AN ALGORITHM TO CONSTRUCT CONTINUOUS AREA CARTOGRAMS*

James A. Dougenik
Bedford, MA

Nicholas R. Chrisman
University of Wisconsin Madison

Duane R. Niemeyer
Reading, MA

Continuous area cartograms distort planimetric maps to produce a desired set of areas while preserving the topology of the original map. We present a computer algorithm which achieves the result iteratively with high accuracy. The approach uses a model of forces exerted from each polygon centroid, acting on coordinates in inverse proportion to distance. This algorithm can handle more realistic descriptions of polygon boundaries than previous algorithms and manual methods, thus enhancing visual recognition. **Key Words: cartograms, thematic cartography, computer cartography, transformations, distortion of maps.**

Cartograms are controversial in part because they are difficult to construct and the results seen to date are crude or imprecise or both. They also may communicate poorly to some audiences. Our computer algorithm attempts to redress the balance by providing a new approach to constructing precise cartograms.

Definition

A cartogram is a map purposely distorted so that its spatial properties represent quantities not directly associated with position on the globe. As thematic maps, cartograms emphasize the distribution of a variable by changing the area (or lengths) of objects on the map. There are two broad categories of cartograms, linear and area [for a more complete discussion see 2]. Linear cartograms express one-dimensional quantities by altering the distance component of maps while area cartograms use two-dimensional distortions to represent thematic information. Since the two forms have distinct methods of construction, we will concentrate on the area case exclusively. Within area cartograms the largest distinction concerns continuity; they can easily be produced by sacrificing continuity and surrounding all zones with varying amounts of blank space [i.e. 3]. Despite this alternative, the traditional form of a cartogram remains the continuous area technique discussed as long ago as 1934 by Raisz [4]. Considering the long-term interest in continuous area cartograms, we believe that an effective computer algorithm to construct them is desired. Our approach maintains continuity, and preserves many local features of cartographic lines that provide visual clues to the identity of the distorted objects.

Chronology of Cartogram Algorithms

The only previous publication presenting an algorithm for continuous area cartograms was produced by Tobler in 1973 [6]. He used a two step process to first fix the base map to a continuous surface representing the thematic variable, then to project the map on that surface onto a new plane introducing some distortion. The projection is specified by minimizing the Jacobian determinant of the surface as an approximation of the new areas, but the new areas relate to a cellular grid, not the original polygons. Through successive iterations involving a quadratic function of differences between desired and actual areas, the approximation is improved.

The quadratic method provides a new area for each cell, but it does not assure that the projection is a continuous function. Tobler describes the final convergence of the method as

slow because a topological test is needed so that cell corners do not cross cell boundaries. In the example (a U.S. state cartogram of 1970 population) in his paper, convergence is far from complete after 99 iterations. The areas achieved only show a correlation coefficient of .6 with the areas intended which implies an explanation of only 36 percent of the variance. This result shows distortion in the correct direction, but the result is a cartogram of a different variable from the one intended.

The insufficient accuracy of Tobler's results led Chrisman to outline a new approach using a rubber sheet distortion [1]. The cartogram process was applied directly to the topological structure of the map, not through an intermediary surface. Each polygon exerted a force on the adjacent boundary nodes, producing a vector result, when summed for all adjacent polygons, that displaced the node's position. The force was proposed to be proportional to the signed square root of the difference between current and desired area. The square root transformation converts a ratio of areas to a ratio of positions. The force was proposed to act from a polygon "center" on the nodes of that polygon's boundary. All the forces of polygons adjacent to a node are summed to displace the node to a new location. Like Tobler, Chrisman planned an iterative cycle with new areas and coordinates replacing the previous ones.

The Proposed Algorithm

Dougenik first attempted to implement Chrisman's ideas in 1981 and discovered difficulties which led to the algorithm presented here. Both previous approaches produced forces acting only on a topological neighborhood and with no concept of distance decay. Dougenik recognized the utility of force field concepts, particularly the distance decay function as applied to electrical, gravitational or even social phenomena. The numerator of the polygon force was defined as Chrisman did, but he introduced a denominator of distance (see formula 1). In close proximity to a centroid, the new distance weighted force is strong. Further away it diminishes, but the force is never ignored. The distance decay concept substitutes for the topological locality suggested by Chrisman. The sum of all forces (from all polygons) is exerted on each coordinate of the map causing it to be displaced. The resulting boundaries avoid topological damage because the distortion field is smooth and twice differentiable. The overall effect produces large distortions in shape, but the property of differentiability preserves conformality in each small area of the map.

$$F_{ij} = (p_j - q_j) \, p_j/d_{ij} \qquad (1)$$

where: F_{ij} = force exerted by polygon j on point i
          p_j = square root (actual area)/square root (π)
          q_j = square root (desired area)/square root (π)
          d_{ij} = distance from centroid of j to point i
          actual area is measured for a polygon and normalized by the sum of the actual areas.
          desired area is the thematic variable, also normalized by its sum.

