Making a GIS Intelligent: CommonGIS Project View

Gennady L. Andrienko, Natalia V. Andrienko
GMD - German National Research Center for Information Technology
Schloss Birlinghoven, Sankt-Augustin, 53754 Germany
WWW: http://allanon.gmd.de/and/
E-mail: gennady.andrienko@gmd.de
Tel: +49-2241-142329 Fax: +49-2241-142242

Submission to AGILE Conference (Rome, April 1999)

Keywords: map visualization design, interactive manipulation, GIS

1. Requirements to an Intelligent GIS

The project **CommonGIS** [1] started in October 1998 with the goal to develop an intelligent system supporting end users in visual analysis of thematic data in a geographic context. The primary focus of the project is automated visualization of spatially referenced data. The partners in the project are 3 research institutes (GMD[1], IGD[2], and JRC[3]), 2 commercial software providers (PGS[4] and Dialogis[5]), national center for geoinformation (CNIG[6]), and international GIS association (GISIG[7]). The project is funded by an ESPRIT programme (thematic call "Information Access and Interfaces", project 28983), the duration is 30 months. The basis of the project is formed by 2 existing systems for knowledge-based visualization design - **Descartes** [2,3] developed in GMD[8] and **Vizard** [4] developed in IGD. Essentials of these systems should be further elaborated and combined with the commercial Internet GIS, **Lava/Magma** [5] developed by PGS.

Our notion of intelligent, user-friendly GIS implies that this should be not only a huge set of powerful instruments potentially useful for data analysis. An intelligent GIS should assist the user in the analysis, and this presupposes the following capabilities:

1. The capability to understand user's information-seeking goals.

---

[1] GMD - German National Research Center for Information Technology, SET - System Design Institute, http://set.gmd.de/

[2] IGD – Fraunhofer Institute for Computer Graphics, http://www.igd.fhg.de/

[3] JRC - Joint Research Center of European Commission, Institute for Systems, Informatics and Safety, http://www.jrc.org/isis/

[4] PGS - Professional Geo Systems, http://www.pgs.nl/

[5] Dialogis Software and Services GmbH, http://www.dialogis.com/

[6] CNIG - Portuguese National Center for Geographic Information, http://www.cnig.pt/

[7] GISIG – European GIS Association, http://gisig.ima.ge.cnr.it/

[8] An online Internet demo is available at the URL http://allanon.gmd.de/and/java/iris/

2. The capability to select and visualize appropriate data in a way productive for achieving these goals.
3. The capability to support the user's analytical activity with the use of the generated presentations.

None of the available GIS possesses such features. A number of research prototype systems capable of automated generation of graphical presentations are known. The system **APT** developed by J.Mackinlay [6] can encode data variables, according to their types and cardinality, by J.Bertin's [7] visual variables and construct graphical displays by combining these visual variables. This approach was adapted by F.Zhan and B.Buttenfield [8] for the selection of an appropriate cartographic presentation method for one spatially referenced data variable. Later V.Jung developed the system **Vizard** capable of automated mapping of several independent variables. **Vizard** accounts not only for data characteristics but also for user's objectives. These objectives are indicated in terms of predefined *generic tasks* (or *user's goals)*: *lookup* (value estimation), *locate* (find a specific value on a map), *compare* (compare values for several objects), see *distribution* (find patterns, trends, clusters, anomalies, etc. in spatial distribution of values).

It should be noted that the possible user's analytical goals considered in Vizard and some other systems for data visualization are either rather primitive (lookup, locate, compare) or rather abstract (see distribution). It would be more convenient for the user to communicate with a system about her/his goals, on the one hand, referring to higher level tasks, on the other hand, using familiar domain notions instead of generic formulations. This is only possible if the system has a model of the domain of the user's activities.

Descartes is a system for automated visualization working over a kind of domain model called "notion base". The notion base enumerates notions of an application domain essential for interpretation of data to visualize. Important relationships among the notions are indicated. For example, the system can "understand" that female and male are parts of the whole population, and that female from 0 to 14 years, female from 15 to 64 years, and female of the age 65 years and more make together female population in total. This kind of knowledge allows grounded selection of particular presentation techniques such as maps with pie charts or segmented bars. The same knowledge could be aptly used to communicate with the user about her/his objectives in domain-specific terms rather than on an abstract level. For example, the system might ask the user whether s/he wishes to compare proportions of age groups 0-14 years, 15-64 years, and 65 years and more in female and male population. Though the present **Descartes** does not account for user's goals, its architecture suggests a good potential for further advancement that is supposed to be done within the **CommonGIS** project.

