Automated Construction of Rectangular Cartograms[1]

Marc van Kreveld, Bettina Speckmann

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 (e.g., population). Good rectangular cartograms are 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.
We present the first algorithms for rectangular cartogram construction. Our algorithms depend on a precise formalization of region adjacencies and build upon existing VLSI layout algorithms. An implementation of our algorithms and various tests show that in practice, visually pleasing rectangular cartograms with small cartographic error can be generated effectively.

Introduction

Cartographers have developed many different techniques to visualize statistical data about a set of regions like countries, states or counties. *Cartograms* are among the most well known and widely used of these techniques. The regions of a cartogram are deformed such that the area of a region corresponds to a particular geographic variable [3]. The most common variable is population: In a population cartogram, the areas of the regions are proportional to their population. 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 [4, 5, 8, 15]. The second type of cartogram is the non-contiguous area cartogram [12]. 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 [13], where each region is represented by a rectangle. This has the advantage that the sizes (area) of the regions can be estimated much better than with the first two types.

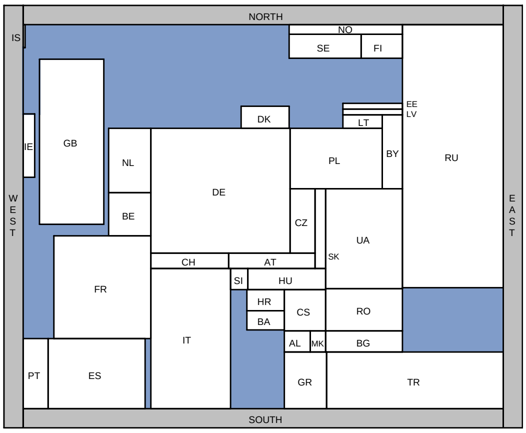

Figure 1: The population of Europe (country codes according to the ISO 3611 standard).

Algorithms for cartograms have been studied for over thirty years, but no method for producing rectangular cartograms has been developed so far [16].

[1]This work has previously been published as part of [11].

Quality criteria. Whether a rectangular cartogram is good is determined by several factors. One of these is the *cartographic error* [4, 5], which is defined for each region as $\mid A_c - A_s \mid / A_s$, where A_c is the area of the region in the cartogram and A_s is the specified area of that region, given by the geographic variable to be shown. The following list summarizes the most important quality criteria:

- 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. Recently, Heilmann et al. [6] presented rectangular map approximations that have zero cartographic error but do not satisfy the other criteria.

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 that 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 [10]:

Theorem 0.1 *A planar graph G has a rectangular dual R with four rectangles on the boundary of R if and only if*

1. *every interior face is a triangle and the exterior face is a quadrangle, and*
2. *G has no separating triangles.*

Most maps give rise to triangulated graphs, because usually at most three regions meet at any one point. Separating triangles occasionally arise, for example, Luxembourg does not border any sea and is incident to only three countries. This implies that a purely rectangular cartogram with correct adjacencies does not exist for Europe. Also 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 .

In [11] we describe three algorithms that compute a cartogram from a rectangular layout. Here we concentrate on the simplest one of these, the so-called *segment moving heuristic*. We evaluated this easy and efficient heuristic experimentally. The results of our implementation can be found in Section . A Java prototype can be seen at `http://www.win.tue.nl/~speckman/demos/carto`.

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 *di-*

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

1. every internal region has at least one North, one South, one East, and one West neighbor, and
2. 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.

A realizable directed edge labeling constitutes a *regular edge labeling* for F as defined in [7] which immediately implies our observation.

3. Rectangular layout: To actually represent a face graph together with a realizable directed edge labeling as a rectangular dual we have to pay special attention to the nodes on the outer face since they may miss neighbors in up to three directions. To compensate for that we add four special regions NORTH, EAST, SOUTH, and WEST, as well as *sea regions* that help to preserve the original outline of the subdivision. Then we can employ the algorithm by He and Kant [7] 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.

