Java Tools for Exploratory Cartographic Visualization in the Internet

Gennady Andrienko and Natalia Andrienko

GMD - German National Research Center for Information Technology

Sankt-Augustin, Germany

Abstract

:

Keywords: visual data exploration, interactive cartography, GIS, thematic maps, Java, Internet

0. INTRODUCTION

Cartographers have long adhered to the view of the map as primarily a medium of communication [1]. Recently, influenced by ideas of scientific visualization and exploratory data analysis, they started to pay more attention to the role of the map as a tool to support visual thinking and decision making [2,3]. To play this role effectively, the map needs two principal additions: interaction and dynamics.

Within AiS Knowledge Discovery research group we are developing Java based tools for exploratory thematic mapping of spatially referenced data. Different Java applets were developed for support of analysis of various types of spatially related information. In our paper [4] we discussed three main approaches to geographic visualization in the Internet:

1. Server-based architectures;

2. Client-based architectures;

3. Client-server architectures

In the first approach, static raster images are generated on the server side and then transmitted to the client. This approach is traditionally used in "Internet Map Server" solutions of traditional GIS companies. It has very strong limitations on interactivity and dynamics of maps and is practically not useful for exploratory data analysis.

Since 1996 we develop Java applets which have client or client-server architectures. In the first section we describe two client-server systems, Descartes and CommonGIS. Both are created for automatic design of thematic maps and are able to support visual data exploration using interactive manipulation of screen maps. In the second section, we present "pure" client Java applets: dMap for publishing interactive maps created by Descartes in the Internet, and applets for visualization of different types of spatio-temporal information. The third section includes descriptions of the systems not available in the Internet: an integrated system Descartes+Kepler for spatial data mining, and Decade for spatial decision support.

1. CLIENT SERVER SOFTWARE

Descartes and CommonGIS systems assist users in the exploration of spatially referenced statistical data, i.e. economical, demographic or ecological data related to geographical locations. The systems incorporate knowledge about thematic cartography to automatically generate thematic maps that correctly visualize selected data. They also support the further, interactive exploration of these maps.

Current Geographic Information Systems specifically support the analysis and manipulation of spatial objects. However, they are weak regarding their support of visualizing spatially referenced statistical data. Their major shortcomings are:

- Users bear full responsibility for a correct selection of presentation techniques. Consequently, they must proficiently know principles of cartographic presentation, have to spend considerable time and efforts for map production, and cannot entirely concentrate on their task of analysis and problem solving.

- GIS are typically complex and difficult to operate; significant training is required.

- GIS do not sufficiently support the further exploitation and exploration of produced maps, i.e. their use in data analysis.

Both our systems overcome these shortcomings. They provide intelligent mapping support and a full spectrum of functions for interactive visual data analysis. Some traditional GIS functions (manipulation of layers, data access through a map, etc.) are also available.

From architectural viewpoint, they have client and server components which are communicating. Descartes uses a server which is running permanently while the client is active. CommonGIS starts server part each time the output from it is needed

1.1 Descartes: Knowledge-based System for Visualization Design

Descartes was created in 1996 on the basis of our earlier system Iris which was also a knowledge-based system for map design implemented in C++ for Windows [5]. The detailed description of Descartes is available in our paper [6]. Demonstrators of Descartes with dozens of applications (example data sets) are available from several servers:

- http://borneo.gmd.de/and/java/iris/ at GMD AiS.KD, Sankt-Augustin, DE

- http://www.mimas.ac.uk/descartes/ at KINDS (Knowledge Interfaces to National Data Sets) Project, Manchester, UK

- http://geolibrary.uidaho.edu/ at University of Idaho, Moscow, Idaho, USA

The server of Descartes (a C++ program available for various dialects of Unix and Windows) is used for accessing and transforming data and for design of visualizations on the basis of data characteristics and relationships. The user interface (client) of Descartes is implemented as a Java applet. It provides table display, map displays (most of traditional thematic mapping styles are implemented), supplementary graphics displays (dot plots, scatter plot, box plot, parallel coordinates, etc.), and supports interactive manipulations with maps and graphics.

