

Research

Open Access

## Geovisualization to support the exploration of large health and demographic survey data

Etien L Koua\*<sup>1,2</sup> and Menno-Jan Kraak<sup>1,2</sup>

Address: <sup>1</sup>International Institute for Geoinformation Science and Earth Observation (ITC), Enschede, P.O. Box 6, 7500 AA The Netherlands and <sup>2</sup>Faculty of Geographical Sciences, Utrecht University, P.O. Box 80115, 3508 TC Utrecht, The Netherlands

Email: Etien L Koua\* - koua@itc.nl; Menno-Jan Kraak - kraak@itc.nl

\* Corresponding author

Published: 04 June 2004

Received: 16 March 2004

*International Journal of Health Geographics* 2004, **3**:12

Accepted: 04 June 2004

This article is available from: <http://www.ij-healthgeographics.com/content/3/1/12>

© 2004 Koua and Kraak; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** Survey data are increasingly abundant from many international projects and national statistics. They are generally comprehensive and cover local, regional as well as national levels census in many domains including health, demography, human development, and economy. These surveys result in several hundred indicators. Geographical analysis of such large amount of data is often a difficult task and searching for patterns is particularly a difficult challenge. Geovisualization research is increasingly dealing with the exploration of patterns and relationships in such large datasets for understanding underlying geographical processes. One of the attempts has been to use Artificial Neural Networks as a technology especially useful in situations where the numbers are vast and the relationships are often unclear or even hidden.

**Results:** We investigate ways to integrate computational analysis based on a Self-Organizing Map neural network, with visual representations of derived structures and patterns in a framework for exploratory visualization to support visual data mining and knowledge discovery. The framework suggests ways to explore the general structure of the dataset in its multidimensional space in order to provide clues for further exploration of correlations and relationships.

**Conclusion:** In this paper, the proposed framework is used to explore a demographic and health survey data. Several graphical representations (information spaces) are used to depict the general structure and clustering of the data and get insight about the relationships among the different variables. Detail exploration of correlations and relationships among the attributes is provided. Results of the analysis are also presented in maps and other graphics.

### Background

The exploration of patterns and relationships in large health statistics and survey data is always a difficult data analysis task due to the volumes of data collected. Examples of widely used surveys include World Bank's Living Standards Measurement Survey (LSM) on economic aspects of well-being such as income and consumption, and the Demographic and Health Survey (DHS) which measures health indicators. These surveys result in useful

and vast geographic data that can help analyze geographical trends on a number of socio-demographic, health and economic situations at community, national as well as international levels. Geographical analysis of these data is based on a combination of indicators usually forming a number of composites of attributes on health, poverty or demography. In such large multidimensional datasets, the extraction of patterns and the discovery of new knowledge may be difficult, as patterns may remain hidden. New

approaches in data analysis and visualization are needed to represent such data in a visual form that can better stimulate pattern recognition and hypothesis generation, and to allow for better understanding and knowledge construction.

A number of approaches are used to address the multidimensional aspect of the analysis of these datasets. For example the Human Development Index (from UNDP), is a relative index based on measures of the life expectancy, education (literacy), and income. Poverty maps are used to compare national level indicators over time or across countries [1]. Information visualization techniques can be used in combination with other data analysis techniques to provide alternative exploration techniques for such data. A number of authors have proposed using Artificial Neural Networks as part of a strategy to improve geographical analysis of large, complex datasets [2-5]. Artificial Neural Networks have the ability to perform pattern recognition and classification, and are especially useful in situations where the data volumes are large and the relationships are unclear or even hidden. This is because of their ability to handle noisy data in difficult non-ideal contexts [6]. Particular attention has been directed to using the Self-Organizing Map (SOM) [7] neural network as a means of organizing complex information spaces [8-10], and for the creation of abstractions where conventional methods may be limited because underlying relationships are not clear or classes of interest are not obvious.

Recent effort in Knowledge Discovery in Databases (KDD) has also provided a window for geographic knowledge discovery. Data mining, knowledge discovery, and visualization methods are often combined to try to understand structures and patterns in complex geographical data [11-13]. One way to integrate KDD framework in geographic data exploration is to combine the computational analysis methods with visual analysis in a process that can support exploratory and knowledge discovery tasks. We explore the SOM in a framework for data mining, knowledge discovery, and spatial analysis, to uncover the structure, patterns, relationships and trends in the data. Some graphical representations based on information visualization techniques and cartographic methods, are then used to portray derived structures and patterns in a visual form that can allow for better understanding of the structures and the geographical processes. The use of these graphical representations (information spaces) play a role by offering visual representations of data that bring the properties of human perception to bear [14].

We present a framework for combining effective pattern extraction with the SOM and graphical representations in an integrated visual-computational environment, to pro-

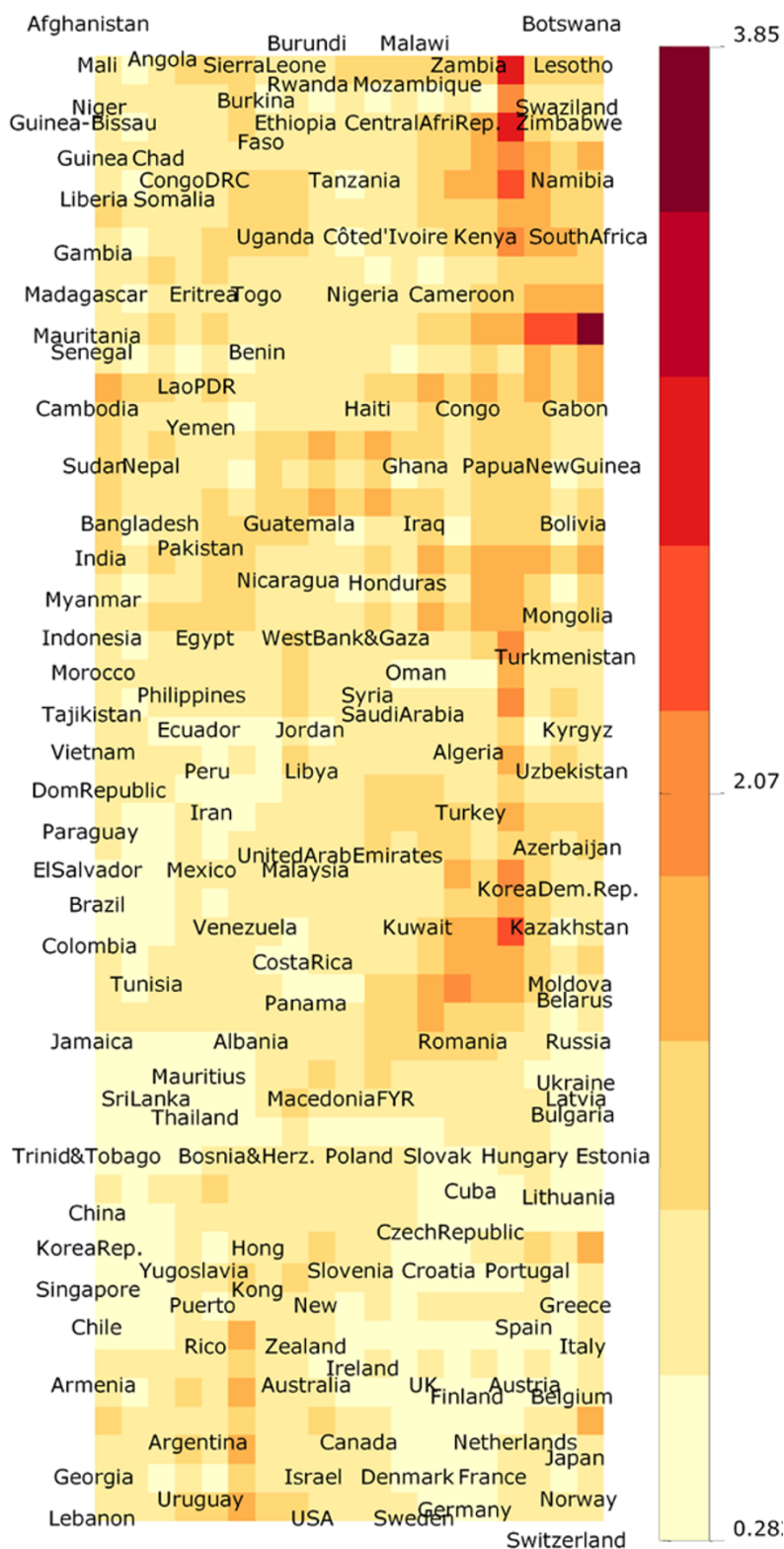
vide exploration of the data and to support knowledge acquisition through interactions. This framework is informed by current understanding of effective application of visual variables for cartographic and information design, developing theories of interface metaphors for geospatial information displays, and previous empirical studies of map and information visualization effectiveness [15]. It suggests a visual exploration of the structure of the dataset in its multidimensional space to allow for orientations for the analysis of relationships and correlations. A number of graphical interface options and exploratory task support are used to guide the user in his hypothesis testing, evaluation and interpretation of patterns from general patterns extracted to specific explorations of selected attributes and spatial locations. In this paper, an application of the method is explored for a health and demographic survey data to provide some understanding of the complex relationships between socio-economic indicators, geographical locations, and health and poverty status. The ultimate goal is to support visual data mining and exploration, and gain insights on appropriate underlying distributions, patterns and trends, and therefore contribute to enhance the understanding of the geographical processes.

