On Rectangular Cartograms

Marc van Kreveld [a] and Bettina Speckmann [b]

[a] *Inst. of Information & Computing Sciences, Utrecht University*

[b] *Dep. of Mathematics & Computer Science, TU Eindhoven*

Abstract

A rectangular cartogram is a type of map where every region is a rectangle. The size of the rectangles is chosen such that their areas represent a geographic variable (for example population). Rectangular cartograms are a useful tool to visualize statistical data. However, good cartograms are generally hard to generate: The area specifications for each rectangle may make it impossible to realize correct adjacencies between the regions and so hamper the intuitive understanding of the map.

Here we present the first fully automated algorithms for rectangular cartograms. Our algorithms depend on a precise formalization of region adjacencies and are building upon existing VLSI layout algorithms. Furthermore, we characterize a non-trivial class of rectangular subdivisions for which exact cartograms can be efficiently computed. An implementation of our algorithms and various tests show that in practice, visually pleasing rectangular cartograms with small cartographic error can be effectively generated.

1. Introduction

Cartograms. Cartograms are a useful and intuitive tool to visualize statistical data about a set of regions like countries, states or counties. The size of a region in a cartogram corresponds to a particular geographic variable [1,6]. Since the sizes of the regions are not their true sizes they generally cannot keep both their shape and their adjacencies. A good cartogram, however, preserves the recognizability in some way. Globally speaking, there are three types of cartogram. The standard type (the *contiguous area cartogram*) has deformed regions so that the desired sizes can be obtained and the adjacencies kept. Algorithms for such cartograms are described in [10,2,3]. The second type of cartogram is the non-contiguous area cartogram [7]. The regions have the true shape, but are scaled down and generally do not touch anymore. The third type of cartogram is the rectangular cartogram introduced by Raisz in 1934 [8] where each region is represented by a single rectangle.

Quality criteria. Whether a rectangular cartogram is good is determined by several factors. One of these is the *cartographic error* of the cartogram [2,3], which is defined for each region as $|A_c - A_s| / A_s$, where A_c is the area of the region in the cartogram and A_s is the area of that region as specified by the geographic variable to be displayed. The following list shows all quality criteria:

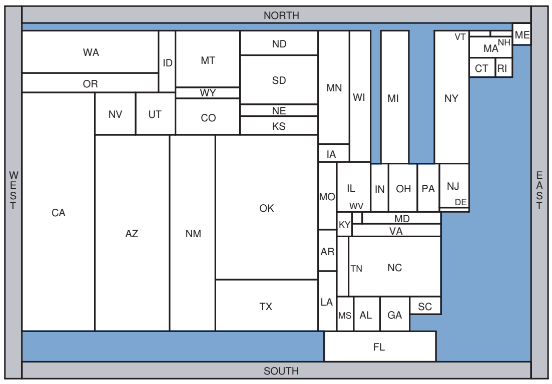

Fig. 1. A cartogram depicting the native population of the United States.

- Average cartographic error.
- Maximum cartographic error.
- Correct adjacencies of the rectangles.
- Maximum aspect ratio.
- Suitable relative positions.

For a purely rectangular cartogram we cannot expect to simultaneously satisfy all criteria well.

Related work. Rectangular cartograms are closely related to *floor plans* for electronic chips. Floor planning aims to represent a planar graph by its *rectangular dual*, defined as follows. A *rectangular partition* of a rectangle R is a partition of R into a set $\mathcal{R}$ of non-overlapping rectangles such no four rectangles in $\mathcal{R}$ meet at the same point. A rectangular dual of a planar graph (G, V) is a rectangular partition $\mathcal{R}$, such that (i) there is a one-to-one correspondence between the rectangles in $\mathcal{R}$ and the nodes in G, and (ii) two rectangles in $\mathcal{R}$ share a common boundary if and only if the corresponding nodes in G are connected. The following theorem was proven in [5]:

Theorem 1 *A planar graph G has a rectangular dual R with four rectangles on the boundary of R if and only if*

(i) every interior face is a triangle and the exterior face is a quadrangle

(ii) G has no separating triangles.

