Article

Optimization-Based Construction of Quadrilateral Table Cartograms

Ryo Inoue * and Mao Li

Department of Human-Social Information Sciences, Graduate School of Information Sciences,
Tohoku University, Miyagi 850-8579, Japan
* Correspondence: rinoue@tohoku.ac.jp;

Abstract: A quadrilateral table cartogram is a rectangle-shaped figure that visualizes table-form data; quadrilateral cells in a table cartogram are transformed to express the magnitude of positive weights by their areas, while maintaining the adjacency of cells in the original table. However, the previous construction method is difficult to implement because it consists of multiple operations that do not have a unique solution and require complex settings to obtain the desired outputs. In this article, we propose a new construction for quadrilateral table cartograms by recasting the construction as an optimization problem. The proposed method is formulated as a simple minimization problem to achieve mathematical clarity. It can generate quadrilateral table cartograms with smaller deformation of rows and columns, thereby aiding readers to recognize the correspondence between table cartograms and original tables. In addition, we also propose a means of sorting rows and/or columns prior to the construction of table cartograms to reduce excess shape deformation. Applications of the proposed method confirm its capability to output table cartograms that clearly visualize the characteristics of datasets.

Keywords: quadrilateral table cartogram; contiguous area cartogram; visualization; table-form data; optimization

1. Introduction

A contiguous (or continuous) area cartogram is a visualization method for geospatial data that records the attributes of positive values of regions, such as population and Gross Domestic Product (GDP). It is a deformed map in which the shape of each region is transformed to fit its area on the cartogram to its size of positive attribute value, while maintaining the adjacency of regions (see [1,2]). The deformed shapes of the regions on cartograms give readers an intuitive impression on the characteristics of datasets.

The special feature of visualization by contiguous area cartograms is the data representation through the deformation of figures. This feature is not only effective for geospatial datasets but also for table-form datasets with positive values. The visualization of a table-form dataset by a contiguous area cartogram-like representation is called a table cartogram. The most common examples of table cartograms are the visualization of characteristics of elements on the periodic chemical table, the table cartograms that exhibit several features of elements created [3–5], and a table cartogram published by the European Chemical Society published in 2019, International Year of the Periodic Table of Chemical Elements, to exhibit the quantity of chemical elements on Earth [6].

Table cartograms are categorized into two types. One type is a 'freely-deformed' table cartogram whose outline border and edges of cells are deformed into curves [3,4,6]. A method of construction of cartograms of this type is the diffusion-based method [7]—one of several contiguous area cartogram construction methods—which imitates the diffusion process of fluids in physics. It is mathematically clear and easy to operate; thus, it has attracted attention in constructing contiguous

area cartograms. Winter [4], for instance, utilized the diffusion-based method. However, when the magnification ratios of adjacent cells—i.e., a cell's attribute value over its area in the original table—are quite different from each other, then the edge between them is deformed into a curved line. The shapes of cells with large values therefore become almost like circles, and the structure of a table in which each cell has quadrilateral shape is not preserved on table cartograms. Consequently, not all rows and columns can be clearly recognized in 'freely-deformed' table cartograms.

The other type is a quadrilateral table cartogram; here, the rectangular outline border is fixed and all cell edges are maintained as straight lines to preserve the quadrilateral shape of the cells. It is an attractive representation of table-form data, being more table-like with each row and column clearly distinguishable than on 'freely-deformed' table cartograms; this allows for a simpler interpretation of the characteristics of the represented data on the cartogram. Its construction method was first proposed in [5]. However, it is difficult to apply to new datasets because it consists of multiple operations, some of which do not have unique solutions and require complex settings.

In this article, we develop a new construction method of quadrilateral table cartograms. To create a construction method with high operability, the quadrilateral table cartogram construction is formulated as an optimization problem, based on the idea of [8], which proposed a contiguous area cartogram construction method. Because the large deformation of rows and columns in the resultant cartogram interrupts figure readers from recognizing the correspondence between the table cartogram and the original table, we attempt to maintain vertical and horizontal edges as much as possible. A brief explanation for the proposal has been made in [9], and this paper gives detailed explanation for the correspondence of the proposal to the idea of [8] and the formulation.

In addition, we propose a means of sorting rows and/or columns of original tables to reduce the shape deformation of table cartograms. The shape deformation of table cartograms becomes prominent if the difference in attribute values of adjacent cells are larger, at which point sorting rows and/or columns based on their similarity is expected to mitigate the shape deformation. Some table-form data have meanings associated to the order of rows and columns; the periodic chemical table and time-series data tables are typical examples. However, not all orders of rows and columns in table-form datasets have meanings. By sorting the rows and/or columns based on their similarity of data distribution, it can reduce the shape deformation and make the table cartogram more interpretable. Hence, we formulate the problem of reordering rows and/or columns based on their similarity as a combinatorial optimization problem and solve it using the simulated annealing algorithm.

The proposed table cartogram construction method involves two types of optimization problems and is evaluated through by applying it to actual dataset.

2. Previous Studies on Table Cartograms

2.1. Table Cartogram, A Special Form of Area Cartograms

Area cartograms are distorted maps on which the attribute value of each region is represented by its size. They can be classified by two classificatory criteria: the representation of adjacency of regions, which can be for contiguous or non-contiguous area cartograms; and the representation of shapes of regions, which can be for 'simple-' or 'complex-' shaped area cartograms.

Non-contiguous area cartograms are relatively easy to construct. Circle cartograms [10,11] and rectangular cartograms [12–15] are examples of simple-shaped non-contiguous area cartograms; they represent each region by a simple shape, a circle or a rectangle, whose area is fit to its attribute value, and place them according to the geographic configuration of regions. Non-contiguous complex-shaped cartograms were proposed in [16]; for these, the regions whose sizes are modified to fit to the attribute values are placed on a plane considering the geographic configuration of regions without overlaps.

Complex-shaped contiguous area cartograms, on which each region is of a similar shape to one on the geographic map, are rather difficult to construct. Up to now, there have been several studies proposing their construction methods [1,2,7,8,17–25]. Among these, the most popular construction

method is the diffusion-based method [7], whose mathematical guarantees its operability. Compared to the complex-shaped contiguous area cartograms, studies on simple-shaped contiguous area cartograms are few. Methods of representing regions by rectangles were discussed in [26–28], and a method of representing regions by rectilinear polygons was discussed in [29]. Because the degree of freedom is smaller in the construction of a simple-shaped contiguous area cartogram than in a complex-shaped one, the construction is comparably more challenging.

