

INTERACTIVE LEARNING MODULE ON SPATIAL VISUALISATION OF STATISTICAL DATA

Prof. Dr.-Ing. Ralf Bill and Dipl.-Ing. Marco Lydo Zehner

Rostock University, Institute for Geodesy and Geoinformatics, D 18051 Rostock,
{ralf.bill,marco.zehner}@auf.uni-rostock.de

**PS WG VI/2-4-1 Computer Assisted Teaching; Internet Resources
and Distance Learning**

KEY WORDS: Spatial Information Sciences, Statistics, Education, Mapping, Learning, Multimedia

ABSTRACT:

Education on basic statistics belongs to the core curricula in various disciplines. The following article introduces the research project "Norddeutscher Methodenlehre-Baukasten" dealing with statistical methods in an E-learning environment for students in psychology, sociology, medicine, educational and economic sciences, and others.

The target of this research project is to develop E-Learning modules on basic statistics. Within this broad range a specific module was designed for the spatial visualisation of administrative statistical data. The didactical concept, which will be described in the paper, is based on explorative learning, a method, where the student tries to develop own hypotheses and evaluates them interactively. Starting with a motivating introduction on the research context the student gets the chance to extend his naive concepts and hypotheses within a search room related to the learning target. By interactive exercises he is improving his own scientific concepts step by step.

The intention of the specific learning unit is to teach the student in using spatial thematic visualisation methods to illustrate statistical data. Especially for the exploration of spatial patterns, graphical visualisation techniques have major advantages. The example used is dedicated to administrative statistics on population in Germany. Statistical data on population are related to different administrative borders in a hierarchy from municipality to the state level. Normally a student is going to calculate statistic measures such as extreme values, average, standard deviation etc. to characterize the data sets. He will learn to differentiate between absolute and relative measurements and between time dependent and time independent data. Different cartographic visualisation techniques are offered. By trial-and-error he will recognize which visualisation technique fits to which data sets. The aggregation along the administrative hierarchy leads to new knowledge, the problems with the visualisation at certain levels are introduced.

The paper describes a typical session which lasts for around 1.5 lecture hours. First, but very promising, experiences with the evaluation of this learning unit within the lectures on Geoinformatics will be presented.

The technical solution is realised mainly using open source software such as MySQL database, SVG and Java geotools.

1. INTRODUCTION

1.1 The research project "Norddeutscher Methodenlehre-Baukasten"

The multi-disciplinary project „Norddeutscher Methodenlehre-Baukasten“ (abbreviated MLBK) is a cooperation of some north Germans universities, funded by the Federal Ministry for education and research (BMBF) in Germany, which promoted in 2000 an initiative on new media in education at universities and colleges. Partners in this project are:

- University of Bremen with the disciplines mathematics, computer science, sociology and psychology,
- University of Hamburg with the disciplines education didactics (principal investigator Professor Schulmeister), medicine, psychology and computer science,
- University of Greifswald with the discipline psychology as well as
- the University of Rostock with the disciplines computer science, medicine, economics, social sciences, geodesy, and geoinformatics.

The part of the Institute for Geodesy and Geoinformatics is the development of the learning module for the spatial visualization of statistic data.

The Methodenlehre-Baukasten (toolkit of methodological education) is a modular learning-teaching program for the main topics of methodic teachings and statistics.

From its requirement profile and its examples, exercises and texts the MLBK targets on studying courses in psychology, sociology, medicine, educating science and economic science.

At the same time the MLBK offers teaching support when planning and executing training meetings in different teach-learn-settings, both in the operational readiness level and in the virtual teachings.

The MLBK offers methods and models for learning how to set up descriptions for real world phenomena, how to acquire data and how to evaluate these data sets including statistical methods.

The modular structure makes it possible for teachers and the students to vary content, examples and exercises at any time related to amount and purpose of the session. A continuous application and example reference should motivate the scholars to constitute an own meaning and to link to own background knowledge in the education field statistics.

