1 *Explainer: Compactness, by the numbers*

MOON DUCHIN

A few years of hands-on work analyzing redistricting has left me convinced that "compactness" is over-emphasized as a cure for gerrymandering. But on the other hand, there is always room to augment the list of metrics for measuring it. The tools of 20th century geometry (let alone 21st!) have been slow to enter the conversation. Let's review the scope of common compactness scores and introduce a new one.[1]

SCORING SHAPES

The history of shape metrics in redistricting is itself long and winding. As we'll see, there are dozens of possible scores that have been proposed in the academic and legal literature, and most of them leverage (literally) ancient mathematics. Still, the most-used scores bear the names of their 20th century popularizers: Reock, writing in 1961 [10]; Schwartzberg, writing in 1966 [11]; and Polsby–Popper, writing in 1991 [8].

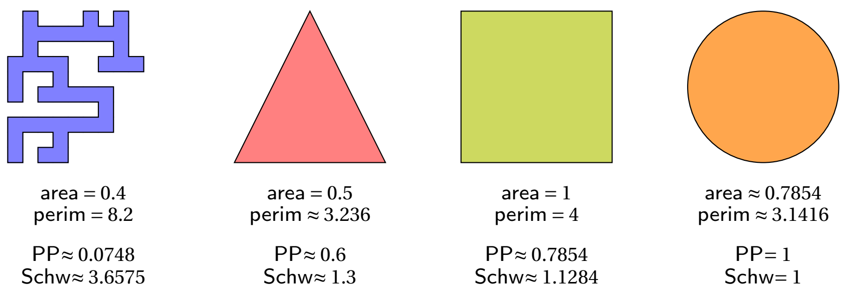


Figure 1: Shapes. If you're simply trying to measure the regularity or area-efficiency of a shape, a standard measure since antiquity has been to compare area to perimeter. The circle is the unique shape that maximizes enclosed area for a fixed perimeter.

The **Polsby–Popper score** of a region is *the ratio of the region's area to the area of a circle with the same perimeter*; bigger is better and 1 is ideal. The **Schwartzberg score** is *the ratio of the region's perimeter to the perimeter of a circle with the same area*. This time smaller is better and 1 is ideal. It's clear from the description that both of them report an ideal score when the region is a circle; it's a classical fact that scores like this *only* report ideal scores for the circle (see Figure 1). Brushing off your high school geometry and writing some formulas, these scores read as follows.

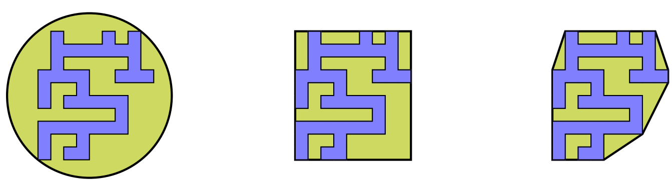
[1] This short treatment draws on ideas from Duchin and Tenner [4], which is written with an audience of political geographers in mind.

$$\mathsf{PP}(\Omega) := \frac{4\pi \cdot \mathsf{area}(\Omega)}{\mathsf{perim}(\Omega)^2}; \qquad \mathsf{Schw}(\Omega) := \frac{\mathsf{perim}(\Omega)}{\sqrt{4\pi \cdot \mathsf{area}(\Omega)}}.$$

But wait—this format makes it clear that $\mathsf{Schw}(\Omega) = \mathsf{PP}(\Omega)^{-1/2}$. Since one score is simply the other score raised to a power, it is immediate that, although specific numerical values will differ, Schwartzberg and Polsby–Popper assessments must rank districts from best to worst in precisely the same way.[2]

But here is an interesting historical quirk. Because Joseph Schwartzberg worried that there was no way (with 1966 technology) to accurately measure perimeters of districts, he also proposed a notion of *gross perimeter*, which comes from a partial discretization, placing points along the boundary and approximating the perimeter by the successive distances between those points [11]. As a result of engineers taking this suggestion literally, software like Maptitude for Redistricting uses this alternative perimeter in the computation of a Schwartzberg score but not in the computation of a Polsby–Popper score, which of course can break the scores' monotonic relationship.[3]

Another collection of scores is based on comparing the district to some idealized relative (Figure 2). For instance, if $\overline{\Omega}$ is the circumcircle of Ω—or a bounding box, or the convex hull, or some other comparison figure—then we can define a score by the proportion of the area filled in by the shape; that is, we compute $\mathsf{area}(\Omega)/\mathsf{area}(\overline{\Omega})$. Of course, redistricting is about people, not acres and trees, so we might prefer to count population rather than land area.



