# Review of Unclassed Choropleth Mapping

Brett Kelly
*University of Nebraska Omaha*
bckelly@unomaha.edu

*Although unclassed choropleth maps lead to a more accurate representation of data, grouping of data into classes is still common. Commonly–used data classification techniques such as equal–interval, quantiles, and natural breaks produce very different and possibly misleading representations. An unclassed map creates a distinct color for each unique value. The method was introduced by Tobler in 1973 using an x, y coordinate plotter that created crossed–line shadings. Tobler's unclassed proposal used grayscale values because color displays were not yet available. Current color monitors have the ability to display 16.7 million colors, while most GIS software packages have limits to their color ramps. QGIS defines color ramps with up to 999 classes. It is also possible to define up to 1000 classes in ArcMap, and ArcGIS Pro has an "Unclassed" option when styling choropleth maps. Utilizing more color classes results in a more truthful map due to minimizing error from the grouping of data. The unclassed method is examined here along with color ramps and classification schemes in QGIS and Esri's ArcMap/ArcGIS Pro. It is demonstrated that it is usually impossible to create a truly unclassed choropleth map using the default color schemes in these programs.*

**KEYWORDS:** unclassed; choropleth; QGIS; ArcGIS; data classification

## INTRODUCTION

TOBLER (1973) DEVELOPED a method of creating choropleth maps in which shading intensity is directly proportional to the data values (Figure 1). This introduction of unclassed choropleth maps challenged the practice of classifying data. Since classification was viewed as a central component of cartographic abstraction, unclassed choropleth mapping was never widely adopted.

Peterson (1979) and Müller (1979) tested the usability of unclassed choropleth maps. Peterson found that subjects had a better understanding of relative and absolute values for individual areas on an unclassed crossed-line map, compared with a classed one. Dobson (1980a) asserted that this method of mapping leads to information overload. Müller found that participants in a study could mentally sort unclassed map units into three classes (low, medium, and high) and that participants could thereby generalize the unclassed maps on their own. In commenting on this study, Dobson (1980b, 107) stated that Müller's task focused on pattern delineation but it should also focus on the map user's ability to memorize the map.

Dobson argued that the classed map is more useful since it creates a "simpler, more efficient communication device."

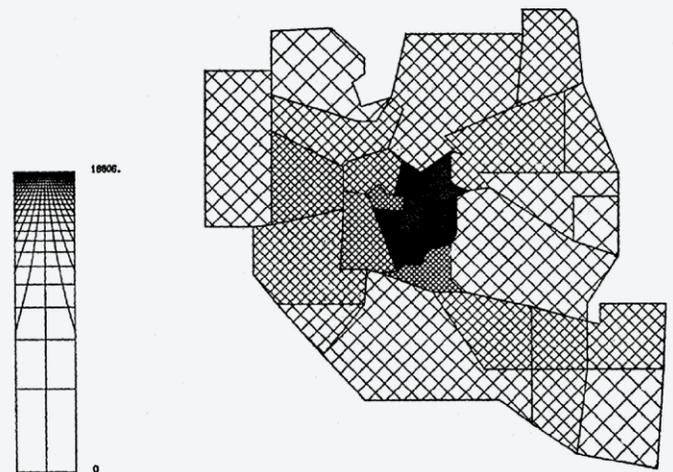The generalization of data is a common method in cartography (Peterson 2009; Axis Maps 2015) and has been



**Figure 1.** *Tobler's (1973) crossed-line shading method first made it possible to create unclassed choropleth maps.*

the focus of much research and discussion (Gilmartin and Shelton 1990; Slocum et al. 2008). Jenks and Caspall (1971) determined different types of error associated with data classification and developed an algorithm that balanced the level of these errors. This resulted in a method that uses statistical "natural breaks" in the data to create classes. This method is usually the preferred classification option, along with equal-interval and quantiles. These different classification schemes can be difficult for the average map-reader to comprehend. The point of unclassed mapping is not to be able to efficiently match data values to a specific color on a legend, but rather to represent a dataset without introducing error.

