Why Weight?
A Cluster-Theoretic Approach to Political Districting

Sam Whittle
Wesley Essig
Nathaniel S. Bottman
University of Washington
Seattle, WA

Advisor: Anne Greenbaum

Summary

Political districting has been a contentious issues in American politics over the last two centuries. Since the landmark case of *Baker v. Carr* (1962), in which the U.S. Supreme Court ruled that the constitutionality of a state's legislated districting is within the jurisdiction of a federal court, academics have attempted to produce a rigorous system for districting a state.

We propose both a modified form of classical K-means clustering and the shortest-splitline algorithm to accomplish impartial redistricting. We apply our methods to redistricting New York, and, as further examples, Texas and Colorado. Both methods use only population-density data and state boundaries as inputs and run in a feasible amount of time.

Our criteria for successful redistricting include contiguity, compactness, and sufficiently uniform population.

The K-means method produces districts similar to convex polygons, and the splitline method guarantees that the resulting districts have piecewise linear boundaries. The K-means method has the advantage of allowing seeding of the district centers. The centers of the generated districts then roughly correlate to the existing districts, by proper seeding, but the resulting boundaries are vastly simpler.

The UMAP Journal 28 (3) (2007) 301–313.

Introduction

The writers of the Constitution created the House of Representatives to be the branch of government most responsive to the people. The reality is just the opposite. Though representatives are elected every two years, almost 400 of the 435 seats are not effectively contestable, because of gerrymandering. With the immensely detailed amount of data and unlimited computing power available to politicians today, gerrymandering has been elevated to an art. With only the requirements that districts be connected and all have equal population, it is possible to pinpoint candidates and place them in a different district than their neighbors [Toobin 2003].

Though undemocratic, gerrymandering is nearly always legal (see, for instance, Backstrom [1986]) and has been used to obtain striking results. In 2002, only four incumbent representatives lost their bid for re-election—the lowest total ever [Toobin 2003]. We will argue that it is certainly true that any attempt to restructure legislative districts fairly needs to ignore the human factors that overwhelmingly determine current redistricting.

Defining a measure of compactness is essential to ensure fair districts. Both methods that we offer produce districts that at first glance are clearly simpler than the existing ones. We use the centers of the existing districts as seeds for a clustering algorithm. Thus, the new districts have some correlation to the existing districts, but their boundaries are determined in a fair manner. The core of many districts will be roughly the same, while the boundaries will be dramatically simpler. This effectively counteracts the effects of gerrymandering, without being overly difficult to implement.

Plan of Attack

Our goal is an algorithm to divide a region into k districts that satisfy some heuristic definition of *fairness*. To accomplish this, we must do the following:

- Define fairness and simplicity.
- State assumptions and constraints.
- Define metrics for comparing algorithms.

Defining Simplicity

We say that district A is *simpler* than district B if A is contiguous and more compact than B.

- **Contiguity.** A district is *contiguous* if it is arcwise-connected; that is, if one can travel from any point a to any other point b in A while remaining entirely within A. If A contains regions separated by bodies of water, A is contiguous if all regions are connected by water and each region is arcwise-connected.

- **Compactness.** Intuitively, a district is compact if it does not meander excessively. This is a hard concept to formalize; many authors give only a hasty definition, and some even argue that compactness is ambiguous to the point of being irrelevant. Nonetheless, we attempt a suitable definition.

Towards a Suitable Definition of Compactness

Young [1988] gives compelling reasons for abandoning all previous definitions of compactness (see the **Appendix**). But Young does not consider the following adjusted version of the Schwartzberg Test, which is alluded to in Garfinkel and Nemhauser [1970]:

Definition 1 *District A is more compact than district B if*

$$\frac{4\pi \text{ Area}_A}{(\text{Perimeter}_A)^2} > \frac{4\pi \text{ Area}_B}{(\text{Perimeter}_B)^2}.$$

We call the quantity 4π Area / Perimeter^2 *the* compactness quotient.

For a circle of radius r, this ratio is equal to 1. It is well-known that the shape with the largest ratio of area to squared perimeter is the circle (see, for instance, Folland [2002]) so the compactness quotient is between 0 and 1.

