for every plan to measure its failure to achieve optimality (perfect compactness, zero splits, and zero population deviation). Now when a random new plan is proposed by the chain, it is accepted according to a probability distribution: definitely accept the new plan if it is better, and accept it with a lower probability derived from its penalty score if it is worse.

The precise formulation of the ways to measure traditional districting principles is described below in an Appendix. We will devote most of the body of this report to the simple random walk implemented in markovchain, but have also explored the space of plans with the Metropolis-Hastings MCMC implementation in redist.

As the chain runs, an *ensemble* is built that accumulates all of the plans encountered by the random walk. This becomes a pool of available alternatives that are comparable to the plan under consideration. I will present data collected by comparing the TS plan to several ensembles of similar alternatives, and will do the same for the current plan and the GOV plan.

[1] Nonetheless I have also considered weighted voter distributions that combine all available election returns, and the results are noisier but not qualitatively different.

[2] These are 2011 Census VTDs, straightforwardly "cleaned" by merging zero-population precincts into their neighbors, by merging precincts when one completely surrounds the other, etc.

2.4 Evaluating partisan performance against the ensemble

We need to select several indicators of partisan performance, given a vote distribution and a districting plan. There are many metrics for partisan skew that can be found in the literature on redistricting. Two of the most popular are the well-established *mean-median score* and the relatively new *efficiency gap*. Each of these measures the amount of advantage enjoyed by one of the political parties. These scores and several others are defined and discussed in an Appendix.

3 Findings

I will give several levels of analysis on the three plans discussed here (TS, Current, and GOV) against three voter geographies (Sen10, Sen16, and SenW).

3.1 General analysis

Given the voter geography recently observed in Pennsylvania, a plan that follows traditional districting principles in a politically neutral way will likely exhibit a tilt toward Republicans relative to the voter proportions. This analysis aims in part to quantify that effect. The full range of possibilities I encountered in trillions of trials against recent Senate vote geography was 4 to 10 seats for Democrats, but the 5-seat outcome is relatively rare and the 4-seat outcome is vanishingly rare.[3] Both the currently enacted plan and the TS plan give 5 seats to Democrats under the Senate 2010 and the averaged Senate distribution and only 4 seats to Democrats under the Senate 2016 distribution. **This immediately suggests that problem with these plans is that they are expressly designed to minimize the Democratic representation.**

3.2 Detailed analysis

The TS plan does improve on the currently enacted plan in terms of several traditional districting principles, especially compactness, but also county and municipality splits. However, it can be seen to be carefully designed to minimize Democratic representation even within those constraints.

In simple seat share, I have algorithmically generated many billions of plans that are similar to the TS plan while improving on compactness. We even find the same result *while keeping intact all of the same counties that are not split by the plan.* To handle population deviation, we note that maps are typically balanced at the end of the production process, and plans with population deviation of 1%, while they would never be enacted into law as-is, are easily *balanceable* by a mapmaker: any such plan can be "zeroed out" (reduced to one-person deviation) without any impact at all on county splits or compactness. All algorithmically generated plans considered in this analysis stay within 1% population deviation, in order to remain easily balanceable. Therefore, the traditional districting principles do not explain the skew in the number of seats obtained by each party.

However, the simple number of seats does not totally capture the partisan dynamics of a plan. An extremely well-established metric that gives a more detailed view of partisan skew is the *mean-median* statistic, which essentially describes how much the party that controls the district lines can fall short of half of the vote while still capturing half of the representation. (Since this analysis is set up to count Democratic seats, a positive mean-median score indicates a systematic advantage for Republicans that is present in the districting plan.)

The figures below (§3.4) show the TS and the currently enacted plan to be wildly extreme in terms of mean-median and efficiency gap scores. Each figure shows a histogram depicting 2^{30}—over 1 billion—steps in a random walk, showing the frequency with which each score occurs. Because there is so much bulk in the histogram, the figures look like smooth curves. The scores of the TS plan and the current plan are so far in the Republican-favoring direction that they appear not to touch the curves at all. By contrast, the GOV plan hits right near the middle of the distribution for the plans with its compactness constraints. In addition, the GOV plan is more slightly compact than TS to begin with. (See §6.2.)

[3] Note that this analysis does not directly address the frequency of districts that are competitive enough to have been winnable by the losing side.

3.3 Rigorous calculations

If a plan is in the most extreme $1/n$ fraction of the plans encountered in a random walk chain with respect to any score, then it has less than $p = \sqrt{2/n}$ probability of being chosen by chance among the other ones that meet the constraints, according to a recent theorem of Chikina-Frieze-Pegden (see Appendix in §5). In this case, the constraints are just adherence to traditional districting principles. Therefore, neither the distribution of voters nor the traditional districting principles can explain the extreme skew.

Recall that Sen10 and Sen16 are the 2010 and 2016 U.S. Senate races, and SenW is the combination of the two (weighted to equalize turnout). Note that $p \leq .05$ is the usual standard for statistical significance, though some prefer the tighter standard of $p \leq .01$. This means that both the TS plan and the currently enacted plan show highly significant levels of partisan gerrymandering, with p values more than 10 times as extreme as the tighter common standard. By contrast, the GOV plan does not meet even the looser standard for statistical significance in this second batch of numbers, and in fact its mean-median scores are slightly to the R-favoring side of its ensemble.

We note that p-values are upper bounds for the probability of an event occurring under the null hypothesis (here, the hypothesis that a districting plan was generated only by traditional districting principles). This means that when you see a p-value greater than 1, you may conclude not only that there is no statistical significance, but that there is *literally no evidence at all* of partisan skew.

