

4-1-2019

(Mis)Communicating with Geographic Information System Mapping: Part 2—Determining Data Cutoffs

Rodrigo Cantarero
University of Nebraska

Maria Rosario de Guzman
University of Nebraska

Sarah Taylor
California State University

Soo-Young Hong
University of Nebraska

Jeong-Kyun Choi
University of Nebraska



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Recommended Citation

Cantarero, R., de Guzman, M. R., Taylor, S., Hong, S., & Choi, J. (2019). (Mis)Communicating with Geographic Information System Mapping: Part 2—Determining Data Cutoffs. *Journal of Extension*, 57(2). Retrieved from <https://tigerprints.clemson.edu/joe/vol57/iss2/9>

This Tools of the Trade is brought to you for free and open access by TigerPrints. It has been accepted for inclusion in *Journal of Extension* by an authorized editor of TigerPrints. For more information, please contact kokeefe@clemson.edu.

(Mis)Communicating with Geographic Information System Mapping: Part 2—Determining Data Cutoffs

Abstract

The increasing use of geographic information system (GIS) technology in various fields suggests the need for professionals, including those in Extension, to be mindful of communicating data accurately and effectively. This article examines approaches to creating classes or groupings within data as well as the weaknesses of each approach. Data break units discussed in the article include equal intervals, quantiles, and units resulting from natural breaks. Ideal situations for each type of data break are presented. The article emphasizes the need for Extension professionals to consider the effects of data grouping to avoid miscommunication when using GIS mapping.

Keywords: [geographic information system \(GIS\)](#), [misrepresentation in GIS use](#), [Extension and geospatial technologies](#)

Rodrigo Cantarero
Associate Professor
University of
Nebraska–Lincoln
Lincoln, Nebraska
rcantarero1@unl.edu

**Maria Rosario T. de
Guzman**
Associate Professor
and Extension
Specialist in
Adolescence
University of
Nebraska–Lincoln
Lincoln, Nebraska
mguzman2@unl.edu

Sarah Taylor
Assistant Professor
California State
University, Long
Beach
Long Beach, California
sarah.taylor@csulb.edu
[u](#)

Soo-Young Hong
Associate Professor
University of
Nebraska–Lincoln
Lincoln, Nebraska
shong5@unl.edu

Jeong-Kyun Choi
Associate Professor
and Early Childhood
Extension Specialist
University of
Nebraska–Lincoln
Lincoln, Nebraska
jchoi@unl.edu

Given the increasing popularity of geographic information system (GIS) technology in various fields, including Extension, it is imperative that professionals be aware of factors that maximize its communicative potential. More importantly, knowledge of these factors can help one avoid miscommunicating or misrepresenting data. In Part 1 of this two-article set, also published in this issue of the *Journal of Extension*, we discuss the importance of choosing the appropriate unit of data representation (i.e., count, percentage, location quotient) and potential implications for data interpretation (see <https://joe.org/joe/2019april/tt3.php>). In this article, we turn our attention to an often overlooked but very important aspect of mapping—namely, how data are organized or broken into classes. How data are broken

down into classes in maps can lead to communication of very different stories; accordingly, it is imperative that those who use GIS mapping be mindful of this issue.

In mapping, data typically are organized into discrete categories (e.g., 0%–25%, 26%–50%, 51%–75%, 76%–100%). The most popular approaches to classifying data are the equal intervals, quantiles, and natural breaks methods (with the last being the default setting in most GIS programs). Although seemingly a minor detail, the choice of methodology for determining breaks can result in drastically different maps. Herein, we illustrate this concept by using the three approaches to represent the concentration of ethnic minorities by county in Nebraska using county data from the 2010–2014 U.S. Census (U.S. Census Bureau, 2014).

