

NEW SOLUTION FOR CONSTRUCTION OF RECTILINEAR AREA CARTOGRAM

Ryo Inoue

Department of Civil Engineering, University of Tokyo
7-3-1, Hongo, Bunkyo, Tokyo 113-8656, Japan
E-mail: rinoue@civil.t.u-tokyo.ac.jp

Kazuki Kitaura

Department of Civil Engineering, University of Tokyo
7-3-1, Hongo, Bunkyo, Tokyo 113-8656, Japan
E-mail: kitaura@trip.t.u-tokyo.ac.jp

Eihan Shimizu

Department of Civil Engineering, University of Tokyo
7-3-1, Hongo, Bunkyo, Tokyo 113-8656, Japan
E-mail: shimizu@civil.t.u-tokyo.ac.jp

Abstract

An area cartogram (hereafter referred to as a cartogram) is transformed a map on which areas of regions are proportional to statistical data values, and is considered as one of powerful tools for the visual representation of statistical data. A rectilinear cartogram, on which regions are represented in the rectilinear format, is one of the most familiar cartograms, however, its computational solutions have not yet been proposed.

The approach for the construction of rectilinear cartograms is similar to that for the construction of continuous cartograms, because both cartograms are contiguous and complex in terms of region shapes. Then, in this study, we elucidate problems faced in the construction of rectilinear cartograms, and provide a computational solution by adopting an approach similar to that adopted by Inoue & Shimizu (2006), which had proposed a solution for continuous cartogram construction.

Based on this approach, we express the construction of a rectilinear cartogram as nonlinear least squares in which variables are coordinates of edges on rectilinear polygons. We then linearize the solution to construct cartograms through the iteration of linear least squares. Further, we apply the solution to datasets of the US population and find that our proposed solution can be successfully used to construct a rectilinear cartogram on which data is accurately represented and deformation of region shapes is rather small. However, on applying the solution to the construction of a rectangular cartogram, which is a similar type of cartogram, we do not obtain an accurate result.

We conclude that the solution is effective for constructing rectilinear cartograms, and it would contribute to the enhancement of the visualization feature of a GIS.

1. Introduction

Nowadays, various statistical datasets compiled by national and local governments are available in the form of digital data at little or no cost. As a result, both analysts and ordinary citizens have ready access to extensive databases. Many statistical datasets can be sourced within geographic frameworks provided by a geographic information system (GIS). One of the primary functions of a GIS is to facilitate the visual presentation of statistical analyses of data. Typical visualization tools that are a part of GIS software packages include methods such as choropleth mapping and dot mapping.

An alternative visualization tool is an area cartogram, which has been discussed in quantitative geography (Monmonier 1977; Dorling 1996; Tobler 2004). Area cartograms are sometimes referred to as “cartograms” or “value-by-area maps”; then, we describe area cartograms hereafter referred to as cartograms in this paper. Cartograms are transformed maps on which areas of regions are proportional to statistical data values. Deformation of the shape of regions and displacement of regions assist map-readers to intuitively recognize the distribution of statistical data represented on cartograms. Therefore, a cartogram is a powerful tool for the visual representation of statistical data.

Cartograms are classified into four types (described in subsequent paragraphs) according to two characteristics—shapes and contiguities of regions on cartograms. First, we focus on the expression of statistical data using different shapes of regions. Some cartograms express statistical data using complex shapes, while others express statistical data using simple shapes such as circles and rectangles. Since it is easy to compare cartograms having complex shapes with geographical maps, it is easy to comprehend the characteristics of spatial distribution of statistical data presented in such cartograms. However, cartograms having complex region shapes are difficult to compare the sizes of regions; in this sense, it is better to use simple shapes to express statistical data. However, considering that the correspondence between regions and shapes on cartograms using simple shapes is unclear, the use of complex and simple shapes is both advantageous and disadvantageous for different reasons.

Cartograms in which statistical data is expressed using complex shapes are classified in two types according to the contiguities of regions. The first type is a cartogram in which contiguities of regions are omitted and is termed a noncontiguous cartogram (Olson 1976). Its construction is quite simple—it involves just the allocation of resized regions. The second type is a cartogram in which contiguities of regions are retained and is termed a contiguous (or continuous) cartogram. It is constructed by deforming regions on a geographical map so as to express the sizes of their statistical data by means of their area on a cartogram. Thus far, many computational algorithms have been proposed for the construction of this type of cartogram (e.g., Tobler 1963; Gastner & Newmann 2004; Inoue & Shimizu 2006).

