A Comparison Study on the mapping and cognition of Area Cartogram

Jinn-Guey Lay[1], Wei-Chia Huang[1], Tian-Yuan Shih[2]

[1]National Taiwan University (NTU),

No. 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan, jglay@ntu.edu.tw

[2]National Chiao Tung University (NCTU),

No. 1001, University Rd., Hsinchu, Taiwan, tyshih@mail.nctu.edu.tw

KEYWORDS: area cartogram, algorithms, map cognition

ABSTRACT: An area cartogram is a thematic map that uses area as a symbol to present certain attribute values. There are many different methods and algorithms to create area cartograms. We identify more than 30 algorithms that have been proposed by researchers since 1973. These algorithms can be categorized into four main types: non-contiguous, contiguous, Dorling and pseudo-cartogram. Within these four types, contiguous cartogram is the most common one, which aims to keep balance between preserving relative location and minimizing shape distortion. Given the variety of algorithms, not many have been included in GIS software packages or public domain software. This research collects several different algorithms and uses them to create a series of maps. We investigate the difference of these maps by comparing their geometric properties and topology. We also investigate the cognition aspect of this issue by testing the reactions of map readers. It is anticipated that such research can enhance the development of area cartogram and help promote its application.

1. BACKGROUND

An area cartogram is a thematic map that uses area as a symbol to present certain attribute values. Its idea matches the graduated symbols in which the size of value is proportional to its area on the map. Graduated symbols are just symbols on a map, while area cartograms distort themselves to reach the expected area.

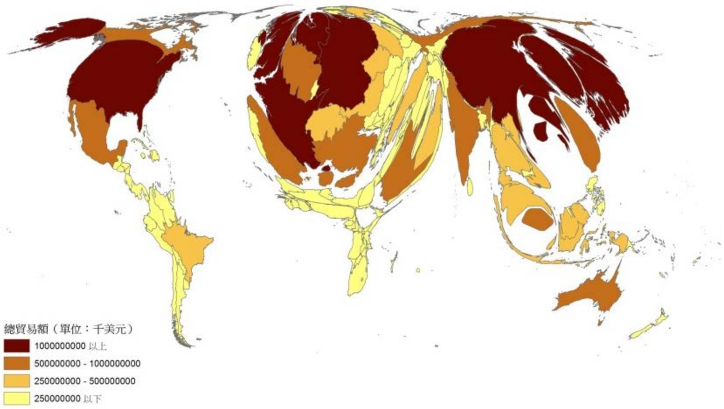

Figure 1: World GDP map shown using graduated symbols (top).

Figure 2: World GDP map shown using area cartogram (bottom).

With the development of computer cartography, computers are able to generate area cartograms using algorithms written by computer scientists. In 1973, Tobler proposed the “rubber map method” to generate area cartograms by computer. Today, more than thirty algorithms have been proposed by researchers across various disciplines, including geography, mathematics, physics and statistics.

The various styles of area cartograms can be categorized into four main types: non-contiguous, contiguous, Dorling and pseudo-cartogram. For the contiguous cartogram, the most common type, each unit must be distorted so that it shares edges with other units. Each algorithm for continuous area cartograms is designed to preserve relative location and minimize shape distortion.

Past studies have rarely explored the differences between algorithms. Today, GIS software offers a range of tools that allow users to freely generate area cartograms. However, users are unlikely to notice the various algorithms that might be applied across different software. For ordinary users, area cartograms just “appear” after a single mouse click. Therefore, it is necessary to clarify the differences between the algorithms and estimate the index in value.

This article begins with a review of the literature. We clarify how many algorithms had been proposed to date and explain how these algorithms are developed. In the following section, we pose four questions about using computers to generate area cartograms. Finally, we use counties in Taiwan to demonstrate the differences between these algorithms.

2. LITERATURE

2-1 Algorithms

With the rise in popularity of computer graphics, computer-based generation of cartography has become increasingly common. An algorithm is needed in order to generate cartograms on a computer. Tobler (1973) first proposed computer-assisted cartogram algorithms, which convert maps into dense grids, then resize the base according to the statistical value. This method is known as the "rubber map." Today, there are over twenty algorithms that have been proposed by researchers.