A table cartogram, a visualization for table-form data, is a special contiguous area cartogram. It is a deformed table on which the sizes of cells are fit to the given positive values, with its cells having the same adjacency relationship to their corresponding cells in the original table. Table cartograms can be classified into two types based on the shapes of cells on them. One type, a free-shaped table cartogram, describes cells with curved edges, and is considered as a complex-shaped contiguous area cartogram. The other type, a quadrilateral table cartogram, describes cells with straight edges, and is categorized as a simple-shaped contiguous area cartogram.

2.2. Construction Methods of Table Cartograms

Because a free-shaped table cartogram can be considered as a complex-shaped contiguous area cartogram, the diffusion-based method [7] is used for its construction [4]. The diffusion-based method imitates the diffusion process of fluids with different initial 'densities', i.e., the ratio of attribute values over areas of regions. According to the diffusion principle, high-density fluids diffuse toward locations where low-density fluids exist. The diffusion process stops when the densities are equalized, and the areas where the fluids are diffused indicate the transformed regions on the contiguous area cartogram. When densities differ greatly by regions, the high-density regions are transformed balloon-like shapes [23], and the low-density regions are reduced to long and narrow streamline shapes. When it is applied to the construction of table cartograms, the shape of an entire table is also deformed. If the data values in the table exhibit differences large enough deform the table significantly, it might be difficult to see the correspondence of rows, columns, and cells on the table cartogram to those of the original table.

It is worth to note the difference between data visualization by contiguous area cartograms and data visualization by table cartograms. When visualizing table-form data by table cartograms, it is important to preserve the table-form structure, in which horizontal and vertical edges form cells. When people look at contiguous area cartograms, the shapes of regions—although deformed—help them to recognize the correspondence between regions on cartograms and on geographical maps. However, the shapes of cells on table cartograms are not as useful because all cells have the same shapes. Hence, it is important to preserve the rows and columns visible on table cartograms.

A quadrilateral table cartogram, whose cells are represented by quadrilaterals, is one answer to improving the recognizability of rows and columns on table cartograms. A quadrilateral table cartogram and its construction were first proposed in [5]. The construction method consists of multiple segmentation and transformation operations. Because information on the sizes of quadrilaterals is not sufficient to determine their shapes, the construction of table cartograms, or any area cartograms, is inherently an ill-posed problem with no unique solutions. The method requires many parameters to be adjusted and is thus difficult to implement when solving this ill-posed construction problem.

2.3. Construction of Contiguous Area Cartograms based on Optimization

As has been described in the previous section, the construction of an area cartogram is an ill-posed problem with no unique answers. The importance of preserving the visibility of rows and columns to create easy-to-comprehend table cartograms has also been well noted.

The technique of controlling the shape deformation for the regularization of an ill-posed problem was utilized in [8], as well as the formulation of the construction of contiguous area cartograms as an optimization problem. This basic idea has been employed in the constructions of circle cartograms [11], rectangular cartograms [14,15], rectilinear cartograms [29], and distance cartograms [30] hitherto. This approach is also considered useful for the construction of quadrilateral

table cartograms, and we introduce the idea and formulation of contiguous area cartogram construction in this section.

The study introduces the regularization term that penalizes the change in bearing angles of edges to solve an ill-posed problem of contiguous area cartogram construction. The regularization term attempts to preserve the similarity of polygonal shapes as much as possible to enhance the readability of cartograms.

To formulate the problem as a simple equation, it is presumed that each polygon is divided into triangles. Let t_{ijk} denote a triangle consisting of three vertices i, j, and k, e_{mn} denote an edge that connects vertices m and n, T denote the set of triangles, E denote the set of edges, D_{ijk} denote an attribute value given to the triangle t_{ijk}, A_{ijk} denote the area of t_{ijk} with sign, and θ_{mn} and θ^G_{mn} denote the bearing angles of edge e_{mn} on the cartogram and on the geographical configuration, respectively. The formulation consists of an objective function that minimizes the differences between squares of areas and squares of data values of triangles and a regularization term that minimizes the changes in bearing angles of edges,

$$\min\left[\sum_{t_{ijk}\in T}\left(1-A_{ijk}^2\,/\,D_{ijk}^2\right)^2+\mu\sum_{e_{mn}\in E}\left(\theta_{mn}-\theta_{mn}^G\right)^2\right] \quad (1)$$

where μ is a weight on the regularization term. The objective function is intended to avoid calculating absolute values when computing the areas of triangles using their vertex coordinates. Equation (1) is obtained using the point coordinates,

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

where x^G_m and y^G_m denote the coordinates of vertex m on the geographical configuration, d_{mn} and d^G_{mn} denote the length of the edge e_{mn} on the cartogram and on the geographical configuration, respectively. This study solves the nonlinear minimization of Equation (2) by linearization. Equation (2) is linearized around the approximate position of vertex n, x'_n and y'_n, respectively, as

$$\min\left[\sum_{t_{ijk}\in T}\left\{1+\frac{3A_{ijk}'^2}{D_{ijk}^2}+\frac{A_{ijk}'^2}{D_{ijk}^2}\left(y_{jk}'x_i-y_{ik}'x_j+y_{ij}'x_k-x_{jk}'y_i+x_{ik}'y_j-x_{ij}'y_k\right)\right\}^2\right.$$
$$\left.+\mu\sum_{e_{mn}\in E}\left\{\frac{(x_n-x_m)y_{mn}'-x_{mn}'(y_n-y_m)}{d_{mn}'^2}\right\}^2\right] \quad (3)$$

where A'_{ijk} denotes the size of triangle by approximate position of vertices i, j, and k, d'_{mn} denotes a length of edges by approximate position of vertices m and n, and $x'_{ij}=x'_i-x'_j$. The solution of Equation (2) is obtained by an iterative process that solves Equation (3) and updates the solutions. The weight μ should be adjusted to balance between the control of shape deformation and the convergence speed of the iterative processes. The study proposes an algorithm to adjust the weight μ automatically.

3. Proposal of Construction of Quadrilateral Table Cartograms

Because quadrilateral table cartograms can be regarded as special contiguous area cartograms, the basic requirements for the construction of contiguous area cartograms are also applicable to them. In this article, we utilize the basic idea of contiguous area cartogram construction introduced in [8].

It must, however, be emphasized that there are differences between the construction of contiguous area cartograms and quadrilateral table cartograms. On contiguous area cartograms, the shape of each polygon is useful in identifying the corresponding polygon on the original geographic maps; however, on quadrilateral table cartograms, the shape of each cell is not useful in identifying

the corresponding cell on the original tables because every cell in the original table has the same shape. Consequently, the prevention of deformation of each cell during the quadrilateral table cartogram construction is of less importance.