Other research has considered alternative methods for creating a continuum of shadings. Peterson (1992) used PostScript to create a large number of dot shadings for printing. Cromley (1995) criticized this method and proposed an alternative classification scheme that used fewer classes. Stewart and Kennelly (2010) created a method of using soft shadows (Figure 2) from an unclassed map, and mashing that up with a classed choropleth map. The free (but discontinued) Java applet *mapresso* allowed the cartographer to create a continuous-tone map as seen in Figure 3. The algorithm in the applet handled extreme values by placing them into their own classes (Herzog 2015). By classifying the outlier values separately, the applet does not correctly represent the true intensity of the data, but does create a more visually appealing map. Kenneth Field (2013) used a diverging color scheme to create what he called an unclassed map of the 2012 US presidential election (Figure 4), using blue and red to represent Democratic and Republican support; many counties have the same color.

## THE 999-CLASS MAP

QGIS, AN OPEN SOURCE program for GIS, allows a maximum number of 999 classes in a graduated color ramp. In some datasets, more than 999 different colors would be needed to ensure a map was truly unclassed. Additionally, limitations on the number of colors in a gradient can make creating unclassed maps difficult. In ArcMap, it is possible to use the "Defined Interval" option to create up to 1,000 classes for maps (Figure 5). The technique to achieve this many classes is to shrink the interval size until the number of classes equals 1,000. An error will result if more than 1,000 classes are specified.



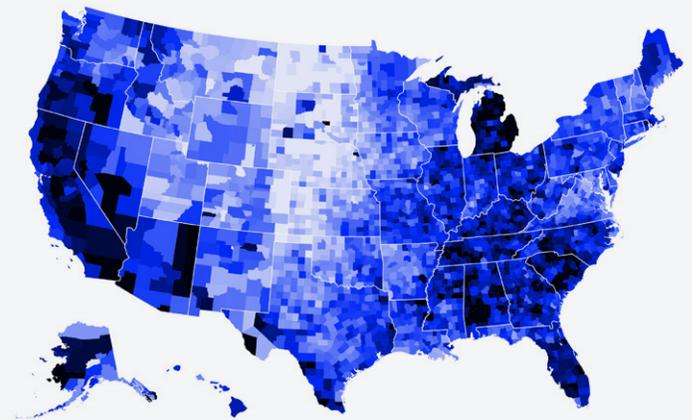*Figure 2.* Stewart and Kennelly's (2010) classed with unclassed soft shadows method.



*Figure 3.* Mapresso continuous-tone map (Herzog 2015). This map shows unemployment rates by county in 2008. This method deals with the issue of extreme values by classing them at the top and bottom. The mapresso applet has since been discontinued due to issues with Java support in web browsers.
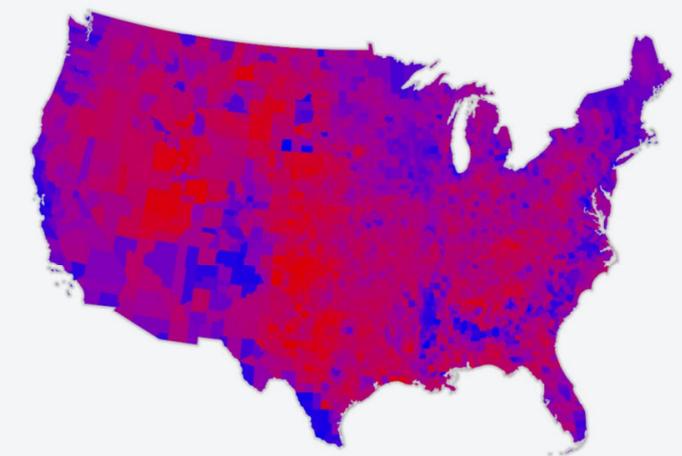


*Figure 4.* Field's (2013) map shows the presidential election results from 2012. Different shades of blue and red in a diverging color scheme are used to indicate the percent of Republican and Democratic votes in each county.

A problem arises in both ArcGIS and QGIS when using the default color gradients. Instead of a specific, unique color for each distinct data value, each color spans a large number of values. The maps are de facto classed, essentially due to too few colors in the gradient. This was determined by finding the RGB values of the color for each county for the two maps in Figure 6 using a color picking program. In QGIS, there are 12 duplicate colors. This means that only 81 colors are being used to represent the 93 values (each county had a different data value). In ArcMap, there were 23 duplicates, two triples, five where four counties had the same color, and four instances where five counties had the same color. This means only 43 unique colors are representing 93 unique values.

