A Contrarian View on New Approaches in Cartography: with Attention to the Choroplethic Conundrum

Aileen R. Buckley

Esri, 380 New York Street, Redlands, CA, USA
abuckley@esri.com

Abstract

The allure of developing new mapping methods and taking new approaches to cartographic design (that is, developing new visualization methods in geovisual analytics) has led cartographers and information designers to explore some interesting though sometimes ultimately fruitless pathways. Meanwhile, commonly-used mapping methods are disregarded and even sometimes dismissed outright. It may be boring to continue working on already established mapping methods, and some might consider that these methods have already been so well researched that there is little left to explore about them. Even so, cartographers and others know that some of the most commonly used mapping methods have some very fundamental problems that have yet to be addressed. Perhaps it is our failure to adequately refine existing mapping methods that causes us to turn in other directions to find solutions to our mapping challenges, thereby effectively avoiding the long standing and well-known problems that some of the most commonly used mapping methods present.

I suggest that we cartographers have not yet dealt adequately or at all with some of the most basic challenges of traditional cartographic approaches. We have not accomplished some basic achievements in cartography that we should have by now. For example, why have we not solved the basic problem of choropleth maps, that is, that the largest areas and thus those with the most color draw the eye, even though that is often not the area that we want people focus on? And why have we not provided a solution to the basic problem of proportional symbol maps, that is, that the congested areas have too many overlapping symbols to be able to decipher? These are two of the most frequently used mapping methods, with problems that have long been known to cartographers, yet no solutions have been developed that have penetrated common usage of these methods. An of-suggested solution is to use choropleth mapping if proportional symbols do not work, or use proportional symbol if choropleth mapping fails. More often than not, a mapped area will have geographic units that defy either solution.

In this paper, I review the fundamental problems with choropleth and proportional symbol maps, I review two frequently suggested solutions and explain why one is sometimes deficient and the other is just plain wrong, and I offer two possible approaches to overcome the limitations of the traditional choroplethic mapping method.

1998 ACM Subject Classification Applied computing → Cartography

Keywords and phrases cartography, choropleth, proportional symbol, graduated symbol

Digital Object Identifier 10.4230/LIPIcs.GVIZ.2018.5

1 Introduction

New developments in a field must be built on an existing foundation. If that foundation is shaky, new developments will also be unstable. A solid foundation provides the support that is required to foster advancements in new directions and uphold the development of new

New Directions in Geovisual Analytics: Visualization, Computation, and Evaluation (GVIZ 2018).
Editors: Anthony C. Robinson and Antoni Moore; Article No. 5; pp. 5:1–5:8
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

techniques. For example, while new approaches in scientific inquiry and analysis have evolved and expanded in response to advancements in technology and knowledge, the scientific method basically remains the same: making observations, identifying a problem or question, formulating a hypothesis, and performing an experiment, which leads to additional observations, problems hypotheses, and experiments. The development of new visual representation types in geovisual analytics presumably would be built on a solid cartographic foundation. However, I challenge the solidness of that foundation and caution that new developments built on this foundation may fail due to a lack of agreement on the basic tenets of cartography.

Cartography is often described as "the art, science, and technology of map making". While most cartographers would modify this definition to be more inclusive of other ideas and activities in which the map is the basic intellectual unit, none would be likely to disagree that the first statement constitutes a large part of cartography. This common cartographic description nearly always starts with "art", and often the creative and aesthetic aspects of cartography are the focus of discussions and practices. Seen as an artistic endeavor, creative license is often sought, and too often granted, when in many cases, map making would benefit from less departure from the norm and more adherence to correctly applied methods. A cartographic approach is not synonymous with an artistic approach, or with a scientific approach, for that matter. The problem, however, is that there is no general agreement on what forms the basis of map making. Thus, there is often little agreement, even amongst cartographers, on how to correctly map a select set of data for a certain set of features or type of phenomena. While it is not the focus of this paper to offer such an expression of a cartographic foundation, I do offer some seeds for thought on the subject.

This lack of agreement on basic cartographic concepts and practices undermines any significant advancements in mapping methods or designs. If a solid foundation for cartography can be expressed and agreed upon, and proven to support the art, science, and technology of map making, then further advancement has a much higher chance of success.

