Gerrymandering and Compactness: Implementation Flexibility and Abuse

Richard Barnes a,b,1 and Justin Solomon b

^aEnergy & Resources Group, UC Berkeley, 310 Barrows Hall, Berkeley, CA, 94720. ORCID: 0000-0002-0204-6040; ^bComputer Science and Artificial Intelligence Laboratory (CSAIL), MIT, Cambridge, USA, 02139

This manuscript was compiled on March 9, 2018

The shape of an electoral district may suggest whether it was drawn with political motivations, or gerrymandered. For this reason, quantifying the shape of districts, in particular their compactness, is a key task in politics and civil rights. A growing body of literature suggests and analyzes compactness measures mathematically, but little consideration has been given to how these scores should be calculated in practice. Here, we consider the effects of a number of decisions that must be made in interpreting and implementing a set of popular compactness scores. We show that the choices made in quantifying compactness may themselves become political tools, with seemingly innocuous decisions leading to disparate scores. We show that when the full range of implementation flexibility is used, it can be abused to make clearly gerrymandered districts appear quantitatively reasonable. This complicates using compactness as a legislative or judicial standard to counteract unfair redistricting practices. This paper accompanies the release of packages in C++, Python, and R which correctly, efficiently, and reproducibly calculate a variety of compactness scores.

geographic information system (GIS) \mid open source software \mid redistricting \mid gerrymandering \mid geometry

Gerrymandering is the practice of designing political districts whose shape serves some agenda. Reasons for gerrymandering range from fundamental concerns like equal representation to less ethical considerations like disenfranchising a minority population. Although contorted shapes can arise from geographic or legal necessity, such as rivers or the boundaries of political superunits, poor geometry is often associated with a political agenda. Thus, detection and quantification of geometric quality is a key consideration in efforts to make redistricting standards systematic, though not the only one. (1)

The *compactness* of a district is a key geometric consideration intended to capture the issues above. Many measures of compactness exist (2–4), and an ongoing discussion between mathematicians and legislators continues to debate their relative merits in promoting desirable district shapes. There has been less discussion, however, about how compactness scores should be implemented in practice.

Here, we the US Census Bureau's 2015 Cartographic Boundary and TIGER/Line data (5) to consider how the variables used to calculate compactness are complicated by reality and how, even once defined, issues including geography, topography, projections, and resolution complicate implementation. Together, the ambiguities we expose provide a high degree of implementation flexibility. We show that this flexibility can be exploited to argue that convoluted and likely gerrymandered districts are quantitatively reasonable.

We conclude with general recommendations for development and fair characterization of compactness scores; we additionally provide a model software implementation intended to avoid the pitfalls we highlight. All of the examples presented and some of the terminology used stem from United States geopolitics, but the ideas presented herein are applicable to redistricting more broadly.

1. Results

We have identified a number of choices that must be made to compute a compactness score. In addition to the choice of (1) compactness definition, it is also important to consider how to handle (2) non-contiguous districts, (3) districts with holes, (4) political superunit boundaries, (5) map projections, (6) topography, (7) data resolution, (8) floating-point uncertainty, and (9) whether alternative choices were possible in drawing a district's boundaries. These are considered independently below

In combination, these choices provide potentially undesirable implementation flexibility. This flexibility can be abused: Different implementation choices applied to what is nominally the same data lead to very different conclusions about fairness of a districting plan.

To demonstrate this effect, we have selected ten U.S. Congressional Districts widely considered to be gerrymandered. Using an optimizer, we apply the full flexibility detailed in this paper and are able to find sets of implementation decisions for which these districts' compactness scores are outliers when compared against the full distribution of districts' scores. We are also able to find sets of decisions which make these districts

Significance Statement

Gerrymandering is the practice of drawing electoral districts whose shape serves a political agenda, often the disenfranchisement of a population of voters. A widely-used measure of shape is compactness; highly non-compact districts are often suspected of having been gerrymandered. The constitutionality of districts in Wisconsin and Maryland is now being considered by the U.S. Supreme Court, with other cases on the way. The court will be pressured to adopt standards, such as a compactness score. Here, we explore the many ways scores can be implemented in practice and show that the ambiguity with which current scores are defined leaves them open to abuse and manipulation. This presents a major challenge to adopting any legislative or judicial standard of compactness.

RB developed the initial writeup, open source tools, and analyses. JS provided guidance on algorithmic and mathematical development. Both authors collaborated on developing the ideas in this paper and on the writing the report.

We have no conflict of interest to declare.

