Exploratory Data Analysis and Decision Making with Descartes and CommonGIS

Hans Voss, Natalia Andrienko, Gennady Andrienko

Fraunhofer Institut Autonome Intelligente Systeme, FhG-AIS, Schloss Birlinghoven, Sankt-Augustin, D-53754 Germany, Tel: +49-2241-14-2532, -2329, Fax: +49-2241-14-2072, E-mail: {Gennady.Andrienko, Natalia.Andrienko, Hans.Voss}@ais.fraunhofer.de, URL: http://www.ais.fhg.de/KD/

Abstract. Our thematic mapping software Descartes has been steadily extended and improved over the past eight years. When it became available in the year 1994 it was the first fully interactive, explorative tool for spatial analysis of statistical geo-referenced data available for use in the Internet. Interactivity was achieved by Java-Applet technology. Beside its powerful presentation and exploration functions one primary research and practical goal of its development continuously was to incorporate cartographic knowledge into the software. Thus expert users are freed from routine activities and casual users are enabled to utilize the functions without much practice. In the previous three years, the internal architecture and implementation of Descartes was completely revised. As a result, it is now very easy to build new system configurations including particular sets and variants of functions. One subset of Descartes' functions has been placed in the Internet for free download and non-commercial use. This configuration is named CommonGIS, according to the EU-funded project of the same name, which provided financial support for creating a GIS accessible and usable for expert and common users alike.

1 Introduction

The thematic mapping software Descartes is available as a stand-alone tool for running on a local PC, and as a web version with client-server architecture. Actually, it was the first thematic mapping software that one could interactively run in the web from within a standard Java-enabled browser when it became first available in 1994. In the web version the server performs knowledge-based design of maps, and sends specifications of appropriate visualization methods to the client. The client software, which is a Java Applet, generates maps according to the specifications and provides various related interactive tools.

Funding of the European Union, with project CommonGIS, resulted in a complete revision of the original Descartes software. In particular, it became completely component oriented, configurable, and embeddable into other software systems. A particular assembly of functions, comprising most of the available visualization methods, but not, for example, data base connectivity, was made available for free download and non-commercial use at www.commongis.de. In the following, we will refer to CommonGIS as the "base system", and only mention Descartes when referring to functionality that is not available in CommonGIS.

CommonGIS can be perceived as an interactive Web-GIS, because it provides some standard GIS functionality, but particularly it is a tool for the visualization, exploratory analysis, and decision support based on geographically referenced statistical data.

Working with CommonGIS functions according to the following scenario: Accessing CommonGIS using a standard web browser, a user may select an application (a territory with associated thematic data). Within the application one may repeatedly choose one or more data variables for analysis. Each time, CommonGIS will **automatically** offer one or more adequate thematic maps visually presenting the data. The maps built by the system comply with sound principles of graphical and cartographical representation of information, i.e. they correspond to (1) characteristics of the data (e.g. whether a variable is qualitative, ordinal, or numeric) and (2) relationships between variables (e.g. comparability or inclusion). Moreover, in designing maps the system also may take into account the analytical goals of the user. As a result, one or more map displays will be created which are cartographically sound, on the one hand, and useful, on the other hand.

The digital maps are highly interactive and dynamically transformable, according to the concept of "geographic visualization" and EDA. For every type of representation method (pie charts, bar charts, choropleth maps, etc.) several interactive techniques are realized, all of them specifically designed to increase the analytical power of the respective representation. The interactivity of the map displays on a web client is achieved by Java and Java Applet technology.

2 Functionality

CommonGIS provides many methods for the visual presentation and analysis of data, either based on maps or using various types of statistical graphs. One outstanding feature of CommonGIS is its control of brushing, where the user may inspect related data by simultaneous highlighting in various linked displays. Here is a summary of available functions:

Data transformations ("calculation")

Classification of one attribute

Classification of two attributes (cross classification)

Classification by dominant attribute

Classification by n-dimensional attribute similarity

Ideal point evaluation (Fig. 3)

Ordering of values

Average/Median/Quartiles/Variance

Calculation of arbitrary formula

Filter objects according to specified attribute values

Map visualizations ("mapping")

Choropleth maps for one numeric attribute

Choropleth maps for one qualitative attribute

Choropleth maps for cross classification

Choropleth maps for other classifications (Fig. 2)

Multiple choropleth maps

Diagrams: utility bars

Diagrams: utility pies

Diagrams: normal bars

Diagrams: normal pies

Diagrams: stacked bars

Diagrams: triangles for two attributes

Graph visualizations ("charting")

Dot plot

Scatter plot

S-plot matrix

Parallel coordinate plots (Fig. 3)

Tukey's box plots

Histograms

Basic map functions

Zoom out (to maximal extent)