2 The Basic Tenets of Cartographic Symbolization, or Not?

Most cartographers would agree that the basis of cartographic symbology lies in three related concepts: visual variables, levels of measurement, and feature dimensionality. Even this short list will be disputed, however. For example, the last item is often described in different ways or called by different names. Here, it means the dimensionality of the feature or phenomena being mapped. For example, a city conceived of as an area has two dimensions, but a city conceived of as a point has zero dimensions. The approaches for mapping areas differs from those for mapping points, so this distinction is an important one and therefore forms one of the basic tenets of cartography. Happily, most cartographers would agree that there are common methods that can be used for mapping zero-, one-, and two–dimensional features, such as proportional symbol, flowline, and choropleth maps, respectively. They may not agree on all the methods that can be used for each type of feature being mapped, and they most likely would not agree on a list of appropriate methods for mapping in three or more dimensions.

Regarding the second item in the "basis of cartographic symbology" short list, that is, levels of measurement, cartographers do not agree on what to call these "levels of measurement", much less what they include. Based on Stevens' [15] taxonomy, the measurement levels would include the nominal, ordinal, interval, and ratio scales. Some cartographers agree with this set of measurement levels [12, 5]. Sometimes, the interval and ratio scales are combined [17] resulting in three levels. Sometimes the levels are reduced to only two categories: qualitative and quantitative [8]. Chrisman [3] took the opposite approach and expanded the list to ten levels. These levels are referred to by different terms, such as levels of measurement [12] [13], data measurement [5], and levels of classification [17].

It would seem that there should at least be agreement on the last item in the "basis of cartographic symbology" short list, which is visual variables. With an unambiguous origin and meticulous description of meaning [1], it might seem that cartographers would agree on the visual variables and their appropriate uses. Most American cartographers do at least agree on the term "visual variables" (though Bertin originally called them "retinal variables"); however, the list of variables derived from Bertin's original set (size, value, texture, color, orientation, and shape) varies, sometimes widely (see [17], page 137, for a table with comparisons). Admittedly, Bertin [1](p. ix) was writing about principles related to a "grammar of graphics", and his concepts were not meant to be applied solely to cartography. When considered in relation only to mapped features, some of Bertin's options are not practical and are not actually used, even though they often appear in cartographic textbooks. For example, orientation is not in practice used to map either line or areas features, and size is not used to map area features. Confusion increases when additional variables are added, such as perspective height, spacing, arrangement, and focus.

If this kind of disagreement and confusion exists even among cartographers, imagine what non-cartographers think of our instruction and guidance in map making and other things cartographic. Without the agreed-upon and practically proven solid foundation our "art" requires, we cannot assume to receive the acceptance and reverence we desire and claim to be a "science".

Clearly disagreement abounds. Perhaps coming up with comprehensive lists for each of the three basic symbolization concepts is not the most productive way to reach agreement on the basic tenets of cartographic symbolization. Taking a minimalist approach, we might get most cartographers to agree on the following:

- The majority of features mapped can be conceived of as having zero, one, or two dimensions; features or phenomena with more dimensions require special attention.
- Many of the commonly used mapping methods can be easily be defined as appropriate for either qualitative or quantitative data.
- The most often used and easily interpreted visual variables for qualitative data include hue, shape, and orientation (the last for point features only); for quantitative data, they include size (for point and line features only), value (or lightness), and saturation (or chroma). Arrangement and pattern texture are also commonly used, but they are slightly different because they involve variation in the graphic elements that make up the symbol rather than variation of a single property of the symbol itself.

3 Mapping Methods

As with the basics of cartographic symbolization, it is likely that not all cartographers would agree on, say, the three most common mapping methods. Likely most would agree that two of them would be choropleth maps and proportional symbol maps (here taken to also include graduated symbol maps). For choropleth maps, the author of an online cartography course suggests that, "Their popularity is due to two main reasons: (1) choropleth mapping capabilities are implemented in most every GIS software packages; and (2) much of the data that geographers and GIScientists work with is collected and aggregated into enumeration units" (https://www.e-education.psu.edu/geog486/node/1864). It could be argued that proportional symbol maps are also popular due to software capabilities, and possibly also to the intuitive ability to decipher the symbology, especially when proportional symbols are used to represent point features.

