A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP)

Diansheng Guo[Member, IEEE],
Department of Geography, University of South Carolina, 709 Bull Street, Rm. 127, Columbia, SC 29208

Jin Chen[Student Member, IEEE],
GeoVISTA Center, Department of Geography, Pennsylvania State University, 302 Walker Building, University Park, PA 16802

Alan M. MacEachren[Member, IEEE], and
GeoVISTA Center, Department of Geography, Pennsylvania State University, 302 Walker Building, University Park, PA 16802

Ke Liao
Department of Geography, University of South Carolina, 709 Bull Street, Rm. 127, Columbia, SC 29208

Diansheng Guo: guod@sc.edu; Jin Chen: jxc93@psu.edu; Alan M. MacEachren: maceachren@psu.edu; Ke Liao: liao4@mailbox.sc.edu

Abstract

The research reported here integrates computational, visual, and cartographic methods to develop a geovisual analytic approach for exploring and understanding spatio-temporal and multivariate patterns. The developed methodology and tools can help analysts investigate complex patterns across multivariate, spatial, and temporal dimensions via clustering, sorting, and visualization. Specifically, the approach involves a self-organizing map, a parallel coordinate plot, several forms of reorderable matrices (including several ordering methods), a geographic small multiple display, and a 2-dimensional cartographic color design method. The coupling among these methods leverages their independent strengths and facilitates a visual exploration of patterns that are difficult to discover otherwise. The visualization system we developed supports overview of complex patterns and, through a variety of interactions, enables users to focus on specific patterns and examine detailed views. We demonstrate the system with an application to the IEEE InfoVis 2005 Contest data set, which contains time-varying, geographically referenced, and multivariate data for technology companies in the US.

Index Terms

Information visualization; multivariate and spatio-temporal data; geovisualization; self-organizing map (SOM); visual analytics; ordering; small multiples

1 Introduction

Complex data sets that contain geographic locations, time series, and multiple variables have become a common but underutilized resource in many domains, from environmental science, through business, to homeland security. Such data hold great potential to provide valuable and previously unknown information that can advance our understanding of complex phenomena and systems [3], [12], [17]. However, visualization and data mining of spatio-temporal data are challenging problems. Existing data analysis approaches, including

both visual and analytical ones, have limited ability to explore complex patterns across all dimensions (i.e., geographic, temporal, and multivariate spaces).

The research reported here focuses on developing a geovisual analytic approach to explore multivariate spatiotemporal data, discover interesting and unknown complex patterns, and present them in an easy-to-understand form to support human interpretation, analytical reasoning, and/ or decision-making. Our approach leverages visual and computational methods to construct an overview of major patterns present in the data. Such an overview allows the analyst to perceive complex patterns across all dimensions and then guides user interactions to explore specific patterns in detail. We demonstrate an application of the approach (and resulting tools) to the analysis of changing characteristics of US industries.

The methods and tools we present are able to:

- perform multivariate clustering and abstraction (including time-series clustering) with a Self Organizing Map (SOM),

- encode the SOM result with colors derived from our ColorBrewerPlus component, which produces a two-dimensional diverging-diverging color scheme,

- visualize the multivariate patterns with an enhanced Parallel Coordinate Plot (PCP) display, which serves as a multivariate "legend" in the integrated system,

- visualize the spatio-temporal variations of multivariate patterns, or the space-variable variations of temporal patterns in a hierarchical, computationally sortable matrix and a temporally or geographically ordered map matrix, and

- support human interactions to explore patterns from different perspectives and at different detail levels.

The remainder of the paper is organized as follows: Section 2 gives a review of related research. We introduce our approach to the visualization of spatio-temporal and multivariate patterns in Section 3. Section 4 demonstrates the variety of interactions that our system supports. We then briefly introduce an interesting extension to visualize spatial interaction data (e.g., companies that relocated from one state to another) in Section 5. Last, we conclude with a discussion of the advantages and limitations of the approach.

2 Related Work

Our geovisual analytic approach directly addresses challenges delineated in the recent research agenda for visual analytics [49] and draws upon research in several related domains. Below, we briefly review four domains upon which we build most directly: multivariate visualization, multivariate and temporal mapping, visualization of large data sets, and computational ordering of multivariate data.