Zoom out (by factor)

Zoom in (with rectangle)

Zoom in (by factor)

Undo last zoom

Shift the map

Select mouse mode (zoom, pan, select or explore)

Change the order of layers

Change layer properties

Change the size and color of symbols

Windows functions

Open new window with current map

Minimize all

Restore all

Close all

Bring an open window to the front

Support Functions

Help functions

General system help

Task support guide (support based on characteristics of data and intended goal of the user)

General information about system and application

Other functions

Show advanced menu items

Show menu items for application building

Range of Data Input and Output possibilities

Data input

Load attribute data from:

DBF, CSV (Excel), TXT (delimited text), ODBC/JDBC, clipboard.

Load geographical data from:

OVL (GMD Descartes), SHP (ESRI Shapefile), JPEG, GIF, FLT (grid data), WKB, Simple Features and GML (OpenGIS)

Open a pre-defined map description MWI

Data output

Display data records on mouse-over

Edit options for display of data records

Print map

Save map as image

Save application

Select objects by mouse-click

Find objects according to specified attribute values

Display table with all objects

---

3 Architecture

The overall system architecture is depicted in Figure 1. The many dots (…) in the picture symbolize the extensibility of the system in various dimensions. In particular, one may add new visualization methods, which can be put under the highlighting/brushing regime of a central supervisor component, which receives mouse pointing or clicking events from registered displays and broadcasts them to other registered displays.

4 Applications

CommonGIS/Descartes is being used in various applications developed inside and outside our organization. For example, it was the basis for the visualization and analysis of statistical and other thematic data of

- Several cities, such as Bonn, Helsinki, Tilburg;
- Eurostat, the European Statistical Office; here also the visualization of time-series data was addressed.
- The UK National Statistical Office; addressing the visualization of the previous census decades.

- European Forest Institute, including statistical data from various sources such as Eurostat, European National forest institutes.
- German Office for Nature Protection (BfN) in project “Naturdetektive”: German children are requested to enter and analyze observations on interactive maps, and to study the flyways of GPS equipped birds (storcks and cranes).

Some of these and further applications can be found as live on-line examples at http://borneo.gmd.de/and/java/iris/.

Some Publications

Andrienko, G. and Andrienko, N. (1999a) Interactive Maps for Visual Data Exploration. *International Journal Geographical Information Science*. v.**13** (4), pp.355-374.

Andrienko, G. and Andrienko, N. (1999b) 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.

Andrienko, G. and Andrienko, N. (1999c) 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.

Andrienko, G. and Andrienko, N. (1999d) Making a GIS Intelligent: CommonGIS Project View. AGILE'99 Conference, Rome, April 15-17, 1999, pp.19-24.

Andrienko, G. and Andrienko, N. (2001a) Exploring Spatial Data with Dominant Attribute Map and Parallel Coordinates. *Computers, Environment and Urban Systems*. v.**25** (1), pp.5-15.

Andrienko, G. and Andrienko, N. (2001b) Interactive Cumulative Curves as a Tool for Exploratory Classification. In D.B.Kidner, G.Higgs (Eds.) *9th annual conference GIS Research in the UK*, University of Glamorgan, Wales, April 18-20, 2001. University of Glamorgan, 2001, pp.439-442.

Andrienko, G. and Andrienko, N. (2001c) 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.

Andrienko, G. and Andrienko, N. (2001d) Interactive Visual Tools to Support Spatial Multicriteria Decision Making. Proceedings UIDIS 2001, Zurich. IEEE Computer Society Press.

Andrienko, G., Andrienko, N., Voss, H., and Carter, J. (1999): Internet Mapping for Dissemination of Statistical Information, Computers, Environment and Urban Systems (Elsevier Science), 1999, v.23 (6), pp.425-441, ISSN 0198-9715

Jankowski, P., Andrienko, N., and Andrienko, G. (2001) Map-Centered Exploratory Approach to Multiple Criteria Spatial Decision Making. *International Journal Geographical Information Science.* v.**15** (2), pp.101-127.

