

Choropleth maps and diagram maps in atlas of cartographic presentation methods

Abstract

The online *Atlas kartograficznych metod prezentacji* [Atlas of cartographic presentation methods, hereinafter the *Atlas*] is a research project being carried out at the Department of Cartography of the University of Warsaw. The aim of the project is to systematize knowledge about the use of cartographic presentation methods.

This study discusses selected issues related to two of the five presentation methods analysed in the project, viz. the choropleth map and the diagram map. A rational application of two quite commonly-used presentation methods leads to a number of problems. These problems are most easily visible during attempts to program its implementation in the web-based Atlas and are largely due to the difficulties with drawing a clear boundary between what is a good and a bad map. For this reason, the system operator's skill and eye for the graphics of semi-automated visualisation seem to be of key importance.

Keywords

Cartographic presentation methods • choropleth map • diagram map
• class selection • graduated symbol scaling

© University of Warsaw – Faculty of Geography and Regional Studies

Jacek Paślowski¹
Jolanta Korycka-Skorupa²
Tomasz Nowacki³
Tomasz Opach⁴

¹Department of Cartography,
Faculty of Geography and Regional Studies,
University of Warsaw
e-mail: jpaslaws@uw.edu.pl

²Department of Cartography,
Faculty of Geography and Regional Studies,
University of Warsaw
e-mail: j.skorupa@uw.edu.pl

³Department of Cartography,
Faculty of Geography and Regional Studies,
University of Warsaw
e-mail: t.nowacki@uw.edu.pl

⁴Department of Cartography,
Faculty of Geography and Regional Studies,
University of Warsaw
e-mail: topach@uw.edu.pl

Received: 24 May 2012
Accepted: 3 July 2012

Introduction

The online *Atlas kartograficznych metod prezentacji* [Atlas of cartographic presentation methods, hereinafter the *Atlas*] is a research project being carried out at the Department of Cartography of the University of Warsaw. The reason for such a publication was the rapidly growing popularity of GIS systems which allow nearly every user to make a map easily and quickly. The aim of the project is to systematise and disseminate knowledge about the use of cartographic presentation methods, including their classification and the formalisation of selected quantitative methods. Formalisation¹ is understood as a process of translating a set of data into a cartographic language and is an attempt at defining the principles and conditions of a proper cartographic presentation, since the clearly formulated assumptions underpinning the individual methods of cartographic presentation would allow almost everyone to draw maps correctly. In

the project, we did not use any of the available GIS software in an attempt to discuss all stages of map-making step-by-step, in a way that helps to understand the very nature of cartographic presentation.

The *Atlas* consists of three parts. The first includes papers on the history and theoretical foundations of cartographic presentation methods. The second discusses the step-by-step procedure showing how a thematic map should be elaborated. An important aspect of this part is a brief description of five forms of presentation, commonly known as cartographic methods. Knowledge of the attributes of each of these forms is a precondition for good map making. The third part offers several dozen examples of adding content to the pattern described in part two. In effect, maps are developed in a way that allows for their modification in order to obtain new (and correct) variants of the basic map.

This study discusses selected issues related to two of the five presentation methods analysed in the project, viz. the choropleth map and the diagram map.

¹ **formalisation** [Latin: *formalis* 'referring to shape, form'], log. procedure intended to translate the content of one theory (system) from a natural, non-formalised language into a formalised one, e.g. that of logic (www.encyklopedia.pwn.pl)

The map-maker can opt for “continuous presentation”, which means that the selected graphic feature (e.g. size in the case of diagram maps, or shade or pattern in the case of choropleth maps) will be an individual attribute of every sign or area. Nevertheless, more frequently a discrete approach is used, in which uniform graphic signs correspond to a specific amount of data. In other words, the author of the map needs to aggregate quantitative data, a procedure that statisticians would refer to as making a graphic array.

At first sight, continuous approaches seem more interesting and precise, allowing the map user to read the map in more detail. In reality however, a greater degree of accuracy can be an obstacle which is the reason why discrete approaches, being in themselves certain forms of generalisation, are usually used in quantitative cartographic presentation methods (Fig. 1). As a result, the map is easier to read and interpret, e.g. class identification facilitates visual regionalisation through mnemonic aggregation of areas belonging to the same class.