## Results and discussion

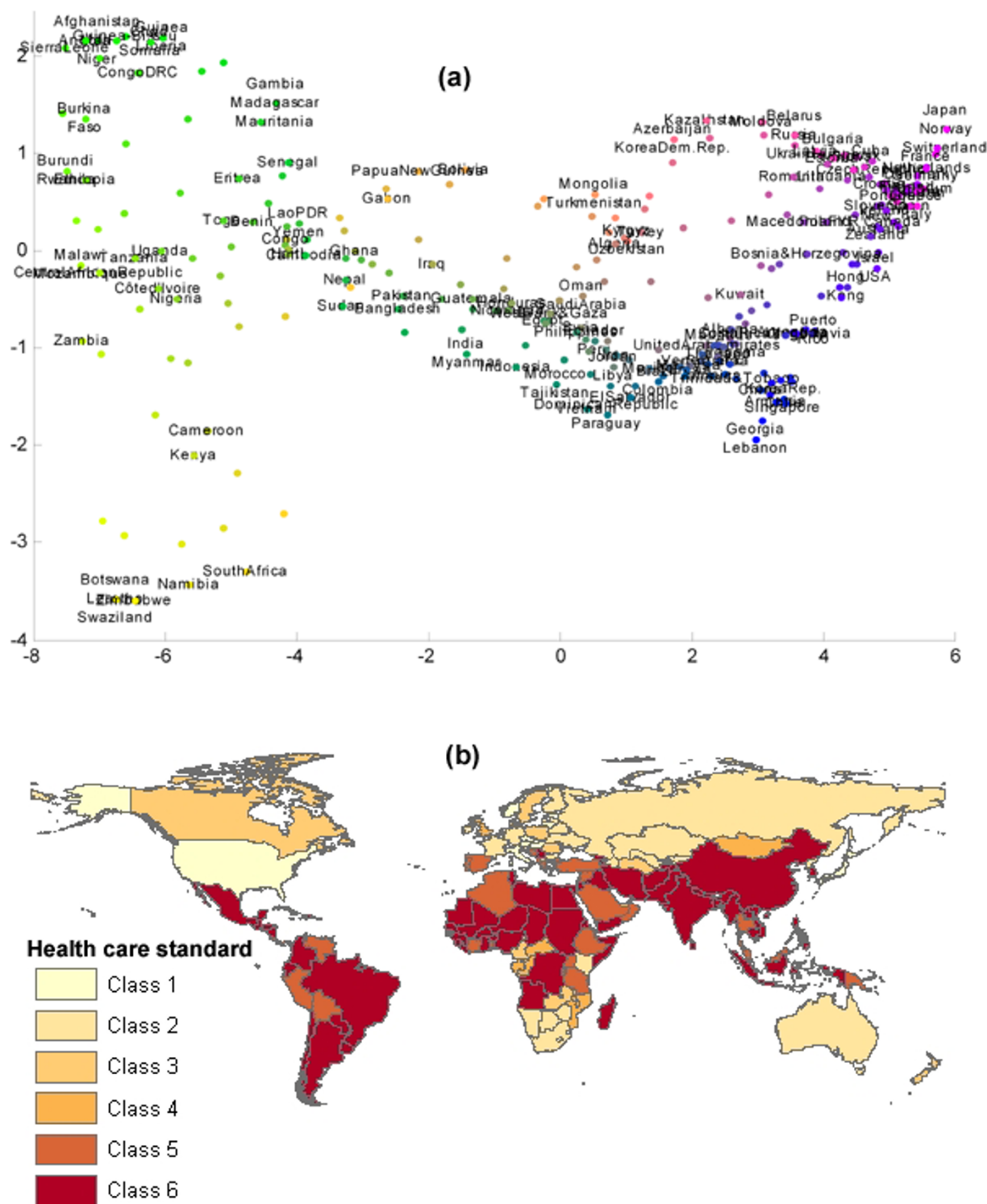
In this section, an example of exploration of (potential) patterns in data is presented using the different visualization techniques described in methodology section. For illustration, a dataset from the human development network (world bank) containing 30 basic indicators for health and living standard including health expenditure, health risk factors, mortality, reproductive health, for 152 countries is used. Although this is not a very large dataset, the approach can be applied for far larger datasets. The idea is to find multivariate patterns and relationships among different attributes and countries. Complex correlations in this kind of statistical data can be portrayed using the Self-Organizing Map to visualize the complex joint effect of the factors related to health as contained in the dataset.