Descartes includes many map manipulation techniques like highlighting, brushing, visual comparison and interactive classification [6], cross-classification [6,9], classification on the basis of dominant attribute [10], etc. Database querying and mathematical calculations are also supported. Their results can be immediately visualized on dynamic maps.
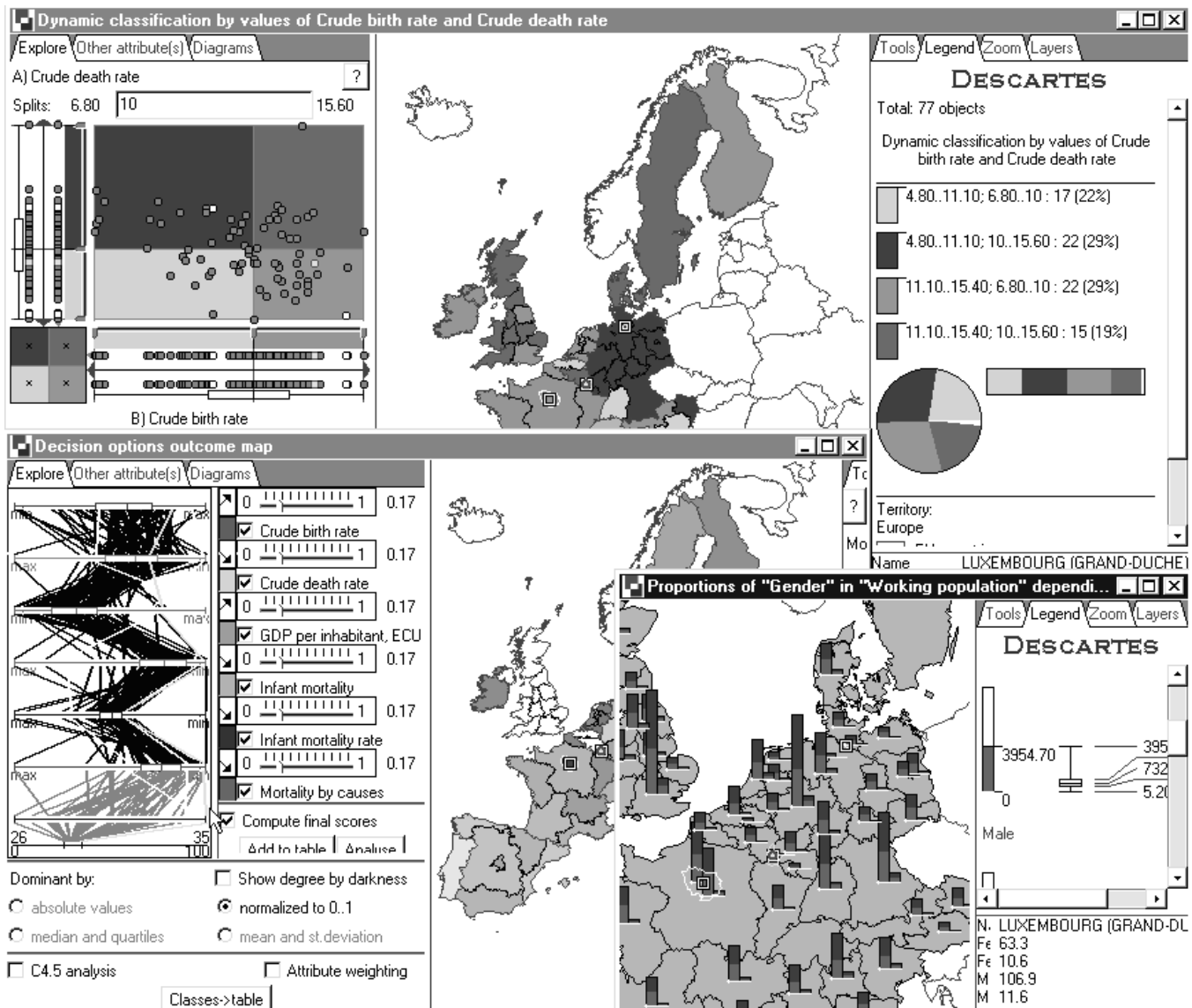
Figure 1. Multiple window user interface of Descartes: maps and statistical graphics linked by dynamic highlighting and brushing

Descartes uses so called data characterization for visualization design, i.e. it is necessary to supply certain description of data to the system. This is done during the process of building an application. To create an application, one should:

- Import geographical data into Descartes' internal format which has a structure isomorphic to the format of ArcView's shape files, but is about two times more compact due to the use of less precise float numbers. Descartes includes a variety of converters from most popular GIS data formats: SHP and E00 by ESRI, MIF/MID by MapInfo, DXF by AutoCad, etc.

- Import table data into DBF format which is used by Descartes.

