www.czasopisma.pan.pl PAN www.journals.pan.pl

GEODESY AND CARTOGRAPHY
Vol. 67, No. 1, 2018, pp. 21–34

DOI: https://doi.org/10.24425/118704

Original research paper

Comparing continuity and compactness of choropleth map classes

Beata Całka

Military University of Technology
Faculty of Civil Engineering and Geodesy
2 gen. W. Urbanowicza St. 00-908 Warsaw 46, Poland
e-mail: beata.calka@wat.edu.pl, ORCID: http://orcid.org/0000-0002-7147-0849

Received: 12 March 2018 / Accepted: 9 May 2018

Abstract: Population density varies sharply from place to place on the whole territory of Poland. The largest number of people per 1 km^2 is 21,531, while uninhabited areas account for about 48% of the country. Such uneven, non-Gaussian distribution of the data causes some difficulty in choosing the classification method in geometric choropleth maps. A thorough evaluation of a geometric choropleth map of population data is not possible using only traditional indicators such as the Tabular Accuracy Index (*TAI*). That is why the aim of the article is to develop an innovative index based on distance analysis and neighbour analysis of grid cells. Two indexes have been suggested in this paper: the Spatial Distance Index (*SDI*) and the Spatial Contiguity Index (*SCI*). The paper discusses the use of five classification methods to evaluate choropleth maps of population data, like head-tail breaks, natural breaks, equal intervals, quantile, and geometrical intervals. A comprehensive assessment of such geometric choropleth maps is also done. The research was conducted for the whole territory of Poland, using data from the 2011 National Census of Population and Housing. Population data are presented in the 1km grid. The results of the analysis are shown on thematic maps. A compatibility of the choropleth maps with urban-rural typology of the OECD (Organisation for Economic Co-operation and Development) was also checked.

Keywords: data classification, choropleth map, neighbourhood analysis, spatial contiguity analysis, head-tail breaks method

1. Introduction

The choropleth map is one of the quantitative methods of cartographic presentation used for data in different types of measurement scales (ratio, or interval). The data are generally related to administrative units or grid cells. It seems that this cartographic method is easy to use, but it requires attention to many factors, such as the nature of the data, adoption of an appropriate reference unit, and an appropriate graphic form of the choropleth map (Müller, 1976; Medyńska-Gulij, 2011; Jiang, 2013; Cromley et al., 2015). Depending on the data distribution it is possible to use different methods of their classification. A method, on the one hand should, yields a map containing relevant information for the user and, on the other, should be easily readable (Bregt et al., 1990; Adrienko et al., 2001). It can therefore be said that these are very important aspects in drawing thematic choropleth maps.

There is a wide range of methods used for data classifications such as quantile, equal intervals, standard deviation, maximum breaks, optimal classification, or Jenks natural breaks method (Jenks and Coulson, 1963; Pasławski, 1984; Coulson, 1987; Cromley, 1995; Robinson et al., 1995). In general, the objective of classification is not only to group data in such a way so that similar objects (observations) are in the same classes, but the classes themselves should be dissimilar (Pasławski, 1984). Detailed reviews of classification methods can be found in Robinson et al. (1995), Murray and Shyy (2000), and Brewer and Pickle (2002). There are two groups of classification methods. The first group is based on data distribution analysis, like natural breaks or head-tail breaks methods. The other group of classification methods use statistical measures, e.g. the quantile method with the equal number of units, or the equal interval method, with each class having an equal range of values. For data with non-Gaussian distribution a classification method called head-tail breaks (Jiang, 2013) and the concentration-based classification method (Cromley et al., 2015) are suggested. The head-tail breaks method is based on dividing data into two parts around the arithmetic mean (Foss et al., 2013; Lin, 2013). The partitioning of the data continues for data above the mean, until data in the head are no longer heavy-tailed distributed. Based on a review of literature on the subject Table 1 presents the methods of data classification most often used, together with their main advantages and disadvantages.

Data classification determines map readability. That is why researchers work intensely on elaborating choropleth map assessing (Jenks and Caspall, 1971; Jenks, 1977; Pasławski, 2003; Armstrong et al., 2003; Medyńska-Gulij, 2010; Wei et al., 2017). Jenks introduced three assessment criteria to evaluate choropleth maps: the Overview Accuracy Index (*OAI*), based on area of entities, the Tabular Accuracy Index (*TAI*), based on sum of absolute deviations about class means, and the Boundary Accuracy Index (*BAI*), based on analysis of neighbourhood of entities (Jenks, 1971). Accuracy of a choropleth map can be measured also using the GVF – Goodness of Variance Fit (Coulson, 1967). Alternative classification accuracy indexes are: the Goodness of Deviation around the Median Fit (*GDMF*), in which squared deviations from the class and array median are calculated, as well as the Goodness of Absolute Deviation Fit (*GADF*), in which absolute deviations from the class and array median are calculated (Robinson et al, 1984). Most indexes intend to minimise data variation within particular classes and maximise differences between classes. Classification of geospatial data should consider two perspectives: statistical and geographical. This allows forming geographical regions with small data variation within (Bregt, 1990).