To enhance the readability of quadrilateral table cartograms, it is important that the rows and columns are identifiable instead. Based on the basic idea in [8], which introduces the regularization term that controls the bearings of edges, this study introduces more strict control on the change in directions of horizontal and vertical edges of table cartograms.

In light of the foregoing discussion, this study sets the following three requirements for the construction of quadrilateral table cartograms.

1. To represent data precisely: the size of each cell should be close to its attribute value as much as possible.
2. To make each row and column recognizable on quadrilateral table cartograms: the direction of horizontal and vertical edges should be kept as much as possible.
3. To create a table-like figures: the outline border should be fixed.

The following section consists of two parts. Section 3.1 introduces a proposal for constructing quadrilateral table cartograms that satisfy the above requirements from a given table-form dataset. Additionally, Section 3.2 introduces a method of sorting rows and/or columns of original tables to construct less deformed table cartograms with high readability.

3.1. Construction

This section explains a two-step approach of constructing quadrilateral table cartograms. The requirement to preserve the direction of horizontal and vertical edges is emphasized in their construction. The first step creates figures that perfectly satisfy our requirements; it only adjusts heights of rows and widths of columns to fit each cell's size to its given values. The second step weakens the constraints on the directions of edges to increase the data representation precision; a similar formulation to that in [8] is set, although the regularization term that penalizes the change in bearings of edges is strengthened.

3.1.1. First step: Adjustment of Heights of Rows and Widths of Columns of Tables

The first step only adjusts the heights of rows and widths of columns to ensure that no changes are made in bearing angles of edges. Let N_r and N_c denote the number of rows and columns, respectively, and D_{ij} denote the attribute value given to the cell of row i and column j, enclosed by horizontal lines $i-1$ and i and vertical lines $j-1$ and j. The initial cells on the original table are set to squares of edge length l, and the table is a rectangular with the height of $N_r{\times}l$ and the width of $N_c{\times}l$ (Figure 1). Because the total size of a table is given by the total sum of attribute values, l is fixed to

$$l = \sqrt{\sum_{i=1}^{N_r}\sum_{j=1}^{N_c} D_{ij} \Big/ \left(N_r \times N_c\right)} \; . \tag{4}$$
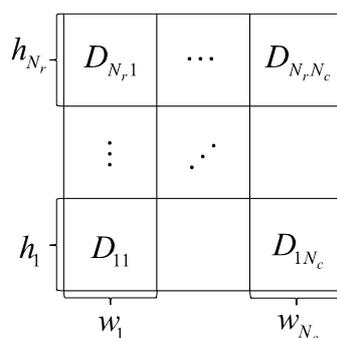

Figure 1. Original table.

Let h_i and w_j denote the height of row i and the width of column j, respectively. The formulation that fits the cell size to their attributes can be written as

$$\min_{\substack{h_1,\ldots,h_{N_r} \\ w_1,\ldots,w_{N_c}}} \sum_{i=1}^{N_r} \sum_{j=1}^{N_c} \left(D_{ij} - h_i w_j\right)^2 \text{ subject to } h_i > 0 \; \forall i, \; w_j > 0 \; \forall j, \; \sum_{i=1}^{N_r} h_i = N_r l, \; \sum_{j=1}^{N_c} w_j = N_c l \tag{5}$$

Now we rewrite Equation (4) using the x-coordinates of the vertical line j, x_j, and the y-coordinates of horizontal line i, y_i. Set the lower-left corner of table (x_0, y_0) at the origin $(0, 0)$, then Equation (4) can be rewritten as

$$\min_{\substack{y_1,\ldots,y_{N_r-1} \\ x_1,\ldots,x_{N_c-1}}} \sum_{i=1}^{N_r} \sum_{j=1}^{N_c} \left[D_{ij} - \left(y_i - y_{i-1}\right)\left(x_j - x_{j-1}\right)\right]^2$$

$$\text{subject to } y_i - y_{i-1} > 0 \; \forall i, \quad x_j - x_{j-1} > 0 \; \forall j, \quad x_0 = y_0 = 0, \quad y_{N_r} = N_r l, \quad x_{N_c} = N_c l. \tag{6}$$

The optimization problem of Equation (5) is solved by an iterative process that solves the linearized problem around the approximate solution of coordinates of vertices and updates the solution. The initial coordinate values are set by allocating the rows' heights based on the proportion of their total attribute values and the widths of columns based on the proportion of their total attribute values.

The vertical and horizontal lines are kept as straight lines and the shapes of all cells are rectangles on the resultant figure. It is easy to read the correspondence between the table cartogram and the original table; however, because the degrees of freedom are low, the data representation precision cannot be high.

3.1.2. Second Step: Deformation of Cells to Increase the Data Representation Precision

The second step weakens the constraints on edge directions to achieve higher data representation precision.

A new quadrilateral table cartogram construction method is developed based on Equation (3), and two modifications are made to fit the method to the quadrilateral table cartogram construction. Firstly, to fix the outer border of table cartograms, the x-coordinates of vertices on the horizontal outer edges and y-coordinates of vertices on the vertical outer edges are not set as parameters. Secondly, to preserve the directions of cell edges in quadrilateral table cartograms to their original horizontal or vertical directions, this study does not update the bearing angles in the regularization term from the previous solutions during the iteration processes of the minimization of linearized problems, although the bearing angles are updated in [8]. The minimization of linearized problems is written as

$$\min_{x,y} f = \min_{x,y} \left[\sum_{t_{ijk} \in T} \left\{ 1 + \frac{3A_{ijk}'^2}{D_{ijk}^2} + \frac{A_{ijk}'^2}{D_{ijk}^2} \left(y_{jk}' x_i - y_{ik}' x_j + y_{ij}' x_k - x_{jk}' y_i + x_{ik}' y_j - x_{ij}' y_k \right) \right\}^2 \right.$$

$$\left. + \mu \sum_{e_{mn} \in E} \left\{ \frac{\left(x_n - x_m\right) y_{mn}^* - x_{mn}^* \left(y_n - y_m\right)}{d_{mn}^{*2}} \right\}^2 \right] \tag{7}$$

where x^* and y^* denote the x- and y-coordinates of vertices of the output of first step, respectively. The iteration is repeated until the preset data representation precision is satisfied.

Here we discuss the computational complexity of the solution by Equation (6). An N_r by N_c table has $(N_r + 1) \times (N_c + 1)$ vertices, and the number of parameters is $2 \times (N_r N_c - 1)$ because the four corners and the outer border lines are fixed. We solve Equation (6) by formulating the linear equations in which the partial derivatives of objective function f with respect to the all parameters are zero. The solution is represented by the square matrix of $2 \times (N_r N_c - 1)$ rows and columns.