- Link records of DBF file to contours or points of some layer in geographical data set. The linkage is done through common identifiers. This means that the table contains a column with identifiers of spatial objects the data in each record refer to. The same identifiers are indicated in the file with coordinates or contours of the spatial objects.

- Describe the semantics of the table data (see [8] for detailed description and full list of data characteristics, paper [7] proposes a more general approach to data characterization). This is done by grouping of columns referring to the same attribute and then by specification of the attribute. The columns referring to the same attribute are differentiated by values of a parameter. For example, columns with population

division refer to different values of parameters "gender" and "age group".

The process of making an application is supported by special Java applet called Application Builder described in [8].

The main shortcoming of Descartes is that it uses socket connection for communication between the server and the client (i.e. the server runs permanently while the client is active) and due to this, its functioning is sometimes prohibited by firewalls. This problem was solved in the next our system, CommonGIS, which is developed within the EU-funded R&D project CommonGIS (Esprit project 28983, November 1998 - April 2001).

1.2 CommonGIS - Common Access to Geographically Referenced Data

The CommonGIS system is developed as integration of two systems: the Internet GIS Lava/Magma and the visual data exploration system Descartes. The integrated system combines merits of both systems:

Lava/Magma:

- opportunity to work with distributed data stored on multiple WWW servers;

- possibility of connection to various relational databases;

- efficient work with large databases due to optimization of access to data;

- relatively easy extension through adding user defined types of spatial objects.

Descartes:

- large variety of visualization techniques for attribute data associated with spatial objects;

- fully automatic design of thematic maps with intelligent selection of visualization techniques according to characteristics of data and

- relationships among data components;

- interactive, dynamic maps, i.e. maps changing their appearance in response to user's manipulation. The manipulation devices are

- designed so as to support visual exploratory analysis of data represented on the maps.

The integrated system also has a client server architecture (see [11]). However, unlike in Descartes, the HTTP protocol is used for communication between the client and the server (i.e. for each transaction the server starts, performs necessary operations, sends the results to the client, and stops). The system is not sensitive to firewalls.

To apply the CommonGIS system to some data set, one should import geographical and table data into Lava/Magma component (which supports most of traditional formats and can be linked to a number of database systems) and provide data characteristics for the Descartes component using the Application Builder [8].

Demonstrators of the CommonGIS prototype are available from two servers:

- http://borneo.gmd.de/descartes/CommonGIS/ at GMD AiS.KD, Sankt Augustin, DE

- http://commongis.jrc.it/commongis/sw/commongis_first/CommonGIS/ at JRC, Ispra, IT

2. CLIENT APPLETS

2.1 dMaps: publishing interactive maps in the Internet

Another way of bypassing firewalls is implemented in our dMap system. Working with Descartes, a user has a possibility to save maps. Descriptions of saved maps are stored at the server side. Later they can be interpreted and displayed by a separate Java applet which supports viewing of maps with necessary interactivity, including basic map manipulation (layering, zooming, panning, etc.) and specific controls for manipulation of visualizations (choropleth maps with and without classification, cross-classification maps, dominant attribute maps, bar charts, pie charts, segmented bars, geometric signs, combined presentations, etc.). Principally, this applet supports all visualization techniques available in full Descartes. The difference is that it is impossible in dMaps to switch to another presentation technique or to another attribute or set of attributes.

Several demonstrators of dMaps are available:

- http://borneo.gmd.de/descartes/IcaVisApplet/ Various maps - Internet supplement to our paper [6]

- http://borneo.gmd.de/and/java/iris/app/elect/indexm.html Election results in the city of Bonn

- http://borneo.gmd.de/and/java/iris/app/itc/indexm.html Overijssel (NL) demographic information

- http://www.dialogis.de/gis/dmap/demo.html Some statistics about the city of Bonn and the region (in German)

In principle, it is possible to create dMaps without using Descartes. A spin-off company of GMD, Dialogis, developed an ArcView extension which allows to export data from ArcView to dMaps directly. The tool also allows to send produced maps together with the applet via e-mail.

2.2 Visualization of temporal information

In our research we consider various aspects that can play a role in visual exploration of spatio-temporal data. Among them, besides different animation modes and techniques of focusing on time moments or intervals, are methods of presentation of data in the map, ways of user-map interaction, alteration of display properties in a way productive for analysis, various transformations of data etc. We realize that diverse tools are needed depending on which aspects of spatial phenomena vary in time: existence, spatial location, geometry (shape and size), or thematic (attribute) data.