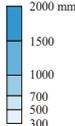
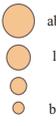
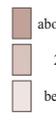
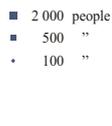
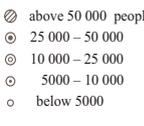
ISOPLETH MAP	DIAGRAM MAP	CHOROPLETH MAP	MULTI SIZE DOT MAP	QUANTITATIVE SYMBOL MAP
Annual precipitation	Passenger traffic in airports	Potato crops	Population distribution	Cities population
				

Fig 1. Class selection is carried out in all basic cartographic presentation methods, including the dot method

The goal of another aspect of cartographic research associated with class identification is to expand the functionality of GIS and similar systems. In this type of software, visualisation of quantitative data (mostly using the classic methods of cartographic presentation i.e. choropleth map, diagram map or isopleth maps) plays an important role, in the same way as for data acquisition and formalisation of methodological procedures. This is why it is important to attempt to operationalise theoretical deliberations through the development of algorithms which allow quantitative data to be correctly (in the cartographic sense) visualised. This issue becomes even more important if we look at the functionality concerning visualisation of quantitative data offered by computer applications. Their evaluation has definitely been negative (Bajer & Korycka-Skorupa 2008, Dmochowski & Paslawski 2000, Spallek 2007) and the list of problems to be resolved is far from short. These problems include a lack of clear terminology, lack of a comprehensive approach (e.g. the isopleth method is often used as a tool accompanying interpolation and not as one of cartographic presentation methods), or the often questionable ways of map legend construction which make it difficult for users to interpret the map rather than facilitating analysis of its content. One of these problems seems also to be the question of how GIS users will cope with class identification.

Practically all major GIS software producers envisage the use of quantitative cartographic presentation methods. Systems which engage in a dialogue with users give them the freedom to

choose the presentation method and modify its parameters, for example by supporting the development of thematic maps using different graphic arrays. However, the fact that they do not restrict the users' freedom of action gives rise to some concerns. What prompts do these map-making tools offer? As a rule, these are terse comments and class selection tools whose algorithms are sometimes difficult to decipher. A person developing the map who is not a cartographer cannot even begin to understand the meaning of his or her actions; whatever they do is done automatically, or intuitively at best. For instance a user of ArcGIS ArcMap, when developing a choropleth map (graduated colours) or a diagram map (graduated symbols) can choose one from the several methods of class selection using a dialog window where a given data set is presented in the form of a histogram (Fig. 2). Thanks to this, the user has some freedom as far as the stemplot structure is concerned, but will only be able to enjoy it when they are aware of the differences between individual class identification methods, their attributes and unique features.

An unprepared user will probably resort to an intuitive approach, acting on the assumption that there are no major differences between the results of using different class selection methods, apart from visual differences. This problem is pertinent not only to class selection but also – and predominantly – to the selection of a proper presentation method, since the differences be-

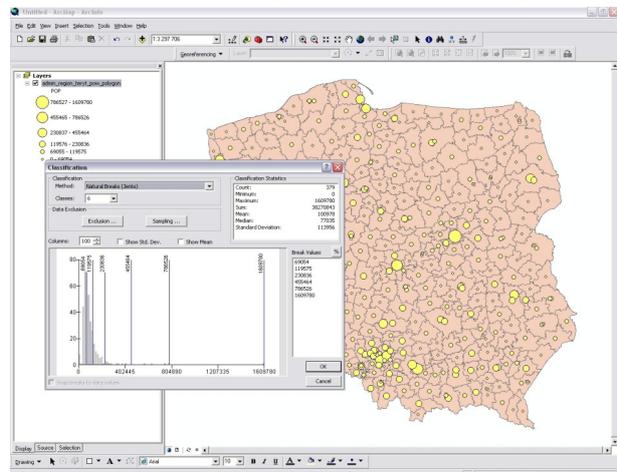


Fig 2. Dialog window of ArcGIS ArcMap 9.2 application, supporting class selection while developing a graduated diagram or choropleth map. The statistical distribution of data in the set is shown in the form of a histogram

