

Intelligent Support of Visual Data Analysis in Descartes

Gennady Andrienko and Natalia Andrienko
Fraunhofer Institute Autonomous Intelligent Systems,
Knowledge Discovery Team AiS.KD,
Schloss Birlinghoven, Sankt Augustin, 53754 Germany
Tel. +49-2241-142329, Fax +49-2241-142072
Gennady.Andrienko@ais.fhg.de

ABSTRACT

Descartes is a knowledge-based system supporting exploratory analysis of spatially referenced data with the use of maps. It is capable of automated generation of maps and other graphical displays that represent data selected for analysis. Besides, the system selects and combines appropriate analytical instruments (graphical displays, data transformations, interactive operations etc.) depending on the analysis task the user needs to perform and instructs the user how to employ these instruments for this task. The intelligent behaviour of the system is based on three kinds of knowledge: the rules of data representation depending on the characteristics of the data and relationships between data components, the possible analysis tasks and correspondence between the tasks and the available instruments, and domain-specific knowledge allowing the system to “understand” the content of data.

1. Introduction

In the past production of maps was an exclusive business of cartographers, i.e. professionals with special knowledge of the principles of data representation in graphical and, in particular, map form. Wide spreading of GIS (geographic information systems) promoted involvement of lots of people into creation of maps for their own purposes. Although these people may be highly qualified specialists in their occupation domain such as demography, marketing, or ecology, they often lack expertise in design of graphics and maps. This may lead to representation of data that does not comply with the principles of graphics and cartography. Such

representation may be completely useless (important information is not visible) and even misleading (wrong conclusions are derived). The problem becomes even more serious with appearance of mapping services in the WWW that engages wide public into mapping activities.

Since it is impractical that professional cartographers and experts in statistical graphics completely serve the growing demand in visual displays of data, it is necessary to incorporate the required expertise into the software used for generation of these visual displays.

Our system Descartes was the first knowledge-based system for thematic mapping that appeared in the WWW [1,2]. It is capable of automatic selection of proper representation methods depending on characteristics of data and relationships among data components. For this purpose Descartes uses a rule base that embodies the established principles of cartographic representation. Recently Descartes was extended to account also for the data analysis tasks that need to be done with the use of maps.

In this article we first give an overview of existing software systems incorporating graphical or/and cartographic expertise. Then we shortly describe the system Descartes that brings such expertise to the service of users in the WWW. At the end we speak of a completely new development in Descartes: a knowledge-based guidance in visualisation and exploratory analysis of data.

2. Knowledge-based systems for graphical representation of data (literature overview)

There exist a number of research systems that apply the generic knowledge on data visualisation for automated graphics generation (see the survey [3]). According to the approach first proposed by Mackinlay [4] and followed by other researchers, the data set to be presented is divided into subsets with less numbers of fields so that it becomes possible to select some *visual primitive* for each subset. Then the primitives are combined with the use of available *composition operators*. For example, two numeric fields

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Smart Graphics '02, June 11-13, 2002, Hawthorne, NY.
Copyright 2002 ACM 1-58113-000-0/00/0000...\$5.00.

can be encoded by positions of marks along X- and Y-axes, sizes of the marks can represent another numeric field, and the marks can vary in color to depict values of some qualitative field.

The Mackinlay's work showed the feasibility of automated knowledge-based graphics design. Later systems further developed his approach. In the system VISTA [5] more visualisation techniques and composition operators are available that allow, in particular, to build 3D graphics. In the project SAGE [6] the register of data characteristics to be accounted for in graphics design is significantly extended. In addition, the design is sensitive to the user's information seeking goals expressed as generic tasks of the kind "accurate lookup of separate data values", "comparison of values of two attributes", "study of the distribution of values", "revealing of functional correlation among attributes".

Casner [7] considers more precisely specified tasks constructed from primitive *logical operators* such as a query for a property of an object. His system BOZ receives a formal description of a task and substitutes the logical operators by primitive *perceptual operators* such as search for an object with a given graphical property or lookup for a specific property of a graphical object. Then the system selects and combines visual primitives that support these operators. The perceptual operators belong to the elementary level of reading, therefore the resulting graphics can serve as good data inventories rather than support exploratory analysis.

All the above-mentioned systems deal with non-spatial data. Cartographic presentation of spatially referenced data has its peculiar features and requires special consideration. In [8] an expert system is described that consults a GIS user what visualisation method should be selected for a given data field. The system does not deal with several fields and does not generate maps. Later the system VIZARD [9] appeared that automatically presents spatially referenced data on maps applying methods developed in cartography.