The concept of use of graphical presentations in data analysis implies that the user has not only the opportunity to look at the picture. S/he should be able also to manipulate the presentation: this can help to reveal important features and patterns in data that would remain latent in a static graphic. Various interactive manipulation techniques for visual data exploration were proposed in the area of statistical graphics [9]. The most widely known is the idea of visual linking of several graphical displays by means of brushing. M.Monmonier [10] suggested to apply this technique to maps linked with non-cartographic displays. Later the idea of linking between different maps and other graphics was implemented by J.Dykes [11] in his **CDV** system. **CDV** also offers facilities for interactive change of map symbolism, investigating contiguity relationships, and some others. It's worth saying that an interactive tool for changing presentation parameters with the aim of making maps more expressive was proposed by T.Yamahira et al. [12] much earlier than the notion of interactive displays emerged. These researchers developed a histogram-based interface to select class intervals for a classed choropleth map. Later S.Egbert and T.Slocum [13] considered the interactive classification as an exploratory task in their **ExploreMap** system.

Another well-known group of interactive manipulation techniques is devoted to database querying. The user can specify and alter query conditions with the aid of convenient graphical widgets and immediately observe corresponding changes in search results presented graphically (**"Dynamic Query"** [14], **"Attribute Explorer"** and **"Influence Explorer"** [15]).

Descartes [2,3] offers a number of interactive exploratory techniques:

- linking of maps and other displays;
- visual comparison with an interactively selected value;
- interactive classification of a single attribute and interactive cross-classification of a pair of attributes with the possibility to study distribution of values of different variables over classes;
- interactive focusing on particular qualitative values or their subsets;
- outlier removal and interactive focusing on numeric value subranges.

In the next chapter we describe why possible interactive manipulation techniques should be considered during visualization design, and how this should be reflected in the system's knowledge base.

2. Interactive visualization: another set of basic user's tasks is needed

One can see that the *primitive user's goals* typically considered in studies on users' perception of graphics and in visualization design systems could be easily achieved by means of interactive manipulation of maps. Let us consider an example with a presentation of a single numeric attribute by an unclassed choropleth map with a single-hue color scheme [16].

In the theory of cartographic presentation of information it is stated that a choropleth map is well suitable for viewing spatial *distributions* and finding spatial trends and clusters, but badly fits the tasks of *lookup*, *location*, and *comparison*. However, in **Descartes** a choropleth map is supplied with interactive tools supporting the latter 3 tasks:

- Pointing on an object in the map results in the corresponding value being shown in a special window.
- Click on an object in the map changes the color scheme from the single-hue to the double-ended [16]. The value associated with the object becomes the *reference value for visual comparison*. Values exactly equal to the reference value are shown in white, larger values are shown by shades of the first hue, and smaller values – by shades of the second hue. The degree of darkness corresponds to the distance to the reference value.
- The map is supplied and dynamically linked with a number line. Movement of the cursor through the number line highlights correspondent objects in the map, and movement through

the map highlights values on the number line.

Having these controls, the user can easily perform each of the tasks by a single mouse operation:

- to retrieve a value associated with an object (*lookup* task), s/he just points on the object in the map:
- to locate a value on the map (*location* task), the user points on this value in the number line;
- to *compare* values for a pair of objects, it is sufficient to click on any of them and see the color hue of the second one.

So, the 3 tasks the static choropleth map was considered as unsuitable for now can be solved in one touch. **Descartes** also offers one more control to facilitate the remaining generic task, study of distribution:

- The number line has a pair of interactive delimiters to limit the number range to be represented. Moving the delimiters removes the objects with values outside the selected interval from the map. The visualization of the remaining objects changes so that the maximum degree of darkness corresponds to the new absolute maximum value (i.e. either the upper or the lower boundary of the selected interval).

This control can be considered as providing a kind of thematic scaling of the map. Scaling into some subinterval of the whole range makes the visualization more expressive: it emphasizes differences among selected objects and at the same time simplifies viewing them as a group since the other objects are removed from the map. With this addition, we can not only visually separate different clusters, but also focus on objects of a cluster and study the distribution of them.

In this example it is visible that the interactive variant of the choropleth map makes it possible to perform the tasks normally not allowed by a static map and further promotes fulfilling the task the static map is recommended for. Similar interactive operations can be proposed as well for other visualization techniques. They will change the applicability of these maps for various analytical activities. We see 2 implications from this:

- interactive cartography requires new knowledge about the use of interactive maps that is not available in the traditional literature on cartography and graphic design;
- more sophisticated analytical tasks should be considered for the formulation of the knowledge about map usage.

3. User guidance: possible scenarios.

Comprehensive data analysis usually requires quite a number of operations with data and their display. Accordingly, the functions and facilities available in GIS are numerous. This means that the user should learn them and always keep in mind. Further, a rather long sequence of operations is often needed to proceed from source data to a useful presentation. For example, it may be necessary to transform absolute values to percentages, calculate differences or ratios, filter database records, etc. We intend to "wrap" such operation sequences into analysis scripts presented to the user as various analytical tasks formulated in terms of analyzed data and domain notions. These scripts will, first, simplify the acquaintance with the system and release the users from memorizing its capabilities and, second, save time and efforts of even experienced users.