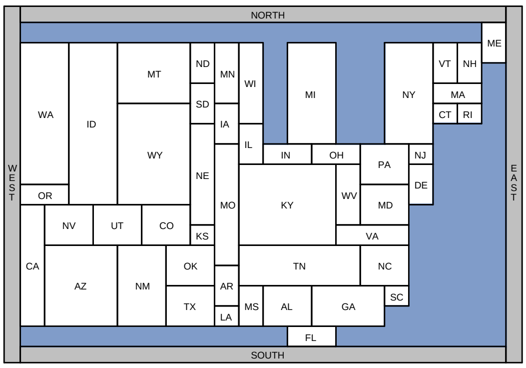

Figure 2: One of 4608 possible rectangular layouts of the US.

4. Area assignment: For a given set of area values and a given rectangular layout we would like to decide if an assignment of the area values to the regions is possible without destroying 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 cartographic error while maintaining reasonable aspect ratios and relative positions.

Segment moving heuristic. A simple but efficient heuristic that works as follows. Consider the maximal vertical segments and maximal horizontal segments in the layout, for example the vertical segment in Figure 2 that has Kentucky (KY) to its left and West Virginia (WV) and Virginia (VA) to its right. This segment can be moved a little to the left, making Kentucky smaller and the Virginias larger, or it can be moved to the right with the opposite effect.

The segment moving heuristic loops over all maximal segments and moves each with a small step in the direction that decreases the maximum error of the adjacent regions. After a number of iterations, one can expect that all maximal segments have moved to a locally optimal position. However, we have no proof that the method reaches the global optimum, or that it even converges.

The segment moving heuristic has some important advantages: (i) it can be used for any rectangular layout, (ii) one iterative step for all maximal segments takes $O(n)$ time, (iii) no area need to be specified for sea rectangles, (iv) a bound on the aspect ratio can be specified, and (v) adjacencies between the rectangles can be preserved, but need not be. Not preserving adjacencies can help to reduce cartographic error.

Implementation and experiments

We have implemented the segment moving heuristic and tested it on several 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. Obviously, we can only answer this question if the segment moving heuristic actually finds a good cartogram if it exist. Secondary objectives of the experiments are to determine to what extent the cartographic error depends on maximum aspect ratio and correct or false adjacencies. We were also interested in the dependency of the error on the percentage of area used by the sea.

Our layout data sets consist of the 36 countries of Europe and the 48 contiguous states of the USA. For

Data set	Sea	Aspect ratio	Ave. error	Max. error
Eu elec.	20%	8	0.071	0.280
Eu elec.	20%	9	0.070	0.183
Eu elec.	20%	10	0.067	0.179
Eu elec.	20%	11	0.065	0.155
Eu elec.	20%	12	0.054	0.137
Eu elec.	10%	10	0.098	0.320
Eu elec.	15%	10	0.076	0.245
Eu elec.	20%	10	0.067	0.179
Eu elec.	25%	10	0.049	0.126

Table 1: Errors for different aspect ratios and sea percentages (correct adjacencies).

Europe, we joined Belgium and Luxembourg, and Ukraine and Moldova, because rectangular duals do not exist if Luxembourg or Moldova are included as a separate country. Europe has 16 sea rectangles and the US data set has 9. For Europe we allowed 10 pairs of adjacent countries to be in different relative position, leading to 1024 possible layouts. Of these, 768 correspond to a realizable directed edge labeling. For the USA we have 13 pairs, 8192 possible layouts, and 4608 of these are realizable. In the experiments, all 768 or 4608 layouts are considered and the one giving the lowest average error is chosen as the cartogram.

As numeric data we considered for Europe the *population* and the *electricity production*, taken from [2]. For the USA we considered *population*, *native population*, number of *farms*, and total length of *highways*. The data is provided by the US census bureau in the *Statistical Abstract of the United States*.[1]

Preliminary tests on all data sets showed that the false adjacency option always gives considerably lower error than correct adjacencies. The false adjacency option always allowed cartograms with average error of only a few percent. A small part of the errors is due to the discrete steps taken when moving the segments. Since cartograms are interpreted visually and show a global picture, errors of a few percent on the average are acceptable. Errors of a few percent are also present in standard, computer-generated contiguous cartograms [4, 5, 8, 9]. We note that most hand-made rectangular cartograms also have false adjacencies and that aspect ratios of more than 20 can be observed.

