A cartographic approach of the process of map symbolization on gvSIG software

Vassilios Krassanakis
Dipl. Eng., PhD Cand. NTUA
krasvas@mail.ntua.gr
http://users.ntua.gr/krasvas

Vasilis Mitropoulos
Dipl. Eng., PhD Cand. NTUA
mitrovas@hotmail.com

Byron Nakos
Professor NTUA
bnakos@central.ntua.gr
http://users.ntua.gr/bnakos

National Technical University of Athens
School of Rural and Surveying Engineering, Cartography Laboratory
9, Heroon Polytechniou Str., Zographos, GR-157 80, Greece

Summary

Maps are very important in many studies related to different fields including geography, history, education etc. The process of map design is an important function of a GIS software. The gvSIG project is one of the open source geospatial tools that can be used in order to visualize different types of spatial phenomena related to geographic space. The development of each open source project is mainly based on the contribution of its user's community. The aim of the present paper is to summarize the fundamental principles related to map symbolization and to provide some critical indications for the improvement of map symbolization in gvSIG.

Keywords

gvSIG, map symbolization, map design, visual variables, data classification methods, thematic mapping

1. Introduction

The development and the use of a GIS software has a great influence in different scientific fields and studies. These studies are mainly related to phenomena which are connected with the geographic space. Generally, GIS systems are differentiated from other information systems as they are focusing on the spatial entities and their relationships (Maguire, 1991). Map can be considered as an abstract image of spatial entities of the three dimensional environment, or simply a "picture" of the ground as it is referred in Keates (1996). It is obvious that maps, as tools for data visualization, are very crucial in different aspects of each common GIS software. Therefore, the process of cartographic symbolization in GIS software must be in agreement with the principles of map design.

Over the last decades, the GIS software receives wide acceptance as basic tool for the creation of maps. Additionally, the open source geospatial software seems to have a great activity in the field (Steiniger & Bocher, 2009) and its applications are well accepted in many studies (e.g. Steiniger & Hey, 2009; Migliaccio et al., 2010; Neteler et al., 2012).

The gvSIG project is one of the open source GIS software that is able to be used for the implementation of map design. The gvSIG software has a wide acceptance in different studies, such as forest fire monitoring (Altobelli et al., 2009), hydrological analysis (Dietrich et al., 2010) etc. Moreover, gvSIG can be used as a training tool (Arago et al., 2012). Despite this fact, the need of improvement is considered appropriate as different studies (e.g. Dietrich et al., 2010; Scholz et al., 2011) report the existence of bugs in gvSIG's functionality. Although gvSIG provides enriched tools for the process of data symbolization, there are also several additional functions that can expand its use.

The aim of the present paper is to summarize the principles of map symbolization process in order to indicate guidelines for the improvement of gvSIG software's existing tools. Possible adds are reported after a general description of the functionality and the abilities that the gvSIG' s tools provide for the production of different map symbols.

2. Visual variables of cartographic design

A map can be either a static or a dynamic (animation) representation of the geographic space and related phenomena. Considering the standard static map, symbols can be classified as points, lines or areas in order to represent different spatial dimensions of phenomena. Additionally, the fundamental list of different types of spatial phenomena consists of points, lines areas, 2.5-D and true 3-D phenomena (Slockum et al., 2005). The perceived differences between map symbols are based on the application of "visual variables" (Slockum et al., 2005). Visual variables were introduced by Bertin (1967/1983) and they are the basic design elements in map symbolization. Bertin' s list includes the variables of position (x,y location), size, color value, texture (spacing), color hue, orientation and shape. An example of the application of visual variables for different type of symbols (points, lines and areas) is presented in Figure 1. Visual variables are able to depict different types of data scales including nominal (e.g. point symbols that represent locations of hotels, shopping centers, restaurants etc), ordered (e.g. small, medium, large values of a phenomenon) and interval/ratio (e.g. different values of population density) differences. In Table 1 the effectiveness of each visual variable, in the visualization of different data scales, is summarized. As it is obvious from table 1, each variable has different abilities in map visualization. More extended descriptions about the function of visual variables and their role in map design are cited in many studies (e.g. Green, 1998; Carpendale, 2003; Deed et al., 2011).

	nominal	ordered	interval/ratio
position (x,y)	effective	effective	effective
size		effective	effective
color value		effective	marginally effective
texture (spacing)	effective	marginally effective	
color hue	effective	marginally effective	
orientation	effective		
shape	marginally effective		

Legend: effective (dark); marginally effective (grey)

Table 1. The effectiveness of Bertin' s visual variables.