Neither the default color ramps in QGIS nor ArcMap can produce an unclassed choropleth map for these 93 values. A separate program, ArcGIS Pro, provides an "unclassed"
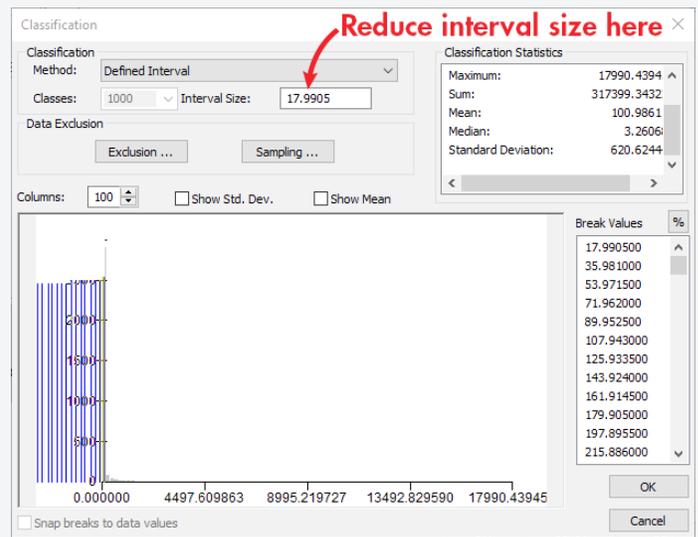


*Figure 5.* The "Interval Size" can be reduced until the number of classes equals 1,000. In this case, 17.9905 is the smallest usable interval size that will create 1,000 classes.
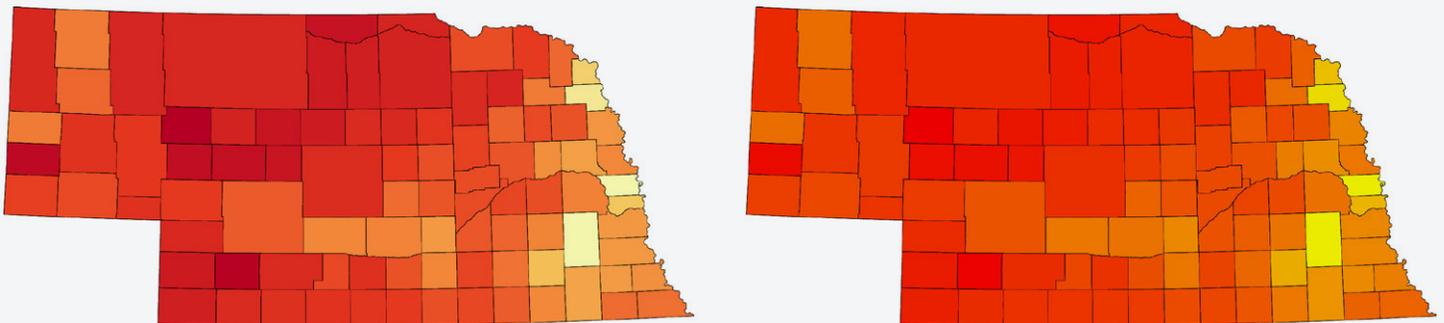


*Figure 6.* A 999-class QGIS map (left) and a 1,000-class ArcGIS map (right) showing percent of vote for Donald Trump by county in Nebraska. The two programs have different default color ramps. The QGIS map has 81 different colors for the 93 different values while the map from ArcMap only has 43 colors. This difference is likely due to the lack of available colors in Esri's ArcMap color palette.
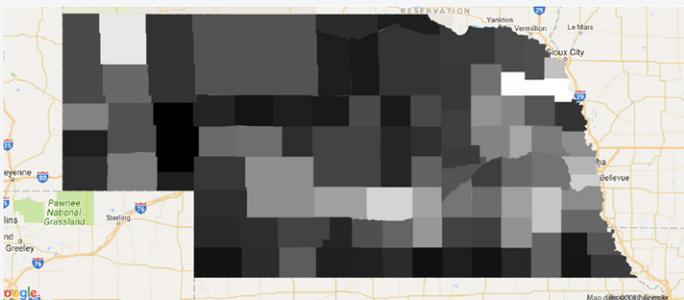


*Figure 7.* This map of median age in Nebraska (2000) has different shades for every unique value. The data set has 24 duplicate values so only 69 different shades are needed (93 counties – 24 duplicate data values = 69). The minimum difference in data values is 0.1 and the range is 24.2, so 242 different gray shades are needed to create an unclassed map (working example available at: *maps.unomaha.edu/cloud*). It should be noted here that ArcGIS and QGIS could also produce unclassed maps of this data set.

option for choropleth mapping (Esri 2017), but the same problem occurs. Multiple areas on the map have the same color. The name of the option in ArcGIS Pro misleads the user into thinking that this option can actually create an unclassed map in all cases. Of the three programs, QGIS's default color schemes come closest to producing an unclassed map with the 999 shadings setting. Note that this research is focused on the default color ramps used in these programs. There are custom options in both ArcGIS and QGIS that may enable a user to incorporate their own color ramps. The use of Javascript alongside some experimentation with the Google Maps API yields an unclassed choropleth map with some datasets.