¹To whom correspondence should be addressed. E-mail: richard.barnes@berkeley.edu

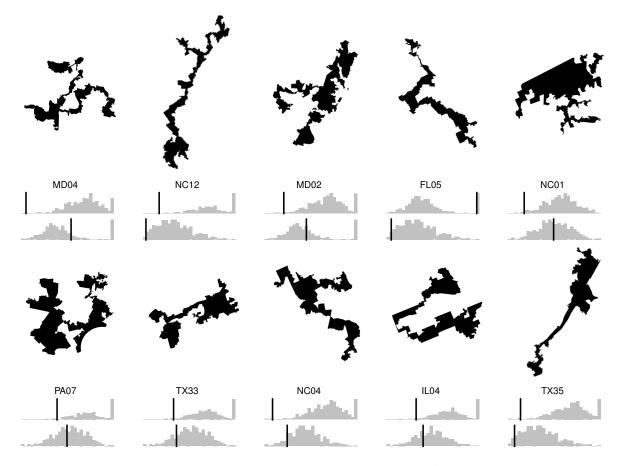


Fig. 1. Applied gerrymandering: abusing implementation flexibility. This figure shows several districts from the 114th Congress that appear incontrovertibly gerrymandered. The compactness scores of all the districts are shown in histograms below the districts' pictures with a black line indicating where the focal district falls on the distribution. Compactness ranges from 0 on the left-hand side of each histogram to 1 on the right-hand side. Scores for districts were generated by performing a grid search over a range of values for each implementation choice and choosing the minimum/maximum value across all choices. The grid search could also choose whether or not to include sole districts (§F) in the histogram. The result for each district is the pair of configurations in which the district appeared both the most gerrymandered as well as the most reasonable. Further details for the figure are in the Supplemental Information (see Table 1).

appear reasonable. That is, we can exploit implementation flexibility to build a seemingly reasonable argument that these districts are both gerrymandered and not.

Figure 1 shows the effects of this adversarial choice of parameters. In the case of NC01, IL04, and PA07, it was possible to move the districts from obvious outliers to middle-of-the-pack status. In other cases, such as NC12, NC04, and TX35, this was not possible, but the districts can still be moved considerably closer to the mean, countering arguments that they are outliers.

The optimizer does not need to use extreme settings to produce the desired results. For example, TX33 appears most gerrymandered using the CvxHullPTB score (scores are defined below) at a 500 m simplification tolerance in a locally-optimized Lambert conformal conic projection all districts included in the distribution; it appears least gerrymandered using the ReockPT score with a 500 m tolerance in a Gall projection with districts comprising an entire state excluded.

A. Open Source Tools. Of the many compactness scores discussed in the literature, some are better able to cope with the complexities discussed here than others. Many of the more robust metrics, however, are also difficult or impossible to calculate using commonly-available software. For instance,

QGIS (6) includes the area of multipolygons as a built-in display field, convex hulls as a function three menu levels deep, and has no functionality to calculate the minimum bounding circles needed for Reock scores. Other scores, such as bizarreness (4) have mathematical descriptions of a complex calculation, but no associated source code.

To address this situation, we have released a family of open source packages which share a common library designed to efficiently, reproducibly, and correctly calculate a variety of compactness scores. The basis of this ecosystem is compactnesslib,* a C++ library and associated command-line interface which ingests bulk or single data in a variety of formats and calculates compactness scores. The python-mander Python package[†] (available via pip[‡]) and the mandeR R package[§] provide high-level interfaces to this library. In addition, a QGIS plugin provides GIS users an easy means of calculating scores (7–10). This stack was utilized to produce the calculations in this paper: The complete source code for generating all the diagrams presented here

^{*}https://github.com/gerrymandr/compactnesslib

[†] https://github.com/gerrymandr/python-mander

[‡]https://pypi.python.org/pypi/mander

[§] https://github.com/gerrymandr/mandeR

¶ https://github.com/gerrymandr/qgis-compactness

is available at github.com/r-barnes/Barnes2018-compactness-implementation.

B. Coda. The measurement of compactness can be used as a tool to help detect and quantify gerrymandering. Numerous engineering and implementation decisions, however, must be made to calculate a score. Whether used unintentionally or maliciously, this flexibility has strong bearing on the quality of compactness measurements and can be leveraged to shape conclusions about the quality of a districting plan.

Beyond providing "best practices" for implementations of compactness standards, we intend the open source software accompanying this paper as a first step toward fair and accurate measurement of compactness, allowing scientists, politicians, and the public to evaluate aspects of their democracy using reproducible, mathematically well-founded, and computationally stable tools.

Finally, we remind the reader that the goal of all of this is to help governments represent their people. Compactness, while attractive as a quantitative metric, is a tool, not the end-game.

2. Discussion

- **A. Best Practices.** The foregoing highlights the importance of being clear about how a score is calculated. In general, a mathematical definition alone is not sufficient: Attention must be paid to data and algorithmic quality. Here, we suggest best practices for the calculation of compactness scores:
 - Scores. Be explicit about what each variable in a compactness score means. Does area include holes? Is it constrained by political superunits? How should noncontiguous districts be handled? Score names should be distinct and informative. Appending a clarifying suffix to the name of a score (e.g. *PTSHp*) informs readers as to what is being done. See our Methods for examples.
 - **Projections.** Scale distortion should be limited to less than 1.25% throughout the region of interest. Reasonable choices of national or local projections usually suffice.
 - Resolution. Use the best available resolution from a trusted source. Simplified or down-scaled data give altered results. Alternatively, choose a score which is robust to changes in resolution: hull-based scores seem to do well in this regard. The U.S. Census Bureau produces reasonable data designed such that all borders that are at the same resolution align. Ideally, districting data should be drawn from a common, trusted, non-partisan source, regardless of who is performing an analysis.
 - Border constraints. Scores which do not explicitly account for constraints imposed by superunit boundaries leave out valuable information about what was possible in drawing a district. That is, they may unfairly penalize a district for having an odd shape when no other shape was possible. Use a score that accounts for superunit borders. Be sure that borders are cropped to features such as major coastlines.
 - Choice. Before doing statistics on a set of district plans, eliminate those districts which encompass an entire political superunit, as no other choices of shape were possible.

