Applying Voronoi Diagrams to the Redistricting Problem

Lukas Svec
Sam Burden
Aaron Dilley
University of Washington
Seattle, WA

Advisor: James Allen Morrow

Summary

Gerrymandering is an issue plaguing legislative redistricting. We present a novel approach to redistricting that uses *Voronoi* and population-weighted *Voronoiesque* diagrams. Voronoi regions are contiguous, compact, and simple to generate. Regions drawn with Voronoiesque diagrams attain small population variance and relative geometric simplicity.

As a concrete example, we apply our methods to partition New York State. Since New York must be divided into 29 legislative districts, each receives roughly 3.44% of the population. Our Voronoiesque diagram method generates districts with an average population of $(3.34 \pm 0.74)\%$.

We discuss possible refinements that might result in smaller population variation while maintaining the simplicity of the regions and objectivity of the method.

Introduction

Defining Congressional districts has long been a source of controversy in the U.S. Since the district-drawers are chosen by those in power, the boundaries are often created to influence future elections by grouping an unfavorable minority demographic with a favorable majority; this process is called *gerrymandering*. It is common for districts to take on bizarre shapes, spanning slim sections of multiple cities and criss-crossing the countryside in a haphazard fashion. The only lawful restrictions on legislative boundaries stipulate that districts must

---

The UMAP Journal 28 (3) (2007) 313–329.

contain equal populations, but the makeup of the districts is left entirely to the district-drawers.

In the United Kingdom and Canada, districts are more compact and intuitive. The success of these countries in mitigating gerrymandering is attributed to turning over boundary-drawing to nonpartisan advisory panels. However, these independent commissions can take two to three years to finalize a new division plan, calling their effectiveness into question. It seems clear that the U.S. should establish similar unbiased commissions yet make some effort to increase the efficiency of these groups. Accordingly, our goal is to develop a small toolbox to aid in the redistricting process. Specifically, we create a model that draws legislative boundaries using simple geometric constructions.

Current Models

The majority of methods for creating districts fall into two categories: ones that depend on a current division arrangement (most commonly counties) and ones that do not depend on current divisions. Most fall into the former category. By using current divisions, the problem is reduced to grouping these divisions in a desirable way using a multitude of mathematical procedures. Mehrotra et al. [1998] uses graph partitioning theory to cluster counties to total population variation of around 2% of the average district size. Hess et al. [1965] use an iterative process to define population centroids, use integer programming to group counties into equally populated districts, and then reiterate the process until the centroids reach a limit. Garfinkel and Nemhauser [1970] use iterative matrix operations to search for district combinations that are contiguous and compact. Kaiser begins with the current districts and systematically swaps populations with adjacent districts [Hamilton 1966]. All of these methods use counties as their divisions since they partition the state into a relatively small number of sections. This is necessary because most of the mathematical tools they use become slow and imprecise with many divisions. (This is the same as saying they become unusable in the limit when the state is divided into more continuous sections.) Thus using small divisions, such as zipcodes (which on average are one-fifth the size of a county in New York), becomes impractical.

The other category of methods is less common. Forrest's method continually divides a state into halves while maintaining population equality until the required number of districts is satisfied [Hamilton 1966; Hess et al. 1965]. Hale et al. create pie-shaped wedges about population centers; this creates homogeneous districts that all contain portions of a large city, suburbs, and less populated areas [Hamilton 1966]. These approaches are noted for being the least biased, since their only consideration is population equality and they do not use preexisting divisions. Also, they are straightforward to apply. However, they do not consider any other possibly important considerations for districts, such as geographic features of the state or how well they encompass cities.

Developing Our Approach

Since our goal is to create new methods, we focus on creating district boundaries independently of current divisions. It is not obvious why counties are a good beginning point for a model: Counties are created in the same arbitrary way as districts, so they may also contain biases, since counties are typically not much smaller than districts. Many of the division-dependent models end up relaxing their boundaries from county lines so as to maintain equal populations, which makes the initial assumption of using county divisions useless and also allows for gerrymandering if the relaxation method is not well regulated.

Treating the state as continuous (i.e., without pre-existing divisions) does not lead to any specific approach. If the Forrest and Hale et al. methods are any indication, we should focus on keeping cities within districts and introduce geographical considerations. (These conditions do not have to be considered if we treat the problem discretely, because current divisions, like counties, are probably dependent on prominent geographical features.)

