Common-GIS: a European project for an easy access to geo-data

Ursula Kretschmer
Fraunhofer-Institute for Computer Graphics (IGD)
Rundeturmstrasse 6
D – 64283 Darmstadt, Germany
ursula.kretschmer@igd.fhg.de

Emanuele Roccatagliata
GISIG - Geographical Information Systems International Group
Via Piacenza, 54
I - 16138 Genova, Italy
gisig@gisig.ima.ge.cnr.it

Abstract. The ESPRIT project CommonGIS[1] supports users by visualizing and analyzing spatially-related statistical data. The data must be entered by a special tool to structure it in an intelligent way. This is the pre-condition for the automatic generation of maps and statistical graphs. As part of the project a data characterization scheme was developed. The user is guided by the system to get a visualization fitting not only to the data but also to a specific problem. To suit the presentation to personal needs interaction techniques are offered to the user.

Introduction

In the area of historic and demographic education the work with quantitative and statistic methods is very common. To answer a historic or demographic question a multitude of different sources must be evaluated, for instance: maps, telephone surveys, public-opinion polls, or collection of historic information like telephone books. The work to be executed can be enabled by use of a Geographical Information System (GIS). Examples would be: connection of available data, evaluation of working hypotheses, and presentation of results.

These systems offer the visualization of spatial data using a broad range of techniques, such as topographic and thematic maps, statistical graphs, perspective rendering of surfaces, and animation. The requirements of the system for the employment in the area of demography and history should be visualization of statistical data as well as an easy to use interface so that users without special knowledge can handle the tool. The purpose of these techniques is to facilitate the understanding and interpretation of spatial data and spatial processes.

[1] URL: http://commongis.jrc.it. The project is partly funded by the European Commission, DGIII, contract No. 28983 (November 1998 – April 2001)

The goal of the ESPRIT project CommonGIS is the development of a system allowing users to view and analyze geographically referenced thematic data. The system is not only oriented to professional GIS users but also to the general public, i.e. people without special training and expertise in map design. This requires understanding of data semantics that, hence, must be formally represented. After selecting a dataset several visualization techniques suiting to the chosen data are offered to the user. The system explains which possible task fits to a visualization method. Therefore, no special background either in statistics or in cartography is required to get truthful results.

The following paper presents in the second and third section the objectives as well as the partnership of the project CommonGIS. An introduction of the data characterization scheme, i.e. the intelligent kernel of the system developed within the project, is given in paragraph four. Paragraph five describes the status of the project after one year. The application of the system in the field of education is presented in the sixth section including an example showing the interaction facilities. Paragraph seven summarizes the possibilities offered by the CommonGIS system.

Objectives

At the moment there is a high demand in many areas for spatially related data. By help of an information system it can be analyzed and by use of the internet it can be presented to a wide audience. The superior goal of CommonGIS is the access and usability of geo-data for everyone from everywhere. Not only a tool to publish information in form of coordinates as well as geometry of spatial objects is needed but also the possibility to submit data in an intelligent way through an exploration tool. Visualization techniques in form of maps and statistical graphs enhance the work with the data. This should be done without the necessity to afford a current expensive GIS software and to spend much time in learning how to operate it. A solution could be a WWW-based GIS only requiring a standard Java-enabled Web browser which automatically generates thematic maps. CommonGIS meets these conditions. By using this system a remote and effortless access to geo-data is possible. As the system applies to casual users without knowledge in graphic and cartographic presentation, the system offers them easy to handle options and assistance to explore and analyze generated maps of the selected geo-data sets.

One way of using CommonGIS is to specify a data set and let the system design visualizations on its own. The user is provided with information why a particular design is suitable for a specific data set in combination with an explicit task and which additional interaction techniques influence its effectiveness favorably. One of these proposed tasks could be, for example, “compare proportions of population of age group in different districts”. The user chooses one visualization method best suited to a special aim and the system calculates the chosen statistical presentation eventually by performing any data transformations and then guides the user by using special interaction facilities to reach the target. Interactive manipulation tools enable the user to adjust the system to personal needs and to perception abilities by map modification.

Partnership

The partners are research institutes as well as enterprises. CNIG (Portugal) is an institution responsible for running the Portuguese network of geo-referenced data. JRC (Italy) contributes its knowledge about developing distributed applications. GMD (Germany) further developed the IRIS system, which is an advanced prototype for the automatic generation of thematic maps. Dialogis (Germany), an independent GMD spin-off, is in the process of producing a commercial version of IRIS, the product Descartes. Fraunhofer-IGD (Germany) has built a software prototype for thematic mapping called VIZARD. PGS (Netherlands) has built a Java-based GIS, which was integrated with Descartes within the project and will appropriately be extended for the specific needs of this project. The European GIS association GISIG uses its excellent dissemination channels to promote the results of the project and to help in the standardization effort of the data characterization scheme. The latter is required to inform the system about the relevant characteristics of the data necessary to make appropriate presentation decisions. The developments of the project base on rich experience of the partners in computer graphics, GIS, knowledge-based systems, and data visualization. This experience is reflected in the existence of the elaborated systems which are all brought into the project as background and basis for further development and integration.

