ARTICLE

An Automated Displaced Proportional Circle Map Using Delaunay Triangulation and an Algorithm for Node Overlap Removal

David Lamb
School of Geosciences / University of South Florida / Tampa / FL / USA

ABSTRACT

Proportional circle maps are a popular method for visualizing quantitative data on a map. Circles are scaled proportionally based on the data provided; larger circles represent larger quantities. The circles require two values, a location and a numeric quantity. For data that have a wide range of values, the resulting map will produce clutter and overlap in which large symbols obscure the information contained in smaller circles. Some previous solutions to this problem are modifying and improving the contrast between symbols, or using stacking algorithms so all symbols are visible. This article proposes the displaced proportional symbol map, which displaces the symbol's location based on the amount of overlap between neighbouring circles. At the same time, it preserves the location of non-overlapping symbols, which distinguishes it from the circular cartogram. The displacement is automated through the proximity stress model algorithm for node overlap removal. This algorithm was originally designed for graph layouts with overlapping nodes, but was modified here for circular symbols and map layout. The result is a map with improved clarity and the ability to add labelling to a cluttered proportional circle map.

Keywords: proportional circles, cartogram, graph layout, automated, displacement

RÉSUMÉ

La carte de cercles proportionnels est une méthode populaire de visualisation de données quantitatives. La dimension des cercles est déterminée proportionnellement aux données fournies, les cercles plus grands représentant les quantités plus importantes. Les cercles exigent deux valeurs, un emplacement et une quantité numérique. Dans le cas de données pouvant afficher un large éventail de valeurs, la carte obtenue présentera des amas et des chevauchements dans lesquels les symboles de grande taille dissimuleront les données contenues dans les cercles de taille plus petite. Les solutions antérieurement proposées à ce problème consistent à modifier et à améliorer le contraste entre les symboles ou à utiliser des algorithmes de superposition permettant de faire en sorte que tous les symboles soient visibles. L'auteur propose la carte de symboles proportionnels décalés, dans laquelle l'emplacement du symbole est décalé en fonction du nombre de superpositions des cercles voisins. Parallèlement, cette carte préserve l'emplacement des symboles qui ne se chevauchent pas, ce qui la distingue du cartogramme circulaire. Le décalage est automatisé au moyen de l'algorithme du modèle PRISM (*proximity stress model*), visant à supprimer les chevauchements en préservant la disposition initiale, en vue de supprimer les superpositions de nœuds. Cet algorithme, conçu à l'origine pour les représentations graphiques présentant des chevauchements de nœuds, est adapté par l'auteur aux symboles et à la représentation cartographique circulaires. La carte ainsi obtenue est d'une plus grande clarté et permet l'ajout d'étiquettes à une carte de cercles proportionnels en superposition.

Mots clés : automatisé, cartogramme, cercles proportionnels, décalage, tracé de graphique

Introduction

Many geovisualization methods are available to display quantitative (and qualitative) values in a succinct manner on a map. One popular method is to scale symbols to reflect the quantitative value being shown, larger symbols representing higher values and smaller symbols representing lower values. This type of representation is broadly referred to as a graduated symbol map, of which the proportional symbol map is a special variety where the size of the symbol is scaled "in proportion to data values" (Brewer 2006, 29; Jenny, Hutzler, and Hurni 2009). This style of map is made up of two components: the location represented as a point and the symbol. For area data, a representative point is provided and assigned the data value of the entire area. The representative point is generated through a centrality-based simplification such as retaining the centroid of the area (Bereuter and Weibel 2013). Common symbols selected for proportional symbol maps are circles and squares, but any symbol may be used (Brewer and Campbell 1998).

More formally, a proportional circle map is defined as an abstract representation of quantitative data where the circle's area is graduated in direct proportion to the quantities represented (Flannery 1971). Studies of the proportional circle map focus on the relationship between the size of the circle and the interpreted value it represents (Flannery 1971; Meihoefer 1969; Gilmartin 1981a, 1981b). Disadvantages of proportional circle maps often derive from the problem of mapping numerical data with a wide range of values, which may cause overlapping symbols or graphical clutter in the map. Clutter is the excessive display of visual information (Touya, Hoarau, and Christophe 2016), such as when small proportional circles are obscured by larger circles. This causes overlap between circles and makes interpretation difficult, particularly estimating the size of the circle (Groop and Cole 1978).