> **Goal:** *Create a method for redistricting by treating the state continuously. We require the final districts to contain equal populations and to be contiguous. Additionally, the districts should be as simple as possible and optimally take into account important geographical features.*

Notation and Definitions

- **Contiguous:** A set R is contiguous if it is pathwise-connected.
- **Compactness:** One way to look at compactness is the ratio of the area of a bounded region to the square of its perimeter:

  $$C_R = \frac{A_R}{p_R^2} = \frac{1}{4\pi}\,\mathcal{Q},$$

  where C_R is the compactness of region R, A_R is the area, p_R is the perimeter and $\mathcal{Q}$ is the isoperimetric quotient. We do not explicitly use this equation, but we keep this idea in mind when we evaluate our model.
- **Simple:** Simple regions are compact and convex.
- **Voronoi diagram:** A partition of the plane with respect to n nodes such that points are in the same region with a node if they are closer to that node than to any other point.
- **Generator point:** A node of a Voronoi diagram.
- **Degeneracy:** The number of districts represented by one generator point.
- **Voronoiesque diagram:** A variation of the Voronoi diagram based on equal masses of the regions.

- **Population center:** A region of high population density.

Theoretical Evaluation of our Model

How we analyze our model's results is tricky, since there is disagreement in the literature on key issues. *Population equality* is well-defined. By law, the populations within districts have to be the same to within a few percent of the average population per district.

Creating a successful redistricting model also requires *contiguity*. In accordance with state law, districts need to be pathwise connected. This requirement is meant to maintain locality and community within districts. It does not, however, restrict districts from including islands if the island's population is below the required population level for a district.

Finally, there is a desire for the districts to be *simple*. There is little to no agreement on this characteristic, and the most common terminology for this is *compact*. Taylor [1973] defines simple as a measure of divergence from compactness due to indentation of the boundary and gives an equation relating the nonreflexive and reflexive interior angles of a region's boundary. Young [1988] provides seven more measures of compactness. The *Roeck* test is a ratio of the area of the largest inscribable circle in a region to the area of that region. The *Schwartzberg* test takes ratio between the adjusted perimeter of a region to a the perimeter of a circle whose area is the same as the area of the region. The *moment of inertia* test measures relative compactness by comparing "moments of inertia" of different district arrangements. The *Boyce-Clark* test compares the difference between points on a district's boundary and the center of mass of that district, where zero deviation of these differences is most desirable. The *perimeter test* compares different district arrangements buy computing the total perimeter of each. Finally, there are the *perimeter* test (shorter = more compact) and the *visual* test. This test decides simplicity based on intuition [Young 1988].

Young [1988] notes that "a measure [of compactness] only indicates when a plan is more compact than another." Thus, not only is there no consensus on how to analyze redistricting proposals, it is also difficult to compare them.

Finally, we remark that the above list only constrains the shape of generated districts without regard to any other potentially relevant features—for example, how well populations are distributed or how well the new district boundaries conform with other boundaries, like counties or zipcodes. Even with this short list, we are not in a position to define simplicity rigorously. What we can do, however, is identify which features of proposed districts which are simple and which are not. This is in line with our goal, since this list can be provided to a districting commission who decide how relevant those simple features are. *We do not explicitly define simple, we loosely evaluate simplicity based on overall contiguity, compactness, convexity, and intuitiveness of the model's districts.*

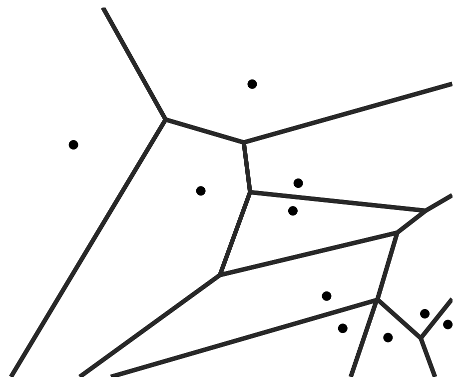


Figure 1. Illustration of Voronoi diagram generated with Euclidean metric. Note the compactness and simplicity of the regions.

Method Description

Our approach depends heavily on using Voronoi diagrams. We begin with a definition, its features, and motivate its application to redistricting.

Voronoi Diagrams

A Voronoi diagram is a set of polygons, called *Voronoi polygons*, formed with respect to n generator points in the plane. Each generator p_i is contained within a Voronoi polygon $V(p_i)$ with the following property:

$$V(p_i) = \{q \mid d(p_i, q) \leq d(p_j, q), \qquad i \neq j,$$

where $d(x, y)$ is the distance from point x to point y. That is, the set of all such q is the set of points closer to p_i than to any other p_j. Then the diagram is given by (see **Figure 1**) $\mathbf{V} = \{V(p_1), \ldots, V(p_n)\}$.