A common feature of all known works on automated design of graphical data displays is that they generate just static, non-interactive images that are fully analogous to "paper" graphics and maps. Being enhanced with interactivity and transformability, such displays could much better support various data analysis tasks.

A unique feature of Descartes is that the maps it generates are highly interactive and can change their appearance in response to user's actions. The user may simultaneously have on the screen several different maps as well as graphical displays of other types (dot plot, scatter plot, parallel co-ordinate plot) supporting viewing of the same data from multiple perspectives. All parallel views of the data are dynamically linked. This means that corresponding elements are simultaneously highlighted in all the displays when the user interactively selects them

with the mouse in one of the graphics. Besides, the maps in Descartes can dynamically change their appearance in response to certain user's actions. Manipulation of maps by the user can make the representation of the data more expressive and help to reveal important relationships initially hidden within the data volume.

All these services provided by Descartes can be accessed through the WWW. The system has a client-server architecture. The server performs knowledge-based design of maps according to data characteristics and conceptual relationships between data components. It sends specification of the appropriate representations to the client. The client generates maps according to the specification and provides various interactive tools for manipulating them.

Knowledge-based map design in Descartes is described in [10,11], the interactive facilities of Descartes are presented in detail in [10,12-14]. Here we focus on the intelligent support of data analysis in the system.

3. Knowledge-based support of data analysis

A new feature that appeared recently in Descartes is a so-called task support guide, the role of which is to help the user to select appropriate analysis instruments according to the analytical tasks to be fulfilled. Here the term "analysis instruments" embraces different methods of cartographic representation of data, non-cartographic displays (e.g. a scatter plot or a cumulative curve), data transformation operations (e.g. calculation of percentages from absolute values), and various interactive facilities such as access to data through a map or a plot, dynamic linking of parallel views, manipulation of visualisation parameters etc. (see [10-14]). The task support guide uses a knowledge base on correspondence between data characteristics, possible tasks, and available instruments.

Creation of this knowledge base occurred to be a very difficult work. We could not reuse the experience of the developers of earlier systems [6,9]. First of all, the knowledge used in these systems and available in literature [15,16] refers to static graphics and maps. Displays produced by Descartes radically differ by their properties (interactivity, possibility of dynamic change) from static presentations and can therefore support much wider range of possible tasks than their static counterparts.

Another deficiency of the existing principles of data representation is that they are much less elaborated with respect to the possible analysis tasks that could be supported by graphical displays than to the characteristics of data these displays can portray. Moreover, we did not find in the literature a sufficiently comprehensive typology or classification of possible tasks. We had to develop our own typology and then to analyse the existing computer-

based instruments for visual analysis in order to link them to the types of tasks.

We were also unsatisfied with the level of the tasks considered in the earlier systems. To our opinion, it is difficult and unnatural for the user to decompose her/his analytical activities into such primitive operations as “lookup exact values”, “locate values on the map”, and “compare values” and then receive a separate instrument for each kind of operation. Although data exploration involves such operations, they are often done in parallel or intermingle and therefore need to be supported by the same instrument or by linked instruments. And, of course, the user should have an opportunity to communicate her/his goals to the system on a higher level, for example, “analyse population structure”.

One more difficulty we encountered was the user interface problem: how to recognise the user’s task in order to recommend her/him the appropriate instruments.

The solution to the problem of task level and the problem of task recognition we found in the idea of generic “scenarios”. The system proposes the user to follow one of the scenarios of data analysis defined in the knowledge base that can be applied to the data under analysis. Examples of possible scenarios are “analyse statistical data referring to territory division” or “select appropriate location”. A scenario is a hierarchy of tasks and subtasks that ultimately refer to the primitive operations. However, these operations are hidden from the user who operates higher level concepts. When the user selects a generic task within a scenario, the system selects and combines the analysis instruments according to the primitive operations involved in the task. Besides providing the user with the instruments, the system supplies her/him with instructions of how to use these instruments for the selected task. The knowledge base for task support is defined in the XML language. It consists of two major parts: definition of scenarios and description of available instruments. Each scenario is a hierarchy of general tasks. The instruments include

- cartographic representation methods such as unclassified and classed choropleth maps or maps with bar charts;
- map manipulation techniques associated with particular representation methods (for example, visual comparison is associated with unclassified choropleth maps);
- other interactive techniques not restricted to a particular map type, such as the dynamic query tool or the tool for accessing exact attribute values;
- non-cartographic displays such as the scatter plot or the plot of parallel coordinates;
- analysis tools based on data transformations and calculations, for example, a tool for automated ranking of options on the basis of multiple criteria

implementing one of algorithms of multi-criteria decision support.