To support interactive exploration of data, we have developed an assembly of widgets further referred to as "time manager". The time manager is connected to a dynamic map display. It allows the user to select time moments or intervals the data presented in the map to refer to. For selecting an interval, the user specifies its starting moment and length. To select a moment, the length of the interval should be set to one time unit.

The current display moment or interval can be shifted forth and back along the time axis either by pressing the buttons "step forth" (">") and "step back" ("<") or by dragging the slider. Each

action causes the map being immediately redrawn to represent the new moment or interval.

This kind of operation may be called user controlled animation. It is also possible to run an automatic animation (the buttons ">>" or ">>…"). In this mode the tool iteratively shifts the current moment/interval by a specified number of time units (step). The user may control the speed of animation by varying the parameter "delay".

This tool was applied to different kinds of maps presenting different types of data: nature observations and earthquakes (point data), seasonal migration routes (route data), and attribute data referring to territory division (see in the following subsections). More information about our approach to spatio-temporal visualization can be found in our papers [12,13]

All the applets presented in the following subsections use geographical data (vector polygons) in Descartes' internal format (which is derived from SHP file format of ArcView and supported by the utilities of Descartes) and thematic data in CSV (comma-separated value) format which is supported by Excel. A linkage between geographical and thematic data is done through identifiers.

2.2.1 Visualization of instant events: nature observations

Within the project "Naturdetektive" (see URL http://www.naturdetektive.de/) German schoolchildren registered when and where they have seen certain plants or birds. For some plants, different states are distinguished: appearing of first leaves, start of blossoming, appearing of fruits. The observations registered were to be presented in map form. One of the requirements to the presentation was to show distribution of observations in time. For this purpose we suggested the dynamic map display manipulated through the time controls described above. The map display is based upon the treatment of the data as instant events. The events are represented by signs (icons) put on the map. Variation of icons is used to encode qualitative information about the events: what plant or animal was observed and, when appropriate, in which state: leaves, flowers, or fruits.

2.2.2 Visualization of instant events: earthquakes in Europe

Another application was visualization of data about earthquakes (data were provided by Dr. R. Peckham, JRC). In this application the events (occurrences of earthquakes) are represented by geometric signs. Characteristics of earthquakes, such as magnitude, depth, or radius, are encoded by the degree of darkness of the signs: darker shades correspond to higher values. The user may interactively choose which of the available characteristics is to be presented.

The readers can run an applet showing the data about earthquakes in Europe from 1980 to 1983 at the URL http://borneo.gmd.de/descartes/java/show1field/eq.html

2.2.3 Visualization of spatial movement

We applied our tool to an example data set containing telemetric observations of migration of four storks to Africa in autumn 1998 and back to Europe in spring 1999. In our tests we found especially interesting and useful displaying route fragments traversed by the objects during time intervals. During the animation the routes look like worms crawling on the map. It is not merely fascinating; these "worms" help to study important dynamic characteristics of movement. The length of a route

fragment shows the speed of movement. The length being reduced signalizes that the movement of the object slows down, and extension of the "worm" means that the movement becomes faster. When an object stops its movement and stays for some time in the same place, the corresponding "worm" reduces to one point. It would be practically impossible to do such observations using an ordinary animated presentation playing a sequence of images showing states (in this case positions) at successive moments of time.

Demonstrators (the same applet was also applied to other data sets):

- http://borneo.gmd.de/descartes/java/birds/   White storks, 1998-1999

- http://borneo.gmd.de/descartes/java/birds/index2000.html White storks, 1999-2000

- http://www.dialogis.de/cms/   Tracked   Flight   of   Eagle Messengers

- http://www.dialogis.de/gis/moviemap/Arbeitsweg.html Route of a car

- http://www.dialogis.de/gis/moviemap/polarstern.html Route of scientific ship "Polarstern"
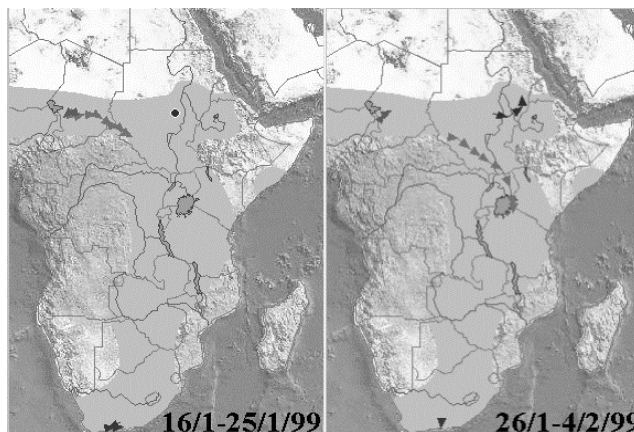

Figure 2. Fragments of interactive maps show different behaviors of white storks in Africa.

