Evaluating the usability of visualization methods in an exploratory geovisualization environment

[1]Etien L. Koua, [2]Alan MacEachren, and [1]Menno-Jan Kraak

[1]International Institute for Geo-Information Science and Earth Observation (ITC)
P.O. Box 6, 7500 AA Enschede, The Netherlands
koua@itc.nl; kraak@itc.nl; maceachren@penstate.edu

[2]GeoVISTA Center
Department of Geography
Penn State University
302 Walker, University Park, PA 16802, USA
maceachren@psu.edu

Abstract

The use of new representation forms and interactive means to visualize geospatial data requires an understanding of the impact of the visual tools used for data exploration and knowledge construction. Use and usability assessment of implemented methods and tools is an important part of our efforts to build this understanding. Based on an approach to combine visual and computational methods for knowledge discovery in large geospatial data, an integrated visualization-geocomputation environment has been developed based on the Self-Organizing Map (SOM), the map and the parallel coordinate plot. This environment allows patterns and attribute relationships to be explored. A use and usability assessment is conducted to evaluate the ability of each of these visual representations to meet user performance and satisfaction goals. In the test, different representations are compared while exploring a socio-demographic dataset.

Keywords: Usability, geovisualization, self-organizing map, visual exploration.

1. Introduction

The need to assess the usefulness and usability of geovisualization tools is increasing as new types of interactions emerge (Muntz, Barclay et al. 2003). The use of new representation forms and interactive means to visualize geospatial data requires an understanding of the impact of the visual tools used for data exploration and knowledge construction. Use and usability assessment of implemented methods and tools are an important part of our efforts to build this understanding. Such assessments focus on the effectiveness, usefulness and performance of a tool. In geovisualization this is needed because use and usability testing can provide insight into how a visual interface can support data exploration tasks.

Increasing research interest in the usability of geoinformation systems has recently linked the Humand Computer Intercation (HCI) field, cognitive science and information science in a few applications of approaches that integrate across these fields (MacEachren and Kraak 2001; Haklay and Tobon 2003; Koua and Kraak 2004b; Fuhrmann, Ahonen-Rainio et al. 2005). The traditional map use studies (MacEachren 1995) conducted in the field of cartography are not necessarily fully applicable in new interactive visualizations that involve new representational spaces and advanced user interfaces. The lack of appropriate evaluation methodology in the geovisualization domain and particularly task specifications for user-based testing in exploratory geovisualization tools (Slocum, Blok et al. 2001) has limited the number of user studies directed at formally assessing geovisualization tools. Since the design of effective visualization tools will depend upon understanding the way users interact with and make interpretations of the information spaces used to represent patterns and relationships in data, the choice of a representation and interaction method is crucial to the success of a visualization environment. Empirical testing of the visualization tools can provide insights into the potential of particular visual displays and interaction paradigms (Fuhrmann, Ahonen-Rainio et al. 2005).

One of the dominant approaches in geovisualization is the integration of several representation methods that provide different perspectives of the data in multiple linked views. Such an integration of views can be more effective if focused on the potential of the individual representations for specific conceptual visualization goals that can better support the exploration, evaluation and interpretation of patterns and ultimately support knowledge construction (Roberts 2005). Based on an approach to combine visual and computational methods for knowledge discovery in large geospatial data, an integrated visualization-geocomputation environment has been developed. It incorporates a self-organizing map (SOM) neural network algorithm for the extraction of patterns and integrates this computational method with graphical representations used to portray extracted patterns to support the understanding of the structures and the geographic processes. This integration of visual representations of the SOM (e.g. views on non-geographic information spaces or attribute space - (Koua and Kraak 2004a)) with maps and the parallel coordinate plot allow (geographic) patterns and attribute relationships to be explored. The tool is designed to facilitate knowledge construction, using a number of steps provided in a data mining and knowledge discovery methodology.

In order to investigate the effectiveness of the design concept, a use and usability assessment is conducted to evaluate the tool's ability to meet user performance and satisfaction goals (Fuhrmann 2005; Tobon 2005). The methodology of the test is based on an understanding of several knowledge discovery activities, visualization operations, and a number of steps in computational analysis used to visualize patterns in the data. In the test, different representation methods are used to explore a socio-demographic dataset; these include maps, a parallel coordinate plot, and interactive visualizations of the SOM output. The study emphasizes the knowledge discovery process based on exploratory tasks and a taxonomy of visualization operations.

The results are organized according to the visual tasks derived from the taxonomy of conceptual visualization goals and operations, and are compared for the different visual representations (maps, parallel coordinate plots, and the SOM-based representations). The taxonomy was used to structure the study. This paper concentrates on the usability evaluation methodology, the test procedures and the results.

2. Exploration and knowledge discovery tasks in the visualization environment

Judging whether a geovisualization (or other) exploratory environment is effective requires answering the question: effective for what? We begin, therefore, with a discussion of exploration and knowledge discovery tasks to which a geovisualization environment can be applied. The model presented in figure 1 emphasizes the exploratory nature of a visualization environment designed to provide support for knowledge construction, from hypothesis formulation to the interpretation of results. This figure focuses on the exploration steps undertaken by users. Some of these steps may be repeated.

2.1. Defining user tasks for usability evaluation

The main goal of geospatial data analysis is to find patterns and relationships in the data that can help answer questions about a geographic phenomenon or process. The geographic analysis process can be viewed as a set of tasks and operations needed to meet the goals of the data exploration (Fotheringham, Brunsdon et al. 2000; Andrienko and Andrienko 2005). The primary tasks in this process include: checking the spatial positioning of elements of interest in order to verify spatial proximity among different elements; verifying their spatial density; and obtaining an overview of how a target value measured at one particular spatial location, or at various neighbouring locations, varies for different attributes. These tasks involve a number of more specific activities and operations that users will perform (Weldon 1996):

- identification of the different clusters in the data, and relationships between elements (within clusters and between different clusters)

- comparison of values at different spatial locations, distinguishing the range of value
- relation of value, position and shape of object identified
- analyze relevance of information extracted.

The above activities are often facilitated by functions that allow selection, scaling, rotation panning, brushing, browsing, filtering, and querying the database.