4. The user interface of task-supporting guide

The guide may be activated on user’s initiative at any time. At the beginning the guide will ask which of the scenario should be applied. Additionally, the user must tell the guide which data sets to use. For example, in the scenario “explore demographic data referring to territory division” the guide needs to know which of the available map layers contains the territory division and what attributes the user is interested in. This information is called the **context** of the scenario. After definition of the context the guide will display the hierarchy of potentially relevant tasks for the selected scenario (Figure 1).

The left part of the window in Figure 1 shows the currently defined context. The user may change this by pressing the “change” button. The upper-right part of the window displays the task hierarchy. Circles denote terminal tasks and “folder” symbols mark higher-level tasks. The lower-right part contains an explanation of the meaning of the currently highlighted task. By traversing the task hierarchy, the user may view explanations of all tasks.

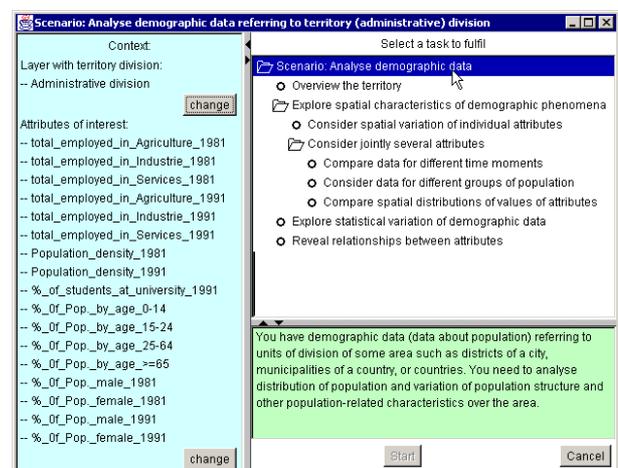


Figure 1. The guide displays the hierarchy of potentially relevant tasks for a user-selected generic scenario.

When the user selects a terminal task for execution, the guide replaces the frame displaying the task hierarchy by a frame devoted to this particular task (Figure 2). As before, on the left the context is shown. On the right the system displays a general instruction how to use the map in order to do the task. Below this instruction the recommended instruments are listed.

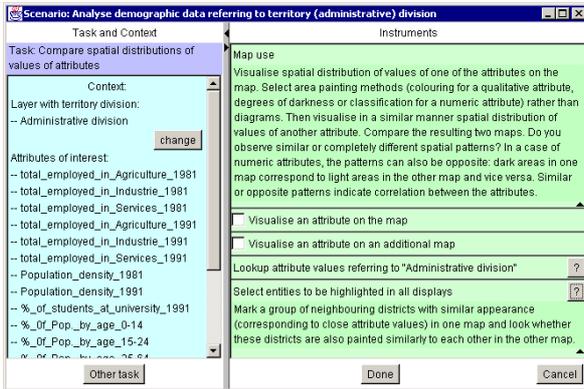


Figure 2. The guide displays a map use instruction and a list of instruments appropriate for the task “Compare spatial distributions of values of attributes”.

Instruments appropriate for a task are of two kinds. One kind is system functions or tools that must be specially activated. The other kind, such as access to attribute values or highlighting of objects, are always available without special activation. The guide marks instruments of the first kind with checkboxes which may be clicked for activating them. Instruments of the second kind are also listed among the recommended instruments in order to at least remind the user that they may be helpful in fulfilling the task. The buttons labeled by question marks provide access to the appropriate pages of the system’s on-line manual where the user can find more extended explanations concerning the use of the tools. If provided in the knowledge base, the guide can additionally display task-specific instructions regarding the permanently available instruments. In our example (Figure 2) there is an instruction about the use of the highlighting facility for the task “Compare spatial distributions of values of attributes” (the instruction appears at the bottom-right of the window). At the same time a “help” button is also available. By pressing it the user gets a general instruction concerning highlighting of objects in the system.

For any instrument included in the list of recommended instruments the user can get a brief explanation of its meaning. The explanation is displayed in a popup window when the user moves the mouse cursor over the name of the instrument (Figure 3).

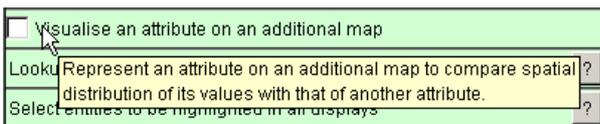


Figure 3. The guide explains the user the meaning of an instrument.