A big collection of demonstrators was developed in Oregon State University   Marine   Mammal   Program, http://hmsc.orst.edu/groups/marinemammal/. Some pages include not only the applet, but also descriptions of findings that were made using interactive graphics:

- Blue Whale Migration and Feeding Animation

  - http://hmsc.orst.edu/groups/marinemammal/AnimBlue-a.htm Feeding season off Southern California

  - http://hmsc.orst.edu/groups/marinemammal/AnimBlue-b.htm Southern fall migration

- Hawaiian Humpback Movements and Migration Animation

  - http://hmsc.orst.edu/groups/marinemammal/AnimHump-a.htm Hawaiian Inter-Island movements

  - http://hmsc.orst.edu/groups/marinemammal/AnimHump-b.htm Northbound migration

- Right Whale Movements Animation

- http://hmsc.orst.edu/groups/marinemammal/AnimRight-a.htm Animated Right whale feeding movements
- http://hmsc.orst.edu/groups/marinemammal/AnimRight-b.htm Right Whale #1421 visits a Warm Core Ring

2.2.4 Analysis of attribute data

The same tool (time manager) combined with interactive and dynamic map display is applicable to attribute data referring to a territory division. In this case interesting is consideration of dynamics, or changes in values of the attribute under investigation. We implemented a Java applet which supports different types of comparison:

- comparison with values for previous moment of time;
- comparison with values for selected moment of time;
- comparison with value of a selected object;
- comparison with a specific number.

The applet was developed within a joint project with JRC and Eurostat. Unfortunately, it is not available in the Internet due to restrictions on data dissemination.
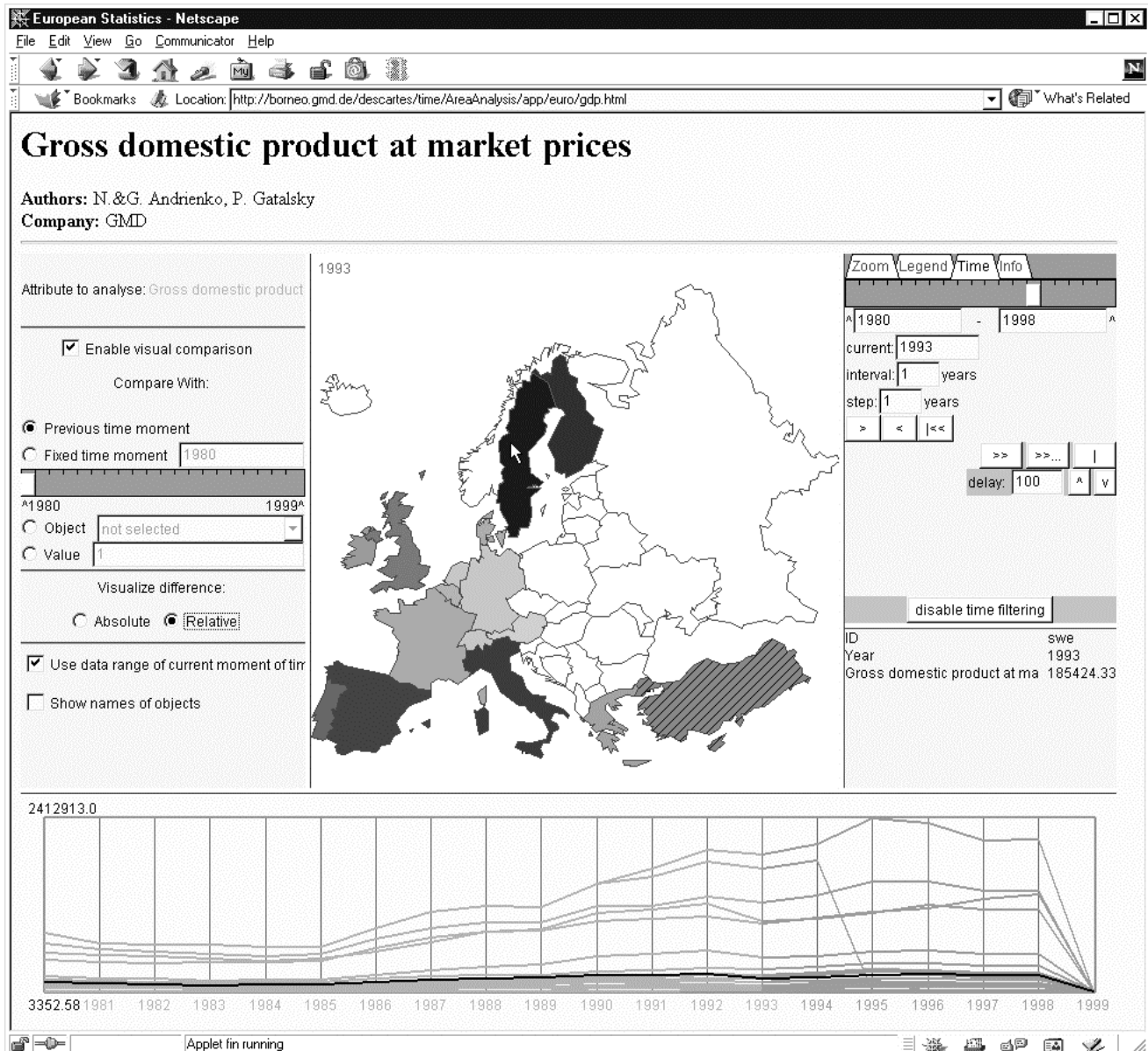
Figure 3. Analysis of change of the gross domestic product in European countries: relative difference between the years 1993 and 1992