When you search a large space by MCMC methods, you will eventually reach a *mixing time* after which you are drawing a representative sample (i.e, you achieve a distribution that is close to the stationary distribution). One reason for seeing stronger results in the second table above is that the search space is more highly constrained so the chains are likely to mix faster. Longer chains would very likely achieve clearer results for the less constrained searches in the first table.

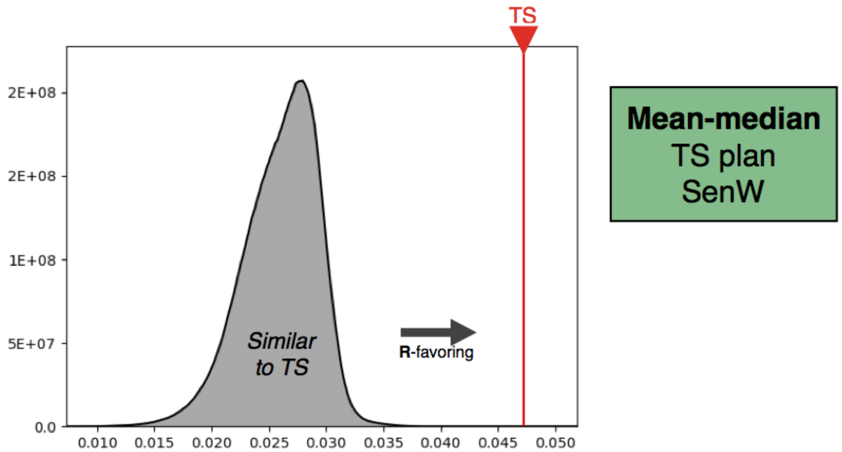
A billion districting plans, at least as compact as initial plan

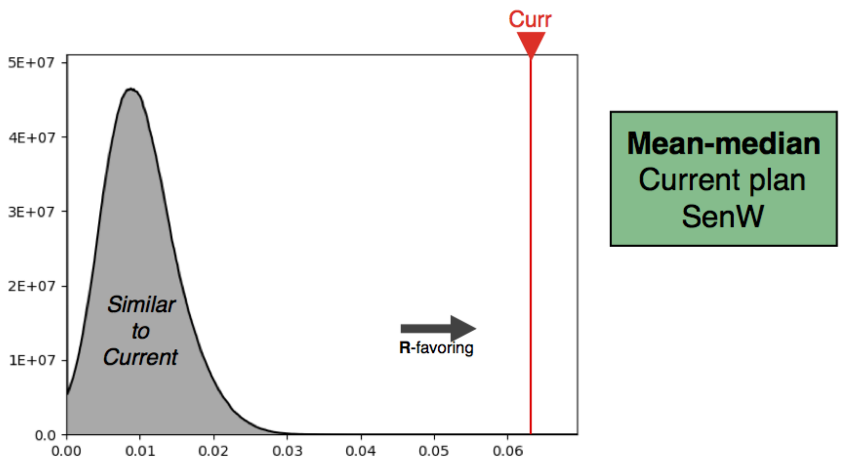
10^{30} steps	D seats	*EG*	frac with higher *EG* than plan	pro-R *p*-value	pro-D *p*-value	bias	*MM*	frac with higher *MM* than plan	pro-R *p*-value	pro-D *p*-value	bias
TS–Sen10	5	.212	.00027	.023	1.41	mild R	4.7%	.00067	.037	1.41	mild R
TS–Sen16	4	.258	.0002	.02	1.41	mild R	4.6%	.013	.16	1.4	—
TS–SenW	5	.21	.047	.31	1.38	—	4.6%	.023	.21	1.4	—
current–Sen10	5	.216	.00027	.023	1.41	mild R	6.2%	.000000006	.00011	1.41	R
current–Sen16	4	.259	.0000042	.0029	1.41	R	4.3%	.015	.17	1.4	—
current–SenW	5	.214	.00021	.02	1.41	mild R	6.2%	.000000021	.0002	1.41	R
GOV–Sen10	6	.149	.92	1.36	.4	—	2.5%	.74	1.2	.73	—
GOV–Sen16	7	.095	.9991	1.41	.042	mild D	3.5%	.15	.54	1.31	—
GOV–SenW	7	.099	.99896	1.41	.046	mild D	3%	.55	1.0	.95	—

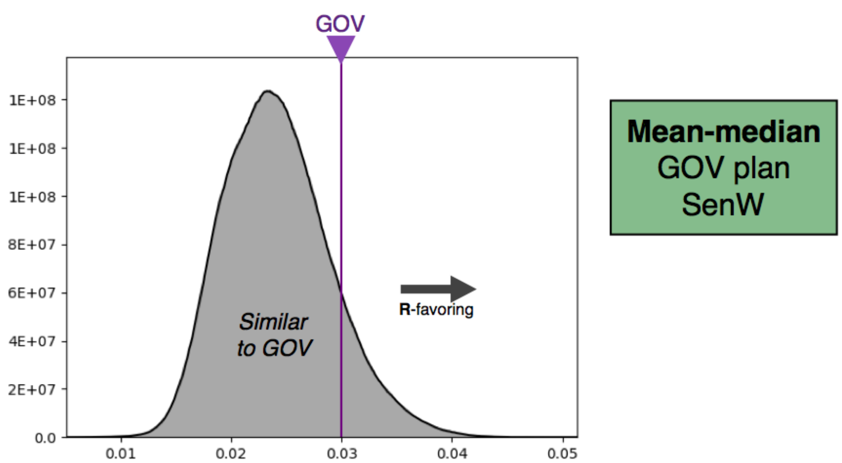
A billion districting plans, at least as compact as initial plan, no more county splits