Data characterization

To enable the automatic generation of presentations the system needs information about the data. To structure it a data characterization scheme was developed based on the data characterization schemes of the existing systems Vizard and Descartes. The synthesis of these data characterizations into a consistent and comprehensive structure was initiated and a data characterization scheme for CommonGIS was designed. According to this scheme, a two-stage mapping from table data as received from data providers to a conceptual, domain specific description was realized. An overview of the data characterization scheme is shown in Figure 4.1.

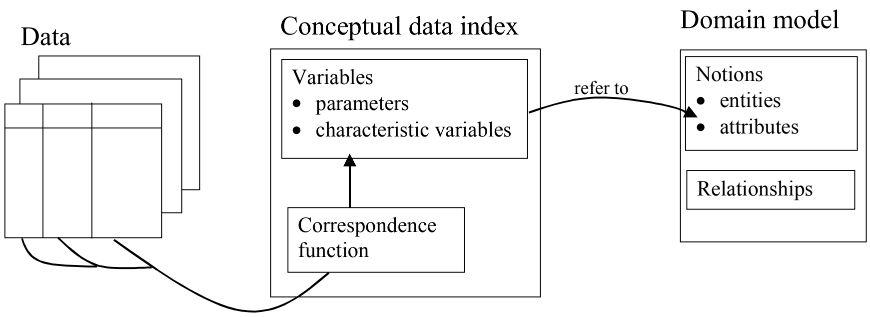


Figure 4.1: Data characterization scheme

In a first step, i.e. the data index, the data columns are divided into parameters and characteristic variables. Parameters define the context of obtaining the data. Characteristic variables represent for example results of measurements, observations, or calculations obtained in the given context. The link to the tables is explicitly given by correspondence functions. The domain model is a collection of notions necessary to describe the meaning of a spatially referenced data set. The notions are linked by relationships. Examples of domain notions are “whole population” and “female population” in a demographic application. These two notions are associated by the inclusion relationship. The conceptual data index establishes links between components of the data and these notions. These links allow the system to interpret the data and to know relationships among data components. This is necessary to provide the software system with knowledge about the data to enable intelligent and user-friendly support in visual data analysis.

First prototype

The two existing systems Descartes on the one hand and Lava/Magma on the other were merged. Descartes developed by GMD is a system automatically producing high-quality thematic maps for statistical data. It applies cartographic knowledge about how to present statistical data in thematic maps, encoded as heuristic rules, together with characterization of the statistical data and the underlying application domain, encoded using an object-oriented knowledge representation language. Lava/Magma, created by PGS is a GIS solution providing a sophisticated architecture to retrieve geo-data through the internet. Being meta-data driven, Lava offers various operations on objects previously defined for them. This includes the automatic invocation of applets provided by the data supplier. On the server side the program Magma receives cgi requests from the http-server, and performs SQL-queries on a geospatial database. To decide which functionality of the systems should be adopted for the CommonGIS system user requirements had to be determined. For that task both systems were checked concerning interface and functionality. The test participants were encouraged to tell what features they would like to have integrated in a future prototype. Based on this evaluation the prototype starts with the presentation of a map as implemented in the Lava/Magma system instead of offering data preparation as realized in Descartes which seemed too sophisticated for most of the users. Figure 5.1. shows the interface of the prototype presenting Portuguese census data.