In the implementation of formula 1, other problems surfaced. When a coordinate was very near a polygon center, the force exerted was tremendously large. An adjustment was employed, patterned on the interpolation search procedure in SYMAP where Shepard attaches a linear function with a continuous derivative to the tail of the distance decay function [5]. The adjustment shown in formula 2 only affects distances less than the "radius" of the polygon. (The term p defined above serves as a radius, considering the polygon to be as compact as possible.)

For d_{ij} greater than p_j use Formula (1)
For d_{ij} less than or equal to p_j:
$$F_{ij} = (p_j - q_j) \, (\, (4p_j - 3d_{ij})/p_j) \, (d_{ij}^2/p_j^2) \qquad (2)$$

The adjustment in formula (2) makes the combined functions continuous and differentiable at the point of crossover, and it also provides a zero value when the distance goes to zero. While formula (2) provides the strength of the displacement, the direction of the vector is determined by the line connecting the centroid to the point. A positive value of the force moves the point away from the centroid, and a negative force pulls towards it.

Unlike Tobler's method, this procedure makes no checks for topological boundary viola-

tions. The basic mathematics of this projection normally avoid deformation of the map to-pology. This property is only theoretical, because polygons are constructed by discrete co-ordinates not as continuous entities. Our algorithm incorporates a number of devices to avoid topological problems. One feature is a "force reduction factor," a number less than 1, used to reduce the impact of cartogram forces in the early iterations of the procedure. The force reduction factor is the reciprocal of one plus the mean of the size error. The size error is calculated by the ratio of area over desired area (if area is larger) or desired area over area in the other case.

In some cases such as panhandles it is possible to produce crossing borders. However, by inserting coordinates at regular intervals along each line, the program can ensure smooth distortion of long lines and will usually avoid crossings. A sure method to avoid overlap is to split polygons so that there are centroids for each convex section of the shape. These functions require a geographical analysis system (such as Harvard's ODYSSEY system) providing polygon overlay and disaggregation. A further alternative would be to recast the force field concept so that the centroids were replaced by a "charged plate."

As long as centroids are used there is potential difficulty from centroids outside the polygon. Rather than investing in complex geometric analysis, it is more reasonable to use interactive graphic correction of centroid location.

Improvements in the efficiency of the algorithm are possible. Currently the procedure (see Appendix) employs a brute force method—the forces of all polygons act upon every boundary coordinate. As long as the number of polygons is relatively small (say under 500), distortions can be computed for rather complex line work (thousands of points). Computation of force effects could be restricted in two ways. A search limitation could be implemented so that infinitesimal forces from far-away polygons are excluded. Alternatively, Chrisman's approach could be adopted in part by performing the full search only for nodes, then interpolating the forces along the rest of the line.

Example of Output

A comparison of a cartogram base map and the standard U.S. map showing the results of the 1960 presidential election is provided in Figures 1 and 2. Use of a cartogram in this situation is appropriate because geographic area is unrelated to winning elections. The cartogram base, by adjusting states' areas to represent electoral vote, creates a clearer impression of the close-ness of that election. From Figure 2 it might be possible to think that Nixon won.

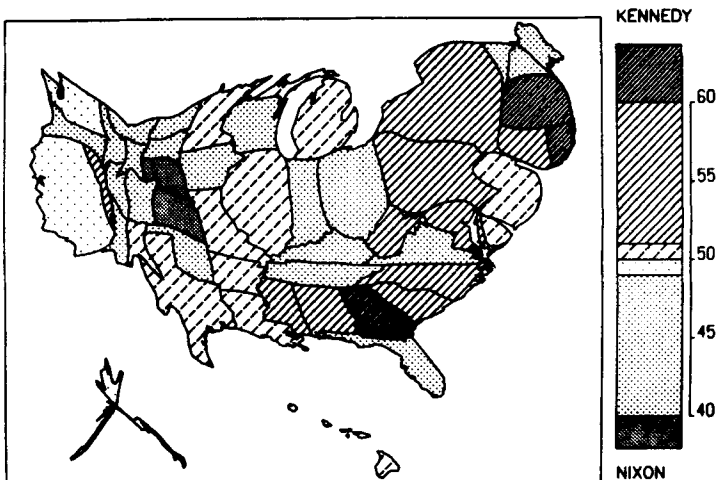Figure 3 shows the base map before distortion, highlighting some examples of percentage

Figure 1. Percentage of popular vote for Kennedy—1960 election (cartogram base).