Year	Author	Algorithm	Type
1973	Tobler	Rubber map method	contiguous
1976	Olson	Projector method	noncontiguous
1978	Kadmon, Shlomi	Polyfocal projection	noncontiguous
1984	Selvin et al.	DEMP (Radial Expansion) method	contiguous
1985	Dougenik et al.	Rubber Sheet Distortion method	contiguous
1986	Tobler	Pseudo-Cartogram method	—
1987	Snyder	Magnifying glass azimuthal map projections	noncontiguous
1989	Cauvin et al.	Piezopleth maps	contiguous
1990	Torguson	Interactive polygon zipping method	contiguous
1990	Danny Dorling	Cellular Automata Machine method	contiguous
1993	Gusein-Zade, Tikunov	Line Integral method	contiguous
1996	Dorling	Circular cartogram	Dorling
1997	Sarkar, Brown	Graphical fisheye views	noncontiguous
1997	Edelsbrunner, Waupotitsch	Combinatorial-based approach	contiguous
1998	Kocmoud, House	Constraint-based approach	contiguous
2003	Keim, North, Panse	Cartodraw	contiguous
2003	Keim, North, Panse	HistoScale	contiguous
2004	Gastner, Newman	Diffusion-based method	contiguous

Year	Author	Algorithm	Type
2004	Sluga	Lastna tehnika za izdelavo anamorfoz	contiguous
2004	Helimann, Keim et al.	RecMap	contiguous
2005	Keim, North, Panse	Medial-axis-based cartograms	contiguous
2007	van Kreveld, Speckmann	Rectangular Cartogram	contiguous
2009	Heriques, Bação, Lobo	Carto-SOM	contiguous
2013	Shipeng Sun	Opti-DCN and Carto3F	contiguous
2014	B. S. Daya Sagar	Mathematical Morphology-Based	contiguous

Figure 3: List of algorithms for generating area cartograms.

In 1985, three students from Harvard University proposed a new algorithm based on "rubber sheet distortion." The improvements are notable and this algorithm is used widely to this day. In 2004, Gastner and Newman, the physicalists, proposed a new "diffusion-based" method algorithm since it borrows ideas from physics.

2-3 Comparison between the algorithms of area cartograms

There is currently little research that compares the algorithms of area cartograms. Previous papers have been written by the authors of an algorithm, with the intent of proving that his or her algorithm is more competitive. Sagar (2014), for example, compared Kocmoud (1997), Keim et al. (2004), Gastner & Newman (2004) and his own mathematical morphology algorithm.

3. RESEARCH QUESTIONS AND RESEARCH APPROACH

3-1 Research Questions

To compare the algorithms, first of all, we must know how many software and algorithms presently exist. Hence, we pose the first question:

(a) How many software and algorithms have been proposed so far, and how does a user execute each of these using a computer?

Furthermore, to compare the algorithms, we must clarify some control variables. Since every algorithm must be executed using software, we pose the second question:

(b) What are the differences of a specific algorithm between different software?

And last but not least, as quantitative researchers, we are interested in estimating the differences between the algorithms with respect to the algorithms' area and distance biases. We pose the third and fourth questions:

(c) What are the differences between the algorithms in view of area bias?

(d) What are the differences between the algorithms in view of distance bias?

3-2 Research Approach

To answer these four questions, we must collect all available cartogram-generating algorithms and explain their algorithms. We use a county map of Taiwan to generate area cartograms based on county population. Finally, we develop statistical indexes to compare the algorithms.

4. RESULTS

4-1 Available algorithms/software

Two of the most popular professional-level GIS software, ESRI ArcGIS and Quantum GIS, both lack embedded cartogram-generating tools. However, they all allow users to download related plug-ins developed by volunteers from an official source. For ESRI ArcGIS, the “Cartogram geoprocessing tool” plug-in, which uses Gastner & Newman (2004)’s algorithm, is available for download. Quantum GIS users can download a plug-in called “Cartogram,” which uses Dougenik et al. (1985)’s algorithm.