Being based on the concept of explorative learning (Bruner, 1961, Neber, 1975), the teaching program offers the possibility to the scholars to work with basic research data in the context of specialized entrances to current scientific questions. This should improve their understanding of the statistics and method

teachings on the basis of their own naive concepts, conquering the whole field in small cognitive steps to achieve at the end a more or less complete scientific understanding.

The MLBK is an interactive didactic program for method teachings and statistics, which tries, with the help of didactical interventions, overcoming the phenomena “fear of statistics” particularly under scholars in humanities and social sciences.

In the core of the software system interactive exercises are located for self learning, which are developed purposeful to address and diminish the cognitive and effective problems of the students when learning method teachings and statistics.

Besides the interactive sessions a book as accompanying text as well as a glossary exists for looking up terms and improving the knowledge.

The *MLBK* supports a computer-assisted learning process, where the scholars decide on the speed of the learning progress and the place for learning. Thus, the project MLBK is an important contribution to the reform of study and teachings, because:

- the method teachings including statistics is a training item in many disciplines,
- the modules of the program cover the main topics in various courses of studies,
- it is possible to structure the sessions in a modular way related to the specialization area or the complexity degree,
- it is open for internationalisation and extension, and
- representatives from several subjects and several universities are involved in the project.

The modularity makes a multilingual employment possible in various specialisation areas. At the same time the MLBK may be extended and upgraded for future interests and may track the constantly growing teaching requirements in the different disciplines easily.

1.2 Overview of the contents

The MLBK is constructed by several components, which thematically include the majority of the method teaching context on an undergraduate level in various disciplines. Main components are the modules *from the reality to the data*, *data collection*, *descriptive statistics* and *inference statistics*.

In one special module components for special methods are integrated such as cluster or time-series analyses. The spatial visualisation part described here is also part of this special module.

The *empirical hands-on training* is a part of the learning environment, where the scholars can deepen their knowledge gained so far applying it by planning and executing own investigations (see Figure 1).

Within these modules there are again several layers in the form of texts, media, exercises and examples, which can be variegated within the specific content and conditions of the session.

In the centre of the MLBK interactive exercises for self learning are located based on the concept of explorative learning. The idea of explorative learning results from a cognitive analysis of the learning processes of scholars. This method is suitable to diminish learning difficulties in complex areas.

Exercises are linked with hypertext-textbooks and hypertext-glossaries. A central navigation location permits the entrance to contents from any component of the system. The users are supported by a monitoring, which on demand informs the user on completed course parts or positions the user, when re-entering the teaching program, at the termination point of the last session.

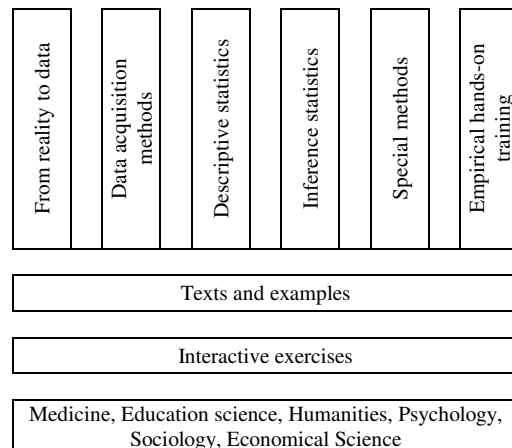


Figure 1: Main structure of the MLBK

2. THE LEARNING MODULE “SPATIAL VISUALISATION OF STATISTIC DATA”

2.1 Technical background

The interactive learning modules for the MLBK are embedded in an on-line Internet environment. The structure is modular and allows to construct own learning paths through the different modules related to the scholars disciplinary background.