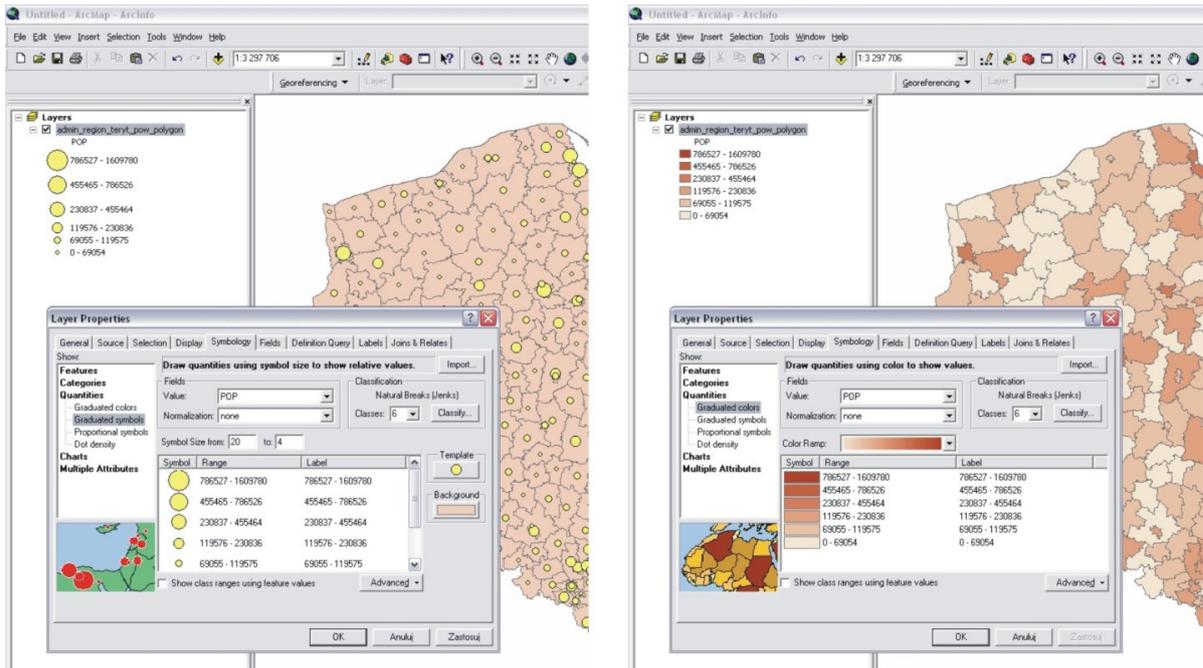


Fig 3. In the “Symbology” tab of the ArcGIS ArcMap 9.2 dialog window, the graduated diagram map (graduated symbols) (left) was placed next to the choropleth map (graduated colours) (right). Such a layout of the dialog window suggests that these methods can be used interchangeably

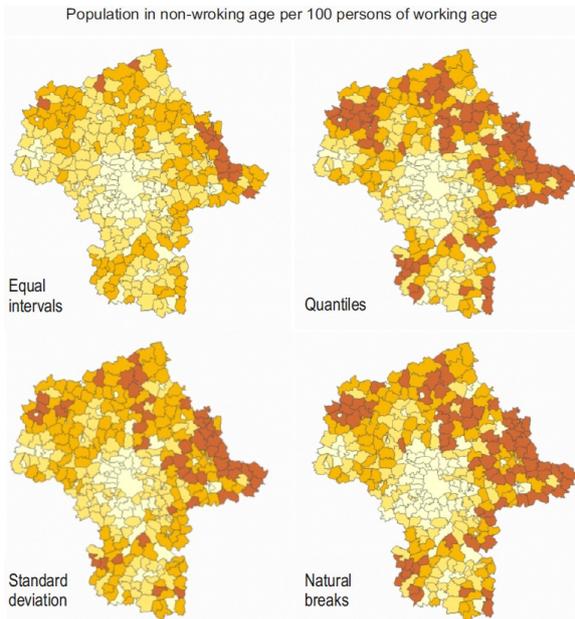


Fig 4. The method of class selection can affect map perception and interpretation. The legends shown above can serve as examples; they have been elaborated using the same data illustrated in Fig. 5

tween individual presentation methods are known only to those with some basic knowledge of cartography. Unprepared users are most likely to choose between the methods based on their aesthetic values. For this reason the poorly-conceived layout of the dialog window can give rise to some concerns, since the juxtaposition of the choropleth map and the diagram map might suggest that these methods are interchangeable (Fig. 3). The layout

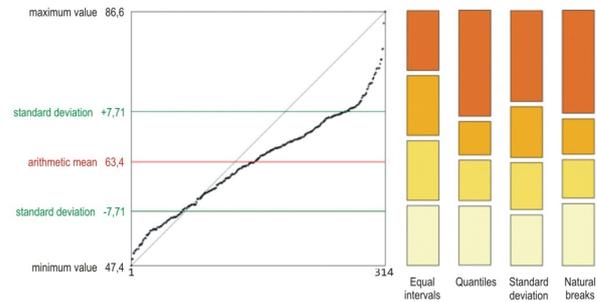


Fig 5. Graphic illustration of data (see graphic array) and rectangles of the map legend illustrating changes in the range of consecutive classes, depending on the way they have been selected

of the dialog window should underline the differences between individual methods and emphasise their specific features, e.g. use of relative data in the choropleth map (Cuff & Bieri 1979, Pravda 1983, Robinson et al. 1995, Kraak & Ormeling 1998, Tomaszewska 2009) and above all, it should be based on a clear and comprehensive classification of cartographic presentation methods.