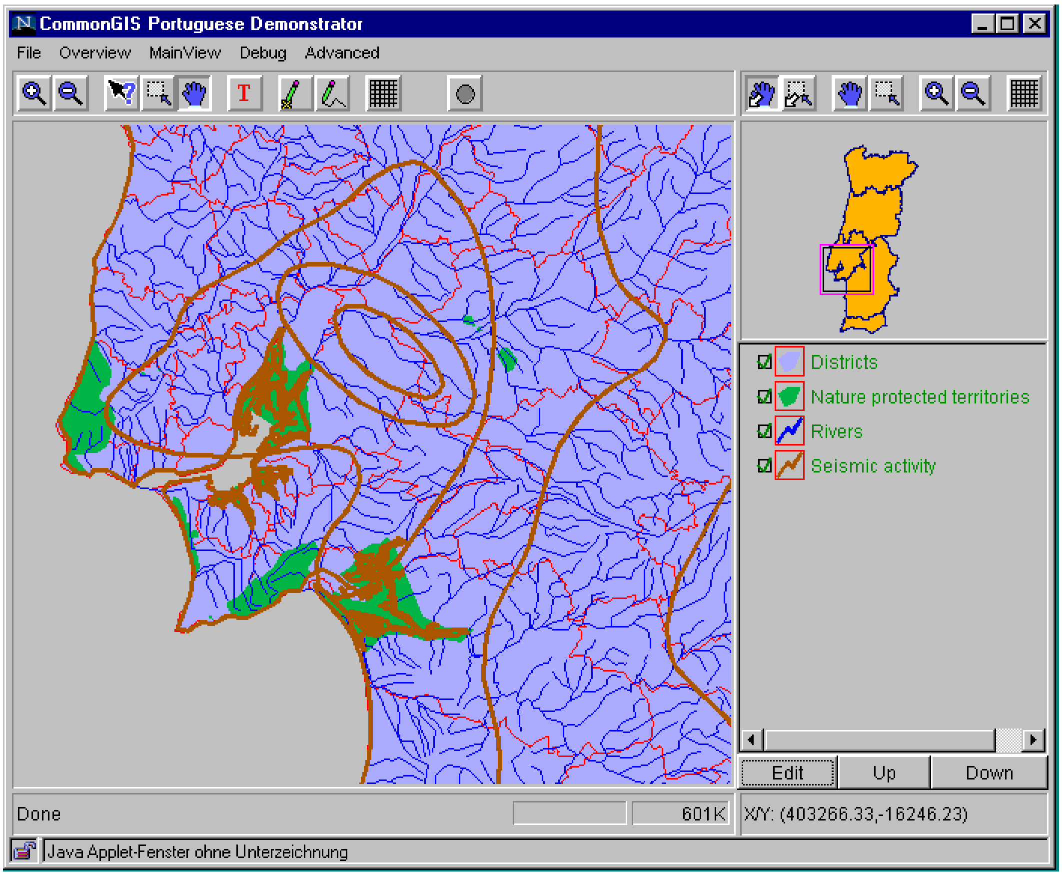


Fig. 5.1 Start interface of the CommonGIS prototype

As the claim of CommonGIS is user-friendliness the powerful tools are presented to the user only in a second step. The first prototype enables the display of visualizations of thematic data using methods and facilities taken from Descartes. The first prototype reflects the status of the project after the first year. It is capable of running the map design module and user interface components of Descartes within the Lava/Magma environment.

Within the second year enhanced functionality will be incorporated into the system. The number of visualization techniques will be increased and the user guidance will be improved. At the moment the user gets a short suggestion what to do with the visualization techniques offered by the system. This is not yet the guidance the system would like to offer. Elaborated explanations are needed and will be integrated into the advanced version of the system.

Education in CommonGIS

Background

To learn at school how to handle statistical data is very important. Not only for the interpretation of results in sciences this knowledge is essential but also for the formation of an independent opinion as a citizen in a democracy it is indispensable.

Biological and psychological experiments serve to get results comparable to experiments in laboratories. As dependencies often can not be experimented but must be simulated by calculating them, the result is not expected to cover all factors. The same is to say for economic and sociological phenomena. Here, the possibility to work under controlled conditions is much more restricted than in most other sciences. To get a meaningful result, material has to be collected, correlations be calculated and time series be analyzed (cf. Pfanzagl). To do so, the student needs not only knowledge about calculation methods but also basic knowledge about how to interpret results.

Furthermore, this information is essential for dealing with media. Newspapers, radio, TV, or the internet try to influence opinions among others by interpreting statistics in one way or the other. Likewise, it is a fact that maps can manipulate opinions. They are just an image of reality, as the three-dimensional world is mapped onto a flat sheet of paper or a screen. Hence, nobody can map the world entirely. The author must limit the presented entities to the ones expressing the message sent to the viewer. Generalization and symbolization are used to emphasize this implication (cf. Monmonier). As not all influences can be uncovered, maps and statistical interpretations are supposed to be watched with reservation. With basic knowledge about how to interpret maps and statistical graphs, they can be examined as nearly objective data. Therefore, misleading interpretations appertaining to them can be determined.

Practices with statistical data are supposed to be learned at school. Subjects as social studies, geography or history guide the students to the ability to handle this data. Here, the employment of easy to use systems support teachers by preparing datasets to visualize specific problems to the students. Dependencies between several features can be presented by the use of graphs. If the data is geo-referenced an underlying map eases the recognition for the students.

The system CommonGIS offers the possibility to store geo-referenced data. The teacher needs to do this by using a specific tool. This process enables the system to get information about the structure of the data. This is used to offer adequate visualization methods to the user. After the teacher had prepared the data the students get the opportunity to choose different visualization methods combined with several tasks. The teacher should guide the students by preparing presentations as well as by interpreting the results. To introduce tasks like comparison of factors or recognition of trends the students should learn how to pick out relative and absolute values, to choose a method to visualize them, and to interpret the results. This enables the students to learn how to deal with maps and statistical data. They can experience how they can be misled by hidden information or by false conclusions deceiving by the presentation.

Example

As an example Portuguese census data is made available for the CommonGIS demonstrator. It consists of data from the years 1981 and 1991 and includes numbers of population groups divided by age, gender, or education. The data is offered to the students in form of tables. An extract is shown in Figure 6.1 below. Each column represents an object that can be visualized by the system.