Focusing on vertex F, the partial derivatives of f with respect to x_F and y_F are written as

$$\frac{\partial f}{\partial x_F} = 2\left[\sum_{t_{Fjk}\in T}\frac{A_{Fjk}'^2 y_{jk}'}{D_{Fjk}^2}\left\{1+\frac{3A_{Fjk}'^2}{D_{Fjk}^2}+\frac{A_{Fjk}'^2}{D_{Fjk}^2}\left(y_{jk}'x_F-y_{Fk}'x_j+y_{Fj}'x_k-x_{jk}'y_F+x_{Fk}'y_j-x_{Fj}'y_k\right)\right\}\right.$$
$$\left.-\mu\sum_{e_{Fn}\in E}\frac{y_{Fn}^*}{d_{Fn}^{*2}}\left\{\frac{\left(x_n-x_F\right)y_{Fn}^*-x_{Fn}^*\left(y_n-y_F\right)}{d_{Fn}^{*2}}\right\}\right] \tag{8}$$

$$\frac{\partial f}{\partial y_F} = 2\left[\sum_{t_{Fjk}\in T}\frac{-A_{Fjk}'^2 x_{jk}'}{D_{Fjk}^2}\left\{1+\frac{3A_{Fjk}'^2}{D_{Fjk}^2}+\frac{A_{Fjk}'^2}{D_{Fjk}^2}\left(y_{jk}'x_F-y_{Fk}'x_j+y_{Fj}'x_k-x_{jk}'y_F+x_{Fk}'y_j-x_{Fj}'y_k\right)\right\}\right.$$
$$\left.+\mu\sum_{e_{Fn}\in E}\frac{x_{Fn}^*}{d_{Fn}^{*2}}\left\{\frac{\left(x_n-x_F\right)y_{Fn}^*-x_{Fn}^*\left(y_n-y_F\right)}{d_{Fn}^{*2}}\right\}\right] \tag{9}$$

The parameters that appear in Equations (7) and (8) are limited to the *x*- and *y*-coordinates of vertices that form the same triangles or edges with vertex *F*. Figure 2a shows the case when the number of parameters is a maximum of 18, and Figure 2b shows the case when the number of parameters is a minimum of 10 if surrounding vertices are not on the outer border of table cartograms. The number of nonzero coefficients is small in the linear system that is composed by Equations (7) and (8), then the system is represented by a sparse matrix.
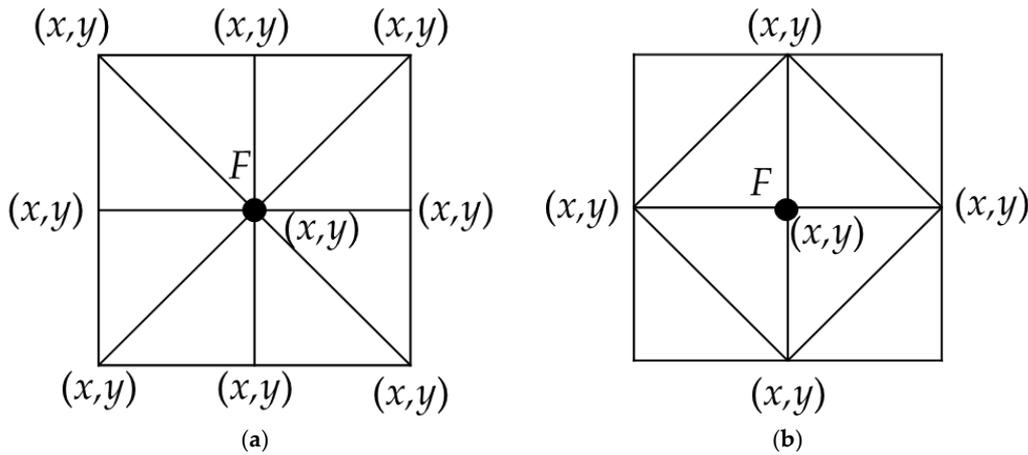

Figure 2. Parameters appear in partial derivative of objective function with respect to *x*- and *y*-coordinates of vertex *F*: (**a**) maximum number of parameters; (**b**) minimum number of parameters.

It is difficult to estimate the computational complexity of the solution as the structure of matrix is not ideal such as symmetric matrices. However, it is clear that the complexity of solution is much smaller than that of 'ordinary' linear equations, whose complexity is $O(n^3)$ where *n* is the number of parameters. The usage of sparse solver enables to apply the proposed method to the large table-form data.

3.2. Sorting of Rows and Columns

If data records with high similarity are placed near, the characteristics of data distribution become clearly visible on the table cartogram with small deformation. If rows and columns are randomly ordered from the viewpoint of similarity of data distribution, such as tables sorted in alphabetical order of names and ascending order of specific rows or columns, the table cartogram would be largely deformed and difficult to be interpreted.

Some tables, such as the periodic table of elements and the time-series data tables, have meanings associated with the order of rows and/or columns. The meanings of order indicate higher similarity of neighborhood cells and contribute to output table cartograms with smaller deformation. Previous studies only discussed constructing table cartograms of given tables without sorting rows and columns.

However, there are many tables that can be sorted to enhance the visibility of characteristics of data. Hence, in this study, we order the rows and/or columns of data based on the similarity of data distribution of cells, following the Kullback and Leibler (K–L) divergence [31].

The K-L divergence $D_{KL}(P \| Q) = \sum_i P(i) \ln(P(i)/Q(i))$ is a measure of the similarity between two probability distributions. This study proposes to search the order of rows and/or columns that minimizes the sum of K–L divergence between adjacent rows and/or columns in a table. Because $D_{KL}(P \| Q) \neq D_{KL}(Q \| P)$, we take an average of both to set the same evaluation measures for the adverse order.

Let $P(i, j)$ denote the proportion of data value of cell ij to the total data value of N_r by N_c table. The similarity indices of the given row order (S_r) and column order (S_c) are as follows.

$$S_r = \sum_{i=1}^{N_r-1} \sum_{j=1}^{N_c} \left(P(i,j) \ln\left(\frac{P(i,j)}{P(i+1,j)} \right) + P(i+1,j) \ln\left(\frac{P(i,j)}{P(i,j+1)} \right) \right) \tag{10}$$

$$S_c = \sum_{i=1}^{N_r} \sum_{j=1}^{N_c-1} \left(P(i,j) \ln\left(\frac{P(i,j)}{P(i,j+1)} \right) + P(i,j+1) \ln\left(\frac{P(i,j+1)}{P(i,j)} \right) \right) \tag{11}$$

The search of orders of rows or columns that minimize S_r or S_c is a combinatorial optimization problem. The simulated annealing (SA) algorithm, a probabilistic method to solve combinatorial optimization problems, is used to solve the problem.