2.1 Multivariate Visualization

Multivariate visualization methods range from commonly used information graphics (e.g., tables, histograms, scatter plots, and charts [26]) through a suite of techniques introduced in the exploratory data analysis and information visualization literature, e.g., scatterplot matrices [1], matrix permutation [39], glyph [42], pixel-oriented approaches [32], and parallel coordinate plots (PCP) [27]. There is also research that combines traditional bar charts with pixel-based techniques to visualize large amounts of data with categorical and numerical types [31]. Due to display space limitations, multivariate data are often projected to a lower dimensional space using dimensional reduction techniques, e.g., multidimensional scaling [53], [55], principle component analysis (PCA), RadViz spring visualization [9], or other projection pursuit methods [13], [55]. It is impractical to provide a comprehensive

review of the range of multivariate visualization methods here, thus the reader is directed to a recent paper [34] that provides a categorization of both data types and visualization methods, with illustrations of most of the methods cited above and as well as others.

2.2 Multivariate and Temporal Mapping

Mapping is essential in visualizing geographic patterns. Multivariate mapping has long been a challenging and interesting research problem. Three primary approaches have been used: 1) multivariate representation that depicts each dimension (variable) independently through some attribute of the display and then integrates all variable depictions into one map using composite glyphs, attributes of color, or other methods [14], [19], [21], [54], [57]; 2) dimension reduction that projects multivariate information to two (or three) dimensions and then maps the result (e.g., [24]), and 3) multiple linked views that show one (or more) variables per view [2], [16], [38], [40].

Subsequent efforts have focused on integration across these approaches and/or their extension through coupling with statistical and/or computational methods, e.g., GGobi [47], Mondrian [48], Orca [46], and ExplorN [51]. Carr et al. [11] present a visualization approach for multivariate data analysis called conditioned choropleth maps (CCmaps) that uses a two-way layout of maps (and matching views of statistical distributions) designed to facilitate comparisons by showing the association between a dependent variable, as represented in a classed choropleth map, and two potential explanatory variables. Guo et al. [24] propose a visual-computational approach designed to detect and visualize multivariate spatial patterns by combining multivariate clustering, multivariate visualization, and a geographic map to present a holistic view of multivariate spatial patterns.

When time is a key attribute in the data analysis, strategies applied to understanding the time component include sequencing methods (either animated or interactive) and three-dimensional approaches where time is visualized as the third dimension over a two-dimensional map, e.g., [37]. However, both approaches have limited ability to visualize multivariate patterns across time. In this research, we combine multivariate abstractions and matrix views to visualize spatial-temporal trends of multivariate patterns or space-variable variations of temporal patterns.

2.3 Visualization of Large Data Sets

Large data sets can cause serious problems for most visualization techniques and these problems can be divided into two groups: the computational efficiency problem and the visual effectiveness problem. Computational efficiency concerns the time needed to process data and render views. A visualization technique has to be computationally efficient and scalable with very large data sets to allow human interactions (e.g., [41]). The visual effectiveness problem concerns the usefulness of data views in revealing patterns. With a large data set, data items can overlap in the visual display (e.g., points overlap in scatter plots or line segments overlap in a parallel coordinate plot—PCP) and make patterns very hard (if possible at all) to perceive [32].

Enhancements to address the visual effectiveness problem have been proposed along two primary directions. One direction is to resolve the overlap in the attribute space or geographic space by sampling, density mapping [29], or repositioning (or shifting) data points [35]. Another direction is to reduce the data size by performing data abstraction (e.g., clustering or aggregation) first so that the visualization component only needs to visualize a relatively small number of data clusters instead of all individual data items, while providing drill-down capabilities that support details for selected clusters [24], [25], [50]. We take the latter approach— data abstraction with drill down—in this research.

2.4 Ordering in Visualization and Data Mining

Ordering is widely used in visualization techniques to accentuate patterns. For example, in the visualization of bacterial genomes, pixel arrangement is used to place adjacent nucleotides as close to each other as possible and thus to help bring out data patterns that otherwise would be difficult to perceive [56]. Ordering is also used in arranging the layout of treemaps [44]. Friendly and Kwan presented a general framework for ordering information in visual displays (tables and graphs) according to the effects or trends. Their framework can be applied to the arrangement of unordered factors for quantitative data and frequency data, and to the arrangement of variables and observations in multivariate displays (e.g., star plots and parallel coordinate plots) [18].

