What do we want to communicate? A few reflections on the choropleth map class ranges.

Elzbieta Bielecka [a,*], Beata Calka [a]

[a] *Military University of Technology, Faculty of Civil Engineering and Geodesy, elzbieta.bielecka@wat.edu.pl, beata.calka@wat.edu.pl*

* Corresponding author

Abstract:

The choropleth map as a cartographic representation of quantitative data has been known for almost 200 years, assuming the year 1826, when Ch. Dupin published a map showing by shading from black to white the distribution of illiteracy in France, as its beginning (Korycka-Skorupa, Pasławski 2017 after Funkhouser 1937). Choropleth mapping gained its popularity because it provides a simple but effective visual presentation of geographical data, in particular statistical data. It allows a mental representation of abstract data related to spatial units, facilitating their understanding. And that is why it can be considered as one of the important factors in scientific creativity. As noted by Maceachren and Kraak (1997) "the nature of maps and their use in science and society has been in the midst of a remarkable shift". This shift was prompted by the rapid development of IT and GI (geographic information) technologies on the one hand, and the growing public demand for geo-referenced information on the other. Over the past few decades, cartographers and GI scientists have devoted much work to refining the map's communicative function as a function that combines the cartographer's vision of the universe of discourse with the user's understanding of "what the map shows." Considering the choropleth mapping the problem of a map communicative function was dealt with, *inter alia*, by Calka (2018), Calka and Bielecka (2017), Andrienko et al (2011) or Kennedy (1994).

Our study aims to use exploratory data analysis in fine scale gridded population data classification, with particular emphasis on such a selection of classification method that underlines data reliability. We hypothesise that data classification based on in-depth analysis of the statistical distribution allows portrayal both compact and consistent ranges of choropleth classes and outlier values suggesting uncertainty in the data or the phenomenon under analysis. The study used two open global raster population datasets, LandScan and GHS-POP as well as Polish gridded population data derived by the Central Statistical Office. The quantification of global raster population (GRPD) data reliability was based on the general assumption that the Polish Population Grid is a reference dataset. The absolute disparity index (ADI) and the deviation rate index (DRI), were introduced and computed on cross-comparison analysis of all corresponding grid cells in considered datasets. They indicate how much the GRPD values deviate from the Polish Population Grid. The absolute disparity index (ADI), similar to absolute estimation error (AEE), measures the total difference in people counts in each i- spatial location (grid cell).

The presented results show how well GRPD data correspond to population distribution derived from statistical census. Although GRPD algorithms are tailored to match the geographical nature and economic conditions of each country and region, the reliability of population distribution in Poland differs. For densely populated regions, global data underestimates the number of people, while for thinly populated ones they rather overestimated.
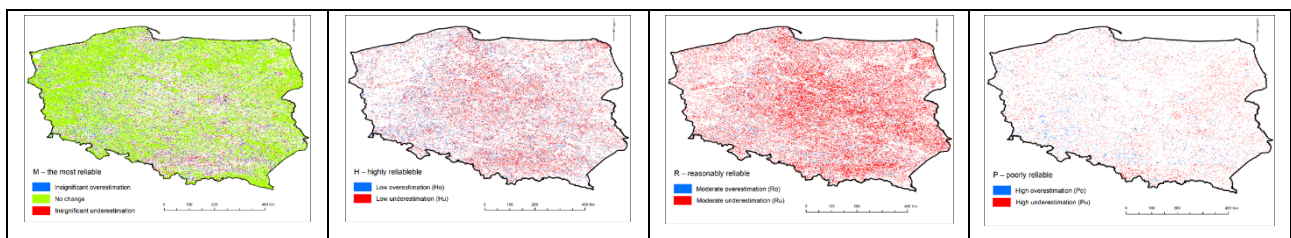
Figure 1. LandScan reliability classes – from the left - the most; high; reasonably; poorly reliable.

The most important conclusion of our research is the high reliability of Global Raster Population data for Poland. The most reliable data for LandScan amount to 56.9% and form irregular clusters, while highly reliable data cover 15.1% and are weakly clustered (Figure 1). Data of definitely insufficient quality cover only 4.2% with an evident trend to underestimation, especially in industrial and commercial zones of big cities. The underestimation results mainly from the

spatial pattern and size of Polish rural settlements, namely a big number of shadowed single households dispersed over agricultural areas and in the vicinity of forests.

Acknowledgements

References

Andrienko, G., Andrienko, N., and Savinov, A. 200. Choropleth maps: classification revisited. In Proceedings ICA (pp. 1209-1219).

Calka, B. 2018. Comparing continuity and compactness of choropleth map classes. *Geodesy and Cartography* 67, 21–34; https://doi.org/10.24425/118704

Calka, B. and Bielecka, E. 2019. Reliability Analysis of LandScan Gridded Population Data. The Case Study of Poland. *ISPRS Int. J. Geo-Inf*. 8, 222. https://doi.org/10.3390/ijgi8050222

Calka, B., Nowak Da Costa, J. and Bielecka, E. 2017. Fine scale population density data and its application in risk assessment. *Geomatics, Natural Hazards and Risk* 8(2), 1440-1455.

Funkhouser, H. G. 1937. Historical development of the graphical representation of statistical data. *Osiris* 3(1):269–405. Reprinted Brugge, Belgium: St. Catherine Press.

Kennedy, S. 1994. Unclassed choropleth maps revisited/Some guidelines for the construction of unclassed and classed choropleth maps. *Cartographica: The International Journal for Geographic Information and Geovisualization* 31(1), 16-25.

Korycka-Skorupa, J. and Pasławski, J. 2017. The beginnings of the choropleth presentation. *Polish Cartographical Review* 49 (4), 187–198.

Kuffer, M, Owusu M, Oliveira, L, Sliuzas, R. and van Rijn, F. T. 2022.he Missing Millions in Maps: Exploring Causes of Uncertainties in Global Gridded Population Datasets. *ISPRS International Journal of Geo-Information* 11(7):403.

Maceachren, A. M. and Kraak, M. 1997. Exploratory cartographic visualization: advancing the agenda. *Computers and Geosciences* 4(23), 335-343.