- **Topography.** We have not found including topography in the calculation of area to be a significant source of error, assuming the use of acceptable map projections.
- Border coalignment. Coalignment of borders is a concern, though the effect was small in our data. To avoid problems, datasets used in an analysis should always be at the same resolution and carefully coaligned during their creation. In the U.S., Census data satisfies these requirements.
- Floating-point considerations. We have not found the choice of single- or double-precision floating-point representations to be important in our calculations.
- Transparency. A compactness score should not be accepted and cannot be interpreted without knowing the steps that went into creating it. From a scientific standpoint this relates strongly to reproducibility: We cannot trust what we cannot reproduce. Therefore, documentation is needed down to the equation level, and the release of source code is critical (11–13).
- B. Policy Implications. While the U.S. court system has declared that egregious gerrymandering is unconstitutional (14-16), the courts have thus far declined to adopt a quantitative standard by which gerrymandering can be judged; however, they have left open the possibility that a "workable standard" exists. (17) This paper demonstrates that any standard must be specified precisely and carefully, since differences in interpretation can have large effects on scores. Furthermore, this paper demonstrates that even a well-specified standard may judge unreasonable districts as being reasonable (see Figure 1). Therefore, any legally-mandated standard of compactness should leave open the possibility of challenges. Additionally, given the implementation flexibility discussed here and its potential for abuse, courts should not accept quantitative arguments unless the code used to build those arguments is made publicly accessible.

3. Materials and Methods

A. Definitions of Compactness. There are over 24 different measures of compactness in the literature, and no doubt many others exist. The measures break down into roughly five categories: (1) length vs. width, (2) area ratios, (3) perimeter-to-area ratios, (4) other geometric measures (moment of inertia, interior angles, &c.), and (5) measures incorporating population or other such information. (2, 3)

In this paper, we consider three widely-used compactness scores and their variants (Figure 2 provides a depiction):

- 1. Polsby-Popper (18): $4\pi A/P^2$ where A is the area of a district and P its perimeter
- 2. Reock (19): the ratio of a district's area to the area of its minimum bounding circle. Note that finding this circle is harder than locating the two most distant points of a district; an efficient algorithm and associated implementation is described in (20).
- 3. Convex Hull (3): the ratio of a district's area to the area of its convex a hull (the minimum convex shape that completely contains the district).

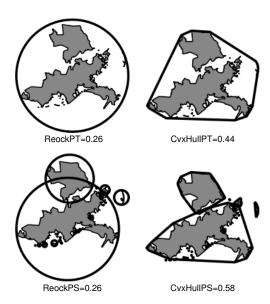


Fig. 2. Reock and Convex Hull scores for Louisiana 01 shown with both the *PS* and *PT* interpretation depicted. It is coincidental that ReockPT and ReockPS are the same here. Note that for both of the ReockPT and CvxHullPT scores the hull polygons overlap; this is potentially problematic since it could be considered double-counting.

All of the above scores are in the range [0, 1] with higher values indicating greater compactness. Districts with relatively low values might be suspected of having been gerrymandered.

Note that these scores are purely geometric; it may be that scores incorporating population densities or other demographic data provide a better means of measuring gerrymandering, but we do not pursue this direction in our discussion. It is likely that incorporating such additional data would exacerbate the issues we discuss.

B. Nomenclature. All of these measures are under-defined: They assume that an electoral district is described by a single polygon without any holes. In reality, districts, such as those with islands (see Figure 2), often are comprised of many polygons. While holes in districts are rarer, they do occur. To resolve these difficulties, we suggest methods be defined with specific reference to multiple polygons and holes.

Here, whether or not contiguity is accounted for in a score will be indicated by the suffixes PT (polygons together) and PS (polygons separate). Whether or not holes are accounted for will be indicated by the suffixes AH (add holes) and SH (subtract holes). If there is ambiguity regarding whether area, perimeter, or some other quantity is being treated in this way, then terms such as PTaSHp (treat the area of the polygons together, subtract the perimeter of holes) may be used. The suffix B indicates that a score accounts for constraints imposed by the boundaries of political superunits.

C. Non-contiguous Districts. There is no federal requirement that districts be contiguous, nor do many states require it. Indeed, the presence of islands (e.g. Hawaii) can make contiguity an impossibility. Non-contiguity may arise in other ways. Civil rights considerations have given Louisiana 01, depicted in Figure 2, two large portions separated by Louisiana 02; Louisiana 02 was drawn as a majority-minority district following the passage of the Voting Rights Act of 1965. Wisconsin's 61st Assembly District (Figure 5) exemplifies a different situa-

tion. The city of Racine, WI, had a non-contiguous boundary as a by-product of annexation, yet Wisconsin required that its districts be composed entirely of wards. As a result, the district itself is non-contiguous and could not legally be drawn in any other way (21). For the 114th Congress 1:500,000 resolution data, 84 of 441 districts are non-contiguous. Of the non-contiguous districts the largest number of subdivisions was 580 (in Alaska) and the median was 5.

The question then is whether a district should be treated as a single unit or several independent units. Treating the district as a single unit by, e.g., enclosing it in a single hull, will tend to result in lower compactness scores indicative of gerrymandering. Treating the district as a separate units and summing the areas of the units' enclosing hulls will result in higher compactness scores indicating less gerrymandering.