The concept of a reorderable matrix [7], [8], [52], as a data table visualization method, has been the focus of several recent research efforts from different perspectives, e.g., testing ordering heuristics for an interactive tool [45], and visualizing time-varying data [43]. In this research, we develop a reorderable matrix and a reorderable map matrix to visualize a spatio-temporal and multivariate data cube.

3 VIS-STAMP: A Visual-Computational Approach

Our geovisual analytic approach and the implemented visualization system for space-time and multivariate patterns (VIS-STAMP) integrates a suite of visual, computational, and cartographic methods, which together are used to construct an overview of major patterns present in the data and support a variety of user interactions to assist the analyst in exploring and interpreting complex spatio-temporal patterns.

Specifically, we integrate:

1. a self-organizing map (SOM) [36] to perform multivariate clustering, sorting, and coloring,

2. a parallel coordinate plot (PCP) to visualize multivariate patterns and serve as a "legend" in the integrated system,

3. a reorderable matrix to organize multivariate patterns in the spatio-temporal space, and

4. a reorderable map matrix to reveal spatial variation of multivariate patterns.

We also use logically constructed, 2D color schemes to encode the SOM result and use two linear ordering methods to reorder the matrices to accentuate patterns.

Our system is designed to be flexible and support the use of different clustering methods (e.g., [22]). However, the SOM is a better fit here for two main reasons. First, an SOM arranges clusters into a 2D layout and, thus, makes it possible to use more colors (e.g., colors from a 2D color scheme instead of a 1D scheme) to represent clusters. Second, an SOM does not impose a hard partition on the data. Similar clusters are close in the SOM layout and colored similarly. Thus, humans can visually determine the number of clusters and the relationship between clusters. The user can also change the SOM size (i.e., the number of nodes) on the fly and compare the results.

Our approach inherently supports both overview and detail analysis. We present the overview and its related methodology in this section and introduce user interactions that lead to detailed views in Section 4.

3.1 Conceptual Data Representation

To simplify the presentation of our methodology, we conceptually represent the data as a data cube (Fig. 1a), which is defined by three components: the geography (e.g., US states, each of which maintains spatial information that can be shown with a map), the time (e.g., years), and a set of numerical variables. Each cell in this cube is defined by a specific spatial object (e.g., Texas), a specific time (e.g., year 2000), and a specific variable (e.g., sales percentage for the energy industry). The value for that cell is the variable value. Each spatial object (e.g., Texas) has a horizontal slice in the cube, which we call a *time-attribute slice* (Fig. 1b). A time-attribute slice can be seen as a series of *multivariate profiles* (one profile for each state/year combination—Fig. 1c), or a set of *time series* (one time series for each state/variable combination—Fig. 1d). For example, suppose we construct a data cube with 50 US states, 16 industries as variables, across 12 years. Then, there are 50 time-attribute slices (one for each state), $50 \times 12 = 600$ multivariate profiles, and $50 \times 16 = 800$ time series. From now on, we will directly refer to these three terms without explanation.

Such a space-time-attribute data cube is often an aggregation of a larger and more detailed data set. For example, the US company data set that we use for demonstration in this paper is from the IEEE InfoVis 2005 Contest [20] and has 563,000+ records. Each record contains the information for a specific company at a specific year, including its location (state name and zipcode), industry type, primary product type, sales, and employees. One possible aggregation of this data set into a data cube is to group data by state, year, and industry type. The value for each cell can be, for example, the sales value for that state/year/industry combination (e.g., California, at 2000, for computer hardware industry). The implementation of our approach allows the user to change the cube configuration interactively, e.g., using product types instead of industry types, or using the number of employees instead of sales values.

To help the reader understand our example analyses, here, we briefly introduce the company data that we use in this paper. We focus on 49 US states, including Washington DC but excluding Hawaii and Alaska for presentation clarity, since including those two states will make other states much smaller in maps. The data span across 12 years, from 1992 to 2003. We select 16 industry types: factory automation (AUT), biotechnology (BIO), chemicals (CHE), computer hardware (COM), defense (DEF), energy (ENR), environmental (ENV), manufacturing equipment (MAN), advanced materials (MAT), medical (MED), pharmaceuticals (PHA), computer software (SOF), subassemblies and components (SUB), telecommunications and internet (TEL), transportation (TRN), and not- primarily-high-tech (NON). This data set and its metadata are available at the IEEE InfoVis 2005 Contest Web site [20].