3. APPLETS NOT AVAILABLE IN THE INTERNET

We started another two projects to investigate how interactive cartographic visualization can be combined with formal methods of data analysis: data mining and decision support.

3.1 Descartes + Kepler for spatial data mining

To support analysis of spatially referenced data, we are integrating traditional data mining instruments with automated cartographic visualization and tools for interactive manipulation of graphical displays. The essence of the approach is that an analyst can view both, source data and results of data mining in the form of maps that convey spatial information to a human in a natural way. This offers at least a partial solution to the challenges caused by spatially referenced data: the analyst can easily see spatial relationships and patterns that are inaccessible for a computer on the present stage of development. In addition, on the ground of such integration various KDD (Knowledge Discovery in Databases) steps can be significantly supported.

The most evident use of cartographic visualization is in interpretation and analysis of data mining results. However, maps can be helpful also in other activities. For example, visual analysis of spatial distributions of different data components can help in selection of representative variables for data mining and, possibly, suggest which derived variables would be useful to produce. On the stage of data preprocessing, a map presentation can expose "strange" values that may be errors in the data or outliers. Discretization, i.e. transformation of a continuous numeric variable into one with a limited number of values by means of classification, can be aptly supported by a dynamic map display showing spatial distribution of the classes. With such a support the analyst can adjust the number of classes and class boundaries so that interpretable spatial patterns arise.

We built an integrated KDD environment on the basis of two existing systems, Kepler for data mining and Descartes for interactive visual analysis of spatially referenced data. Kepler includes a number of data mining methods and, what is very important, provides a universal plug-in interface for adding new methods. Besides, the system contains some tools for data and formats transformation, access to databases, querying, and is capable of graphical presentations of some kinds of data mining results (trees, rules, and groups).

Kepler also has a client server architecture: the server is responsible for data manipulation and computations, the client provides the user interface. The link between the two systems is implemented using socket connections. There are two level of linking:

1. The servers communicate about data and control flow, i.e. Descartes sends to Kepler table data and makes it run a specific method of data mining.

2. The clients communicate about user's actions: Descartes informs Kepler about selection of objects on maps or statistical graphics displays, Kepler informs Descartes about user's actions in windows with results of data mining (for example, focusing on a node of a tree).