Figure 7 shows an unclassed map of the median age of Nebraska's population, made with the Google Maps API. This data is linear in nature, and with duplicates, there
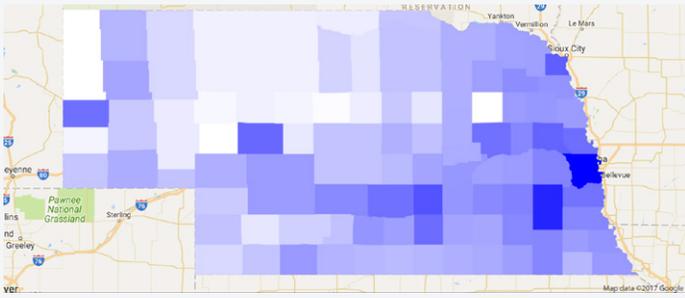
*Figure 8. Population density in Nebraska in 2010. In this map, there are 80 different colors used for 85 different data values (including 8 duplicate values). It is not an unclassed map, so a "percent unclassed" value can be calculated. There are five counties that are classified, and 88 counties that are not. Dividing the number of unclassed counties by the total number of counties equals 94.62%. In this case a simple disclaimer, "94.62% Unclassed" OR "Percent Classed: 5.38%," should be included with the map to describe the amount of classification caused by an insufficient number of colors. Working example available at:* maps.unomaha.edu/cloud.

are only 69 different values among the 93 counties. The number of unique colors needed to display this map as unclassed is found by dividing the range of values, 24.2, by the minimum difference between two non-duplicate data values, which is 0.1. 242 different colors are needed. Using the minimum difference between two values as the interval size ensures that there are no identical colors being used for different values. Since the range is small enough, and the difference between the two closest values is large enough, 256 brightness values or gray shades are sufficient for an unclassed map of this data distribution. This formula can be used to determine the number of colors or shades of gray that are needed for any data set. It is also possible to provide a percentage of classed or unclassed values for any given choropleth map (Figure 8), by dividing the number of classed or unclassed values by the total number of values. Working examples of these two figures on the web are available at **maps.unomaha.edu/cloud**, the companion website to Michael Peterson's *Mapping in the Cloud* (2014), under Code 14. Note that through customization, both ArcGIS and QGIS should be able to create the maps seen in Figures 7 and 8.

## THE 16.7 MILLION-COLOR SOLUTION

A typical computer screen can display 16.7 million colors. This is because most computer screens use a 24-bit color depth of red, green and blue (RGB) values. Each hue (red, green and blue) has a value between zero and 255, resulting in 16.7 million (or $2^{24}$) possible colors. With 16.7 million colors, an unclassed color gradient with more than 1,000 colors is conceivable.

Simultaneous contrast, or the psychological effect of differences in color perception based on surrounding colors, is certainly an issue when using this many colors (Gruver 2017). Although our screens display 16.7 million colors, we may have difficulty determining whether two colors are the same or different. This has been one justification for data classification. Of course, the amount of error in interpretation caused by simultaneous contrast, that is, the inability of the map reader to differentiate colors, is minor compared to the error introduced by classification. The common classification methods in QGIS and ArcMap use five to seven colors, and group large swaths of data values into the same category. Since these methods vary on how they divide and represent the same data, more interpretive

error is introduced by relying on a default classification method and color gradient, than by using a greater number of colors to distinguish between different data values.

The misrepresentation of data is common with classification, as demonstrated by Monmonier (2005) with different maps showing crude birth rate in the United States. He described how a person with malicious intent could purposefully deceive their audience by manipulating class intervals to show too few or too many births. A more common scenario is a mapmaker using a common classification method without understanding how it may misrepresent the data. This can happen every time a classed choropleth map is made, if the cartographer is not familiar with these methods.

The 999-class QGIS option provides the map user with more information because it displays a map with many colors. Although the method usually does not actually create 999 unique colors, it still promotes a less-generalized graphic representation. The default color ramps in ArcMap and ArcGIS Pro provide far fewer colors.