10^{30} steps	D seats	*EG*	frac with higher *EG* than plan	pro-R *p*-value	pro-D *p*-value	bias	*MM*	frac with higher *MM* than plan	pro-R *p*-value	pro-D *p*-value	bias
TS–Sen10	5	.212	.000048	.0098	1.41	R	4.7%	.0000005	.00099	1.41	R
TS–Sen16	4	.258	.000004	.0089	1.41	R	4.6%	.00000031	.00078	1.41	R
TS–SenW	5	.21	.000043	.0093	1.41	R	4.6%	.0000004	.0009	1.41	R
current–Sen10	5	.216	.0007	.12	1.41	—	6.2%	.000000014	.00017	1.41	R
current–Sen16	4	.259	.000046	.03	1.41	mild R	4.3%	.000049	.0099	1.41	R
current–SenW	5	.214	.00065	.036	1.41	mild R	6.2%	.00000049	.00099	1.41	R
GOV–Sen10	6	.149	.074	.38	1.36	—	2.5%	.065	.36	1.37	—
GOV–Sen16	7	.095	.998	1.41	.063	—	3.5%	.12	.5	1.32	—
GOV–SenW	7	.099	.78	1.25	.66	—	3%	.12	.49	1.33	—

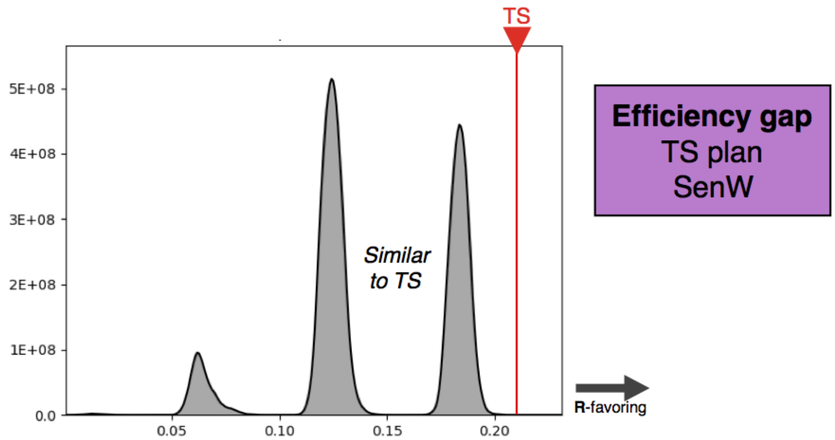
3.4 Figures

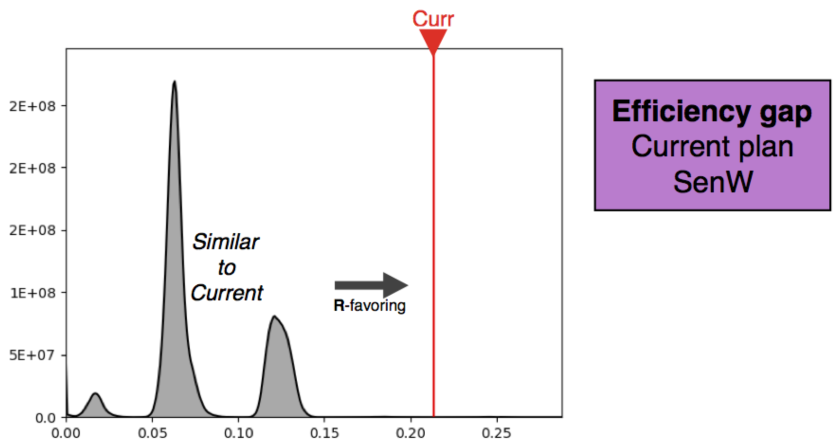


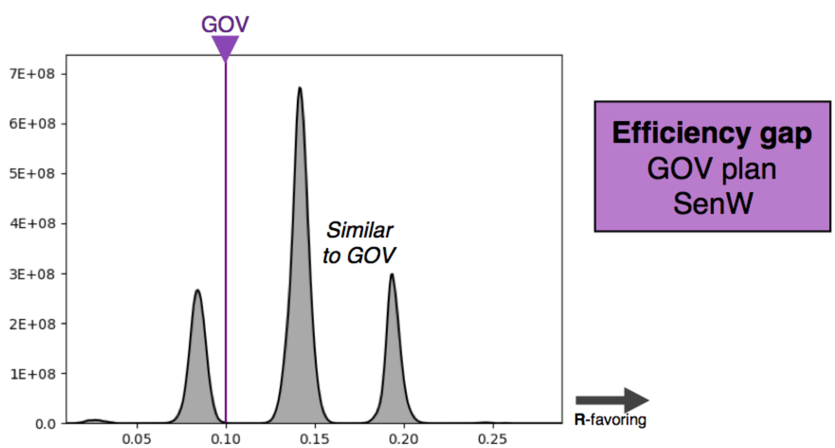




The plots in this group are histograms from the mean-median scores for the TS, currently enacted, and GOV plans, respectively. The positive direction is more systematically favorable to Republicans across a range of vote assumptions. Only the GOV plan falls within reasonable parameters among similar maps.







The plots in this group show histograms of efficiency gap scores for the three ensembles. We see that all three plans pass up on options that are otherwise similar, but would be more favorable to Democrats.

Note that the *EG* values divided up into individual bell curves that shift as the number of Democratic seats varies. This is because of the close relationship of *EG* to seat share (see §7.2). The TS plan is visibly extreme, both for minimizing the number of Democratic seats and even compared to other plans with the same number of Democratic seats awarded.