Mathematically speaking, although Polsby-Popper is usually calculated as being proportional to $\frac{A}{P^2}$, there are at least two possibilities for extending this formula to non-contiguous districts, in particular $\sum_i \frac{A_i}{P_i^2}$ and $(\sum_i A_i)/(\sum_i P_i^2)$, where i enumerates the non-contiguous subregions of the district. Note that although the original Polsby-Popper score is bound to the range [0,1], this is not true of the first of these alternatives. Here, we use the latter alternative.

Ultimately, special attention should be given to non-contiguous districts to determine whether they result from natural features, legitimate legal requirements, or electoral engineering. Figure 7 shows the effect the foregoing interpretations can have on compactness scores. The wide gap between different interpretations of what is nominally the same score supports the need for exactitude in both language and implementation.

D. Holes. Holes are relatively rare in districting, but many of the same considerations apply. Wisconsin 61, discussed previously, has a legally-mandated hole (Figure 5). Texas 18 very nearly surrounds the urban core of Houston and could, in a low-resolution dataset, be assigned a hole. Holes also appear as artifacts of the digitization process (Figure 8). For the 114th Congress 1:500,000 resolution data, four of 441 districts have holes as artifacts.

E. Borders. Districts are constrained by borders imposed by higher geopolitical units as well as by nature. Compactness scores that do not account for such constraints may assign inappropriately low scores to a district. The panhandles of Florida and Oklahoma, as well as Kentucky's border with the Ohio River (see Figure 4), contain electoral districts whose shape, at least in part, cannot be dictated by politics. The same is true of almost any coastal district since islands and peninsulas must be included, but lengthen their perimeters. Louisiana (see Figure 4) exmplifies this.

Some scores can be modified to account for this issue. They can be marked with the suffix B (borders accounted for). For example, in the case of the convex hull and Reock scores, if the hull or minimum bounding circle is intersected with a state polygon, the result is a better representation of what was possible and, therefore, a better indicator of whether gerrymandering took place. Taking this into account can have a considerable impact on compactness scores (Figure 9). Those scores, such as the Polsby-Popper, which cannot be modified to account for borders, are calculated as described elsewhere without consideration of borders.

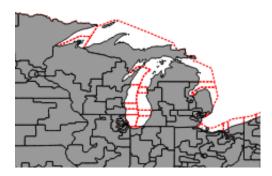


Fig. 3. Electoral districts of the 114th Congress including maritime regions. Two datasets of electoral districts are overlaid. The gray area depicts electoral district boundaries cropped to coastlines whereas the dashed red line indicates the full extent of the electoral districts. Note the growth of the district's areas and the relative smoothness of the perimeters. Data was drawn from the US Census Bureau (5): cropped data is from the Cartographic Boundaries dataset, e.g., cb_2015_us_cd114_rr.zip, whereas uncropped data is from the TIGER/Line dataset, e.g., tb_2015_us_cd114.shp.

The boundaries of electoral districts, states, and countries may include large maritime regions, as shown in Figure 3. Insofar as these regions generally cannot be populated, save for areas immediately adjacent to the shore, their inclusion in compactness calculations may serve to hide the effects of gerrymandering. Input data should be cropped to major coastlines to account for this, though, doing so is not a panacea: coastlines tend to be fractal (see Figure 14).

As Figure 6 shows, border data, especially when drawn from disparate sources, may not always co-align. We attempted to quantify this effect by overlaying high-resolution district data with medium-resolution state data and found that the impact was usually small (see Figure 10 for details). Problems can be avoided entirely by using data which is co-aligned, such as is available from the U.S. Census.

- **F. Choice.** If only one possible plan exists for a district, that district cannot be gerrymandered and should be excluded from analysis. In the Census Bureau data (5) used here, 13 congressional districts, including Alaska, Delaware, and Vermont, had only one congressional district. No matter how oddly shaped these districts are, they are not gerrymandered.
- **G. Projections.** Although scores are often defined as though districts exist on a plane, in reality districts are wrapped around the curvature of the Earth and local topographical features. Several interpretations of scores are possible: districts could be mapped to the plane using a projection designed to minimize distortion across an entire country, a subdivision of a country such as a state, or even the district itself. Alternatively, variables could be calculated on the sphere or WGS84 ellipsoid. As Figure 11 shows, despite all the possibilities, compactness measures appear to be stable to reasonable choices among localized (country-scale) map projections used in practice. Alaska demonstrates what happens when an unreasonable choice is made: its score in a projection suitable for the conterminous United States differs that in an Alaska-specific projection by up to 20%.

Clearly, using a global projection such as the standard Mercator induces too much distortion. This implies that Web Mercator (EPSG:3857) should never be used for compactness

calculations, despite its ubiquitous use on the internet. Across all districts, scores, and projections, the absolute score difference between a district as measured in a locally-optimal projection versus a conterminous projection was less than 0.009 in 99% of cases. The other 1% of cases comprise districts such as Alaska and American Somoa, which are outside the region of interest for the conterminous projections. Given this, nation-sized projections—excluding outlying states and territories—are likely reasonable choices. Quantitatively, the conterminous Albers Equal Area (EPSG:102003) projection has a maximum scale distortion of 1.25% (22): this value hence can be taken as an upper limit on what is acceptable for any projection and is our recommended choice for districts in the conterminous United States.