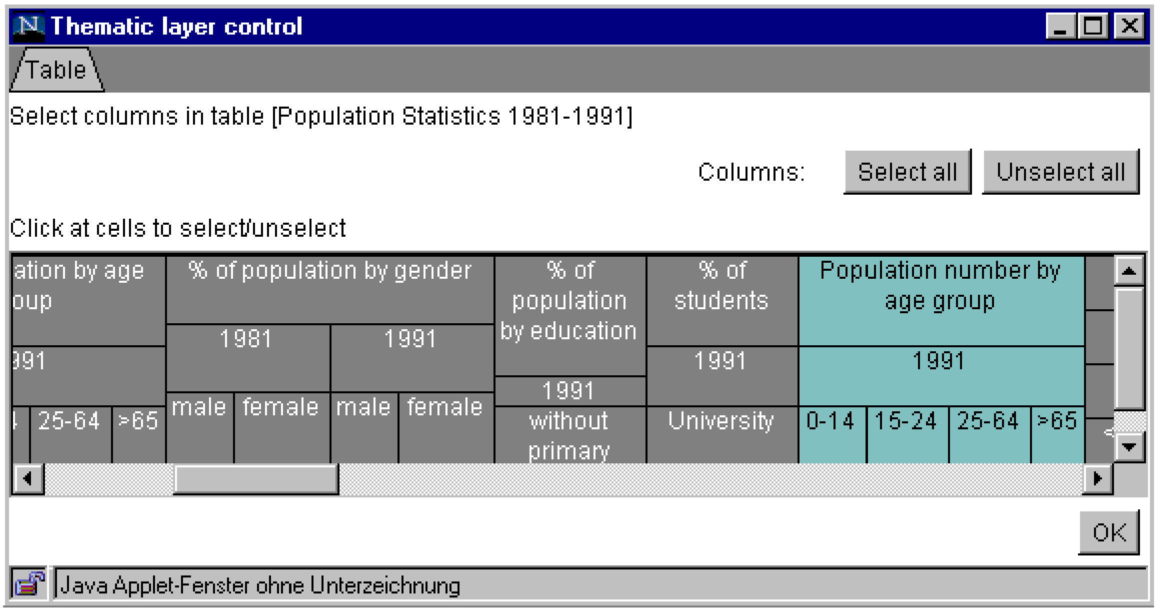


Fig.6.1: Data selection from table

As a first step the student should pick out one or more columns of the tables, i.e. features being visualized by the system. A choice could be ‘Population number by age group’. The respective columns are painted in green whereas all the others remain gray. After choosing data the system offers the user the possibility to visualize these features. In the described case pie charts or bar charts fit to the chosen data. The possible visualization methods are listed on the screen shown in Figure 6.2.

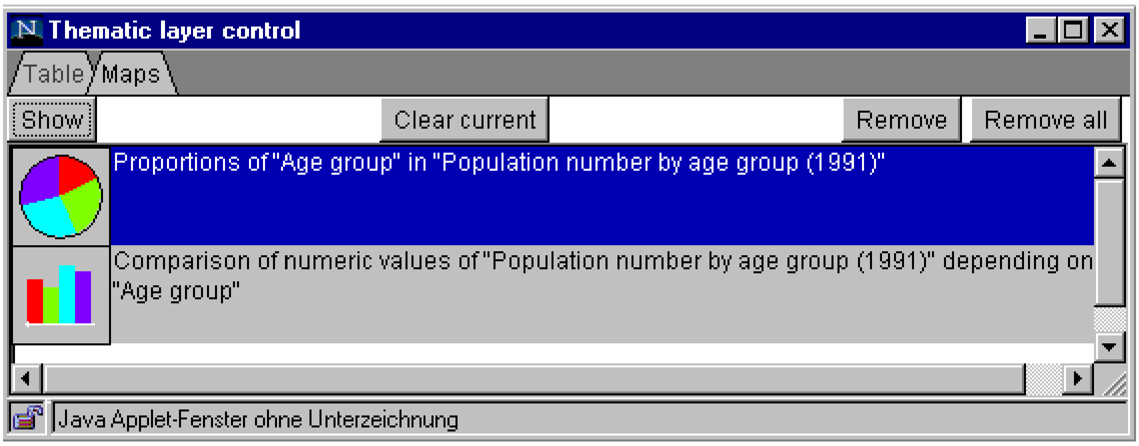


Fig.6.2: Visualization methods for the chosen data set

The chosen example, i.e. numeric data, leads to the visualization using bar charts. This diagram should be chosen, among others, to compare, find, or read values. As the data was handed to the system using an input-tool based on the characterization scheme, CommonGIS recognizes these relationships and automatically executes data transformations like the derivation of relative from absolute values. In the described example the selected data is linked by the relationship "included in a common total". As a consequence, the system automatically transforms the absolute values into relative ones. This leads for the visualization by use of pie charts.

For the first prototype the user guidance within the window listing possible visualization techniques is restricted to a short sentence describing the selected values and occasionally proposing a task to be fulfilled by the method. An example are the explanatory sentences in Fig.6.2: 'Proportions of "Age group" in "Population number by age group (1991)"' and 'Comparison of numeric values of "Population number by age group (1991)" depending on "Age group"' The selection of the visualization methods depends on the data characteristics including relationships. If the latter are heterogeneous or if some are incomparable, they are separated automatically into groups with elements connected by homogenous relationships.