The analysis of spatial patterns is used by many authors in their research (Diggie, 1983; Li and Zhang, 2007; Calka et al., 2017). Diggie (1983) presented the definition of a spatial pattern as a set of locations, irregularly distributed within a region of interest, which have been generated by random mechanisms. The analysis is based on measuring the distance between each feature centroid and its nearest neighbour's centroid location. In case when the average distance is less than the average for a hypothetical random

Table 1. Classification methods with advantages and disadvantages

Classification method	Advantages	Disadvantages
Equal Intervals	The legend limits contain no missing values (or gaps). The results are easy to interpret for the map user. The class limits can be computed manually.	The class limits fail to consider how data are distributed. May produce classes with zero observations. Features with similar values may end up in different classes, exaggerating their differences. Wide range of values can end up in the same class, minimizing differences.
Quantile	Classes are not left empty nor do they have limited or excessive number of values. The class limits can be computed manually. Applicable to ordinal data.	The class limits fail to consider how data are distributed. Variable class width and/or gaps in legend. Distribution unequal when division of observations by the number of classes does not result in a whole number. Duplicate data values at class break requires manual adjustments. Value of an observation could be closer to value in a different class than its own. Features with similar values may end up in different classes, exaggerating their differences. Wide range of values can end up in the same class, minimizing differences.
Standard Deviation	Seeing which features are above or below an average value. The class limits consider how data are distributed. Produce constant class intervals.	Map doesn't show actual values, only how far values are from the mean. Very high or low values can skew the mean. Good only for normal distributed data.
Natural Breaks (Jenks)	Consider the distribution of the data. Good for the data not evenly distributed. Uses grouping based on break points visible graphically. Intuitive.	Class breaks are subjective. Breaks are not necessarily obvious. Difficulty to determine breaks increases with large data set. Can be difficult to choose optimum number of classes, especially if data are evenly distributed.
Geometrical Intervals	It works well on data that are not distributed normally. Good for data with significantly large ranges. Break points determined by rate of change in the data. This ensures that each class range has approximately the same number of values with each class and that the change between intervals is fairly consistent.	Not appropriate for data with small ranges or linear trends.
Head-tail Breaks	It works well on data that are not distributed normally. It is easy to choose optimum number of classes. Good for data with significantly large ranges.	Not appropriate for data with normal distribution.

distribution, the distribution of the features being analysed is considered clustered. Otherwise, the features are considered dispersed (Mitchell, 2005). The author intends to treat fields of reference ascribed to a given class as geographical objects occurring in a given location. Spatial patterns of choropleth maps has been studied by several cartographers (Olson, 1972; Müller, 1975; Bregt, 1990). Müller (1975) defined the aggregation index (*AG*) based on an analysis of adjacent units, and the boundary contrast index (*BI*) based on the amount of adjacent units.

Problems with the development of choropleth maps presenting population density, signalled earlier in the paper of Calka et al. (2017), was an inspiration to undertake the present research. A large number of small mapping units, a very high range of data (from 0 to more than 23,000 per 1 km^2) and non-Gaussian distribution cause difficulty in data classification, and thus the development of choropleth maps. Literature shows that the head-tail breaks method brings the best results for the classification of data with other than the normal distribution. However, this does not resolve the difficulty in visualization, because grid cells assigned to each class do not form dense clusters. The problem, however, could not be identified with the help of the most commonly used *TAI* index. Therefore, the purpose of the article is to develop novel indicators or to apply existing ones to assess geometric choropleth maps of population density, with an emphasis on continuity and compactness of grid cells in each class. Continuity refers to the fact that there are no gaps in spatial data (Goodchild, 1992), and the compactness is acknowledged as one of the most intriguing and important properties of a shape (Angel et al., 2010). Those indicators are based on distance analysis and neighbourhood analysis of grid cells. The *SCI* index has been developed by the author of this paper, while *SDI* is an innovative application of the existing ANN indicates (Average Nearest Neighbor) for the assessment of class division on choropleth mapps. Additionally, the correctness of rural and urban area portrayal in line with the OECD (Organisation for Economic Co-operation and Development) are analysed (Brezzi, 2011). The paper is structured as follow: section 2 describes methods, section 3 deals with study area and data used in analysis, section 4 demonstrates population data classification and its assessment with different methods using the indexes proposed in this paper. Finally, conclusions are drawn in section 5.