- H. Topography. A different effect of mapping electoral districts to a plane is that topography, such as mountains, is left out of quantities such as area and perimeter. As a result, the true land area and overland distance between points is under-estimated. Using the 30 m USGS National Elevation Dataset (23), we calculated the surface area of districts using RichDEM's implementation (24) of an algorithm by Jenness (25) and modeled perimeter as the summed length of all the raster elevation cells at the edge of a district. The difference in Polsby-Popper scores between the topographic and non-topographic data was less than 0.03 for all districts, with 75% of districts having deviations less than 0.005. This should be expected given that Kansas (and every other state) is provably flatter than a pancake. (26)
- I. Resolution. Resolution can be thought of as the density of points describing a boundary. Figure 4 shows the same district at several resolutions; lower resolutions lead to simpler shapes usually, but not always, by reducing the length of the perimeter. The U.S. Census Bureau releases boundary data of Congressional Districts in four resolutions: full, 1:500k, 1:5M, and 1:20M (5). The full-resolution data is available as "TIGER/Line" data whereas the other resolutions are available as "Cartographic Boundary Shapefiles." At these resolutions the perimeters of the districts of the 114th Congress are defined by an average of 8914, 1531, 322, and 70 points, respectively.

As services move online and onto mobile devices with constrained processing, it will be tempting for practitioners to introduce lower-resolution or simplified data into compactness measurements. Even in the high-performance environments used for automated redistricting efforts (27), low-resolution data is tempting as it may yield substantial savings on compute time. Ultimately, we find that the choice of resolution has a substantial impact on compactness scores (Figure 13 and 15) with the Polsby-Popper score especially affected. This adds to a growing list of criticisms of the Polsby-Popper score. (1, 4)

Since data may be supplied to users by outside sources, adversarial inputs are possible: A high-frequency wave applied to the boundary of a district may be visually imperceptible while introducing substantial alterations to a district's score. The Koch snowflake is an extreme example of this: It has an arbitrarily-long perimeter surrounding a finite area (Figure 14). More practically, data may contain digitization or simplification artifacts that only become apparent under significant magnification, as shown in Figure 8.

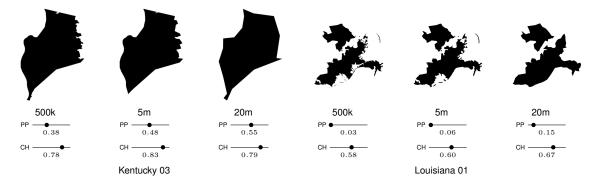


Fig. 4. Effect of polygon simplification on districts and their compactness scores. Districts from the 114th Congress are shown at 1:500,000 (500k), 1:5,000,000 (5m), and 1:20,000,000 (20m) resolution. Simplification was performed by the US Census Bureau using in-house algorithms that ensure border alignment. Here, PP stands for PolsbyPTAH while CH stands for CvxHullPT; note how these scores change with resolution. Kentucky 03 encompasses metropolitan Louisville and is bounded on the north by Kentucky's state border and the Ohio River. Louisiana 01 is bounded by the Mississippi Delta, divided by Louisiana 02, and includes unexpected parts of New Orleans. Note that under simplification the rough edges of Kentucky 03 disappear, as do entire bays and islands in Louisiana 01.

4. Ordering

The foregoing considerations change not only what the values of the calculated scores are, but also the relative ordering of the scores (Figure 16). If this is quantified using Spearman's rank correlation coefficient (Figure 17), it is apparent that different scores give markedly different rankings. Thus, any ranking of districts by compactness is thoroughly tied to and arises from choices made in developing the scores. Figure 1 explores this issue further.

5. Floating-point Issues

Computers generally store fractional values based on the IEEE754 specification using either the 32-bit single-precision type, which gives about 7 decimal places of precision, or the 64-bit double-precision type, which gives about 15 decimal places of precision. In terms of decimal degrees, the former provides approximately centimeter accuracy while the latter provides nanometer accuracy; thus, single-precision is sufficient for storing geographic coordinates. However, performing mathematics on fractional numbers, especially 32-bit types, is known to give potentially erroneous results (28).

We tested for this by computing all of the scores mentioned here using both 32-bit and 64-bit IEE754 compliant types, with the latter taken as the "true" value. No score had a percent difference between the two of more than 0.027%.

ACKNOWLEDGMENTS. The open source software described here had its genesis in the *Geometry of Redistricting* workshop held at Tufts University August 7–11, 02017. John Connors helped develop the mander package. Max Gardner, Aaron Dennis, Daniel McGlone, and Ariel M'ndange-Pfupfu helped develop the python-mander package. Ariel M'ndange-Pfupfu and Vanessa Archambault helped develop the QGIS plugin. Computation and data utilized XSEDE's Comet supercomputer (29). Travel funding for RB and research support for JS was provided by a Prof. Amar G. Bose Research Grant and an Amazon Research Award. In-kind support was provided by Isaac B., Hannah J., Kelly K., Vivian L., and Jerry W.