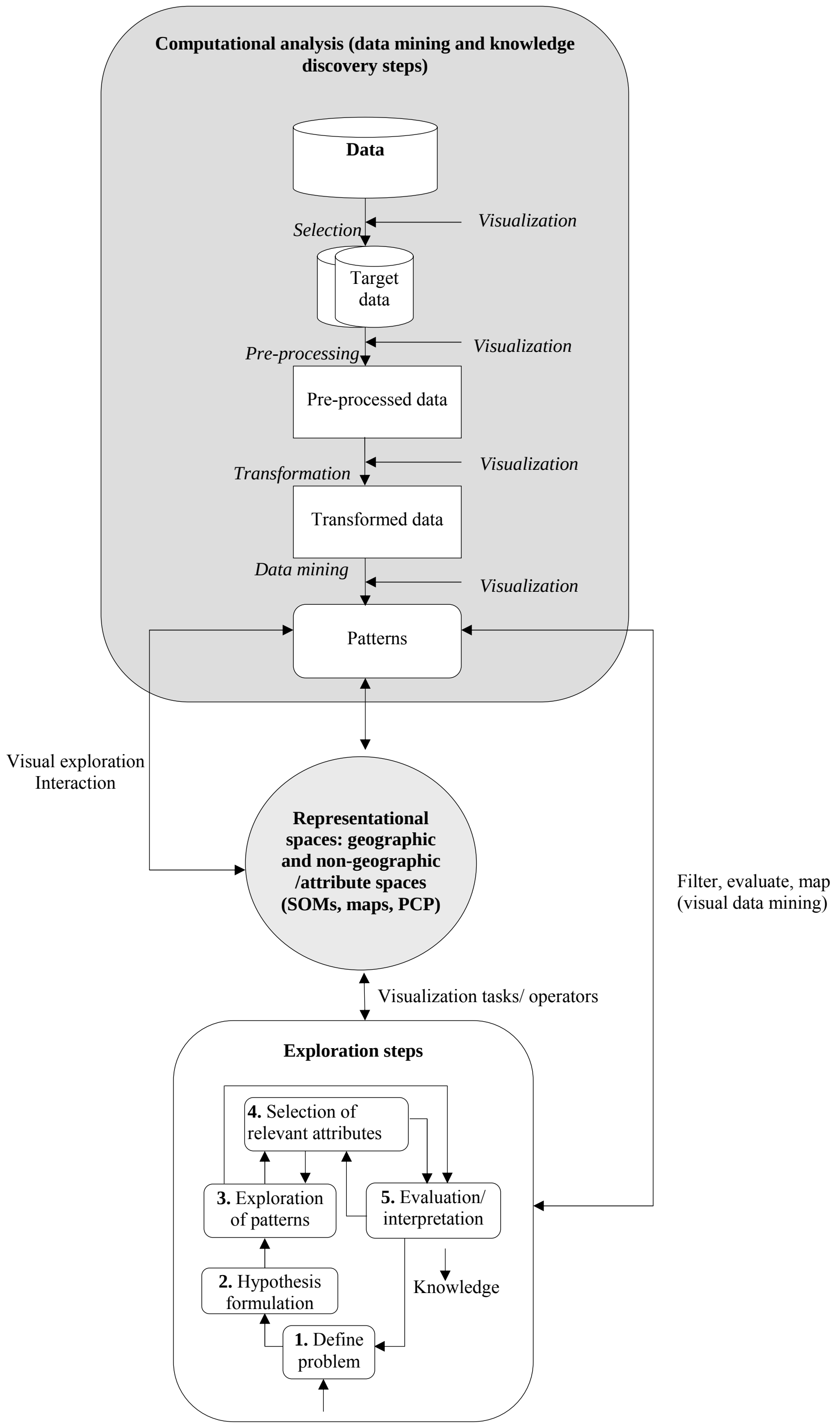


Figure 1. Data mining, exploratory visualization and knowledge discovery processes. The first part of this process consists of the general data mining and knowledge discovery steps (computational analysis). Each of the steps of the computational analysis can allow visualization. Patterns extracted as a result of the computational process can be explored using graphical representations (geographic and non-geographic information spaces). This exploration is guided by a number of steps to support knowledge construction. The steps presented in this figure correspond to the classification of Gvis and KDD operations presented by MacEachren (1999).

The exploration steps described in figure 1 are supported by basic visualization tasks and operators, as users manipulate the graphical representations and initiate actions during the different steps. These visualization operations are the basis for the success of the exploration process.

2.2. Exploration tasks and visualization operators

To complete the tasks described above, the user will have to execute a number of visualization operations during the exploration process described in figure 1. Several authors have suggested taxonomies for visualization operations (Keller and Keller 1992; Qian, Wachowicz et al. 1997; Zhou and Feiner 1998; Ogao and Kraak 2002). The most comprehensive list (Keller and Keller 1992; Wehrend and Lewis 2000) includes: identify, locate, distinguish, categorize, cluster, distribution, rank, compare, associate, and correlate:

- Identify: to establish the collective characteristics by which an object is distinctly recognizable
- Locate: to determine the absolute or relative position
- Distinguish: to recognize as different or distinct
- Categorize: to place in specifically defined divisions in a classification; this can be done by colour, position, type of object (shape)
- Cluster: to join into groups of the same, similar or related type
- Distribution: to describe the overall pattern. This is closely related to cluster in the same way that locate and identify are related. The cluster operation asks that the groups be detected, whereas the distribution operation requires a description of the overall clustering.

- Rank: to give an order or position with respect to other objects of like type
- Compare: to examine so as to notice similarities, differences, order
- Associate: to link or join in a relationship
- Correlate: to establish a direct connection (correlation).

A set of representative tasks derived from the steps described in figure 1 and key visualization operations described above are identified in visualization task scenarios for the evaluation study. This results from a decomposition of the basic visualization tasks and is presented in the next section. The rationale behind the use of scenarios is that they can represent how the system is intended to be used by end users. Task scenarios provide a task-oriented perspective on the interface and represent a structure and flow of goals and actions that participants are supposed to evaluate. Such scenarios ensure that certain interface features are evaluated (Caroll and Rosson 1992; Caroll, Rosson et al. 1998; Caroll and Rosson 2003).

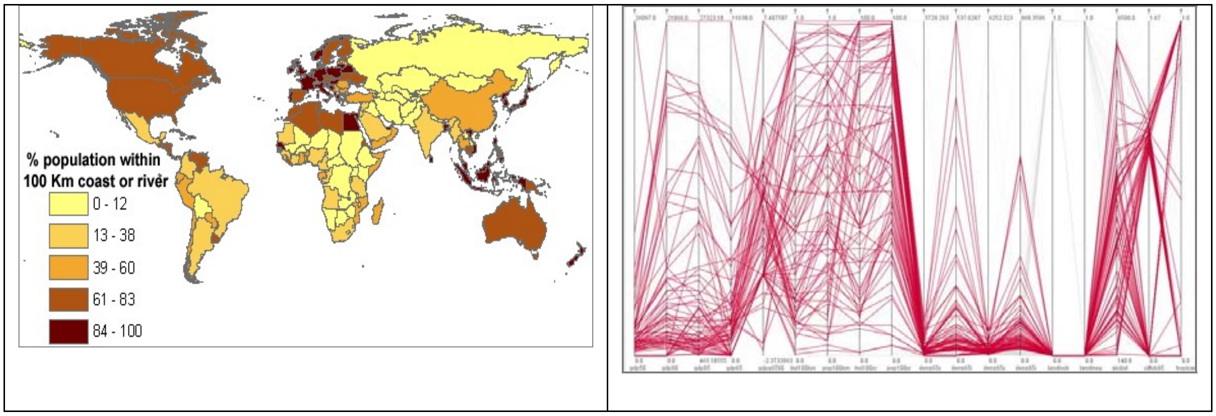
2.3. Evaluation tasks model

The conceptual goals and the different steps of the exploration and knowledge discovery process described earlier are used as the basis for defining low-level (operational) tasks that users need to perform to meet the conceptual goals.

Examples of the visual representations used are shown in Figure 2. Next to the map (Figure 2a) and the Parallel Coordinate Plot (PCP) (Figure 2b) several SOM based visualization have been used. These include unified distance matrix representation (Figure 2c), 2D/3D surface (Figure 2d), component plane displays (Figure 2e) as well as 2D/3D projection

