

HistoScale: An Efficient Approach for Computing Pseudo-Cartograms

Daniel A. Keim, Christian Panse, Matthias Schäfer, Mike Sips
University of Konstanz, Germany

{keim,panse,schaefer,sips}@informatik.uni-konstanz.de

Stephen C. North

AT&T Shannon Laboratory, Florham Park, NJ, USA

north@research.att.com

1 Motivation

Nowadays, two types of maps, the so-called *Thematic Map* and *Choropleth Map*, are used in Cartography and GIS-Systems. *Thematic Maps* are used to emphasize the spatial distribution of one or more geographic attributes. Popular thematic maps are the Choropleth Maps (Greek: choro = area, pleth = value), in which enumeration or data collection units are shaded to represent different magnitudes of a variable. Besides, the statistical values are often encoded as colored regions on these maps. On both types of maps, high values are often concentrated in densely populated areas, and low statistical values are spread out over sparsely populated areas. These maps, therefore, tend to highlight patterns in large areas, which may, however, be of low importance. A cartogram can then be seen as a generalization of a familiar land-covering choropleth map. According to this interpretation, an arbitrary parameter vector gives the intended sizes of the cartogram's regions, that is, a familiar land-covering choropleth map is simply a cartogram whose regions sizes proportional to the land area. In addition to the classical applications mentioned above, a key motivation for cartograms as a general information visualization technique is to have a method for trading off shape and area adjustments. Pseudo-Cartograms provide an efficient and convenient approximation of cartograms, since a complete computation of cartograms is expensive. In this poster, we propose an efficient method called *HistoScale* to compute Pseudo-Cartograms.

2 HistoScale Approach

The basic idea of the *HistoScale* method is to distort the map regions along the two euclidian dimensions x and y . The distortion depends on two parameters, the number of data items which are geographically located in this map area, and the area covered by this map region on the underlying familiar land-covering map. The distortion operations can be efficiently performed by computing a histogram with a given number of bins in two euclidian dimensions x and y to determine the distribution of the geo-spatial data items in these dimensions. The two histograms are independent from each

other, that means, the computation of the histograms can be random. The two consecutive operations in the two euclidian dimensions x and y realize a grid placed on a familiar land-covering map. The number of histogram bins can be given by the user. For a practicable visualization we suggest 256 histogram bins for both histograms.

Each of the histogram bins covers an area on the underlying familiar land-covering map. To determine this area, the *HistoScale* method computes the upper and the lower point of intersection with the underlying map. The minimal bounding box containing the points of intersection and the preceding bin approximate the covered area for each histogram bin. The next step is to rescale the minimal bounding box of each histogram bin and, at the same time, the associated map regions in such a way, that our *HistoScale* method fulfills the cartogram condition. That is, the covered map area is equal to the number of geographically located data items in the map region. The area covered by the minimal bounding box is determined by the width (equal for all histogram bins) and the height $H_{M\!B\!B}$ of the minimal bounding box (this one being different for each histogram bin). Therefore, we compute new widths for each of the minimal bounding boxes while, at the same time, the heights of the minimal bounding boxes remain unmodified. The new lengths of the minimal bounding boxes can be determined using the following formula:

$$\Rightarrow \overrightarrow{|hb'_{i-1}hb'_i|} = \frac{\sum_{j=1}^{|DB|} |p = (x_j, y_j) \in hb_i|}{|\overrightarrow{up_i lp_i}|} \cdot \frac{\sum_{k=1}^{|hb|} (A_{M\!B\!B})^k}{\sum_{k=1}^{|hb|} (A'_{M\!B\!B})^k}$$

with $\forall i \in \{1, \dots, |hb|\} : (A_{M\!B\!B})^i = \overrightarrow{|hb_{i-1}hb_i|} \cdot |\overrightarrow{up_i lp_i}|$

$$\forall i \in \{1, \dots, |hb|\} : (A'_{M\!B\!B})^i = \sum_{j=1}^{|DB|} |p = (x_j, y_j) \in hb_i|$$

where $A'_{M\!B\!B}$ is the area of a minimal bounding box *MBB* of each histogram bin, $hb = \{hb_1, \dots, hb_m\}$ the end points of all histogram bins, and lp , up the lower and upper points of intersection. To compute the new boundaries in an efficient way, our *HistoScale* algorithm only needs to compute the new end points of each histogram bin. The original and new

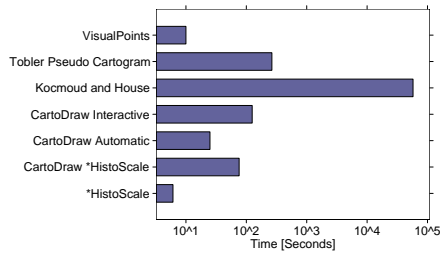


Figure 1: Time Comparison - we have assumed a 120MHz Intel CPU to compute the US-State Cartograms

end points of each histogram bin are stored in an array in ascending order.

