Learning Proper and Ethical Data Classification for Thematic Maps

R. Cammack

Geography and Geology, University of Nebraska Omaha, Omaha, USA
rcammack@unomaha.edu

Keywords: Cartographic Education, Data Classification, Ethical Cartography

Introduction – Challenges of Data Classification

When thematic maps became part of the cartographic framework, cartographers grew to understand that generalizing information into classes and setting fixed symbology for the data created a trade-off between factual information and spatial representation. This trade-off is widely referred to as a part of map generalization. In the traditional map world this led to a loss of the exact data values, but it gained by displaying simplified spatial patterns. This process is one part of what has been referred to as thematic mapping (Dent,1996; Slocum et al., 2009), cartographic data visualization (Kraak and Ormeling, 1996), or visualization (Maceachren and Kraak, 1997).

Though this aspect of cartography is fundamental to the map communication dynamic, new cartographic practitioners' must experience this process and understand how its many nuances develop different outcomes. The cartographer's decision regarding data classification changes the map and in turn changes the map reader's experience. As a cartography educator, getting new cartographers to develop a refined method of reasoning for classification is paramount to mapping and spatial communication. Cartographic communication is the medium where cartographic ethics are applied.

In this ongoing research project, an educational tool is being developed regarding data classification of thematic maps. In this first phase of the project a cartographic educational tool is focused on data classification methods for choropleth maps. The intent of the tool is to facilitate the process for a novice cartographer to gain experience in understanding the merits and drawbacks of different data classification methods. The goal of the education tool is to aid in learning how to classify data that provides both a proper and ethical classification.

Method

Developing education tools for cartographic education can take numerous forms. The goals in building these education tools are multifaceted. The primary purpose is to create an environment that helps novice cartographers understand both the numerical and ethical implications of data classification. Another of these goals is showing how computer programming is an important tool for cartographers to understand and, for those who embrace it, a powerful skill to build a cartographer's portfolio. With this goal in mind a review of the development environment is pertinent to the research project. In this project, Python programming language (Python Software Foundation) was chosen as the development language. In addition to Python several powerful packages are included in the development environment: NumPy (NumPy Project), Pandas (Pandas Developer Team), Bokeh (Bokeh Development Team). For project management and development Anaconda® (Anaconda Software Distribution) was used to manage the development environment for Python and the packages. For the coding and the phase one user experience the Spyder (Spyder IDE) Integrated Development Environment (IDE) was used. This combination of development technology was used to have an open source like approach to the learning tool. This open-source approach encourages students to see behind the tool into the underlaying computer code. In future phases of this research this tool will be used, and students can develop their own ideas and integrate it with the tool.

For this phase of the project the primary factors for the classification are classification method and number of classes. The interplay of these two factors creates a classification that results in a numerical generalization and a spatial pattern. All cartography software supports this approach for choropleth mapping. How this approach varies is that it creates a set of data visualization that examines the data classification in non-spatial manner. The creation of this visualization enables the cartographer to explore how the classification is converting the exactness of the data into symbolic groupings. Classification methods included in the tool so far include: Equal Steps, Equal Observation, Quantile, Standard Deviation, and Natural Breaks. The classification tool starts with a classically hard data set to classify but also students to load their chosen data set.

In addition to the data visualization outputs based on the number of classes and classification method, students can also get all the computations in an easy-to-read report. This report also includes variations of how to report the classification as a legend element. Many times, students never consider how the classification is shown in the legend has a major effort on the map reading experience.

Results

At the time of this submission the cartography learning tool was not tested with cartography students. A set of cartography students will have used the tool to learn about the classification and data generalization prior to the presentation of this research. After the students used the tool for an assignment in the cartography design lab, a monitored discussion will be conducted. This qualitative interview will cover topics of use of the tool, classification methods, classification visualization, classification reports, assessment of classification and its role in understanding a data sets. A part of this discussion will examine how decisions lead to positive or negative cartographic ethics.

Discussion and Conclusion

Map generalization is a key component to understand how maps work to communication information. To do this ethically is not learned easily. In cartographic education or cartographic software use a focused effort needs to be made to aid in understanding how cartographic decision making for map generalization is part of the foundations to cartographic ethics. The aim in developing this thematic map classification tool is to assist the novice cartographer in understanding how their choices in method of classification and number of classes interplay to make a representation. The two principal products of classification allow cartographers to fairly represent a dataset in both a numerical and spatial pattern.

References

Anaconda Software Distribution. Computer software. Version. 2-2.4.0. Anaconda, Nov. 2016. Available at https://anaconda.com.

Bokeh Development Team. (2022) Bokeh: Python Library for interactive visualization, version 2.43 Available at https://bokeh.org.

Dent, B. D. (1996). *Cartography: Thematic Map Design*. 4th eds. Wm. C. Brown.

Pandas Development Team. Pandas, version 1.4.2. Doi.org/10.5281/zenodo.3509134.

Python Software Foundation. Python Language Reference, version 3.10. Available at http://www.python.org.

Maceachren, A.M., & Kraak. M. (1997). Exploratory cartographic visualization: Advancing the agenda. Computers & Geosciences, 23(4). Doi:10.1016/S0098-3004(97)00018-6.

NumPy Developers. NumPy Reference, version 1.22. Available at https://numpy.org/doc/stable/index.html.

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2009). *Thematic Cartography and Geovisualization*. 3rd eds. Prentice Hall.

Spyder IDE (2022). Spyder: The Scientific Python Development Environment. version 5.1.5 Available at https://www.spyder-ide.org/.