Similarly, cartograms in which statistical data is expressed using simple shapes are also

classified into two types— noncontiguous and contiguous cartograms. Examples of noncontiguous cartograms are circle area cartograms (Dorling 1996) and rectangular cartograms (Upton 1991), and heuristic algorithms for the construction of these cartograms have been proposed. One type of contiguous cartograms is a rectangular cartogram, proposed by Rasiz (1934). In rectangular cartograms, regions are represented by rectangles and rectangles representing adjacent regions are contiguous. Since it is easy to perceive areas of regions in a rectangular cartogram, their use is effective for the visualization of statistical data. However, due to difficulties in the construction of rectangular cartograms, not many mathematical solutions have been proposed (van Kreveld & Speckmann 2004, 2007; Speckmann *et al.* 2006; Heilmann *et al.* 2004).

Another type of cartogram termed rectilinear cartogram expresses statistical data using rectilinear polygons. This cartogram is similar to the rectangular cartogram as well as the continuous cartogram. Rectilinear cartogram has advantages of both these cartograms: (1) it is easy to perceive areas of regions, because of the simple shapes of regions and (2) it is easy to comprehend the spatial distribution of data, because of the ease of comparison to a geographical map. Because of high data comprehensibility, rectilinear cartograms are frequently used to represent data in newspaper articles, particularly in articles on election results (e.g., Nagourney *et al.* 2008). However, thus far, no computational solution has been proposed for their construction.

The construction of rectilinear cartograms is similar to that of continuous cartograms, in the sense that both these cartograms are contiguous and complex in terms of region shapes. Then, we consider that an approach similar to that for the construction of a continuous cartogram would be effective for constructing a rectilinear cartogram. Inoue & Shimizu (2006) have proposed a solution for the construction of a continuous cartogram, which involves triangulation of regions and imposition of a restriction on changes in bearings (hereafter referred to as “bearing changes”) of edges. This solution is user friendly, since it does not require the setting of more than one parameter, whose mathematical meaning is clear; further, this solution involves only a short calculation for cartogram construction.

In this study, we obtain a solution for the construction of rectilinear cartograms by adopting an approach similar to that adopted by Inoue & Shimizu (2006) and verify the applicability of the solution to the construction of rectilinear cartograms. Moreover, since rectangular cartograms are only a specific form of rectilinear cartograms, a solution for rectilinear cartograms might be acceptable even for rectangular cartograms. Then, we also verify the applicability of the proposed solution for their construction.

2. Previous Studies on Construction of Rectangular and Rectilinear Cartograms

In this section, we review previously proposed solutions for the construction of rectangular and rectilinear cartograms.

Rasiz (1934) first proposed a rectangular cartogram, depicting information such as population of the United States. These handwritten cartograms, however, do not necessarily represent the correct contiguity of the states. In fact, it is impossible to represent contiguities on rectangular cartograms correctly. Rectangles have four edges; thus, if a region shares borders with less than four other regions, it is adjacent to other regions that are not geographically adjacent. This characteristic makes the allocation of rectangles complex and the construction of a rectangular cartogram difficult. As pointed out by Biedl & Genc (2005), its construction is almost NP-hard. Consequently, no computational solutions for rectangular cartograms had been proposed until recently.

Van Kreveld & Speckmann (2004, 2007) first proposed and implemented a computational solution for the construction of rectangular cartograms. The solution consists of a method for creating rectangular layouts that are similar to geographical contiguities, out of many available rectangular layouts, and a heuristic method for moving edges of rectangles in order to fit the area of rectangles to assigned data. Then, Speckmann *et al.* (2006) improved the latter method of the abovementioned solution as linear programs on coordinates of vertexes of rectangles; this in turn was achieved by imposing error of area constraints, aspect ratio constraints, planarity preserving constraints, and adjacency preserving constraints. This improved solution yields cartograms that represent data more accurately than the previous solution. Further, Speckmann *et al.* (2006) also proposed a solution that is applicable to the construction of cartograms that include L-shaped regions, and found that the use of L-shaped regions caused a decrease in errors in area representation. However, this solution cannot be used for constructing rectilinear cartograms. Heilmann *et al.* (2004) proposed another solution in which constraints were imposed on empty space and on the area, shape, topology, and relative positions of rectangles. This solution consists of a heuristic method and can be used to construct both continuous and noncontiguous rectangular cartograms by changing the combination of the constraints. It is also applicable to the construction of cartograms that include L-shaped, C-shaped, and O-shaped regions; however, it is not applicable to the construction of rectilinear cartograms. De Berg *et al.* (2006) also proposed the solution that constructs rectangular cartograms in which rectilinear regions are formed only when the contiguities of input rectangles are not preserved, however it is also not applicable to the construction of rectilinear cartograms.