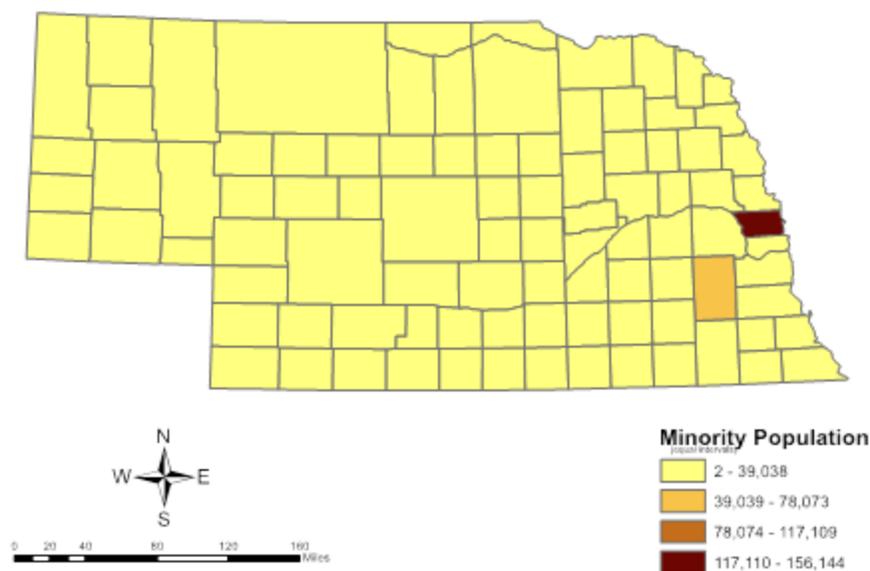
Equal Intervals

The term *equal intervals* simply means that all data classes are about the same size (e.g., 0%–25%, 26%–50%, 51%–75%, 76%–100%). Figure 1 exemplifies the use of the equal intervals method, with classes calculated by dividing the highest count by the number of categories we want. Some scholars published in the *Journal of Extension* have used equal intervals in mapping various data (e.g., Veregin, 2015). In our example, Douglas County had the highest number of minorities (156,144). Thus, we divided this figure by the number of classes we had predetermined ($156,144 \div 4$), creating intervals of approximately 39,000. One county falls in the class with the highest minority population range, one falls in the class with the third highest range, and the remaining counties are in the class with the lowest range.

The equal intervals approach (Figure 1) highlights the influence of high concentrations of population (skewness). Because Nebraska's population distribution is highly concentrated in two counties (Douglas County, where Metro Omaha is located, and Lancaster County, where the capital of Lincoln is located), so is its minority population.

Figure 1.

Concentrations of Ethnic Minorities in Nebraska (Equal Intervals)

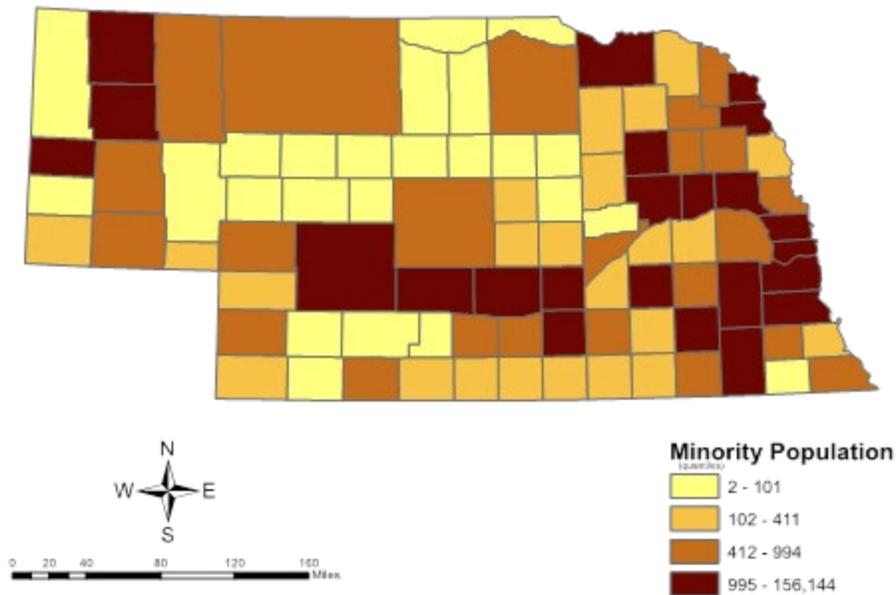


Quantiles

In the quantiles approach, the number of groups is predetermined. Cases (e.g., counties) are then sorted by the target variable (e.g., number of ethnic minorities) and then divided into equal groups, each containing the same number of cases. Because there are 93 counties in Nebraska and we had predetermined four data classes, each class has approximately 23 counties. As indicated by the breaks in the data in Figure 2, 25% of the counties fall within the 2–101 range for number of minorities, 25% in the 102–411 range, 25% in the 412–494 range, and 25% in the 995–156,144 range.

Figure 2.

Concentrations of Ethnic Minorities in Nebraska (Quantiles)



The stories from the two maps shown in Figures 1 and 2 could not be more different despite the fact that both represent the same data and the same unit of representation (i.e., counts/frequencies). Someone examining Figure 1 might surmise that there is ethnic minority concentration in only two counties, whereas someone examining Figure 2 would likely think that there is great dispersion of ethnic minority populations across the state.

An important factor to consider in choosing whether to use equal intervals or quantiles is whether the population distribution is skewed. In our example, one county in Nebraska has only two ethnic minorities, and another has 156,144. In addition, there are two counties with somewhat large populations and correspondingly large ethnic minority populations. All other counties have low numbers of ethnic minorities (less than 39,000). Thus, based on use of equal intervals, Figure 1 has only two counties highlighted. In contrast, the use of quantiles (Figure 2) forces cases into each category. Thus, the highest quantile group is overrepresented (weighted in favor of the group with the highest minority population range)—making it seem as though there are many "hot spots" of minority population.