% OF POPULAR VOTE FOR KENNEDY
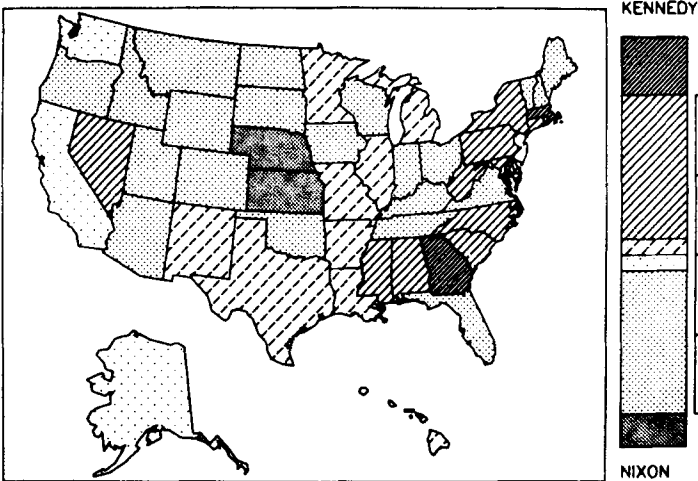
— 1960 PRESIDENTIAL ELECTION

Figure 2.   Percentage of popular vote for Kennedy—1960 election (equal area projection).

differences between desired "electoral area" and the area shown. Rhode Island is 18 times too small, Alaska starts out 13 times too large, while Florida is only 7 percent smaller than desired. Figures 4, 5, and 6 show the results of iterations leading to the final result. After eight iterations (in Figure 6) the mean percentage difference between actual and desired state areas is 1.7 percent. This cartogram is considered sufficiently accurate to use for Figure 1.
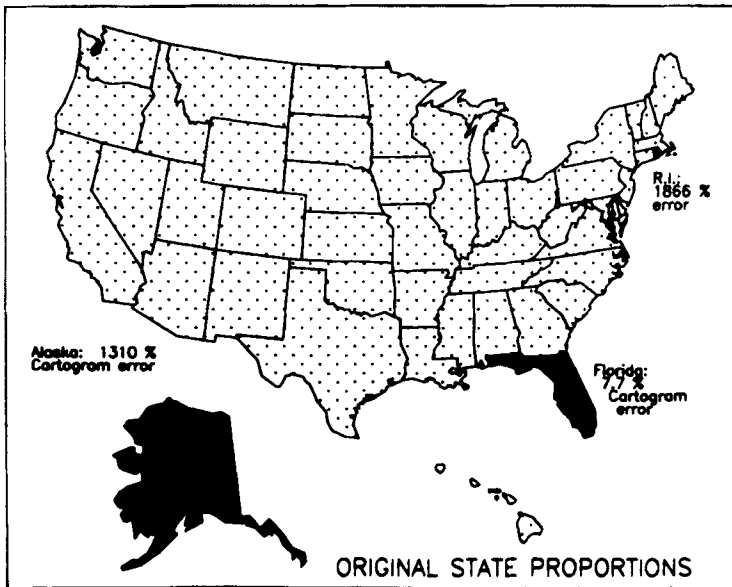

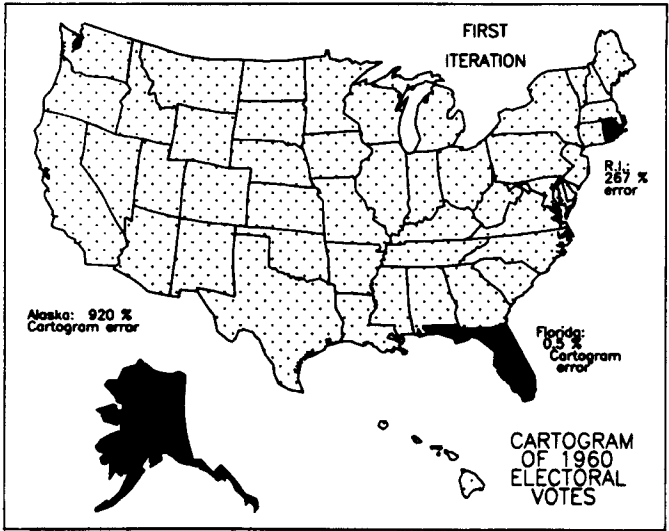Figure 3.   Selected proportionate error (equal area projection).