The following example explains our idea. Suppose that a dataset under analysis contains earlier cited fields with absolute population number in sex-age population groups for different countries. The system can foresee several analytical tasks that can be done with the use of these data:

- study how sex structure varies depending on age;
- study how age structure varies depending on sex;
- study sex (or age) structure across countries irrespective of age (or sex);
- examine a particular age group, etc.

These or similar formulations are proposed to the user as alternatives to select from. Standing behind each task is a sequence of operations resulting in potentially useful presentation or several presentations and, possibly, some recommendations how to use them and how to proceed further.

Suppose that the user has selected the first task, study of dependency of sex structure on age. In response the system automatically calculates percentages of male and female in all age groups and creates a map with segmented bars: bars correspond to age division, and segments show proportions of male and female. Note that automation of calculating percentages and selection of this type of presentation really requires knowledge of conceptual relationships among fields.

Displaying the map to the user, the system supplies it with a brief comment explaining that this map is suitable for seeing local differences in sex structure depending on age in each country or for pairwise comparison of countries. It does not help in seeking for spatial patterns and trends. Thus the system offers as a direction for further investigation to take separately percentages of male or female and consider their spatial distributions for different ages. Alternatively, the user may be proposed to concentrate on studying differences in percentages of male and female population depending on age. For the first task a series of choropleth maps would be suitable. In the second case the system would automatically calculate the differences and represent them by bar chart map. At the next step the system may propose the user to study spatial distributions of differences for the age groups.

User guidance applies also to the utilization of interactive manipulation facilities for data analysis. Again, the system can help the user not only by a general description of this or that tool (static, pre-prepared on-line help) but also with some data- and analysis context-specific recommendations. For instance, if in the course of analysis a ratio of two numeric fields was calculated and presented, the system can propose to apply visual comparison with the value 1; for a difference of two fields visual comparison with 0 is reasonable. In both cases the map will change so that the geographical objects will be visually classified into 3 groups: 1) field1<field2; 2) field1=field2; 3) field1>field2 (sometimes another classification may be useful: field1<<field2; field1<>field2; field1>>field2). The system can also automatically detect cases when interactive outlier removal is necessary and propose the user to do this.

It should be noted that the use of guidance is optional: the user does not have to analyze data according to proposed scenarios. S/he always has the possibility to apply any of the available functions in any order. This is important, as we cannot guarantee that it is possible to foresee all imaginable analysis tasks. Yet, since the guidance is proposed stepwise, the scripts may occur to be useful for partial automation of rather sophisticated investigations.

4. User guidance: which kinds of knowledge will be utilized in the project

In guiding the user the system will utilize the following kinds of knowledge:

1. *Generic analysis tasks* such as "Local comparisons of values of attributes", "Looking at spatial distribution of values of an attribute", "Study of dynamics of data", "Local consideration of proportions" etc. The tasks may have applicability conditions. For example, the latter task makes sense for a set of data fields that together constitute a meaningful whole. Unlike the generic tasks in the **Vizard** system, our tasks are patterns rather than simply abstract statements. The patterns have slots filled with appropriate domain notions when the system proposes analysis scenarios to the user.

2. *Knowledge about methods of cartographic and graphical presentation*: which generic analysis tasks are enabled by each of the methods. For example, "Parallel bars" => "Local comparisons of values of attributes", "Choropleth map" => "Study spatial distribution of values of an attribute", "Scatter plot" => "Look for relationships between two attributes". Some presentation methods offer different opportunities depending on data they applied to. For example, "Pie charts"/absolute quantities => "Local consideration of proportions", "Comparison of totals"; "Pie charts"/percentages => "Local consideration of proportions", "Comparison of proportions for pairs of geographical objects".

3. *Knowledge about potentially useful operations with data*: for what generic tasks they can be applied and how to perform each operation with the use of available functions. An example of such an operation is proceeding from absolute values to percentages. This operation is helpful, in particular, in the task of studying proportions (other variants of application are also possible). It is performed with the use of the calculation function of the system.

4. *Knowledge about interactive manipulation facilities* available in the system: possible ways of use depending on the analysis context. Here belong the earlier mentioned heuristics about visual comparison with 1 for calculated ratios and with 0 for calculated differences. Another example concerns the application of interactive classification tool for investigating relationships between one attribute selected as a base of classification and some other attributes for that class statistics is calculated and displayed. A reasonable strategy is to try to increase the number of classes and move class boundaries to probe the robustness of the demonstrated relationship, if any.