Up until now, proposed solutions have revised symbols to improve clarity, or modified the stacking of symbols algorithmically. Graphic design solutions enhance the contrast between symbols by using a "hollow" symbol (outline, no fill), cut-outs, or transparency to show symbols beneath (Meihoefer 1969; Groop and Cole 1978; Dent 1993). Others approached the problem algorithmically through optimal stacking of the circles so the smaller circles sat on top, producing a clearer layout (Cabello and others 2010; Kunigami and others 2011).

Figure 1a presents a standard proportional symbol map displaying the total number of individuals aged 15–49 infected with human immunodeficiency virus (HIV). Only countries with data have symbols associated with them, placed at the centroids of the countries. The symbols are stacked so that the circles with smaller area sit on top of larger symbols. In addition, transparency is applied to improve the clarity of the information over the clutter (particularly in Africa).

Even with the improvement in transparency, the spatial relationship between circles remains unclear. Which countries share which borders is lost in the clutter, and it is difficult to associate symbols with their original countries. As Slocum and Gilmartin (1979) suggest, "One of the primary purposes of a thematic map is to illustrate general spatial relationships between individual units of the phenomenon being mapped" (133). An alternative may be to use a circular cartogram to represent the information, as shown in Figure 1b (the algorithm of Dorling 1991, 1996 scales the symbols differently). Cartograms in general are difficult for map readers to interpret (Sui and Goodchild 2011). Sun and Li (2010) found that, for the participants of their study, the Dorling cartogram was the least effective in representing quantitative and qualitative data. In addition, Inoue (2011) notes that "the relative positions of circles on cartograms sometimes differ greatly from the geographical maps; the displacement of circles then causes difficulty in distinguishing which circles represent which regions" (147). For example, in Figure 1b, Kenya (three-digit code KEN) becomes part of East Asia.

Another approach to reducing the clutter of proportional symbols is to displace the symbols' spatial locations to reduce or completely remove any spatial conflict between symbols. While overlapping symbols are displaced, non-overlapping symbols retain their places relative to the original spatial location. In essence, this might be called a pseudo-circular cartogram, or a displaced proportional circle map. This short article proposes to use a modified graph layout algorithm called the proximity stress model (PRISM) to displace circular symbols, reducing clutter and improving the readability of the map.

The PRISM Algorithm

Gansner and Hu (2009) proposed the PRISM algorithm for efficient node overlap removal in drawing graph layouts. The goal of the PRISM algorithm is "to remove overlaps while preserving the shape of the initial layout by maintaining proximity relations among the nodes" (Gansner and Hu 2009, 209). Although originally developed for rectangular-shaped nodes, it is adjusted here to work with circle symbols.

The PRISM algorithm is reliant on a proximity graph generated from a Delaunay triangulation (DT). A DT creates non-overlapping triangles connecting sets of three points for all points in a data set (such as country centroids) (Bourke 1989). The DT creates a rigid structure that maintains relationships between nodes and is converted to a temporary proximity graph, where the centre of the circle is a node and the triangles are edges (see Figure 2 for an example DT). This is ideal, since it creates a structure that maintains the original spatial relationship between points in physical space. Finally, the algorithm minimizes a stress function defined by the number of overlapping circles and the ideal distance between nodes along the edge (Gansner and Hu 2009).

The radius, r_i, of the circle symbol is defined proportional to the data (D_i) based on a minimum symbol size ($r_{\min}$), as in the equation

$$r_i = \sqrt{\frac{D_i}{D_{\min}}} r_{\min}. \tag{1}$$

Scaling the circles to the minimum symbol sizes ensures that the smallest values will still be visible. If the symbols