[1] http://www.census.gov/statab/www/

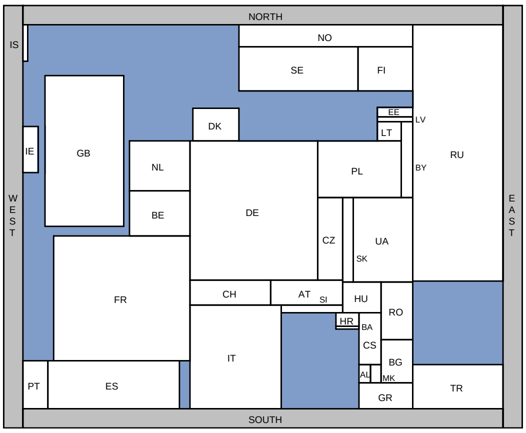

Figure 3: A cartogram depicting the electricity production of Europe.

Table 1 shows errors for various settings for the electricity production data set. The rectangular layout chosen for the table is the one with lowest average error. The corresponding maximum error is shown only for completeness. In the table we observe that the error goes down with a larger allowed aspect ratio, as expected. For Europe and population (not shown in the table), errors below 0.1 on the average with correct adjacencies were only obtained for aspect ratios greater than 15. The table also shows that a larger sea percentage brings the error down. This is as expected because sea rectangles can grow or shrink to reduce the error of adjacent countries, while a sea rectangle cannot have an error in its area. So, more sea means more freedom to reduce errors. However, sea rectangles should not become so small that they visually (nearly) disappear.

Table 2 shows errors for various settings for two US data sets. Again, we choose the rectangular layout giving the lowest average error. In the US highway

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

Table 2: Errors for different aspect ratios, and correct or false adjacencies. Sea 20%.

data set, aspect ratios above 8 do not seem to decrease the error below a certain value. In the US population data set, correct adjacencies give a larger error than is acceptable. Even an aspect ratio of 40 gave an average error of over 0.3. We ran the same tests for the native population data and again observed that the error decreases with larger aspect ratio. An aspect ratio of 7 combined with false adjacency gives a cartogram with average error below 0.04 (see Fig. 4). Only the highways data allowed correct adjacencies, small aspect ratio, and small error simultaneously.

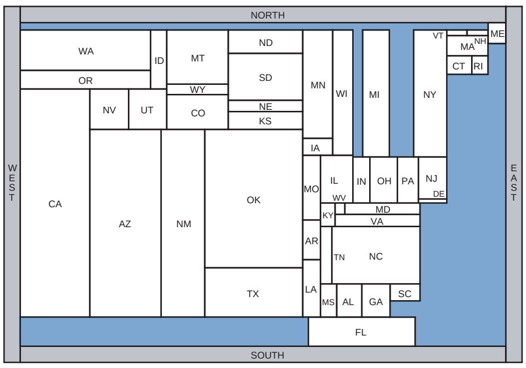

Figure 4: A cartogram depicting the native population of the United States.

Figures 4, 5, 6, and 7 show rectangular cartograms of the US. Three of them have false adjacencies, but we can observe that adjacencies are only slightly disturbed in all cases (which is the same as for hand-made rectangular cartograms). The data sets allowed an aspect ratio of 10 or lower to yield an average error between 0.03 and 0.06, except for the farms data. Here an aspect ratio of 20 gives an average error of just below 0.1. Figures 1 and 3 show rectangular cartograms for Europe. The former has false adjacencies and aspect ratio bounded by 12, the latter has correct adjacencies and aspect ratio bounded by 8. The average error is roughly 0.06 in both cartograms.