In the following parts, we will discuss some of the many topics associated with the correct and informed drawing of thematic maps, viz. the purpose of the different methods of class selection in choropleth maps and issues related to the size and location of diagrams on the map. One of the overriding goals of the *Atlas* is to explain the basic principles underpinning the use of cartographic presentation methods.

When embarking on the work on the *Atlas*, we were not fully aware of the range of the problems that will need to be resolved. The ability to display relevant information suitably on the map

is a token of genuine cartographic mastery. Proper presentation of significant information pertains to different, both topical and graphic, aspects of the cartographic methodology.

Some problems related to class selection on choropleth maps

Class selection in choropleth maps and diagram maps is frequently underestimated, as is proven by incorrectly-drawn maps one can find, mostly online. Part of the responsibility for this can be attributed to the lack of relevant instructions provided to users by the GIS software.

This is illustrated by the four apparently different maps shown in Fig. 4. They have been drawn on the basis of the same data but the differences in the actual representations are a result of the use of different methods of class selection (Fig. 5).

Several research issues can be identified in connection with the structure and perception of choropleth maps – the nature of the data, space units of reference or the structuring of the map legend (Pasiński 1982, 1991, 2003, Tomaszewska 2009). One of the more interesting – and more difficult – issues is the formalisation of procedures associated with the construction of stemplots. This issue seems particularly relevant when the topic being considered is the interrelationship between the features of the statistical distribution of the set being mapped and the optimisation of the class limits. The question of whether identification of the data distribution will allow (e.g. by analysing the value distribution) the unequivocal identification of the class selection method remains open. This is probably not the case, but analysis can help preclude certain class selection methods. This issue is expounded by an analysis of four statistical distributions shown in Fig. 6.

The data referring to spatial units in the sphere of socio-economic phenomena usually make up sets that statisticians describe as skewed. In such a set, the vast majority of data shows minor differences, and their small number represents much bigger values. Fig. 6 A is a good illustration of such a set and the four boxes on the diagram illustrate variability in the ranges of four classes that arise as a result of using different ways to select them.

By using the **equal ranges** method, we will always obtain (regardless of the data distribution or range) bars of equal heights, i.e. classes of equal ranges. The choropleth map whose classes were identified in this manner is nearly uniform, as over 98% municipalities (308 of 314) were included in the first, or lowest class. To remedy such a situation, classes can be delineated to ensure that **the same number** of municipalities is included in each class (Quantiles). This will produce a choropleth map with three classes in the lower range of small-extent data, but the resultant map will be distinctly differentiated. Although there will not be many differences in the value of the classes, we will have much more information about the differentiation of a given phenomenon than in the case of the first map.

These are the two basic concepts concerning the identification of equal-range and equal-number classes. When we look at the class ranges in the remaining distributions (Fig. 6 B, C and D), we will see that the more skewed the distribution the more varied the range of the classes, if the number of the classes remains the same. Differentiating the class ranges may make it difficult to in-

terpret the map correctly. In case of equal ranges, the differences in the values between the classes are the same – they resemble steps made up of boxes of equal height - which helps us to properly evaluate the quantitative relationships between the areas included in different classes. In the case of different class ranges however, estimating quantitative relationships is more difficult. At this point, the role of the map legend should be highlighted. For a situation of variable class ranges in particular, showing the differentiation graphically and with the support of the description of the values means this becomes information that is extremely useful for the map reader.

Standard deviation is the class boundary should only be used with nearly normal distributions as these are not frequently encountered in spatial socio-economic research. In this situation, individual classes on the map are areas with values remaining within one or two standard deviations. It is worth noting that in the normal distribution, 34% of data usually falls within the first range of standard deviation. These possibilities for class selection are envisaged in the GIS software user manuals, although as a rule the explanations given are not sufficient.

The last of the proposed options for functionality are “**natural intervals**”. This refers to more pronounced differences within the data which are usually easily visible on the graphic array or histogram. Naturally, the adoption of such a principle produces classes with varying ranges of classes and numbers of space units. One advantage of this method is highlighting similar areas, i.e. areas with data that has relatively similar values, but a weakness is that the differentiation of the range of the classes makes it difficult to correctly determine the quantitative relationships.