2. Methods

The spatial pattern of choropleth maps was assessed using two new indexes: the Spatial Distance Index (*SDI*) and the Spatial Contiguity Index (*SCI*), while readability was evaluated with the Tabular Accuracy Index (*TAI*). The Spatial Distance Index is based on existing index called Average Nearest Neighbor (*ANN*) existing in literature. It is calculated as the observed average distance divided by the expected average distance. The expected average distance is based on a hypothetical random distribution with the same number of features covering the same total area. The statistical significance of obtained results is tested with the use of p-value. The *SDI* ratio value lower than 1 means that the pattern of each class exhibits clustering. Otherwise, the trend is toward dispersion (Clark and Evans, 1954).

$$SDI = \frac{D_o}{D_E} \tag{1}$$

where: D_o is the observed mean distance between each feature and its nearest neighbour, and D_E is the expected mean distance for the features given in the random pattern.

Polygon contiguity (the Spatial Contiguity Index) is determined using a statistic called Polygon Neighbours. Spatial contiguity was defined by Lai et al. (2009) as a measure to check contiguity between areas (whether polygons share border). The neighbour relationship can loosely represent a measure of potential interaction between adjoining areas. Contiguity measures evaluate the characteristics of spatial units that are connected. The Polygon Neighbours tool available in ArcGIS 10.3.1 follows a hierarchical path to determine the type of neighbour and the statistic showing the number of shared borders. To evaluate the classification, standardized the Spatial Contiguity Index (*SCI*) based on polygon contiguity was developed by the author. The *SCI* Index was specified by using the following formula:

$$SCI = \frac{\sum_{n=1}^{n} \frac{s}{4 \cdot m}}{n} \tag{2}$$

where:
s – the number of neighbours in the same class determined with the Polygon Neighbours tool,
m – the number of observations in a given class,
n – the number of classes in the choropleth map,
4 – the number of adjacent neighbours for each grid cell.

The index takes vales from 0 to 1, with 0 meaning that that the objects in various choropleth map classes are characterized by a high dispersion (low proximity), and 1 meaning that the objects in the choropleth map for each class are characterized by a high concentration (high proximity).

Five population density maps of Poland were assessed using this spatial indexes. All maps were 1 km geometrical choropleth maps portraying the same set of population data in five classes. The number of classes was determined with the head-tail breaks method. Different classification method were employed to different maps: natural breaks, quantile, equal intervals, geometrical intervals and head-tail breaks methods.

The *TAI* index based on minimising the variation of data within the classes and maximizing differences between classes is computed according to the formula (Janks and Caspall, 1971):

$$TAI = 1 - \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} \left| x_{ji} - \bar{x}_j \right|}{\sum_{i=1}^{n} \left| x_i - \bar{x} \right|} \tag{3}$$

where:

x_i – observed values,
n – the total number of observations, the number of grid cells,
$\bar{x}$ – the mean of all values,
k – the number of classes,
x_{ji} – class j values,
n_j – the number of observations in class j,
$\bar{x}_j$ – the class mean for class j.

The Tabular Accuracy Index takes values form 0 to 1. The numerator is the sum of absolute deviation values to be found in a class and the denominator is the sum of absolute deviation of all values.

3. Area and data

The study area covers Poland, a central-eastern European country. With nearly 38.5 million people living in Poland, its area is 312.7 thousand km^2, making it the 6th largest country in Europe. Population data was taken from the Census of Population and Housing. The last census was carried out in 2011. The population data are available in a grid cell of 1 km^2, in Lambert Azimuthal Equal Area coordination system (LAEA). The people counts were attributed to grid cells based on address points (Calka et al., 2017).

Population data are characterized by high spatial differentiation and the logarithmic distribution (Figure 1). Nearly 40% of the Polish land is uninhabited, while 70% of the total population is located in 5% of the country's area.