The user guidance in the advanced version will be more detailed. Not only a sentence with a proposed task is offered to the user, but also explanations how to set in interaction techniques within a visualization method to reach more goals than foreseen by use of only static statistical maps. As an example pie charts are considered. They are used among others to compare amounts of values. The presentation of only these charts in a map can not be used to get exact values. CommonGIS offers the possibility to point in a special area. As a result the respective precise values stored in the table are presented in another window to the user. The user should be guided through these possibilities to use as few methods as possible to reach as plenty goals as possible. These interactive manipulations offer the possibility to watch the behavior of data.

To get a general view of absolute values, bar charts are selected. For the chosen data the map shown in Figure 6.3 is presented to the user.

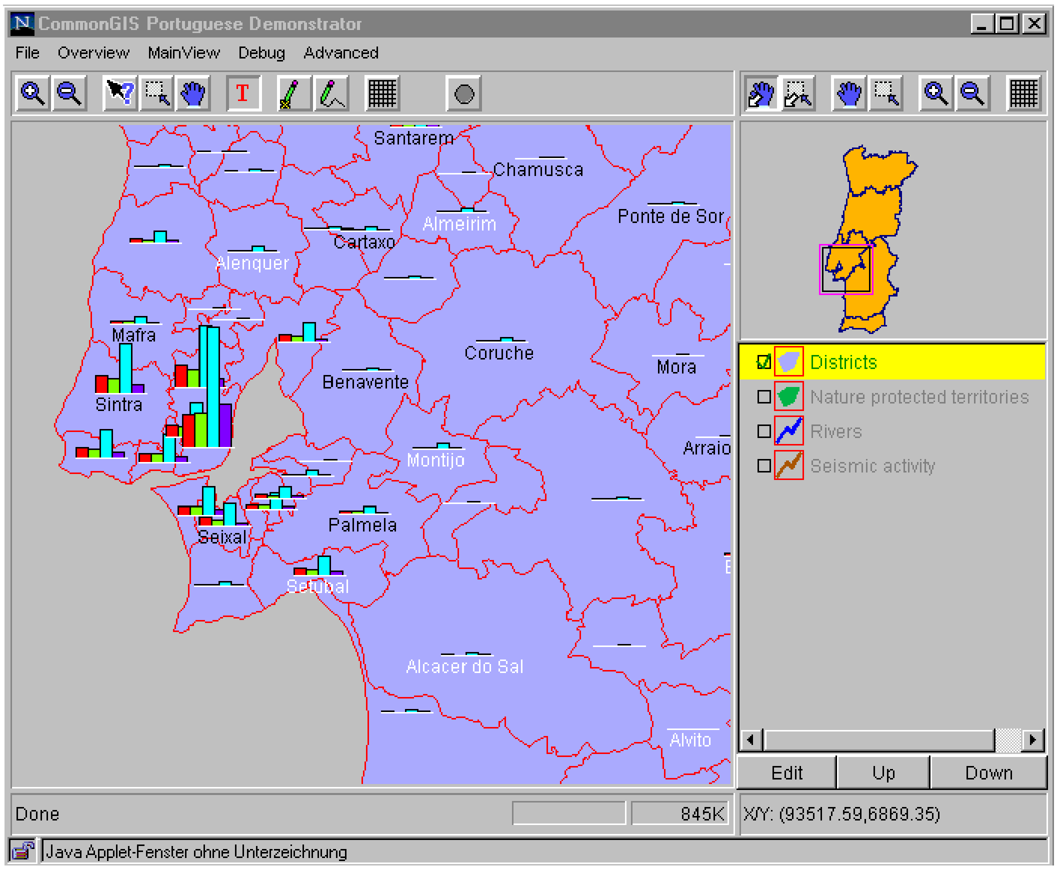


Fig.6.3: Data representation on a map

This map can be used by the student to get knowledge about how to deal with a map, how to zoom in, zoom out, i.e. to change the scale. To analyze the data the presented part of the map should be extended to an area covering Portugal as a whole. Thereafter, one recognizes some areas within Portugal where the population number is higher than in most of the other regions. This is remarkably true for the regions around Lisbon as well as around Porto. In most of the other territories the bars are not visible as they are very small. By zooming into the chosen area they will get observable. The use of the system could serve as an introduction for younger students to occupy with dependencies between population density and employment, with differences between rural and urban areas or with problems of migration within a country as well as between countries.

Interaction techniques

The interface of the system consists of two windows: the map window and the thematic layer control. The latter can be used to select data in form of columns from a table, as well as to select a visualization method. A third function, the exploration tool, is presented in Figure 6.4.