4. Evaluation and Discussion

This section evaluates the proposed methods—the construction method of quadrilateral table cartogram from the given tables and the row (or column) sorting method—by applying them to the visualization of table-form data. The proposed methods were implemented in MATLAB 2019 by the MathWorks, Inc., MA, USA and tested on a computer with 2.93-GHz CPU.

4.1. Evaluation of a Quadrilateral Table Cartogram Construction Method

4.1.1. Initial Values of Weight μ

The effect of initial values of weight μ was tested by a simple 3×3 table. Figure 3a,b show, respectively, the original table for input and the table cartogram output by Step 1 with vertical and horizontal lines. Figure 3c–f show the resultant quadrilateral table cartograms by different initial μ values.
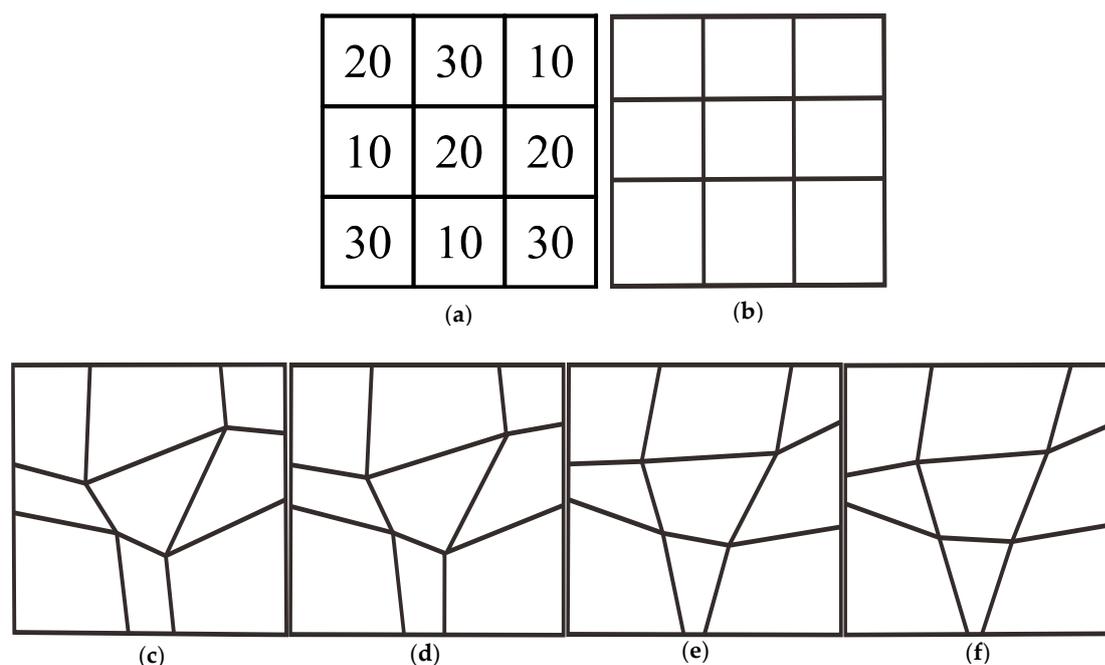
Figure 3. Quadrilateral table cartogram construction by different initial value of μ: (**a**) Original table; (**b**) Output of first step; (**c**) $\mu = 0.005$; (**d**) $\mu = 0.01$; (**e**) $\mu = 0.05$; (**f**) $\mu = 0.1$.

Table 1 indicates the root mean square error (RMSE) between values and sizes of cells as the indices of data representation precision, the maximal and minimal inner angles, the RMSE of bearing angles of edges between table cartograms and original tables, and the RMSE between right angles and inner angles as indices of deformations. When the deformation is small, inner angles approach right angles, and the bearings angles of edges approach horizontal or vertical directions.

It is confirmed that all the resultant figures represent the data at high precision; the RMSEs are smaller than 0.01% of the smallest data values. The shapes of resultant figures are different, but they preserve the table-like shapes. The larger initial weight μ outputs the figure with smallest deformation according to all the deformation indices in Table 1.

Table 1. Data representation precision and deformation with different values of μ

μ	RMSE between values and sizes of cells	Maximal inner angle (degree)	Minimal inner angle (degree)	RMSE of bearing angles of edges between on table cartogram and original table (degree)	RMSE between inner angles and right angles (degree)
0.005	5.05e-04	147.45	40.92	12.69	22.74
0.01	4.02e-04	141.58	41.74	11.71	21.44
0.05	3.32e-04	141.99	51.95	11.01	20.78
0.1	3.78e-04	133.68	50.54	10.81	20.29

4.1.2. Comparison with Inoue and Shimizu (2006) [8]

The proposed method was modified to reduce the bearing changes of horizontal and vertical edges, compared to the method in [8]. To check the validity of the modification, we compare the output of proposed method with that of the previous method that is modified to be able to fix the outline border.