These pictures all have all of the cited districting principles turned ON, and each plot has over a billion maps in it. Images of this kind for all calculations discussed here are available to the court upon request.

4 Conclusion

The TS plan is shown to be an extreme outlier in the partisan advantage afforded to the Republican party. This is true even when it is compared only to plans that closely resemble it which were found by a neutral algorithmic search encoding the stated principles set out by the Court.

The GOV plan, by contrast, falls squarely within the ensemble of similar plans created using nonpartisan criteria, and therefore gives no reason at all to believe that it was drawn with Democratic-favoring partisan intent.

5 Appendix: Rigorous bounds for statistical significance

We appeal above to the theorem of Chikina-Frieze-Pegden which assesses the likelihood that a given plan appears to be an extreme outlier by chance rather than by careful design. This can be applied to either the simple random walk, whose stationary distribution is uniform, or to the Metropolis-Hastings algorithm, which has the Gibbs distribution (more heavily weighting plans that better conform to traditional districting principles) as a stationary distribution.

The simple random walk we employ here is heavily constrained by traditional districting principles at the level of the starting map, so in effect the Markov chain is searching a much smaller state space that is a single connected component of a disconnected space, making it likelier to achieve mixing. The weighted random walk is preferentially seeking plans that more closely adhere to traditional districting principles. In either case, the null hypothesis ($X_0 \sim \pi$) is that the plan was chosen by the stated principles laid out by the Court. For such a plan to be an ϵ-outlier after k steps of the chain could occur with probability at most $\sqrt{2\epsilon}$.

Theorem 1 ([1]) *Let $M = X_0, X_1, \ldots$ be a reversible Markov chain with a stationary distribution π on its state space Ω, and consider a labeling function $G : \Omega \to \mathbb{R}$. If $X_0 \sim \pi$, then for any fixed k, the probability that $G(X_0)$ is an ϵ-outlier from among the list of values observed in the trajectory $X_0, X_1, X_2, \ldots, X_k$ is at most $\sqrt{2\epsilon}$.*

In statistical science, results are often reported with a p-value which indicates the fit of the observed data with the null hypothesis. A frequent standard for journal publication is to have a p-value below .05, which has traditionally represented adequate statistical significance to reject the null hypothesis. Note that $\epsilon = .00125$ gives $p = \sqrt{2\epsilon} = .05$, so to meet that standard of significance we would need an assessed map to fall in the worst one-eighth of a percent of the values encountered in a chain. The results documented above are much stronger than that.

6 Appendix: Quantifying traditional districting principles

The Court has asked for a plan that reports on splits of political boundaries; population parity; and compactness. In this appendix I discuss metrics that measure adherence to these traditional districting principles.

6.1 Splitting

6.1.1 How much do the districts split the counties?

Suppose the 67 counties of Pennsylvania are labeled $\mathcal{C} = \{C_1, \ldots, C_{67}\}$. Let w_j be the weight of county j, defined as the population of C_j divided by the population of the state, and let $p_i^{C_j}$ be the population of $D_i \cap C_j$ over the population of C_j (that is, the fraction of county j that is contained in district i). Then we define $SqEnt(\mathcal{D}|C_j) = \sum_i \sqrt{p_i^{(C_j)}}$ and $SqEnt(\mathcal{D}|\mathcal{C}) = \sum_j w_j \sum_i \sqrt{p_i^{(C_j)}}$.

This is a modification of the classical Shannon entropy which measures how much two different partitions cut each other into pieces; if for a function f you consider

$$Ent_f(D|C) = \sum_j \left[w_j \sum_i p_i^{C_j} \cdot f\left(1/p_i^{(C_j)}\right) \right],$$

then Shannon entropy uses $f(x) = \ln x$ and ours uses $f(x) = \sqrt{x}$. The reason to use square roots instead of logs is that we want to substantially penalize small "nibbles" that cut off the corner of a county, whereas Shannon entropy considers a 99–1 split to be negligibly worse than an intact county.

To illustrate how this works, consider the following choices of how to split county j.

	A	B	C	D	E	F	G	H
Splitting	97–3	88–12	50–50	96–2–2	50–25–25	33.3–33.3–33.3	25–25–25–25	25–20–1–1–...–1
Score	1.16	1.28	1.41	1.26	1.7	1.73	2	6.45

Note that these scores behave well under refinement: if one piece is broken down into two or more parts while leaving the other pieces alone (such as in moving from C to E to G above) then the score always goes up.

6.1.2 How much do the counties split the districts?

Typically, districts are much bigger than counties, so we try to keep small and medium-sized counties intact in a good districting plan. However, in the case of large counties (Philadelphia, Montgomery, Allegheny), the counties are larger than the ideal district size. In this case, we should try to keep the districts intact within the counties, to enact respect for political boundaries. The score $SqEnt(\mathcal{C}|\mathcal{D})$ looks at each district in turn and measures how much the districts are cut up by the counties, as opposed to $SqEnt(\mathcal{D}|\mathcal{C})$ which looks at each county and scores how much it is cut up by the districts.