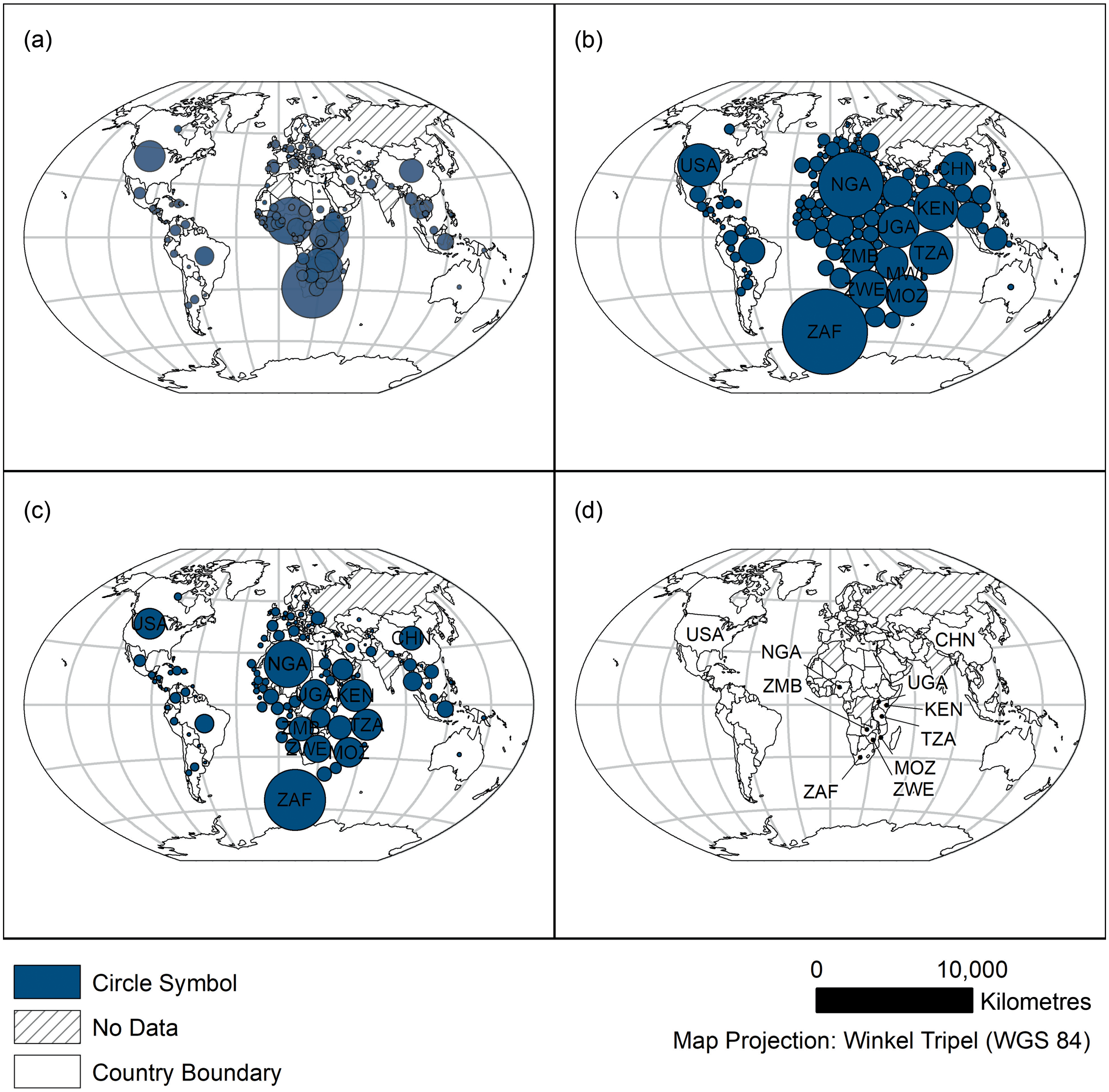

Figure 1. Number of individuals 15–49 infected with HIV displayed as (a) a proportional circle map, (b) circular cartogram, and (c) displaced proportional circle map
Note: Panel (d) is for label reference.
Data Sources: UNAIDS, AIDSInfo, and Natural Earth.

were scaled to a fixed maximum symbol size, this might affect the visibility of circles representing the smallest values. The overlap factor, t_{ij}, between two nodes with radii r_i and r_j is defined as the sum of the radii divided by the distance between those nodes' physical locations, $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$:

$$t_{ij} = \max\left(\frac{r_i + r_j}{\|\boldsymbol{v}_i - \boldsymbol{v}_j\|}, 1\right). \tag{2}$$

If no overlap exists, then t_{ij} is equal to one. The ideal length between two nodes becomes $t_{ij}\|\boldsymbol{v}_i - \boldsymbol{v}_j\|$. For two non-overlapping nodes, t_{ij} is one, and the ideal distance does not change their current positions. If the nodes overlap, the ideal distance is the current distance times the overlap factor. A larger overlap will increase the distance between the nodes. To dampen the effect of the overlap factor so that the distance is not increased by too high a