This evaluation utilizes a 7×7 table that records the GDP of the main affected countries and regions by the Asian financial crisis in 1997. The data was obtained from a publication by the World Bank. Figure 4 indicates the outputs of two cartogram construction methods, and Table 2 indicates the indices of data representation precision and shape deformation.
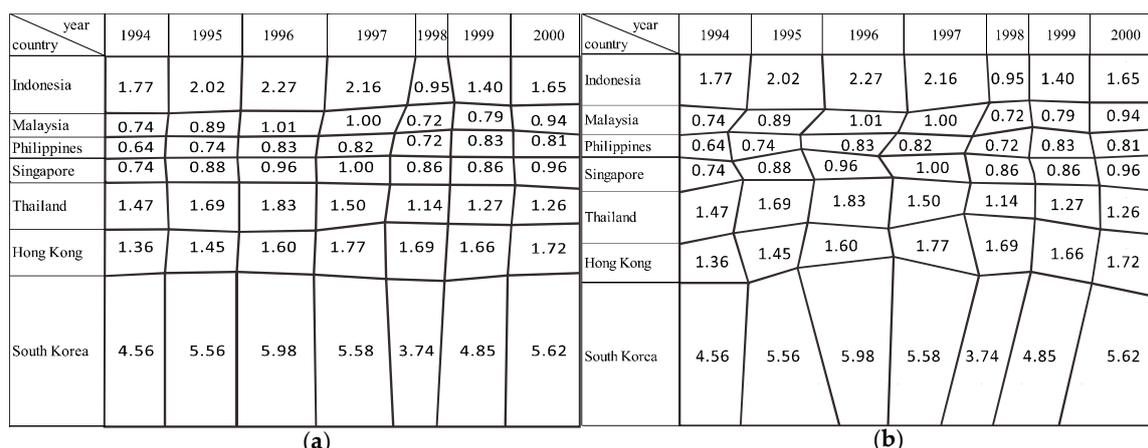
year country	1994	1995	1996	1997	1998	1999	2000	year country	1994	1995	1996	1997	1998	1999	2000
Indonesia	1.77	2.02	2.27	2.16	0.95	1.40	1.65	Indonesia	1.77	2.02	2.27	2.16	0.95	1.40	1.65
Malaysia	0.74	0.89	1.01	1.00	0.72	0.79	0.94	Malaysia	0.74	0.89	1.01	1.00	0.72	0.79	0.94
Philippines	0.64	0.74	0.83	0.82	0.72	0.83	0.81	Philippines	0.64	0.74	0.83	0.82	0.72	0.83	0.81
Singapore	0.74	0.88	0.96	1.00	0.86	0.86	0.96	Singapore	0.74	0.88	0.96	1.00	0.86	0.86	0.96
Thailand	1.47	1.69	1.83	1.50	1.14	1.27	1.26	Thailand	1.47	1.69	1.83	1.50	1.14	1.27	1.26
Hong Kong	1.36	1.45	1.60	1.77	1.69	1.66	1.72	Hong Kong	1.36	1.45	1.60	1.77	1.69	1.66	1.72
South Korea	4.56	5.56	5.98	5.58	3.74	4.85	5.62	South Korea	4.56	5.56	5.98	5.58	3.74	4.85	5.62

(**a**)　　　　　　　　　　　　　　　　　　(**b**)

Figure 4. Quadrilateral table cartograms of Gross Domestic Product(GDP) (10 billion current USD) of Eastern and Southeastern Asian countries before and after the Asian financial crisis in late 1990s: (**a**) Proposed method; (**b**) Previous method in [8].

Table 2. Comparison of data representation precision and deformation by the proposed method and the method in [8]

Methods	RMSE between GDP and sizes of cells (10 billion current USD)	Maximal inner angle (degree)	Minimal inner angle (degree)	RMSE of bearing angles of edges between on table cartogram and original table (degree)	RMSE between inner angles and right angles (degree)
Proposed method	2.93e-05	113.10	68.65	3.35	5.74
Inoue and Shimizu (2006)	3.72e-05	130.07	50.90	10.31	15.15

Although Table 2 indicates that the both figures represent values by areas of cells at high precision, it is apparent from Figure 4 that the proposed method output a lesser deformed table than that obtained using the Inoue-Shimizu method. The output obtained by the proposed method almost preserves the horizontal and vertical lines, and it is easy to recognize the characteristics of data distribution; the '1998' column shrinks drastically and the degree of decline is larger for Indonesia than for other countries and regions. It is difficult to read the characteristics from Figure 4b owing to the unnecessary deformation of vertical and horizontal lines, which hinders the understandings of figure readers.

Through the comparison, it is confirmed that the two-step procedure and the enforcement of controls on the bearing angle changes in edges by the proposed method is effective in enhancing the readability of quadrilateral table cartograms.

4.1.3. Comparison with Evans et al. (2018) [5]

The proposed method was applied to the visualization of properties of chemical elements, which is the most common applications of table cartograms. The quadrilateral table cartograms of properties of elements are presented in [5]. Figure 5 indicates the resultant table cartograms that visualize the relative thermal conductivity and the relative boiling points, and Table 3 shows the indices of the precision, deformation, and calculation times.
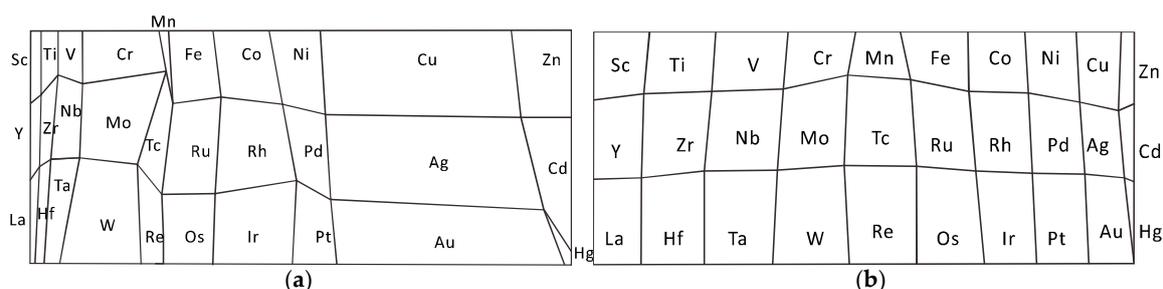
Figure 5. Quadrilateral table cartograms of periodic table of chemical elements: (**a**) Relative thermal conductivity; (**b**) Relative boiling point.

Table 3. Data representation precision, deformation and calculation time by the proposed method when $\mu = 0.15$

Data	RMSE between values and sizes of cells	Maximal inner angle (degree)	Minimal inner angle (degree)	RMSE of bearing angles of edges between on table cartogram and original table (degree)	RMSE between inner angles and right angles (degree)	Calculation time (sec)
Thermal conductivity	2.43e-05 (W·cm^{-1}·K^{-1})	175.49	11.04	15.92	25.51	1.84
Boiling point	0.0618 (℃)	117.51	66.40	5.32	9.63	3.79

The resultant figures are alike those obtained in [5], although the data representation precision is not discussed in that study. The RMSEs of Figure 5a,b is 0.031% and 0.017% of the minimum data values, respectively; it is confirmed that the proposed method can output table cartograms that represent data at high precision. The feasibility of proposed method is higher than that of the previous method as initial weight μ is the only parameter to be decided when constructing table cartograms. The different initial values of μ output the different shapes of table cartograms; however, because of the fast calculation of the proposed method, users can try several initial values of μ and select a figure based on the data representation precision, the indices of deformation, and the shape.