When the user activates some instrument by checking the corresponding checkbox, the guide may ask to specify the input for the instrument. Thus, for the instrument “Visualize an attribute on the map” the user needs to select an attribute. When the input is specified, the guide launches the appropriate tool. This may cause appearance of a new window, for example, with an additional non-cartographic data display. At the same time information about the tool is inserted into the guide’s window (Figure 4). This information consists of two parts: 1) general explanation of how the instrument works (e.g. for a graphical display – how data are represented by graphical symbols); 2) task-specific instruction, i.e. how to use the instrument for the current task. The latter is constructed from the instructions for the particular basic operations supported by the instrument (these instructions are specified in the knowledge base). The guide will show only instructions for those operations that are relevant for the current task.

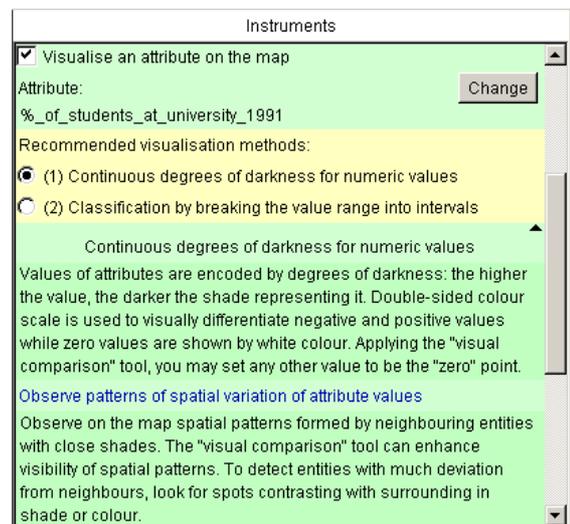


Figure 4. The guide explains how numeric values are represented using a continuous scale of degrees of darkness.

For the visualization of attributes on maps the guide only proposes cartographic representation methods that are appropriate with regard to the selected attributes and the current task. The guide can also provide explanations about each of the selected presentation methods in a popup window. If the user selects one of the methods the data are immediately represented on the map by means of this method, and appropriate map manipulation tools are attached to the map. At the same time the guide inserts information about this method in the window for task support (Figure 4). Information about the manipulation tools is also provided.

To avoid overloading the user with masses of text, all the explanations and instructions appear initially in a “folded” state. A triangular mark resembling downward-oriented arrow indicates presence of a text. Clicking on the mark opens the text for reading. The user can “fold” the text again by clicking on the upward-oriented arrow that appears when the text is opened.

In summary, the user interface of the guide serves the following purposes:

- Remind the user about potentially relevant subtasks of a task, recommend appropriate sequence of subtasks;
- Suggest analysis instruments when they are useful, i.e. appropriate for the subtask being currently fulfilled;
- Activate the instruments when the user decides to apply them, i.e. visualize data, create corresponding controls, etc.;
- Explain how to use the instruments that are relevant for the current subtask.

5. Directions for further work

The current design and implementation of the guide is a prototype to be further developed in several directions. First, the guide must be extended in order to process input-output links between instruments, from instruments to tasks, and between tasks. On this basis the guide must be enabled, when necessary, to induce a logical sequence of performing tasks. Second, we need to extend the knowledge base, which currently only contains knowledge for two general scenarios, to a broader spectrum of scenarios. Third, there was no opportunity to test the usability of the guide. In particular, we need to validate the following hypotheses:

- The concepts of general scenarios, tasks, and task hierarchies are understandable and appropriate for users, i.e. users are able to relate their specific tasks to the general tasks the system deals with;
- The guide encourages the use of available interactive techniques, i.e. more users will actively apply more of the interactive techniques;
- The instructions provided by the guide are effective, i.e. people can and want to learn how to apply the new exploratory instruments without previous oral presentation and live demonstration;
- The user interface of the guide is understandable and convenient.

Regarding the user interface, we are in particular worried about the large amounts of texts provided by the guide. From our experience and the results of the usability tests we learned that common users are reluctant to read any manuals or explanations, even when they are available on-line. On the other hand, it is necessary not only to

prompt people into using the novel exploratory techniques but also to explain how these techniques work and how to apply them. It is hard to do this without texts. Since the guide shows texts in relatively small portions and only when applicable, we hope that the users will not feel overloaded. Of course, this needs validation.

We will also pursue an idea of complementing textual information with some visual ways of demonstrating the work of the tools. For example, when Descartes generates a choropleth map, the guide could automatically play an animation demonstrating how this map can be changed through the visual comparison tool. Such a visual demonstration may already be sufficient for an advanced user to understand how to use the tool, and in this case there would be no further need in reading the corresponding texts. At the same time a visual mini-demonstration could help casual users in understanding the textual explanations. Besides, it might raise interest and willingness to learn more about the tool.