Of the many possible metrics, we use the three most common:

- Euclidean metric: $d(p, q) = \sqrt{(x_p - x_q)^2 + (y_p - y_q)^2}$
- Manhattan metric: $d(p, q) = |x_p - x_q| + |y_p - y_q|$
- Uniform metric: $d(p, q) = \max\{|x_p - x_q|, |y_p - y_q|\}$

Useful Features of Voronoi Diagrams

- The Voronoi diagram for a set of generator points is unique and produces polygons, which are path-connected.
- The nearest generator point to p_i determines an edge of $V(p_i)$
- The polygonal lines of a Voronoi polygon do not intersect the generator points.

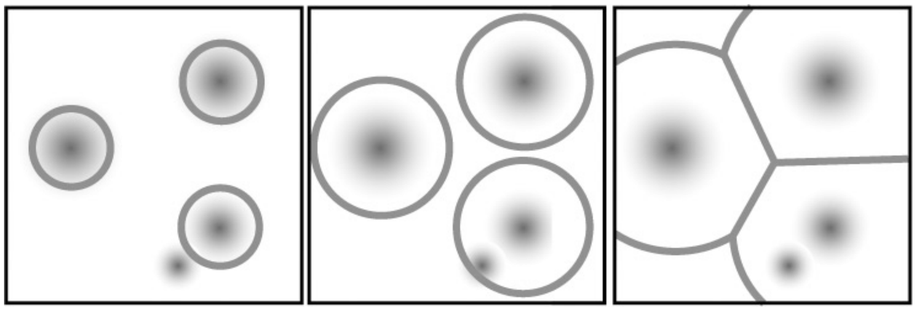


Figure 2. The process of "growing" a Voronoiesque diagram with respect to a population density. Only three three generator points are used. Figures from left to right iterate with time.

- When working in the Euclidean metric, all regions are convex.

The first property tells us that regardless of how we choose the generator points, we generate unique regions. The second property implies that each region is defined in terms of the surrounding generator points, while in turn each region is relatively compact. These features of Voronoi diagrams effectively satisfy two of the three criteria for partitioning a region: contiguity and simplicity.

Voronoiesque Diagrams

The second method that we use to create regions is a modification of Voronoi diagrams; we call them *Voronoiesque diagrams*. One way to visualize the construction of Voronoi diagrams is to imagine shapes (determined by the metric) that grow radially outward at the same constant rate from each generator point. In the Euclidean metric, these shapes are circles; in the Manhattan metric, they are diamonds; in the Uniform metric, they are squares. As the regions intersect, they form the boundary lines for the regions.We define Voronoiesque diagrams to be the boundaries defined by the intersections of these growing shapes. The fundamental difference between Voronoi and Voronoiesque diagrams is that Voronoiesque diagrams grow the shapes radially outward at a constant rate like Voronoi diagrams. Their radial growth is defined with respect to a real function on a subset of $\mathbb{R}^2$ (representing the space on which the diagram is being generated) (see **Figure 2**).

More rigorously, we define a Voronoi diagram to be the intersections of the $\mathcal{V}_i^{(t)}$s, where $\mathcal{V}_i^{(t)}$ is the Voronoiesque region generated by the generator point p_i at iteration t. With every iteration,

$$\mathcal{V}_i^{(t)} \subset \mathcal{V}_i^{(t+1)}$$

and

$$\int_{\mathcal{V}_i} f(x,y)\,dA = \int_{\mathcal{V}_j} f(x,y)\,dA$$

for all $\mathcal{V}_i$, $\mathcal{V}_j$ representing different regions. The manner in which the $\mathcal{V}_i^{(t)}$s are grown radially from one iteration to the next is determined by the metric used.

What's useful about Voronoiesque diagrams is that their growth can be controlled by requiring that the area under the function f for each region is the same at every iteration. In our model, we take f to be the population distribution of the state. Thus, the above equation is a statement of population equality. Also, when f is constant, the regions grow at a constant rate until they intersect, so the resulting diagram is Voronoi.

The final consideration for using Voronoiesque diagrams is determining the location for generator points.

Determining Generator Points

Our first approach is to place generator points at the m largest set of peaks that are well distributed throughout the state, where m is the required number of districts. Doing this keeps larger cities within the boundaries that we will generate with Voronoi or Vornoiesque diagrams and makes sure that the generator points are well dispersed throughout the state. One problem that arises is when a city is so large that for districts to hold the same number of people, a city must be divided into districts. A perfect example is New York City, which contains enough people for 13 districts.

Our second approach is to choose the largest peaks in the population distribution and assign each peak a weight. The weight for each generator point is the number of districts into which the population surrounding that peak needs to be divided. We call this weight the *degeneracy* of the generator point. We begin assigning generator points to the highest populated cities with their corresponding degeneracies until the sum of all the generator points and their respective degeneracies equals m. As we will see with New York, this method works well.