### Exploration of the general patterns and clustering

The proposed approach offers a number of visualizations to show the clustering structure and similarity (patterns). These techniques use a distance matrix to show distances between neighbouring SOM network units. The most widely distance matrix technique used is the U-matrix [16]. It contains the distances from each unit center to all of its neighbours. The neurons of the SOM network are represented here by rectangular cells (see Figure 1). The distance between the adjacent neurons is calculated and presented with different colourings. A dark colouring between the neurons corresponds to a large distance and thus represents a gap between the values in the input



**Figure 1**  
**Representation of the general patterns and clustering in the input data with a distance matrix visualization** The Unified Distance Matrix showing clustering and distances between positions on the map. Countries having similar characteristics according to the multivariate attributes (health and health care situation) are found close to each other.



**Figure 2**  
**Alternative representations of the SOM general clustering of patterns** Alternative representations of the SOM general clustering of the data with projection of the SOM results in 2D space (a). This graphic can be interactively manipulated (rotation, panning, zooming, walkthrough) and be viewed in 3D space. The output of the SOM (the similarity coding extracted from the computational analysis) can be projected on a geographic map, to display the common characteristics of the countries according to the multivariate attributes (b).

space. A light colouring between the neurons signifies that the vectors are close to each other in the input space. Thus light areas represent clusters and dark areas cluster separators (a gap between the values in the input space). This representation can be used to visualize the structure of the input space and to get an impression of otherwise invisible structures in a multidimensional data space. This distance matrix representation shows not only the values at map units but also the distances between map units. In Figure 1, the structure of the data set is visualized in a distance matrix representation. Countries having similar characteristics based on the multivariate attributes are positioned close to each other, and the distance between them represents the degree of similarity or dissimilarity. The different clusters on the map can be automatically encoded with different colors, to allow a geographic view of the similarity mapping based on the multivariate attributes. These common characteristics representation can be regarded as the health standard for these countries (Figure 2b). On the geographic map, each country is assigned a color describing the its health standard type in relation to other countries. In contrast to other projection methods in general, the SOM does not try to preserve the distances directly but rather the relations or local structure of the input data. While the distance matrix representation is a good method for visualizing clusters, it does not provide a very clear picture of the overall shape of the data space because the visualization is tied to the SOM grid. Alternative representations to the distance matrix representation can be used: 2D and 3D projections (using projection methods such as the Sammon's mapping and PCA to project SOM results), 2D and 3D surface plots, and component planes. These techniques will be described in the next paragraphs. In Figure 2a, the projection of the SOM offers a view of the clustering of the data with data items depicted as coloured nodes. Similar data items are grouped together with the same type or colour of markers. Size, position and colour of markers can be used to depict the relationships between the data items. This gives an informative picture of the global shape and the overall smoothness of the SOM in 2D or 3D space.

Exploration can be enhanced by rotation, zooming, and selection. The clustering structure can also be viewed as 2D or 3D surfaces representing the distance matrix (Figure 3 and Figure 4) using color value to indicate the average distance to neighbouring map units. This is a spatialization [17] that uses a landscape metaphor to represent the density, shape, and size or volume of clusters. Unlike the projection in Figure 2a that shows only the position and clustering of map units, areas with uniform color are used in the surface plots to show the clustering structure and relationships among map units. In the 3D surface (Figure 4), color value and height are used to represent the region-

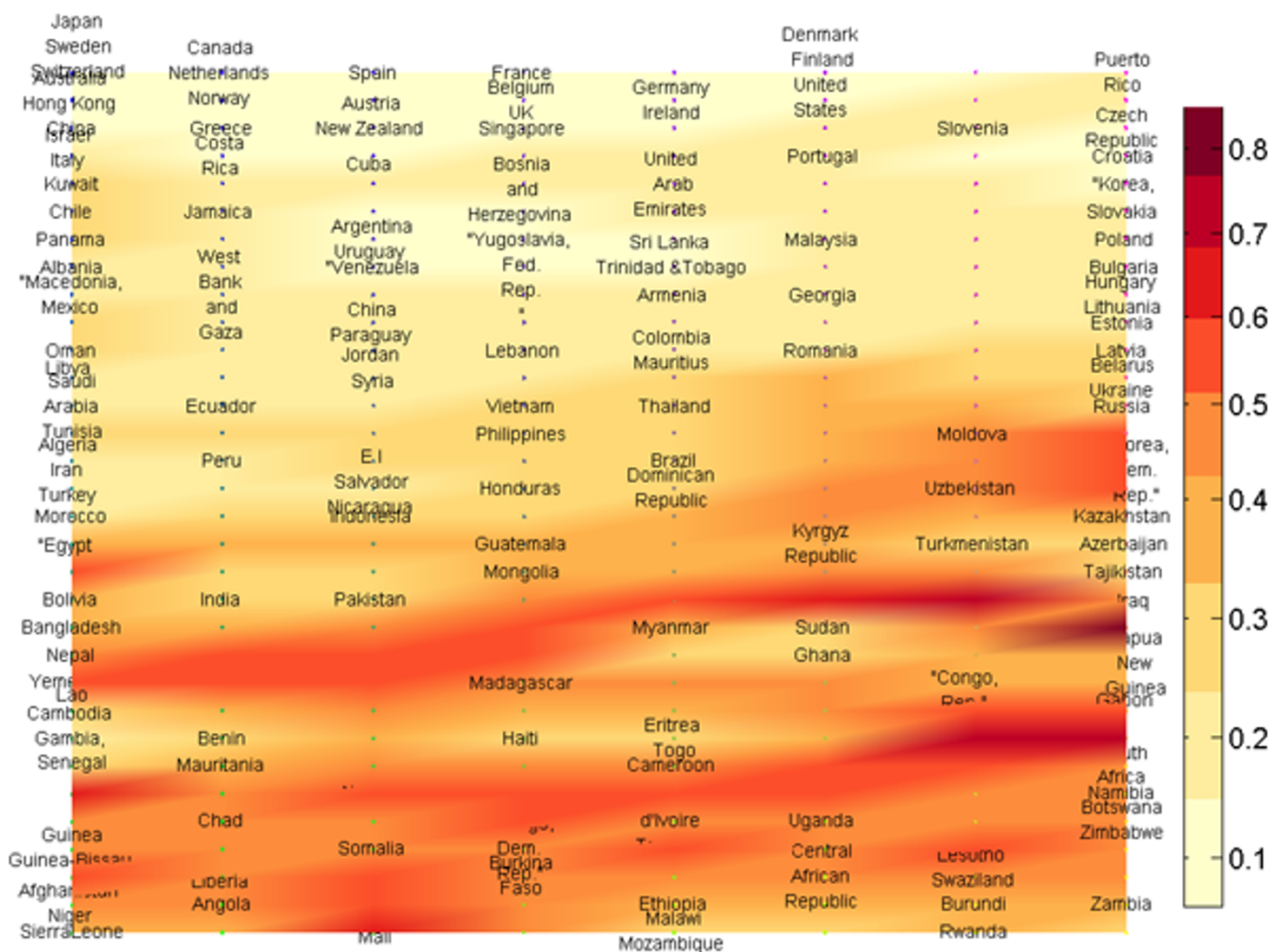
alization of map units according to the multidimensional attributes.

To illustrate the use of the 2D and 3D surface plots, we present a selection of variables of the dataset related to mortality (Figure 3 and Figure 4).