Examples and exercises are dedicated to the special disciplines, but may be selected freely. The framework and the data are arranged in such a way that they are exchangeable by the respective lecturer. The high degree of interactivity in this module can not be developed with simple Internet technologies. The problem lies on the one hand in the visualisation of the changeable spatial data with its attributes and on the other hand with the real-time computations of statistic procedures. Both requirements could easily be attained with existing software products, but would cause additional license costs for all education sites using these modules. Thus, one goal of the project was to use only cost-free products or self-developments. This guarantees that the total package can freely be distributed without auxiliary costs.

Learning contents are stored with the help of an author system into a MySQL database, which likewise supports the modular structure. The complex structures are represented on a Website with the server side script language PHP. Most of the interactive exercises had to be programmed by our own.

The available subproject for spatial visualisation is developed based on a specialised example of population statistics. In the first phase of the project different new technologies were tested to realise the interactive exercises. We tested clients Java applications, SVG and Flash elements. Java applications can be developed with appropriate programming and offer high-grade interactivity on the client side. For the representation of spatial data in the web the JAVA-library Geotools is very prominent. Geotools is a project of the University of Leeds, started in 1996, and contains an extensive class library for the selection and representation of spatial and other data. The functionalities of Geotools could easily be used for the interactive environment. In the module a set of learning units were implemented with Geotools. The conversion was quite simple, because only the different components in a Java applet must be rearranged. Additional interactions were developed with the Java library Swing. All applets are parameterised and are called dynamically, so that reusability is given. A major performance problem comes with larger application data sets resulting in

load time and the increasing requirements to the client computer. The representation of more than 1.000 municipalities in Mecklenburg-Western Pomerania could not be achieved within a reasonable time. These performance issues limited the representation on the borough level in the administrative hierarchy in Germany. On the client side a version of the Java Runtime environment is needed.

A further open environment for the interactive visualization of spatial data is SVG. SVG (Scalable Vector Graphics) is a language to describe two-dimensional graphics by XML. Already different standardisations are given by the W3C and it can be assumed that SVG intersperses as a standard in all future WWW-viewers. Highly developed interactive SVG applications are converted with the help of a supplementing script language. The full access to the SVG Document Object Model (DOM) is possible e.g. with JavaScript. Thus at the same time access to all further XHTML and SVG elements in the same web page is given. SVG applications are quality-free scalable. The file size is small using text compression and the loading time is short. SVG is a kind of XML and is based on pure ASCII files. These files can be built dynamically with PHP and be embedded in the Web application. Hereby the spatial data are stored in the open standard GML (Geographic Markup Language) in the MySQL database and visualised depending upon interaction. The principle of the open SVG map server according to www.carto.net is used.

For the computation of statistical measures the language R is used. R is a development of the University of Auckland and a computer language for statistic analyses. R can be merged into different application packages. Via the CGI interface of the Web server the functions can be called dynamically, R can access the data in the database directly. For our local applications the combination of MySQL/PHP and SVG was suitable and the performance was satisfying.

In the future employment the developed software components are merged with the respective teach and learning platforms of the university e.g. WebCT, Ilias, Blackboard, or stud.IP. The learning-platform offers the basic functionalities such as user administration and communication tools.

2.2 Target, content and data set in the learning module

The target of the learning unit is to learn about spatial visualisation and spatial analysis with respect to population statistics. Subject of population statistics is to apply statistical methods and procedures for the numerical collection, representation, analysis and interpretation of the development on population in a special region. Statisticians know that diagrams of statistical or calculated data are easier to interpret. With increasing number of data sets graphical visualisations are also easier to be interpreted than tables. Particularly for the representation of spatial distributions of the population map techniques are very helpful, because spatial conditions and relations can be better recognized.