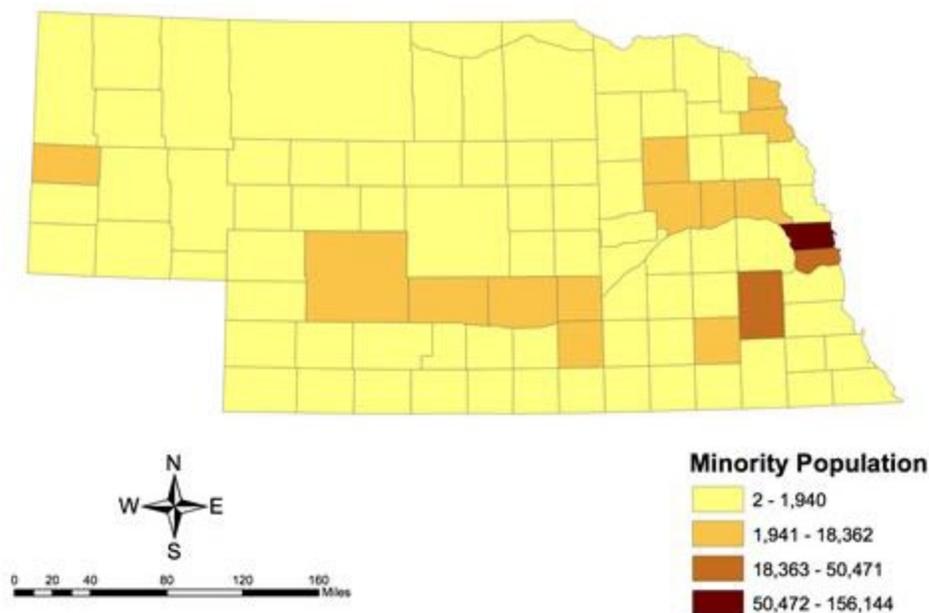
Natural Breaks

The natural breaks approach ameliorates the extremes affecting the equal intervals and quantiles methods as they are based on the average. Natural breaks in the data are determined through an algorithm that

maximizes between-class variance and minimizes within-class variance. In other words, the natural breaks method makes classes where there are clusters in the data (population) distribution. The natural breaks method is the default setting in many mapping programs, and its calculation can be found in earlier scholars' descriptions (e.g., Jenks, 1967). Several examples of the use of the natural breaks approach can be found in articles published in the *Journal of Extension* (e.g., Harris, Aboueissa, Jacobus, Dharod, & Walter, 2010; Rebori & Burge, 2017). Regarding our example of mapping ethnic minority concentrations in Nebraska, Figure 3 shows one county in the group with the highest minority population range; two in the group with the second highest range; 13 in the group with the third highest range; and the rest in the group with the lowest range.

Figure 3.

Concentrations of Ethnic Minorities in Nebraska (Natural Breaks)



Conclusion

Reflecting on such issues as the purpose of a mapping exercise, the geographic spread of the data (e.g., are the data of interest highly concentrated in a handful of areas?), and existence of outliers can guide Extension professionals in their interpretation and development of maps for programming. In our example, if our intention is to highlight locations with high concentrations of ethnic minorities, the equal intervals approach (Figure 1) may be most helpful. However, equal intervals would be a poor choice if we want to communicate where in the state we can find more ethnic minorities. If our need is to highlight counties in which special attention should be paid to new audiences or topics relevant to ethnic minorities, the natural breaks approach (Figure 3) might be most useful. As for quantiles (Figure 2), it is difficult to envision a situation in which this would be most useful. Given the nature of our data, the use of quantiles (Figure 2) is in fact quite misleading, exaggerating the presence of minorities and suggesting greater dispersion across the state. Nonetheless, quantiles may be useful in other scenarios, particularly if one is mapping data that are more evenly distributed across the range. Understanding issues around the seemingly small detail of determining data breaks is extremely important in effectively communicating with maps and avoiding miscommunication.

Acknowledgment

Our study was funded by a Research and Engagement Grant from the Rural Futures Institute at the University of Nebraska.

References

Harris, D. E., Aboueissa, A. M., Jacobus, M. V., Dharod, J., & Walter, K. (2010). Mapping food stores & people at risk for food insecurity in Lewiston, Maine. *Journal of Extension*, 48(6), Article 6RIB3. Available at: <https://joe.org/joe/2010december/rb3.php>

Jenks, G. F. (1967). The data model concept in statistical mapping. *International Yearbook of Cartography*, 7, 186–190.

Rebore, M. K., & Burge, P. (2017). Using geospatial analysis to align Little Free Library locations with community literacy needs. *Journal of Extension*, 55(3), Article 3TOT3. Available at: <https://www.joe.org/joe/2017june/tt3.php>

U.S. Census Bureau. (2014). *2010–2014 American Community Survey 5-year estimates*. Retrieved from <https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=CF>

Veregin, H. (2015). Using maps in web analytics to evaluate the impact of web-based Extension programs. *Journal of Extension*, 53(3), Article 3IAW2. Available at: <https://www.joe.org/joe/2015june/iw2.php>

Copyright © by Extension Journal, Inc. ISSN 1077-5315. Articles appearing in the Journal become the property of the Journal. Single copies of articles may be reproduced in electronic or print form for use in educational or training activities. Inclusion of articles in other publications, electronic sources, or systematic large-scale distribution may be done only with prior electronic or written permission of the Journal Editorial Office, joe-ed@joe.org.

If you have difficulties viewing or printing this page, please contact JOE Technical Support