5. *Knowledge about data and underlying problem domain*. This knowledge, besides selection of proper visualization methods, allows formulating analysis tasks in a way easily understandable by the user. Thus, the generic task "Local estimation of proportions" may have a formulation "Consider proportions of age groups 0-14 years, 15-64 years, 65 and more years in population of each country of Europe" or "Consider proportions of classes of industry *X*, *Y*, ..., *Z* in overall industrial product of main cities of Germany", depending on the application domain. The knowledge about data is used in automatic application of such system functions as calculations, querying, and classification according to the pursued analysis scenario.

The utilization of these kinds of knowledge for generating guidance proposals on different steps of user's work may be governed by <u>*rules*</u> with following structure:

IF [*applicability conditions*] **THEN** [*recommendation*], where

[*applicability conditions*] may include one or more of the following:
a) required data characteristics and relationships;
b) characteristics of currently considered presentation;
c) currently pursued generic task;
[*recommendation*] may be either one or more generic tasks to proceed to or a hint concerning the use of interactive map manipulation facilities.

Acknowledgments

We thank all the partners of the project for interesting and productive discussions during the preparation of the project proposal and further meetings. Constructive critics and advises of the EU project officer, Mr. U.Boes, are kindly appreciated. Numerous discussions within GMD SET.KI Cooperative Design group gave us a wonderful opportunity to probe our ideas.

References

1. Perdigao, J. CommonGIS: GIS for Everyone, from Everywhere. *4th EC - GIS Workshop* Budapest, Hungary, 24 - 26 June, 1998. URL http://commongis.jrc.it/CommonGIS/
2. Andrienko, G., and Andrienko, N. Intelligent Visualization and Dynamic Manipulation: Two Complementary Instruments to Support Data Exploration with GIS. *Proceedings of AVI'98: Advanced Visual Interfaces Int. Working Conference* (L'Aquila Italy, May 24-27, 1998), ACM Press, pp.66-75
3. Andrienko, G., and Andrienko, N. Visual Data Exploration by Dynamic Manipulation of Maps. T. Poiker and N.Chrisman (eds.*) 8th International Symposium on Spatial Data Handling, SDH'98*, July 11-15, 1998, Vancouver, Canada, pp.533-542
4. Jung, V. Knowledge-Based Visualization Design for Geographic Information Systems. ACM-GIS'95. *Proceedings of the 3rd ACM International Workshop on Advances in Geographic Information Systems* (held in Baltimore, Maryland, on Dec. 1-2, 1995), ACM Press, New York, 1995, pp.101-108
5. Van der Berg, C., Tuijnman, F., Vijlbrief, T, and others. Multi-server Internet GIS: Standardization and Practical Experiences. *International Conference on Interoperating Geographic Information Systems*. Santa-Barbara, CA, December 3-4, 1997, NCGIA, pp.296-299.
6. Mackinlay, J. Automating the Design of Graphical Presentation of Relational Information, *ACM Transactions on Graphics*, **5**(2), 1986, pp.110-141.
7. Bertin, J. *Semiology of Graphics. Diagrams, Networks, Maps.* The University of Wisconsin Press, Madison WI, 1983.
8. Zhan, F.B. and Buttenfield, B.P. Object-Oriented Knowledge-Based Symbol Selection for Visualizing Statistical Information, *International Journal of Geographical Information Systems*, **9**(3), 1995, pp.293-315.
9. Cleveland, W.S. *Visualizing Data*. Hobart Press, Summit NJ, 1993
10. Monmonier, M.S. *How to Lie with Maps*. University of Chicago Press, 1991.
11. Dykes, J.A. Exploring Spatial Data Representations with Dynamic Graphics. Computers & Geosciences, **23**(4), 1997, pp.345-370
12. Yamahira, T., Kasahara, Y., and Tsurutani, T.: How Map Designers can Represent their

Ideas in Thematic Maps. *The Visual Computer*, **1**, 1985, pp.174-184
13. Egbert, S.L., and Slocum, T.A.: EXPLOREMAP: An Exploration System for Choropleth Maps. *Annals of the Association of American Cartographers* **82** (2), 1992, pp.275-288.
14. Ahlberg, C., Williamson, C., & Shneiderman, B. Dynamic Queries for Information Exploration: an Implementation and Evaluation. *In Proceedings of ACM CHI'92 Conference*. ACM Press, 1992, pp.619-626
15. Tweedie, L.: Characterizing Interactive Externalizations, in *Proceedings of CHI’97* (Atlanta GA, March 1997), ACM Press, 1997, pp.375-382.
16. Brewer, C.A. Color Use Guidelines for Mapping and Visualization, in A.M.MacEachren and D.R. Fraser Taylor (eds.) *Visualization in Modern Cartography*. Elsivier Science Ltd, 1994, pp.123-147.

Last updated 30th of November, 1998