#### **Exploration of correlations and relationships**

The correlations and relationships in the input data space can be easily visualized using the component planes visualization (Figure 5). The component planes show the values of the map elements for different attributes and how each input vector varies over the space of the SOM units (here representing countries). This can be appropriate for viewing and exploring correlations and relationships. Comparatively with the maps, patterns and relationships among all the attributes can be easily examined in a single visual representation using the SOM component planes visualization (Figure 5a). Since the SOM represents the similarity clustering of the multivariate attributes, the visual representation becomes more accessible and easy to use for exploratory analyses. This kind of spatial clustering makes it possible to conduct exploratory analyses to help in identifying the causes and correlates of health problems [18] when overlaid with environmental, social, transportation, and facilities data. These map overlays have been important hypothesis-generating tools in public health research and policy-making [19].

Unlike standard choropleth maps, the position of the map units (which is the same for all displays) is determined during the training of the network, according to the characteristics of the data samples. A cell here can represent one or several political units according to the similarity in the data. Two variables that are correlated will be represented by similar displays. In Figure 5a, all the components are displayed and a selection of one example attribute is made more visible for the analysis in Figure 5b. From the view in Figure 5a, correlations and relationships can be explored, and hypotheses can be made. It becomes easy to see for example that that the HIV prevalence rate has had an impact on life expectancy for the countries most affected by the epidemic (Botswana, Lesotho, Swaziland, Namibia, Zimbabwe and South Africa) if we consider life expectancy in 1980 and that of 2001; or that the death rate is relatively higher in countries with less developed health care (number of hospital beds, number of physicians per 1000 inhabitants) such as the African countries. To enhance visual detection of the relationships and correlations, the components can be ordered so that variables that are correlated are displayed next to each other. The kind of visual representation (imagery cues) provided in the SOM component planes visualization can be used as an effective tool to visually detect correlations among operating variables in a large



**Figure 3**  
**Exploration of mortality in 2D surface** The 2D surface shows a flat organizing of the clusters and indicates the average distance among the different countries according to several attributes related to mortality in the dataset (Life expectancy in 1980, Life expectancy in 2001, Infant mortality rate in 1980, Infant mortality rate in 2001, Under 5 mortality rate in 1980, Under 5 mortality rate in 2001, Adult male mortality rate in 2001, Adult female mortality rate in 2001, Percentage of male over 65 years old 2002, Percentage of female over 65 years in 2002).

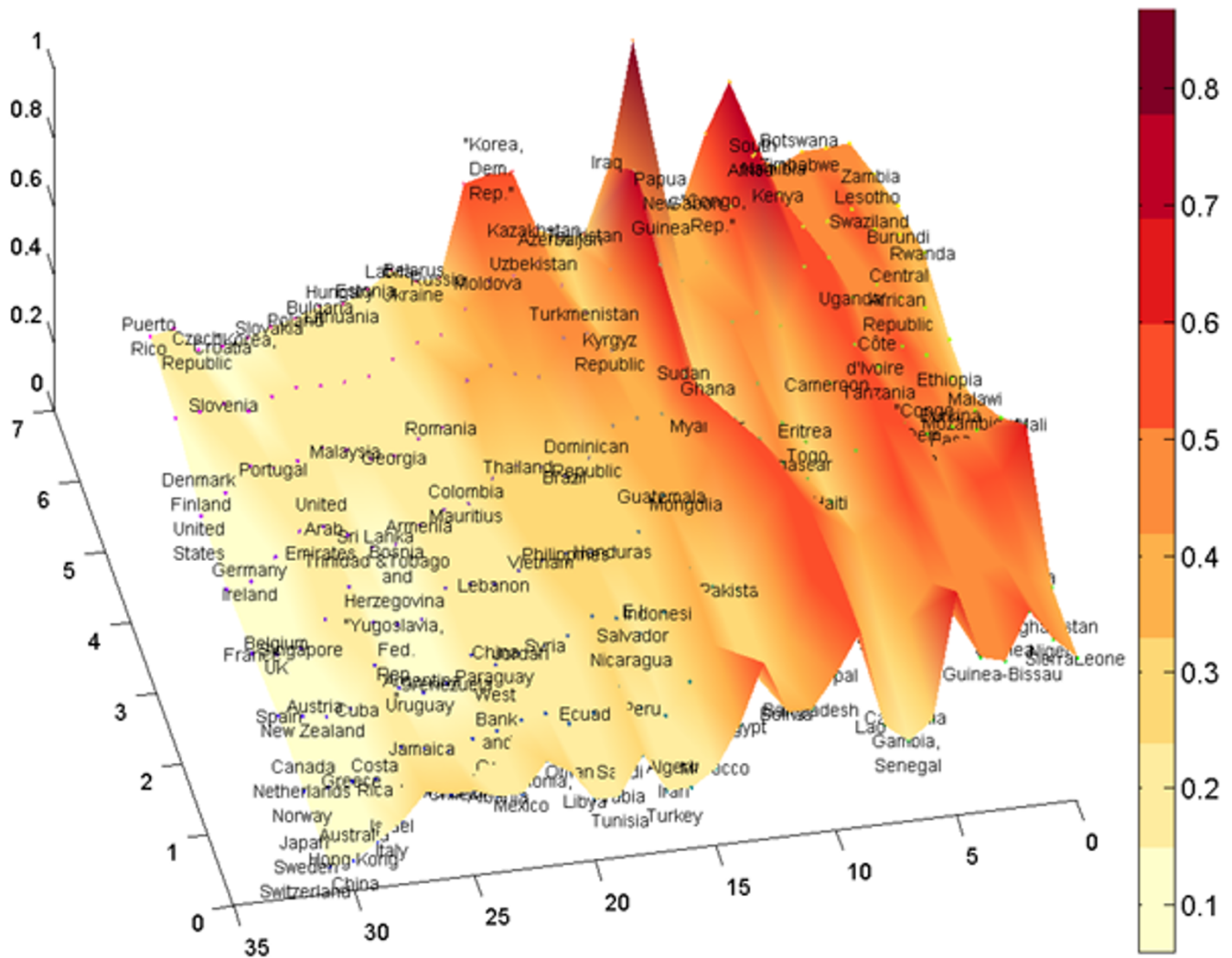
volume of multivariate data, facilitates visual detection, and has an impact on knowledge construction [20]. New knowledge can be unearthed through this process of exploration, which can be followed by the identification of associations between attributes, and finally the formulation and ultimate testing of hypotheses.

Geographic maps can be made to represent the result of this reasoning process for better geographical exploration and comparison (Figure 5c).