As seen in **Figure 1**, a compactness quotient of 0.13 is visually quite bad. Using the fact given in Bourke [1988] that the area of a non-self-intersecting closed N-gon (with the kth vertex in counterclockwise order equal to (x_k, y_k)) is equal to

$$\tfrac{1}{2}\sum_{i=1}^{N-1}(x_i y_{i+1} - x_{i+1} y_i),$$

we calculated the compactness quotients of several actual districts by approximating their boundaries by piecewise linear segments. Two of New York's more sprawling districts have compactness quotients 0.097 and 0.101 (**Figure 2**)—even worse than the gerrymander in **Figure 1**! The two most compact districts in New York, the 26th and 21st, have compactness quotients 0.406 and 0.498.

We decide that the mean for any state should be at least 0.6, so that the average district would be better than the best current districts in New York. Furthermore, we insist that 0.25 should be more than 2 standard deviations from the mean. It is not possible to require that all districts be greater than 0.25, since several districts inevitably have most of their border coincide with the state border.

Defining Fairness

Almost all unfairness occurs when political and social measures factor into redistricting decisions. Concentrating supporting voters in a single district,

Figure 1. The compactness quotients of the circle, square, and gerrymander are 1, $\pi/4 \approx 0.79$, and $23\pi/576 \approx 0.13$, respectively.

Figure 2. Current New York districts 8 and 28 (in dark shade), with compactness quotients of 0.097 and 0.101. Source: U.S. Department of the Interior [2007].

diluting opposing voters over several districts, placing two incumbents in the same district and forcing them to run against each other, and isolating minorities (see Toobin [2003] and Hayes [1996]) are all the result of districting being controlled by those who attempt to skew voting patterns. In general:

- *Unfair districting stems from either human biases or poorly designed algorithms.*

Our computer simulations use only population density and the boundary of the state, so the determination of districts is completely unbiased. While a district may be unfair on a local scale, in that it divides up a community with a common interest—for instance, a community of apple-growers may be split between two districts—on the national scale, such imbalances will even out. Because of this, there will be no pathological examples of disproportionate representation.

Applying the Theory of Data Clustering

Data clustering is classifies observations (or objects) into groups. The main benefits of a cluster-theoretic algorithm are:

- *Data clustering often reveals an internal structure that may not have been initially apparent.*
- *It is easier to work with a small number of clusters than with a large number of raw data points.*

The philosophy of data clustering is to divide data into a (not necessarily fixed) number of clusters, with the elements in a cluster somehow *similar*. Data clustering is often applied to problems that deal with a large number of variables, and it is usually very difficult to determine the "proper" way to cluster data [Afifi and Clark 1984]. We apply data clustering in the following way:

- *Split the state into small, discrete units.* Our units correspond to geographic locations of interpolated census population measurements [Center for International Earth Science Information Network 2007].
- *Determine some partition of these units into clusters.* Note that the only variables present are the location and population of each unit.

After defining a method for ordering the preference of cluster arrays, we might suppose we are done with the problem: All that is left is to look at all possible cluster arrays and choose the best one. However, this turns out not to be feasible. Abramowitz and Stegun [1968] prove that the number of ways of sorting n observations into m groups is a Stirling number of the second kind:

$$S_m^{(n)} = \frac{1}{m!}\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k}k^n.$$

For instance, there are more than 10^{15} ways to sort 25 objects into 5 groups. We need an algorithmic process to determine an appropriate array of clusters.

Cluster-Theoretic Districting

The K-means Algorithm

Standard Algorithm

The K-means algorithm is an iterative method for data clustering [Shapiro and Stockman 2001]. Let $D = \{\mathbf{x}_j\}_{j=1}^{N} \subset \mathbb{R}^n$ be the data to be clustered, and let $S = \{\mathbf{s}_j\}_{j=1}^{K}$ be a set of seeds. Suppose that we desire to partition D into K clusters; let the ith cluster be denoted by C_i. Associate to the ith cluster a *geographical center*, denoted by $\mathbf{c}_i$. Given a distance function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, the K-means algorithm proceeds as follows.

- **Initialization:** For all C_i, let $\mathbf{c}_i = \mathbf{s}_i$

- **Iteration:**
  - **Assign points to clusters:** For all $\mathbf{x} \in D$, associate $\mathbf{x}$ to a cluster C_i whose center $\mathbf{c}_i$ minimizes $f(\mathbf{x}, \mathbf{c}_i)$.
  - **Update cluster centers:** Redefine $\mathbf{c}_i = \left(\sum_{\mathbf{x}\in C_i} \mathbf{x}\right) / \left(\sum_{\mathbf{x}\in C_i} 1\right)$.
  - **Repetition:** If updating cluster centers changes at least one cluster center, repeat the iteration step. Otherwise, stop.