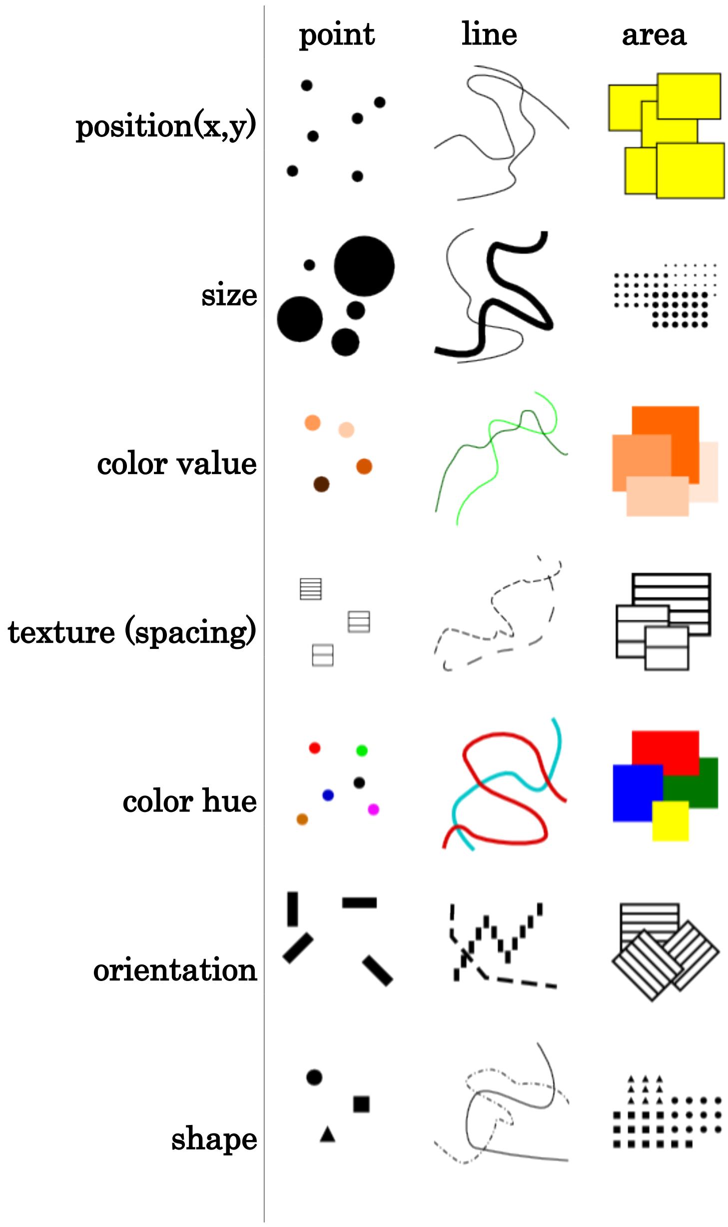

Figure 1. The basic visual variables in cartographic symbolization (after Bertin 1967/83 modified).

3. Data classification

In many cases, cartographers use the visual variable of color value in order to depict quantitative data on a thematic map. A significant problem arises here on how humans are perceiving the visual variable of color value. Specifically, the human eye can effectively discriminate no more than five to eight different shades of a color. Thus, cartographers ought to classify the numeric values before depict them on a map. There are many methods for classifying data values. Classification methods are well summarized in many studies (e.g. Kraak and Ormenling, 2003; Slockum et al., 2005). A brief description of the most common ones are presented below.

- **The equal intervals classification**: According to this method, the range between the lowest and the highest value of the numerical data is divided into sections of equal width. The number of sections is predefined and corresponds to the number of classes the cartographer wants to generate. The technique is suitable when the data are uniformly distributed.

- **The mean standard deviation classification**: The method is appropriate for numerical data that show an approximately normal distribution. The distribution's parameters, that are the mean value and the standard deviation of the numerical data, are used for the definition of the limits of the generated classes.
- **The quantiles classification**: In this method, the cartographer defines the number of generated classes. The observations of the numeric data are sorted in ascending or descending order. The classes are derived by placing the same number of observations in each class. Quantiles are suitable for ordered data.
- **The areal equal intervals classification**: In this method, the area of the surface referred to the data is used as the classification measure. The numerical values are classified in ascending or descending order. The limits are determined in a way that each class to include equivalent in area observations.
- **The maximum breaks classification**: In this method, the numerical values are sorted in ascending order. Then the difference between each neighboring value is calculated. The largest value differences define the class limits. The cartographer chooses the "n" largest value differences in order to generate "n+1" number of classes.
- **The natural breaks classification**: In this method, the cartographer uses logical and subjective criteria in order to group the numerical data. Two principles that are usually taken into account are to minimize value differences between data within the same class and, at the same time, to emphasize the differences between the classes.