**Conclusions**

In this paper we have presented an approach to combine visual and computational analysis for exploratory visualization intended to contribute to the analysis of large volumes of health survey data. The approach focuses on the effective application of computational algorithms to extract patterns and relationships in large datasets, and visual representation of derived information that involves effective use of visual variables used in such complex information spaces to facilitate knowledge construction. A number of visualization techniques were explored with the objective to support visual exploration and knowledge construction. Users can perform a number of exploratory

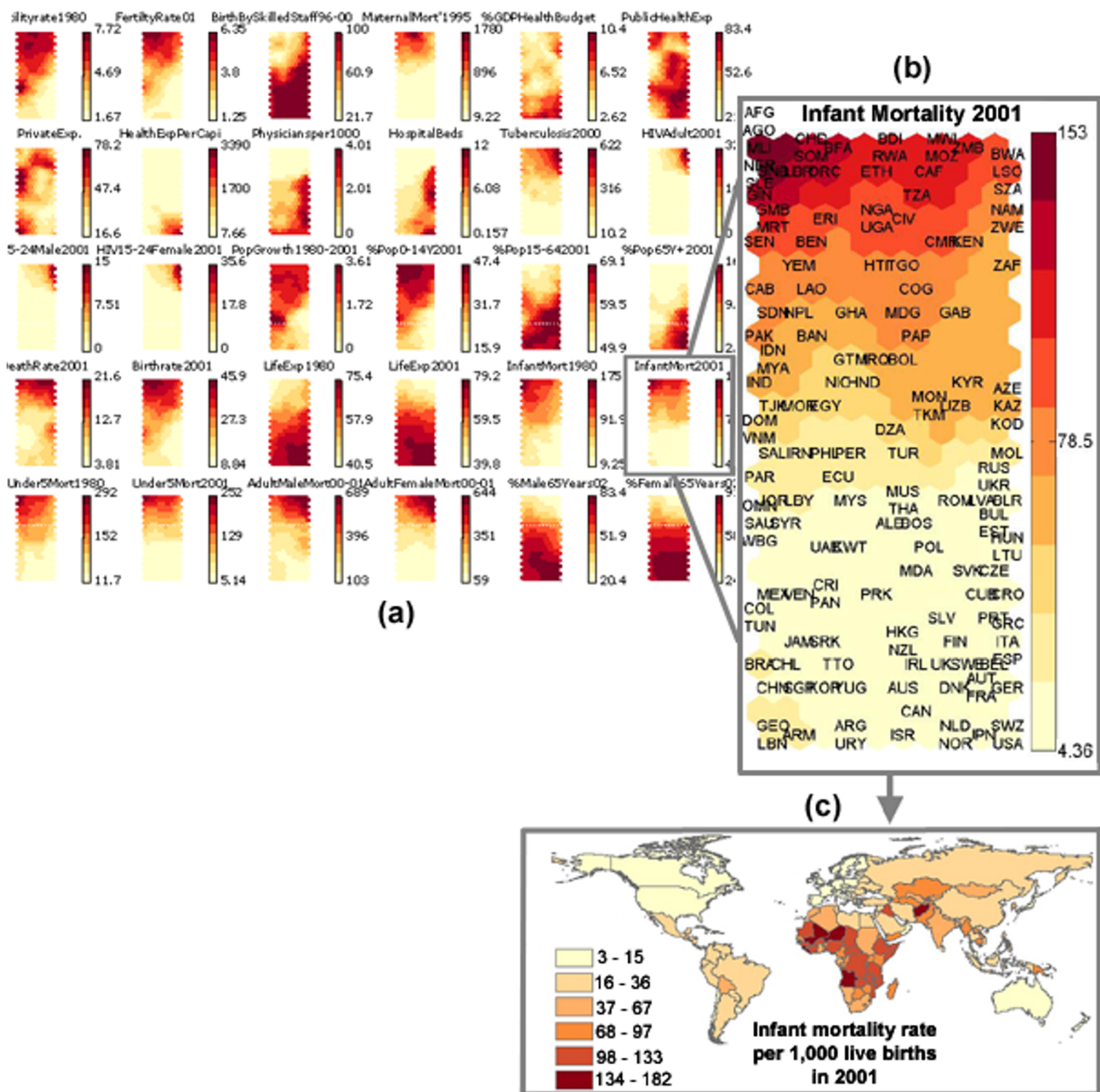




**Figure 4**  
**Exploration of mortality in 3D surface** The 3D surface plot shows a better view of the distance between the clusters for same selected of variables on mortality in the dataset shown in Figure 3 (Life expectancy in 1980, Life expectancy in 2001, Infant mortality rate in 1980, Infant mortality rate in 2001, Under 5 mortality rate in 1980, Under 5 mortality rate in 2001, Adult male mortality in 2001, Adult female mortality rate in 2001, Percentage of male over 65 years old 2002, Percentage of female over 65 years in 2002). Here distance between map units is represented by height value along the Z axis and color. It becomes easier to see the different clusters and view the distances between the different countries according to the select attributes related to mortality in the dataset.

tasks to understand the structure of the dataset as a whole and also to explore detailed information on individual or selected attributes of the dataset such as finding correlations and the relationships among attributes. With this respect, the SOM computational analysis can support exploratory visualization and the knowledge discovery process when integrated with appropriate visual exploration tools. Interactive manipulation (zooming, rotation, panning, filtering, and brushing) of the graphical repre-

sentations can enhance user goal specific querying and selection from the general patterns extracted to more specific user selection of attributes and spatial locations for exploration, hypothesis testing, and ultimately knowledge construction. We have extended these alternative representations of the SOM results used to highlight different characteristics of the computational solution and integrated them with other graphics into multiple views to allow brushing and linking, for exploratory analysis and



**Figure 5**  
**Detail exploration of the dataset using the SOM component visualization** All the components can be displayed to reveal the relationships between the variables and the spatial locations (countries) in (a). Selected components related to a specific hypothesis can be further explored and to facilitate visual recognition of relationships among selected variables. Here we display one example of attribute on child mortality (b). Geographic maps of components corresponding to hypothesis found in the exploration can be displayed (c).

knowledge discovery purposes. Results of the exploratory process can be presented in interactive maps. This link

between the attribute space visualization based on the SOM, the geographic space with maps representing the



SOM results, and other graphics such as parallel coordinate plots, in multiple views can provide alternative perspectives for better exploration, evaluation and interpretation of patterns and ultimately support for knowledge construction.

## Methods

### **A framework to support visual exploration of large geographic data**

The proposed framework explores ways to effectively extract patterns using data mining based on the SOM and to represent the results using graphical representations for visual exploration. As presented in Figure 6, the data mining stage allows to construct a clustering (similarity matrix) of the multidimensional input space using the SOM training algorithm tool (SOM Toolbox) and graphics processing with Matlab software. From this computational process, global structure and patterns can be represented with graphical representations and maps (geographic view) of similarity results. More exploration can be made on relationships and correlations among the attributes. The framework includes spatial analysis, data mining and knowledge discovery methods, supported by interactive tools that allow users to perform a number of exploratory tasks to understand the structure of the dataset as a whole and also explore detailed information on individual or selected attributes of the dataset. Different representation forms are integrated and support user interaction for exploratory tasks to facilitate the knowledge discovery process. They include some graphical representations based on the SOM, maps and other graphics such as parallel coordinate plots. Cartographic methods support this design for effective use of visual variables with which the visualizations are depicted. The graphical representations can be interactively manipulated in Matlab graphical interface using rotation, zooming, panning, and brushing.