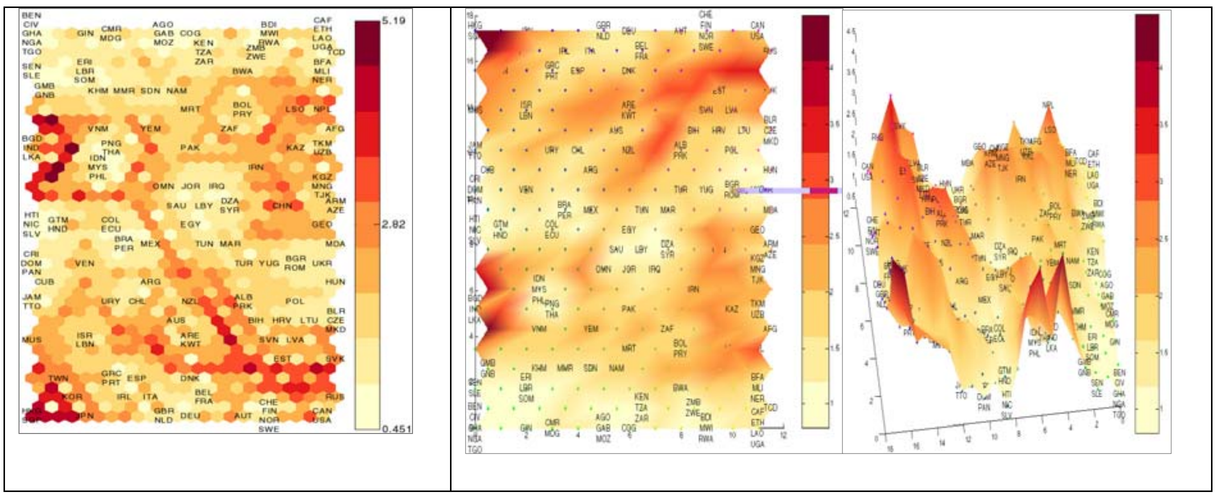
(Figure 2f). The map was selected because it provides a visual representation of the real world that participants are used to. The PCP was selected because it is becoming a prominent tool used in geovisualization. The background of each of the SOM visualization has been described in (Koua and Kraak 2004a).

The visualizations are based on a dataset that represents the relationship between geography and macroeconomic growth (Gallup, Sachs et al. 1999). The dataset contains 48 variables on economy, physical geography, population and health for 150 countries. This dataset was separately explored by the test designer as an experiment and the conclusions of the exploration were used to validate the test participant's results.



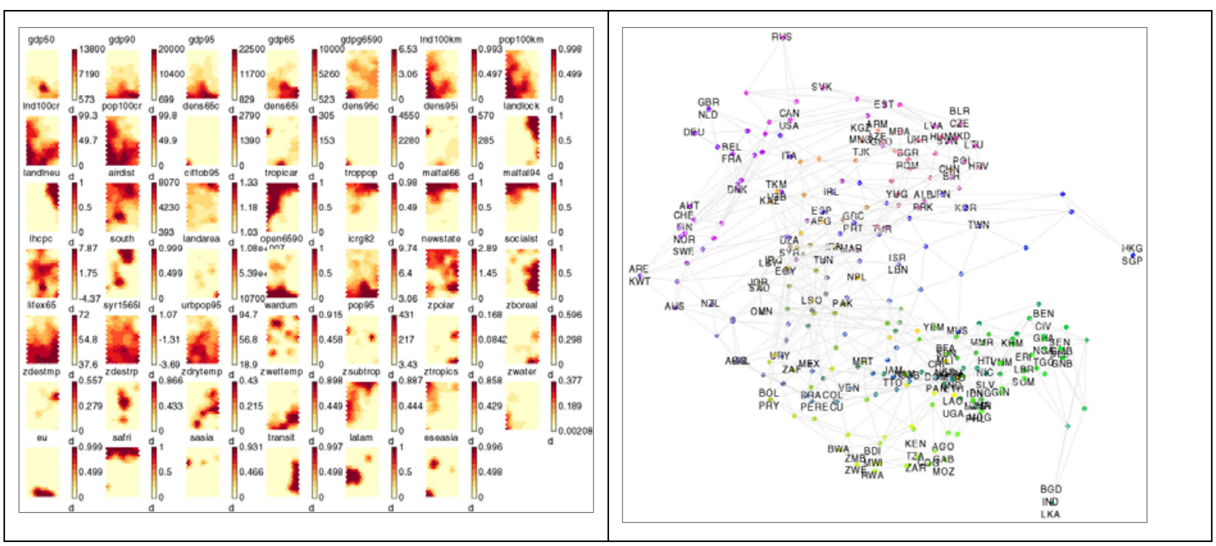
a) map

b) parallel coordinate plot



c) SOM distance matrix representation

d) SOM 2d/3D surface



e) SOM component plans

f) SOM projection

Figure 2. Visual representation used in the test

The test is based on a low-level taxonomy of tasks derived by decomposition of basic visualization operators that users might be applied a visual environment (see table 1). This decomposition of the basic visualization operators was obtained by analyzing task structures of real world visualization problems, representing the collection of subtasks, developing related taxonomy or classification as well as a set of semantic relationships among the concepts, and other entities necessary to perform the task.

The defined taxonomy mapped on the different representation methods used to represent each task contains too many tasks. Since each task is executed with 3, 4, 5 or 6 different representations, much time is needed to complete the test. In order to create a test that could be handled both by the authors as well as the test person (a maximum duration of 1 hour and half for each test person), it was necessary to review the task structure. This was realized based on a visual tasks taxonomy by Zhou and Feiner (Zhou and Feiner 1998) that includes a set of dimensions by which the tasks can be grouped. The major dimensions of this taxonomy include visual accomplishments and visual implications. Visual accomplishments refers to the type of presentation intents that a visual representation might help to achieve while visual implications specify a particular type of visual action that a visual task may carry out. The following experimental tasks are derived for the test (table 1 and 2).

Table 1. List of operational tasks derived from the taxonomy, and specific example tasks for the evaluation.

Conceptual goals / visualization operators	**Operational visualization task**	**Specific task explored in the study**	**Task number**
Locate	Indicate data items of a certain range of value	Indicate the poorest countries (reference to the 1995 GDP lower than 750)	1
Identify	Identify relationships between attributes	Identify possible relationships between the following attributes: population density in the coastal region and in the interior, and GDP per capita 95	2
Distinguish	Distinguish how a target value measured at one particular spatial location, or at various neighboring locations, varies for different attributes (e.g. different values of the same attribute at different spatial locations, and the value of different attributes at a specific spatial location)	How does income (GDP 1995) of the countries vary across space? Define differences and similarities between the countries	3
Categorize	Define all the regions on the display, and draw boundaries. Indicate spatial positioning of elements of interest and spatial proximity among the different elements	Define all the regions on the display, and draw boundaries. Define categories of countries such as rich, and poor countries on the display, and indicate to which category South Africa belongs. Are there African countries in this category? List the countries	4
Cluster	Find gaps in the data on the display	Find gaps in the data and indicate the different clusters	5
Distribution	Describe the overall pattern (overview)	What are the common characteristics of low-income countries (GDP lower than 750)?	6
Rank	Indicate the best and worst cases in the display for an attribute	Indicate the 5 lowest GDP countries and the 5 highest GDP	7
Compare	Compare values at different spatial locations, and the order of importance of objects (data items) accordingly	Compare population density on coastal regions (within 100 km of the coastline) and inland regions (beyond 100 km from the coastline)	8
Associate	Form relationships between data items in the display; Identify relationships between data items (within clusters and between different clusters)	Form relationships between economic development (GDP 1995) of countries in the geographic tropics as compared with other countries.	9
Correlate	Discern which data items share similar attributes	Examine economic development (GDP 95) across the countries: landlocked countries and countries that have access to the sea	10

The operational tasks described in table 1 are tested against all three usability indicators and corresponding measures discussed in the next section. Specific domain exploration tasks related to the dataset explored are used to illustrate each operational task as defined in table 2).

Table 2. Specification of user tasks and visual representation method used to represent task.