The general idea of our approach to spatial data mining and software architecture of the integrated system are described in our paper [14]. Specific information concerning visualization support for C4.5 classification tree method are presented in [15].
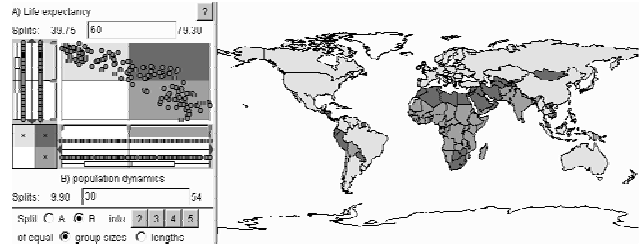
Figure 4. Cross-classification of countries of the world on the basis of two attributes, "Birth rate" and "Life expectancy". The light shade corresponds to countries with high life expectancy and small birth rate, middle dark to countries with small life expectancy and high birth rate, and the darkest to high values of both attributes. The scatter-plot on the left is dynamically linked with the map through simultaneous highlighting of objects pointed at by the cursor either in the map or in the plot.
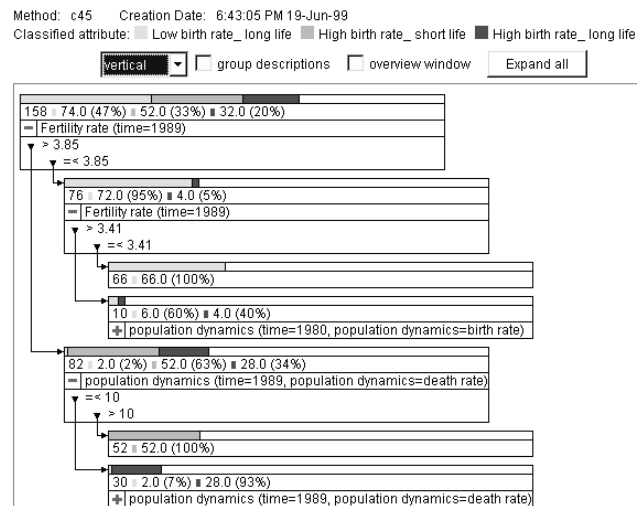
Figure 5. Classification tree calculated for the map in Figure 2. It shows that all 52 countries with high birth rate and small life expectancy can be characterized by fertility rate higher than 3.85 and death rate more than 10.

3.2 Decade for spatial decision support

In August and September 1999 our group was hosting Prof. P. Jankowski from Univ. Idaho. Together with him, we studied how interactive maps and data mining techniques can support spatial decision making procedure. For this purpose, two methods (Ideal point and Aspiration level) of multicriteria optimization (MCO) were implemented in Java and. integrated into Descartes. We found, that maps could act as a key element in spatial decision support [16]. For this purpose, maps should be dynamic and interactive. Maps are useful for initial studying of data, preparation of input variables for MCO, and for analysis of results. Another interesting finding was that using data mining methods, it is possible to select most important criteria for decision making and in this way decrease the dimensionality of the problem [17].

Further research is planned in this direction. In particular, we want to extend our approach for collaborative spatial decision support.

4. CONCLUSIONS AND FUTURE WORK

Big variety of systems and applets mentioned in the paper forced us to think about a component architecture for the software. Such architecture is under development now. We hope that implementation of it will allow us to combine tools into specific task oriented systems without software reengineering.

Knowledge based automated visualization design in Descartes and CommonGIS systems can be extended to more general support in problem solving. Some more intelligent systems could guide end-ures through the whole process of solving a data analysis problem: selection of data (including attribute, spatial and temporal aspects), preprocessing of data, visualization design, use of interactive maps for problem solving, and, finally, reporting (or communicating) the findings [18].

Within such a system, intelligent guidance can also be provided in selection and setting up of mathematical methods of data mining or spatial statistics. We plan to develop specific visualization methods to support viewing and interpretation of results of data mining and spatial statistics. Such visualization can be enhanced by interactivity and dynamics of graphic displays. Data mining-related issues will be further investigated within the EU-funded SPIN project (URL: http://www.ccg.leeds.ac.uk/spin/, IST Program, project No. IST-1999-10536, January 2000 - December 2002). As to spatial statistics, we plan to do a joint project together with Prof. R. Haining and Dr. S. Wise from University of Sheffield.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1]MacEachren, A.M. "How Maps Work: Representation, Visualization, and Design" (NY: The Guilford Press), 1995

[2]MacEachren, A.M. "Visualization in modern cartography: setting the agenda". In Visualisation in Modern Cartography (NY: Elsevier Science Inc.), pp.1-12, 1994

[3]MacEachren, A.M. and Kraak, M.-J. "Exploratory cartographic visualization: advancing the agenda". Computers and Geosciences, 23 (4), pp.335-344, 1997

[4]Andrienko, G., Andrienko, N., Voss, H., and Carter, J. "Internet Mapping for Dissemination of Statistical Information", Computers, Environment and Urban Systems, 23 (6), pp.425-441, 1999