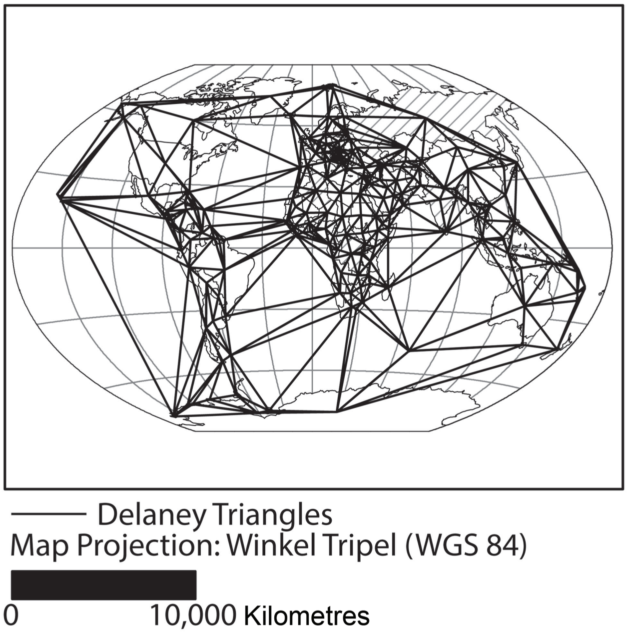

Figure 2. Example of a Delaunay triangulation for the centroids of countries
Data Source: Natural Earth.

value, a new parameter is created by comparing t_{ij} with 1.5, as in the equation

$$s_{ij} = \min(t_{ij}, 1.5). \tag{3}$$

This is Gansner and Hu's (2010) recommended value.

The goal of the algorithm is to minimize the proximity stress model iteratively as in the equations

$$\text{psm} = \sum w_{ij}\left(\|\boldsymbol{v}_i - \boldsymbol{v}_j\| - d_{ij}\right)^2, \tag{4}$$

$$d_{ij} = s_{ij}\|\boldsymbol{v}_i - \boldsymbol{v}_j\|, \tag{5}$$

$$w_{ij} = \frac{1}{d_{ij}^2}. \tag{6}$$

Through each iteration, a new DT is generated after the nodes are displaced. The ideal distance, d_{ij}, is the dampening parameter, s_{ij}, multiplied by the physical distance between the nodes (v_i and v_j) in Equation (5). This ideal distance is used in the minimization equation (Equation [4]) and is used to determine how much each overlapping point should be displaced, as in the equation

$$\boldsymbol{v}_n = \boldsymbol{v}_i + \frac{\boldsymbol{v}_i = \boldsymbol{v}_j}{\|\boldsymbol{v}_i - \boldsymbol{v}_j\| d_{ij}}. \tag{7}$$

The algorithm iterates over each node and each neighbour of that node according to the DT graph. The overlap is calculated and each node's displacement is adjusted based on the overlap (Equation [7]). The overall overlap is added to the total *psm*, and then each node is moved by its total displacement. When the *psm* is below a threshold value, ε, such as zero, the initial loop ceases. Gansner and Hu (2010) also propose a second loop following the same steps to minimize *psm* and any remaining overlap.

Conclusion

Figures 1, 3, and 4 provide several examples of the displaced proportional symbol map as compared with the proportional circle map. Figure 1c presents the HIV data discussed above as a displaced proportional circle map. There is an approximate 12.5 million square kilometres of overlap in Figure 1a, with a large portion of this in sub-Saharan Africa. This overlap is removed in both the cartogram (Figure 1b) and the displaced proportional circle map (Figure 1c). However, the cartogram displaces the circles' centres from the country centroids by over 100,000 km, and the displaced proportional circle map by approximately 63,000 km (see Figure 1d for the country locations). This is because the PRISM algorithm only removes overlap between the circles, and preserves the original circle points when possible. Like the cartogram, the displaced proportional circle map maintains spatial relationships; for example, Zimbabwe (ZWE) is to the west of Mozambique (MOZ), and Tanzania (TZA) is to the south of Kenya (KEN). Unlike the cartogram, the displaced proportional symbol map does not obscure the Indian subcontinent, or create a false neighbour relationship between China (CHN) and Kenya by forcing them to be adjacent.

The overlap in the proportional circle map makes identifying the associated countries beneath the circles difficult. However, it would still be difficult in either the cartogram or the displaced proportional circle map to identify the countries without some of the labels and callouts. This cannot be done easily with Figure 1a because the overlap makes it difficult to call out specific circles. Since the displaced proportional circle map preserves the centre points of many of the circles, and their locations, fewer labels would be needed than in the cartogram.