While these two forms of maps have been used and studied widely (for example, [9] and [4]), there are some fundamental problems with both that cartographers still have not solved. Before (or in additional to) turning our attention to new mapping methods and design issues, it would behoove us to address the incessant problems with these prevalent mapping methods. Perhaps the suggestions in this paper could be considered "new approaches", but certainly they do not apply to "nascent challenges". Nonetheless, the importance of improving existing mapping methods is paramount because these methods (and their problems) are evidenced in many of the maps that have been created (and will be created), ingrained in the knowledge that people have about how to make maps, and encoded in the software that is used to make maps.

3.1 The Choropleth Conundrum

On a choropleth map, each enumeration unit (such as a census unit or political designation) is shaded according to the magnitude of a quantitative variable. Where there is more of something, the unit generally has a darker shade, and units with less of the thing will be lighter. Assumptions of choropleth maps, and indeed the visual impression they impart, is that values in the enumeration units are homogenous within each unit (that is, the phenomenon is present at the same intensity throughout each unit) and values can change abruptly between units (that is, at the units' borders).

An important reality is that the size of enumeration units on the map can and often do vary greatly. This creates a fundamental problem for choropleth maps because larger areas will dominate the visual appearance of the map. This is especially problematic when the importance of these particular enumeration units is low yet the mapping method exaggerates their appearance and thus their perceived importance. This happens more often than might be expected, especially for enumeration units based on population (such as census tracts), because the smaller units are where more people are located and the variation is of interest, while the larger areas have fewer people and are thus of less interest.

3.2 The Proportional Symbol Predicament

When mapping areal features, proportional symbol maps substitute area fills with point symbols, often circles, of different sizes. Like darker shades on choropleth maps, the larger symbols on proportional symbols maps are intuitively interpreted as representing higher values.

Proportional symbol maps have a problem the is almost the opposite that of choropleth maps. On these maps, large concentrations of small areas result in so many overlapping symbols that the geographic patterns in these areas are obscured. Symbols in the larger outlying areas are easy to see, but the clustered symbols in the areas that are generally of more importance and interest cannot be deciphered. This is especially true for population-related data because the values for smaller areas with higher concentrations of people will often be represented by larger circles, thus resulting in even more egregious overlap.

4 Solutions

While the problems posed by choropleth and proportional symbol maps are grave, they are not insurmountable. Here I explore two commonly advised solutions, one of which is a hit or miss solution and the other of which may be considered patently incorrect. I also offer two new solutions that have not been previously introduced.

4.1 Normalization

One solution has been to advise people to normalize the data; the hope is that in the normalization process, the values for the smaller units will increase and the values for larger units will decrease. Then, when the normalized values are mapped, the problem of larger areas having darker shades is solved. This is not always the case, however, so this is not always the solution. This is a hit or miss solution and its success depends entirely upon the data and units being mapped.

4.2 Mapping Method Substitution

If it is the non-normalized data that should be mapped, a solution that is often proposed is to substitute a choropleth map with a proportional symbol map. For example, the Wikipedia entry for choropleth maps contains this statement: "Another solution is to represent total amounts using a proportional symbol map". In the lesson for choropleth maps in an online cartography course, students are advised: "This is not to say that there is never a good reason for mapping raw counts, just that other symbolization methods may be more appropriate (e.g., graduated or proportional symbols)" (https://www.e-education.psu.edu/geog486/node/1864). However, there are situations in which this substitution will not work, primarily when the areal units being mapped include both large, less important areas, and small but important congested areas. This is typical of many urban areas where smaller census enumeration units are concentrated in central urban areas and larger less populated enumeration areas surround the city. In this case, as stated earlier, proportional symbol maps do not provide a solution.

Other substitutions may provide better results, including dot density and dasymetric mapping [18, 11]. The challenge with using either of these methods is the requirement for ancillary, related data. With each of these methods, the unit in which the symbols are drawn is reduced based ancillary data that identifies exclusion and sometimes inclusion zones. The value that is symbolized is recalculated to take the reduced area into account, thereby resulting in higher values (such as population density). Because the areas are reduced in size, the overwhelming appearance of larger areas may be mitigated; however, the higher values shown with darker colors or denser symbols may offset any advantages gained by the areal reduction. Nonetheless, it can be argued that the veracity of the mapped data is maintained. Additionally, isopleth maps generated using pycnophylatic interpolation [16] may also offer some promising results [10].

A far graver problem with substituting choropleth maps with proportional symbols maps is that when mapping areal units, the data for BOTH these maps should be normalized [2]. This concept, neither well understood nor adopted by cartographers, relates to whether the variable being mapped is spatially intensive or spatially extensive.