Rectilinear cartograms have been constructed manually; Hunter & Young (1968) reported that it required 16 man-hours to construct the rectilinear population cartogram of the 62 counties in England and Wales in 1961 using approximately 9,000 wooden blocks. Although they are frequently used (e.g., Nagourney *et al.* 2008), their computational solutions have not proposed yet.

In summary, although several methods have thus far been developed for the construction of rectangular and rectilinear cartograms, they are generally not sufficiently accurate for the construction of rectilinear cartograms.

3. Proposal of New Solution for Construction of Rectilinear Cartogram

The construction of rectilinear cartograms involves the transformation of rectilinear regions such that the size of regions is proportional to their statistical data. There are two basic problems in this construction:

1. It is impossible to derive the shapes of rectilinear polygons only from information on their area. That is, the construction of rectilinear cartograms is an ill-posed problem and has no unique answer; some regularization is required to solve this problem.
2. While map-readers read rectilinear cartograms, they note the differences in the sizes and shapes of regions through a comparison with those on geographical maps. Then, if the map distortion is very large, it is difficult for them to interpret data. Therefore, to preserve visual elegance, it is necessary to retain the shapes and alignments of regions in rectilinear cartograms.

As described earlier, we resolve these problems by adopting an approach similar to that adopted by Inoue & Shimizu (2006). Here, we briefly describe this approach. It yields a solution for solving the ill-posed problem and for retaining region shapes. This solution involves the triangulation of regions and imposition of a restriction on the bearing changes of triangle edges; it also yields a unique cartogram shape whose deformation is quite small. The triangulation of regions and the imposition of a restriction on the bearing changes of triangle edges address both the ill-posed nature of cartogram construction and the requirement for small distortions in region shape to yield visually elegant cartograms. Another advantage of triangulation is that it enables the expression of construction in the form of a simple equation.

The abovementioned approach is also applicable to the construction of rectilinear cartograms. The basic concepts of this approach are shown in Figure 1. We consider constructing a rectilinear cartogram of polygon OABCDE to represent the given data D_{OABCDE} . Since a cartogram should be fixed in a coordinate system of cartogram, vertex O is set at the origin; then, unknown variables are the coordinates of the five vertices. Adjacent vertices must share x- or y-coordinates on rectilinear cartograms; i.e., $x_A = x_O = 0$, $y_E = y_O = 0$, $y_A = y_B = y_{AB}$, $x_B = x_C = x_{BC}$, $y_C = y_D = y_{CD}$, and $x_D = x_E = y_{DE}$. Finally, the number of unknown variables is four—x-coordinates of edges BC and DE and y-coordinates of edges AB and CD. On the other hand, only one observation equation exists in the absence of triangulation; i.e., A_{OABCDE} should be equal to D_{OABCDE} . Therefore, it is obvious that the number of unknown variables exceeds the number of observations; then, the problem does not have a unique answer.

Here, we carry out triangulation. The polygon OABCDE is divided by the Delaunay triangulation procedure, and the data D_{OABCDE} is distributed according to the area of predeformed triangles. The number of observation equations increases from one to four, which is equal to the number of unknown variables. In fact, the number of triangles is always greater than or equal to the number of unknown coordinates of edges; in such a case, the problem has a unique answer. However, this does not hold true for the case

