ABSTRACT

In 2004 van Kreveld and Speckmann presented the first algorithms for the automated construction of rectangular cartograms. The first step for these algorithms is to construct a partition of a large rectangle - the map - into a set of smaller rectangles that represent the regions of the administrative subdivision to be depicted ("countries") as well as (parts of) adjacent oceans, large lakes, or neighboring regions ("sea regions"). Countries have a specified target area that is based on a geographic variable like population. Sea rectangles do not have a specified target area but solely serve to increase the recognizability of the map by preserving a global coast or border outline.

In this paper we investigate different ways of partitioning sea regions into rectangles. We also study the effects of different cartographic error measures for sea rectangles. Our experimental results are given both in tabular form and as cartograms, to be able to evaluate both the error and the visual quality.

1. INTRODUCTION

Cartograms, which are also referred to as value-by-area maps, are a useful and intuitive tool to visualize statistical data about a set of regions like countries, states or provinces. The size of a region in a cartogram corresponds to a particular geographic variable [1]. 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 four 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 [3][4][7][10]. The second type of cartogram is the non-contiguous area cartogram [8]. 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 [9], where each region is represented by a rectangle (see Figure 1). This has the advantage that the sizes (area) of the regions can be estimated much better than with the first two types. The fourth type of cartogram is based on circles [2]. Finally, hybrid versions of these cartogram types also exist.

Tobler states in a recent survey, “Thirty-five years of computer cartograms” [11], that none of the existing cartogram algorithms are capable of generating rectangular cartograms. However, even more recently the authors of this paper presented the first algorithms for rectangular cartogram construction [12].

Figure 1: The population of Europe (country codes according to the ISO 3611 standard).
Quality criteria. Whether a rectangular cartogram is good is determined by several factors. One of these is the cartographic error [3][4], 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 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.
- Reasonable outer shape, or coastline.

For a purely rectangular cartogram we cannot expect to simultaneously satisfy all criteria well. Recently Heilmann et al. [5] presented rectangular map approximations that have zero cartographic error but do not satisfy the other criteria.

Cartogram construction. Following [12] we first formalize the region adjacencies based on their geographic location and then use this information together with an algorithm by Kant and He [6] to construct a rectangular layout (see for example Figure 2). A rectangular layout is a map where each region (country, state, etc.) is represented by a rectangle and adjacent regions are depicted by adjacent rectangles. We also use additional sea rectangles that help to preserve the original shape of the coastline of the regions.

We then apply one of three methods to try and give the rectangles their desired sizes. The simplest one of these is the segment moving heuristic. The segment moving heuristic loops over all maximal segments in the layout 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, although it often gives aesthetically pleasing cartograms with small error.

In our previous paper we presented various test results for rectangular cartograms for the United States and for the countries of Europe. We did not consider which role the sea regions play with respect to the visual quality or the error of the cartogram that was computed. For the USA, where the outer shape is nearly rectangular, one can expect that the sea does not play a large role. But for Europe, several sea regions of complex shape are needed to guarantee a reasonable coastline shape for the countries. The segment moving heuristic requires that the sea region is partitioned into rectangles as well. This can be done in several ways, leading to different rectangular layouts. We address the question how different layouts lead to cartograms with different errors and visual quality. Secondly, we showed in [12] that allowing a mild form of incorrect adjacencies between rectangles results in cartograms with considerably lower error. We examine the question whether false adjacencies of only sea rectangles already lead to rectangular cartograms with low error. This results in three options for the adjacencies, which we compare. Thirdly, the segment moving heuristic seems to work better if sea rectangles also have an error that is taken into account. On the other hand, this error should not be equally important as error in country rectangles, because sea rectangles do not play a role in the human interpretation of a cartogram. We test how much less important the error of sea rectangles should be with respect to country rectangles.

Section 2 provides more details on the segment moving heuristic and on the options that our algorithm allows. Section 3 states the research questions addressed in this paper more explicitly and describes the experimental set-up. Section 4 contains the results and analysis of the tests. Section 5 summarizes and concludes the paper.