3.2 Multivariate Clustering and Visualization

3.2.1 Abstraction and Encoding of Multivariate Patterns—We use a self-organizing map (SOM) to cluster multivariate profiles (Fig. 1c), each of which is a multivariate vector for a specific state/year combination. More importantly, the SOM orders clusters (nodes) in a two-dimensional layout so that nearby clusters (nodes) are similar (in the multivariate space). Thus, the SOM effectively transforms the multivariate data into a two-dimensional space. We then use a systematically designed two-dimensional color scheme to assign a color to each SOM node so that nearby (and, therefore, similar) clusters have similar colors. Below, we briefly introduce this color-coded SOM. Readers are referred to [24] for details.

Our implementation of the SOM uses a traditional hexagonal layout and normally has 9×9 or fewer nodes (clusters) since it is difficult to construct a two-dimensional color scheme

with more than $9 \times 9 = 81$ colors and grouping data into more clusters is seldom a sufficient abstraction to be useful. SOM clusters are visualized using a U-Matrix [36] with several new added features (Fig. 2b). Each cluster is depicted with a circle, whose size (area) is linearly scaled and proportional to the number of data items it contains. Each hexagon is shaded (in shades of gray) to show the multivariate dissimilarity between immediate neighboring nodes, with darker colors showing greater dissimilarity. Thus, the shaded U-Matrix reveals the nonlinear mapping between the multivariate space and the regular 2D layout of nodes, which are not evenly distributed in the multivariate space.

Our two-dimensional diverging-diverging color scheme uses a systematic variation in both hue and lightness to provide a 2D array of logically ordered but discriminable colors [10] (see Fig. 2). This 2D color scheme differs from the color scheme proposed by Kaski et al [30] in that it has more color variations and makes the ordered sequence more clear. The size of the 2D color array is always the same as the size of the SOM. The two-dimensional array of colors is then folded onto the regular 2D layout of the SOM nodes (not onto the regression surface in the actual data space). Thus, each node corresponds to a unique color in the color array. As mentioned earlier, although the SOM nodes are ordered on a regular 2D layout, they are not evenly distributed in the multivariate data space and the distances (dissimilarities) between neighboring nodes may vary. Therefore, the color differences only represent the relative dissimilarity between two clusters of data items.

3.2.2 Visualization of Multivariate Patterns—The meaning of the colors in the SOM (Fig. 2), however, cannot be defined by a simple legend (as it might be on a typical geographic map), since each color represents a multivariate cluster. Thus, the colors, which signify the relative similarity of clusters, must be supplemented by a multivariate visualization method that allows analysts to understand the characteristics of each cluster and thus the meaning of each color. To accomplish this, we extend an earlier version of a parallel coordinate plot (PCP), introduced in [24], to visualize the data clusters identified by the SOM.

The earlier version of the PCP visualizes clusters instead of original data items, and thus partially avoids the overlap problem. Each string (representing a cluster with its mean vector) has the same color as it does in the SOM, which in turn dramatically improves the visual effectiveness of the PCP in presenting multivariate patterns. The PCP uses a nested-means scaling on each axis and, thus, further alleviates the overlapping problem. Nested-means is a nonlinear scaling method that recursively calculates a number of mean values (and submeans) and uses these values as break points to divide each axis into equal-length segments. Therefore, nested-means scaling always puts the mean value at the center of each axis and thus makes axes defined by different units and data ranges comparable (Fig. 3). The thickness of each string represents the cluster size (i.e., the number of data items contained in the cluster). It is also possible to use the string thickness to represent the data variance within each cluster (which is not implemented here).

We extend the earlier version of the PCP by adding features to support user interaction and information inquiry at different levels. Bertin defined three "levels of reading," i.e., the elementary level (allowing users to view the information about a single data element), intermediate level (revealing summary information about a group of elements), and global level (presenting an overall picture of all items in the data) [7]. As shown in Fig. 3, the colored PCP at the cluster level presents a global view of the overall patterns. A user can then select one or more clusters in the PCP (or the SOM), switch to the data item level (instead of the cluster level), and examine all the data items in the cluster(s) (Fig. 4). Selection can be made on either data item or cluster level. For example, one can show data at the item level in the PCP and then select a single data item to read its exact variable

values. One can also switch back to the cluster level and see to which cluster the selected item belongs. That cluster may contain many other items as well—thus, its circle will become a wedge to show the partial selection. We demonstrate these interaction features in Section 4.