- 1. Alexeev B, Mixon DG (2017) An impossibility theorem for gerrymandering. ArXiv e-prints.
- Altman M (1998) Chapter 2: The Consistency and Effectiveness of Mandatory District Compactness Rules in *PhD Thesis: Districting principles and democratic representation*. (California Institute of Technology).
- Niemi RG, Grofman B, Carlucci C, Hofeller T (1990) Measuring Compactness and the Role
 of a Compactness Standard in a Test for Partisan and Racial Gerrymandering. The Journal
 of Politics 52(4):1155–1181.

- Chambers C, Miller A (2010) A Measure of Bizarreness. Quarterly Journal of Political Science 5(1):27–44.
- United States Census Bureau (2016) Cartographic boundary shapefiles (https://www.census. gov/geo/maps-data/data/cbf/cbf_cds.html access on 2017-08-26).
- QGIS Development Team (2017) QGIS Geographic Information System (Open Source Geospatial Foundation).
- 7. Barnes R (2018) compactnesslib (https://github.com/gerrymandr/compactnesslib)
- 8. TODO, TODO (2018) python-mander (https://github.com/gerrymandr/python-mander)
- Barnes R, Connors J (2018) mandeR: Compactness measures (https://github.com/ gerrymandr/mandeR).
- Archambault V, M'ndange-Pfupfu A (2017) qgis-compactness (https://github.com/gerrymandr/qgis-compactness).
- 11. Barnes N (2010) Publish your computer code: it is good enough. Nature 467:753.
- 12. Merali Z (2010) Why scientific programming does not compute. *Nature* 467:775–777.
- Ince DC, Hatton L, Graham-Cumming J (2012) The case for open computer programs. Nature 482:485–488.
- 14. Supreme Court of the United States (1986) Davis v. Bandemer, 478 U.S. 109 (1986).
- United States Federal Courts (2016) Whitford v. Gill, 218 F. Supp. 3d 837 (W.D.Wis. 2016), stayed pending appeal 137 S. Ct. 2289 (June 19, 2017).
- Supreme Court of Pennsylvania (2018) League of Women Voters v. Commonwealth, No. 159 MM 2017, 2018 Pa. LEXIS 438 (January 22, 2018) supplemental opinion at League of Women Voters v. Commonwealth, 2018 Pa. LEXIS 771 (February 7, 2018).
- 17. Supreme Court of the United States (2004) Vieth v. Jubelirer, 541 U.S. 267 (2004).
- Polsby DD, Popper RD (1991) The third criterion: Compactness as a procedural safeguard against partisan gerrymandering. Yale Law & Policy Review 9(2):301–353.
- Reock EC (1961) A note: Measuring compactness as a requirement of legislative apportionment. Midwest Journal of Political Science 5(1):70–74.
- Gärtner B (1999) Fast and robust smallest enclosing balls. *Algorithms-ESA'99* pp. 693–693.
- Altman M, McDonald MP, , et al. (2011) Bard: Better automated redistricting. Journal of Statistical Software. Forthcoming, URL http://www. jstatsoft. org 42(4):1–28.
- Deetz CH, Adams OS (1934) Elements of map projection with applications to map and chart construction. (US Gov Printing Office for Department of Commerce's Coast and Geodetic Survey), 4 edition.
- U.S. Geological Survey (USGS) (2016) USGS national elevation dataset (NED) 30m version (https://prd-tnm.s3.amazonaws.com/index.html?prefix=StagedProducts/Elevation/1/IMG/ accessed 02017-08-26).
- 24. Barnes R (2016) RichDEM: Terrain Analysis Software.
- Jenness JS (2004) Calculating landscape surface area from digital elevation models. Wildlife Society Bulletin 32(3):829–839.
- Fonstad M, Pugatch W, Vogt B (2003) Kansas is flatter than a pancake. Annals of Improbable Research 9(3):16–18.
- Tam Cho WK, Liu YY (2016) Toward a talismanic redistricting tool: A computational method for identifying extreme redistricting plans. Election Law Journal 15(4):351–366.
- Goldberg D (1991) What every computer scientist should know about floating-point arithmetic ACM Computing Surveys (CSUR) 23(1):5–48.
- Towns J, et al. (2014) XSEDE: accelerating scientific discovery. Computing in Science & Engineering 16(5):62–74.
- Legislative Technology Services Bureau (2017) Wisconsin state assembly districts (https://data-ltsb.opendata.arcgis.com/datasets/a907d137f96b49289d83172db8cf96f0_0 acquired 02017-08-28).
- Snyder JP (1984) A low-error conformal map projection for the 50 states. The American Cartographer 11(1):27–39.
- Von Koch H (1904) Sur une courbe continue sans tangente obtenue par une construction géométrique élémentaire. Arkiv för matematik, astronomi och fysik utgifvet af Kungl Svenska vetenskapsakademien 1:681–702.
- 33. Gillies S, , et al. (2007-) Shapely: manipulation and analysis of geometric objects.
- OSGeo (2017) GEOS geometry engine, open source (https://trac.osgeo.org/geos).

Supplemental Information

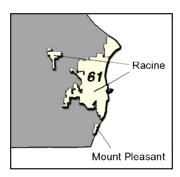


Fig. 5. Wisconsin's 61st Assembly District showing non-contiguous regions. See text for discussion. Figure drawn from (30).