National and federal offices for statistic data supply official population data on different aggregation stages. The smallest administrative unit, on which the information is published, is the municipality. A lot of characteristic data are raised annually on the municipality level at statistical offices. In this teaching material comparatively the population existence Mecklenburg-Western Pomerania are graphically represented for the time stamp 31.12.1990 and 31.12.2000. These data were published by the statistic national office in the year 2001. In Mecklenburg-Western Pomerania there were altogether 1203 administrative districts (municipalities, districts, urban areas and country) in the year 2000. Thus the table generated from

that data set with the appropriate population information has over 1200 lines and according to the acquired population characteristics many columns. It is very hardly readable. The following table shows a cut-out of these data from Mecklenburg-Western Pomerania in the reference to the population numbers. Each municipality represents a regional date with its attributes. Each municipality has its specific size, to which the values refer.

Gemeinde-schlüssel	Gemeinde	Kreis	Amt	Bevstd. 1971	Bevstd. 1981	Bevstd. 1990	Bevstd. 1998	Bevstd. 1998m	Dichte 1998
13051015	Broderstorf	Bad Doberan	Carbäk	1 359	1 242	1 304	2 576	1 324	125
13051037	Klein Kassewitz	Bad Doberan	Carbäk	639	512	508	514	259	36
13051045	Mandelshagen	Bad Doberan	Carbäk	342	223	195	257	140	16
13051054	Poppendorf	Bad Doberan	Carbäk	542	305	268	723	365	52
13051064	Roggentin	Bad Doberan	Carbäk	648	514	502	1 684	875	176
13051075	Steinfeld	Bad Doberan	Carbäk	379	300	292	410	201	30
13051079	Thulendorf	Bad Doberan	Carbäk	593	428	410	448	225	42
13051003	Altenhagen	Bad Doberan	Kröpelin	420	279	300	291	152	26
13051029	Jennewitz	Bad Doberan	Kröpelin	708	529	492	565	300	29
13051032	Karin	Bad Doberan	Kröpelin	419	324	292	443	267	30
13051041	Kröpelin, Stadt	Bad Doberan	Kröpelin	4 485	4 823	4 441	4 285	2 096	160
...

Table 1: Table excerpt - official statistic data of Mecklenburg-Western Pomerania

2.3 A session with the learning module

At the beginning of the learning unit the student tries to conquer the data set with the acquired statistic knowledge from preceding chapters. For instance questions such as “find the largest and smallest municipality, municipality with the highest, middle and lowest total population” can easily be answered by using statistical measures such as span width, average value etc. The student can compute also new data e.g. the population density as a quotient from population conditions to the surface size of the administrative unit. Thereby the scholar learns to differentiate between absolute (e.g. population conditions) and relative values (e.g. population density). He may compare the results of different municipalities. As long as these are identified with their names, the student has no problems, particularly if he has a certain local knowledge in that region. But the scholar will recognize that it is not easy to find patterns on spatial distributions in these data.

For each municipality and/or each district in the table the geometric borders are stored. This is now linked to the attribute data of the municipality and/or the district making use of the primary key, the name of the municipality. The polygons are individual geometries associated to the appropriate data record. In the following lesson the student sees the spatial allocation of the municipalities in the country Mecklenburg-Western Pomerania. For the first time the scholar makes himself familiar with different types of representation. Absolute values are applied for entities like the population existence to the surfaces to a map. Relative values are used, if statistic proportionality factors have to be shown such as per cent or values relative to an uniform base factor (e.g. habitant/km²). For absolute value representations signatures for point or area elements may be used, which possess qualitative and quantitative attributes. The signatures are represented at standard positions related to the unit of the area. In order to represent several values in the comparison, one uses diagrams. The values are represented in addition in different scale, so that the values are recognizable despite different dimensions (e.g. one point per 1000 person).