Regardless of the classification method the cartographer should take into consideration: **(a)** All data values (including the minimum and maximum) must be included in the process. **(b)** Each data value must belong to one class. **(c)** Each class must contain at least one data value. **(d)** The classes should be homogeneous. In other words, the data values must be divided in reasonably equal groups of observations. **(e)** The number of classes must be great enough to be useful, but limited to make the map easily readable. For example, suppose that a cartographer has to represent a numeric data phenomenon (e.g. population) for the 14 regions of Central Greece. If he/she divides the numeric data into ten groups, the produced map is ineffective in cartographic terms. As shown in Figure 2a, it is difficult for the reader to discriminate the ten color values used to depict the ten classes. On the other hand, it is also ineffective to generate two classes (Figure 2b). As shown in Figure 2c, a cartographically effective map could be produced if he/she chose to classify the numeric data into five groups. **(f)** The purpose and the audience of the map should be taken into account.

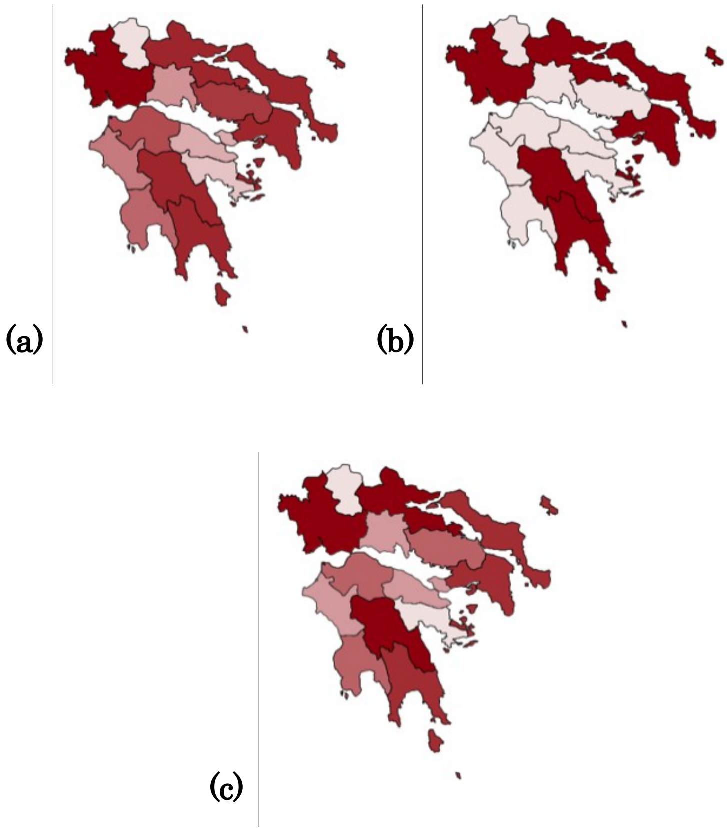

Figure 2. Example of classifying the same numeric data into different number of classes.

4. Map symbolization in gvSIG

The gvSIG software provides a large number of data analysis and representation tools. Regarding to the data classification process, the gvSIG software supports the Equal, the Natural and the Quantile Intervals methods. These techniques satisfy a wide range of classification needs. Nevertheless, it would be useful to add some commonly used techniques in the platform (e.g. the maximum breaks or the areal equal intervals), in order to provide a wider range of preprocessing tools. In addition, it is important for the user to know the theoretical background of the method he/she applies. Thus, it would more practical if a short description of the applied method is accompanying each tool.

When a user implements the data classification tools in order to classify and depict areal phenomena that are differentiated in ordered or interval/ratio scale (for example the population of the seven regions of Peloponnese, as shown in Figure 3), the gvSIG platform visualizes automatically the outcomes. It produces a choropleth map like the one shown in Figure 3a. This map is cartographically ineffective, as long as the color hue is not the proper visual variable to represent interval differentiated data. However, the platform provides the ability to depict the data using the visual variable of color value, so as to create an effective map, like the one shown in Figure 3b. It would be more effective, in similar cases the gvSIG platform to utilize automatically (by "default") the color value as visual variable. A very helpful site that could be easily linked in the gvSIG software is the following: http://colorbrewer2.org/. It contains advices of how to use the color in order to represent classified data during the map making process. A full description of this on line toolbox can be found in

Harrower and Brewer (2003).

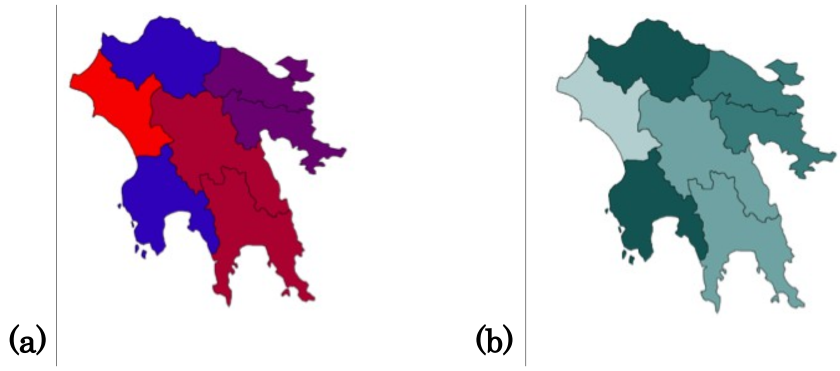

Figure 3. Ineffective (a) and effective (b) representation of interval differentiated data.