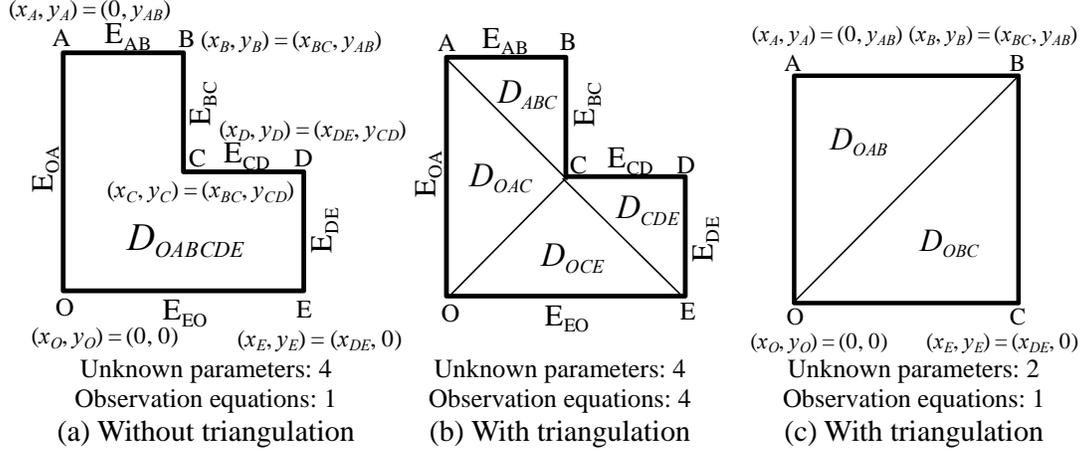


Figure 1 Number of unknown parameters and observation equations without restriction on bearing changes of triangle edges

shown in Figure 1 (c). The number of unknown variables is two—the x-coordinate of edge BC and the y-coordinate of edge AB. The polygon OABC is triangulated, and data is assigned to each triangle. However, since D_{OAB} always equals to D_{OBC} , triangulation and the assignment of data to triangles do not increase the number of observation equations; therefore, the cartogram does not have a unique shape.

Then, we impose a restriction on the bearing changes of the triangle edges. Since this restriction increases the number of observations, it is possible to obtain rectilinear cartograms. Moreover, the restriction constrains the excess don rectilinear cartograms; then, it becomes possible to construct visually elegant cartograms.

Next, we show the formulation of the construction problem of rectilinear cartograms by the triangulation of regions and imposition of a restriction on bearing changes. The objective function and regularization term can be described as

$$\min \left[\sum_{t_{ijk} \in T} \left(1 - \frac{A_{ijk}}{D_{ijk}} \right)^2 + \mu \sum_{e_{mn} \in E} \left(\theta_{mn}^C - \theta_{mn}^G \right)^2 \right] \quad s.t. \textcircled{a} \begin{cases} x_p = x_q = x_{pq} \quad \forall e_{pq} \in E_{Px} \\ y_r = y_s = x_{rs} \quad \forall e_{rs} \in E_{Py} \end{cases} \quad (1)$$

where t_{ijk} is a triangle consisting of vertices i, j , and k ; T , the set of triangles included in the cartogram; A_{ijk} , the size of the triangle t_{ijk} ; D_{ijk} , the data allocated to the triangle t_{ijk} ; e_{mn} , the edge between the vertices m and n ; E , the set of all edges included in T ; θ_{mn}^C , the bearing of e_{mn} on the geographical map; θ_{mn}^G , the bearing of e_{mn} on the cartogram; μ , the weight on the regularization term; further, x_i and y_i are the x- and y-coordinates of the vertex i on the cartogram, and E_{Px} and E_{Py} are the sets of vertical and horizontal edges that compose the rectangular polygons.

Next, we rewrite equation (1) using the vertex coordinates (x_i, y_i) . First, the size of the triangle t_{ijk} on the cartogram A_{ijk} may be expressed as

$$A_{ijk} = \left| (x_j - x_i)(y_k - y_i) - (x_k - x_i)(y_j - y_i) \right| / 2 \quad (2)$$

To remove the absolute sign from the objective function, we define D_{ijk-w_sign} as

$$D_{ijk-w_sign} = \begin{cases} D_{ijk} & \text{if } (x_j - x_i)(y_k - y_i) - (x_k - x_i)(y_j - y_i) \geq 0 \\ -D_{ijk} & \text{if } (x_j - x_i)(y_k - y_i) - (x_k - x_i)(y_j - y_i) < 0 \end{cases} \quad (3)$$

