# Preserving attribute value differences of neighboring regions in classified choropleth maps

Jochen Schiewe*

* HafenCity University Hamburg, Lab for Geoinformatics and Geovisualization (g2lab), Germany

## Extended Abstract

In **choropleth maps** the underlying attribute values might be represented either in an unclassified or in a classified manner. While the first choice maintains numerical values and their relations (which might be necessary for some exploration tasks), the – more frequently used – second choice has advantages for overall impression purposes by enhancing value differences and allowing for an easier visual matching with the legend.

For a **classification** commonly used methods perform a division into intervals purely along the number line. Methods like equal intervals, quantiles, mean-standard deviation, maximum breaks, or natural breaks are typical representatives which are implemented in standard GI and cartography software packages. It is well known that applying different methods – together with varying the number of classes, the display of border classes (open vs. closed) and the consideration of extreme values – influence the overall visual impression and with that the message of a map. Depending on the task related to a map (or the goal of the cartographer, respectively) specific classifications should be selected on good grounds.

A **limitation** of the aforementioned classification methods, performing a division into intervals along the number line, is that they fail to consider the actual spatial context of underlying data. By doing so, possibly existing spatial relationships of neighboring units or patterns across the entire map space might get lost. Only little work has been performed on this topic, most of it is motivated by the idea of just simplifying the map pattern.

In this contribution we will take up the idea of classifying data for choropleth mapping considering the spatial context. The specific goal is to keep large gradients or similarities of neighboring regions after the data classification step. For evaluating the preservation of attribute value differences a

ICA European Symposium on CARTOGRAPHY VIENNA, 10–12 NOVEMBER 2015

Published in "Proceedings of the 1st ICA European Symposium on Cartography", edited by Georg Gartner and Haosheng Huang, EuroCarto 2015, 10-12 November 2015, Vienna, Austria

suitable measure is needed. As a consequence of the shortcomings of existing indices (like the Tabular Accuracy Index, TAI; or the Boundary Accuracy Index, BAI), a new **Edge Preservation Index (EPI)** is proposed here. It relates the original value difference to the corresponding class value difference – following the idea that there should be a linear correlation between these two differences. The EPI is designed in such a manner that its value and algebraic sign give an indication whether there is a loss, preservation or amplification of value differences of neighboring classes. Furthermore, it allows for the consideration of a flexible number of neighbors – either the preservation at all boundaries is considered, or the preservation of the edges with the p % largest value differences is considered only (i.e., an emphasis on large gradients can be quantified).

Applying this new index it is now possible to find an optimal data classification for the given purpose out of the many options that emerge by varying the number of classes and class boundaries. The figure below shows the results for the best preservation as well as for the largest amplification of neighboring value differences – in comparison to "standard" classification methods the effect of considering the spatial context becomes clearly obvious.

**Current work** is concerned with an optimization of the implementation of the complex and time consuming algorithm. Based on such an implementation user studies using various data sets will test the our core hypothesis which states that the classification results considering the spatial context will lead to an improved effectiveness for respective tasks.

**Figure 1**. Classification of population density values for German states (3 classes) applying different classification methods – from left to right: equidistant, quantiles, Jenks, EPI (best preservation), EPI (largest amplification)