Conceptual goals /visualization operators	Operational visualization task	Task n°	Method used in the prototype to represent task	Representation number
Locate	Indicate data items of a certain range of value	1	Maps	1
			Parallel coordinate plot	2
			Component planes	3
Identify	Identify relationships between attributes	2	Maps	1
			Parallel coordinate plot	2
			Component planes	3
Distinguish	Distinguish how a target value measured at one particular spatial location, or at various neighboring locations, varies for different attributes (e.g. different values of the same attribute at different spatial locations, and the value of different attributes at a specific spatial location)	3	Maps	1
			Parallel coordinate plot	2
			Component planes	3
Categorize	Define all the regions on the display, and draw boundaries. Indicate spatial positioning of elements of interest and spatial proximity among the different elements	4	Unified distance matrix	4
			2D/3D projection	5
			2D/3D surface	6
Cluster	Find gaps in the data on the display	5	Unified distance matrix	4
			2D/3D projection	5
			2D/3D surface	6
			Parallel coordinate plot	2
Distribution	Describe the overall pattern (overview)	6	Map	1
			Parallel coordinate plot	2
			Component planes	3
			Unified distance matrix	4
			2D/3D projection	5
			2D/3D surface	6
Rank	Indicate the best and worst cases in the display for an attribute	7	Map	1
			Parallel coordinate plot	2
			Component planes	3
Compare	Compare values at different spatial locations, and the order of importance of objects (data items) accordingly	8	Maps	1
			Parallel coordinate plot	2
			Component planes	3
Associate	Form relationships between data items in the display; Identify relationships between data items (within clusters and between different clusters)	9	Maps	1
			Parallel coordinate plot	2
			Component planes	3
			Unified distance matrix	4
			2D/3D projection	5
			2D/3D surface	6
Correlate	Discern which data items share similar attributes	10	Maps	1
			Parallel coordinate plot	2
			Component planes	3

3. A user-based and task-based usability evaluation of exploratory geovisualization

There are several objectives for the proposed usability evaluation. The evaluation intends to assess the visualization tool's ability to meet goals for user performance and satisfaction with regard to the general task of exploring patterns and relationships in data. Examples would be the percentage of users that will be able to complete representative tasks within a certain time or without requiring assistance, or the percentage of users that will be satisfied with the usability of the tool. It is realized that evaluations will not lead to absolute answers, and that exploratory tasks are rather open, but still we are convinced the evaluation can result in clear indications.

3.1. Test measures

The proposed assessment methodology includes three criteria (see table 3): effectiveness/user performance, usefulness, and user reactions (attitude).

1. Effectiveness focuses on the tool functionality and examines the user's performance of the tasks, and how to manipulate any parameters or controls available to complete the tasks. Effectiveness can be measured by the time spent on completing tasks, the percentage of completed tasks (Sweeney, Maguire et al. 1993; Rubin 1994; Fabricant 2001), the correctness of outcome of task performance and response, the success and accuracy (error rate and error types), the amount of time spent for help and questions, the range of functions used and the level of success in using each, the ease of use or level of difficulty, time spent to access the documentation or for help.

2. Usefulness refers to the appropriateness of the tool's functionality and relates to whether the tool meets the needs and requirements of users when carrying out tasks, the extent to which users view the tools as supportive for their goals and tasks, and the individual user's level of understanding and interpretation of the tool's results and

processes. It includes flexibility, compatibility in relation to the user's expectations (finding patterns in data, relating different attributes, comparing values of attributes for different spatial locations). This is gathered through task performance, verbal protocols, post-hoc comments and responses on a questionnaire.

3. User reactions refer to the user's attitude, opinions, subjective views, and preferences about the flexibility, compatibility (between the way the tool looks and works compared to the user's conventions and expectations). It can be measured using questionnaires and survey responses, and comments from interviews and ratings.

The specific usability measures and measuring methods used for the different tasks is described in table 3 below.

Table 3. Usability indicators used in the assessment.

	Usability indicators used		
	Effectiveness / user performance	**Usefulness**	**User reactions (attitude)**
Specific usability measures	- Correctness of outcome of task performance and response (success, percentage of completed tasks, accuracy or error rate) - Time to complete tasks - Time spent for help, documents access, guidance and support	- Compatibility and appropriateness in relation to user's expectations and goals - User's level of understanding and interpretation of the tool's results and processes	- Opinions, subjective views on the flexibility, compatibility (between the way the tool looks and works and the user's expectations), functionality, and appropriateness of the tool for the tasks - User preferences
Measuring method	- Examines tool functionality and the user's performance of the tasks and response to specific questions	- Task performance - Verbal protocols - Post-hoc comments - Responses on questionnaire - Answers to comprehension questions	- Questionnaires, interviews and survey responses, - Ratings

3.2. *Test environment and procedure*

The operational tasks described in table 2 were used in the experiment with sample cases from the dataset explored in the test. This dataset was separately explored by the test designer as an experiment and the conclusions of the exploration were used to validate the test participant's results.

The test environment consisted of a computer installed with ArcGIS, Matlab software, and the prototype visualization tool. The test environment has been selected so that noise levels are minimum, in order to avoid disrupting the test. The test sessions were individual sessions in which the participant worked in the presence of only the test administrator on the tasks using each of the different representations. Two first candidate users were used as pilot test subjects to ascertain any deficiencies in the test procedures, such as tasks descriptions, timing of each test session, the rating system, and instructions for test tasks. A revision was made based on the problems detected during pilot testing, particularly of the task description and timing. Twenty participants, including geographers, cartographers and environmental scientists, with experience in data analysis and the use of GIS were invited to take part in the test. The dataset used is related to a general geographic problem, for which all the participants have the knowledge to conduct the analyses.

The individual SOM-based graphical representations were programmed to be used separately in a window with interactive features provided in the Matlab interface (zooming, panning, rotation and 3D view). ArcGIS was used for tasks involving maps, and a free and fully functional Java-based interactive parallel coordinate plot was used, with the basic features needed for the test (brushing, automatic display of names of countries and values of variables as the mouse moves over the data records, and adding and removing attributes from the display). Participants were encouraged to interact with the interface. While completing the tasks, they were asked to report their preferences and viewpoints about the representation forms.

To ensure that participants were at ease, were fully informed of any steps, and inquiries were answered, an introduction to each session was given. The introduction explainsed the purpose of the test, and introduced the test environment and the tools to be used. Participants were informed that the test consists of testing the design and tools, not their abilities. At the end of the introduction, participants' questions were answered. The tasks were written on separate sheets and were given one at a time according to the random numbers assigned. Individual test sessions were conducted using random numbers for the order of task presentation of the graphical representations for the 10 tasks, and for the order of the graphical representations used for each task. The rationale behind the use of random numbers for the order of task presentation and the graphical representations for each of the 10 tasks (three to four graphical representations were used for each task) was to reduce the learning effect for the sample size. In the introduction, the participants were informed about the total number of tasks, but the tasks were given one at a time according to the random numbers assigned. Participants were assured that they have the option to abandon any tasks that they were unable to complete. They were left to work quietly, without any interruption unless necessary. Participants were asked to report, as they work, any problems they find or things they don't understand and were allowed to ask questions during the test.

The introduction and all the steps of the test were contained in a script so that all the participants were treated in the same way during the session and received the same information. The script describes the steps of the test in detail, and was read to each participant at the beginning of the session in order to ensure that all participants receive the same information. To allow the participants to refer back to the list of tasks as they attempt a task, a written description of the task was handed to each participant.

A logging sheet for each participant (at each session) was used to record timing, events, participant actions, concerns and comments. At the end of the session, a brief questionnaire was given to the participants to collect other comments they need to communicate.

Two forms were used to record user task performance and the different ratings. Task performance was reported by the test administrator. User ratings on usefulness (compatibility, ease of use, understanding) and user reactions (satisfaction and preferences) were reported by the participants on the second form for the different tasks and representations used.

The average time required to complete all the tasks was 90 minutes. On a logging sheet, the time for user task performance for each representation was recorded, as well as participants' opinions and comments. Participants were allowed to ask questions during the test.