6. Conclusion

In providing various services in the WWW it is necessary to take into account that the wide public expected to access these tools often does not possess the necessary expertise to operate them. Therefore there is a need in this expertise being incorporated into the tools. This can help to prevent misuse of the tools and promote better utilisation of the available facilities for the tasks they are designed for.

Descartes is a system for visual exploration of spatially referenced data. With its client being a Java applet, it offers its services in the WWW (URLs <http://borneo.gmd.de/and/> and <http://www.commongis.com/>). The system incorporates knowledge of three kinds: the rules of data representation depending on the characteristics of the data and relationships between data components, the possible analysis tasks and correspondence between the tasks and the available instruments, and domain-specific knowledge allowing the system to “understand” the content of data. On the basis of this knowledge the system automatically visualises data on maps and other graphical displays and provides intelligent guidance to the user in the process of analysis.

The critical issue for the system is acquisition of knowledge necessary for “understanding” of the meaning of data components and relationships between them. We are now working on development of an interactive tool that interviews the user about the data when this is necessary in the process of analysis and builds a knowledge base about data from such fragments acquired “on-line”.

7. References

- [1] Andrienko, G. and N. Andrienko. Knowledge-Based Support for Visual Exploration of Spatial Data. *Extended Abstracts of Int. Conf. CHI'97*. ACM Press (1997) 16-17
- [2] G.L. Andrienko, and N.V. Andrienko. Intelligent cartographic visualisation for supporting data exploration in the IRIS system. *Programming and Computer Software*, 23 (1997) pp.268-282
- [3] B.S. Murray, Intelligent Information Presentation Systems. *The Knowledge Engineering Review* 9 (1994), pp.269-286
- [4] J. Mackinlay. Automating the Design of Graphical Presentation of Relational Information. *ACM Transactions on Graphics* 5 (1986), pp.110-141
- [5] H.Senay, and E. Ignatius. A Knowledge-based System for Visualization Design. *IEEE Computer Graphics and Applications* 14 (1994), pp.36-47
- [6] Roth, S.M. and J. Mattis. Data Characterization for Intelligent Graphics Presentation. *Proc. SIGCHI'90: Human Factors in Computing Systems*. ACM Press (1990) 193-200
- [7] S.M Casner. A Task-analytic Approach to the Automated Design of Graphic Presentations. *ACM Transactions on Graphics* 10 (1991), pp.111-151
- [8] F.B. Zhan, and B.P. Battenfield. Object-oriented Knowledge-based Symbol Selection for Visualizing Statistical Information. *International Journal of Geographical Information Systems* 9 (1995), pp.293-315
- [9] Jung, V. Knowledge-based Visualization Design for Geographic Information Systems. *Proc. 3rd ACM Workshop on Advances in GIS*. ACM Press (1995) 101-108
- [10] G. Andrienko, and N. Andrienko. Interactive maps for visual data exploration. *International Journal Geographical Information Science* 13 (1999), pp.355-374
- [11] Andrienko, G. and N. Andrienko. Knowledge Engineering for Automated Map Design in DESCARTES. In C.B.Medeiros (ed.) *Advances in Geographic Information Systems. Proceedings of the 7th International Symposium ACM GIS'99*. ACM Press, New York (1999) 66-72
- [12] G. Andrienko, and N. Andrienko. Exploring Spatial Data with Dominant Attribute Map and Parallel Coordinates. *Computers, Environment and Urban Systems* 25 (2001), pp.5-15
- [13] N. Andrienko, G. Andrienko, and P. Gatalisky. Exploring Changes in Census Time Series with Interactive Dynamic Maps and Graphics. *Computational Statistics* 16 (2001), pp.417-433.
- [14] Andrienko, G. and Andrienko, N. Constructing Parallel Coordinates Plot for Problem Solving. In A.Butz, A.Krueger, P.Oliver, and M.Zhou (Eds.) 1st International Symposium on Smart Graphics, Hawthorne, New York, USA, March 21-23, 2001, ACM Press, 2001, pp.9-14
- [15] Bertin, J. *Semiology of Graphics. Diagrams, Networks, Maps*. Madison: The University of Wisconsin Press (1967, 1983)
- [16] Robinson, A.H., J.L. Morrison, P.C. Muehrcke, A. Jon Kimerling, and S.C. Guptil. *Elements of Cartography*. Wiley, New York (1995)