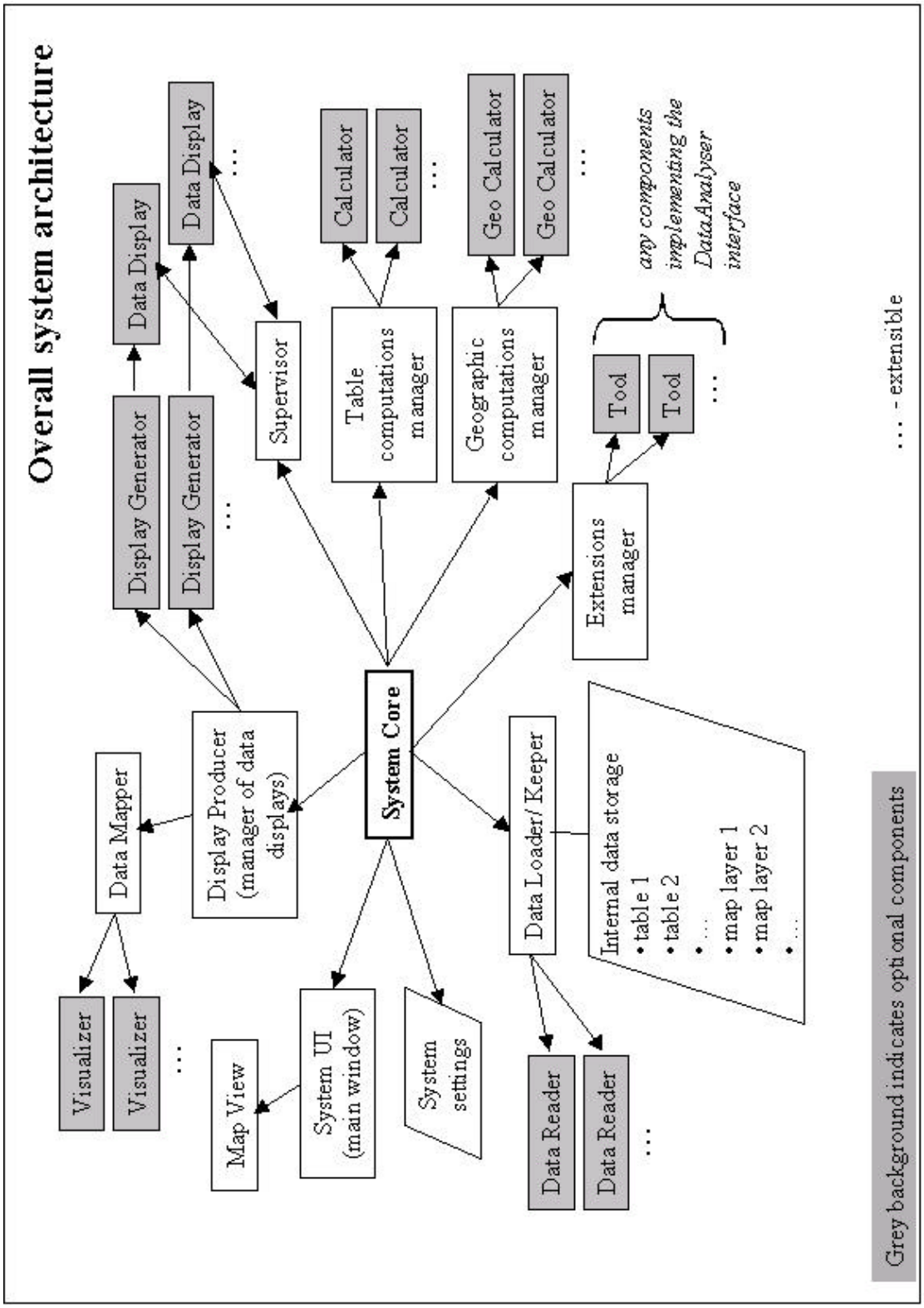

Figure 1. Overall System Architecture

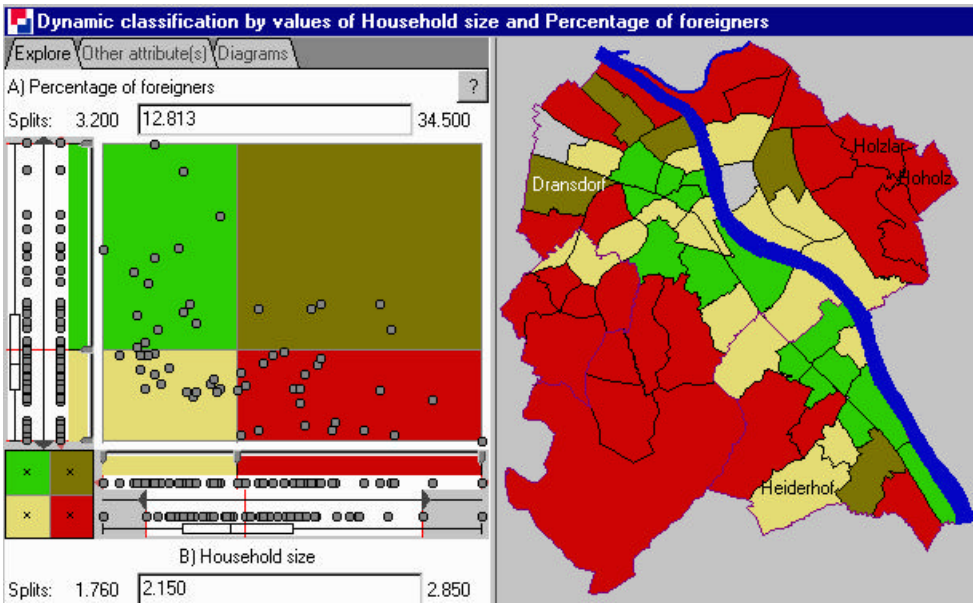

Fig. 2: A scatter plot showing a weak inverse correlation between percentage of foreigners and mean household size for the City of Bonn.