3.3. Participants

The profile of test participants was a target population that included geographers, demographers, environmental scientists, and epidemiologists – likely users of such a geovisualization environment. The selected participants were GIScience domain specialists, with knowledge of the application domain (of economic development) and of similar datasets. Twenty participants from an initial list of 25 who met the profile set for the test agreed to make time for the test. They included geographers, cartographers, geologists, and environmental scientists, and all had had experience in data analysis and the use of GIS. They also had both the motivation and the qualifications to carry out the kinds of analysis being studied. Most of the participants are pursuing PhD research. The selection of the sample size (20 participants) was based on recommendations from usability engineering literature (Nielsen 1994) regarding final testing that involves actual use.

The first two candidate users were used as pilot test subjects to ascertain any deficiencies in the test procedures, such as tasks descriptions, timing of each test session, the rating system, and instructions for test tasks. A revision was made based on the problems detected during pilot testing, particularly of the task description and timing.

4. Results

The analysis of the test data is organized according to the usability measures described above: effectiveness/performance, usefulness, and user reactions. Detailed analysis of the test data was conducted using a paired-wise t-test with the different representations to compare the mean scores for the different measures. The results are also presented by experimental tasks and corresponding conceptual visualization goals. The tasks are grouped into clustering (cluster and categorize) and exploration (locate, identify, distinguish, compare, rank, distribution, associate, correlate).

4.1 Analysis of effectiveness

4.1.1. Correctness of response

Correctness of response was used as a measure of performance. A task completed with the correct response is given 1 and a task not completed or completed with the wrong response is assigned 0. The analysis of the correctness of response shows that the parallel coordinate plot performed poorly compared with maps and SOM component planes. The SOM component plane display performed well for all tasks. The map performed well generally, except for task 6 (distribution), task 2 (identify) and task 8 (compare).

The component plane display performed better than maps and the parallel coordinate plot for visualization tasks such as '*identify*', '*distribution*', '*correlate*', '*compare*' and '*associate*'. The maps were as good as component planes for tasks such as '*locate*', '*distinguish*' and '*rank*'. For these tasks (*rank*, *associate* and *distinguish)* the parallel coordinate plot performed poorly.

For the tasks '*cluster*' and '*categorize*', the SOM-based representations (unified distance matrix, 2D/3D surface and 2D/3D projection) performed equally well and far better than the parallel coordinate plot. For revealing the categories, the unified distance matrix was found less effective than the 2D/3D projection and 2D/3D surface. The 2D/3D projection was found to be more effective for finding the categories.

Further analysis of the correctness of response measure was conducted using a pair-wise comparison of the mean scores for the different representations for each conceptual visualization goal examined. Statistics of the paired sample tests are presented in table 4. The paired sample tests show significant differences ($p<0.05$) in the mean scores for the different tasks. For the task '*locate*', the map and the component plane display performed equally well (with 100% successful task completion by users), compared with the parallel coordinate plot (75% successful task completion by users). For this task, a significant difference was found between the map and the parallel coordinate plot ($P=0.021$), and between the component plane display and the parallel coordinate plot ($P=0.021$).

For the task '*identify*', the map and parallel coordinate plot performed relatively poorly (60% and 55% successful task completion by users respectively), compared with the component plane display (90%). The component plane display shows a significant difference in performance in comparison to the map ($p=0.030$) as well as to the parallel coordinate plot ($p=0.005$).

Table 4. Paired samples test for correctness of response. (*) indicates that the difference is significant (P<0.05). Udm = unified distance matrix, Pcp = parallel coordinate plot, Comp = SOM component plane display, Proj = 2D/3D projection, Surf = 2D/3D surface plot.

Conceptual visualization goal	***Representation method***	***Paired differences***					***t***	***df***	***P-value Sig. (2-tailed)***
		Mean	***Std. deviation***	***Std. error mean***	***95% confidence interval of the difference***				
					Lower	***Upper***			
Locate	Map - Pcp	0.25	0.444	0.099	0.042	0.458	2.517	19	0.021*
	Pcp - Comp	-0.25	0.444	0.099	-0.458	-0.042	-2.517	19	0.021*
Identify	Map - Pcp	0.05	0.510	0.114	-0.189	0.289	0.438	19	0.666
	Map - Comp	-0.3	0.571	0.128	-0.567	-0.033	-2.349	19	0.030*
	Pcp - Comp	-0.35	0.489	0.109	-0.579	-0.121	-3.199	19	0.005*
Distinguish	Map - Pcp	0.65	0.489	0.109	0.421	0.879	5.940	19	0.000*
	Map - Comp	0.05	0.224	0.050	-0.055	0.155	1.000	19	0.330
	Pcp - Comp	-0.6	0.598	0.134	-0.880	-0.320	-4.485	19	0.000*
Categorize	Udm - Proj	-0.15	0.489	0.109	-0.379	0.079	-1.371	19	0.186
	Udm - Surf	-0.1	0.308	0.069	-0.244	0.044	-1.453	19	0.163
	Proj - Surf	0.05	0.394	0.088	-0.134	0.234	0.567	19	0.577
Cluster	Udm - Pcp	0.45	0.510	0.114	0.211	0.689	3.943	19	0.001*
	Proj - Pcp	0.45	0.510	0.114	0.211	0.689	3.943	19	0.001*
	Surf - Pcp	0.45	0.510	0.114	0.211	0.689	3.943	19	0.001*
Distribution	Map - Pcp	-0.05	0.759	0.170	-0.405	0.305	-0.295	19	0.772
	Map - Comp	-0.65	0.489	0.109	-0.879	-0.421	-5.940	19	0.000*
	Pcp - Comp	-0.6	0.503	0.112	-0.835	-0.365	-5.339	19	0.000*
Rank	Map - Pcp	0.1	0.308	0.069	-0.044	0.244	1.453	19	0.163
	Pcp - Comp	-0.1	0.308	0.069	-0.244	0.044	-1.453	19	0.163
Compare	Map - Pcp	0.15	0.671	0.150	-0.164	0.464	1.000	19	0.330
	Map - Comp	-0.3	0.470	0.105	-0.520	-0.080	-2.854	19	0.010*
	Pcp - Comp	-0.45	0.510	0.114	-0.689	-0.211	-3.943	19	0.001*
Associate	Map - Pcp	0.25	0.716	0.160	-0.085	0.585	1.561	19	0.135
	Map - Comp	-0.2	0.410	0.092	-0.392	-0.008	-2.179	19	0.042*
	Pcp - Comp	-0.45	0.510	0.114	-0.689	-0.211	-3.943	19	0.001*
Correlate	Map - Pcp	0.35	0.671	0.150	0.036	0.664	2.333	19	0.031*
	Map - Comp	-0.15	0.366	0.082	-0.321	0.021	-1.831	19	0.083
	Pcp - Comp	-0.5	0.513	0.115	-0.740	-0.260	-4.359	19	0.000*

4.1.2. Time to complete tasks

Time to complete the tasks was used as a second variable for the performance measure. The analysis of the time taken to complete the tasks reveals some important differences between the different representations used (see figure 3). In general the component plane display required less time than the maps and the parallel coordinate plot for the different tasks. The map was faster for '*distinguish'*, but a far slower medium for comparison tasks (see figure 3).