Some authors of algorithms have also released cartogram-generating software to the public. For instance, Gastner & Newman (2004) developed software called “Cart.” Moreover, a new algorithm recently proposed by Sun (2013) also has software, “Software for Unified Network Analysis” (SUNA), developed by himself.

Aside from the aforementioned professional-level GIS software, independent software is also available. For example, Scape Toad, which is specifically designed for generating cartograms, is an independent software developed by Choros Laboratory. It uses the Gastner & Newman (2004) algorithm. MAPresso is a free Java applet for generating choropleth map and area cartograms. It uses both the Dougenik et al. (1985) and Dorling (1996) algorithms.

Software	Embedded Cartogram	Algorithm
Cart	Yes	Gastner & Newman (2004)
Scape Toad	Yes	Gastner & Newman (2004)
MAPresso	Yes	Dorling (1996)
Protovis	Yes	Dorling (1996)
Software for Unified Network Analysis (SUNA)	Yes	Sun (2013)
ESRI ArcMap	Cartogram geoprocessing tool	Gastner & Newman (2004)
QGIS	Cartogram	Dougenik et al. (1985)
GeoDA	Yes	Unidentified

Figure 4: List of available algorithms/software.

4-2 The differences of a specific algorithm between different software

From the former section, we found that the Gastner & Newman (2004) algorithm was used in three software: Cart, Scape Toad and Cartogram geoprocessing tool, a plug-in of the ESRI ArcMap software. To clarify the differences of a specific algorithm between different software, we generate two area cartograms from Scape Toad and ESIR ArcMap. The original map is of counties in Taiwan and the value is population.

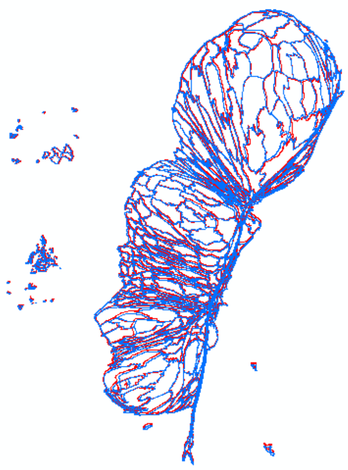

Figure 5: Overlay of two area cartograms from Scape Toad (red) and ESRI ArcMap (blue).

Figure 5 shows the overlay of two area cartograms from Scape Toad and ESRI ArcMap. The result is without coordinate adjustment, which means that we just added two shapefiles from two different software in a GIS to get Figure 5. From this image, the result from Scape Toad is in accordance with the result from ArcGIS. Since these two software generate substantially similar results, we believe that algorithms, rather than software, play the most important role in developing computer-based cartograms. In other words, software using the same algorithm will generate cartograms consistent with one another.

4-3 Comparison of the algorithms

To measure the differences between the area cartograms from variant algorithms, we use two indexes – area bias and distance bias.

4-3-1 Area bias

The index of area bias helps us know if each unit in an area cartogram really matches its proportional value. For example, if a county accounts for 1% of a nation's population, the expected area in the area cartogram should also be 1%. If the actual area does not match, a bias appears.

For every unit, we use the following index:

$$Area\ bias = The\ area\ proportion\ in\ area\ cartogram\ map\ - The\ propotion\ of\ a\ specific\ value$$

It can be formed as:

$$Area\ bias = \frac{Expected\ area - actual\ area}{Expected\ area + actual\ area}$$

Figure 6 shows our original map and an overlay of two area cartograms – Dougenik et al. (1985) and Gastner & Newman (2003).

Figure 6: The area cartograms of Dougenik et al. (1985) (blue), Gastner & Newman (2003) (red) and the original map (gray).