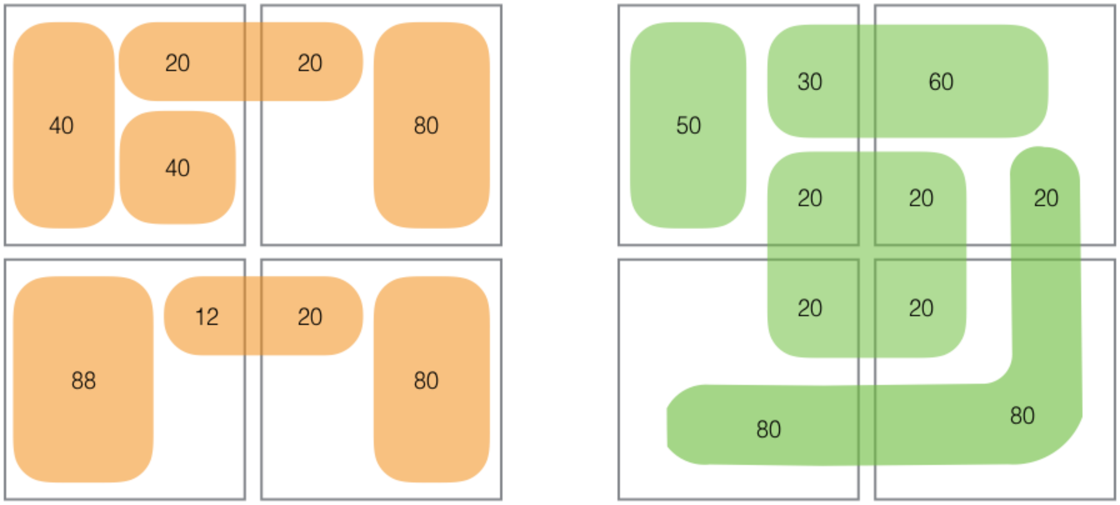
In order to take both of these into account, our overall splitting penalty will be

$$\mathsf{Split}(\mathcal{D}) = SqEnt(\mathcal{D}|\mathcal{C}) + SqEnt(\mathcal{C}|\mathcal{D}).$$

The definition is precisely the same for municipality splits rather than county splits, replacing $\mathcal{C}$ with $\mathcal{M}$.

6.1.3 Example

The images depict two districting plans, where the four squares are districts, the colored regions are counties, and the numbers are populations.



In the first plan (with counties in orange), $SqEnt(\mathcal{C}|\mathcal{D}) = 5.67$, while the second (with counties in green) has $SqEnt(\mathcal{C}|\mathcal{D}) = 6.05$. That means the first one is a little bit better at keeping districts from being overly cut up.

On the other hand, the first plan has $SqEnt(\mathcal{D}|\mathcal{C}) = 1.008$ while the second has $SqEnt(\mathcal{D}|\mathcal{C}) = 2.065$ That means that the first one does a significantly better job of keeping large counties from being chopped up too badly.

Taken together, our penalties read $\mathsf{Split}(\mathcal{D}) = 6.68$ for the first plan and $\mathsf{Split}(\mathcal{D}) = 8.12$ for the second.

In the weighted random walk, this means that a proposed transition from green to orange would automatically be accepted by a Metropolis algorithm, while a transition in the other direction would occur with lower probability.

in the simple random walk, we can simply constrain the splitting penalty at the level of a plan to be studied in order that it is compared to an ensemble with at least as good of a splitting score.

6.1.4 Communities of interest and voting rights

Our algorithmic treatment of the problem can activate a feature that labels identified communities of interest as *geoclusters* and treats them like counties within counties. This has the effect that the algorithm can either enact a light preference for steps that do not create a split within these zones (in weighted random walk), or can require that they be kept intact (in simple random walk).

For instance, the city of Philadelphia has two long-recognized historical Black neighborhoods, generally known as West Philadelphia and Southwest Philadelphia, each with several hundred thousand people. We can designate precincts covering those areas to be geoclusters in our algorithms. By freezing them intact, we are able to study districts that are built around these cores. I regard this as substantially more flexible and more responsive to the language of the Voting Rights Act of 1965 than previous algorithmic alternatives, which either constrain minority percentages at previously observed levels or freeze majority-minority districts wholesale. Our algorithmic methods are able to consider any of these alternatives, however, and to confirm that none of these explains the partisan skew of the TS plan.

6.2 Compactness

The most-cited compactness score in the redistricting literature and in expert testimony is the **Polsby-Popper** score $4\pi A/P^2$, which compares area to perimeter. The Isoperimetric Theorem guarantees that this quantity varies from 0 to 1 (for all measurable shapes with rectifiable boundaries).

Our primary constraint on compactness in the algorithmic treatment is derived from what mathematicians would call an L^{-1} average Polsby-Popper score: we average the reciprocals of the PP scores of the 18 districts. The reason to average reciprocals instead of the straight scores is to attach a heavier penalty to plans with one extremely low score among the districts. (This averaging is the sense in which I assert that the GOV plan is slightly more compact than the TS plan.)

The **Schwartzberg** score is similar. It is commonly defined as *the ratio of the perimeter of a district to the perimeter of a circle with the same area as the district*, which works out to $\frac{1}{2\sqrt{\pi}} \cdot \frac{P}{\sqrt{A}}$. As such, it is exactly equal to the Polsby-Popper score raised to the $-\frac{1}{2}$ power. That means that the way it ranks districts is completely redundant with the Polsby-Popper score—because exponentiation is order-preserving, *they rank districts exactly the same way.* This is sometimes obscured by the fact that Maptitude, the industry-leading software for redistricting, does not use this formulation to report its Schwartzberg scores. Instead of perimeter, Maptitude uses a notion of "gross perimeter" that was proposed by Schwartzberg himself in the 1960s, when computers were not yet able to report perimeters reliably.

We have additionally defined a discrete compactness score as follows: $\mathsf{Cpct}(D_i) = Pop/BPop^2$, where Pop is the population of the district and $BPop$ is the population of the precincts of the district that are on the boundary with other districts or on the edge of the state. This is a population-based version of the classical Polsby-Popper and Schwartzberg scores, which both compare the area A of a district to the perimeter P of a district via comparison of A to P^2, and it is available as a feature in our algorithms.