2. THE SEGMENT MOVING HEURISTIC

The segment moving heuristic for computing rectangular cartograms starts with a rectangular layout: a partitioning of a rectangular region into rectangles. The layout consists of a set of maximal horizontal and vertical segments. Any horizontal segment has one or more rectangles directly above it and one or more rectangles directly below it. Moving such a segment a unit up or down influences the size of these rectangles only. We move a segment a unit up or down if: (i) no undesired adjacencies are created or removed, (ii) the maximum error of the adjacent rectangles goes down, and (iii) no adjacent country rectangle gets an aspect ratio larger than specified.
For example, consider the horizontal segment in Figure 2 with Germany (DE) and the Czech Republic (CZ) above it, and Switzerland (CH) and Austria (AT) below it. Moving this segment up or down influences the errors and aspect ratios of these four countries only. Undesired adjacencies that may be created are between Belgium (BE) and Switzerland, or between the Czech Republic and Hungary (HU).

The segment moving heuristic iterates over all maximal horizontal and vertical segments, moving each a unit up or down if the three conditions are met. The total number of iterations determines how often each maximal segment is tested in total. The following settings are needed to generate a rectangular cartogram from a rectangular layout:

- Number of iterations.
- Maximum aspect ratio (height:width or width:height ratio, whichever is at least 1).
- Percentage of the map area for the sea.
- Sea error division value (explained later).
- Adjacency options for rectangles: correct for sea and country rectangles, correct only for country rectangles, possibly incorrect for all rectangles.

The number of iterations is chosen in such a way that no more improvements are obtained in further iterations. In all cases, we needed less than 500 iterations on a map of 800 by 600 units.

**Initial layout.** Several choices need to be made during the cartogram construction. First of all, we must choose an initial rectangular layout. The choices are often clear for rectangles that represent countries. For example, the Netherlands should be represented by a rectangle that is west of the rectangle for Germany. But for the seas and other water regions it is less clear which partition into rectangles to choose. Currently the adjacencies of all sea and country rectangles are chosen manually during preprocessing. We expect that the sea partition influences the quality of the cartogram that will be obtained, but it is not clear to what extent. We examine this question in this paper. Three initial layouts with different sea partitions for Europe are shown in Figure 2 (mixed layout) and Figure 3.

**Figure 2:** A rectangular layout for Europe (left) and an illustration of the segment moving heuristic on one maximal segment (right).

**Figure 3:** Horizontal and vertical layout (sea partitions) for Europe.
**Desired area of sea rectangles.** A second choice that must be made is the desired area of each sea rectangle. We cannot allow sea rectangles to become arbitrarily small, because then countries separated by a sea can become adjacent. More generally, preserving the size of the seas to some extent helps to preserve the global shape of the coastline between countries and seas. We choose to let sea rectangles have as desired areas the area that was given to them by the layout produced by the algorithm of Kant and He. We scale these values so that their sum corresponds to the percentage of the map area for the sea. We also scale the desired country rectangle areas so that they together claim the remaining area of the map.

**Error of sea rectangles.** Since sea rectangles have a desired area, they also have an error. This error should not be treated the same way as the errors of the countries. Errors of countries influence the correctness of the cartogram, whereas errors of sea rectangles only influence the aesthetic aspects. Therefore, we divide the error of a sea rectangle by some factor, the sea error division value. For example, if the sea error division value is 8, a sea rectangle and a country rectangle both need to grow, and they share a segment, then the error of the sea rectangle is more relevant for the move of the segment only if its error is more than 8 times as large as the error of the country. It is not clear what the optimal sea error division value should be. We will examine this question in this paper.

**Rectangle adjacency preservation.** Three adjacency options are possible. Consider the layout for Europe in Figure 2, in particular the maximal horizontal segment that lies above Italy (IT) and Slovenia (SI), and below Switzerland (CH) and Austria (AT). Moving this segment up several times can eventually lead to creating an adjacency between Slovenia (SI) and Slovakia (SK), and removing the adjacency between Austria (AT) and Hungary (HU) at the same time. Such a segment moving step is not allowed when correct adjacencies for countries are required, but it is allowed in the false adjacency option. Next, consider the horizontal sea partition for Europe, and the maximal horizontal segment that contains the top of Great Britain (GB). Possibly pushed upward by the growing of the rectangle of Great Britain, the right endpoint of this maximal segment may create different adjacencies of sea rectangles. This happens when it passes the horizontal segment containing the top of Estonia (EE). This is allowed if adjacencies of sea rectangles need not be preserved. The horizontal segment may move up even further, creating a new adjacency of a sea rectangle with Sweden (SE). This is allowed if sea adjacencies need not be preserved, because no country-country adjacencies are influenced. We call the three options the correct adjacency, false sea adjacency, and false adjacency options.