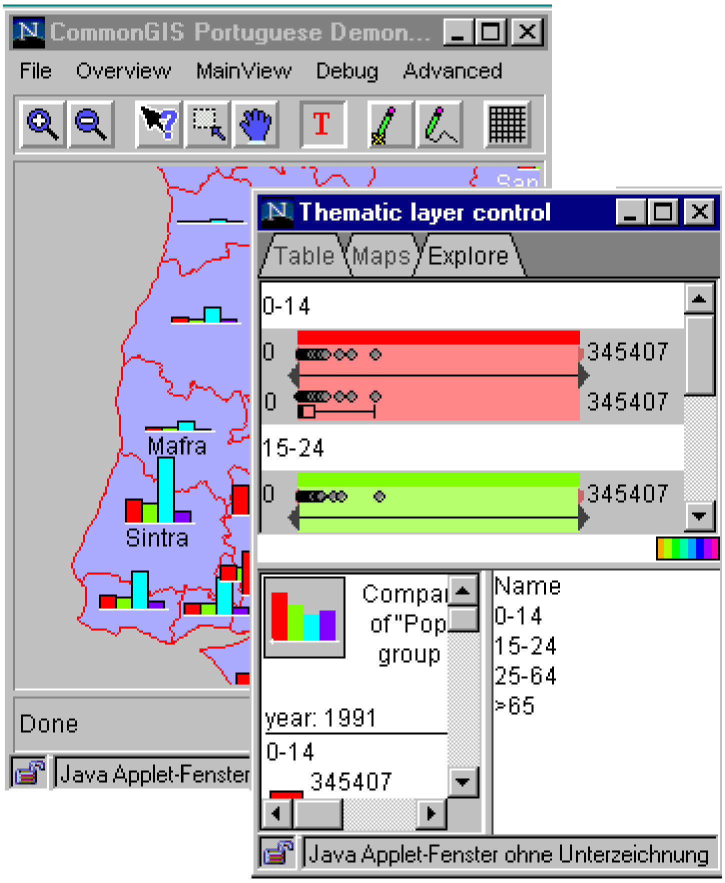


Fig. 6.4: Map window and thematic layer control

Using the exploration tool, the appearance of the map output can be matched to individual preferences. This is one of the user-friendly characteristics of the CommonGIS system. To do so, the exploration window is explicitly shown in Figure 6.5.

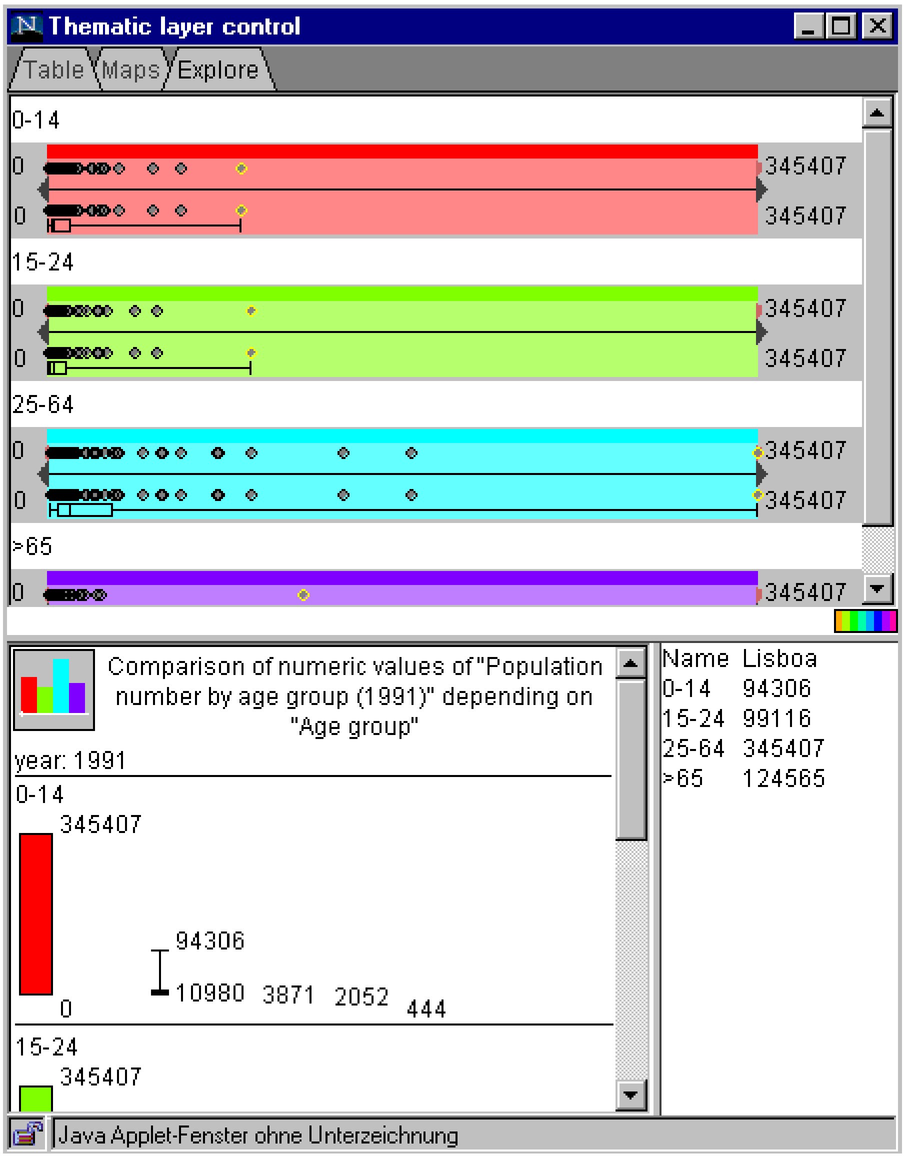
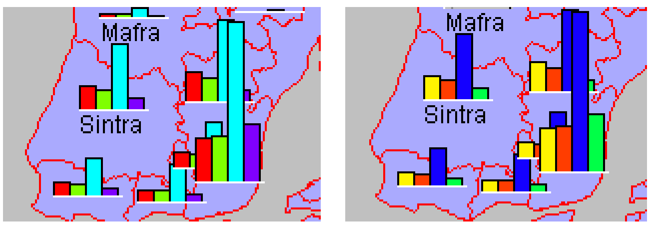