## COLOR RAMPS

AN EXAMPLE OF many of the default color ramps from QGIS and ArcMap can be seen in Figure 9. Each color ramp consists of 11 boxes ranging from zero to 100 (e.g., 0%, 10%, 20%, etc.). These color ramps were created using the maximum number of classes for both QGIS and ArcGIS. Both packages lack sufficient colors for unclassed choropleth maps. They also use very different default color schemes. The QGIS color ramps span a greater visual range, with much lighter and much darker values than their equivalents in ArcGIS. Consider the lightest value of blue in the "Blues" color ramp below; it is much lighter in QGIS, in fact it is nearly white. Compare this to the blue gradient in ArcGIS.

## CONCLUSION

THE USE OF COLOR RAMPS to create a 999-class map in QGIS is a straightforward task, but, depending on the dataset, does not usually yield 999 different colors. In ArcMap, it is not as easy to create the 1,000-class map, and the software does not use 1,000 different colors. ArcGIS Pro has a misleading "unclassed" option, and it does not provide unique colors for every unique data value. In fact, it uses fewer colors than QGIS's 999-class option. Future research should study the potential of developing color ramps with many more colors to facilitate true un-classed choropleth mapping.

It is nearly impossible to create an unclassed choropleth map with the currently available software unless the data values permit it (i.e., there are a small number of unique data values that are sufficiently spaced). A method to calculate whether an unclassed map is possible for a particular dataset is included in the Google Maps API example showing median age in Nebraska. This method that displays the level of classification caused by an insufficient number of colors should be incorporated in all choropleth mapping software.
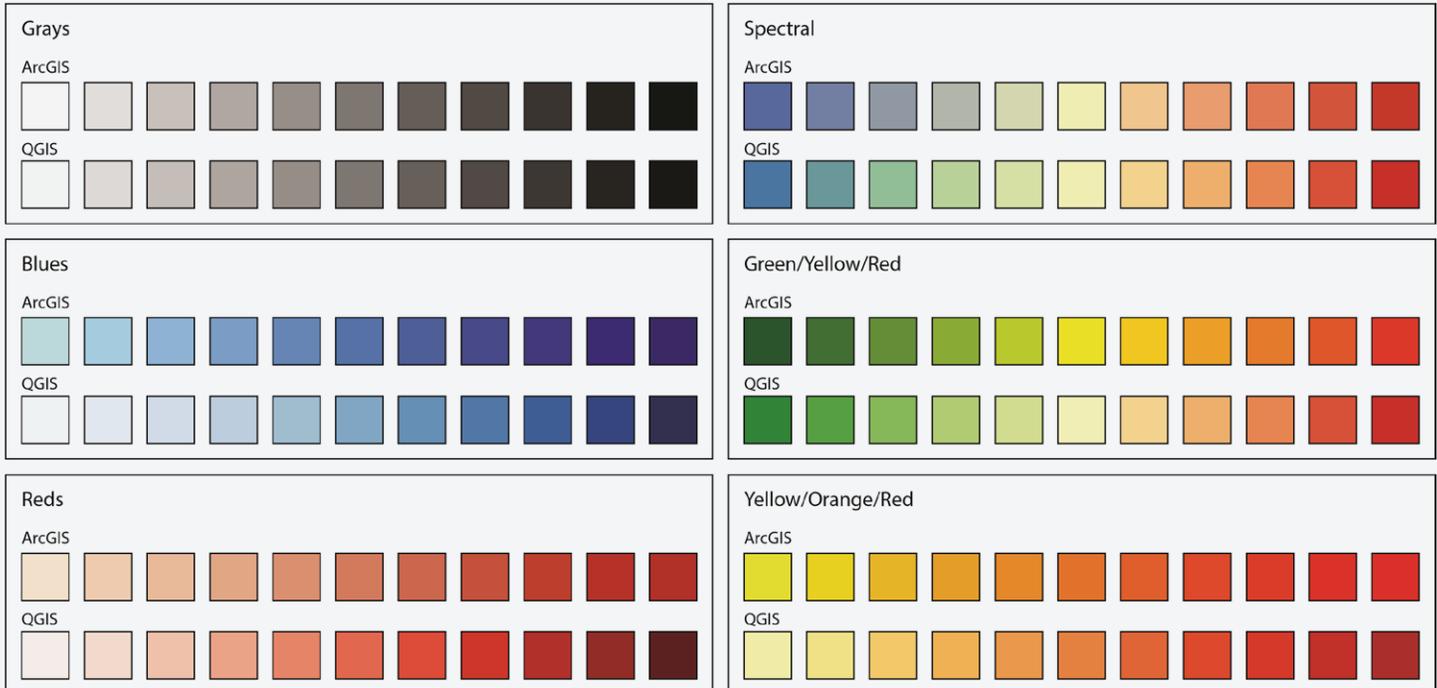