Spatially extensive variables are true only of the place as a whole; that is, they are dependent on the spatial units [14]. Examples include counts such as total population and amounts such as total income. These variables have summative values that are inseparable from their enumeration units [7]. For example, census block population data are collected in census blocks as the summative population in each block, and each data value represents the entirety of and cannot be detached from its census block.

Spatially intensive data is independent of the spatial units. Examples include data that has been normalized in some way, such as densities, rates, proportions, or percentages. These variables could potentially be true of every part of an area, if the area were truly homogeneous. Spatially intensive variables can be further applied to locations within an enumeration unit. For example, a population density value is applicable to every location in the area, rather than to the entire area as a whole [7].

If data are spatially extensive, you can add the values for multiple units to get a correct value. For example, the values for two enumeration units, one with 100 people and the other with 200 people, can be added to get a total of 300 people. On the other hand, data that are spatially extensive cannot be divided if the enumeration unit is subdivided. For example, if an enumeration unit is divided in half, the value associated with it cannot be equally distributed amongst the halves by dividing the value by two.

Because spatially extensive variables are determined by the size of discrete geographic features from which the measurements are taken [14], they must only be applied to the entirety of the discrete geographic features. Thus, caution must be taken to avoid mapping these variables incorrectly. Data that is spatially extensive (counts and amounts) should not be mapped using the choroplethic method, nor any other method that applies a single symbol to the entire area (including nearly all traditional area mapping methods, including proportional symbol, dot density, dasymetric, and filled isoline maps). For all of these maps, spatially extensive variables should be converted to spatially intensive variables. Luckily, it is possible to derive spatially intensive data from spatially intensive data. For example, dividing counts by area gives density, or dividing the count for one unit by the sum of counts for all units gives a proportion.

4.3 Masking

One solution is to apply a mask to filter out symbology in areas that do not have properties that relate to the variable being mapped. This mask is overlaid on a traditional choropleth map. For example, using a filter that relates to the percentage of developed impervious surface from the U.S. National Land Cover Dataset [19], Stuart Allan, of Raven Maps and Benchmark Atlases, and I effectively identified locations which likely have larger numbers of people. For population-based variables, such as per capita income or population density, the choropleth symbology (area fills) in the populated areas gives the correct impression of the true geographic pattern. Outlying areas with few people are masked, thus focusing attention on the patterns within the areas where there are likely to be people.

To create the mask, a cut-off point is selected for the percentage of impervious surface. For maps of all urban areas of the United States, we found through trial and error that areas with more than 15 percent impervious surface could be used to create the mask. This cut-off would likely be different for select urban areas. For other types of phenomena, different masks could be used. Again, drawing on the NLCD data [6], a percent tree canopy cover mask could be used to focus attention on vegetated areas when mapping themes such as insect or invasive plant infestation, timber yields, or annual harvests.

4.4 Symbol Substitution

Another method Allan and I have explored is use of the choropleth method with point symbols substituted for the area fills of the larger outlying areas; this is a method that was developed by Stuart Allan (personal communication) and Gene Martin. Allan describes the method as such: "Any polygon larger than (n) is replaced by a circle of slightly smaller diameter. When using this method, we do not print any boundary lines (stroke), unless the palette includes very light colors, in which case, the boundary line is thin and screened (or transparency is applied). The unit boundaries will not show up very well, but this is appropriate since no one polygon is that important. An advantage of this method is the no one mistakes a circle for a polygonal enumeration unit, so it is intuitively clear how the symbology changes. It is important to add a note to the effect in the legend.

The circles do not need to be as small as the smallest enumeration units, but they should not be larger than the cut-off size (the point at which the symbols shift to circles from choroplethic polygons). At some point, it can be determined that these small isolated circles are too small for the color to be clearly seen. This also happens to be true for the small (densely packed) area polygons, but there is nothing that can be done about that as the area sizes are what they are. Determining the cut-off size is achieved through trial and error. When each additional change makes the solution, we stop because the cut-off point has been discovered."

5 Conclusions

While we consider the methods experimented with to be promising, finer details about their application beg further exploration. For example, how can the cut-off points be determined more objectively? Is trial and error the best way to create the masks and identify the size of areas for which color fills should be substituted with circle symbols? How do readers perceive maps using these methods? Would other symbol substitutions be more effective, such as dasymetric area fills and proportional symbols or dot density area fills and proportional symbols (per Bernard Jenny, personal communication)? What is the best way to indicate in the legends that the basic and likely familiar mapping methods have been modified? Can these solutions be easily implemented in a map making workflow and as a software solution?