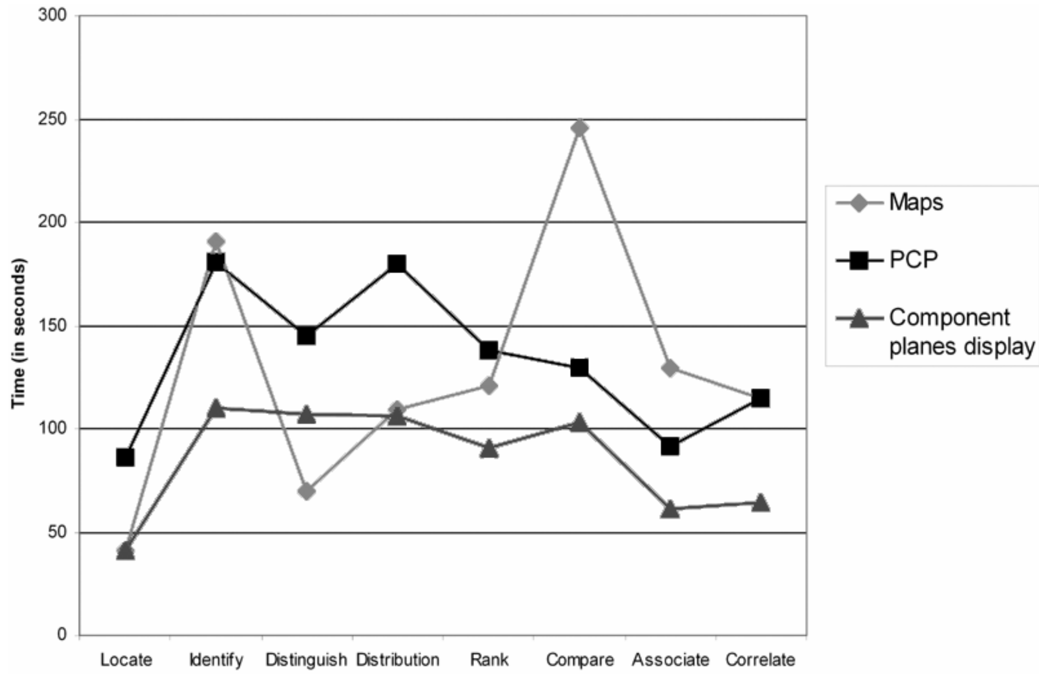


Figure 3. Time to complete tasks for three exploratory tools: map, parallel coordinate plot (PCP), and component plane display.

For the task '*locate'*, the parallel coordinate plot required double the time needed by the map and the component plane display for the same task. Thus a significant difference was found between the parallel coordinate plot and the map (p=0.005) and the component plane display (p=0.002). Table 5 provides detailed statistics on the comparison of time spent on the tasks using the different representations.

For the task '*identify'* the difference is significant between the component plane display and the parallel coordinate plot (p=0.020).

The map required a lot more time (245 seconds) than the component plane display (103 seconds) and the parallel coordinate plot (129 seconds) for the task '*compare'*. A significant difference was found between the map and the component plane display (p=0.012).

Table 5. Paired samples test for the time taken to complete the tasks. (*) indicates that the difference is significant (P<0.05). Udm = unified distance matrix, Pcp = parallel coordinate plot, Comp = SOM component plane display, Proj = 2D/3D projection, Sur f= 2D/3D surface plot.

Task		***Paired differences***					***t***	***df***	***P-value Sig. (2-tailed)***
		Mean	***Std. deviation***	***Std. error mean***	***95% confidence interval of the difference***				
					Lower	Upper			
Locate	Map - Pcp	-46.267	53.526	13.820	-75.909	-16.625	-3.348	14	0.005*
	Map - Comp	0.300	52.374	11.711	-24.212	24.812	0.026	19	0.980
	Pcp - Comp	50.200	52.757	13.622	20.984	79.416	3.685	14	0.002*
Identify	Map - Pcp	22.000	132.990	42.055	-73.136	117.136	0.523	9	0.614
	Map - Comp	83.167	140.597	40.587	-6.164	172.498	2.049	11	0.065
	Pcp - Comp	75.273	90.459	27.274	14.502	136.044	2.760	10	0.020*
Distinguish	Map - Pcp	-90.000	82.595	31.218	-166.388	-13.612	-2.883	6	0.028*
	Map - Comp	-37.200	118.804	26.565	-92.802	18.402	-1.400	19	0.178
	Pcp - Comp	3.143	151.201	57.149	-136.695	142.981	0.055	6	0.958
Categorize	Udm - Proj	-1.133	58.481	15.100	-33.519	31.252	-0.075	14	0.941
	Udm - Surf	-5.313	83.028	20.757	-49.555	38.930	-0.256	15	0.801
	Proj - Surf	2.882	82.722	20.063	-39.649	45.414	0.144	16	0.888
Cluster	Udm - Proj	-1.350	74.025	16.553	-35.995	33.295	-0.082	19	0.936
	Udm - Surf	12.100	43.086	9.634	-8.065	32.265	1.256	19	0.224
	Udm - Pcp	11.273	42.626	12.852	-17.364	39.910	0.877	10	0.401
	Proj - Surf	13.450	64.779	14.485	-16.868	43.768	0.929	19	0.365
	Proj - Pcp	25.273	106.432	32.091	-46.229	96.775	0.788	10	0.449
	Surf - Pcp	-6.636	35.870	10.815	-30.734	17.461	-0.614	10	0.553
Distribution	Map - Pcp	13.000	107.480	76.000	-952.672	978.672	0.171	1	0.892
	Map - Comp	-1.571	142.078	53.701	-132.972	129.829	-0.029	6	0.978
	Pcp - Comp	67.667	193.987	64.662	-81.445	216.778	1.046	8	0.326
Rank	Map - Pcp	-14.889	84.948	20.023	-57.133	27.355	-0.744	17	0.467
	Map - Comp	30.450	75.701	16.927	-4.979	65.879	1.799	19	0.088
	Pcp - Comp	55.944	67.753	15.969	22.252	89.637	3.503	17	0.003*
Compare	Map - Pcp	126.333	245.452	100.205	-131.253	383.919	1.261	5	0.263
	Map - Comp	132.077	160.423	44.493	35.134	229.020	2.968	12	0.012*
	Pcp - Comp	30.444	110.263	36.754	-54.312	115.200	0.828	8	0.432
Associate	Map - Pcp	40.444	83.539	27.846	-23.769	104.658	1.452	8	0.184
	Map - Comp	66.250	101.841	25.460	11.983	120.517	2.602	15	0.020*
	Pcp - Comp	32.833	83.512	24.108	-20.228	85.895	1.362	11	0.200
Correlate	Map - Pcp	25.222	98.776	32.925	-50.704	101.148	0.766	8	0.466
	Map - Comp	54.500	86.186	20.314	11.641	97.359	2.683	17	0.016*
	Pcp - Comp	54.000	48.360	15.293	19.406	88.594	3.531	9	0.006*

4.2. Usefulness and user reactions

Usefulness and user reactions were reported using a five-point scale (5 = very good, 4 = good, 3 = fairly good, 2 = poor, 1 = very poor). Usefulness includes compatibility, ease of use/flexibility, and perceived user understanding. User reactions include user satisfaction and preferences. A combined view of the different measures of usefulness and user reactions is presented in figure 4 for the tasks.