Creating Regions

Once we have our generator points, we generate our diagrams. First, generate the diagram using the given generator points. Within each generated region, called a *subdivision*, with some degeneracy r, create r new generator points within that subdivision by finding the r largest population density peaks and create another diagram (**Figure ??**).

Redistricting in New York State

We begin by explaining our choice of generator points at population centers, since these points uniquely determine a Voronoi diagram for the state. Then we describe several methods for generating Voronoi and Voronoiesque diagrams

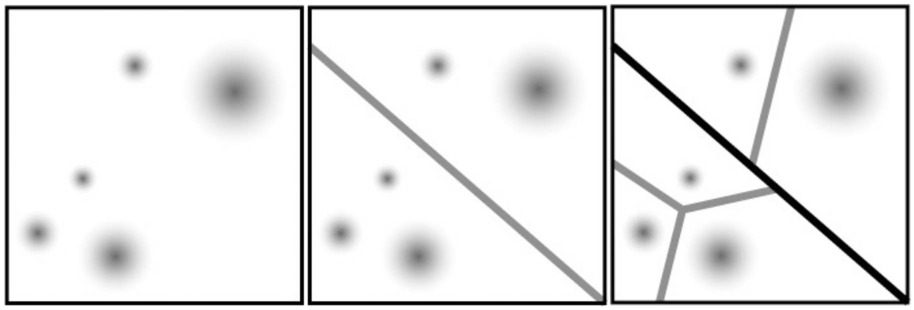


Figure 3. Creating divisions by first subdividing the map. Left: Population density distribution of hypothetical map with five desired districts. Middle: A subdivision of the map into two regions generated from two unshown generator points. Right: Final division of each subregion from the middle figure into desired final divisions.

from these points and present the corresponding results. Finally, we discuss how to use these diagrams to create political districts for New York.

Population Density Map

The U.S. Census Bureau maintains a database of census tract-level population statistics; when combined with boundary data for each tract, it is possible to generate a density map with a resolution no coarser than 8,000 people per region [**?**]. Unfortunately, our limited experience with the Census Bureau's data format prevented us from accessing these data directly, and we contented ourselves with a 792-by-660 pixel approximation to the population density map [Irwin 2006].

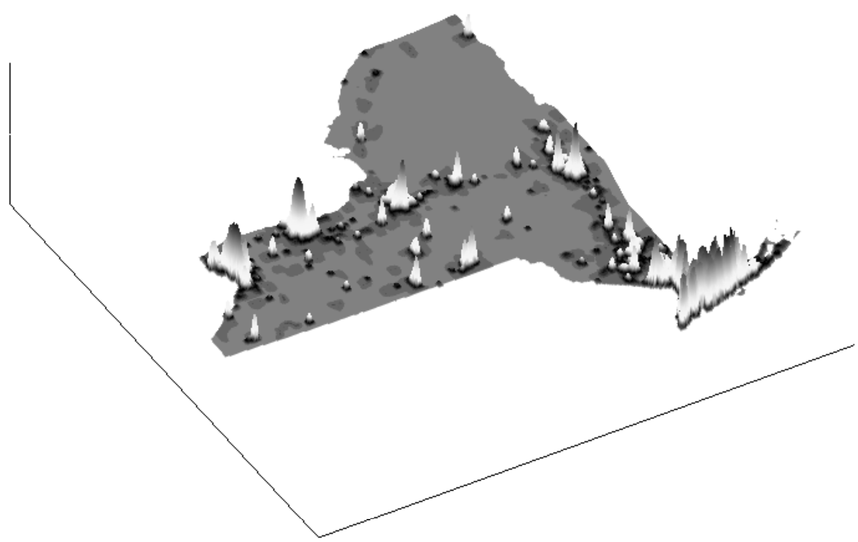
We loaded this raster image into Matlab and generated a surface plot where height represents population density. To remove artifacts introduced by using a coarse lattice representation for finely-distributed data, we applied a 6-pixel moving average filter to the density map. The resulting population density is shown in **Figure 6**.

Limitations of the Image-Based Density Map

The population density image that we used yields a density value, for every one-third of a square mile, from the following set (measured in people per square mile):

$$\{0,\ 10,\ 25,\ 50,\ 100,\ 250,\ 500,\ 1000,\ 2500,\ 5000\}\,.$$

This provides a decent approximation for regions with a density smaller than 5,000 people/sq. mi. However, the approximation will break down at large population centers; New York City's average population density is 26,403 people/sq. mi. [Wikipedia 2007].



b. Angled View: Clearer view of population distribution over New York.