Then, the construction of a rectilinear cartogram can be expressed as

$$\min \left[\sum_{t_{ijk} \in T} \left\{ 1 - \frac{(x_j - x_i)(y_k - y_i) - (x_k - x_i)(y_j - y_i)}{2D_{ijk-w_sign}} \right\}^2 + \mu \sum_{e_{mn} \in E} \left\{ \frac{(x_n - x_m)(y_n^G - y_m^G) - (y_n - y_m)(x_n^G - x_m^G)}{d_{mn} d_{mn}^G} \right\}^2 \right] \quad (4)$$

$$s.t. \begin{cases} x_p = x_q = x_{pq} \quad \forall e_{pq} \in E_{Px} \\ y_r = y_s = y_{rs} \quad \forall e_{rs} \in E_{Py} \end{cases}$$

where (x_i^G, y_i^G) are the coordinates on the geographical map, d_{mn} is the distance between the vertices m and n on the cartogram, and d_{mn}^G is the distance on the geographical map. We can construct rectilinear cartograms by solving the nonlinear least squares of equation (4). While solving this equation, it is essential to control the size of weight μ for the regularization terms, in order to prevent the overturning of any triangles; the overturning of triangles destroys the topology of rectilinear polygons. The method described in Inoue & Shimizu (2006) is able to control the value of μ automatically; then, the users are not bothered by any parameter settings.

Let V_{Px} and V_{Py} denote the set of vertices on the edges in E_{Px} and E_{Py} . For convenience, we linearize equation (4) around the approximate values of vertices coordinates (x'_i, y'_i) :

$$\min f(x_{i \in V_{Px}}, y_{j \in V_{Py}}) = \min \left[\sum_{t_{ijk} \in T} \left\{ 1 + \frac{2A'_{ijk} + y'_{jk}x_i - y'_{ik}x_j + y'_{ij}x_k - x'_{jk}y_i + x'_{ik}y_j - x'_{ij}y_k}{2D_{ijk-sign}} \right\}^2 + \mu \sum_{e_{mn} \in E} \left\{ \frac{(x_n - x_m)y'_{mn} - x'_{mn}(y_n - y_m)}{d_{mn}^{\prime 2}} \right\}^2 \right] \quad (5)$$

$$s.t. \begin{cases} x_p = x_q = x_{pq} \quad \forall e_{pq} \in E_{Px} \\ y_r = y_s = y_{rs} \quad \forall e_{rs} \in E_{Py} \end{cases}$$

$$\text{where } x'_{ij} = x'_j - x'_i, y'_{ij} = y'_j - y'_i, A'_{ijk} = (x'_{ij}y'_{ik} - x'_{ik}y'_{ij})/2, d_{mn}^{\prime 2} = x_{mn}^{\prime 2} + y_{mn}^{\prime 2}$$

Now it is possible to construct rectilinear cartograms by solving the linear least squares of equation (5) iteratively. Then, we can reduce the amount of calculation by solving the following equations, instead of solving equation (5):

$$\frac{\partial f(x_{i \in V_{Px}}, y_{j \in V_{Py}})}{\partial x_{i \in V_{Px}}} = 0 \quad \forall i \in V_{Px}, \quad \frac{\partial f(x_{i \in V_{Px}}, y_{j \in V_{Py}})}{\partial y_{j \in V_{Py}}} = 0 \quad \forall j \in V_{Py}. \quad (6)$$

Equation (6) is the system of simultaneous linear equations. The number of equations is the number of rectangular polygons' edges minus two, since two edges should be fixed to the cartogram's coordinate system.

Generally, the amount of calculation to solve a system of simultaneous linear equations is proportional to the cubic of the number of equations; then in this case, the calculation time is proportional to the cubic of the number of rectilinear edges. However, the system of equation (6) is composed of a sparse matrix; it is possible to reduce the amount of calculation using numerical computation libraries, such as “pardiso” function in Intel Math Kernel Library.