6.3 Population parity

If Pop_i is the population of district i and I is the ideal district population (i.e., the population of the state divided by 18), then there are several reasonable ways to evaluate the deviation from population parity. For the simple random walk trials I constrained population deviation to 1%, meaning that the only maps considered were those that satisfied

$$.99I \leq Pop_i \leq 1.01I \quad \forall i.$$

When creating an associated penalty, I considered the population deviation score

$$\mathsf{PopDev}(\mathcal{D}) = \sqrt{\left(\frac{Pop_1 - I}{Pop_1}\right)^2 + \cdots + \left(\frac{Pop_{18} - I}{Pop_{18}}\right)^2},$$

which is just the (L^2) distance of the populations from ideal. An L^∞ distance, taking into account only the worst deviation from ideal, is a reasonable alternative.

6.4 Combinations and tuning

For defining an energy to use in a Metropolis-Hastings search, an overall penalty score of a districting plan $\mathcal{D}$ can be defined as a linear combination of a county-splitting penalty, a compactness penalty, and a population deviation penalty:

$$\alpha \cdot \mathsf{Split}(\mathcal{D}) + \beta \cdot \frac{1}{\mathsf{Cpct}(\mathcal{D})} + \gamma \cdot \mathsf{PopDev}(\mathcal{D}),$$

The weights α, β, γ are arrived at by a tuning protocol: initial runs are made with fixed values for those parameters, compared against runs with other relative weights, until a steady level with a high acceptance ratio is found. This is a standard protocol for tuning parameters in MCMC. We used this combined penalty to explore the space of possible districting plans with the redist package developed by Fifield et al. [3, 4]

7 Appendix: Quantifying partisan skew

There are many metrics for partisan skew that can be found in the literature on redistricting. Two of the most popular are the well-established *mean-median score* and the relatively new *efficiency gap*. Each of these measures the amount of advantage enjoyed by one of the political parties.

To compute each of these scores, we fix a districting plan $\mathcal{D}$ and a geographic distribution of votes Δ. (For instance, $\mathcal{D}$ = TS plan, and Δ = Sen16 vote distribution.) Writing V_i^D for the Democratic vote total in district i (and likewise V_i^R for Republicans), we then have $V_i = V_i^D + V_i^R$ for the total major party vote. Let $X_i = V_i^D/V_i$, which is the Democratic percentage of the head-to-head vote in district i.

Then the number of seats awarded to Democrats in plan $\mathcal{D}$ and voting pattern Δ is simply $\#\{i : X_i > \frac{1}{2}\}$.

7.1 Mean-Median score

This is just the mean (average) of the $\{X_i\}$ minus the median (50th percentile value). The interpretation is this: the mean Democratic vote share over districts is a proxy for the statewide Democratic vote share. The median is a vote share X for which half of the districts have $X_i \le X$ and half have $X_i \ge X$. A gap of m means that Republicans could earn half of the representation with $\frac{1}{2} - m$ of the statewide vote share, under the assumption of uniform partisan swing (i.e., we add or subtract the same share to each district to assess performance under voting swings). For instance, a mean-median score of .05 means that for Republicans to be awarded half of the representatives, they only need 45% of the vote.

This is the main score relied on in this report. It is the longest-standing and most well established of all the partisan scores. It has many features that make it well-suited to this analysis, such as varying with sufficiently small increments over the random walks. I do not intend to endorse it as the most meaningful of partisan scores, but I have selected it as a reliable and uncontroversial score with a long pedigree.

Given the context of Pennsylvania redistricting, in which the major parties frequently have roughly equal levels of support but the congressional representation has been about two-thirds Republican, it would be even more interesting in the future to study a different score that compares the mean to the 33rd percentile value. This could be interpreted as how much Republicans can fall short of half the vote while still securing 12 out of 18 seats.

7.2 Efficiency gap

The efficiency gap formula relies on a definition of wasted votes. Suppose that party A loses district i. Then its wasted votes in that district are V_i^A, i.e., all votes were wasted. On the other hand, suppose A wins the district. Then its wasted votes are $V_i^A - \dfrac{V_i}{2}$, the votes cast for that party in excess of what was needed to win. With these definitions, we calculate W^D and W^R, the total wasted votes for each party summed over the districts. Then the efficiency gap is defined as $EG = \dfrac{W^D - W^R}{V}$, which measures the wastage for Democrats minus the wastage for Republicans as a proportion of the total vote in the state.

When this is positive, it means that the map is keyed to waste more Democratic votes. This effect is more exaggerated as EG grows higher. A rule of thumb that was proposed by the creators of the EG score is that magnitudes over .08 should be presumptively disallowed.

As is well-documented in the growing literature on efficiency gap, it is closely tied to the number of seats won by each party. If the voting turnout were equal across districts, then EG would precisely equal $2v - s - \frac{1}{2}$, where v is party A's statewide vote share (head-to-head) and s is party A's fraction of the representation. However, when turnouts are unequal, the effect of EG is similar, but with different weights to different districts according to their turnout. This is why holding Δ constant it is possible to see different values of EG with the same proportion s of the representation.

In summary, efficiency gap should be thought of as just reporting the seat allocation, with a fairly arbitrary smoothing rule. Thus the plots of efficiency gap over an ensemble should have multiple bell curves, one for each possible number of D seats encountered in the ensemble.

We further note that the findings above make a strong case against using $|EG| > .08$ as a stand-alone test. EG gives interesting information about a districting plan, but should only be assessed in the context of other possible districting plans and not against an abstract ideal.