Figure 4. New York State population density map. Data from 792-by-660 pixel raster image at Irwin [2006]; color and height indicate the relative population density at each point.

Selecting Generator Points

Our criteria for redistricting stipulate that the regions must contain equal populations. New York State must be divided into 29 congressional districts, so each region must contain 3.45% of the state's population. Since a state's population is concentrated primarily in a small number of cities, we use local maxima of the population density map as candidates for generator points.

If we were to simply choose the highest 29 peaks from the population density map as generator points, they would be contained entirely in the largest population centers and would not be distributed evenly over state. For the largest population centers, we assign a single generator point with a degeneracy. After all the generator points have been assigned, we generate a Voronoi diagram for the state. Then, we return to the regions with degenerate generator points and repeat the process of finding generator points for that region and generate a Voronoi diagram from them. See **Figure 3** for an illustration of the decomposition before and after subdivision.

We subdivide the region around New York City into 12 subregions, Buffalo into 3 subregions, and Rochester and Albany into 2 subregions. This roughly corresponds to the current allocation, where New York City receives 14 districts, Buffalo gets 3, and Rochester and Albany both get roughly 2. New York City's population is underestimated, since the average density there far exceeds our data's density range. With a more detailed data set, our method would call for the correct number of subdivisions.

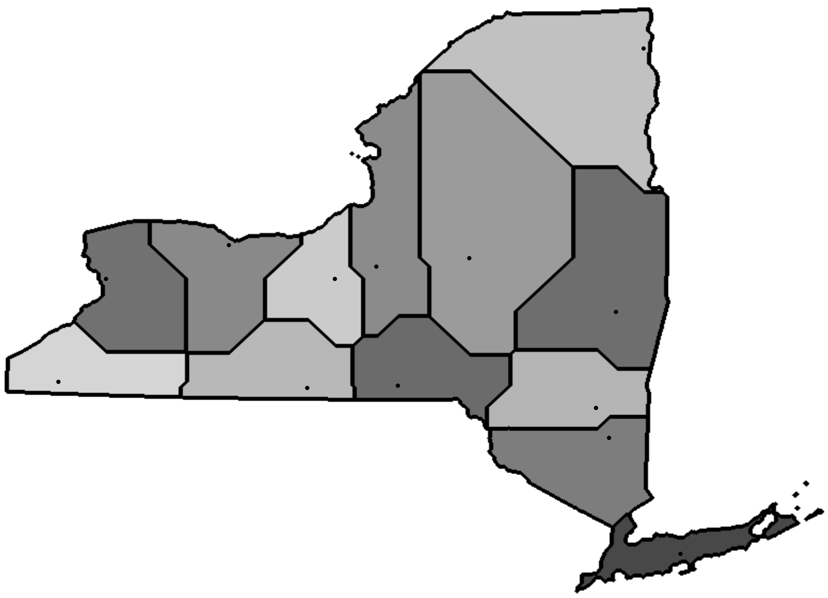
Applying Voronoi Diagrams to NY

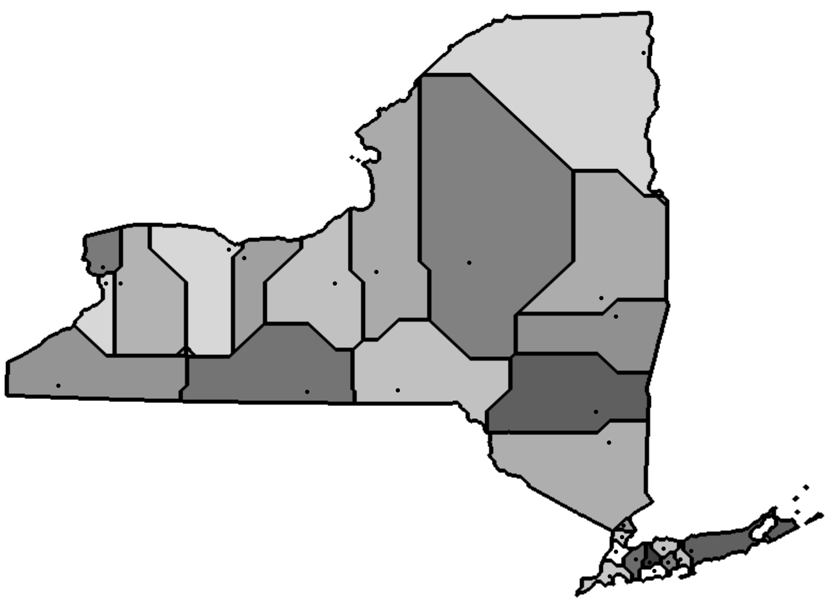
The simplest method that we consider for generating congressional districts is to generate the discrete Voronoi diagram from a set of generator points. We achieve this by iteratively "growing" regions outward with the function f constant. That way the regions grow at a constant rate, and hence the resulting diagram is Voronoi. A region's growth is limited at each step by its radius in a certain metric; we considered the Euclidean, Manhattan, and Uniform metrics. Once the initial diagram has been created, a new set of generator points for dense regions are chosen and those regions are subdivided using the same method. Unrefined decompositions can be seen in **Figure 13**.