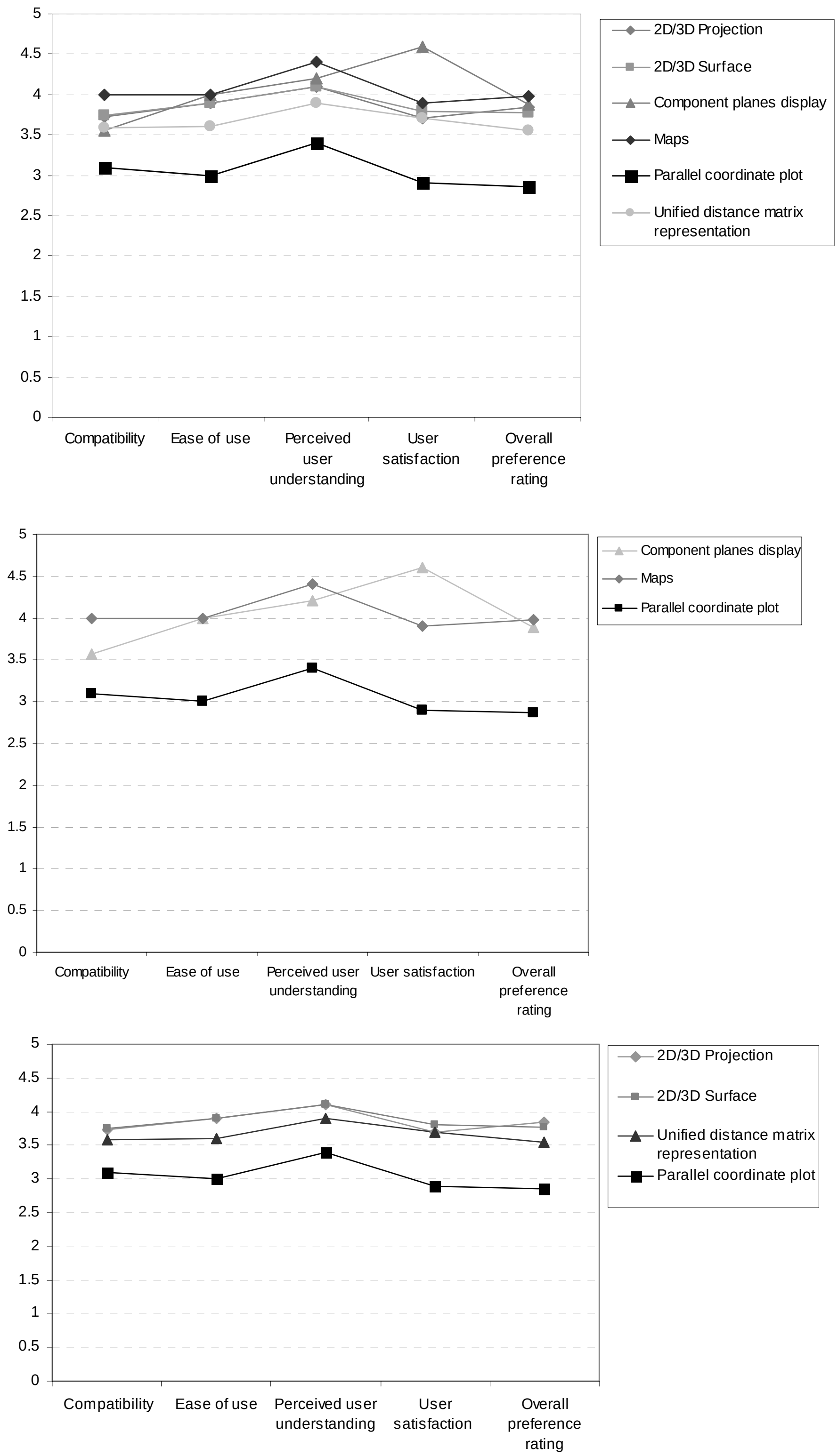


Figure 4. Overall ratings of the representations for all the different tasks combined: (a) shows all the representations for all the tasks; (b) shows tools used for detailed exploration tasks; and (c) shows tools used for visual grouping (clustering) tasks. The vertical axis represents the rating scale (5 = very good, 4 = good, 3 = fairly good, 2 = poor, 1 = very poor).

4.2.1. Compatibility with the user's expectations for the different tasks

For compatibility with the user's expectations of the tool for the tasks, the map was found more suitable (mean = 4.85 and median = 5 on the five-point scale) for the tasks *'locate', ''distinguish' and 'rank'*. The component plane display was found more appropriate for the tasks '*identify*', '*distribution*', '*compare*', '*associate*' and '*correlate*'. The parallel coordinate plot was rated generally poor (2 on the five-point scale) or fairly good (3 on the five-point scale) for all the tasks. The best ratings of the parallel coordinate plot were for the tasks '*rank*' and '*locate*', where the mean score = 3.75 and the median = 4 (same result for both tasks). These results for compatibility confirm the performance analysis presented in section 4.1. for correctness of response and time taken.

4.2.2. Flexibility/ease of use

As with compatibility, the map was found easier for the tasks '*locate*', '*distinguish*' and '*rank*'. The component plane display was found easier to use for the tasks '*identify*' and '*distribution*'. The parallel coordinate plot was generally found difficult to use, especially for the tasks '*distinguish*', '*associate*', and '*compare*', but less difficult to use for the tasks '*rank*' and '*locate*'.

4.2.3. Perceived user understanding of the representations used

The map and the component plane display were generally well understood for all the tasks The parallel coordinate plot was not well understood for some of tasks such as '*compare*', ‘*associate’*, ‘*distinguish’*, ‘*distribution’* and ‘*correlate’,* but relatively well understood for the task ‘*rank’*.

4.2.4. User satisfaction

In general users were satisfied with the component plane display and the map. The parallel coordinate plot was not satisfactory for the tasks ‘*distinguish’*, ‘*associate’*, ‘*correlate’* and ‘*distribution’*.

4.2.5. User preference rating

The overall preference rating of the tools for the different tasks revealed that the map was preferred for the tasks ‘*locate’*, ‘*distinguish’* and ‘*rank’*. The component plane display was preferred for the tasks ‘*identify’*, ‘*distribution’*, ‘*compare’* and ‘*correlate’*. The map and the component plane display were generally equally rated with regard to preference for the task ‘*associate’*. The parallel coordinate plot was not preferred for any of the tasks in the test.

5. Discussion

The analysis of the test results presented in the previous section reveal some important differences between the SOM-based representations, the map and the parallel coordinate plot as they are applied to the taxonomy of visualization tasks used for the evaluation. As proposed by Wehrend and Lewis (2000) for visual representations generally, each of the representation methods by its inherent structure seems to emphasize particular attributes and support a particular set of visual tasks or inferences.

Maps were more effective for certain visual tasks such as locate and distinguish, but less effective for the tasks of comparison, correlation, and for relating many attributes (see figure 3). Although easy to use in general for all the test participants since they are used to such visual representation of the world a major problem was that the map can show only a limited number of attributes, which is not appropriate for investigating many attributes for the dataset in a reasonable time. This would require many maps to complete some of the tasks. For visual comparison, the map was not as effective as the component plane display. It required more time for tasks that involve viewing relationships, since differences between classes geographically are not noticeable despite the colour scheme used for classification.

Component plane displays were more effective for visual perception and were also found easier to use for finding relationships and understanding the patterns. This representation was especially effective and suitable for tasks involving visual composition (Zhou and Feiner 1998), such as associate, correlate, identify, and compare. Participants reported that the component plane display did not require much effort to view the patterns and to relate different attributes in a single view. Relationships between the attributes were found to be very apparent in component planes. This ability to permit immediate information extraction at a single glance with less effort is one of the measures of the quality of a visualization (Bertin 1983). The component plane display was less effective for the task of ranking among similar data items because of the clustering. Participants needed some guidance in using the component planes, but generally found the tool easier to use after a short introduction.

Parallel coordinate plots required the participants to keep track of a lot of information before they could summarize answers for the tasks. This is an important issue in visual encoding and perception (Cleveland and McGill 1984; Cleveland and McGill 1986; MacEachren 1995), key elements in knowledge construction using visual representations. This difficultly in keeping track of the information perceived makes the parallel coordinate plot difficult for the test participants to understand. Some participants reported they found the parallel coordinate plots confusing: too many lines were used and thus the picture provided was not clear, despite the brushing feature used. Much effort was needed, patterns were difficult to see, and it required more time to examine a particular variable. This aspect was critical in the user rating (compatibility, ease of use, understanding, satisfaction and preference rating) for the effectiveness of the tool, and may explain the poor results. The visual processing of graphical displays by users (visual recognition and visual grouping) is an important factor in graphical perception (Cleveland 1993). The display of the parallel coordinate plot was found difficult to understand, although good for relating multiple variables, with its dynamic, interactive features. It was particularly inappropriate for tasks such as cluster, distinguish, and locate for patterns that are found at different locations, tasks that are related to visual attention (Zhou and Feiner 1998).