Figure 4.   Selected proportionate error (after first iteration).

Virtually all the error is located in Nevada and Alaska which are still too large. Alaska, being a disjoint polygon, exhibits a star like contraction due to the operation of the distance weighting. It would be a bit more pleasing to scale Alaska beforehand, but this result shows the actual effect on raw maps. A pre-scaled Alaska, combined with a split-up California, could produce a map of near perfect accuracy.
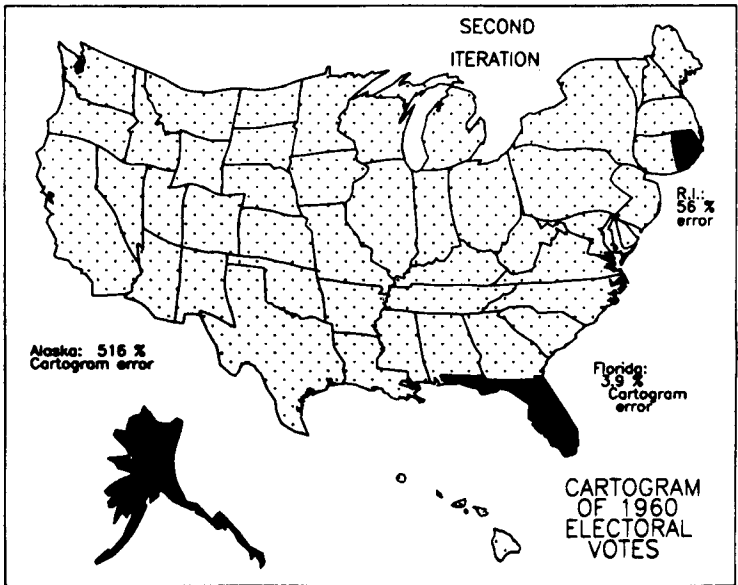

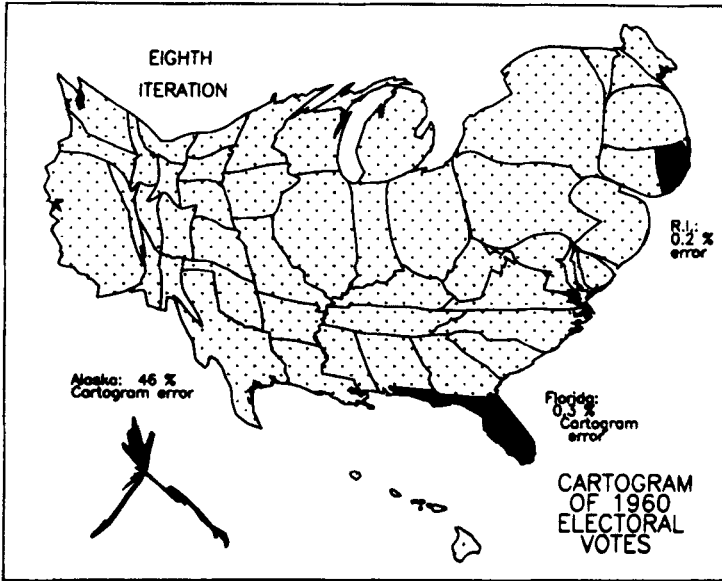Figure 5.   Selected proportionate error (after second iteration).

Figure 6.   Selected proportionate error (after eighth iteration).

Accuracy and Convergence

The electoral variable used for the cartogram examples here, being based on the population distribution over the states, is rather well behaved for this algorithm. In general, the results are more pleasing when the variable is spatially autocorrelated, less pleasing in cases like California/Nevada where the difference is sharp. In addition, the use of electoral votes transforms population figures (due to the two Senate seats) so that states do not fall too close to zero, relative to the maximum value (47). The algorithm will operate for less well-conditioned data, but the results may not be quite as pleasing.

Cartograms have been produced by Dougenik for the U.S. states using other, less autocorrelated variables, such as egg production. In this case, the low values are much closer to zero (relative to the maximum). For example, Nevada nearly vanishes, while Arkansas expands remarkably. The egg production achieved reasonable convergence, but it took more iterations than the electoral vote map.