4.2. Evaluation of Sorting Based on the Similarity of Data

We evaluate the effect of sorting of rows and/or columns using the data of crude death rate per thousand people published by the World Bank. Figure 6 shows the resultant table cartograms. Figure 6a is constructed from the table whose rows are sorted by the proposed sorting method, and Figure 6b is constructed from the table whose rows are sorted by the alphabetical order of country names. Table 4 indicates the indices of data representation precision and shape deformation.
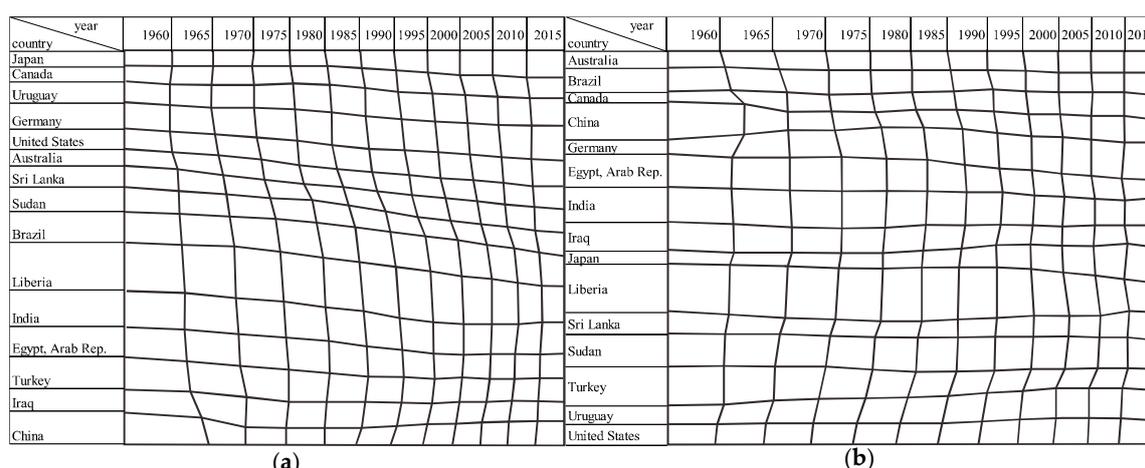
Figure 6. Quadrilateral table cartograms of death rate crude per thousand people by proposed method: (**a**) Rows are sorted by the proposed sorting method; (**b**) Alphabetical order of country names.

Table 4. Data representation precision and shape deformation using arranged table and unarranged table by the proposed method (initial μ = 10)

Dataset	RMSE between death rates and sizes of cells (%)	Maximal inner angle (degree)	Minimal inner angle (degree)	RMSE of bearing angles of edges between on table cartogram and original table (degree)	RMSE between inner angles and right angles (degree)
Sorted by proposed method	1.44e-04	129.05	57.15	6.12	11.32
Alphabetical order of country names	1.35e-04	140.45	50.88	5.37	8.49

Figure 6 clearly indicates that the table deformation becomes smaller when the table is sorted by the similarities of data distribution. Because countries with similar socio-economic condition are placed near, the sorting of rows (or columns) cannot only improve the appearance of table cartograms but also make the output more interpretable. The data representation precision and shape deformation in Table 4 show that the table cartogram from the sorted dataset and that from the original table have the same level of precision, while the shape deformation of the sorted table cartogram is smaller than the table cartogram with the original order of rows. We therefore confirm that the sorting of rows and/or columns reduces deformation, thereby enabling figure readers to focus on the change in shapes in table cartograms.

4.3. Discussion on the Applicability of Visualization by Quadrilateral Table Cartograms

A table cartogram is one of visualization options, such as coloring of cells and drawing of bar charts in cells. An advantage of the visualization by table cartograms is that they can indicate the differences of data sizes. Another advantage is that they can visualize two types of information simultaneously by classifying cells of table cartograms by colors. When visualizing time-series data, it would be effective to show the values by cell sizes and their rates of change by colors.

A quadrilateral table cartogram is useful to visualize table-form data that have clear characteristics of rows/columns. Figure 5a is the example; the data distribution patterns of columns, rows, and cells are clearly visualized by the linear edges. The cell sizes fit the data almost exactly,

even though the large difference exists in data size such that the thermal conductivity of silver (Ag) is 55 times as large as that of manganese (Mn). Even if the original table-form data do not indicate clear data distribution patterns, the quadrilateral table cartogram construction with sorting of rows/columns might be able to clarify and visualize the patterns, as is shown in Figure 6a.

However, if there is no clear pattern in the data distribution of the sorted table, a table cartogram is not an option for the visualization. When the distribution of data varies in rows or columns, it is expected to be difficult to fit cell sizes to their given data values at high precision. For example, an origin–destination matrix usually has large values in diagonal cells, and the similarity of data distribution between rows or columns is rather small. Consequently, it would be difficult to visualize the table by a quadrilateral table cartogram with fixed borders.

Although the proposed method is possible to fit the cell sizes to the data, the resultant figures are deformed largely, and difficult to be interpreted because readers cannot find any patterns in them. Figure 3 is the example. The values given to cells do not differ much, only three-times difference at maximum. The values are almost perfectly represented by the cell sizes, however, the changes in direction of edges in Figure 3 are larger than that in Figures 4,5b,6, where the differences of data values are larger than the data in Figure 3. As no clear patterns are shown in the figures, the figures are difficult to be interpreted.

As we discussed in Section 3.1.2, the proposed construction method is applicable to the large table-form data, although the sorting would take time to get the order with smaller similarity indices. However, the limitation of applying table cartograms to large table-form data lies in the effectiveness as a visualization tool. When the number of rows and columns are large, it would be difficult to find the data of interest.

The above discussion reveals the limitations in the usages of quadrilateral table cartograms.

5. Conclusion

In this article, we proposed a method to construct quadrilateral table cartograms from the given table-form data, along with a method to sort rows and columns of tables, to output table cartograms with small deformations.

The proposed method for the construction consists of two steps: the first step adjusts the heights of rows and widths of columns of tables by preserving the direction of horizontal and vertical edges; the second step transforms the shapes of cells to fit their sizes to their given data values, while controlling the degrees of deformation. Both steps were formulated as a non-linear minimization problem, and they are solved by iterative processes of linearization around the approximate solutions.

This study also proposed a method to sort rows and columns of tables according to the similarity of data distribution. Utilizing the K–L divergence as an index of similarity of rows and columns, the sorting problem was formulated as a combinatorial optimization problem and was solved using the simulated annealing algorithm.

Applying this method to the data confirmed that the proposed construction method can generate highly accurate quadrilateral table cartograms with smaller deformations of shape compared to previous construction methods, and the sorting of rows and/or columns enables the output of table cartograms with smaller deformations while improving the visibility of the characteristics of data distribution.