Among the clustering tools, the 2D/3D surface was found to be more comprehensible for visual grouping (proximity, similarity), and helpful for finding small differences within clusters, although it was reported that the use of fuzzy boundary made it a bit difficult to see cluster borders. The 2D/3D surface is generally preferred above the unified distance matrix. The 2D/3D projection was more used for representing proximity among data items. The unified distance matrix was found clear and helpful with the use of the hexagonal grid. These SOM-based tools for visual clustering were found better than the parallel coordinate plot.

6. Conclusion

In this paper we have presented an approach for assessing the usability and usefulness of the visual-computational analysis environment. The evaluation method emphasizes exploratory tasks and knowledge discovery support. It is based on the examination of a taxonomy of conceptual visualization goal and tasks. These tasks were decomposed into operational visualization tasks and experimental tasks related to the particular dataset used in the evaluation. New representation forms used to visualize geospatial data such as the SOM use new techniques to represent the attribute spaces. An important step in the design of such visualization tools is to understand the way users make interpretations of the information spaces. The choice of a proper representation metaphor is crucial to the successful use of the tools. To investigate the usability of the different representations, it was necessary to examine the subject's ability to perform visual tasks such as identifying clusters and relating the visual features to problems in the data exploration domain. This was realized by applying the visual taxonomy-based evaluation methodology in order to compare the use of SOM-based representations with that of maps and parallel coordinate plots.

The results of the usability testing provided some insight into the performance, and usefulness of the SOM-based representations (unified distance matrix, 2D/3D projection, 2D/3D surface, and component plane display) compared with the map and the parallel coordinate plot for specific visual tasks. For visual grouping and clustering, the SOM-based representations performed better than the parallel coordinate plot. For detailed exploration of attributes of the dataset, correlations and relationships, the SOM component plane display was found more effective than the map for visual analysis of the patterns in the data and for revealing relationships. The map was generally a better representation for tasks that involve visual attention and sequencing (locate, distinguish, rank).

The results of this test can serve as a guideline for designing geovisualization tools that integrate different representations such as maps, parallel coordinate plots and other information visualization techniques. The integration of visual tools can for example use tools such as the SOM component plane display for visual processing of relationships and correlations in the data. Results of users' exploration with such exploratory tools, can be presented in maps as the final output of the exploration process.

It is also obvious from the test that for each task a particular visual representation, being SOM visualizations, maps or even parallel coordinate plots, performs best. The avaialbitiy of the combination of the visualization result is the best possible environment to support exploratory activities.

Acknowledgments

This research was supported, in part, by the U.S. NSF (grant # EIA-9983451) and by the U.S. National Cancer Institute (grant CA95949).

References

Andrienko, N. and G. Andrienko (2005). Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach. Berlin, Springer Verlag.

Bertin, J. (1983). Semiology of graphics: diagrams, networks, maps. Madison, WI, University of Wisconsin press.

Caroll, J. M. and M. B. Rosson (1992). "Getting around the task-artifact cycle: how to make claims and design scenario." ACM Transactions on Information Systems **10**(2): 181 - 212.

Caroll, J. M. and M. B. Rosson (2003). Design Rationale as Theory. Toward a multidisciplinary science of human-computer Interaction. J. M. Caroll. San Francisco, Morgan-Kaufmann.

Caroll, J. M., M. B. Rosson, et al. (1998). "Requirements development in scenario-based design." IEEE Transactions on Software Engineering **24**(12).

Cleveland, W. S. (1993). "A Model for Studying Display Methods of Statistical Graphics." Journal of Computational and Graphical Statistics **2**: 323--364.

Cleveland, W. S. and R. McGill (1984). "Graphical perception: Theory, experimentation and application to the development of graphical methods." Journal of the American Statistical Association **79**: 531-554.

Cleveland, W. S. and R. McGill (1986). "An experiment in graphical perception." International Journal of Man-Machine Studies **25**: 491-500.

Fabricant, S. I. (2001). Evaluating the usability of the scale metaphor for querying semantic information spaces. Spatial Information Theory: Foundations of Geographic Information Science. D. R. Montello. Berlin, Germany, Springer Verlag**:** 156-171.

Fotheringham, A. S., C. Brunsdon, et al. (2000). Quantitative Geography: Perspectives on Spatial Data Analysis. London, SAGE Publications.

Fuhrmann, S., P. Ahonen-Rainio, et al. (2005). Making useful and useable Geovisualization: design and evaluation issues. Exploring Geovisualization. J. Dykes, A. M. MacEachren and M. J. Kraak. Amsterdam, Elsevier.

Fuhrmann, S. P. (2005). User-centred Design of Collaborative Geovisualization Tools. Exploring Geovisualization. J. Dykes, A. M. MacEachren and M. J. Kraak. Amsterdam, Elsevier**:** 553-576.

Gallup, L. J., J. D. Sachs, et al. (1999). Geography and economic development, Center for International Development, Havard University.

Haklay, M. and C. Tobon (2003). "Usability evaluation and PPGIS: toward a user-centered design approach." International Journal of Geographical Information Science **17**(6): 577-592.

Keller, P. and M. Keller (1992). Visual clues: Practical Data Visualization. Los Alamitos, CA, IEEE Computer Scociety Press.

Koua, E. L. and M. J. Kraak (2004a). Alternative visualization of large geospatial datasets. Cartographic Journal. **41:** 217-228.

Koua, E. L. and M. J. Kraak (2004b). Evaluating Self-organizing Maps for Geovisualization. Exploring Geovisualization. J. Dykes, A. M. MacEachren and M. J. Kraak. Amsterdam, Elsevier.

MacEachren, A. M. (1995). How maps work: representation, visualization, and design. New york, The Guilford Press.

MacEachren, A. M. and M. J. Kraak (2001). "Research challenges in geovisualization." Cartography and Geoinformation science **28**(1).

Muntz, R. R., T. Barclay, et al. (2003). IT road map to a Geospatial Future, report of the Committee on Intersections Between Geospatial Information and Information Technology. Washington, D C, National Academics Press.

Nielsen, J. (1994). Usability Engineering. San Francisco, Morgan Kaufmann.

Ogao, P. J. and M. J. Kraak (2002). "Defining visualization operations for temporal cartographic animation design." International Journal of Applied Earth Observation and Geoinformation **4**: 11-22.

Qian, L., M. Wachowicz, et al. (1997). Delineating operations for visualization and analysis of space-time data in GIS. GIS / LIS, Cincinnati.

Roberts, J. C. (2005). Exploratory Visualization with Multiple Linked Views. Exploring geovisualization. M. J. Kraak. Amsterdam, Elsevier**:** 159-180.

Rubin, J. (1994). Handbook of usability testing: How to plan, design, and conduct effective tests. New York, John Wiley & Sons, Inc.

Slocum, T. A., C. Blok, et al. (2001). "Cognitive and usability issues in geovisualization: a research agenda." Cartography and Geographic Information Science **28**(1): 61-76.

Sweeney, M., M. Maguire, et al. (1993). "Evaluating user-computer interaction: a framework." International Journal of Man-Machine Studies **38**.

Tobon, C. (2005). Evaluating Geographic Visualization Tools and Methods: An approach and Experiment Based upon User Tasks. Exploring Geovisualization. J. Dykes, A. M. MacEachren and M. J. Kraak. Amsterdam, Elsevier**:** 645-666.

Wehrend, S. and C. Lewis (2000). A problem-oriented classification of visualization techniques. IEEE Visualization.

Weldon, J., L. (1996). "Data mining and visualization." Database programming and design **9**(5).

Zhou, M. X. and S. K. Feiner (1998). Visual task characterization for automated visual discourse synthesis. Computer Human Interaction, Los Angeles, CA.