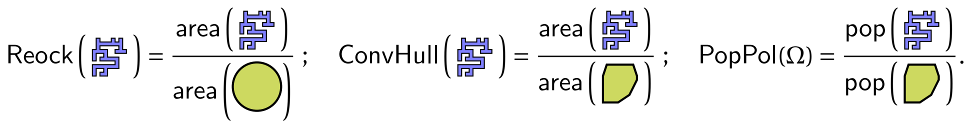
≈ 31.25% of area ≈ 44.44% of area ≈ 52.63% of area

Figure 2: Comparing a shape to its circumcircle (left), bounding box (middle), or convex hull (right—intuitively called the "rubber band enclosure"). For each of these relative area scores, the ideal would be to fill in 100% of the comparison figure.

The **Reock** score, **convex hull** score (AKA "minimum convex polygon"), and **population polygon** score are, respectively,

[2] This is because for positive values of x and y, we have $x > y \iff x^{-1/2} < y^{-1/2}$. Therefore, a higher (and thus better) PP score corresponds to a lower (and thus better) Schw score.

[3] Maptitude is so dominant in the industry that all but one brief about a proposed remedial plan in the Pennsylvania case simply included a printout of the Maptitude report to describe the compactness of the plan (as well as for other metrics like county splitting). The only exception was written by me! [5] This is a good reminder of how thoroughly software mediates our interaction with the quantifiable side of redistricting.



And now we have met the five scores that were cited by the Pennsylvania Supreme Court in the 2017–2018 redistricting challenge: Polsby–Popper, Reock, Schwartzberg, convex hull, and population polygon. Every party that submitted a remedial map was required to report these five scores for all 18 Congressional districts in the proposed map.

CONTEXT AND AGGREGATION

Armed with scores, we are ready to go! Here's a district: is it good or bad?

Well, it's got one very straight edge, which might be good. But it's pretty elongated rather than plump, so that might be bad. It's got some pretty "thin necks," which seems bad. And it's pretty windy and erratically formed, which is probably bad.

Our mystery district turns out to be Maryland's 6th (Figure 3), which was successfully challenged in district court as a pro-Democratic gerrymander in the case that eventually became Benisek v. Lamone. But its sins, such as they are, are not primarily geometric. MD-6 looks much more benign overlaid on a precinct map of the state.[4]

One more major problem demands attention: most of the zoo of compactness scores consists of individual district metrics. Even if it's clear how to compare two shapes head-to-head, how do you compare a set of 8 or 18 scores? For instance, we heard above that the Pennsylvania Supreme Court required petitioners to report 5 scores for each of the 18 districts in their plans. That means that each plan is assessed by 90 numbers, making it totally-not-obvious how to compare one plan to the next! Most petitioners reported the *average* of the 18 individual scores in each

[4] Caveat: the district is not actually made of whole precincts, but its precinct-level approximation, made using geospatial approximation based on population [6], is depicted here.

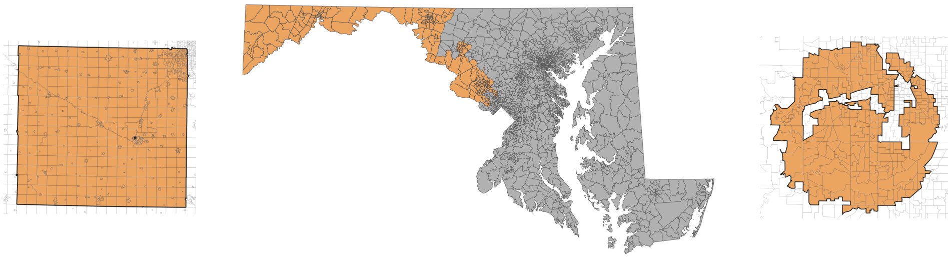


Figure 3: Left: You can build a square district when the units cooperate! But imagine trying to do that from the Maryland precinct units shown in the middle. Right: unreasonable districts can have good scores—in this case, a stellar Reock score. (Image credit: Amy Becker.)

metric, but it's rather unclear that that's reasonable. If you have a really terrible district, is it actually balanced out by a nice plump district somewhere else in the state?[5] The Pennsylvania remedial plan reported an average Polsby–Popper score of 0.33 across its 18 districts. Coincidentally, that's the same average Polsby–Popper score of the enacted Congressional plan in Minnesota across its 8 Congressional districts.[6] Given all the differences in the natural landscape, number of districts, shapes of the units, and all the rest, this does not feel like a particularly meaningful comparison.