Figure 3 presents two more examples of the displaced proportional circle map compared with the traditional proportional circle map. Figures 3a and 3b show the population 50 years of age or older for the United States. In Figure 3a, a lot of overlap exists in the northeast that makes it difficult to associate the smaller circles with their respective states. In Figure 3b, the states' circles are separated so the smaller symbols are easier to read. This method creates a more readable map and allows the map reader to interpret the distribution of 50-year-olds and over in the United States.

In Figures 3c and 3d, the total population of each county in the US state of Florida is shown as a circle symbol. Here the overlap is primarily in the central and southern parts of Florida. In Figure 3c, the overlap obscures and

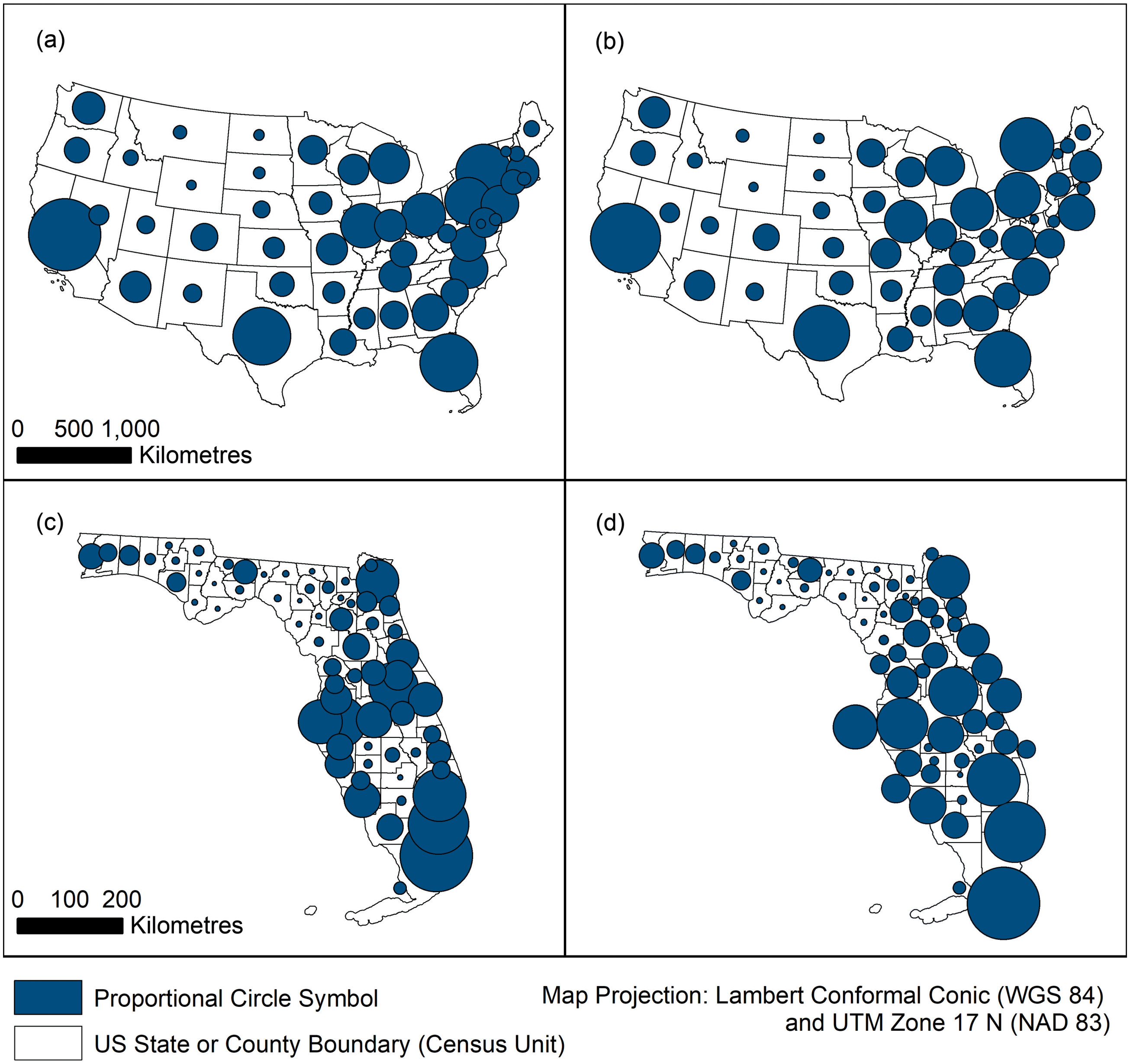

Figure 3. Total population 50 years old and over for the United States mapped as (a) a proportional circle map and (b) a displaced proportional circle map; and the total county population for the state of Florida in the United States using (c) a proportional circle map and (d) a displaced proportional circle map
Data Source: US Census Bureau.