Fig. 6. Misaligned borders. The border shown lies between Maryland 06 and West Virginia 01. The "true boundary" is drawn from data at 1:500,000 resolution and is shown by the transition between greys, while the black line represents the state boundary between the two districts and is drawn from 1:5,000,000 data. Note that differing data resolutions is only one way whereby such a mismatch might occur: shifts in data (as from projections), differing collection procedures, or deliberate manipulation are all possible.

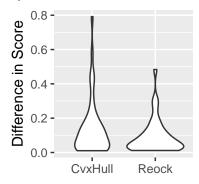


Fig. 7. Absolute value of differences in definitions of scores for districts of the 114th Congress. Shown are CvxHullPT vs. CvxHullPS and ReockPT vs. ReockPS. Districts are only shown if their score changed between definitions and they were part of a multi-district state, giving 47 data points for CvxHull and 38 for Reock. The data resolution was 1:500,000.

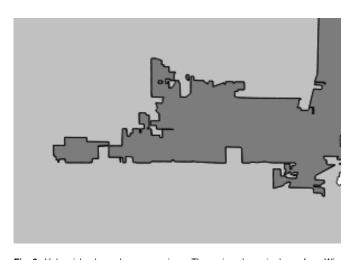


Fig. 8. Holes, islands, and narrow regions. The region shown is drawn from Wisconsin's Assembly Districts (30) and shows how bad digitization or subsequent simplification can lead to subtle data issues which may not be visually apparent without significant magnification. Note that many borders are axially-aligned, this may reflect reality, but may also be an artifact arising from discretization of input data, demarcation choices, simplification algorithms, or even our own visualization software. Regardless, axial alignment renders useless many simple geometric algorithms and invokes special cases in others; therefore, implementations of scoring algorithms require care.

District	Score Value	Diff from Mean	Score Name	Tolerance	Projection	Choice
MD03	0.28	0.48	Mercator	5000	CvxHullPTB	nochoice
MD03	0.58	0.09	Local LCC	500	ReockPTB	nochoice
NC12	0.29	0.5	Local AEA	5000	CvxHullPTB	nochoice
NC12	0.03	0.17	Mollweide	0	PolsbyPopp	choice
MD02	0.42	0.35	Robinson	5000	CvxHullPTB	nochoice
MD02	0.49	0	Local LCC	1000	ReockPTB	nochoice
FL05	1	0.52	Local AEA	5000	ReockPTB	choice
FL05	0.04	0.16	Mollweide	0	PolsbyPopp	choice
NC01	0.26	0.32	EPSG:102003	5000	Schwartzbe	nochoice
NC01	0.37	0	Gall	100	ReockPS	choice
PN07	0.47	0.31	Local LCC	5000	CvxHullPTB	nochoice
PN07	0.35	0.02	Gall	500	ReockPT	choice
TX33	0.43	0.33	Local LCC	500	CvxHullPTB	nochoice
TX33	0.25	0.12	Gall	500	ReockPT	choice
NC04	0.34	0.43	Gall	5000	CvxHullPTB	nochoice
NC04	0.15	0.14	Mollweide	50	ReockPT	nochoice
IL04	0.42	0.34	Local LCC	50	CvxHullPTB	nochoice
IL04	0.27	0.06	Robinson	1000	ReockPT	nochoice
TX35	0.36	0.41	Mollweide	5000	CvxHullPTB	nochoice
TX35	0.05	0.15	Mollweide	0	PolsbyPopp	choice

Table 1. Applied gerrymandering: abusing implementation flexibility. This table shows the choices made to produce the histograms shown in Figure 1. Recall that each of district which appeared incontrovertibly gerrymandered was paired with two histograms, one of which made the district's compactness score seem like an outlier and the other of which made it seem reasonable. The districts' scores are listed here, along with the absolute value of their difference from the mean of the distribution. The set of implementation choices made for each distribution is also shown: the compactness score, the simplification tolerance of the data, the map projection, and whether or not districts which comprised the entirety of their political superunit (districts in which a choice of boundaries was not possible) were included.

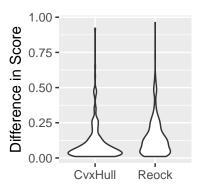


Fig. 9. Effects of constraining compactness measures using the boundaries of political superunits for the 144th Congress. The convex hull and, for the Reock score, the minimum bounding circle were cropped to state borders before being used to calculate scores. Only districts which were part of multi-district states and whose scores changed are shown: 215 for the convex hull and 320 for the Reock, of 441 total. District and state data were at 1:500,000 resolution.

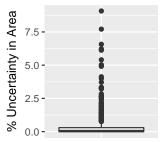


Fig. 10. Approximate percent uncertainty in area introduced by border misalignment. Each dot represents one congressional district. Areas with especially high uncertainty are usually coastal where the lower resolution data introduce significant areas of water into a district. Using the data from Figure 6, an exclusive-or on each district and state yielded areas of misalignment. Districts and states were shrunk and expanded to form border outlines which were intersected with the exclusive-or thereby limiting misalignment to border areas. The remaining area divided by the original, high-resolution areas gives the percentage. A few especially small districts ($< 10 \, \mathrm{km}^2$) were culled from the analysis as this method made the entirety of the districts uncertain.