Each metric produces a relatively simple decomposition of the state, though the Manhattan metric has simpler boundaries and yields a slightly smaller population variance between regions.

Applying Voronoiesque Diagrams to NY

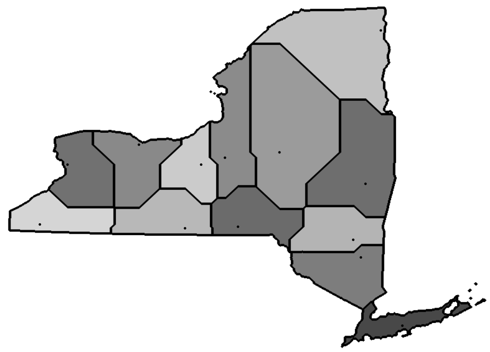
Though our simple Voronoi diagrams produced simple regions with a population mean near the desired value, the population variance between regions is enormous. In this sense, the simple Voronoi decomposition doesn't meet one



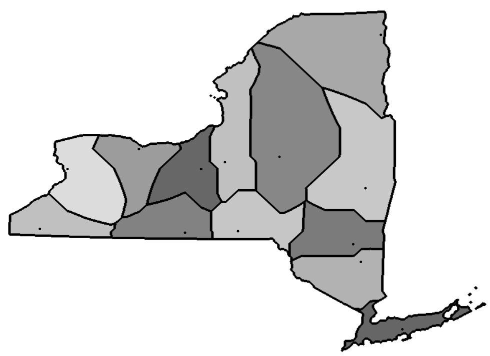


b. Regions created using the Manhattan metric after subdivisions are implemented. Subdivisions are created in New York City, Buffalo, Rochester, and Albany.

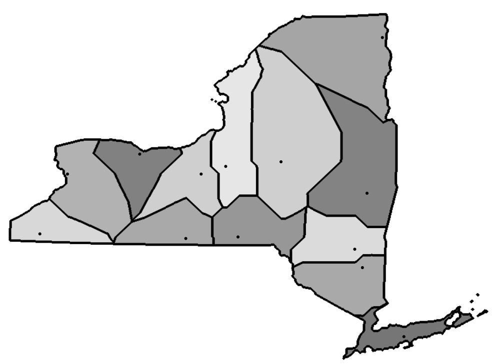
Figure 5. Implementation of Voronoi diagrams with the Manhattan metric. in three steps: assigning degeneracies to generator points, using the points to generate regions, and creating subregions generated by degenerate points. Only the last two steps are depicted. (Dots in each region represent generator point locations.)



a. Regions created using the Manhattan metric before subdivisions. Average Population $= (3.5 \pm 2.2)\%$.



b. Regions created using the Euclidean metric before subdivisions. Average Population $= (3.7 \pm 2.6)\%$.



c. Regions created using the Uniform metric before subdivisions. Average Population $= (3.7 \pm 2.6)\%$.

Figure 6. Voronoi diagrams generated with three distance metrics before subdivision of densely populated regions. (Dots in each region represent generator point locations.)

of the main goals. However, the Voronoi regions are so simple that we prefer to augment this method with population weights rather than abandon it entirely.

Figure 17 shows the result of this decomposition, along with exploded views of the two regions which were subdivided more than twice in the refinement stage of the diagram generation. The population in each region varies from 2.44% to 6.15%.

Precisely Defining Boundary Lines

It is not satisfactory to say that the regions created by our models should define the final boundary locations. At least, boundaries should be tweaked so that they don't accidentally divide houses into two districts. However, given the scale at which the Voronoi and Voronoiesque diagrams were drawn, it seems reasonable to assume that their boundaries could be modified to trace existing boundaries—like county lines, zipcodes, or city streets—without changing their general shape or average population appreciably. As an example, the average area of a zipcode in New York state is 10 sq. mi. and roughly 200 city blocks per square mile in Manhattan, while the minimum size of one of our Voronoi regions is 73 sq. mi. and the average size is 2,000 sq. mi. Therefore, it seems reasonable that we could approximate the boundaries of our Voronoi and/or Voronoiesque diagrams by pre-existing boundaries.