hides some of the counties, which is partly because of the stacking order. In Figure 3d, the displacement separates the circles to make these areas clearer, while preserving their spatial relationships. Figure 3d also shows a case where the threshold factor, ε, was modified to allow overlap between some of the circles. This threshold prevented the algorithm from continuing until all overlap was removed.

This last example shows how the algorithm is easily adapted to fit different scenarios and scales. This is done through two parameters: minimum symbol size and ε. Setting ε to values greater than zero will allow different levels of overlap, depending on the effect desired. The cartographer may only wish to reduce some of the clutter without forcing circles too far from their original locations. In addition, the minimum symbol size changes depending on the scale of the map and study area. Changing these two parameters can create many variants on the same map. This flexibility allows both large and small area maps to be created, depending on the cartographer's needs.

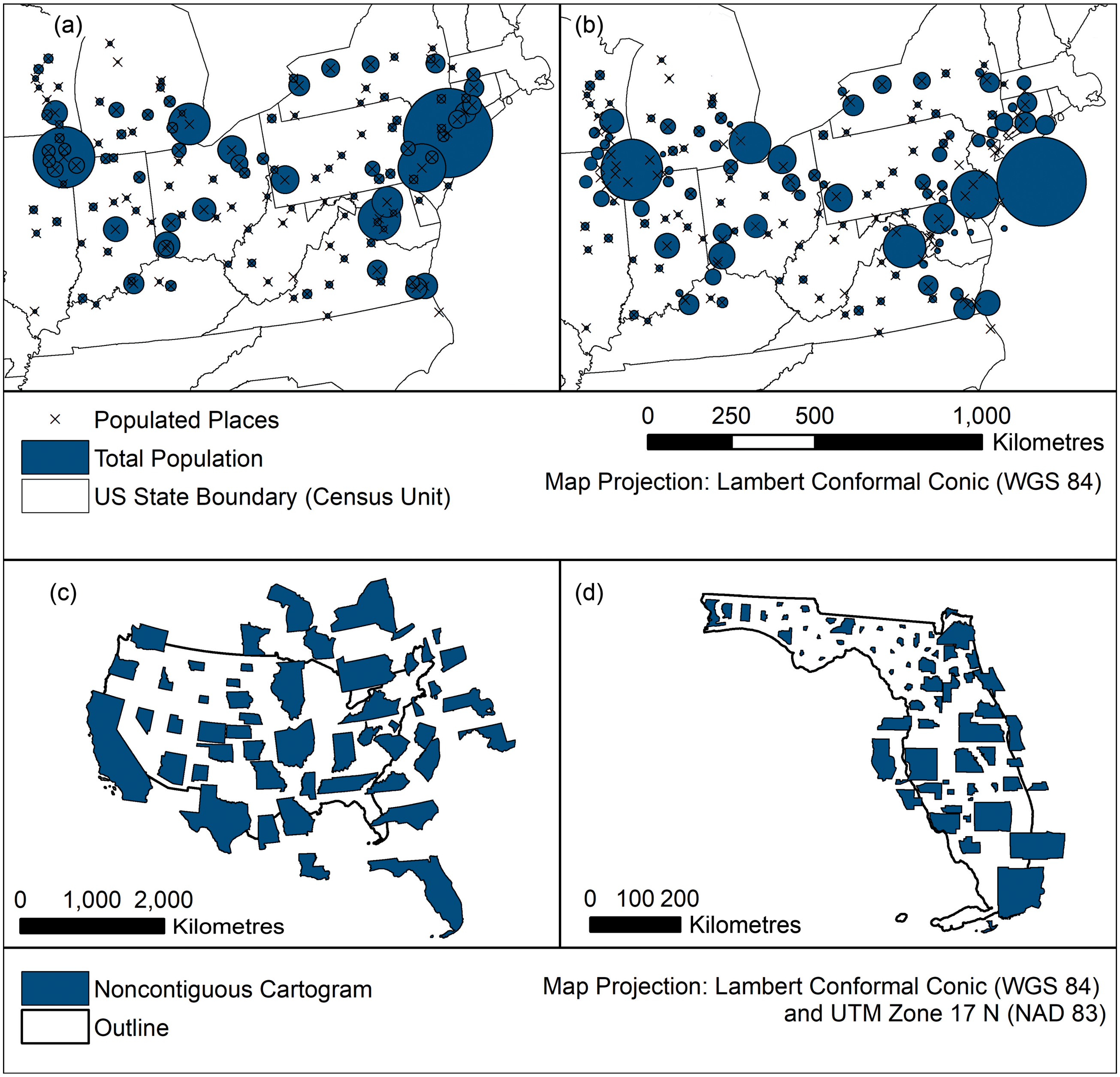

Figure 4. Total population of various populated places in part of the United States mapped as (a) a proportional circle map and (b) a displaced proportional circle map; (c) total population 50 years old and over for the United States using the PRISM algorithm and states scaled to their values; and (d) total county population for the state of Florida in the United States using the PRISM algorithm and counties scaled to their values
Data Sources: (a) and (b) US Census Bureau; (c) and (d) US Census Bureau and Natural Earth.

Figures 4a and 4b show the use of proportional circle symbols for point data. These points represent populated places in part of the United States. The circles represent the total populations of those places. Here, the displaced proportional symbol shifts the centre of the circle away from the original location of the populated place. This is less effective than when the circle represents an area. An enumeration area allows the centre of the circle to be placed easily anywhere inside its boundaries (Tyner 2010). However, the circles in Figure 4b may sit far from their original locations, and may be misleading to the map reader. The overlap in Figure 4a provides valuable information about the proximity of cities and their population sizes. In this case, it would seem that the traditional proportional symbol map is more appropriate.