Weighted Algorithm

To generate districts of appropriate population, we add a weighting system to the standard algorithm. Let each cluster correspond to a legislative district. Let $D = \{\mathbf{x}_j\}_{j=1}^{N} \subset \mathbb{R}^2$ be the set of census coordinates. Thus, $x \in D$ corresponds to the position of a population measurement. Define a population function $p : D \to \mathbb{R}$ such that p_i is the population at the coordinates specified by $\mathbf{x}_i$. A cluster C_j is defined by its points $\mathbf{x} \subset \mathbb{R}^2$, its center $\mathbf{c}_j \in \mathbb{R}^2$, and some *weight* α_j. Define f to be the Euclidean distance function in $\mathbb{R}^2$. Our weighted K-means algorithm proceeds as follows:

- **Initialization:** Using the *standard* K-means algorithm, assign points to clusters and centers to appropriate positions.

- **Iteration:**
  - **Assign points to clusters:** For all $\mathbf{x} \in D$, associate $\mathbf{x}$ to a cluster C_i whose center $\mathbf{c}_i$ minimizes $f(\mathbf{x}, \mathbf{c}_i)$.
  - **Update cluster centers:** Redefine $\mathbf{c}_i = \left(\sum_{\mathbf{x}\in C_i} p_i\mathbf{x}\right) / \left(\sum_{\mathbf{x}\in C_i} p_i\right)$.
  - **Update cluster weights:** Redefine

$$\alpha_j = g\left(\sum_{\mathbf{x}_i\in C_j} p_i\alpha_j\right),$$

where g is defined below.

- **Repetition:** If the properties of the clusters are within tolerance, stop. Otherwise, repeat the iteration step.

By adjusting the weights, we control the growth or decay of the clusters. If the weight of a cluster increases, data points are more likely to be grouped in other clusters. Similarly, decreasing the weight helps to increase the population of a cluster. Thus, the weight function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is crucial in the performance of the algorithm. We define:

$$g(p, w) = w\sqrt{\frac{i}{i_0}} + w \cdot \frac{p}{p_0} \cdot \sqrt{1 - \frac{i}{i_0}},$$

where i is the current iteration, i_0 is the maximum number of iterations, and p_0 is the desired population for each cluster. Towards the beginning of the algorithm, i/i_0 is small, causing the second term to dominate the weight function. As i increases, the weight fluctuates less because the first term begins to dominate. This formula enables the weights to change rapidly at the beginning of the iterative process, causing the clusters to vary greatly between iterations. However, by the end of the algorithm, the weights do not change as readily, allowing stabilization over an optimal clustering. This is somewhat similar to simulated annealing, where initial negative actions allow the algorithm to escape local optima and the probability that a negative action is taken decreases over time.

Splitline Algorithm [Smith 2007]

Method

The idea behind the shortest splitline algorithm is quite simple:

- Start with the number of districts for the state. Divide that number in two as evenly as possible, using integers.
- Find the *shortest* line that divides the state into two parts such the ratio of their populations is the same as the ratio determined in the previous step.
- Repeat this process recursively on the subdivided parts until the number of parts is the same as the number of districts. At every step, the division is just a line, so the resulting districts have piecewise linear boundaries. Using the *shortest* line ensures that the districts will have a good compactness quotient.

Demonstration

Figure 3 is a demonstration of the splitline algorithm creating 5 districts from 15 people; there need to be 3 people in each district. The ratio 3:2 is the most balanced integer ratio that 5 can be divided into. At step 1, the algorithm divides the state into two regions with 9 and 6 people, the correct ratios for 3 and 2 districts. At step 2, it acts recursively on the 2 subdivisions. Thus, the region with 6 people is divided into regions with 3 each, with no more subdivision needed. The other region is divided into regions with 6 and 3 people. At the third and final step, the last region is split in two and the process is complete. By using the shortest line at each step, none of the shapes ends up with an unsatisfactory compactness quotient.