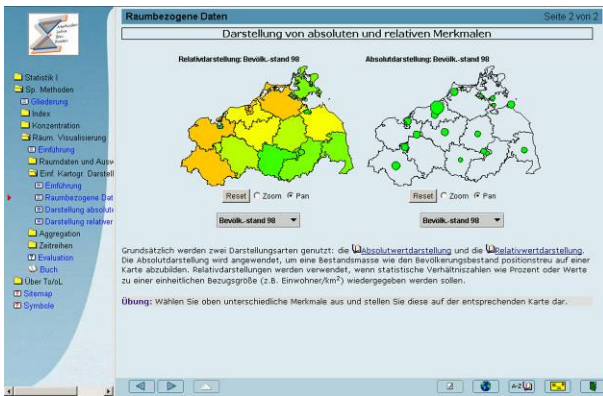


Figure 2: Visualisation of absolute and relative values

A meaningful choice of the colours is particularly important with thematic evaluations. Colour-optically and psychological-correctly selected colours facilitate the legibility and interpretation ability of the map. Depending upon application sliding colour scales or opposite colours should be used. Sliding colours are suitable for values of equal sign. Opposite colours express positive and negative effects better. Alternatively they can be worked out with hatching. Opposite values can be represented also very well with traffic light colours. Apart from this well-known red-yellow-green colour combination one uses frequently blue-white-red, because this corresponds to the human sense for cold and warm. Similar to the first lessons the scholars work on the same question again, but now they are applying spatial visualisation techniques. By this, the student learns, which representation form is suitable and that cartographic mapping makes phenomena and relations such as population movements (like rural exodus or suburbanization) visible.

The formation of classes becomes necessary if the entities, being either metrically scaled or with constant or discrete characteristics, show a wide variety of possible values. The class formation allows to recognize the characteristics of the distribution and to determine regularities in the data sets. In such a lesson the student learns, how many classes of meaningful values may be constructed and how many classes humans are able to separate visually. The number of classes and the class borders affect the derived measures as well as the visualisation.

Municipalities are the smallest units in the official statistics. They can be hierarchically combined into larger administrative units by means of aggregation and later be also disaggregated from larger units to smaller units. In Germany there are administrative subdivisions, which follow the international framework of statistics. Each municipality is assigned to an districts and a country. The key of the municipality is used to distinguish between the municipalities. Additionally there is an allocation in regions, which may be governmental districts or planning regions. Several districts may belong to a region, again a hierarchical relation. In this lesson the scholars explore the fact that certain absolute values may be aggregated from one administrative level to the next level. On the other hand relative values such as population by area can not be aggregated directly. Therefore one has to go back to absolute values for aggregation and calculates the relative values afterwards. The students also recognize that aggregation and classification affects the measures as well as the graphical visualisation.

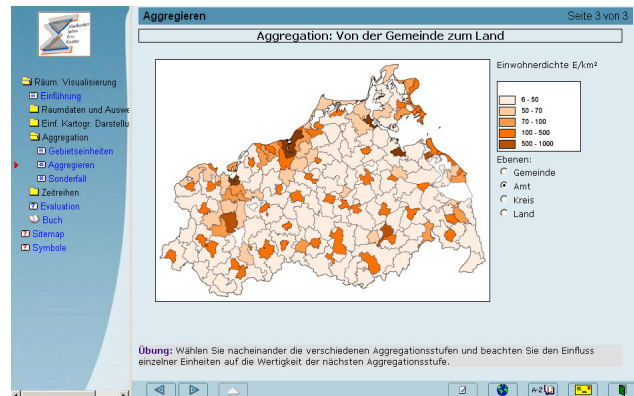


Figure 3: Evaluation representation aggregation

Population statistics is collected for a long period of time which allows answering questions on population changes over time. In the last lesson the scholar is working on time-series. These are usually provided for one period from one year and made available by the official statistics. To check longer periods in time new sets of values such as balance rates are computed. For the representation of the time series existing procedures known from the visualisation of absolute and relative measures can be used. For absolute values these may be diagrams or cartograms on a map as well as signatures. Relative values may be indicated in sequences post or next to each other. Alternatively a relative characteristic may be determined for the respective period of time. Nearly all presentation methods can be combined with each other.