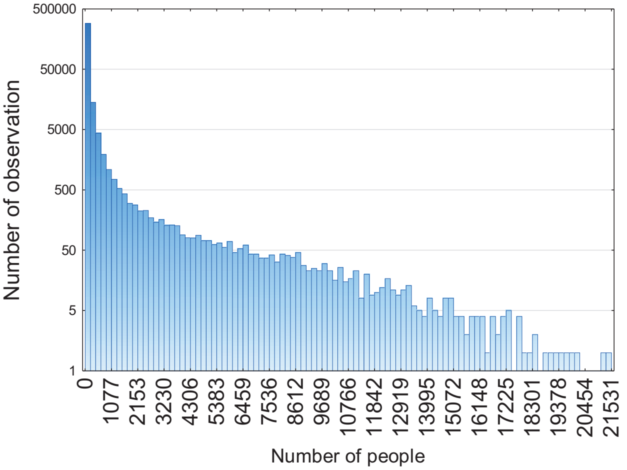

Fig. 1. Histogram of population density in Poland

The highest number of 33.014 people per 1 km^2 was attributed to the grid cell located in Szczecin. After examining this area this value was this value has been considered as outlying, and then replaced by a similar value to the data from four neighbouring grid cells equals 13.014 people per 1 km^2.

The descriptive statistics calculated for Polish population data are presented in Table 2. The average population per 1 km^2 is almost 122 people. The data do not have a normal distribution, being skewed to the right, with the skewness coefficient vale of 13.9. This is also confirmed by a low value of the first and the third quartile, and a low value of median. Standard deviation equalling to 654.1 shows that variability of population counts in grid cells is highly heterogeneous.

Table 2. Descriptive statistics of population data

Descriptive statistics	Population Data
Number of observations	315,857
Min	0
The first quartile (Q1)	0
Median	12.0
The third quartile (Q3)	64
Max	21,531
Sum	38,492,223
Mean	121.9
Standard deviation	654.1
Coefficient of skewness	13.9

4. Results and discussion

Choropleth maps together with histograms are presented in Figure 2. The head-tail map (Figure 2a), as relatively new in visualising population density, describes, in the author's opinion, more details than others.

Class 1 (Figure 2a) presents sparsely populated area with 122 inhabitants per 1 km^2. It well reflects rural areas according to the OECD typology. Moreover, it is the most numerous, comprising 85% of all values. The *SDI* index for class 1 is greater than 1 ($SDI = 1.56$), underlying the spatial dispersion of grid cells belonging to this class. The values of *SDI* index for the other classes are lower than 1, and are as follows: Class 2 – $SDI = 0.81$; Class 3 – $SDI = 0.48$; Class 4 – $SDI = 0.38$; Class 5 – $SDI = 0.33$ (Table 3). The low value of the *SDI* index confirms that the spatial pattern of classes 2, 3, 4, and 5 exhibits clustering. Class 2, ranging from 123 to 697 people in a grid cell, well illustrates small cities, with the share of population about 12.4% out of the total population. Classes 3, 4, and 5 are the smallest, with the number of observations below 2%. They represent urban areas, with large and medium sized cities.

Classes 1 in the natural breaks and the equal intervals methods are the biggest (grid cells of those classes constitute, respectively, 96%, and 99% of all objects). The values