### **The Self-Organizing Map algorithm**

The Self-Organizing Map [7] is an Artificial Neural Network used to map multidimensional data onto a low dimensional space, usually a 2D representation space (see Figure 7). The network consists of a number of neural processing units (neurons) usually arranged on a rectangular or hexagonal grid, where each neuron is connected to the input. The goal is to group nodes close together in certain areas of the data value range. Each of the units  $i$  is assigned an  $n$ -dimensional weight vector  $m_i$  that has the same dimensionality as the input patterns.

What changes during the network training process, are the values of those weights. Each training iteration  $t$  starts with the random selection of one input pattern  $x(t)$ . Using Euclidean distance between weight vector and input pattern, the activation of the units is calculated.

The unit with the lowest activation is referred to as the winner,  $c$ , of the training iteration:

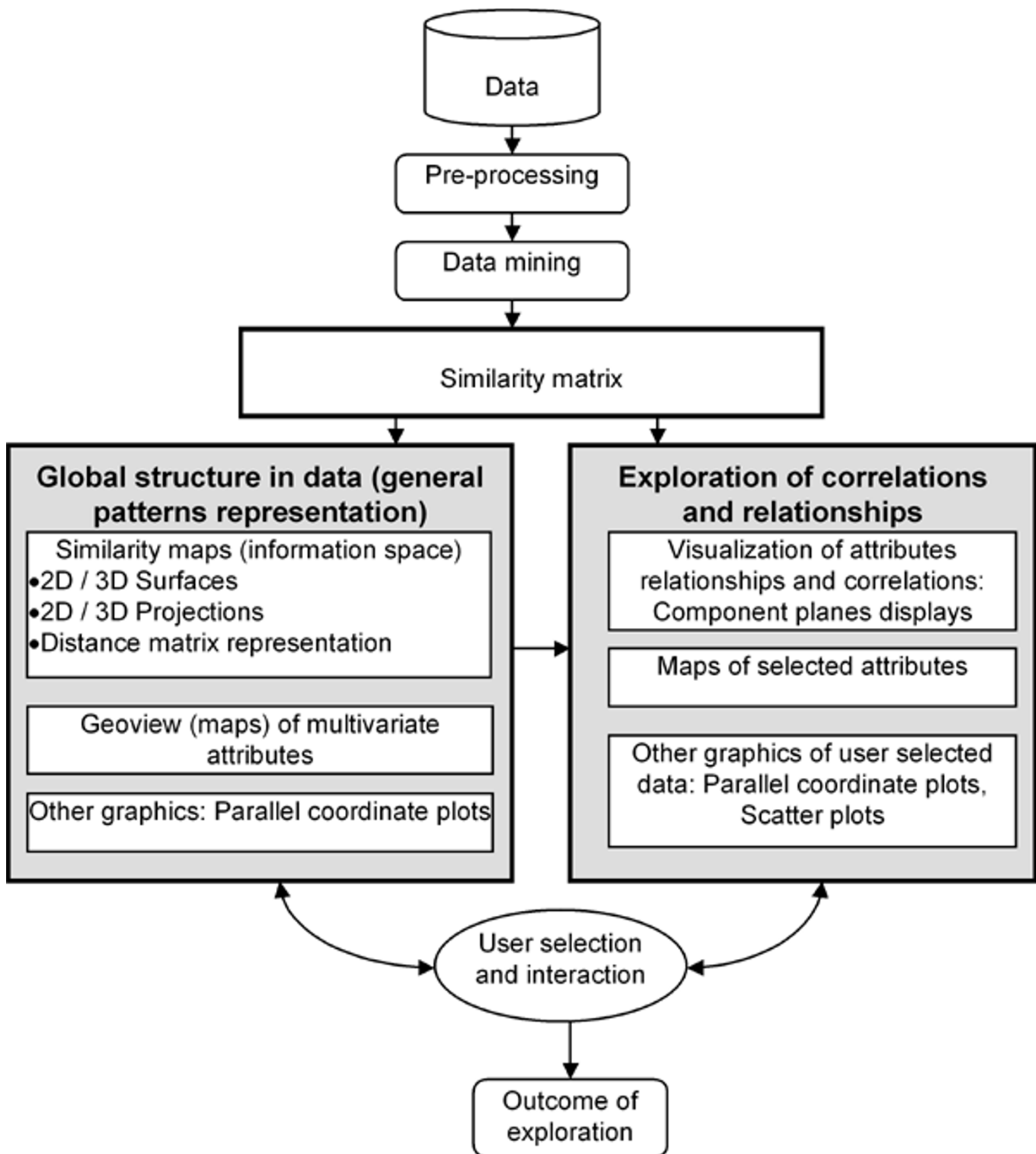
$$m_c(t) = \min_i \{ \|x(t) - m_i(t)\| \} \quad (1)$$

Finally the weight vector of the winner as well as the weight vectors of selected units in the neighbourhood of the winner are adapted to represent the input pattern. At each step  $t$  of the random sequence of the given  $x(t)$  values, the values of  $m_i$  are gradually and adaptively changed in the following adaptation process:

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)] \quad (2)$$

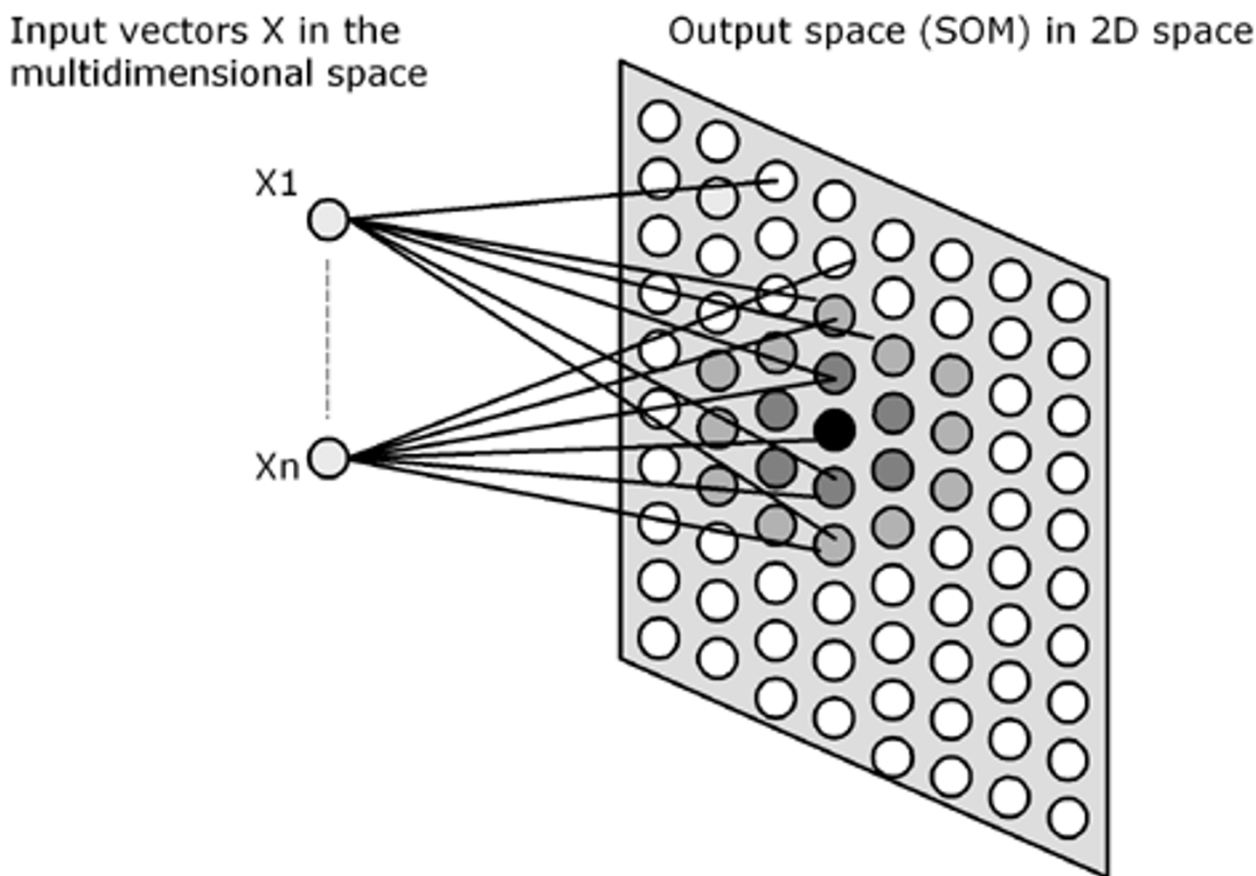
The degree of adaptation in the neighbourhood is characterized by a neighbourhood function  $h$ , which is a decreasing function of the units from the winning unit on the map lattice until no noticeable changes are observed. As a result of a general adaptation process, a number of units in the neighbourhood of the winner lead to a spatial clustering of similar input patterns in neighbouring parts of the SOM.