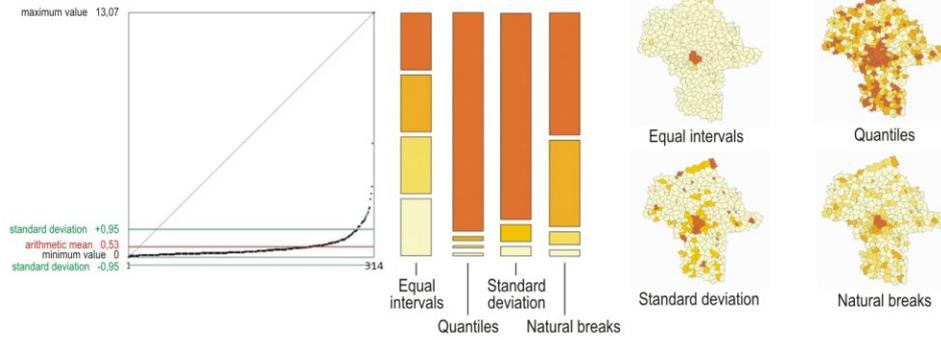
In the *Atlas*, the users can choose between data and analyse their statistical distribution. They can decide on the number of classes and on the method of class selection. The user of the *Atlas* can choose between two and six classes, depending on the degree of detail they want in the presentation. The number of classes is limited as an unlimited number of classes in GIS programs is inefficient. The system of prompts and comments will inform users whether a given presentation makes sense or how it could be read. Simple propositions for colour scales are also provided in the *Atlas*, including the popular BrewerColor system (www.ColorBrewer.org).

Some issues related to graduated symbol scaling

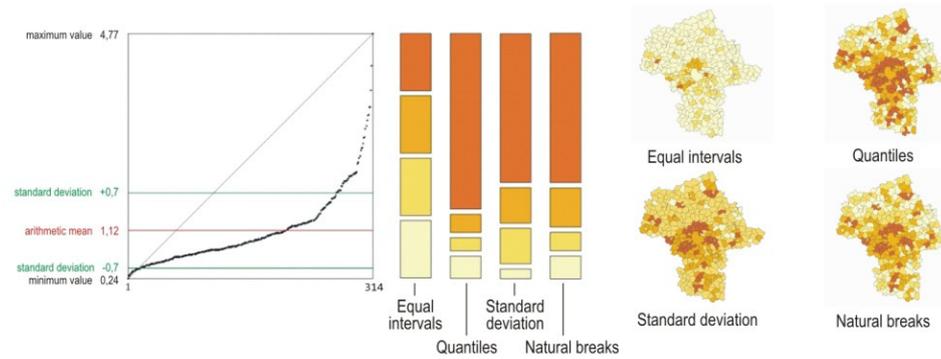
Graduated symbol scaling, or the correct presentation of size, is an important issue in the creation of diagram maps. When we decide to present data in a continuous manner, or one in which every statistical value has its representative on the map in the form of a diagram with a specific size that is proportional to the statistical value, we need to ensure that diagrams are properly scaled. In this case, the notion “properly” refers to two significant aspects of cartographic presentation, viz.:

a. The size of the diagrams should be directly proportion to the statistical values. A diagram that is half as big as another corresponds to a value half as big. We do not recommend using the arbitrary scale as described by Salishchev (1998), which involves changing the size of the sign using a pre-adopted rule (e.g. the

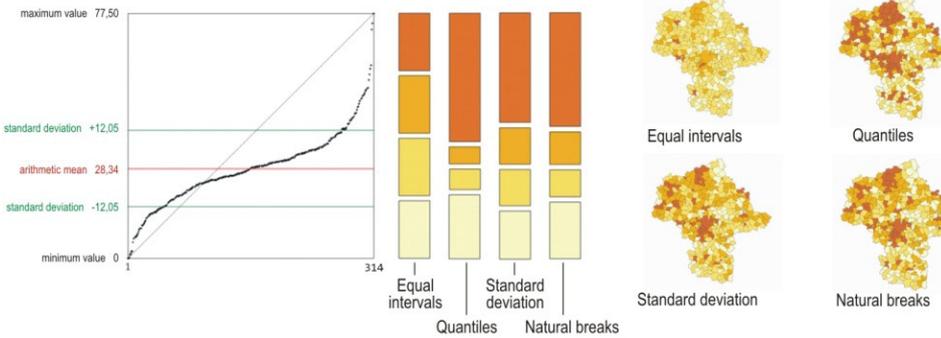
A. Number of commuters to one departing person (clearly skewed distribution)