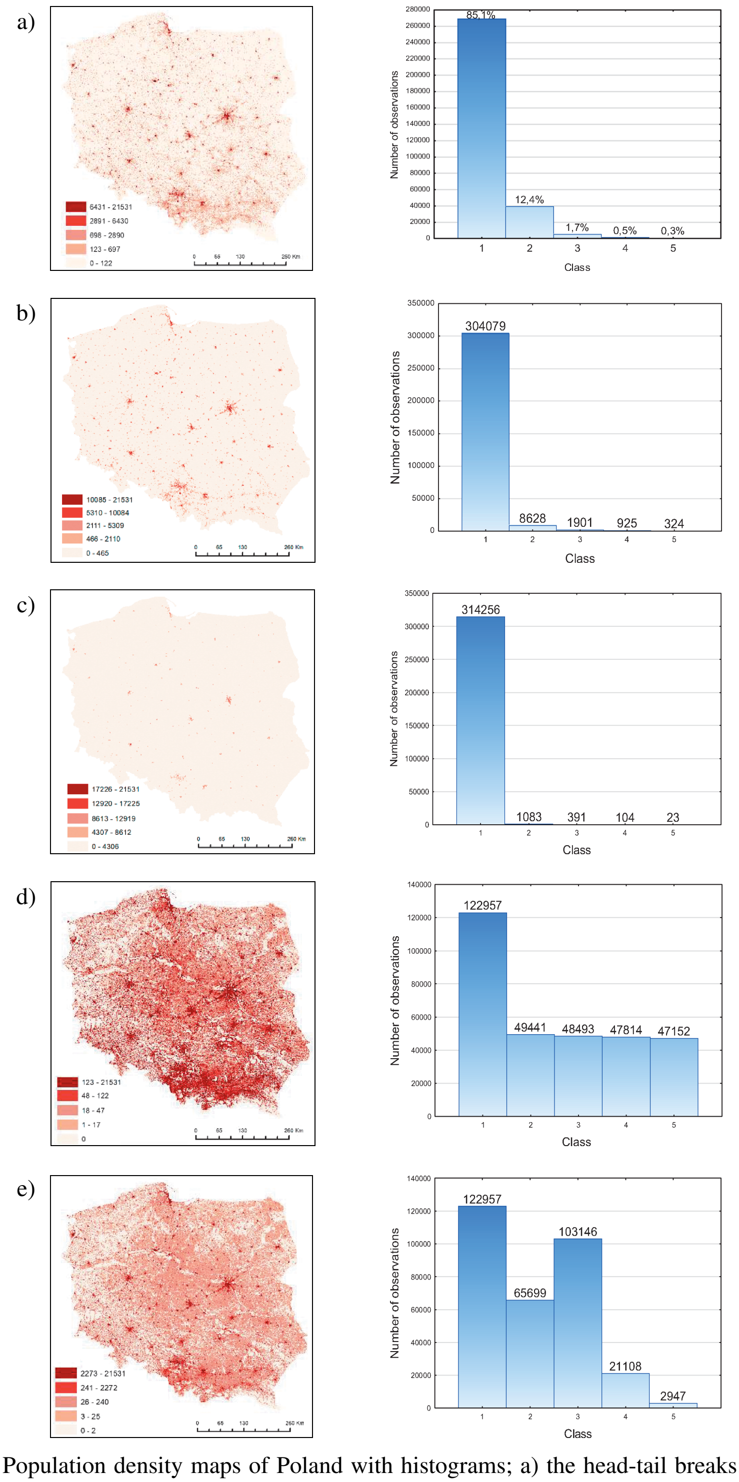

Fig. 2. Population density maps of Poland with histograms; a) the head-tail breaks; b) the natural breaks; c) the equal intervals; d) the quantile method; e) the geometrical intervals

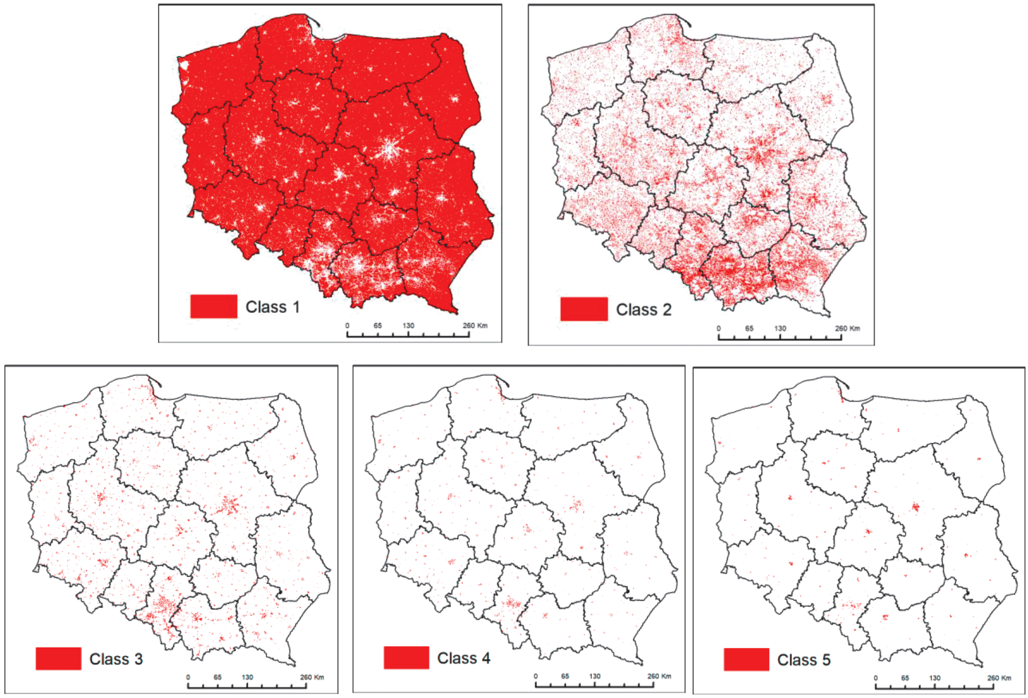

Fig. 3. Maps of population density in Poland developed using head-tail breaks classification, a division into classes 1 to 5.