The resultant maps (SOMs) are organized in such a way that similar data are mapped onto the same node or to neighboring nodes in the map. This leads to a spatial clustering of similar input patterns in neighboring parts of the SOM and the clusters that appear on the map are themselves organized internally. This arrangement of the clusters in the map reflects the attribute relationships of the clusters in the input space. For example, the size of the clusters (the number of nodes allotted to each cluster) is reflective of the frequency distribution of the patterns in the input set. Actually, the SOM uses a distribution preserving property which has the ability to allocate more nodes to input patterns that appear more frequently during the training phase of the network configuration. It also applies a topology preserving property, which comes from the fact that similar data are mapped onto the same node, or to neighboring nodes in the map. In other words, the topology of the dataset in its  $n$ -dimensional space is captured by the SOM and reflected in the ordering of its nodes. This is an important feature of the SOM that allows the data to be projected onto the lower dimension space while roughly preserving the order of the data in its original space. Another important feature of the SOM for knowledge discovery in complex datasets is the fact that it is an unsupervised learning network meaning that the training patterns have no category information that accompany them. Unlike supervised methods which learn to associate a set of inputs with a set of outputs using a training data set for which both input and output are known, SOM adopts a learning strategy where the similarity relationships between the data and the clusters are used to classify and categorize the data. The SOM can be



**Figure 6**

**Data exploration framework** From the computational process, global structure and patterns can be visualized with graphical representations and maps of similarity results. Relationships and correlations among the attributes are presented with interactive graphical representations, maps and other graphics such as parallel coordinate plots.



**Figure 7**  
**The Self-Organizing Map structure.** Selection of a node and adaptation of neighbouring nodes of the neural network to the input data.

useful as a knowledge discovery tool in database methodology since it follows the probability density function of underlying data. It also offers visual representations that enable easy data exploration.

**Visual data mining and knowledge discovery**

One approach to analysis of large amount of data is by using data mining and knowledge discovery methods. The main goal of data mining is identifying valid, novel, potentially useful and ultimately understanding patterns in data [21]. Generally, three general categories of data mining goals can be identified [22]: explanatory (to explain some observed events), confirmatory (to confirm a hypothesis), exploratory (to analyze data for new or unexpected relationships). Typical tasks, for which data mining techniques are often used, include clustering, classification, generalization and prediction. These techniques vary from traditional statistics to artificial

intelligence and machine learning. The most popular methods include decision trees (tree induction), value prediction, and association rules often used for classification [23]. Artificial Neural Networks are particularly used for exploratory analysis as non-linear clustering and classification techniques. For example, unsupervised neural networks such as the Self-Organizing Map are a type of neural clustering and neural architecture such as back-propagation and feedforward are neural induction methods used for classification (supervised learning). The algorithms used in data mining are often integrated into Knowledge Discovery in Databases (KDD), a larger framework that aims at finding new knowledge from large databases. While data mining deals with transforming data into information or facts, KDD is a higher-level process using information derived from data mining process to turn it into knowledge or integrate it into prior knowledge. In general, KDD stands for discovering and visualiz-

ing the regularities, structures and rules from data [23], discovering useful knowledge from data [21] and for finding new knowledge. It consists of several generic steps, namely data pre-processing, transformation (dimension reduction, projection), data mining (structure mining) and interpretation / evaluation.

New developments in data mining and KDD has provided a window for geographic data mining and knowledge discovery, which has become an established field in geographic visualization [11,12,23-27]. This framework has been used in geospatial data exploration [11-13,23,28] to discover unexpected correlation and causal relationships, and understand structures and patterns in complex geographical data. The promises inherent in the development of data mining and knowledge discovery processes for geographical analysis include the ability to yield unexpected correlation and causal relationships. Since the dimensionality of the dataset is very high, it is often ineffective to work in such high dimension space to search for patterns. We use the SOM algorithm as a data mining tool to project input data into an alternative measurement space based on similarities and relationships in the input data that can aid the search for patterns. It becomes possible to achieve better results in such similarity space rather than the original attribute space [29]. As described in the previous section, the SOM adapts its internal structures to structural properties of the multidimensional input such as regularities, similarities, and frequencies. These properties of the SOM can be used to search for structures in the multidimensional input. Graphical representations are then used to enable visual data exploration allowing the user to get insight into the data, evaluate, filter, and map outputs. This is intended to support visual data mining [30] by allowing several variables and their interactions to be inspected simultaneously, and receive feedback from the knowledge discovery process by means of interaction techniques that support the process [31].