B. Number of registered persons to one checked out person (skewed distribution)



C. Water usage per 1 inhabitant (nearly normal distribution)



D. Population in non-working age per 1000 of persons in working age (nearly rectangular distribution)

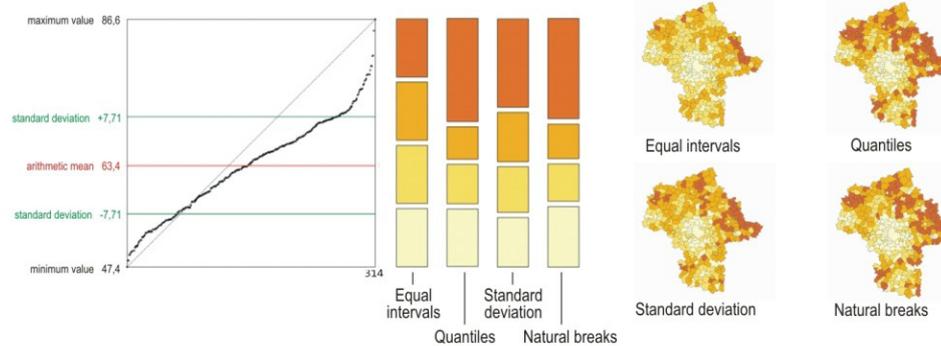


Fig 6. Four different data distributions (A – clearly skewed distribution, B – skewed distribution, C – nearly normal distribution, D – nearly rectangular distribution), which produce different graphic illustrations of four basic types of identification of classes in choropleth maps

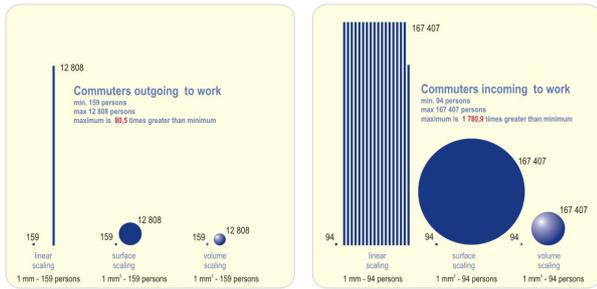


Fig 7. With narrow data ranges (A – the ratio of the extreme values in this example is 1:80), it is possible to use bars, but with broader ranges (B – in the illustration, the 1:1781 ratio), it is better to use volume scaling, or cube in this particular example

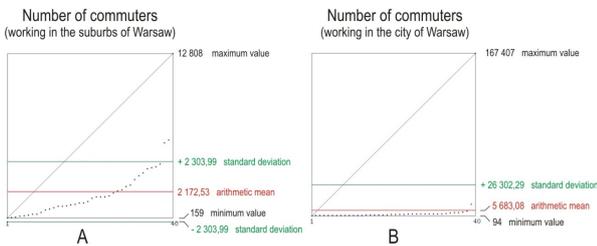


Fig 8. Class selection is carried out in all basic cartographic presentation methods, including the dot method

value of the element) showing only such a change of value, without retaining the proportion rule. In particular, bar scaling using a logarithmic scale has a misleading effect (Buczowski & Garstka 1991).

b. Diagrams should be placed in such a way to allow their spatial reference (point, surface area) to be correctly read and to allow the statistical value to be estimated.

The overall correlation between the data range and the selection of the diagram is well-known. With narrow ranges of data sets, bars can be used to allow for an optimal estimation of quantitative data. When the range of the data is broader (which is typically the case), geometric shapes are used whose surface areas illustrate statistical values. With broad data ranges, we use volume scaling in the form of spheres or cubes.

Maps showing outgoing and incoming commuters by district (powiat) can serve as good illustrations of the scaling method being adjusted to the data range. This is well illustrated in Fig. 7 which shows the largest and the smallest diagrams that were developed using different data ranges for different scaling methods.

The distribution of data in the set, i.e. whether the number of average or high values (Fig. 8A) or whether low values with several outliers prevail (Fig. 8B), affects the selection of the diagram scaling method and thereby the formalisation of the diagram map method.

Depending on the data distribution and diagram scaling method, the discernibility of diagrams in the map may be adequate or poor (Figs. 9 and 10). Naturally, the lower or upper threshold may always be modified for the size of the diagram depending on preferences as far as the visual effect is concerned, but this aspect is difficult to optimise or formalise. The map author's intuition seems to be of crucial significance in this respect.