Table 3. Values of the *SDI* index

Classification Method	*SDI* (Spatial Distance Index)				
	Class 1	Class 2	Class 3	Class 4	Class 5
Head-Tail Breaks Method	1.56	0.81	0.48	0.38	0.33
Natural Breaks Method	1.66	0.54	0.39	0.40	0.36
Equal Intervals Method	1.69	0.39	0.34	0.41	0.50
Quantile Method	1.10	0.91	0.90	0.89	0.84
Geometrical Intervals Method	1.10	0.95	1.05	0.67	0.32

of the *SDI* index (Table 3), with $SDI = 1.66$ and $SDI = 1.69$, indicate strong dispersion. Although the other classes for those methods of classification exhibit clustering, the clear dominance of the first class negatively affects the readability of maps. Both the natural breaks and the equal intervals methods do not highlight rural or urban areas clearly enough. The quantile method (Figure 2d) is characterized by a similar size of each class. Classes 2 to 5 comprise 15% of all population grid cells and are characterized by high clustering, as shown in Table 3. Class 1 constitutes 38% of all values. The *SDI* index for class 1 is greater than 1 ($SDI = 1.10$), so the trend is toward dispersion. Classes 1 to 4 represent rural areas. With the geometrical intervals method (Figure 2e) class 1 has the largest number of grids, with 39% of all values. This class, with the value of the *SDI* equals to 1.10, is strongly dispersed. Class 3, with the number of observations of 33%, has the *SDI* value of 1.05 and lacks clustering properties. The results of the p-value in Table 3 constitute 0.000, showing that there is less than 1% likelihood that this clustered or dispersed pattern could be random. It means that the obtained results are statistically significant.

Spatial contiguity is measured with the number of neighbours in the same class, hence the computation of the *SCI* index should be preceded by Neighbourhood Analysis (Table 4). Class 1 in all methods is the most numerous, and therefore shows the greatest spatial contiguity, which positively affects the readability of this class on the maps.

Table 4. Polygon Neighbours analysis results

Head-tail Breaks	1	2	3	4	5
1	305785	11169	61131	73959	31728
2	11169	1014	6447	7515	2538
3	61131	6447	47771	67064	20100
4	73959	7515	67064	150682	52291
5	31728	2538	20100	52291	81909

Natural Breaks	1	2	3	4	5
1	1182247	22622	2405	590	76
2	22622	8434	2372	880	197
3	2405	2372	1562	962	303
4	590	880	962	848	420
5	76	197	303	420	300

Equal Intervals	1	2	3	4	5
1	1244953	2821	710	140	17
2	2821	912	444	128	27
3	710	444	306	87	17
4	140	128	87	42	23
5	17	27	17	23	4

Quantile	1	2	3	4	5
1	305785	63942	43322	38996	31728
2	63942	45428	39165	29954	19089
3	43322	39165	48208	40851	22362
4	38996	29954	40851	47942	33477
5	31728	19089	22362	33477	81909

Geometrical Intervals	1	2	3	4	5
1	305785	79743	85878	11993	374
2	79743	79001	93673	9846	326
3	85878	93673	200158	31574	1204
4	11993	9846	31574	26696	4302
5	374	326	1204	4302	5582

The values of the *SCI* index and Tabular Accuracy Index (*TAI*) for different methods are presented in Table 5.

The results of the analysis show that the head-tail breaks method is characterised by the best selected class ranges, shown by a high rate of *TAI* (Table 5). At the same time, for this classification the *SCI* index is the lowest, indicating that the classes are

Table 5. *SCI* and *TAI* of classification methods

Classification method	*SCI*	*TAI*
Head-tail breaks method	0.29	0.70
Natural breaks	0.38	0.60
Quantile classification	0.37	0.34
Equal intervals	0.31	0.24
Geometrical intervals	0.44	0.65

of low spatial contiguity, which negatively affects the perception of maps by the user. This involves difficulty in identifying individual classes on the map. The best spatial contiguity and good *TAI* results are achieved using the geometrical intervals method, with a value of the Spatial Contiguity Index reaching 0.44, and with the *TAI* of 0.65. This means that the geometrical intervals method allows large diversity between classes, and small diversity inside classes, and they, at the same time, are relatively easily identifiable on the map by a user. The natural breaks method generally produces good results in terms of the *TAI* (0.60), and quite a high score for *SCI* (0.38). The ranges of classes result in quite a large diversity between classes and quite low diversity in individual classes. The equal interval and quantile methods both lead to classes with large within-class variations for population data sets and to a low value of *TAI*. This is due to the fact that in the quantile method class 5 has a very large value range (from 123 to 21,531), while in the equal intervals method class 1 is dominant, with 99% of all observations. At the same time those classification methods have quite a low or average value of the *SCI* index (0.31 and 0.37), due to the fairly low spatial contiguity of the classes.