While we do not claim to have solved the fundamental problem with choropleth maps, we have at least risen to the challenge of trying to find a solution. As stated previously, I suggest that before adding new methods to our arsenal for mapping, we cartographers should rectify the problems with existing methods, thereby creating a defensible, agreed upon suite of mapping methods that can be accurately and appropriately applied to the types of data that are commonly mapped.

Acknowledgements. I want to thank Stuart Allan for his contributions to the methods described in this paper and Jon Kimerling for his ideas and suggestions. Thanks also to Bernie Jenny and an anonymous reviewer for their helpful comments and recommendations.

References

1 Jacques Bertin. *Semiology of Graphics: Diagrams, Networks, Maps.* Esri Press, Redlands, California, 2011.

2 Aileen R. Buckley. Understanding statistical data for mapping purposes. *ArcUser*, Winter:54–59, 2013.

3 Nicholas R. Chrisman. Rethinking levels of measurement for cartography. *Cartography and Geographic Information Science*, 25(4):231–242, 1998.

4 Robert G. Cromley, Shuowei Zhang, and Natalia Vorotyntseva. A concentration-based approach to data classification for choropleth mapping. *International Journal of Geographical Information Science*, 29(10):1845–1863, 2015.

5 Borden D. Dent, Jeffrey S. Torguson, and Thomas W. Hodler. *Cartography: Thematic Map Design.* McGraw Hill, Boston, Massachusetts, sixth edition, 2009.

6 Collin Homer, Jon Dewitz, Limin Yang, Suming Jin, Patrick Danielson, George Xian, John Coulston, Nate Herold, James Wickham, and Kevin Megown. Completion of the 2011 national land cover database for the conterminous united states - representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, 81(5):346–354, 2015.

7 Karen K. Kemp, editor. *Encyclopedia of Geographic Information Science*, chapter Discrete Versus Contiuous Phenomena: Spatially Extensive Versus Spatially Intensive Variable, page 112. Sage Publications, Los Angeles, California, 2008.

8 A. Jon Kimerling, Aileen R. Buckley, Phillip C. Muehrcke, and Juliana O. Muehrcke. *Map Use: Reading, Analysis, Interpretation.* Esri Press, Redlands, California, eighth edition, 2016.

9 Mitchel Langford and David J. Unwin. Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal*, 31:21–26, 06 1994.

10 Jie Lin and Robert Cromley. Inferring spatial scale change in an isopleth map. *The Cartographic Journal*, 54(1):48–60, 2016.

11 Jeremy Mennis. Generating surface models of population using dasymetric mapping. *The Professional Geographer*, 55(1):31–42, 2003.

12 Arthur H. Robinson, Joel L. Morrison, Phillip C. Muehrcke, A. Jon Kimerling, , and Stephen C. Guptill. *Elements of Cartography.* John Wiley and Sons, New York, New York, sixth edition, 1995.

13 Terry. A. Slocum, Robert B. McMaster, Fritz C. Kessler, and Hugh H. Howard. *Thematic Cartography and Geographic Visualization.* Pearson/Prentice-Hall, Upper Saddle River, New Jersey, third edition, 2005.

14 Michael J. De Smith, Michael F. Goodchild, and Paul A. Longley. *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques, and Software Tools.* Troubador Publishing Limited, Leicester, United Kingdom, sixth edition, 2018. Accessible online at http://www.spatialanalysisonline.com/HTML/attributes.htm.

15 S.S. Stevens. On the theory of scaling of measurement. *Science*, 103(2684):677–680, 1946.

16 Waldo R. Tobler. Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association*, 74(367):519–530, 1979.

17 Judith A. Tyner. *Principles of Map Design.* The Guilford Press, New York, New York, 2010.

18 J.K. Wright. A method of mapping densities of population with cape cod as an example. *Geographical Review*, 26(1):103–110, 1936.

19 George Xian, Collin Homer, Jon Dewitz, Joyce Fry, Nazmul Hossain, and James D. Wickam. The change of impervious surface area between 2001 and 2006 in the conterminous united states. *Photogrammetric Engineering and Remote Sensing*, 77(8):758–762, 2011.