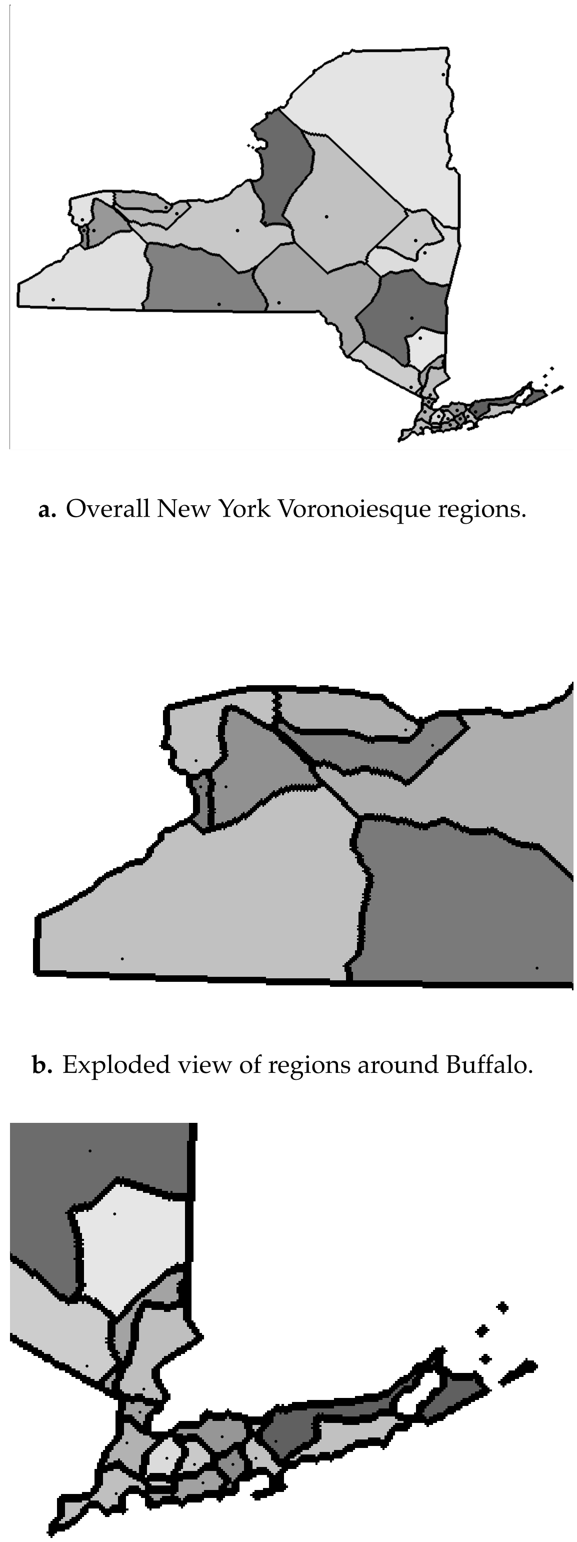
Analysis

New York State Results

In terms of simplicity of generated districts, our Voronoi-diagram method is a clear winner, particularly when applied with the Manhattan metric: The generated regions are contiguous and compact, while their boundaries—being unions of line segments—are about the simplest that could be expected. However, this method falls short in achieving equal population distribution among the regions.

When we modify the Voronoi diagram method to generate population-weighted Voronoiesque regions, we cut the population variance by a factor of four—from $\pm 2.8\%$ to $\pm 0.7\%$—while suffering a small loss in the simplicity of the resulting regions. In particular, regions in the Voronoiesque diagrams appear to be less compact, and their boundaries are more complicated, than their Voronoi diagram counterparts, though contiguity is still maintained.

Any implementation of a diagram generated from either method would have to make small localized modifications to ensure that the district boundaries make sense from a practical perspective. Though this would appear to open the door to politically-biased manipulation, the size of the necessary deviations (on the order of miles) is small enough when compared to the size of a Voronoi or Voronoiesque region (on the order of tens or hundreds of miles)



a. Overall New York Voronoiesque regions.

b. Exploded view of regions around Buffalo.

c. Exploded view of regions around Long Island.

Figure 7. Districts created by the Voronoiesque diagram for New York. Average population per region $= (3.34 \pm 0.74)\%$. (Dots in each region represent generator point locations.)

to make the net effect of these variations insignificant. Therefore, though we have provided only a first-order approximation to the congressional districts, we have left little room for gerrymandering.