Fig. 6.5: Interaction facilities

To fit the map output to individual preferences concerning colors, they can be changed by clicking on the little palette on the middle of the right side. As a result, a new window opens. Here, for each bar of the map the color can be changed. Figure 6.6 presents this adjustment.

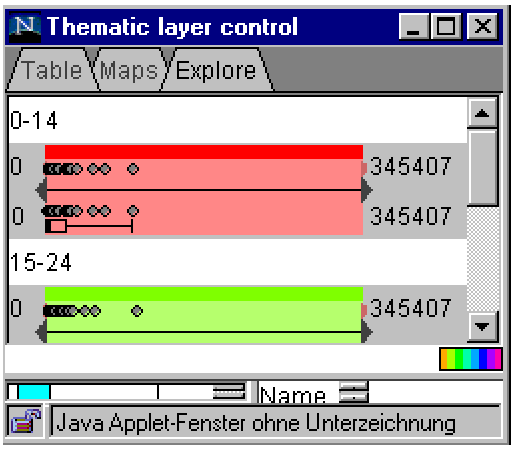


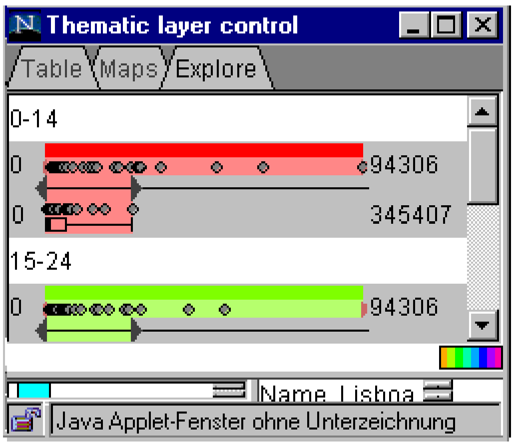
a) Original presentation b) Modified presentation