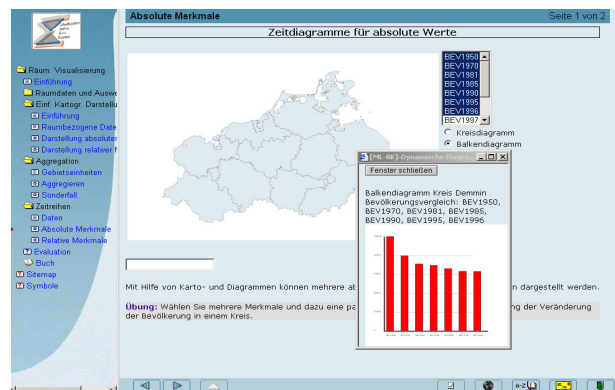


Figure 4: Time-series analyses

By means of diagrams or cartograms several absolute values and temporal successions can be represented. For the direct comparison of the regional areas over time the values for arbitrary periods (e.g. balance rates from 1995 to 2000) can be computed. Thereby the relative values always refer to the last conditions of the respective statistical collection. With such derived relative measures again meaningful visualisations can be created.

3. FIRST EXPERIENCES

In the context of the standard lecture on geoinformatics in a course of studies on land management and environmental protection (32 students in the sixth semester) we evaluated the learning module "Spatial Visualization of Statistic Data" in summer 2003. 22 students participated in the evaluation. Statistics is not a central component of that lecture and the statistical knowledge of the undergraduates in this engineering course are not comparable to those of the targeted courses of

studies (e.g. a fear of statistics could not be determined). Nevertheless this first evaluation should examine the general suitability of the developed learning module. A complete evaluation takes place in the next time in several courses of studies by the evaluation group involved in the project.

In our evaluation the average participant needed 1:30 h for the entire revised lesson. The time includes the answering of a large evaluation form. Some of the results should be explained here. More than 50% of the students found the training aims clearly described which can be achieved with the help of this teaching program. The program allows a critical investigation of the topic. About two thirds of the scholars were satisfied with the human interface (e.g. simple and understandable operation 65%, self-describing 72%, user guidance clear and understandable 68%, good information arrangement 73%, useful guidance and assistance 65%, simple orientation 65%, clear arrangement 60%, understandable feedback 75%, clear structure of the learning module 75%). The reactions on the assumed pre-conditions gave also a high agreement (tying to well-known background 51%, sufficient computer knowledge 91%, necessary pre-conditions given 81%). The auxiliary information offered such as the glossary and the text-book found large acceptance (technical terms explained sufficient 83%, glossary great help 70%). Regarding the knowledge acquisition by the selected method of explorative learning there was a clear deadlock (understanding promoted 56%, remembering supported 45%, active tackling with the topic given 45%, self learning and exploring supported 44%, too little range for own ideas and solution attempts 55%, ability to define learn speed 72%, work with genuine data promotes understanding 65%). In summary we state that the given learning module founds quite good agreement from the scholars. Many of them however have problems with the method of explorative and even steered learning. Additional discussions with the students showed that a large portion of them have to become more familiar with electronic learning in such a form given here. The computer as spare/substitution of the lecturer is not wanted by the students.

4. LITERATURE AND LINKS

Bruner, J.S. (1961). The Act of Discovery. In: Harvard Educational Review 31, 2132

Neber, H. (Hg.) (1975). Entdeckendes Lernen. (2nd edition). Weinheim.

<http://www.izhd.uni-hamburg.de/baukasten.html> - project side of the MLBK

<http://methoden.informatik.uni-rostock.de> - subproject Rostock

<http://www.medienbildung.net/> - loan program BMBF

<http://www.bmbf.de/> - Federal Ministry for education and research

<http://www.geotools.org/Geotools> - the Geotools library

<http://www.carto.net/> - Cartography with SVG

<http://www.r-project.org/> - The R Project for Statistical Computing

ACKNOWLEDGEMENTS

The authors thank the BMBF for the promotion of the work on the project under the loan program no 08 Nm 10.