

A proposition to define boundaries based on the smoothness and the continuity of voluminous point data phenomena

Raphael Gonçalves de Campos^{a, *}, Caio Paiva^b, João Vitor Mesa Bravo^c, Silvana Philippi Camboim^d

^{a, b, d} Federal University of Parana, Graduate Program in Geodetic Sciences, raphageoc@gmail.com, anjospaiva@gmail.com, silvanacamboim@gmail.com

^c Federal University of Uberlandia, Institute of Geography, Graduate Program in Geography and Graduate Program in Agriculture and Geospatial Information, jmbravo@gmail.com

* Corresponding author

Keywords: Thematic Maps, Geospatial Big Data

Abstract:

Cartography is now facing with the emerging challenge of representing voluminous sets of geographic data, within the technological context of Big Data. Therefore, here we developed a method to optimize the process of delimiting boundaries while observing a voluminous point data set, based on the automatic quantification of smoothness and continuity concepts (MacEachren and DiBiase, 1991). According to MacEachren and DiBiase (1991), the smoothness and the continuity must be considered when choosing the technique of thematic cartographic representation. Smoothness describes the dependence or independence between the phenomenon values in the spatial units, with variation between smooth and abrupt. Continuity reflects the filling of geographic space by the phenomenon, which differs from the strict mathematical sense, even though the nature of the phenomenon is an essential factor in choosing the representation method. Both characteristics can vary for the same distribution depending on the scale at which they are represented, so a suitable representation method to be applied to distribution at a specific scale may not be ideal for a different scale.

Nowadays, exploring maps is a relatively simple task because tools allow navigation between different scales or view different portions of geographic space. Additionally, geospatial data is widely available, which allows for detailed representations at large scales. However, when observing the smoothness and the continuity of voluminous point data, it is hard to delimit boundaries, especially when the decision is made by a human. That means, human eyes have limits and humans have different cognitive backgrounds, which makes even those experienced cartographers to select different approaches for defining the boundaries associated with voluminous point data. Then, we understand this as a subjective process that could benefit from Artificial Intelligence as an iterative solution for choosing the best boundaries configuration.

Furthermore, representing the Big Data on maps requires new methods that consider its volume, speed in creation, and variety of types. Robinson et al. (2019) list several research opportunities related to the challenges posed by Big Data in the context of cartography. Among them is the need to combine computational methods and best practices in cartographic representation, such as computational solutions to automatically assess the phenomenon's characteristics to define the best representation method.

Therefore, here we propose two measures to characterize the smoothness and continuity of the spatial distributions of point phenomena aggregated into geographic units. The calculations of the two measures were implemented using the Python language to be executed as a script within the QGIS 3.x software. The difference in values of the phenomenon between geographical units that touch each other is calculated to quantify the smoothness. The average of all differences is calculated and normalized so that this value varies between 0 and 1, so it can be compared with other distributions.

The continuity property is measured considering the "holes" in the spatial distribution. For this, it is counted how many geographic units do not present occurrences of the phenomenon, and the result is divided by the total number of geographic units that make up the division of the territory used. As this parameter is a proportion, we can compare it with the same parameter of distributions of other phenomena.

To validate this methodology, we used six spatial distributions of different phenomena. For each one, we calculated the smoothness and continuity parameters and order them according to each parameter. These orders are confronted with a manual ordering performed by experienced cartographers through a visual analysis of an unclassified choropleth representation of the phenomenon and its 3D representation. In addition, we created an online form to collect information and responses from volunteers and a web page for viewing representations that allows comparison between two distributions at a time (Figure 1), available at the following address: http://www.labgeolivre.ufpr.br/teste_raphael/.