General Results

Population Equality

The largest problem with this requirement occurs when we try to make regions too simple. Typically, our Voronoi method has the most room for error here. If a state has high population density peaks with a relatively uniform decrease in population density extending away from each peak, then the regions will differ quite a bit. This is because in this situation, ratios of populations are then roughly equal to the ratios of areas between regions. However, our final method focuses primarily on population, so equality is much easier to regulate.

Contiguity

Contiguity problems arise if the state itself has little compactness, like Florida, or has some sort of ocean sound, as Washington has. The first two methods focus on population density without acknowledging the boundaries of the state. So it's possible for a district to be divided by a geographic obstruction, such as a body of water or a mountain range. Our final method fixes this by growing in increments, which allows for regions to grow not over but around specified obstacles.

Compactness

The Voronoi diagram method creates convex regions. Though the Voronoi-esque method cannot guarantee convexity, the form of a region is similar in shape and size to the Voronoi region. A nice property of Voronoi regions is that we can make slight adjustments to the boundaries while still maintaining convexity (see below). This is good for taking population shifts across districts into account between redistrictings.

Improving the Method

Boundary Refinement

The Voronoi approach is good at generating polygonal districts but not as successful at maintaining population equality. One improvement is vertex repositioning. Adjacent districts generated by this method all share a vertex common to at least three boundaries. From this vertex extends a finite number of line segments that partially define the boundaries of these adjacent regions.

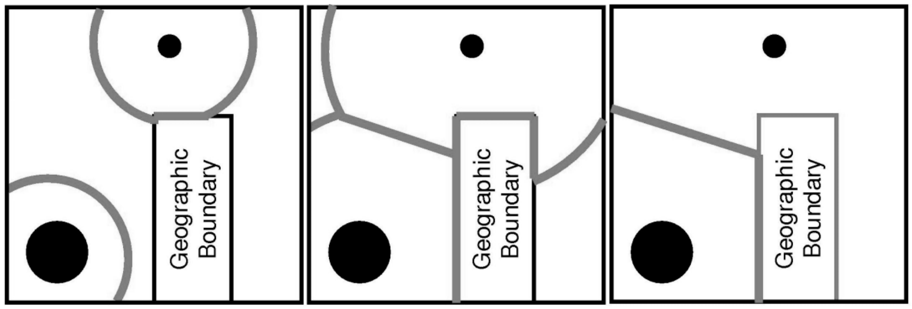


Figure 8. Illustration of Voronoiesque diagram generation that takes geographic obstacles into account.

Connecting the endpoints of these segments yields a polygon. Now we are free to move the common vertex anywhere in the interior of this polygon while still maintaining convexity; we can redraw boundaries between regions to equalize population size.

In the Voronoiesque method, too, there are ways to adjust population inequality: Looking at the region with the lowest population, systematically increase the area of the low-population regions while decreasing the area of the neighboring high-population regions.

Geographic Obstacles

Our methods don't consider geographic features such as rivers, mountains, canyons, etc. The Voronoiesque method, however, has the potential to implement these features. The same algorithm that detects intersections between Vornoiesque regions can detect a defined geographic boundary and stop growing in that direction. An illustration of this idea is shown in **Figure 18**.

Conclusion

Our model requires the use of only a state's population distribution but as an option can incorporate county, property, and geographic considerations.

Our Voronoiesque model satisfies our proposed goal. We supply a model for a redistricting committee to generate district boundaries that are simple, contiguous, and produce districts with equal populations. In particular, Voronoiesque diagrams redistrict New York very well. What is particularly attractive about our methods is that generating the districts is intuitive and accessible to the general public. The computer generation process takes less than 10 s.

References

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

Hamilton, Howard D. 1966. *Reapportioning Legislatures*. place???: Charles E. Merril Books, Inc.

Hess, S.W., J.B. Weaver, H.J. Siegfeldt, J.N. Whelan, and P.A. Zitlau. 1965. Nonpartisan political redistricting by computer. *Operations Research* 13 (6): 998–1006.

Irwin, Jim. 2006. `http://upload.wikimedia.org/wikipedia/en/e/e2/New_York_Population_Map.png` .

Mehrotra, Anuj, Ellis L. Johnson, and George L. Nemhauser. 1998. An optimization based heuristic for political districting. *Management Science* 44 (8): 1100–1114.

Taylor, Peter J. 1998. A new shape measure for evaluating electoral district patterns. *American Political Science Review* 67 (3): 947–950.

________. 2007. Census 2000 Summary File 1 [New York]. `http://www2.census.gov/census_2000/datasets/Summary_File_1/New_York/` .

U.S. Census Bureau. 2005. 2005 first edition tiger/line data, Feb. 2007.

Wikipedia. 2007. New York City. `http://en.wikipedia.org/wiki/New_York_City` .

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