Last, Figures 4c and 4d show the application of the algorithm to other types of symbols: in this case, the shapes of the states in the United States and the shapes of counties in the state of Florida. This creates a non-contiguous cartogram layout by removing overlap between scaled enumeration areas. The overlap factor, t_{ij}, was modified from measuring the overlap between circles to measuring the overlap area between the irregular shapes. In addition to other symbols, the approach would be applicable to Web-based mapping of symbols that need to be placed in an automated fashion.

General guidelines for proportional circle maps would be applicable to the displaced proportional circle map. For example, proportional circle maps typically perform poorly when the circles have a uniform size, making it difficult for a map reader to interpret their values. The minimum symbol size varies by region, scale, and the size of the area units. Tyner (2010) suggests that circles should not exceed the sizes of their enumeration areas, but a displaced proportional circle map will reduce the overlap when the circle crosses these area boundaries. The flexibility of the PRISM algorithm may extend the use of different types of symbols by only modifying how the overlap factor is calculated. This can increase the number of options available to the cartographer while properly mapping the data.

Author Information

David Lamb is an adjunct instructor in the School of Geosciences at the University of South Florida, Tampa, Florida. He holds a Ph.D. in geography and environmental science and policy and a B.A. in geography from the University of Northern Colorado. His research interests are in applying GIScience and quantitative methods to health and medical geography problems. E-mail: davidlamb@usf.ed.

References

Bereuter, P., and R. Weibel. 2013. "Real-Time Generalization of Point Data in Mobile and Web Mapping Using Quadtrees." *Cartography and Geographic Information Science* 40(4): 271–81. https://doi.org/10.1080/15230406.2013.779779.

Bourke, Paul. 1989. "Efficient Triangulation Algorithm Suitable for Terrain Modelling." Read at the Pan Pacific Computer Conference, Beijing, China, January. http://paulbourke.net/papers/triangulate/.

Brewer, C., and A.J. Campbell. 1998. "Beyond Graduated Circles: Varied Point Symbols for Representing Quantitative Data on Maps." *Cartographic Perspectives* 29(29): 6–25. https://doi.org/10.14714/CP29.672.

Brewer, C.A. 2006. "Basic Mapping Principles for Visualizing Cancer Data Using Geographic Information Systems (GIS)." *American Journal of Preventive Medicine* 30(2): S25–S36. https://doi.org/10.1016/j.amepre.2005.09.007.