### 3. EXPERIMENTAL SETUP

The experiments we have performed address the following questions:

1. Does the partition of the sea regions into rectangles have influence on the error and visual quality of the rectangular cartogram that is produced?
2. Does the use of correct adjacencies, false sea adjacencies, and false adjacencies have influence on the error and visual quality of the rectangular cartogram that is produced?
3. Does the sea error division value have influence on the error and visual quality of the rectangular cartogram that is produced?

To answer these questions we ran the segment moving heuristic many times. The following settings were tested in all combinations, leading to 216 rectangular cartograms and average error values for the countries.

1. Different regions: Europe and the countries adjacent to the Mediterranean.
2. Different data sets: true area, population, and highway length in kilometers.
3. Different sea partitions (layouts): mostly horizontal, mostly vertical, and mixed horizontal and vertical.
4. Different adjacency options: correct, false sea adjacencies, and false adjacencies.
5. Different sea error division values: 4, 8, 12, and 16.

Our implementation of the segment moving heuristic allows the choice of several other parameters, including the number of iterations, the aspect ratio, and the percentage of the area for sea regions. We chose the number of iterations to be 500 in all cases. This is large enough to obtain a rectangular cartogram in which no improvements occur any more. The aspect ratio was set to 16 in all cases, which is a value that gives visually good output but is still large enough to result in cartograms with low errors. The area for the sea was chosen to be 20% for Europe and 30% for the Mediterranean region.

To be able to judge the results better, we applied a coloring of the country rectangles by their error. White rectangles have an area that is correct within a small margin. A rectangle that is too small is colored red, where the intensity of red
shows by how much it is too small. Similarly, blue rectangles are too large. The coloring helps to understand the operation of the algorithm and why it cannot reduce the remaining error any further.

4. EXPERIMENTAL RESULTS AND THEIR INTERPRETATION

In this section we answer the three experimental questions listed in the previous section. We study both the average error of the countries and the visual quality of the cartograms. During our experiments we grouped the results by layout and data sets. One of the 18 resulting sheets can be seen in Figure 7 at the end of this paper.

Partitioning the sea. We begin with the effects of different partitions of the sea, or initial layouts. Although there are differences in error between different layouts, the results are not conclusive. Differences are largest in the case of correct adjacencies. For the map of Europe with correct adjacencies, the mixed sea partition is always better than the horizontal partition, and usually better than the vertical partition, see the table below. For false sea adjacencies and false adjacencies, the mixed sea partition is slightly better on the average, but the differences are small, and all three sea partitions have the smallest average error in some test.

<table>
<thead>
<tr>
<th></th>
<th>Horizontal</th>
<th>Vertical</th>
<th>Mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>div 4</td>
<td>0.141</td>
<td>0.164</td>
<td>0.137</td>
</tr>
<tr>
<td>div 8</td>
<td>0.096</td>
<td>0.096</td>
<td>0.076</td>
</tr>
<tr>
<td>div 12</td>
<td>0.077</td>
<td>0.079</td>
<td>0.057</td>
</tr>
<tr>
<td>div 16</td>
<td>0.070</td>
<td>0.070</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>0.202</td>
<td>0.122</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td>0.214</td>
<td>0.109</td>
<td>0.141</td>
</tr>
<tr>
<td></td>
<td>0.240</td>
<td>0.104</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>0.238</td>
<td>0.101</td>
<td>0.131</td>
</tr>
<tr>
<td></td>
<td>0.175</td>
<td>0.128</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>0.154</td>
<td>0.098</td>
<td>0.141</td>
</tr>
<tr>
<td></td>
<td>0.151</td>
<td>0.093</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>0.142</td>
<td>0.097</td>
<td>0.097</td>
</tr>
</tbody>
</table>

Table 1: Average error values for Europe; correct adjacencies only, different data sets, division values, and layouts.

For the Mediterranean map and correct adjacencies, the horizontal sea partition is usually best. For the other adjacency options, horizontal, vertical, and mixed are all the best for some tests; there is no clear pattern. We can only observe that the average error differs considerably, depending on which sea partition is chosen initially.

Visually, there can be considerable differences in the cartogram produced from different initial sea partitions. In Figure 4 the highway length of the Mediterranean region with false sea adjacencies and sea division value 4 are shown, starting from the vertical and from the horizontal layout. The cartograms have average errors 0.130 and 0.111, respectively. The difference in shape of the African countries and of Cyprus is clear. However, in general it is not clear which initial layout gives the most aesthetic cartograms for the European and Mediterranean data sets and often the cartograms are similar.