The same drawback occurs when the gvSIG' s user wants to visualize spatial phenomena with point nature, differentiated in ordered or interval/ratio scale. The platform visualizes by default the outcomes using the color hue as visual variable. As mentioned above, it would be more effective if the color value was automatically used. Another visual variable that is considered the most suitable for the depiction of similar phenomena is the size of the symbol. The gvSIG platform enables the user to change manually the point symbol's size. However, there are specific methods, such as the method of square roots or its extension the Flannery' s method (Flannery, 1971), that compute the size of the symbol according to the data numeric values. It would be useful if such techniques were included and executed automatically in the gvSIG software.

5. Conclusion

The gvSIG software provides the opportunity of map design as it consists of a variety of options for the creation of different symbols. The present study reports several issues about the theoretical framework that can be used for the implementation of map symbolization. More specifically, the main principles for the process of map design and symbolization are cited. Thence, some suggestions which can be implemented in gvSIG platform in order to improve the process of map symbolization are provided. The fact that gvSIG software has an open source approach is very important as it can be improved with the contribution of different users from different fields.

Acknowledgment

Authors would like to thank Victoria Agazzi (Community Manager of gvSIG project) and Manuel Madrid (Product and Testing Manager of gvSIG project) for the fruitful discussions and the excellent collaboration.

Bibliographic References

Altobelli A., Sgambati A., Bader F., Fior G., Magajna B., Ferrazzo L., Braut R., Urrutia P., Ganis P., Orlando S. (2009), "Application of gvSIG in a study related to forest fire monitoring", 5th International gvSIG Conference, Valencia.

Arago P., Kasie T., Grandio A., Clark S., Huerta J. (2012), "gvSIG education for Disaster Risk Management within an international coorporation project", 8th International gvSIG Conference, Valencia

Bertin J. C. (1967/1983), "Semiology of Graphics: Diagrams Networks Maps, Wisconsin Press, (French Edition 1967).

Carpendale M. S. T. (2003), "Considering Visual Variables as a Basis for Information Visualization", Technical Report 2001-693-16, Department of Computer Science, University of Calgary.

Deeb R., Ooms K., De Maeyer P. (2011), "Typography in the Eyes of Bertin, Gender and Expertise Variation", Cartographic Journal, 49(2), 176-185.

Dietrich S., Hildah M., David F. (2010), "Evaluation of gvSIG and SEXTANTE Tools for Hydrological Analysis", 6th International gvSIG Conference, Valencia.

Flannery J. J. (1971), "The relative effectiveness of some common graduated point symbols in the presentation of quantitative data", Cartographica, 8(2), 96-109.

Green M. (1998), "Toward a Perceptual Science of Multidimensional Data Visualization: Bertin and Beyond", ERGO/GERO Human Factors Science.

Harrower M., Brewer C. A. (2003), "ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps", The Cartographic Journal, 40(1), 27-37.

Keates J. S. (1996), "Understanding maps", 2nd ed. Harlow, Essex, England: Longman.

Kraak M. J., Ormeling F. (2003), " Cartogrpahy: Visualization of Geospatial Data" (2nd ed.), Prentice Hall, London.

Maguire D. J. (1991), "An overview and definition of GIS". In: Maguire D. J., Goodchild M. F. , Rhind D.W. (eds.) Geographical Information Systems: principles and applications, Longman, London, 1, 9-20.

Migliaccio F., Carrion D., Zambrano C. (2010), "A comparison between free/open-source and commercial geospatial software tools, based on a case study", FOSS4G 2010, Barcelona.

Neteler M., Bowman M. H., Landa M., Metz M. (2012), "GRASS GIS: A multi-purpose open source GIS", Environmental Modelling & Software, 31, 124-130.

Schloz J., Weiser P., Abdalla A. (2011), "Teaching FOSS-GIS at Technical University of Vienna – gvSIG vs. QGIS", Proceedings of leGIO- WORKSHOP: GIS-education in a changing academic environment, Leuven.

Slocum T. A., McMaster R. B., Kessler F. C., Howard H. H. (2005), "Thematic Cartography and Geographic Visualization", 2nd ed. Pearson Prentice Hall.

Steiniger S., Bocher E. (2009), "An overview on current free and open source desktop GIS developments", International Journal of Geographical Information Science, 23(10), 1345-1370.

Steiniger S., Hay G. F. (2009), "Free and open source geographic information tools for landscape ecology", Ecological Informatics, 4, 183-195.