In addition to the nested-means method, the new version of the PCP also supports several other linear scaling methods, including data min-max scaling-using the minimum and maximum data values to linearly scale each axis, cell min-max scaling-using the minimum and maximum cluster (node) mean values, and global min-max scaling-using the minimum and maximum for all variable values. The global min-max scaling is especially useful when the values on different axes are directly comparable, for example, percentage values as used in this research (Fig. 4).

3.3 Spatio-Temporal Visualization of Multivariate Patterns

The SOM and PCP together can visualize and present multivariate patterns effectively. However, multivariate patterns often vary over the geographic space and evolve over time. It is critical to visualize the data cube across all dimensions (i.e., space, time, and multiple variables) and construct a holistic view of the complex patterns present in the data cube. We develop a form of reorderable matrix to organize multivariate patterns (represented with colors) across time and space. This reorderable matrix, when one of the two dimensions represents geography, will be accompanied with a reorderable map matrix. Below, we first introduce the two matrices and then introduce the ordering methods used to reorder the matrices.

3.3.1 Reorderable Matrix and Map Matrix—The reorderable matrix we implemented supports computational sorting of both columns and rows. In the application shown in Fig. 5 (top-left), columns represent time (years) and rows represent places (states of the US). Ordering of time is fixed (for these applications) in normal temporal order. Ordering of places is computationally derived with several cluster-based ordering methods, which we present in the next subsection. Users can interactively choose any of the implemented sorting methods to reorder the rows. After the reordering, states that have similar industry patterns over time are next to each other in the matrix and, thus, form homogeneous spatio-temporal "regions." The reorderable map matrix we implemented essentially converts each column in the reorderable matrix to a map and these maps are arranged in the same order as that of the columns (Fig. 5, top-right). The advantage of a map matrix over a reorderable matrix is that the spatial topology is preserved and it can better support the perception of spatial distribution patterns. However, the disadvantage is that the temporal trend for a specific spatial object (e.g., California) is not as clear as in the reorderable matrix.

The data used in Fig. 5 includes six industry types, which are selected for demonstration—a more complete analysis with 16 industry types is included in Section 4. Each cell in the reorderable matrix or state in a map has a multivariate profile, which is a vector of sales values for the six industries for that state/year (see Fig. 1c). Then, each profile is converted to percentages, i.e., each value in the vector is divided by the vector total. The SOM takes all multivariate profiles as input, groups similar profiles into clusters, and assigns each cluster (and each multivariate profile) a color. Thus, similar colors represent similar multivariate profiles.

Therefore, both the reorderable matrix and the map matrix are actually "3D" views of the data cube, showing information across space (represented with the vertical dimension in the spatio-temporal matrix or with maps in the map matrix), time (represented with the horizontal dimension), and multiple variables (represented by colors).

In the middle of the reorderable matrix as depicted in Fig. 5, for example, we notice a "region" of purple colors and it contains Louisiana (LA), Oklahoma (OK), and Texas (TX), across all years (except Texas and Oklahoma at 1992). From the PCP, we understand that purple colors signify industry compositions dominated by the energy industry (ENR). Similarly, we also notice a dark red "region" (right above that purple "region") consisting of four states: Washington DC for 1992–1997, Arkansas (AR) for 1992–1999, New Mexico (NM) for 1994–2000, and Missouri (MS) for 1995–2003. From the PCP, we know that dark red or red colors represent a high percentage of the telecommunication and Internet industry (TEL). We also see that, around the late 1990s, three of those red states (except Missouri) changed to not-primarily-high-tech (NON) industry (represented with blue colors). Another overall pattern we can easily perceive is that many states shifted to the NON industry (represented by blue colors) since 2001. This pattern is evident in both the reorderable matrix and the map matrix.

Such an overview, prepared by the clustering, coloring, and ordering methods and presented with the four visual components, is a rich and yet clear representation of the major spatio-temporal and multivariate patterns present in the data cube. Even without user interactions, one can already perceive, interpret, and understand a variety of patterns by visually examining and linking those four views. Thus, it can allow the presentation and communication of complex patterns in static forms, e.g., images or printed papers, when interactive presentation is not possible.