Note that although every triangulated planar graph without separating triangles has a rectangular dual this does not imply that an error free cartogram for this graph exists.

Results. We present the first fully automated algorithms for the computation of rectangular cartograms. We formalize the region adjacencies based on their geographic location and are so able to enumerate and process all feasible *rectangular layouts* for a particular subdivision (i.e., map). The precise steps that lead us from the input data to an algorithmically processable rectangular subdivision are sketched in Section 2.

We describe three algorithms that compute a cartogram from a rectangular layout. The first is an easy and efficient (*segment moving*) heuristic which we evaluated experimentally. The visually pleasing results of our implementation can be found in Section 3. Secondly, we show how to formulate the computation of a cartogram as a bilinear programming problem. The details concerning these two methods can be found in the full paper.

For our third, exact, algorithm we introduce an effective generalization of sliceable layouts, namely *L-shape destructible* layouts. We show that:

Theorem 2 *An L-shape destructible layout with a given set of area values has exactly one or no realization as a cartogram.*

The proof of Theorem 2 (which is given in the full paper) immediately implies an efficient algorithm to compute an exact (up to an arbitrarily small error) cartogram for subdivisions that arise from actual maps.

Finally we also establish the following theorem:

Theorem 3 *A subdivision S with n vertices whose face graph is outerplanar can be represented by a rectangular cartogram with the correct adjacencies and any pre-specified area values. This cartogram can be computed in linear time.*

2. Algorithmic Outline

Assume that we are given an administrative subdivision into a set of regions. The adjacencies of the regions can be represented in a graph F, which is the face graph of the subdivision.

1. Preprocessing: The face graph F is in most cases already triangulated (except for its outer face). In order to construct a rectangular dual of F we first have to process internal vertices of degree less than four and then triangulate any remaining non-triangular faces.

2. Directed edge labels: Any two nodes in the face graph have at least one direction of adjacency which follows naturally from their geographic location. While in theory there are four different directions of adjacency any two nodes can have, in practice only one or two directions are reasonable.

Our algorithms go through all possible combinations of direction assignments and determine which one gives a correct or the best result. While in theory there can be an exponential number of options, in practice there is often only one natural choice for the direction of adjacency between two regions. We call a particular choice of adjacency directions a *directed edge labeling*.

Definition 4 (Realizable directed edge labeling)

A face graph F with a directed edge labeling can be represented by a rectangular dual if and only if

(i) every internal region has at least one North, one South, one East, and one West neighbor.

(ii) when traversing the neighbors of a node in clockwise order starting at the western most North neighbor we first encounter all North neighbors, then all East neighbors, then all South neighbors and finally all West neighbors.

3. Rectangular layout: A realizable directed edge labeling constitutes a *regular edge labeling* for F as defined in [4]. Therefore we can employ the algorithm by He and Kant [4] to construct a *rectangular layout*, i.e., the unique rectangular dual of a realizable directed edge labeling. The output of our implementation of the algorithm by He and Kant is shown in Figure 2.

4. Area assignment: For a given set of area values and a given rectangular layout we would like to decide if an exact assignment of the area values to the regions is possible without destroying

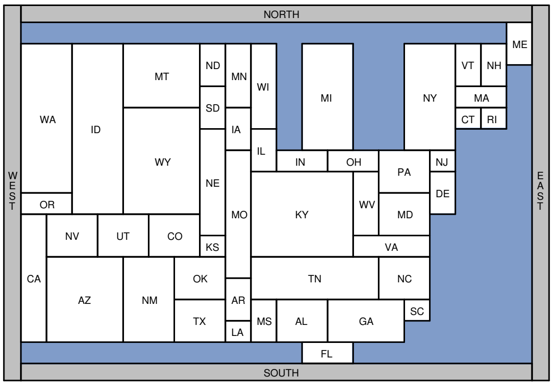

Fig. 2. One of 4608 possible rectangular layouts of the US.

the correct adjacencies. Should the answer be negative or should the question be undecidable, then we still want to compute a cartogram that has a small amount of cartographic error while maintaining reasonable aspect ratios and relative positions.