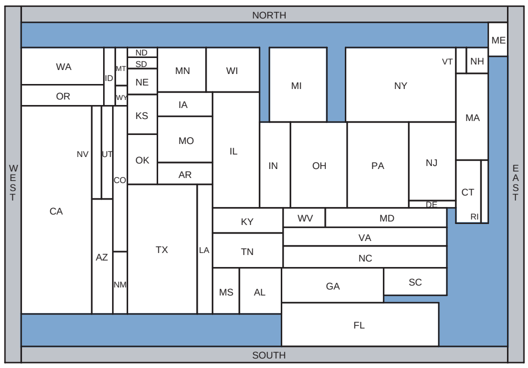

Figure 5: The population of the US.

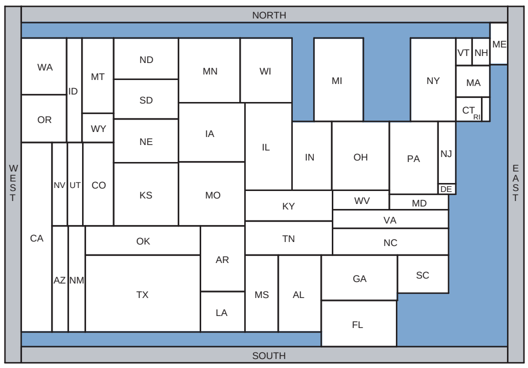

Figure 6: The highway kilometers of the US.

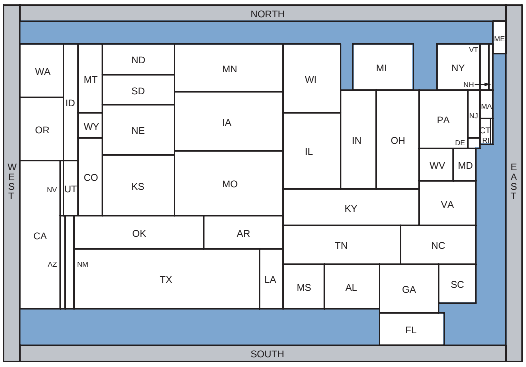

Figure 7: The farms of the US.

Conclusion

In this paper we presented the first algorithms to compute rectangular cartograms. We showed how to formalize region adjacencies in order to generate algorithmically processable layouts. An interesting open problem is whether rectangular cartogram construction (correct or minimum error) can be done in polynomial time.

We experimentally studied the quality of our segment moving heuristic and showed that it is very effective in producing aesthetic rectangular cartograms with only small cartographic error. Our tests show the dependency of the error on the aspect ratio, correct adjacencies, and sea percentage. The quality of the cartograms generated is comparable to hand-made rectangular cartograms.

References

[1] M. Bazaraa, H. Sherali, and C. Shetty. *Nonlinear Programming - Theory and Algorithms*. John Wiley & Sons, Hoboken, NJ, 2nd edition, 1993.

[2] *De Grote Bosatlas*. Wolters-Noordhoff, Groningen, 52nd edition, 2001.

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

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

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

[6] R. Heilmann, D. A. Keim, C. Panse, and M. Sips. Recmap: Rectangular map approximations. In *Proceedings of the IEEE Symposium on Information Visualization (INFOVIS)*, pages 33–40, 2004.

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

[8] D. Keim, S. North, and C. Panse. Cartodraw: A fast algorithm for generating contiguous cartograms. *IEEE Transactions on Visualization and Computer Graphics*, 10:95–110, 2004.

[9] C. Kocmoud and D. House. A constraint-based approach to constructing continuous cartograms. In *Proceedings of the Symposium on Spatial Data Handling*, pages 236–246, 1998.

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

[11] M. v. Kreveld and B. Speckmann. On rectangular cartograms. *Computational Geometry: Theory and Applications*, 2005. (to appear).

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

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

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

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

[16] W. Tobler. Thirty-five years of computer cartograms. *Annals of the Association of American Geographers*, 94(1):58–71, 2004.

Contact Information

Marc van Kreveld

Institute for Information and Computing Sciences
Utrecht University
P.O. Box 80.089 3508 TB Utrecht
The Netherlands
`marc@cs.uu.nl`

Bettina Speckmann

Department of Mathematics and Computer Science
TU Eindhoven
P.O. Box 513
5600 MB Eindhoven
The Netherlands
`speckman@win.tue.nl`