Fig 9. In case of the distribution shown in Fig. 8A, linear scaling, i.e. bars, produces good results. In order to enhance diagram discernibility in surface and volume scaling, all diagrams should be increased proportionally

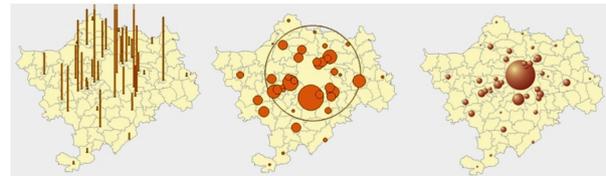


Fig 10. In case of a skewed distribution as in Fig. 8B for example, linear scaling is inappropriate. Manipulating the area can allow the quantitative relations to be represented quite efficiently, if the diagram's maximum values are not filled with colour. Although manipulating volume produces the most desirable graphic effect, assessing values on the basis of diagram sizes is much more difficult

In the online *Atlas of cartographic presentation methods* (Korycka-Skorupa et al. 2010), in addition to selecting the scaling method users will be able to select the lower or upper thresholds for the size of the diagram (Fig. 11). Such a solution offers only partial formalisation, involving the indication and parametrisation of attributes which are relevant for a given presentation method i.e. the diagram map in this particular case. Such an approach can hardly be regarded as complete automation since the role of the cartographer, as the operator of the system who has to select the parameter values based on his or her intuition, cannot be ignored. In the case of the *Atlas*, formalisation will be manifested through alerting the users of if they are taking actions that, in our opinion, do not make sense and will not make the map any more readable.

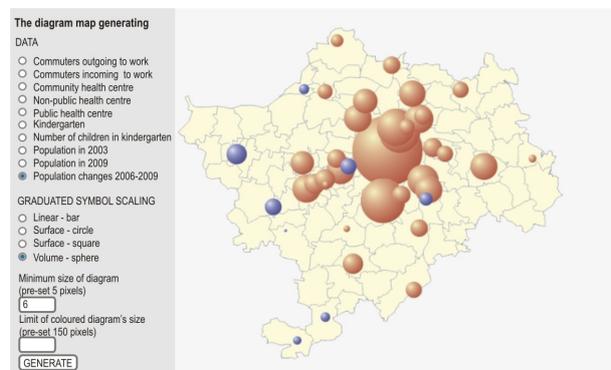


Fig 11. Working interface of the online *Atlas of cartographic presentation methods*. The web-based application makes it possible to select the method of graduated symbol scaling for example, or the lower/upper threshold of the diagram size

In terms of preparing a web-based cartographic application for automated visualisation of quantitative data using the diagram map method, other problems can arise that are associated with the formalisation of diagram scaling. When developing the relevant algorithms, it is necessary to adopt hard criteria describing such issues as:

Adequate discernibility of diagrams. Diagrams should not be too small and should ensure that the differences between them are noticeable. This also applies to high values in volume scaling, where the increase in diagram size and the increase in value is relatively small. Popping units are a special example here. This problem is usually associated with large agglomerations (e.g. Warsaw) and neighbouring areas with relatively low values of a given index, such as population density. In such a case, different symbology may be used for divergent values in order to maintain the discernibility of units with low values of a given phenomenon. In case of the *Atlas*, such an arrangement was used for linear scaling, where the bars reaching beyond the map are represented as a dotted line (Fig. 10 A). Naturally, such a graphic operation is selected automatically by the system (web-based application) when the values go beyond a pre-defined threshold. In some cases, when one statistical value is conspicuous in a given data set the use of double scaling is necessary.

Appropriate location of diagrams. The visual attribution of diagrams should be unambiguous and accurate, although it is difficult to develop an algorithm producing satisfactory results where the reference units are scattered and have a complicated distribution. Some problems can also arise with bar diagram maps where the highest bars need to be shifted so that they do not reach beyond the map. In general terms, the distribution of bars should be intelligent i.e. it should be dependent on a number of factors.

Avoiding the overlapping of diagrams. Naturally, diagrams may slightly overlap, but this should be done in a way that al-

lows their size to be estimated. Although various solutions are used, devising an appropriate algorithm is difficult, even for major software manufacturers (Bajer & Korycka-Skorupa 2008). In case of the algorithm used in the *Atlas*, smaller diagrams are always displayed "above" bigger ones. In addition, in area scaling large diagrams are not coloured in but are displayed only as a colour ring.

Summary

A rational application of two quite commonly used presentation methods leads to a number of problems. In both these cases, a study of the statistical distribution should be undertaken as the first step and this is done best using value charts or histograms. User manuals for GIS software do not seem to address this stage of preparing statistical maps with sufficient attention.

Whilst identification of classes in choropleth maps is easy to program, this is not the case with diagram maps.