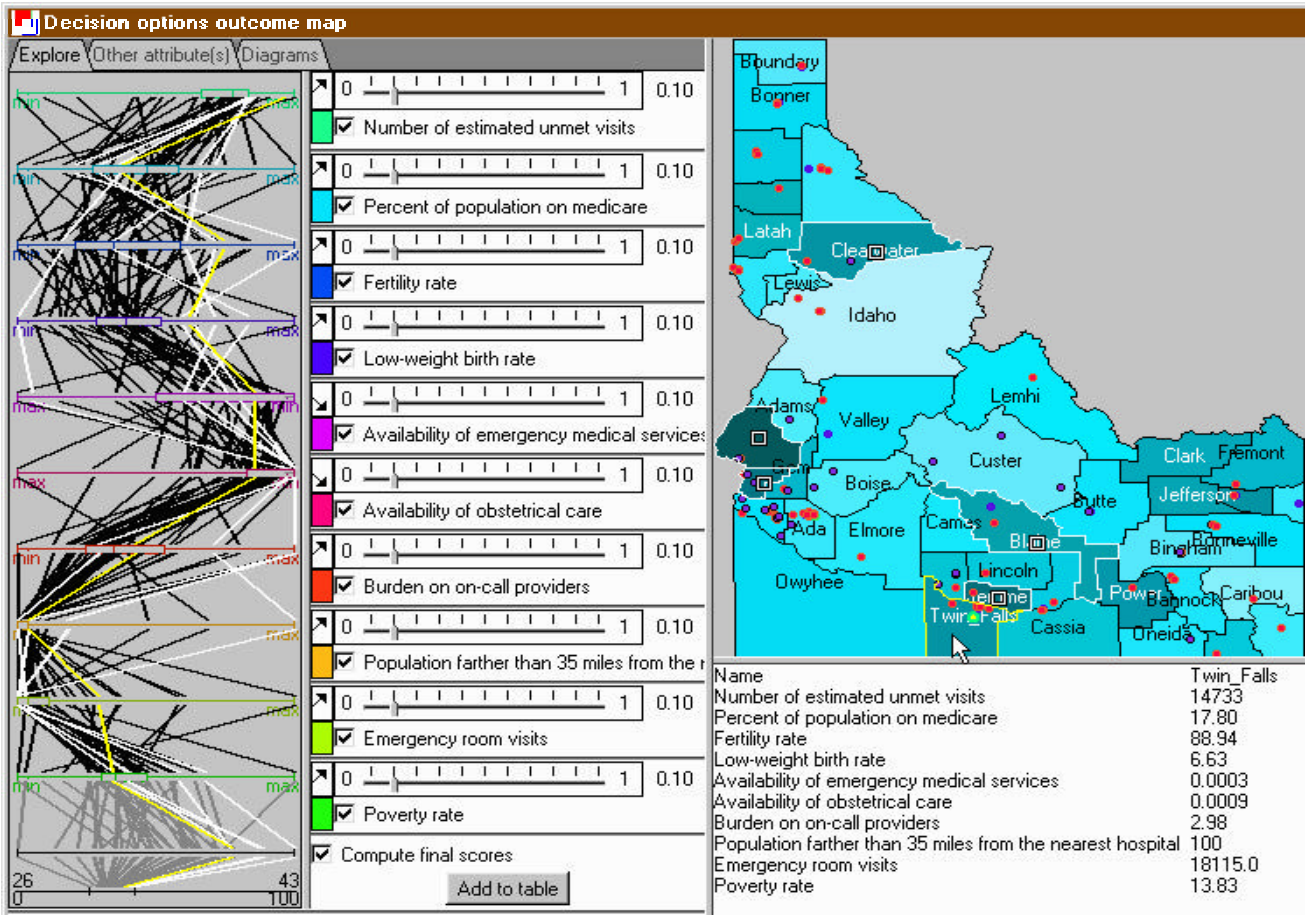

Fig.3: Parallel coordinate plot with prioritized attributes for counties of Idaho. The map shows the ranking of the counties according to the integrated weighted criteria.