Cabello, S., H. Haverkort, M. van Kreveld, and B. Speckmann. 2010. "Algorithmic Aspects of Proportional Symbol Maps." *Algorithmica* 58(3): 543–565. https://doi.org/10.1007/s00453-009-9281-8.

Dent, B.D. 1993. *Cartography: Thematic Map Design*, 3rd ed. Dubuque, IA: Wm. C. Brown.

Dorling, D. 1996. *Area Cartograms: Their Use and Creation. Concepts and Techniques in Modern Geography (CATMOG) 59*. Modern Geography Series. London: Institute of British Geographers.

Dorling, D.F.L. 1991. "The Visualization of Spatial Social Structure." PhD diss., University of Newcastle upon Tyne.

Flannery, J.J. 1971. "The Relative Effectiveness of Some Common Graduated Point Symbols in the Presentation of Quantitative Data." *Canadian Cartographer* 8(2): 96–109. https://doi.org/10.3138/J647-1776-745H-3667.

Gansner, E., and Y. Hu. 2010. "Efficient, Proximity-Preserving Node Overlap Removal." *Journal of Graph Algorithms and Applications* 14(1): 53–74. https://doi.org/10.7155/jgaa.00198.

Gansner, Emden R., and Yifan Hu. 2009. "Efficient Node Overlap Removal Using a Proximity Stress Model." In *Graph Drawing: Lecture Notes in Computer Science* 5417. ed. I.G. Tollis and M. Patrignani, 206–17. Berlin: Springer. https://doi.org/10.1007/978-3-642-00219-9_20.

Gilmartin, P.P. 1981a. "Influences of Map Context on Circle Perception." *Annals of the Association of American Geographers* 71(2): 253–58. https://doi.org/10.1111/j.1467-8306.1981.tb01351.x.

Gilmartin, P.P. 1981b. "The Interface of Cognitive and Psychophysical Research in Cartography." *Cartographica* 18 (3): 9–20. https://doi.org/10.3138/BN50-8GU9-226R-62K1.

Groop, R.E., and D. Cole. 1978. "Overlapping Graduated Circles/Magnitude Estimation and Method of Portrayal." *Canadian Cartographer* 15(2): 114–22. https://doi.org/10.3138/Q5Q5-N244-8462-NG25.

Inoue, R. 2011. "A New Construction Method for Circle Cartograms." *Cartography and Geographic Information Science* 38(2): 146–52. https://doi.org/10.1559/15230406382146.

Jenny, B., E. Hutzler, and L. Hurni. 2009. "Self-Adjusting Legends for Proportional Symbol Maps." *Cartographica* 44(4): 301–04. https://doi.org/10.3138/carto.44.4.301.

Kunigami, G., P.J. de Rezende, C.C. de Souza, and T. Yunes. 2011. "Optimizing the Layout of Proportional Symbol Maps." In *Computational Science and Its Applications - ICCSA 2011*, ed. B. Murgante, O. Gervasi, A. Iglesias, D. Taniar, and B.O. Apduhan, 1–16. Berlin: Springer. https://doi.org/10.1007/978-3-642-21931-3_1.

Meihoefer, H.J. 1969. "The Utility of the Circle as an Effective Cartographic Symbol." *Canadian Cartographer* 6(2): 105–17. https://doi.org/10.3138/J04Q-1K34-26X1-7244.

Slocum, T.A., and P.P. Gilmartin. 1979. "Cartographic Analysis of Proportional Circle Maps Using Delaunay Triangles." *Canadian Cartographer* 16(2): 133–44. https://doi.org/10.3138/E183-2337-076V-3614.

Sui, D., and M. Goodchild. 2011. "The Convergence of GIS and Social Media: Challenges for GIScience." *International Journal of Geographical Information Science* 25(11): 1737–48. https://doi.org/10.1080/13658816.2011.604636.

Sun, H., and Z. Li. 2010. "Effectiveness of Cartogram for the Representation of Spatial Data." *The Cartographic Journal* 47(1): 12–21. https://doi.org/10.1179/000870409X12525737905169.

Touya, G., C. Hoarau, and S. Christophe. 2016. "Clutter and Map Legibility in Automated Cartography: A Research Agenda." *Cartographica* 51(4): 198–207. https://doi.org/10.3138/cart.51.4.3132.

Tyner, J.A. 2010. *Principles of Map Design*. New York: Guilford.