7.3 Duke Gerrymandering Index

Finally, the *gerrymandering index* of Mattingly et al [5] can be computed as follows: for a given distribution of voters and a given districting plan $\mathcal{D}$ and vote distribution Δ, we adopt the convention that the districts of $\mathcal{D}$ are re-indexed so that D_1 has the lowest Democratic head-to-head share against Republicans, with the Democratic share increasing up to its highest value D_{18}. (That is, $X_1 \le \cdots \le X_{18}$, where $X_i = V_i^D / V_i$.)

Then let $X_i(\mathbb{E}_{\mathcal{D}})$ be the median value of X_i over the local ensemble based on a districting plan $\mathcal{D}$, and let $X_i(\mathcal{D})$ be the value in the initial districting plan. The Duke index is

$$G(\mathcal{D}, \Delta) = \sqrt{\sum_i \big(X_i(\mathbb{E}_{\mathcal{D}}) - X_i(\mathcal{D})\big)^2}.$$

The meaning of this score is to report how atypical the plan $\mathcal{D}$ is with respect to the ensemble's reported district-by-district vote shares. A carefully gerrymandered plan will exhibit signs of both *packing* (the last few $X_i(\mathcal{D})$ will be very high, over 0.7) and *cracking* (there will be several $X_i(\mathcal{D})$ at levels just enough under 0.5 to represent reasonably safe wins). The ensemble itself may show no such tendency, in which case those telltale district values will cause sizeable contributions to the sum in $G(\mathcal{D}, \Delta)$.

The simple random walk used here can also report this score, giving yet another piece of persuasive evidence that a plan is a gerrymander.

References

[1] Maria Chikina, Alan Frieze, and Wesley Pegden, *Assessing Significance in a Markov Chain without Mixing*, Proceedings of the National Academy of Sciences, March 14, 2017, vol. 114 no. 11, 2860–2864.

[2] Chikina et al, markovchain package: http://www.math.cmu.edu/∼wes/files/markovchain.tgz

[3] Ben Fifield, Michael Higgins, Kosuke Imai, and Alexander Tarr, *A New Automated Redistricting Simulator Using Markov Chain Monte Carlo*, preprint.

[4] Fifield et al, redist package: https://cran.r-project.org/web/packages/redist/index.html

[5] Greg Herschlag, Robert Ravier, and Jonathan Mattingly, *Evaluating Partisan Gerrymandering in Wisconsin*, preprint. https://arxiv.org/pdf/1709.01596.pdf

8 Appendix: Compactness scores for GOV plan

	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9
Polsby-Popper	.27	.35	.34	.45	.44	.26	.39	.27	.33
Schwartzberg	1.92	1.7	1.7	1.49	1.5	1.95	1.6	1.93	1.75
Schwartzberg*	1.85	1.68	1.53	1.41	1.47	1.82	1.56	1.85	1.64
Reock	.32	.33	.33	.39	.47	.47	.47	.29	.44
Minimum convex polygon	.7	.69	.66	.91	.81	.71	.88	.82	.78
Population polygon	.88	.75	.84	.9	.7	.71	.77	.7	.86

	D_{10}	D_{11}	D_{12}	D_{13}	D_{14}	D_{15}	D_{16}	D_{17}	D_{18}
Polsby-Popper	.35	.31	.16	.37	.31	.19	.28	.35	.38
Schwartzberg	1.68	1.8	2.49	1.64	1.8	2.29	1.88	1.69	1.62
Schwartzberg*	1.64	1.68	2.29	1.62	1.67	2.21	1.73	1.54	1.54
Reock	.53	.54	.26	.46	.6	.27	.52	.52	.51
Minimum convex polygon	.82	.79	.61	.78	.81	.74	.78	.78	.86
Population polygon	.54	.76	.34	.81	.84	.77	.71	.93	.63

All compactness scores were computed in Maptitude except for *minimum convex polygon*, which was computed in ArcGIS. As noted in §6.2, the built-in Maptitude functionality uses a slightly different definition of the Schwartzberg score than the one commonly defined in expert reports. We use Schwartzberg* to denote the Maptitude Schwartzberg score.

9 Appendix: County and municipality splits in GOV plan

Political Subdivisions Split Between Districts

Thursday February 15, 2018 2:29 PM

Number of subdivisions not split:

County 51

Number of subdivisions split into more than one district:

County 16

Number of subdivision splits which affect *no* population:

County 0

Split Counts

County

Cases where a County is split among 2 Districts: 13

Cases where a County is split among 3 Districts: 3

Number of times a County has been split into more than one district: 19

Total of County splits: 35

County	District	Population
Split Counties :		
ALLEGHENY	12	195,085
ALLEGHENY	14	705,688
ALLEGHENY	18	322,575
BEAVER	3	86,795
BEAVER	12	83,744
BERKS	6	64,981
BERKS	15	218,608
BERKS	16	127,853
BUCKS	8	545,535
BUCKS	13	79,714
CENTRE	5	84,293
CENTRE	9	69,697
CUMBERLAND	4	169,309
CUMBERLAND	11	66,097
DELAWARE	1	417,158
DELAWARE	6	141,821
LEBANON	11	75,179
LEBANON	16	58,389
LEHIGH	8	113,428
LEHIGH	15	236,069
LUZERNE	10	109,700
LUZERNE	17	211,218
MIFFLIN	9	152
MIFFLIN	11	46,530

Page 1

County	District	Population
Split Counties (continued):		
MONTGOMERY	7	705,688
MONTGOMERY	13	94,186
NORTHAMPTON	8	46,725
NORTHAMPTON	15	251,010
PHILADELPHIA	1	288,530
PHILADELPHIA	2	705,687
PHILADELPHIA	13	531,789
SOMERSET	9	16,053
SOMERSET	12	61,689
TIOGA	5	1,886
TIOGA	10	40,095

The count of municipality splits is very sensitive to the precise data source used. We identified municipalities from the US Census place dataset (TIGER files), including all incorporated cities and boroughs.