4. Applications of Proposed Solution

We evaluated the applicability of the proposed solution by using it to construct a rectilinear cartogram of data of the US population in 2005, obtained from the US Census Bureau, United States Department of Commerce. First, we constructed a rectilinear cartogram by using simplified rectilinear regions (Figure 2(a)). We placed two dummy

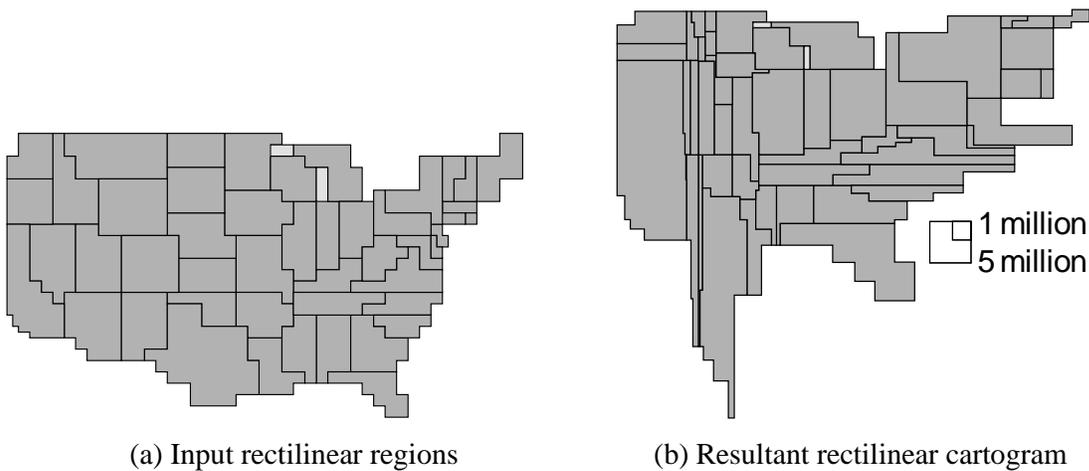


Figure 3 Construction of rectilinear cartogram of US population in 2005

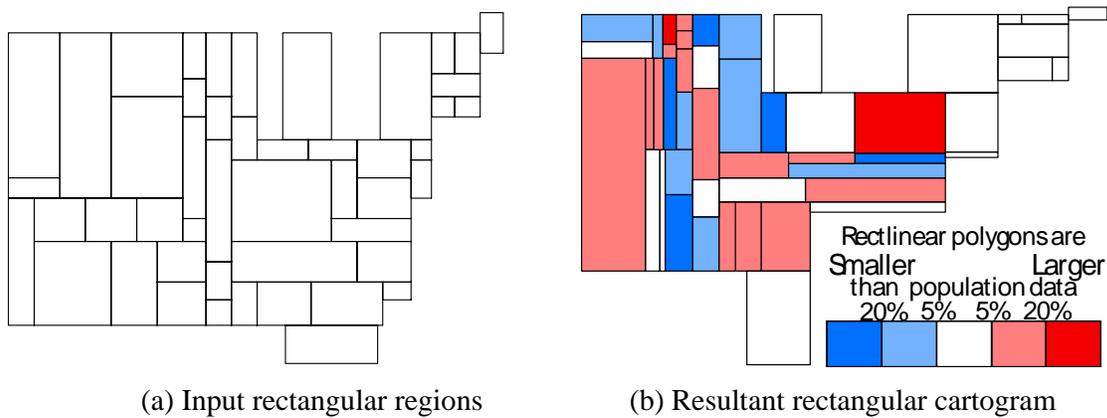


Figure 3 Construction of rectangular cartogram of US population in 2005

polygons at Lake Superior and Lake Michigan in order to prevent the intersection of the regions across lakes. The rectilinear polygons consisted of 210 vertices, 84 vertical edges, and 76 horizontal edges for 48 continental states.

Figure 2(b) shows the resultant cartogram. The shape of rectilinear cartogram looks similar to that of input regions' shape; you may be able to distinguish which rectilinear polygons represent which states' population data. The calculation was completed in less than two seconds on a computer with 1.2 GHz Pentium processor. The root mean square errors (RMSEs) between the input population data and the areas of rectilinear polygons on the resultant cartogram were 167, which were 0.003% of the average state population. This implies that the areas of rectilinear polygons perfectly fit the population data.