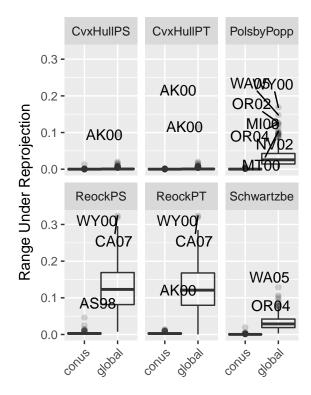


Fig. 11. Change in score between a locally-optimized projection and regionally- and globally-optimized projections for all electoral districts of the 114th Congress. Each district was projected into locally-fitted Lambert Conformal Conic and Albers Equal Area Conic projections; into conterminous US-fitted Albers Equal Area (EPSG:102003), Lambert Conformal Conic (EPSG:102004), and Equidistant Conic (EPSG:102003) projections; and into globally-fitted Mercator, Robinson, Molleweide, and Gall stereographic projections. For each district, the maximum range between any value in the local group and any value in the conterminous (CONUS) and global groups was calculated. For the conterminous projections, boxplot bodies appear as thin black lines indicating the bulk of districts experienced negligible change under different projections: in fact the 99th quantile score across all districts was 0.009. The outlier is Alaska, for which a conterminous projection should never be used due to excessive distortion. If the entire United States, including Hawaii and Alaska, needs to be processed at once, Snyder's GS50 projection (31) is a good choice as it provides < 2% scale distortion throughout the this region. Data was at 1:500,000 resolution.

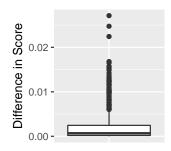


Fig. 12. Difference in Polsby-Popper scores calculated accounting for topography and assuming planar topography. Topography-inclusive area of districts was calculated using the 30 m USGS National Elevation Dataset. (23) Districts were cropped using 1:500,000 resolution boundaries from the US Census Bureau for the 114th Congress. (5) Surface area was calculated using RichDEM's implementation (24) of an algorithm by (25). Perimeter was taken as the summed length of all the cells at the edge of a district and was constant with respect to topographic considerations.

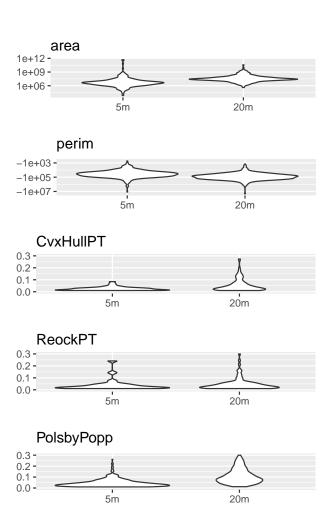


Fig. 13. Effect of resolution on compactness scores. Scores were calculated for districts from the 114th Congress at resolutions 1:500,000 (500k), 1:5,000,000 (5m), and 1:20,000,000 (20m). Score differences versus the 1:500,000 values are shown for those districts whose scores changed. Area and perimeter values are log transformed.

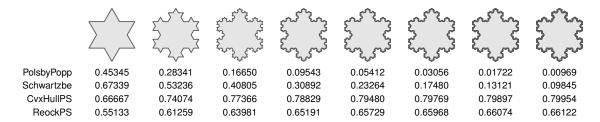


Fig. 14. The Koch Snowflake (32) shown for its first 8 levels of resolution (the 0th level is omitted). Note that at each resolution both the shape and boundary of the snowflake are visually similar, especially at higher resolutions. Despite this, levels have markedly different scores. For each increase in resolution the Polsby-Popper score decreases by 77% and the Schwartzberg score by 33%. After initial increases, the Convex Hull and Reock scores stabilize.

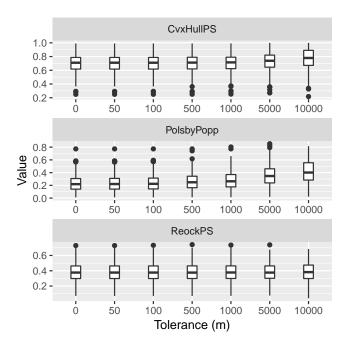


Fig. 15. Effect of polygon simplification on compactness scores. Districts from the 114th Congress were simplified by shapely (33) using a topologically-preserving algorithm from GEOS (34) with the indicated tolerances.

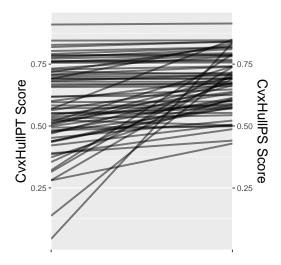


Fig. 16. Implementation affects ranking. Here, the compactness scores for the 114th Congress at 1:500,000 resolution are plotted for two different interpretations of the convex hull score. Only the 136 districts whose score changed as a result of the differing interpretations are shown.

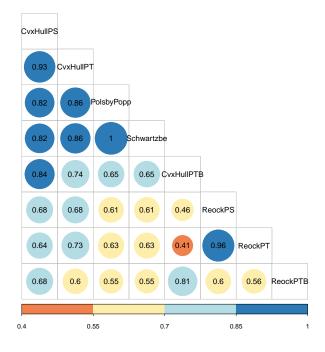


Fig. 17. Correlations of rankings. Rankings of compactness scores for the 114th Congress at 1:500,000 resolution are compared against each other using Spearman's rank correlation coefficient. A value of one indicates perfect agreement of relative rankings while a value of zero indicates no correlation.