After rescaling certain map regions our *HistoScale* algorithm computes the new coordinates of the map polygon mesh. The basic idea is to determine for each polygon node the original histogram bin in which this polygon node is geographically located. The search for this bin can be done in logarithmic time using binary search.

3 Application and Evaluation

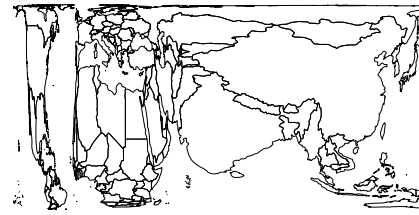
The resulting output maps are referred to as pseudo-cartograms, since they are only approximations to the true cartogram solution. On the other hand our approach generates interesting maps and good solutions in least square sense. The computation of pseudo-histograms using our *HistoScale* algorithm can be done in real-time (see figure 1). Due to the run time behavior, *HistoScale* can be used as a preprocessing step for other cartogram algorithms. Figure 1 shows, that the computation time of the *CartoDraw* algorithm can be reduced without losing any quality. Figure 2 shows several interesting applications using our *HistoScale* algorithm. The world population pseudo-cartogram shows clearly, that China and India are the most populated world regions. This fact has for example an important influence on the evolution of epidemics such as SARS, as unknown epidemics in such areas can be dangerous for the whole world population. The USA Pseudo-Cartogram clearly shows the two most populated areas, which are, New York City and Los Angles County.

References

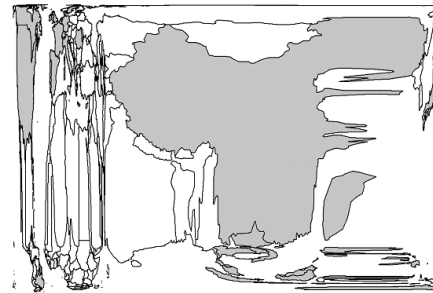
Christopher J. Kocmoud and Donald H. House. Continuous cartogram construction. In *IEEE Visualization, Research Triangle Park, NC*, pages 197–204, 1998.

Daniel A. Keim, Stephen C. North, Christian Panse, and Jörn Schneidewind. Visualizing geographic information: VisualPoints vs CartoDraw. *Palgrave Macmillan – Information Visualization*, 2(1):58–67, March 2003.

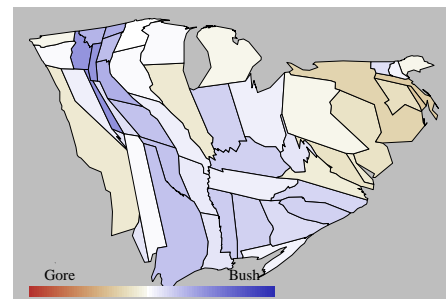
W.R. Tobler. Pseudo-cartograms. *The American Cartographer*, 13(1):43–40, 1986.



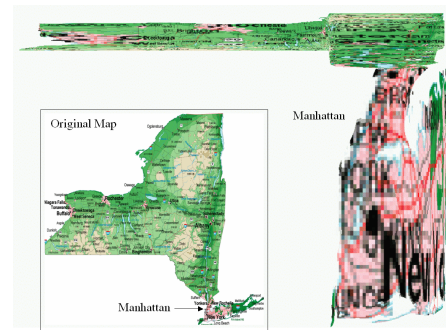
(a) World Population Pseudo-Cartogram



(b) World SARS Pseudo-Cartogram (gray indicates countries with SARS cases)



(c) The area of the states in the cartogram corresponds to population and the color corresponds to the percentage of the votes. A bipolar colormap is used to show which candidate won the state.



(d) NY-State Pseudo-Cartogram with texture mapping

Figure 2: Application examples