As a further example, the algorithm has been applied to a population cartogram of Massachusetts by municipality. These 351 cities and towns range in population from Boston's hundreds of thousands to many Berkshire towns less than 100. The population surface at the more local level has some sharp drops. Although Boston is surrounded by rings of suburbs, some of the smaller cities particularly in the west are not. When applied to the Massachusetts case, the algorithm achieved 7 percent average deviation after twenty iterations. The smallest towns were flattened beyond recognition, but the overall shape was remarkably clear. Most of the error came from small towns which were still too large. The proper solution would be to aggregate the small towns before applying the procedure so that the spatial unit would be large enough to be visible.

These few applications do not constitute a formal proof of convergence or accuracy for all applications. But they point to a few practical rules that will lead to more useful results. First, the perception of shape will be best when the variable is spatially autocorrelated. Second, zones with complex shapes should be cut into separate, more nearly convex portions for computation, then reaggregated for display. Third, standard rules of mapping must be extended to incorporate cartogram problems. Traditionally scale translates into line weights,

distance tolerances and minimum mapping units. In the case of a cartogram, scale puts limits on the thematic variable. Small zones should be aggregated into nearby zones, preferably sharing other characteristics.

Conclusion

The algorithm outlined here and listed in the Appendix provides an effective means to construct continuous area cartograms. It operates substantially more quickly than Tobler's previous algorithm (8 versus 99 iterations) and it comes much closer to providing the desired areas. Availability of this algorithm should make it possible to generate continuous area cartograms without the problems of overgeneralization.

Appendix: A Procedure for Producing Cartograms

For each polygon
  Read and store PolygonValue (negative value illegal)
  Sum PolygonValue into TotalValue
For each iteration (user controls when done)
  For each polygon
    Calculate area and centroid (using current boundaries)
  Sum areas into TotalArea
  For each polygon
    Desired   = (TotalArea * (PolygonValue/TotalValue))
    Radius   = SquareRoot (Area/π)
    Mass   = SquareRoot (Desired/π) - SquareRoot (Area/π)
    SizeError   = Max(Area, Desired) / Min(Area, Desired)
  ForceReductionFactor = 1 / (1 + Mean (SizeError))
  For each boundary line; Read coordinate chain
    For each coordinate pair
      For each polygon centroid
        Find angle, Distance from centroid to coordinate
        If (Distance > Radius of polygon)
          Fij = Mas * (Radius/Distance)
        Else
          Fij = Mass * (Distance $\wedge$2 / Radius $\wedge$2) *
                      (4 − 3 * (Distance / Radius))
      Using Fij and angles, calculate vector sum
      Multiply by ForceReductionFactor
      Move coordinate accordingly
    Write distorted line to output and plot result

Literature Cited

1. Chrisman, Nicholas R. "Cartogram Projections of Planar Polygon Networks." Internal report, Laboratory for Computer Graphics and Spatial Analysis, Harvard University, 1977.
2. Monmonier, Mark S. "Maps, Distortion and Meaning." *AAG Resource Paper 75–4*, Washington: 1977.
3. Olson, Judy M. "Noncontinuous Area Cartograms." *Professional Geographer*, 28 (1976), 371–380.
4. Raisz, Erwin. "The Rectangular Statistical Cartogram." *Geographical Review*, 24 (1934), 292–296.
5. Shepard, Donald S. "SYMAP Interpolation Characteristics." Vol. 2, Report L in *Computer Mapping as an Aid in Air Pollution Studies*. Edited by John Goodrich. Cambridge, MA: Laboratory for Computer Graphics and Spatial Analysis, Harvard University, 1970.
6. Tobler, Waldo R. "A Continuous Transformation Useful for Districting." *Annals*, New York Academy of Sciences, 219 (1973), 215–220.

JAMES A. DOUGENIK is a software engineer for Verbex, Bedford, MA. His research interests include speech recognition and algorithms for automated cartography. NICHOLAS R. CHRISMAN (Ph.D., Bristol University) is an Assistant Professor, Department of Landscape Architecture, University of Wisconsin-Madison, 53706. His current research concerns measures of map error and applications of information systems to the operations of local governments, along with the development of data structures and algorithms. DUANE R. NIEMEYER (M.A., Akron) is a member of the technical staff for The Analytic Sciences Corporation, Reading, MA, working in the field of automated cartography.