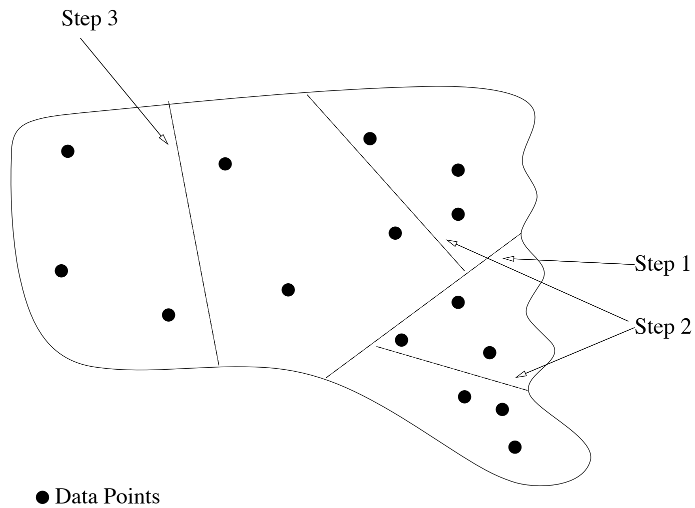

Figure 3. An illustration of the splitline method.

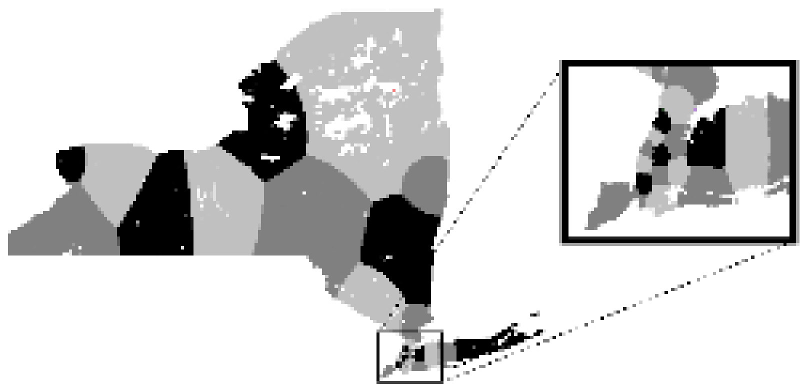

Figure 4. A proposed redistricting of New York, using the K-means algorithm.

Districting of New York State

K-means Algorithm

The results given by the K-means algorithm in **Figure 4** are generally quite good. Traditionally, when applying cluster-theoretic algorithms to redistricting, it is common practice to split off any regions with particularly high population density and apply the algorithm to those regions separately (see, for instance, Garfinkel and Nemhauser [1970]). This was not needed for the K-means algorithm: Even though the maximum population density of New York City is roughly 2,000 times the mean population density of the state of New York, the K-means algorithm produced results within our tolerance levels.

To confirm that the weighted K-means algorithm is an effective aid for determining districts, we also used the algorithm to redistrict Texas too. Texas is a good choice because it is large and contains a variety of population densities.

The K-means algorithm worked overall well with only a few districts outside our target tolerance.

Splitline Algorithm

To obtain results within our desired tolerance, it was necessary to calculate the districts of New York City separately from the remainder of the state. A limitation of the splitline algorithm is that it does not guarantee contiguity of districts (see **Figure 5**). However it produces contiguous (and, furthermore, convex) districts for a convex state.

Conclusions

We conclude that both the K-means algorithm and the splitline algorithm are viable methods for fair and simple redistricting. K-means produced much better results for New York: The greatest value of

$$\max_{\text{alldistricts}} \left(\frac{1 - \text{cluster population}}{\text{target population}} \right),$$

is no more than 2.5%. As an interesting note, while the unweighted K-means method clusters data into regions with piecewise-linear boundaries, inclusion of the weight function effectively rounds the boundaries of the produced districts. These rounder districts have superb compactness coefficients. K-means also has a visually appealing output and meets all other criteria.

The splitline algorithm results are not quite as satisfactory; however, we believe that this is a result of our implementation and not of the algorithm itself. Even our flawed version of splitline produced districts simpler than the current districts in New York. Our implementation could achieve districts with either even population or high compactness coefficients, but not both simultaneously. It is also difficult to enforce the contiguity requirement in regions with a highly irregular border. When the splitline algorithm is applied to states with convex boundaries, there are no discontiguities; furthermore, every district is convex. In the case of simple states, the splitline algorithm works well, perhaps even better than the K-means algorithm. Its intuitive simplicity is also likely to make shortest-splitline more appealing to the public.