BUT WHAT DO THE NUMBERS TELL US? AND WHAT ELSE CAN YOU DO?

Four out of the five scores mentioned above are pure shape scores, with no reference at all to the units, the population, or the particularities of the state and its districting setup—they just use area, perimeter, and some kind of old-school (we're talking *millennia*-old) plane geometry. You could compute all those scores for a Rorschach blot or a coffee stain just as easily as for a voting district. Even the population polygon score, which does make reference to the people and where they live, is still *contour-based*, in the language of Duchin and Tenner [4], which means that it is determined by the outline of the district on a flat plane. All contour-based scores will be sensitive—and sometimes highly sensitive—to things that don't matter for district quality, like the choice of projection from the Earth to a plane (see Chapter 13 and Bar-Natan et al. [1]) and the measurement precision of the winding boundaries [2]. Most will be majorly impacted by other electoral irrelevancies, like the assignment of unpopulated areas to one district or the next. On the other hand, these contour-based scores are insensitive to the physical geography (mountains, rivers, and other features of the natural and built environment) and to the units that were actually available to the districter as building blocks. It is not a great state of affairs when your metrics are heavily impacted by irrelevant factors but

[5] To drive this point home, consider that most of the scores are valued between 0 and 1. So if you raise them to any $\alpha > 0$, you get a new score between 0 and 1, but where averaging behaves differently!

[6] See `https://www.gis.leg.mn/redist2010/Congressional/C2012/reports/compactness.pdf` for the Minnesota report and `https://www.pubintlaw.org/wp-content/uploads/2017/06/attachment-1.zip` for the Pennsylvania remedial plan files and report.

not impacted at all by important features of the problem you are studying.

If we want to modernize the geometry of district shape, we should strive to build approaches that are (a) keyed to the geographical units, and (b) designed for ease of comparison to relevant alternatives. At the same time, any reasonable score should (c) comport with the all-important "eyeball test," both because of public optics and because the case law tells us that this matters [7]. Bonus points if the score is (d) attached to a mathematical formulation with good theory and good algorithms behind it. And finally a quantitative approach will always succeed better if it is (e) simply motivated and easily described.

There will not be a perfect compactness score for the 21st century, but there will be new ideas.[7] Throughout this book you will hear about the *cut edges* score of a districting plan. First, choose a redistricting setting (like Maryland Congressional districts or South Carolina House districts) and fix the units of the problem (like precincts or census blocks that you'll be using to build districts). Then you simply score a plan by the number of pairs of units that were adjacent in the state but were severed from each other by the division into districts (see Figure 4). This is directly based on the geographic units and doesn't care at all about how jagged the units themselves are. It doesn't need any averaging or summing to give you a whole-plan score, so it's set up very well for within-state comparisons. It does a good job of tracking with visual district appearance, as you can see in Figure 4. And it is extremely natural from a mathematical point of view: in combinatorics terms, it is the size of the cut-set for the graph partition. A host of theorems and algorithms exist that reference and leverage this notion.[8] Finally, I think it does pretty well on the simplicity scale: It measures the number of neighbors that are separated when cutting out the plan with scissors!

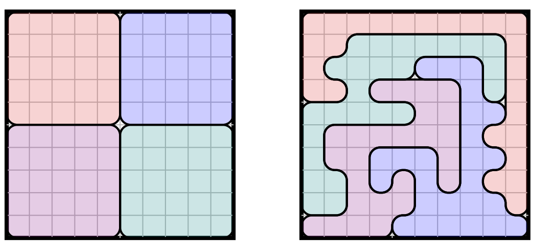


Figure 4: Cut edges as a measure of compactness. The plan on the left has just 20 cut edges, whereas the plan on the right cuts far more (73 out of the 180 edges in the grid). (Reproduced from DeFord et al. [3].)

[7] Having been asked about this point in a recent deposition, I'd like to be very clear that I'm proposing discrete geometry to *complement* traditional contour-based geometry, giving another vantage point on the efficiency or complexity of district shapes. Also, as you'll learn in this book, discrete geometry undergirds the powerful algorithms used to explore the universe of districting plans.