5. Conclusion

The article presents an analysis of the application of five classification methods (natural breaks, equal intervals, geometrical intervals, quantile and head-tail breaks method) to present Polish population data obtained from the Main Statistical Office. These data have the high right-hand coefficient of skewness of 13.9, which negatively affects the choice of appropriate classification methods. The data are presented by means of geometric choropleth map. An evaluation of the methods is done in two ways: with the *TAI* index, and with two innovative indexes: the Spatial Distance Index (*SDI*) and Spatial Contiguity Index (*SCI*). The article shows how important a selection of data classification methods is in the process of developing a choropleth map. Depending on the selection of classification methods, a variety of population density maps can be developed. Therefore, it is so important to use not only statistics, but also spatial analysis methods to choose an appropriate one for data classification.

The head-tail breaks classification method shows its large utility in the process of mapping population density. Maps developed with this method clearly highlight rural areas in accordance with the OECD typology. In further studies on data classification it seems to be reasonable to exclude areas with a population value equal to zero from the analysis, and in consequence these areas will form a separate class. Classes developed by using this method are clustered, although they have a low *SCI* value, indicating low spatial contiguity. The geometrical intervals method is a useful tool for population data classification, with the highest value of *SCI* and a high quality confirmed by a high *TAI* value. Classes 4 and 5 present urban areas in accordance with the OECD rural-urban typology. When assessed with the *TAI* index, the quantile method and the equal intervals methods show the smallest utility in population data classification.

The results how that choosing a method of data classification is important because the contiguity and compactness of choropleth classes affect the readability of the map. Depending on the selection of class ranges in mapping population density, rural areas can be made prominent or the focus can be directed at urban areas with small, medium, and large town and cities. An evaluation of geometric choropleth maps for population data is not possible using only such traditional indicator as the Tabular Accuracy Index (*TAI*). In choropleth map analysis the pattern complexity of grid cells should also be taken into account. This paper shows that complex rating of geometric choropleth maps allows a selection of the best method of data population classification.

Acknowledgement

The study was conducted at the Faculty of Civil Engineering and Geodesy of Military Technical University in Warsaw, and financed from statutory research No 933/2018.

The author would sincerely like to extend her gratitude to Professor Elzbieta Bielecka for support and guidance.

References

Angel, S., Parent, J. and Civco, D.L. (2010). Ten compactness properties of circles: measuring shape in geography. *Canadian Geographer-Geographe Canadien*, 54 (4), 441–461.

Andrienko, G., Andrienko, N. and Savinov, A. (2001). Choropleth maps: classification revisited. Proceedings of ICA, 6–10 August 2001, Beijing, China, 2: 1209–1219.

Armstrong, M.P., Xiao, N., Bennett, D.A. (2003). Using genetic algorithms to create multicriteria class intervals for choropleth maps. *Annals, Association of American Geographers*, 93 (3), 595–623. DOI: 10.1111/1467-8306.9303005.

Bregt, A.K. and Wopereis, M.C.S. (1990). Comparison of complexity measures for choropleth maps. *The Cartographic Journal*, 27, 85–91.

Brewer, C.A. and Pickle, L. (2002). Evaluation of Methods for Classifying Epidemiological Data on Choropleth Maps in Series. *Annals of the Association of American Geographers*, 92 (4), 662–681.

Brezzi, M., Dijkstra L. and Ruiz, V. (2011). OECD Extended Regional Typology: The Economic Performance of Remote Rural Regions, OECD Regional Development Working Papers, 2011/06, OECD Publishing. 10.1787/5kg6z83tw7f4-en.

Calka, B., Bielecka, E. and Figurski, M. (2017). The spatial pattern of ASG-EUPOS sites. *Open Geosciences*, 9 (1), 613–621. DOI: 10.1515/geo-2017-0046.

Calka, B., Nowak Da Costa, J. and Bielecka E. (2017). Fine scale population density. *Geomatics, Natural Hazard and Risk*, 8 (2), 1440–1455. 10.1080/19475705.2017.1345792.

Clark, P.J. and Evans, F.C. (1954). Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology*, 35, 445–453. 10.2307/1931034.