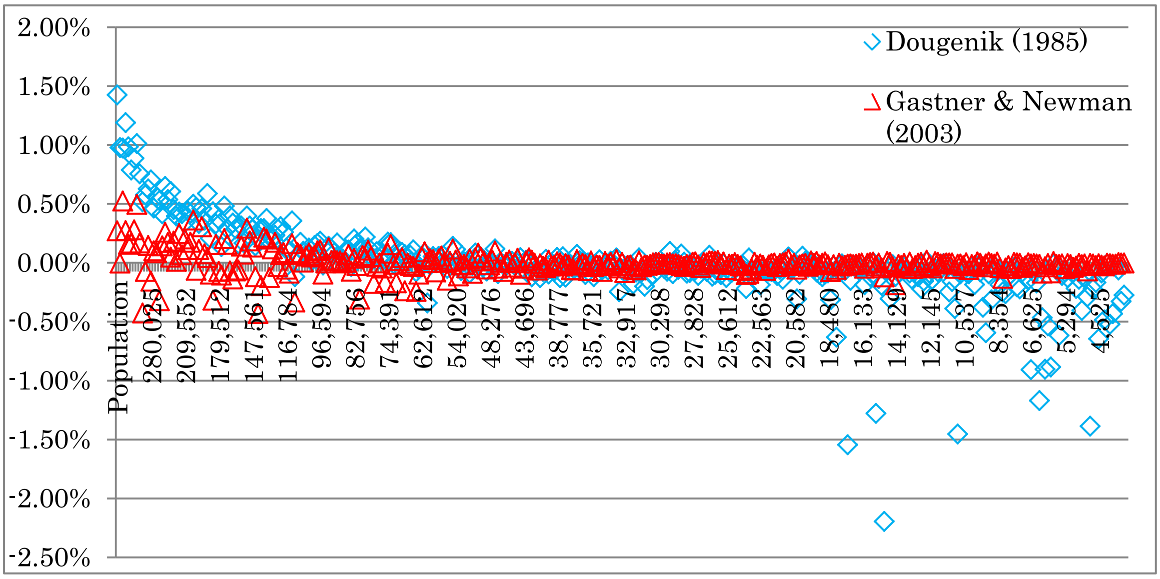

Figure 7: The area biases of Dougenik et al. (1985) and Gastner & Newman (2003).

Figure 7 shows the area bias of the two algorithms. The x-axis represents the population of each of the counties in Taiwan, sorted from highest to lowest. The y-axis represents the area bias of each county. Positive values in the y-axis indicate that the area on the cartograms has been underestimated, that is, it appears smaller than it should be. Negative values indicate that the area on the cartogram appears larger than it should be—the area on the cartogram has been overestimated.

Based on Figure 7, it is evident that Dougenik et al. (1985)'s algorithm has a much bigger area bias than Gastner & Newman (2004)'s – even exceeding ±2. On the other hand, the Gastner & Newman (2004) algorithm is relatively accurate with a bias of around ±0.6%.

The western half of Taiwan is a highly and densely populated area while the eastern half of Taiwan is more sparely populated with a smaller population. The eastern half of Taiwan distorts to a line almost in the area cartogram generated by Gastner & Newman (2004) algorithm to demonstrate the expected area. In the area cartogram by Dougenik et al. (1985) algorithm, counties in the eastern half of Taiwan retain their original shapes, but counties in the more populated areas in the western half of Taiwan have been underestimated.

4-3-2 Distance bias

The index of distance bias helps us to know how units "move" for area cartograms. In this case, the measurement of the centroid offset has been chosen as the index.

The original map uses two-degree zone transverse Mercator projection (TWD97), hence the original distance and the generating area cartograms all use meter as their distance unit. Figures 8 and 9 shows centroid offset of the two algorithms and Figure 10 shows the statistical values.