[8] To name just a few settings for cut-sets: the *max flow–min cut* problem is one of the foremost mathematical models of the 20th century, and important algorithms like Karger's algorithm were built to find minimum cuts. The *Cheeger constant* in geometry relates the sizes of cut-sets to the sizes of the separate pieces of the partition; there are theorems connecting the values of this constant h to various notions of curvature, expansion, and to spectral properties. Multiple authors, including Clelland et al., Procaccia–Tucker-Foltz, and Tapp, show evidence that the cut edge count correlates closely with the spanning tree count discussed in Chapter 17.

No score is going to do the hard, human, deliberative work of finding fairness in representative democracy. But there are still best practices for designing metrics, and they call for thinking about *the work that you want the score to do for you.* Your score should track the distinctions that you are most interested in flagging, and not superfluous ones. Being attentive to sensitivity, robustness, and the incentives created by a score—or the flip side of incentives, namely gameability—will help you to troubleshoot and improve it. And you won't get anywhere without trying your new metric out on real data. These are sound guidelines for *critical modeling* that will be valuable throughout the study of redistricting, and beyond.

WHERE TO READ MORE

In addition to the references cited in this chapter, those interested in learning more should check out the following resources:

- Stephen Ansolabehere and Maxwell Palmer. *A Two Hundred-Year Statistical History of the Gerrymander.* Ohio State Law Journal, vol. 77, 2016: 1–23.
- Daryl DeFord, Hugo Lavenant, Zachary Schutzman, and Justin Solomon. *Total Variation Isoperimetric Profiles.* SIAM Journal on Applied Algebra and Geometry 3.4, 2019.
- Aaron Kaufman, Gary King, and Mayya Komisarchik. *How to Measure Legislative District Compactness If You Only Know It When You See It.* American Journal of Political Science 3 (2017): 55-58.
- Richard G. Niemi, Bernard Grofman, Carl Carlucci, and Thomas Hofeller. *Measuring Compactness and the Role of a Compactness Standard in a Test for Partisan and Racial Gerrymandering.* The Journal of Politics 52 (4) 1990: 1155–81.

REFERENCES

[1] Assaf Bar-Natan, Elle Najt, and Zachary Schutzmann, *The Gerrymandering Jumble: Map Projections Permute Districts' Compactness Scores.* Cartography and Geographic Information Science, Volume 47, Issue 4, 2020, 321–335.

[2] Richard Barnes and Justin Solomon, *Gerrymandering and Compactness: Implementation Flexibility and Abuse.* Political Analysis, Volume 29, Issue 4, October 2021, 448–466.

[3] Daryl DeFord and Moon Duchin. *Redistricting Reform in Virginia: Districting Criteria in Context.* Virginia Policy Review, Vol. 12 No. 2 (2019), 120–146.

[4] Moon Duchin and Bridget Tenner. *Discrete Geometry for Electoral Geography,* preprint 2018. `https://arxiv.org/abs/1808.05860`

[5] Moon Duchin. *Outlier Analysis for Pennsylvania Congressional Redistricting.* Technical report, 2018. Expert report submitted in *LWV v. Pennsylvania.*

[6] MGGG Redistricting Lab, *MAUP: A geospatial toolkit for redistricting data.* GitHub repository. `https://github.com/mggg/maup`

[7] Richard H. Pildes and Richard G. Niemi. *Expressive Harms, "Bizarre Districts" and Voting Rights: Evaluating Election-District Appearances after Shaw v. Reno.* Michigan Law Review, Vol. 92, No. 3 (Dec., 1993), 483–587.

[8] D.D. Polsby and R.D. Popper, *The Third Criterion: Compactness as a Procedural Safeguard Against Partisan Gerrymandering.* Yale Law & Policy Review Vol. 9 (1991), 301–353.

[9] Ariel D. Procaccia and Jamie Tucker-Foltz. *Compact Redistricting Plans Have Many Spanning Trees,* Preprint. 2021. `https://arxiv.org/abs/2109.13394`

[10] Ernest C. Reock. *Measuring Compactness as a Requirement of Legislative Apportionment.* Midwest Journal of Political Science, Vol. 5, No. 1 (Feb., 1961), 70–74.

[11] Joseph E. Schwartzberg. *Reapportionment, Gerrymanders, and the Notion of "Compactness."* Minnesota Law Review, Vol. 50 (1966), 443–452.