Next, we applied the proposed solution to the construction of a rectangular cartogram. Figure 3(a) shows the input shape of regions; it consists of 109 vertices, 30 vertical edges, and 31 horizontal edges. Figure 3(b) shows the resultant cartogram. Because of the small degree of freedom, the solution cannot be successfully used to construct a rectangular cartogram that represents data accurately. The brighter colors in Figure 3(b) indicate data represented with low accuracy, i.e., RMSEs of around 2 million, which is almost 40% of the average state population. Since the solution retains the input contiguities of regions, it follows that it is not possible to construct rectangular cartograms when the deformation of region shapes is large.

5. Conclusions and Future Work

In this study, we have proposed a new solution for the construction of rectilinear cartograms. This solution involves the division of rectilinear-shaped regions into triangles using Delaunay triangulation and the transformation of these triangles to fit them to data. In order to construct rectilinear cartograms that are readily interpretable, we imposed a regularization condition that constrained changes in the bearings of triangle edges. We then linearized the solution to construct rectilinear cartograms through the iteration of linear least squares.

The proposed solution can be used to construct rectilinear cartograms that represent data accurately, although it cannot be used to construct accurate rectangular cartograms. The solution is user friendly since it rapidly constructs rectilinear cartograms and does not require any parameter settings. Using this solution, rectilinear cartograms can be constructed using only input rectilinear shapes and a data set for visualization.

In the future, we intend to develop software for designing rectilinear shapes automatically and an algorithm for the construction of rectangular cartograms.

References

- de Berg, M., Mumford, E. & Speckmann, B. 2006. Optimal BSPs and rectilinear cartograms. *Proceedings of the 14th Annual ACM International Symposium on Advances in Geographic Information Systems*, 19–26.
- Biedl, T. & Genc, B. 2005. Complexity of octagonal and rectangular cartograms. *Proceedings of the 17th Canadian Conference on Computational Geometry*, 117–120.
- Dorling, D. 1996. *Area cartograms: Their use and creation*. Concepts and Techniques in Modern Geography (CATMOG), 59.
- Gastner, M.T. & Newman, M.E.J. 2004. Diffusion-based method for producing density-equalizing maps. *Proceedings of the National Academy of Sciences of the United States of America*, 101(20): 7499–7504.
- Heilmann, R., Keim, D.A., Panse, C. & Sips, M. 2004. RecMap: Rectangular map approximations. *Proceedings of the IEEE Symposium on Information Visualization*, 33–40.
- Hunter, J.M. & Young, J.C. 1968. A technique for the construction of quantitative cartograms by physical accretion models. *The Professional Geographer*, 20(6): 402–407.
- Inoue, R. & Shimizu, E. 2006. A new algorithm for continuous area cartogram construction with triangulation of regions and restriction on bearing changes of edges. *Cartography and Geographic Information Science*, 33(2): 115–125.
- van Kreveld, M. & Speckmann, B. 2004. On rectangular cartograms. *Lecture Notes in Computer Science*, 3221: 724–735.
- van Kreveld, M. & Speckmann, B. 2007. On rectangular cartograms. *Computational Geometry: Theory and Applications*, 37(3): 175–187.
- Monmonier, M.S. 1977. *Maps, Distortion, and Meaning*. Washington: Association of American Geographers.
- Nagourney, A., Zeleny, J. & Carter, S. 2008. The electoral map: key states, Election Guide 2008, *The New York Times*, November 4, viewed January 5, 2009, <<http://elections.nytimes.com/2008/president/whos-ahead/key-states/map.html>>.
- Olson, J.M. 1976. Noncontiguous area cartograms. *The Professional Geographer*, 28(4): 371–380.
- Rasiz, E. 1934. The rectangular statistical cartogram. *The Geographical Review*, 24: 292–296.
- Speckmann, B., van Kreveld, M. & Florisson, S. 2006. A linear programming approach to rectangular cartograms. In Riedl, A., Kainz, W. & Elmes, G. (eds.) *Progress in Spatial Data Handling*, Berlin: Springer. 529–546.
- Tobler, W.R. 1963. Geographical area and map projections. *The Geographical Review*, 53(1): 59–78.
- Tobler, W.R. 2004. Thirty five years of computer cartograms. *Annals of the Association of American Geographers*, 94(4): 58–73.
- Upton, G.J.G. 1991. Rectangular cartograms, spatial autocorrelation, and interpolation. *The Journal of Regional Science Association International*, 70(3): 287–302.