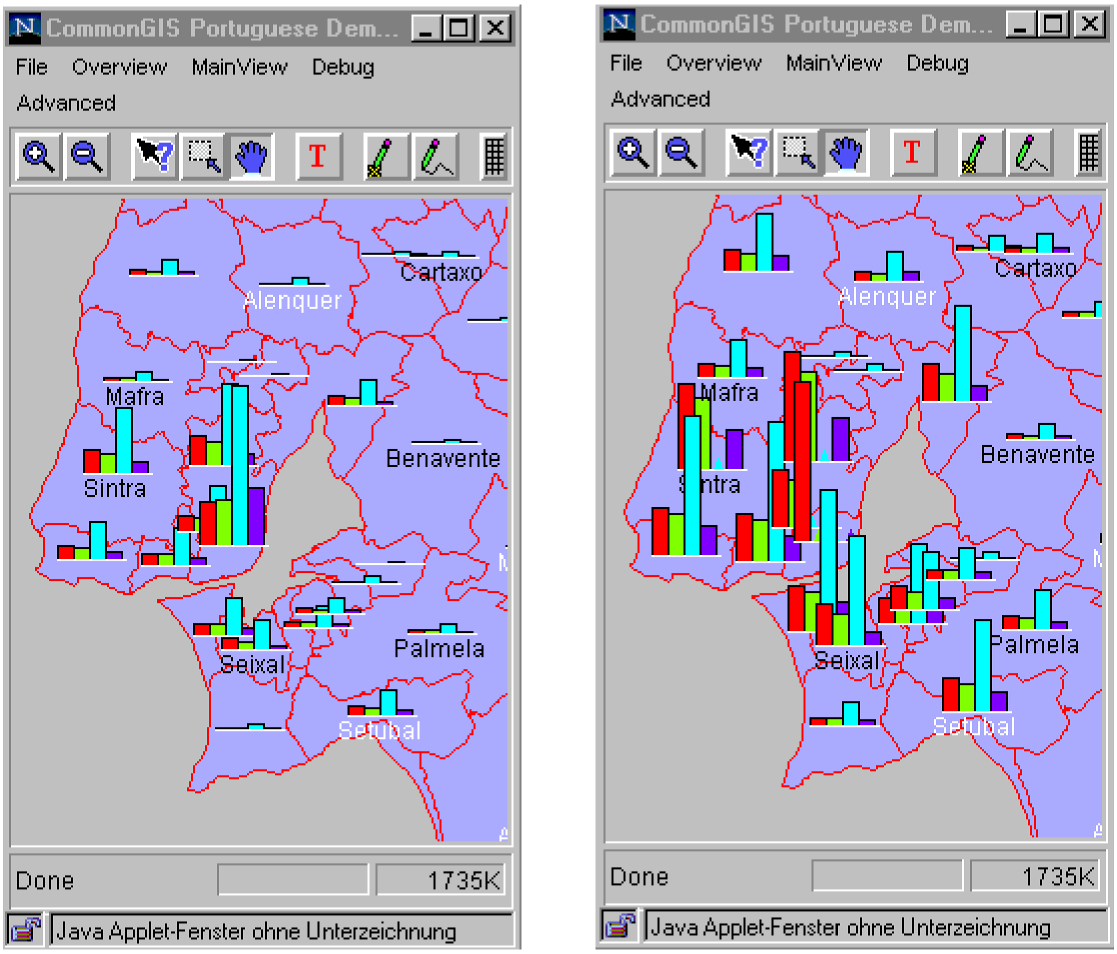
Fig. 6..6: Change of color

The bottom part of the exploration window corresponds to a legend as can be seen in figure 6.5 above. The four age groups of the selected data are represented by four bars in different colors. In the original map the colors picture the following age groups: red for 0-14 years, green for 15-24 years, turquoise for 25-64 years, and violet for older than 65 years.

With the mouse the user can point on a special district within the map. On the right side of the thematic layer control window (figure 6.5) data of this region is specified. While indicating another region on the map, data of the respective district is listed. Data for each age group for all of Portugal is denoted on the top part of the screen. Each dot within the age groups indicates a value stored in the data tables. While choosing a particular district in the map the respective dot is highlighted by a yellow margin. The reference size of the visualization determines which values presented by bar charts are visible and which ones are not, i.e. the change of this reference size affects the height of the bars of all values. This modification can be done by moving the slider indicated by an arrow to the favored rate or by clicking another dot to refer all of the values to this new value. Figure 6.7 shows the difference between two maps with diverse reference sizes.







a) b)

Fig. 6.7: Presentation with different reference sizes

a) Exploration and map window for the reference size 345.407

b) Exploration and map window for the reference size 94.306

By changing the reference size in the chosen example from 345.407 to 94.306 the heights of the bars are raising. This leads to a better overview in areas with small values but also to more confusion in areas with great values. Depending on the task the user would like to fulfill, the presentation can be matched to personal needs very easily.

As the modifications are done immediately, users without background in statistics can observe how visualizations change by manipulating several conditions. The knowledge about the behavior of statistical data is essential for understanding and interpreting statistical visualizations.

Summary

The presented CommonGIS system is an easy-to–handle tool for the visualization and analysis of statistical data. The teacher can use data already supplied by a data provider within the system or input data by using a special tool based on the data characterization scheme developed within the project. Students on the other hand need this data stored in tables to select features they want to visualize. The system then offers appropriate visualization methods to them by considering the data structure and eventually transforming the data. After choosing one of these techniques, the user gets a map enriched by the desired statistical visualization. Manipulation by use of basis GIS functionality like zooming or panning are offered to the user. Sophisticated interaction facilities like the change of reference sizes enable students to evaluate the magnitude of manipulation. This enables the interpretation of visualizations in a special way. This knowledge is not only essential for sciences like biology or chemistry but also for everyday use of media like newspapers or the internet.

Acknowledgements

We hereby would like to thank our partners in the CommonGIS project, especially Natalia and Gennady Andrienko, Hans Voss, Robert J. Peckham, Frank Tuijnman, and Fatima Bernardo for fruitful discussions and their enthusiasm to reach the common goal.

Reference

Andrienko, Gennady and Natalia, 1998: *Interactive Maps for Visual Data Exploration,* ICA Commission on Visualization meeting

Jung, Volker, 1992: *Historie auf dem Bildschirm*, Computer Graphik Topics Vol. 4

Jung, Volker, 1998: *Integrierte Benutzerunterstützung für die Visualisierung in Geo-Informationssystemen,* Dissertation am Fachbereich Informatik der Technischen Universtität Darmstadt

Monmonier, Mark, 1996: *How to Lie with Maps*. 2nd ed. Chicago: The University of Chicago Press

Pfanzagl, Johann, 1983: *Allgemeine Methodenlehre der Statistik I.* 6th ed. Berlin, New York: de Gruyter