3.3.2 Hierarchical Clustering and Matrix Ordering—In this section, we introduce the ordering methods that we developed to order the matrix rows (or columns). Let $A = \{a_1\ a_2, \ldots, a_n\}$ be a set of objects (either all the rows or all the columns in the matrix). All pair-wise dissimilarity values within A form a symmetric matrix (hereafter, dissimilarity matrix). In this paper, we define the dissimilarity between two spatial objects as the Euclidean distance between their time-attribute slices (see Fig. 1b), each of which is a 2D array of numerical values. Our system provides the flexibility to incorporate other similarity measures since similarity definitions are often application dependent. To render the display of the reorderable matrix shown in Fig. 5, all rows (i.e., states) are ordered according to their dissimilarity matrix.

There are several existing methods for sorting a matrix based on dissimilarity or other measures [4], [6], [18], [22]. An ordering can be derived based on a hierarchical clustering result. However, as seen in Fig. 6, a cluster hierarchy cannot determine a unique ordering. There are 2^{n-1} (n is the number of objects to be ordered) different orderings that are consistent with the same cluster hierarchy. Bar-Joseph et al. [5], [6] propose a method to find the shortest ordering from a given cluster hierarchy and the ordering method is of $O(n^3)$ complexity (in addition to the computational cost of the hierarchical clustering method adopted to derive the cluster hierarchy).

We propose and implement a simple, fast, and yet satisfying ordering strategy based on a hierarchical clustering result. Given a dendrogram (Fig. 6, middle), we process the hierarchy from the bottom up. At the beginning, each cluster contains a single item (e.g., a state). When two clusters are merged into one, the closest (i.e., most similar) ends of the two clusters should be connected. For example, when B is merged with cluster $\{C, D\}$, B should be next to D because it is closer to D than to C (Fig. 6, left). When cluster $\{A, E\}$ is merged with cluster $\{B, D, C\}$, C and E should be next to each other since CE is the closest among the four connection options: AB, AC, BE, and CE. Once all data items are in the same cluster, an ordering is achieved (Fig. 6, right).

Theoretically, any hierarchical clustering method can be combined with the above ordering strategy to derive a unique ordering. We implemented two ordering methods, one based on the single-linkage clustering and the other based on the complete-linkage clustering [28]. Fig. 7 shows three reorderable matrix views of the same data (as used in Fig. 5). Generally, the complete-linkage ordering produces a better result [23]. The matrix also includes a programming interface that supports connection with other sorting methods.

An interesting advantage here is that we can crosscheck the ordering result and the SOM result (i.e., colors) since they are constructed independently. From the matrices shown in Fig. 7, we can see that the ordering and the SOM result match very well as rows with similar colors (which is the SOM result) are also ordered next to each other (which is the result of the ordering).

3.4 Space-Variable Visualization of Temporal Patterns

Our system has the flexibility to present patterns in the space-time-multivariate data cube from different perspectives. For example, we can use industry types and states (instead of years and states) to organize the reorderable matrix (now, we call it space-variable matrix). In other words, each column represents an industry type and each row represents a state. Temporal series (see Fig. 1d) will be treated as multivariate vectors and clustered (and colored) by the SOM. To characterize a temporal trend, we convert each time series (one for each state/industry combination) to percentages, e.g., the percentage of one year's sales against the total sales for all years for that specific state and industry.

Now, we are able to examine the variation of temporal patterns across geography and multiple categories (e.g., industry types) (Fig. 8). The colors represent similar temporal trends. From the PCP (as the legend), we can tell that green colors represent trends that peaked in the early 1990's, but declined since then. Blue colors represent a combination of rising trends in earlier years and declining trends for recent years. Purple colors represent a rise during 1998–2001 and a decrease in 2002 and 2003. Red and dark red colors represent a very recent growth in sales—with low sales for most years but rapid rises in 2003.