With these definitions there are 14 municipality splits, though 5 of them are due to the fact that the municipalities themselves cross county lines. By contrast, the currently enacted plan splits 20 municipalities by this definition, with 5 due to municipality/county crosses.

Split municipalities in GOV plan, with which districts intersected

Adamstown borough – 6,16
Baldwin borough – 14,18
Bristol borough – 8,13
Carnegie borough – 14,18
Central City borough – 9,12
Clairton city – 14,18
Jefferson Hills borough – 14,18
Philadelphia city – 1,2,13
Plum borough – 12,14
Seven Springs borough 12,18
Shippensburg borough – 4,9
Telford borough – 7,8
Trafford borough – 12,14
Whitehall borough – 14,18

Split precincts in GOV plan, with which districts intersected

Split VTDs

ALLEGHENY	CARNEGIE WD 01 DIST 03	14	237
ALLEGHENY	CARNEGIE WD 01 DIST 03	18	933
ALLEGHENY	INDIANA TWP DIST 04	12	1,250
ALLEGHENY	INDIANA TWP DIST 04	14	104
ALLEGHENY	INDIANA TWP DIST 05	12	875
ALLEGHENY	INDIANA TWP DIST 05	14	402
BEAVER	NEW SEWICKLEY TWP VTD FREEDOM	3	1,942
BEAVER	NEW SEWICKLEY TWP VTD FREEDOM	12	352
BERKS	ONTELAUNEE TWP DIST 01	15	241
BERKS	ONTELAUNEE TWP DIST 01	16	1,405
BERKS	SOUTH HEIDELBERG TWP PCT 01	6	28
BERKS	SOUTH HEIDELBERG TWP PCT 01	16	2,215
BUCKS	BRISTOL VTD WEST ED 01	8	538
BUCKS	BRISTOL VTD WEST ED 01	13	0
BUCKS	BRISTOL VTD WEST ED 03	8	854
BUCKS	BRISTOL VTD WEST ED 03	13	3
CENTRE	PATTON TWP VTD NORTH ED 02	5	2,708
CENTRE	PATTON TWP VTD NORTH ED 02	9	217
CENTRE	PATTON TWP VTD SOUTH ED 03	5	12
CENTRE	PATTON TWP VTD SOUTH ED 03	9	2,578
CUMBERLAND	HAMPDEN TWP PCT 02	4	2,253
CUMBERLAND	HAMPDEN TWP PCT 02	11	12
CUMBERLAND	HAMPDEN TWP PCT 10	4	4,403
CUMBERLAND	HAMPDEN TWP PCT 10	11	0
CUMBERLAND	HAMPDEN TWP PCT 12	4	2,005
CUMBERLAND	HAMPDEN TWP PCT 12	11	645
DELAWARE	HAVERFORD TWP WD 03 PCT 03	1	625
DELAWARE	HAVERFORD TWP WD 03 PCT 03	6	608
DELAWARE	MARPLE TWP WD 03 PCT 03	1	25
DELAWARE	MARPLE TWP WD 03 PCT 03	6	825
LEBANON	WEST CORNWALL TWP	11	938
LEBANON	WEST CORNWALL TWP	16	1,038
LEHIGH	SOUTH WHITEHALL TWP DIST	8	322
LEHIGH	SOUTH WHITEHALL TWP DIST 04	15	1,457
LEHIGH	SOUTH WHITEHALL TWP DIST 05	8	1,661
LEHIGH	SOUTH WHITEHALL TWP DIST 05	15	937
LUZERNE	NEWPORT TWP WD 03	10	495
LUZERNE	NEWPORT TWP WD 03	17	366
MIFFLIN	MENNO TWP Voting District	9	152
MIFFLIN	MENNO TWP Voting District	11	1,731
MONTGOMERY	UPPER MORELAND TWP VTD 02 ED 02	7	1,374
MONTGOMERY	UPPER MORELAND TWP VTD 02 ED 02	13	91
MONTGOMERY	UPPER MORELAND TWP VTD 05 ED 01	7	968
MONTGOMERY	UPPER MORELAND TWP VTD 05 ED 01	13	286
NORTHAMPTON	BETHLEHEM TWP WD 04 DIST 01	8	2,495
NORTHAMPTON	BETHLEHEM TWP WD 04 DIST 01	15	387
NORTHAMPTON	BETHLEHEM TWP WD 04 DIST 02	8	1,043
NORTHAMPTON	BETHLEHEM TWP WD 04 DIST 02	15	921
NORTHAMPTON	PALMER TWP VTD MIDDLE ED 02	8	862
NORTHAMPTON	PALMER TWP VTD MIDDLE ED 02	15	1,309
PHILADELPHIA	PHILADELPHIA WD 31 PCT 10	2	437
PHILADELPHIA	PHILADELPHIA WD 31 PCT 10	13	147
PHILADELPHIA	PHILADELPHIA WD 36 PCT 07	1	29
PHILADELPHIA	PHILADELPHIA WD 36 PCT 07	2	567
SOMERSET	CENTRAL CITY	9	1,124
SOMERSET	CENTRAL CITY	12	0
SOMERSET	CONEMAUGH TWP VTD 03	9	859
SOMERSET	CONEMAUGH TWP VTD 03	12	675
TIOGA	WESTFIELD TWP VTD 01	5	401
TIOGA	WESTFIELD TWP VTD 01	10	646