Coulson, M.R.C. (1987). In the matter of class intervals for choropleth maps: with particular reference to the work of George F. Jenks. *Cartographica*, 24 (2), 16–39.

Coulson, M.R.C. (1987). In the matter of class intervals for choropleth maps: with particular reference to the work of George F. Jenks. *Cartographica*, 24 (2), 16–40.

Cromley, R.G. (1995). Classed versus Unclassed Choropleth Maps: A Question of How Many Classes. *The International Journal for Geographic Information and Geovisualization*, 32 (4), 15–27. DOI: 10.3138/J610-13NU-5537-0483.

Cromley, R.G., Zhang, S. and Vorotyntseva, N. (2015). A concentration-based approach to data classification for choropleth mapping. *International Journal of Geographical Information Science*, 29 (10), 1845–1863. 10.1080/13658816.2015.1058388.

Diggle, P.J. (1983). *Statistical analysis of spatial point patterns*. Academic Press, London.

Foss, S., Korshunov, D. and Zachary, S. (2013). *An Introduction to Heavy-Tailed and Subexponential Distributions*. Springer.

Goodchild, M. (1992). Geographical information science. *International Journal of Geographic Information Systems*, 6 (1), 31–45.

Jenks, G. and Coulson., M. (1963). Class Intervals for Statistical Maps. *International Yearbook of Cartography*, 3, 119–134.

Jenks, G. (1977). Optimal Data Classification for Choropleth Maps. Occasional paper No. 2, Department of Geography, University of Kansas.

Jenks, G.F. and Caspall, F.C. (1971). Error on choroplethic maps: Definition, measurement, reduction. *Annals of the Association of American Geographers*, 61, 217–244, DOI: 10.1111/j.1467-8306.1971.tb00779.x.

Jiang, B. (2013). Head/tail breaks: A new classification scheme for data with a heavy-tailed distribution. *The Professional Geographer*, 65 (3), 482–494.

Lai, P., So F., and Chan, K. (2009). *Spatial Epidemiological Approaches in Disease Mapping and Analysis*. Taylor & Francis Group.

Li, F. and Zhang, L. (2007). Comparison of point pattern analysis methods for classifying the spatial distributions of spruce-fir stands in the north-east USA. *Forestry: An International Journal of Forest Research*, 80 (3), 337–349. 10.1093/forestry/cpm010.

Lin, Y. (2013). A Comparison on Natural and Head/tail Breaks Involving Digital Elevation Models. Available on-line: http://www.diva-portal.org/smash/get/diva2:658963/FULLTEXT02.pdf.

Medyńska-Gulij, B. (2010). Map compiling, map reading and cartographic design in Pragmatic pyramid of thematic mapping. *Quaestiones Geographicae*, 29 (1), 5763. DOI: 10.2478/v10117-010-0006-5.

Medyńska-Gulij, B. (2011). *Kartografia i geowizualizacja*, PWN Warszawa.

Mitchell, A. (2005). *The ESRI Guide to GIS Analysis*, Volume 2. ESRI Press.

Murray, A.T. and Shyy, T.K. (2000). Integrating attribute and space characteristics in choropleth display and spatial data mining. *International Journal of Geographical Information Science*, 14, 649–667.

Müller, J.C. (1975). Definition, measurement and comparison of map attribute in choropleth mapping. In Proceedings of the Association of American Geographers, 7, 160–164.

Müller, J.C. (1976). Objective and Subjective Comparison in Choroplethic Mapping. *The Cartographic Journal*, 13, 156–166.

Olson, J. (1972). Autocorrelation as a measure of map complexity. In Proceedings of the American Congress on Surveying and Mapping, 111–119.

Pasławski, J. (1984). In search of a general method of class selection for choropleth maps. *International Yearbook of Cartography*, 34, 159–69.

Pasławski, J. (2003). *Jak opracować kartogram*. Uniwersytet Warszawski, Warszawa.

Robinson, A.H., Morrison, J., Muehrcke, P.C., Kimerling, A. and Guptill, S. (1995). *Elements of cartography*. New York, USA,John Wiley & Sons.

Robinson, A.H., Sale, R.D., Morrison, J.L. and Muehrcke, P.C. (1984). *Elements of cartography*. New York, John Wiley and Sons.

Wei, R., Tong, D., and Phillips, J.M. (2017). An integrated classification scheme for mapping estimates and errors of estimation from the American Community Survey. *Computers, Environment and Urban Systems*, 63, 95–103. DOI: 10.1016/j.compenvurbsys.2016.04.003.