Figure 4: Two cartograms depicting the highway length in km of the Mediterranean, based on a vertical layout (left) and a horizontal layout (right).

Adjacency options. The second experimental question concerns differences resulting from correct adjacencies, false sea adjacencies, and false adjacencies. In van Kreveld and Speckmann (2004), correct and false adjacencies for the USA and for Europe were already considered, and it was clear from the experiments that false adjacencies help considerably to bring the average error down. Here we confirm these observations and establish that allowing only false sea adjacencies already gives a reduction in the average error when compared to correct adjacencies.
In the European cartograms, false adjacencies always yield the lowest average error, usually a factor 2-3 better than the false sea adjacency option, and a factor 3-4 better than the correct adjacency option. In several cases the same cartograms and errors are obtained with correct adjacencies and false sea adjacencies, which were apparently not used. It may seem surprising that sometimes, correct adjacencies performs better than false sea adjacencies; this is probably due to the fact that the segment moving heuristic locally reduces the error of the rectangle with largest error in order to globally get a small average error. In some cases a reduction in the largest error may cause the error of several rectangles go up a little, leading to a larger average error. It may be that such a segment move was only allowed with the false sea adjacencies, but not with correct adjacencies.

<table>
<thead>
<tr>
<th>Area</th>
<th>Population</th>
<th>Highway km</th>
</tr>
</thead>
<tbody>
<tr>
<td>div 4</td>
<td>div 8</td>
<td>div 12</td>
</tr>
<tr>
<td>Correct</td>
<td>0.249</td>
<td>0.145</td>
</tr>
<tr>
<td>FalseSea</td>
<td>0.052</td>
<td>0.079</td>
</tr>
<tr>
<td>False</td>
<td>0.052</td>
<td>0.077</td>
</tr>
</tbody>
</table>

Table 2: Average errors for the Mediterranean region; mixed layout only, different data sets, division values, and adjacency options.

In the Mediterranean cartograms we observe roughly the same results. Table 2 shows the values for the mixed sea partition, where the differences between the three adjacency options are most clear.

The visual quality is affected by the use of correct adjacencies, false sea adjacencies, or false adjacencies, where cartograms produced with false adjacencies are sometimes clearly worse. In some cartograms, Denmark drifts left or right to become adjacent to the Netherlands or Poland, and Turkey sometimes grows upward quite far. The difference in visual quality between correct adjacencies and false sea adjacencies is not so clear. In Figure 5, the left cartogram is the result of correct adjacencies or false sea adjacencies, there is no difference. The right cartogram has the same settings, except for allowing false adjacencies, which can clearly be seen for Turkey, Denmark, and Portugal. The average error of this cartogram is only 0.027, whereas the correct adjacency cartogram has an average error of 0.132.

Figure 5: Two cartograms depicting the population of Europe with correct adjacencies (left) and false adjacencies (right).

In the Mediterranean cartograms, the false adjacency option can disturb the outer shape chosen initially: the coast line of the sequence of countries SI, HR, BA, CS, AL, and GR is not a staircase shape anymore descending to the Southeast.

Although the visual quality for false adjacencies is worse, in our tests the results were never dramatically wrong (such as the Iberian peninsula separating from France, which could happen theoretically).

**Division value.** Thirdly, we study the effects of the sea error division value. On the one hand, a high value (which means that the sea error is hardly taken into account when moving a segment) implies that a country with some error remaining can easily cause a segment to move at the expense of an adjacent sea, thus lowering the average error of the countries. On the other hand, a low value helps sea rectangles to keep their initial size to some extent, helping to maintain the shape of the coastline. Furthermore, a low value may have better properties of passing surplus of area via the sea to other parts of the cartogram, which again has a positive influence on the average error of the countries.
Usually the average error of the European data sets decreases with increasing division value. But there are cases where the highest average errors are obtained for the middle two division values 8 and 12, and for other cases the average error decreases until division value 12, and increases again for 16. For the Mediterranean data sets the situation is not conclusive either. The two tables presented earlier show several average errors for different division values. Overall, a division value of 4 gives the worst results. Visually, a division value of 16 may result in very thin sea rectangles; this sometimes happens with a division value of 12 as well.