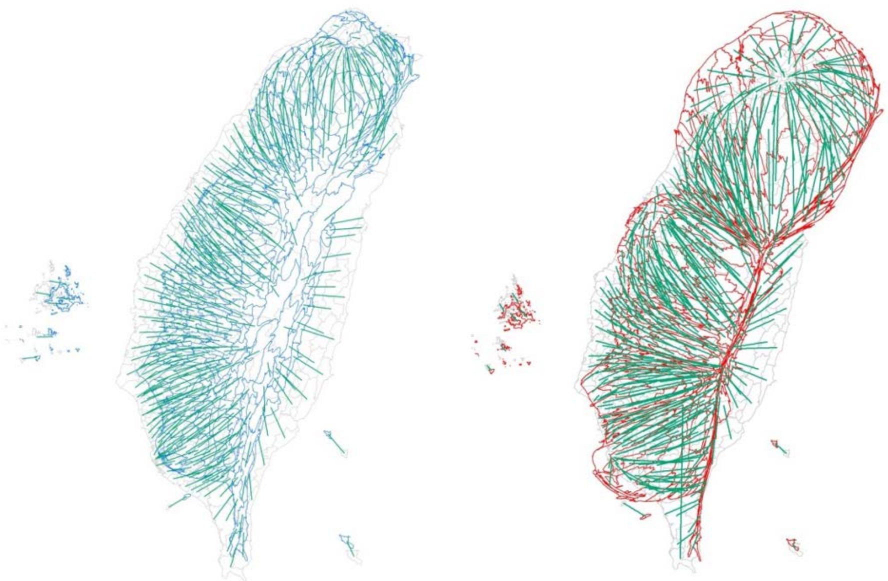

Figure 8: The centroid offset of Dougenik et al. (1985) (left).

Figure 9: The centroid offset of Gastner & Newman (2004) (right).

Algorithm	Dougenik et al. (1985)	Gastner & Newman (2004)
Min	1,126	461
Max	40,601	108,858
Sun	7,582,149	12,090,371
Average	21,180	33,772

Standard Deviation	6,733	15,733

Figure 10: Statistical values.

Dougenik et al. (1985)'s algorithm has a much smaller distance bias in terms of centroid offset. On the other hand, the Gastner & Newman (2004) algorithm has much larger variations, which may be attributable to the algorithm's "diffusion-based method."

5. DISCUSSION

In this paper, we identified four algorithms that are open to the public. Some authors of the algorithms developed their own cartogram software. Many of the software share the same algorithms but the differences in software do not necessarily result in significant outcomes.

For the two algorithms tested in this paper, we found that Gastner & Newman (2004)'s algorithm is relatively accurate because the area bias of it is less then 0.2%. However, in choosing between minimizing shape distortion and minimizing area bias, it seems that it leans in favor of the latter. Given a statistical data with higher dispersion, some units with low values may "disappear" through this algorithm.

The result of the Dougenik et al. (1985) algorithm is opposite: it tends to minimize shape distortion. Every unit can retain its original shape as much as possible, even in extreme cases. However, we also noticed that this algorithm can take up to five minutes to run using Quantum GIS.

We observe that the differences between the algorithms are noteworthy. Generating cartograms should not simply be treated as a one-click operation. We propose that various types of data each have their most appropriate algorithm. Hence, a professional GIS should support more than one algorithm.

6. Future development

In this paper, we clarify the differences between two major algorithms. In the future, more algorithms should be considered because they are based on different theories and may have variant results. However, because only few cartogram-generating software are released to the public, it is a challenge to contact those authors to in order to retrieve and subsequently compare the algorithms.

On the other hand, we will investigate the cognition aspect of this issue by testing the reactions of map-readers. Past papers have emphasized the important of accuracy, but the cognition aspect for map-readers is missing. In this paper, for example, we demonstrate that the eastern half of Taiwan distorts to a line almost in the area cartogram generated by Gastner & Newman (2004). Although this cartogram is accurate, it may not be easy for the average reader to comprehend. We hope that studying the cognition aspect will allow researchers to enhance the development of area cartograms and help promote their application in the future.

REFERENCES

1. Tobler W., 1963. Geographic area and map projections. Geogrev, 53 pp. 59-78.
2. Dorling D.,1996. Area cartograms: their use and creation. 1st ed. Bristol: Department of Geography, University of Bristol.
3. Gastner MT and Newman M., 2004. Diffusion-based method for producing density-equalizing maps. National Academy of Sciences 101(20), pp. 7499-7504.
4. Sagar B. D., 2013. Cartograms via mathematical morphology. Information Visualization 9, pp. 42-58.