#### **Integrating the computational analysis and visualization**

One of the advantages of the SOM over other types of neural network is that it is suitable for visualization. The outcome of the computational process can easily be portrayed through visual representation. The first level of the computational analysis described above provides a mechanism for extracting patterns from the data. The output of this computational process is depicted using graphical representations (information spaces) to facilitate human perception and cognitive processes [14,32], by offering visualizations of the general structure of the dataset (clustering), as well as the exploration of relationships among attributes. Several graphical representations provide ways for representing similarity (patterns), relationships, including a distance matrix representation, 2D and 3D projections, 2D and 3D surfaces, and component

planes visualization. They highlight different characteristics of the computational solution and integrate them with other graphics into multiple views to allow brushing, linking, panning, zooming, rotation and 3D views for exploratory analysis and knowledge discovery purposes and enhance exploration. The resulting information spaces suggest and take advantage of natural environment metaphor characteristics such as '*near = similar, far = different*' [12], which is epitomized by Tobler's first law of geography [33]. This is an example of spatialization, an approach discussed more generally by [17]. Distance (similarity between data items), regions (aggregation of similar data items), and scale (level of detail in a database) are examples of spatial metaphors used in these representation spaces [34]. The scale allows exploration of the information space at multiple levels of detail, and provides the potential for hierarchical grouping of items, and revealing categories or classifications. Coordinate systems allow determining distance and direction, from which other spatial relationships (size, shape, density, arrangement) may be derived. We integrate the graphical representations mentioned above and maps to represent the attribute space. Using multiple views, interactions between several variables can be presented simultaneously over the space of the SOM, maps and parallel coordinate plots. This can emphasize visual change detection and the monitoring of the variability through the attribute space. These alternative and different views on the data can help stimulate the visual thinking process that is characteristic for visual exploration. Four goals of the exploration are emphasized:

- Patterns discovery (through similarity representations)
- Correlations and relationships exploration for hypothesis generation
- Exploration of the distribution of the dataset on the map
- And the detection of irregularities in the data

#### **Authors' contributions**

EK carried out the research and drafted the manuscript. MK reviewed the manuscript.

#### **References**

1. Deichmann V: **Geographic aspects of inequality and poverty**. 1999.
2. Schaale M, Furrer R: **Land surface classification by Neural Networks**. *International Journal of Remote Sensing* 1995, **16**:3003-3031.
3. Openshaw S, Turton I: **A parallel Kohonen algorithm for the classification of large spatial datasets**. *Computers-and-Geosciences* 1996, **22**:1019-1026.
4. Gahegan M: **On the application of inductive machine learning tools to geographical analysis**. *Geographical Analysis* 2000, **32**:113-139.
5. Gahegan M, Takatsuka M: **Dataspace as an organizational concept for the neural classification of geographic datasets**.

- Fourth International Conference on GeoComputation Fredericksburg, Virginia, USA; 1999:25-28.
6. Openshaw S, Openshaw C: **Artificial Intelligence in geography**. Chichester, John Wiley & Sons; 1997.
  7. Kohonen T: **Self-Organization and Associative memory**. 3rd Edition edition. Spring-Verlag; 1989.
  8. Girardin L: **Mapping the virtual geography of the World Wide Web**. Fifth International World Wide Web conference Paris, France; 1995.
  9. Chen C: **Information visualization and Virtual Environments**. London, Springer-Verlag; 1999.
  10. Fabricant SI, Buttenfield B: **Formalizing semantic spaces for information access**. *Annals of the Association of American Geographers* 2001, **91**:263-280.
  11. Gahegan M, Harrover M, Rhyne TM, Wachowicz M: **The integration of geographic visualization with Databases, Data mining, Knowledge Discovery Construction and Geocomputation**. *Cartography and Geographic Information Science* 2001, **28**:29-44.
  12. MacEachren AM, Wachowicz M, Edsall R, Haug D, Masters R: **Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with knowledge discovery in databases methods**. *International Journal of Geographical Information Science* 1999, **13**:311-334.
  13. Wachowicz M: **The role of geographic visualization and knowledge discovery in spatio-temporal modeling**. *Publications on Geodesy* 2000, **47**:27-35.
  14. Card SK, Mackinlay JD, Shneiderman B: **Readings in Information Visualization. Using Vision to Think**. San Francisco, Morgan Kaufmann Publishers; 1999.
  15. Koua EL, Kraak MJ: **Evaluating Self-organizing Maps for Geovisualization**. *Exploring Geovisualization* Edited by: Dykes J, MacEachren A M and Kraak M J. Amsterdam, Elsevier; 2004.
  16. Ultsch A, Siemon H: **Kohonen's self-organizing feature maps for exploratory data analysis**. *Proceedings International Neural Network Conference INNC'90P* Dordrecht, The Netherlands; 1990:305-3308.
  17. Fabrikant SI, Skupin A: **Cognitively Plausible Information Visualization**. *Exploring GeoVisualization* Edited by: Dykes J, MacEachren A and Kraak MJ. Amsterdam, Elsevier; 2004.
  18. Cromley EK, McLafferty SL: **GIS and public health**. New York, The Guilford Press; 2002.
  19. Croner C, Pickle L, Wolf D, White A: **A GIS approach to hypothesis generation in epidemiology**. *ASPRS/ACSM technical papers Volume 3*. Edited by: Voss A W. Washington, DC, ASPRS/ACSM; 1992:275-283.
  20. Keller P, Keller M: **Visual clues: Practical Data Visualization**. Los Alamitos, CA, IEEE Computer Society Press; 1992.
  21. Fayyad U, Piatetsky-Shapiro G, Smyth P: **From data mining to knowledge discovery in databases**. *Artificial Intelligence Magazine* 1996, **17**:37-54.
  22. Weldon J,L: **Data mining and visualization**. *Database programming and design* 1996, **9**.
  23. Miller HJ, Han J: **Geographic data mining and knowledge discovery**. London, Taylor and Francis; 2001.
  24. Weijan W, Fraser D: **Spatial and temporal Classification with Multiple Self-Organizing Maps**. *Society of Photo-optical instrumentation* 1996, **2955**:307-314.
  25. Sibley D: **Spatial applications of exploratory data analysis**. Norwich, Geo Books; 1988.
  26. Roddick JF, Lees BG: **Paradigms for Spatial and Spatio-Temporal Data Mining**. *Geographic Data Mining and Knowledge Discovery* Edited by: Miller H J and Han J. London, Taylor and Francis; 2001:33-49.
  27. Liu W, Gopal S, Woodcock C: **Spatial data mining for classification, visualization and interpretation with ARTMAP Neural Network**. *Geographic data mining and knowledge discovery* Edited by: Miller H J and Han J. London, Taylor and Francis; 2001.
  28. Openshaw S, Cross A, Charlton M: **Building a prototype geographical correlates machine**. *International Journal of Geographical Information Systems* 1990, **4**:297-312.
  29. Strehl A, Ghosh J: **Relationship-based clustering and visualization for multidimensional data mining**. *INFOMS Journal on Computing* 2002, **00**:1-23.
  30. Keim DA: **Information Visualization and Visual Data Mining**. *IEEE transactions on Visualization and Computer Graphics* 2002, **7**:100-107.
  31. Cabena P, Hadjnia P, Stadler R, Verhees J, Alessandro Z: **Discovering data mining: From concept to implementation**. New Jersey, Prentice Hall; 1998.
  32. MacEachren AM: **How maps work: representation, visualization, and design**. New York, The Guilford Press; 1995.
  33. Tobler W: **A Computer Movie Simulating Urban Growth in the Detroit Region**. *Economic Geography* 1970, **46**:234-240.
  34. Fabrikant SI, Skupin A: **Cognitively Plausible Information Visualization**. *Exploring GeoVisualization* Edited by: Dykes J, MacEachren A and Kraak MJ. Amsterdam, Elsevier; 2003.
  35. Fabrikant SI: **Visualizing region and scale in information spaces**. *Proceedings of the 20th International Cartographic Conference* Beijing, China; 2001.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