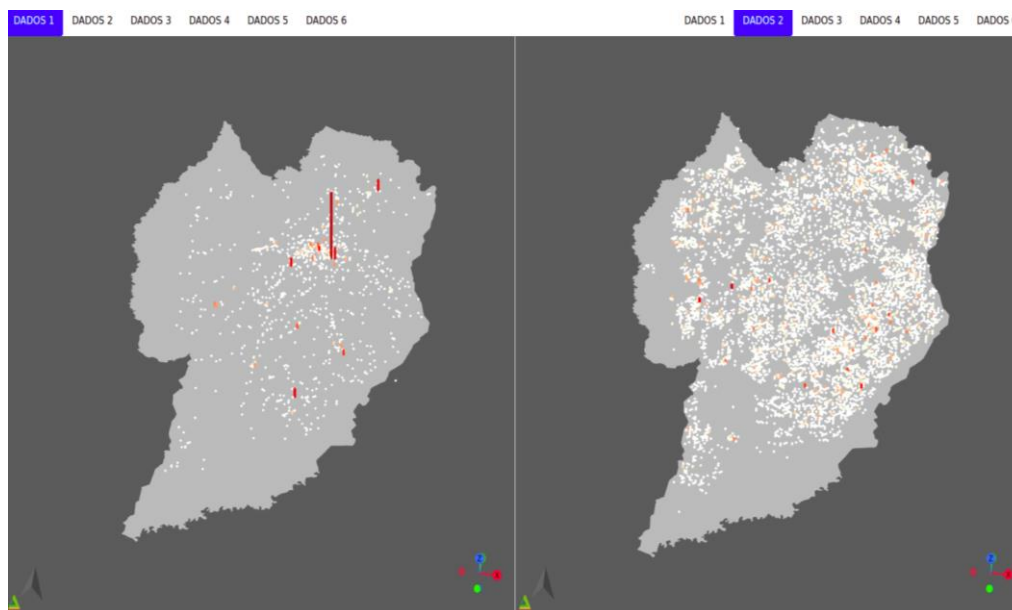


Figure 1. Web viewer for comparing two different spatial distributions. The perspective representations are for data A (left) and B (right) contained in Table 1.

This work is still in progress, but we have already obtained 24 responses from specialists with experience in cartography ranging from six to thirty-three years, with the results indicated in Table 1.

Position divergence	Smoothness						Continuity					
	A (%)	B (%)	C (%)	D (%)	E (%)	F (%)	A (%)	B (%)	C (%)	D (%)	E (%)	F (%)
-5	4,2	0	0	0	0	0	0	0	0	0	0	0
-4	25	0	0	0	0	0	0	0	0	4,2	0	4,2
-3	0	62,5	0	0	0	0	0	0	0	0	0	0
-2	4,2	12,5	0	8,3	0	0	0	0	0	16,7	0	4,2
-1	4,2	8,3	0	12,5	29,2	25	4,2	0	4,2	20,8	8,3	45,8
0	62,5	12,5	8,3	66,7	41,7	4,2	83,3	91,7	58,3	25	70,8	45,8
1	0	4,2	29,2	12,5	25	12,5	8,3	4,2	20,8	33,3	16,7	0
2	0	0	29,2	0	4,2	33,3	0	0	16,7	0	4,2	0
3	0	0	12,5	0	0	4,2	4,2	0	0	0	0	0
4	0	0	16,7	0	0	20,8	0	0	0	0	0	0
5	0	0	4,2	0	0	0	0	4,2	0	0	0	0

Table 1. Proportion of user responses with the degree of divergence between the perceived position and the one defined by the algorithm, for each data set. Larger values close to the line with zero position divergence means that there is more agreement between users and the algorithm.

While observing the values we found (Table 1), we conclude that there was a general agreement in the classification of variables by the specialists and by the algorithm. However, a more accurate statistical analysis is still needed, but this step is promising as a work in progress. Therefore, in the following steps we will propose ranking the possible visualizations based on characteristics we mentioned to perform perceptual clustering tests with users. Suppose the importance of smoothness and continuity is validated to create more adequate representations for each data set. In that case, it will be possible, among other applications, to train artificial intelligence algorithms so that the generation of geospatial Big Data visualization options is automated.

References:

1. MacEachren, AM and DiBiase, D. (1991). Animated maps of aggregate data: conceptual and practical problems. *Cartography and Geographic Information Systems*, 18(4), 221-229.
2. Robinson, AC, Demšar, U., Moore, AB, Buckley, A., Jiang, B., Field, K., Kraak, MJ., Cambodia, SP and Sluter, CR (2017). Geospatial big data and cartography: research challenges and opportunities for making maps that matter. *International Journal of Cartography*, 3(S1), 32-60.