Figure 6 shows four cartograms of Europe by true area and different division values. The average errors are 0.146, 0.074, 0.054, and 0.054. The cartogram for division value 16 has a very thin sea between Estonia and Finland. In the Mediterranean cartograms for highway length, the seas between Spain, France, and Italy on the one side, and Morocco, Libya, and Tunisia on the other side, also became very thin for division values 12 and 16. The problem of thin seas hardly occurs for division values 4 and 8. If a sea division value is 4 or 8, and a low error cartogram is computed, then no thin sea rectangles occur. This behavior is as expected.

Summary. The next three tables show average errors that are averaged over several cartograms. The first table shows the errors for different layouts, averaged over the four division values and three data sets. The top row is also averaged over the three adjacency options, so it is the average of 36 cartograms; the second row is the average over 12 cartograms. The second table shows the average errors for different adjacency options. Each number is the average over 36 cartograms. The third table shows the average errors for different division values. Each number is the average over 27 cartograms. It is clear that the adjacency option heavily influences the error in the expected way. The division value also influences the error, but less strongly. The layout also has significant differences in error, but it is not clear which layout is to be preferred. Note that we only tested three layouts, but many more are possible.

<table>
<thead>
<tr>
<th></th>
<th>Europe</th>
<th>Mediterranean</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>vertical</td>
<td>horizontal</td>
<td>mixed</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td>0.076</td>
<td>0.097</td>
<td>0.066</td>
</tr>
<tr>
<td><strong>False sea only</strong></td>
<td>0.086</td>
<td>0.094</td>
<td>0.085</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Correct</th>
<th>False Sea</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Europe all</strong></td>
<td>0.116</td>
<td>0.089</td>
<td>0.034</td>
</tr>
<tr>
<td><strong>Mediterranean all</strong></td>
<td>0.128</td>
<td>0.083</td>
<td>0.044</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>div 4</th>
<th>div 8</th>
<th>div 12</th>
<th>div 16</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Europe all</strong></td>
<td>0.091</td>
<td>0.080</td>
<td>0.076</td>
<td>0.072</td>
</tr>
<tr>
<td><strong>Mediterranean all</strong></td>
<td>0.106</td>
<td>0.086</td>
<td>0.075</td>
<td>0.072</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS

The handling of the sea in the segment moving heuristic to construct rectangular cartograms is a non-trivial matter. Many choices exist to try and obtain small error and aesthetically pleasing cartograms at the same time. In this paper we examined three of these choices in detail by analyzing the error and the cartograms produced in various settings and on several data sets. We observed that the initial layout, the adjacency options, and the sea error division values all influence the cartogram that is generated.

For the initial layout, we observed that it is important for the average error which one to choose. We compared three initial layouts, but it did not become clear how the choice of layout influences the average error, and which layout to use in general. More research in this direction is needed.
For the adjacency options, the false sea adjacency yields lower error cartograms than correct adjacencies in many cases, while the output is visually the same in quality. False adjacencies give lower errors, but the visual quality is often worse. Therefore, both false sea adjacencies and false adjacencies can be appropriate choices.

For the sea error division values, there is a trend to lower errors for higher sea division values. The trend becomes less clear when the value increases, and at the same time the visual quality goes down due to seas getting very thin. A choice of 12 for the sea division value seems the best.

Our tests were performed on two small to medium size rectangular cartograms. The most challenging open problem is to generate good rectangular cartograms of the World. The handling of the sea is essential in this task and we believe that the insight obtained by the experiments in this paper will help to address this challenge.

ACKNOWLEDGEMENTS

The authors thank Harald Scheper for generating part of the layouts and data sets, and Sander Florisson for several contributions to the implementation.

REFERENCES

Figure 7: All cartograms for the vertical partition of Europe and the highway km data set inclusive maximum and average cartographic error.
RECTANGULAR CARTOGRAM COMPUTATION
WITH SEA REGIONS

Marc van Kreveld
Inst. of Information and Computing Sciences
Utrecht University
marc@cs.uu.nl

Bettina Speckmann
Dep. of Mathematics & Computer Science
TU Eindhoven
speckman@win.tue.nl

Biography

Bettina Speckmann received her Ph.D. in computer science from the University of British Columbia, Vancouver, Canada, in 2001. Her thesis concerns kinetic data structures for collision detection. She was a postdoctoral researcher at the institute of theoretical computer science at ETH Zurich from 2001 until 2003. Since 2003 she is an assistant professor at the department of mathematics and computer science of the TU Eindhoven, the Netherlands. Her main research interests are computational geometry and algorithms for automated cartography.