Both K-means and splitline are deterministic: that is, when each algorithm is applied to a fixed problem, and all parameters are constant, the final result is unique. Some authors have expressed the opinion that any good districting algorithm is deterministic [Hayes 1996]. There is one human element involved in the K-means algorithm: The choice of seeds is made, in some sense, subjectively, by the person implementing the algorithm. This factor could be completely eliminated by randomly picking the seeds, but this is not the most desirable solution. Random seeds can produce solutions far from the global

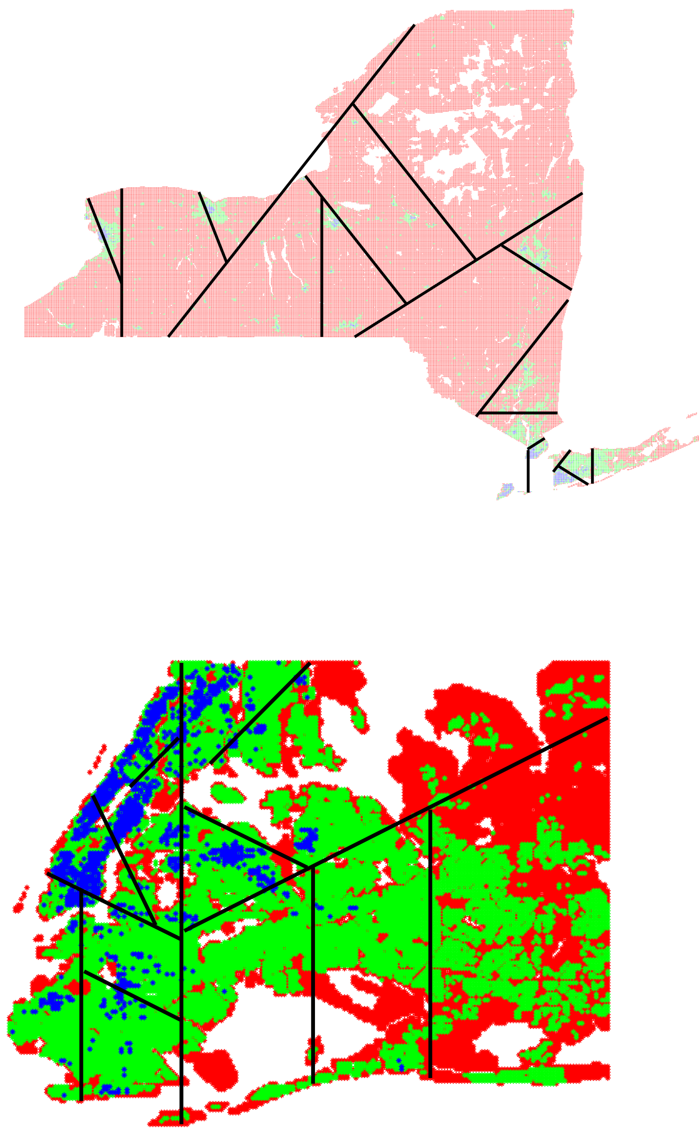

Figure 5. A proposed redistricting of New York, using the splitline algorithm and calculating the districts within New York City separately from the remaining districts.

optimum of the optimization function and can require many more iterations to get an answer within a given tolerance level. The natural choice is to use the approximate centers of existing districts as seeds. At first, this may seem contrary to our goal of reversing the effects of gerrymandering. A closer analysis of gerrymandering shows that this is not true. Gerrymandering relies on intricately carving districts based on data that are invisible to our algorithm—say, ethnicity, income level, or political affiliation.

The K-means algorithm clearly performs better on more-complex data sets. The splitline algorithm should not be abandoned, but our final recommendation is that

> *The K-means algorithm quickly and deterministically produces districts that are simple and fair, and applying this algorithm would produce a drastic improvement over current districts in any state.*

References

Afifi, A.A., and Virginia Clark. 1984. *Computer-aided Multivariate Analysis*. Belmont, CA: Lifetime Learning Publications.

Anderberg, Michael R. 1973. *Cluster Analysis for Applications*. New York: Academic Press.