Exact algorithm. For certain types of rectangular layouts we can compute an exact cartogram. We first determine for a given rectangular layout a *maximal rectangle hierarchy.* The maximal rectangle hierarchy groups rectangles that together form a larger rectangle, as illustrated in Figure 3. It can be computed in linear time [9]. All groups in the hierarchy are independent and we will determine areas separately for each group.

A node of degree 2 in the hierarchy corresponds to a sliceable group of rectangles. If the maximal rectangle hierarchy consists of slicing cuts only, then we can (in a top-down manner) compute the unique position of each slicing cut.

Nodes of a degree higher than 2 require more complex cuts (see for example the four thick segments in Fig. 3). Here we introduce a type of non-sliceable layout for which the coordinates are still uniquely determined by the specified areas. We say a rectangular layout $\mathcal{R}$ is *irreducible* if no proper

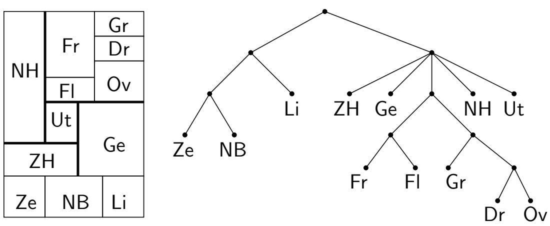

Fig. 3. Rectangular layout of the 12 provinces of the Netherlands and a corresponding maximal rectangle hierarchy.

subset of $\mathcal{R}$ (of size greater than one) forms a rectangle. Furthermore, we call a rectilinear simple polygon with at most 6 vertices *L-shaped.* We say that an L-shaped polygon is *rooted* at a vertex p if one of its convex vertices is p.

Definition 5 (L-shape destructible) *An irreducible rectangular layout $\mathcal{R}$ of a rectangle R is* L-shape destructible *if there is a sequence starting at a corner s of R in which the rectangles of $\mathcal{R}$ can be removed from R such that the remainder forms an L-shaped polygon rooted at the corner t of R opposite to s after each removal.*

Fig. 4. An L-shape destructible layout of the Eastern US. The shaded area shows the L-shaped polygon after the removal of rectangles $1-4$.

3. Implementation and experiments

We have implemented the segment moving heuristic and tested it on some data sets. The main objective was to discover whether rectangular cartograms with reasonably small cartographic error exist, given that they are rather restrictive in the possibilities to represent all rectangle areas correctly. Secondary objectives were to determine how the cartographic error depends on maximum aspect ratio and correct or false adjacencies.

Our layout data sets consist of the 12 provinces of the Netherlands and the 48 contiguous states of the USA. Here we present only some of the results pertaining to the US data sets, all other results can be found in the full paper. We allowed 13 pairs of adjacent states of the USA to be in different relative positions leading to 8192 possible layouts. Only 4608 of these correspond to a realizable directed edge labeling. We considered all 4608 layouts and chose the one giving the lowest average error as the cartogram.

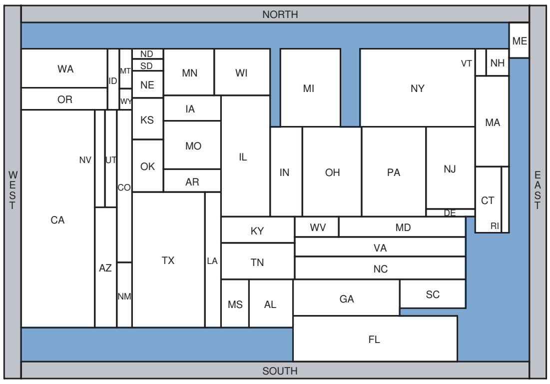

Fig. 5. The population of the US.