Although the distribution of data in diagram maps does have major influence on the type of scaling used in the map, it is difficult to formalise this relationship. These problems are most readily visible in attempts to program its implementation in the web-based *Atlas*. This is largely due to the difficulty of defining what is a good and a bad map. For this reason, the system operator's skill and eye for the graphics of a semi-automated visualisation seem to be of key importance

This research was funded from the science budget in the years 2010-2012 as research project N N526 073838 titled: Formalisation of cartographic presentation of quantitative data and its implementation in the online Atlas kartograficznych metod prezentacji.

English translation: Dorota Szmajda

References

- Bajer, A & Korycka-Skorupa, J 2008, 'Kartodiagram w wybranych programach komputerowych' [Diagram Map in Selected Computer Software]. *Polski Przegląd Kartograficzny*, vol. 40, no. 3, pp. 247–266. English summary.
- Buczkowski, K & Garstka, A 1991, 'Pragmatyczne aspekty skalowania diagramów', [Pragmatic Aspects of Graduated Symbol Scaling]. *Polski Przegląd Kartograficzny*, vol. 23, no. 3, pp. 79-85. English summary.
- Cuff, DJ & Bieri, KR 1979, 'Ratios and Absolute Amounts Conveyed by a Stepped Statistical Surface'. *The American Cartographer*, vol. 6, no. 2, pp. 157–168.
- Dmochowski, M & Paślowski, J 2000, 'Kartogram w wybranych programach komputerowych' [Choropleth Map in Selected Computer Programs]. *Polski Przegląd Kartograficzny*, vol. 32, no. 4, pp. 288–301. English summary.
- Encyklopedia PWN. Available from: <<http://www.encyklopedia.pwn.pl>>. [7 May 2012].
- Jenks, GF 1973, 'Visual Integration in Thematic Mapping: Fact or Fiction?' *International Yearbook of Cartography*, vol. 13, pp. 27–35.
- Korycka-Skorupa, J & Nowacki, T & Opach, T & Paślowski, J 2010, 'Internetowy „Atlas metod kartograficznych” czyli krok ku formalizacji kartograficznych metod prezentacji' [Internet Atlas of cartographic methods – a step towards formalising cartographic methods of presentation]. "Prace i Studia Kartograficzne", vol. 3 *Polska kartografia w dobie przemian metodycznych i technologicznych* [Polish cartography in the Era of Methodological and Technological Change], Warsaw: Polish Geographical Society – Section of Cartography, pp. 187–190, English summary.
- Kraak, MJ & Ormeling, F 1998, *Kartografia. Wizualizacja danych przestrzennych* [Cartography. Visualisation of Spatial Data]. Warsaw. Wydawnictwo Naukowe PWN (In Polish)
- Paślowski, J 1982, 'O konstrukcji objaśnień kartogramów i map izoliniowych' [On the Construction of Choropleth and Isoleth Map Explanations]. *Polski Przegląd Kartograficzny*, vol. 14, no. 3,

- pp. 114–122. English summary.
- Pasławski, J 1991, 'Dane statystyczne a kartogram' [Statistical Data and the Choropleth Presentation]. *Polski Przegląd Kartograficzny*, vol. 23, no. 3, pp. 73–79. English summary.
- Pasławski, J 2003, *Jak opracować kartogram* [How to Make a Choropleth Map]. 2nd ed. Warsaw: University of Warsaw, Faculty of Geography and Regional Studies. English Summary.
- Pravda, J 1983, Zákon kartogramu a problem vyjadrovania nerozlohovych charakteristik. *Geogr. Časopis*, vol. 35, no. 2, pp. 136–159.
- Robinson, AH & Morrison, JL & Muehrcke, PC & Kimerling, AJ & Guptill, SC 1995, *Elements of Cartography*. 6th ed. New York. John Wiley & Sons Inc.
- Salishchev, KA 1998, *Kartografia ogólna* [General cartography]. Warsaw. Wydawnictwo Naukowe PWN
- Spallek, W 2007, 'Funkcje kartograficzne MapViewer' [Cartographic Functions of MapViewer]. *Kartograficzne programy komputerowe* [Computer Software for Cartography], eds. W. Żyszkowska, W. Spallek. University of Wrocław, pp. 123–134. English Summary.
- Tomaszewska, M 2009, 'Sposoby ujęcia danych a poprawność map opracowanych metodą kartogramu' [Data Character and Correctness of Choropleth Maps]. *Polski Przegląd Kartograficzny*, vol. 41, no. 3, pp. 209–22. English summary.