Abramowitz, M., and I.A. Stegun. 1968. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Washington, DC: U.S. Government Printing Office.

Baker v. Carr, 369 U.S. 186 (1962).

Backstrom, Charles H. 1986. The Supreme Court prohibits gerrymandering: A gain or a loss for the states? *The State of American Federalism* 17 (3): 1–33.

Bourke, Paul. 1988. Calculating the area and centroid of a polygon. `http://local.wasp.uwa.edu.au/~pbourke/geometry/polyarea/` . Accessed 11 February 2007.

Center for International Earth Science Information Network. 2007. Socioeconomic data and applications center. `http://sedac.cisesin.columbia.edu/` . Accessed 8 February 2007.

Folland, Gerald B. 2002. *Advanced Calculus*. Upper Saddle River, NJ: Prentice Hall.

Garfinkel, R.S., and G.L. Nemhauser. 1970. Optimal political districting by implicit enumeration techniques. *Management Science* 16 (8): B495–B508.

Hayes, Brian. 1996. Machine politics. *American Scientist* 84 (6): 522–526.

Shapiro, Linda G., and George C. Stockman. 2001. *Computer Vision.* Upper Saddle River, NJ: Prentice Hall.

Smith, Warren B. 2007. Examples of our unbiased district-drawing algorithm in action. `http://rangevoting.org/RangeVoting.html`. Accessed 12 February 2007.

Toobin, Jeffrey. 2003. The great election grab. *New Yorker* 79 (38) (8 December 2003): 63–80.

U.S. Department of the Interior. 2007. National Atlas of the United States. Printable Maps. Congressional Districts—110th Congress. `http://nationalatlas.gov/printable/congress.html`.

Young, H.P. 1988. Measuring the compactness of legislative districts. *Legislative Studies Quarterly* 13 (1): 105–115.

Appendix: Definitions of Compactness

The following definitions of compactness are said in Young [1988] to be representative of those definitions favored in past and present scholarship.

As discussed by Young [1988], the following definitions of compactness are often used or cited in the literature.

- **Visual test.** A district is more compact if it appears to be more compact.
- **Roeck test.** Find the smallest circle containing the district and take the ratio of the district's area to that of the circle. This ratio is always between 0 and 1; the closer to 1, the more compact the district.
- **Schwartzberg test.** Construct the adjusted perimeter of the district by connecting with straight lines points on the district boundary where three or more constituent units (*i.e.*, census tracts) from any district meet. Divide the length of the adjusted perimeter by the perimeter of a circle with area equal to that of the district.
- **Length-width test.** Find a rectangle enclosing the district and touching it on all four sides, such that the ratio of length to width is a maximum. The closer the ratio to 1, the more compact the district.
- **Taylor's test.** Construct the adjusted perimeter of the district by connecting by straight lines those points on the district boundary where three or more constituent units (*i.e.*, census tracts) from any district meet. At each such point the angle formed is "reflexive" if it bends away from the district and "non-reflexive" otherwise. Subtract the number of reflexive from the number of non-reflexive angles and divide by the total number of angles. The resulting number is always between 0 and 1; the closer to 1, the more compact the district.

- **Moment of Inertia test.** Locate the geographical center $\mathbf{c}_i$ of each census tract i in the district. Select an arbitrary point $\mathbf{x}$ and calculate the square of the distance from $\mathbf{x}$ to $\mathbf{c}_i$, multiplied by the population of tract i. The sum of these numbers is the district's moment of inertia about $\mathbf{x}$. The point that gives the minimum moment of inertia is the center of gravity of the district. The smaller the moment of inertia about the center of gravity, the more compact the district.

- **Boyce-Clark test.** Determine the center of gravity of the district and measure the distance from the center to the outside edges of the district along equally-spaced radial lines. Compare the percentages by which each radial distance differs from the average radial distance, and find the average of the percentage deviations over all radials. The closer the result is to 0, the more compact is the district.

- **Perimeter test.** Find the sum of the perimeters of all the districts. The shorter the total perimeter, the more compact the districting plan.

Members of both Outstanding teams from the University of Washington in the Gerrymander Problem. Top row, from left: Sam Whittle, Aaron Dilley, Sam Burden, advisor Jim Morrow; bottom row: Lukas Svec, Wesley Essig, Nate Bottman. Not shown: Advisor Anne Greenbaum.