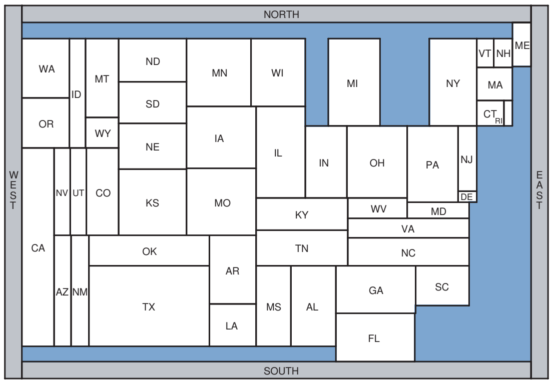

Fig. 6. The highway kilometers of the US.

Data set	Adjacency	Aspect ratio	Av. error	Max.
US pop.	false	8	0.104	0.278
US pop.	false	9	0.085	0.193
US pop.	false	10	0.052	0.295
US pop.	false	11	0.030	0.091
US pop.	false	12	0.022	0.056
US pop.	correct	12	0.327	0.618
US pop.	correct	13	0.319	0.608
US pop.	correct	14	0.317	0.612
US pop.	correct	15	0.314	0.569
US pop.	correct	16	0.308	0.612
US hw.	correct	6	0.073	0.188
US hw.	correct	7	0.059	0.111
US hw.	correct	8	0.058	0.101
US hw.	correct	9	0.058	0.101
US hw.	correct	10	0.058	0.101

Table 1. Errors for different aspect ratios, and correct/false adjacencies. Sea 20%.

As numeric data we considered *population*, *native population*, number of *farms*, number of *electorial college votes*, and total length of *highways*.

Table 1 shows errors for various settings for two US data sets. Only the average error is significant in the table, since the rectangular layout chosen for the table is the one with lowest average error. The corresponding maximum error is shown only for completeness. Since cartograms are interpreted visually and show a global picture, errors of a few percent on the average are acceptable. Such errors are also present in standard contiguous cartograms.

The error decreases with a larger aspect ratio, as expected. For the native population data set an aspect ratio of 7 combined with false adjacency results in a cartogram with average error below 0.04 (see Fig. 1). Figures 5 and 6 show rectangular cartograms for two of the five US data sets. The data sets allowed an aspect ratio of 10 or lower to yield an average error between 0.03 and 0.06.

References

[1] B. Dent. *Cartography - thematic map design*. McGraw-Hill, 5th edition, 1999.

[2] J. A. Dougenik, N. R. Chrisman, and D. R. Niemeyer. An algorithm to construct continous area cartograms. *Professional Geographer*, 37:75–81, 1985.

[3] H. Edelsbrunner and E. Waupotitsch. A combinatorial approach to cartograms. *Comput. Geom. Theory Appl.*, 7:343–360, 1997.

[4] G. Kant and X. He. Regular edge labeling of 4-connected plane graphs and its applications in graph drawing problems. *Theor. Comp. Sci.*, 172:175–193, 1997.

[5] K. Koźmiński and E. Kinnen. Rectangular dual of planar graphs. *Networks*, 5:145–157, 1985.

[6] NCGIA / USGS. Cartogram Central, 2002. `http://www.ncgia.ucsb.edu/projects/Cartogram_Central/index.html`.

[7] J. Olson. Noncontiguous area cartograms. *Professional Geographer*, 28:371–380, 1976.

[8] E. Raisz. The rectangular statistical cartogram. *The Geographical Review*, 24:292–296, 1934.

[9] S. Sur-Kolay and B. Bhattacharya. The cycle structure of channel graphs in nonslicable floorplans and a unified algorithm for feasible routing order. In *Proc. IEEE International Conference on Computer Design*, pages 524–527, 1991.

[10] W.Tobler. Pseudo-cartograms. *The American Cartographer*, 13:43–50, 1986.