Both the columns and the rows of the space-variable matrix are ordered, separately, with the complete-linkage ordering. Patterns in the space-variable matrix (Fig. 8, top-left) are not as clear as we saw in Fig. 5. This indicates that it is rare for two states to have similar temporal trends for each industry type, and that it is also rare for two industries to develop in the same way in all states. However, we do see that after the ordering red columns are shifted to the right side in the reorderable matrix and accordingly "hot" maps are shifted to the lower part in the map matrix. That gives us a clear perception of the industries that had rising sales in recent years and the states that those rapid increases occurred. For example, four industries had a significant growth recently almost nationwide, including the telecommunication and Internet (TEL), biotechnology (BIO), energy (ENR), and the nonprimary high-tech (NON) industry. The reorderable map matrix shows us the spatial distribution and regional differences in the growth of each industry. For example, the material industry (MAT-the third map on the top row in the map matrix) had a recent growth in the northwest region, while rising sales of the manufacturing industry (MAN—the second map on the last row)— focused on the Midwest and East coast.

4 Human Interactions with VIS-STAMP

In addition to constructing a holistic view of patterns in the spatio-temporal and multivariate data cube, VIS-STAMP also supports a variety of human interactions that allow the analyst to examine patterns in detail. We specifically design and implement the system to support three main interaction features. First, each visual component should be able to support user

selections and the selection made in one component should be highlighted in all other components simultaneously. Second, the user can make a selection in one component and then refine that selection in the same or another visual component by adding or subtracting new selection(s). Third, each component should be able to respond to selections made at different levels (i.e., data elements or clusters). Three examples are included to demonstrate interactions at three different levels, which correspond to Bertin's three reading levels, i.e., the elementary level, intermediate level, and the global level [7].

4.1 Overview of Patterns

Fig. 9 shows an overview of patterns in the US company data. There are two major differences between this analysis and the one presented in Fig. 5. First, this analysis includes 16 industry types instead of six. Please refer to Section 3.1 for a list of these industries. Second, the reorderable matrix is organized into five geographic regions, i.e., the Pacific (Pac), Southwest (SW), Midwest (MID), Northeast (NE), and the Southeast (SE). The matrix supports such concept hierarchies for both the columns and rows (Section 5 has another example). States (rows) in each region are then ordered with the complete-linkage ordering.

From this overview, we can perceive a variety of patterns across the multivariate space, time, and geography. Those patterns identified in Fig. 5 are still evident here, although with different colors. Wealso see many new patterns, as more industries are included in this overview. For example, from the PCP, we understand that the white color represents high percentages (> 30 percent) for Advanced Materials (MAT) and from the reorderable matrix (and/or the map matrix) we can see that the MAT industry was dominant in West Virginia (WV) since 1993, Utah (UT) for 1999–2002, New Hampshire (NH) and Tennessee (TN) for 1992.

4.2 Interaction at the Elementary Level

In addition to the above overview, user interactions are particularly useful when we want to understand each pattern precisely. For example, we can select Nevada (NV) to understand how its industry composition changed over the 12 years (Fig. 10). The selection is made in the reorderable matrix by dragging the mouse across those cells. The selected cells are highlighted by shrinking other cells to a quarter of their original sizes. The PCP shows data at the data item level (instead of the cluster level). With colors as identifiers, we can easily link the same data element across different views.

Therefore, we can perceive precisely that, for 1992–1994 Nevada's industry sales were primarily from the energy industry (ENR) (> 40 percent). Then, for 1995–2000, energy percentages dropped to 25 percent while computer hardware (COM) increased to about 40–50 percent. For this same period, Nevada also had a moderate growth for subassemblies and components (SUB). Since 2001, however, Nevada's industry sales were dominated by not-primarily-high-tech (NON) type (> 60 percent).

4.3 Interaction at the Intermediate Level

We can also examine a group of data elements or compare groups of data elements. For example, to focus on those states (and years) that had a high percentage of sales from the transportation industry (TRN), we make a selection in the PCP to include all strings with high values (> 35 percent) on TRN. Five states meet this criterion: Washington (WA) for 1996–2002, New Mexico (NM) for year 1993, Rhode Island (RI) for 1992–2003, Kansas (KS) for 1992–1998, and Missouri (MS) for 1993–1994 (Fig. 11).

We notice that only Rhode Island kept that composition for all the years, while other four states all changed eventually before 2003. We then want to understand how (and to what other industries) these states had shifted. Therefore, we add several selections from the reorderable matrix to extend the four states (except Rhode Island) by five years or up to year 2003 (Fig. 11). Clearly, both Kansas and Missouri changed to telecommunication and Internet (TEL) when their transportation industry diminished. Washington also changed in 2003 to a high share of Internet business (about 30 percent). New Mexico only had one year (1993) dominated by transportation sales and since then has changed to a combination of other industries (including TEL). The greater variation observed for New Mexico is probably due to its relative small economy size and, thus, even a small change in one industry may cause the overall composition shifted.