[5]Andrienko, G. and Andrienko, N. "Intelligent Cartographic Visualization for Supporting Data Exploration in the IRIS System", Programming and Computer Software, 23 (5), pp.268-282, 1997

[6]Andrienko, G. and Andrienko N. "Interactive Maps for Visual Data Exploration", International Journal Geographical Information Science, 13 (4), pp.355-374, 1999

[7]Andrienko, G. and Andrienko, N. "Data Characterization Schema for Intelligent Support in Visual Data Analysis", In Freksa, C., & Mark, D. M. (eds.) Spatial information theory - Cognitive and computational foundations of geographic information science COSIT'99, Lecture Notes in Computer Science, vol. 1661. Berlin: Springer, pp. 349-366, 1999

[8]Andrienko, G. and Andrienko, N. "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, Kansas-City, November 5-6, 1999, NY: ACM Press, pp.66-72

[9]Andrienko, G. and Andrienko, N. "Dynamic Categorization for Visual Study of Spatial Information", Programming and Computer Software, 24 (3), pp.108-115, 1998

[10]Andrienko, G. and Andrienko, N. "Exploring spatial data with dominant attribute map and parallel coordinates", In Proceedings of GIS Research UK'2000, York, April 5-7, 2000, pp.104-109. Also to be published in "Computers, Environment and Urban Systems", special issue on GIS Research UK'2000, 2001, v.25 (accepted)

[11]Andrienko, G. and Andrienko N. "The Architecture of the CommonGIS System (ESPRIT Project 28983)", Proceedings of 5th EC-GIS Workshop, JRC ISIS, European Communities, EUR Report, 2000

[12]Andrienko, G., Andrienko N., and Gatalsky, P. "Spatio-Temporal Visualization in Naturdetective and Beyond", K. Fullerton (ed.) Proceedings of 5th EC-GIS Workshop, Stresa, Italy 28-30 June, 1999. Report EUR 19018 EN, European Communities, 2000, pp. 499-507

[13]Andrienko, N., Andrienko, G., and Gatalsky, P. "Supporting Visual Exploration of Object Movement", V. Di Gesu, S. Levialdi, L. Tarantino (eds.) Proceedings of the Working Conference on Advanced Visual Interfaces AVI 2000, Palermo, Italy, May 23-26, 2000. ACM Press, 2000, pp.217-220, 315

[14]Andrienko, G. and Andrienko, N. "Knowledge-Based Visualization to Support Spatial Data Mining", In Hand, D.J., Kok, J.N., and Berthold, M.R. (Eds.), Advances in Intelligent Data Analysis, 3rd International Symposium, IDA-99, Amsterdam, The Netherlands, August 9-11, 1999, Proceedings. Lecture Notes in Computer Science, vol. 1642. Berlin: Springer-verlag, pp.149-160, 1999

[15]Andrienko, G. and Andrienko, N. "Data Mining with C4.5 and Cartographic Visualization", In N.W.Paton and T.Griffiths (eds.) User Interfaces to Data Intensive Systems, IEEE Computer Society, Los Alamitos, CA, pp.162-165, 1999

[16]Jankowski, P., Andrienko, N., and Andrienko, G. "Map-Centered Exploratory Approach to Multiple Criteria Spatial Decision Making", International Journal Geographical Information Science (accepted)

[17]Andrienko, G., Jankowski, P., and Andrienko, N. "Data mining approach to reducing cognitive complexity of multiple criteria spatial decision problem", In Proceedings Spatial Data Handling SDH'2000 (accepted)

[18]Andrienko, G. and Andrienko, N. "Making a GIS Intelligent: CommonGIS Project View", In AGILE'99 Conference Proceedings, Rome, April 15-17, 1999, pp.19-24

About the authors

Drs. Natalia and Gennady Andrienko received their PhD equivalent from Moscow State University in 1993 and 1992 respectively. Since 1997 they are working as researchers in German National research Center for Information Technology. They are members of the Visualization Commission of the International Cartographic Association. Their research interests are interactive computer graphics, visual data exploration, knowledge engineering, Internet programming

Tel. +49-2241-142486, -142329

Fax +49-2241-142072

Snail mail:

Drs. Andrienko
GMD AiS.KD
Schloss Birlinghoven
Sankt Augustin
53754 Germany

WWW http://borneo.gmd.de/and/

E-mail: **gennady.andrienko@gmd.de**