*Figure 9.* A comparison of color palettes from ArcGIS and QGIS using the generic graduated color symbols provided by each software. The boxes represent 11 polygons with attributes ranging from zero to 100 (0 on the left, 50 in the middle, and 100 on the right for each ramp).

# REFERENCES

Axis Maps. 2015. "Thematic Cartography Guide." Accessed March 6, 2017. http://www.axismaps.com/guide/choropleth.

Cromley, Robert G. 1995. "Classed versus Unclassed Choropleth Maps: A Question of How Many Classes." *Cartographica* 32 (4): 15–27. doi: 10.3138/j610-13nu-5537-0483.

Dobson, Michael W. 1980a. "Unclassed Choropleth Maps: A Comment." *The American Cartographer* 7 (1): 78–80. doi: 10.1559/152304080784522928.

———. 1980b. "Commentary: Perception of Continuously Shaded Maps." *Annals of the Association of American Geographers* 70 (1): 106–107. doi: 10.1111/j.1467-8306.1980.tb01301.x.

Esri. 2017. "Unclassed Colors." Accessed March 29, 2017. http://pro.arcgis.com/en/pro-app/help/mapping/symbols-and-styles/unclassed-colors.htm.

Field, Kenneth. 2013. "Choropleth - Unclassed." Accessed March 6, 2017. https://www.arcgis.com/home/item.html?id=4753cd8f1cf143eeb1890deb0bbfc415.

Gilmartin, Patricia, and Elisabeth Shelton. 1990. "Choropleth Maps on High Resolution CRTs: The Effects of Number of Classes and Hue on Communication." *Cartographica* 26 (2): 40–52. doi: 10.3138/w836-5k13-1432-4480.

Gruver, Adrienne. 2017. "Choropleth Maps." Accessed April 10, 2017. https://www.e-education.psu.edu/geog486/node/1864.

Herzog, Adrian. 2015. "Unclassed Choropleth Map." Accessed March 6, 2017. http://bl.ocks.org/aquilo/f39dec65b716fad5bb1e.

Jenks, George F., and Fred C. Caspall. 1971. "Error on Choroplethic Maps: Definition, Measurement, Reduction." *Annals of the Association of American Geographers* 61 (2): 217–244. doi: 10.1111/j.1467-8306.1971.tb00779.x.

Monmonier, Mark. 2005. "Lying with Maps." *Statistical Science* 20 (3): 215–222. doi: 10.1214/088342305000000241.

Müller, Jean-Claude. 1979. "Perception of Continuously Shaded Maps." *Annals of the Association of American Geographers* 69 (2): 240–249. doi: 10.1111/j.1467-8306.1979.tb01254.x.

Peterson, Gretchen N. 2009. *GIS Cartography: A Guide to Effective Map Design*. Boca Raton: CRC Press. doi: 10.1201/9781420082142.

Peterson, Michael. 2014. *Mapping in the Cloud*. New York: Guilford Press.

———. 1979. "An Evaluation of Unclassed Crossed-line Choropleth Mapping." *The American Cartographer* 6 (1): 21–37. doi: 10.1559/152304079784022736.

———. 1992. "Creating Unclassed Choropleth Maps with Postscript." *Cartographic Perspectives* 12: 4–6. doi: 10.14714/cp12.1028.

Stewart, James, and Patrick J. Kennelly 2010. "Illuminated Choropleth Maps." *Annals of the Association of American Geographers* 100 (3): 513–534. doi: 10.1080/00045608.2010.485449.

Slocum, Terry A., Robert B. McMaster, Fritz C. Kessler, and Hugh H. Howard. 2008. *Thematic Cartography and Geovisualization, 3rd Edition*. Upper Saddle River, NJ: Prentice Hall.

Tobler, Waldo R. 1973. "Choropleth Maps Without Class Intervals?" *Geographical Analysis* 5 (3): 262–265. doi: 10.1111/j.1538-4632.1973.tb01012.x.