The proposed methods are applicable for the visualization of tables with positive values. However, as we discussed in Section 4.3, there are limitations where the visualization by quadrilateral table cartograms are not applicable. We evaluated the proposed method based on data representation precision and shape deformation; however, we did not analyze the interpretability of data. It is necessary to check the usefulness of visualization by quadrilateral table cartograms through the comparison with other data visualization methods by user studies. These discussions are left for future studies.

Author Contributions: Conceptualization, Ryo Inoue; Methodology, Ryo Inoue; Software, Mao Li; Validation, Mao Li and Ryo Inoue; Formal Analysis, Mao Li and Ryo Inoue; Investigation, Mao Li; Resources, Ryo Inoue;

References

1. Tobler, W.R. Thirty-five years of computer cartograms. *Ann. Am. Assoc. Geogr.* **2004**, *94*, 58–73, doi:10.1111/j.1467-8306.2004.09401004.x.
2. Nusrat, S.; Kobourov, S. The state of the art in cartograms. *Comput. Graph. Forum* **2016**, *35*, 619–642.
3. Sheehan, W.F. Periodic table of elements with emphasis. *Chemistry* **1976**, *49*, 17–18.
4. Winter, M.J. Diffusion cartogram for the display of periodic table data. *J. Chem. Educ.* **2011**, *88*, 1507–1510, doi:10.1021/ed1000203.
5. Evans, W.; Felsner, S.; Kaufmann, M.; Kobourov, S.G.; Mondal, D.; Nishat, R.I.; Verbeek, K. Table cartogram. *Comput. Geom.* **2018**, *68*, 174–185, doi:10.1016/j.comgeo.2017.06.010.
6. Element Scarcity—EuChemS Periodic Table. Available online: https://www.euchems.eu/euchems-periodic-table/ (accessed on 30 October 2019).
7. Gastner, M.T.; Newman, M.E.J. Diffusion-based method for producing density-equalizing maps. *PNAS* **2004**, *101*, 7499–7504, doi:10.1073/pnas.0400280101.
8. Inoue, R.; Shimizu, E. A new algorithm for continuous area cartogram construction with triangulation of regions and restriction on bearing changes of edges. *Cartogr. Geogr. Inf. Sci.* **2006**, *33*, 115–125, doi:10.1559/152304006777681698.
9. Li, M.; Inoue, R. Table cartogram generation as an optimization problem. *Abstr. Int. Cartogr. Assoc.* **2019**, *1*, 1–2, doi:10.5194/ica-abs-1-214-2019.
10. Dorling, D. *Area Cartograms: Their Use and Creation*, 1st ed.; Department of Geography, University of Bristol: Bristol, UK, 1996.
11. Inoue, R. A new construction method for circle cartograms. *Cartogr. Geogr. Inf. Sci.* **2011**, *38*, 146–152, doi:10.1559/15230406382146.
12. Rasiz, E. The rectangular statistical cartogram. *Geogr. Rev.* **1934**, *24*, 292–296, doi:10.2307/208794.
13. Upton, G.J.G. Rectangular cartograms, spatial autocorrelation, and interpolation. *J. Reg. Sci. Assoc. Int.* **1991**, *70*, 287–302, doi:10.1007/BF01434423.
14. Inoue, R. Generalized approach to construction of simple-shaped non-contiguous area cartograms. In *Proceedings of the 25th International Cartographic Conference*, Paris, France, 3–8 July 2011; pp. CO–272.
15. Inoue, R. An approach to the construction of simple-shaped non-contiguous area cartograms. *Theory Appl. GIS* **2012**, *20*, 11–22, doi:10.5638/thagis.20.11. (In Japanese)
16. Olson, J.M. Noncontiguous area cartogram. *Prof. Geogr.* **1976**, *28*, 371–380, doi:10.1111/j.0033-0124.1976.00371.x.
17. Tobler, W.R. A continuous transformation useful for districting. *Ann. N. Y. Acad. Sci.* **1973**, *219*, 215–220, doi:10.1111/j.1749-6632.1973.tb41401.x.
18. Tobler, W.R. Pseudo-cartograms. *Am. Cartogr.* **1986**, *13*, 43–50, doi:10.1559/152304086783900194.
19. Dougenik, J.A.; Chrisman, N.R.; Niemeyer, D.R. An algorithm to construct continuous area cartograms. *Prof. Geogr.* **1985**, *37*, 75–81, doi:10.1111/j.0033-0124.1985.00075.x.
20. Gusein-Zade, S.M.; Tikunov, V.S. A new technique for constructing continuous cartograms. *Cartogr. Geogra. Inf. Syst.* **1993**, *20*, 167–173, doi:10.1559/152304093782637424.
21. House, D.H.; Kocmoud, C.J. Continuous cartogram construction. *Proc. IEEE Symp. Inf. Vis.* **1998**, 197–204, doi:10.1109/VISUAL.1998.745303.
22. Keim, D.A.; North, S.C.; Panse, C. CartoDraw: A fast algorithm for generating contiguous cartograms. *IEEE Trans. Vis. Comput. Graph.* **2004**, *10*, 95–110, doi:10.1109/TVCG.2004.1260761.
23. Sun, S. A fast, free-form rubber-sheet algorithm for contiguous area cartograms. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 567–593, doi:10.1080/13658816.2012.709247.
24. Sun, S. An optimized rubber-sheet algorithm for continuous area cartograms. *Prof. Geogr* **2013**, *65*, 16–30, doi:10.1080/00330124.2011.639613.
25. Gastner, M.T.; Seguy, V.; More, P. Fast flow-based algorithm for creating density-equalizing map projections. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E2156–E2164, doi:10.1073/pnas.1712674115.

26. Heilmann, R.; Keim, D.A.; Panse, C.; Sips, M. RecMap: Rectangular map approximations. *Proc. IEEE InfoVis* **2004**, 33–40, doi:10.1109/INFVIS.2004.57.

27. Speckmann, B.; van Kreveld, M.; Florisson, S. A linear programming approach to rectangular cartograms. In *Progress in Spatial Data Handling*; Riedl, A., Kainz, W., Elmes, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 529–546.

28. van Kreveld, M.; Speckmann, B. On rectangular cartograms. *Comput. Geom.* **2007**, *37*, 175–187, doi:10.1016/j.comgeo.2006.06.002.

29. Inoue, R.; Kitaura, K.; Shimizu, E. New solution for construction of rectilinear area cartogram. In *Proceedings of the 24th International Cartographic Conference*, Santiago, Chile, 15–21 November 2009; pp. 20_3: 1–10.

30. Shimizu, E.; Inoue, R. A new algorithm for distance cartogram construction. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 1453–1470, doi:10.1080/13658810802186882.

31. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.