Please notice how the SOM view and map matrix respond to this selection. The SOM view primarily focuses on clustersinstead of data elements. The selected data items belong to six different nodes in the SOM, among which two are fully selected and four are partially selected and shown with wedges, scaled to show the selected proportion of each cluster.

4.4 Interaction at the Global Level

Here, we demonstrate interactions at the cluster level, more towards a global view. We continue with the overview presented in Fig. 8, in which temporal series are clustered with the SOM, matrices are organized with states and industry types, and the PCP shows temporal trends. Both the rows and columns in the reorderable matrix are ordered with the complete-linkage ordering. Maps in the matrix are in the same order as that of columns in the reorderable matrix.

To examine those rising trends in recent years (from 1998 to 2003), we select all the "hot" clusters on the right side in the SOM view (Fig. 12). These clusters represent temporal trends that were low in early years but increased rapidly in recent years, as interpreted from the PCP. Specifically, purple colors represent fast growth from 1998 to 2000, while red and dark brown colors represent a rapid growth since 2001.

In Fig. 12, we can easily perceive from the reorderable matrix and the map matrix that NON (not primarily high-tech) and TEL (telecommunications and Internet) were the fastest-growing industries nationwide in recent years. Moreover, we can also tell that the TEL industry had its growth mainly during 1999–2001, because there are more purple colors than red/brown colors in its column or its map. On the other hand, the biggest growth for the NON industry started after 2001, as its colors are primarily red. In addition to the NON and TEL industries, Energy (ENR) and Biotechnology (BIO) also witnessed a recent growth in many states (but not as widespread as NON and TEL). There are many other patterns evident in the snapshot shown in Fig. 12, but we are not able to enumerate all of them here due to space limitation.

5 VIS-STAMP Extended: Visualization of Spatial Interactions

The reorderable matrix and the map matrix introduced earlier are designed to visualize the space-time-attribute data cube, which describes multivariate characteristics for each state and each year. To address the unique challenges in analyzing and visualizing spatial interaction information (e.g., companies that relocated from an origin state to a destination state), we extend both the reorderable matrix and the map matrix to construct two novel variants. We singled out all the companies that have relocated once or more from the company data we used earlier. Each record in this new data set has the year, origin state, destination state, sales, and the number of employees for each relocated company. If one company moved more than once, each move will be a record in the new data set.

5.1 Spatial Interaction Matrix

The reorderable matrix introduced earlier is directly applicable for this new data set by having both columns and rows representing geography. Specifically, the matrix now has its rows representing origin states (where companied moved out) and columns representing destination states (where companies moved into). Each cell in this matrix represents the number of companies that moved from the row state to the column state (Fig. 13). Therefore, this matrix is asymmetric. Both columns and rows are first organized into five geographic regions (i.e., Pacific-Pac, Southwest-SW, Midwest-MID, Northeast-NE, and South-east-SE). Then, columns and rows are ordered, separately, using the single-linkage method within each region. The similarity between two states is defined as the total number of companies relocated between them. We can see, for example, that many companies moved from the Northeast to the Southeast, but many fewer from the Southeast to the Northeast.

5.2 Map^2 Matrix-Maps within a Schematic Map

To extend the map matrix introduced earlier to visualize spatial interactions, we develop a new form of map matrix-Map^2, which is essentially a "map" of maps (Fig. 14). The overall view is a schematic "map" that contains multiple component (small) maps. Each component map in the matrix represents all the companies that moved from all other states into a specific state, which is labeled above that component map and highlighted in yellow. For example, the top-left map shows companies moving into Washington (WA) from each other state, with the darkest shade of green representing number of companies. These individual maps are ordered into an abstract map layout in which the location of the component map in the matrix is similar to the actual geographic location of that state (e.g., WA at the northwest corner, Florida in the Southeast). Thus, this layout could be considered as a form of discontiguous cartogram [15], [33].

When we view the map matrix as a single (abstract) map, we can see regions with almost no influx of companies (e.g., the upper Great Plains) and others with a relatively large company influx (e.g., NY, NJ, PA, and MA). When looking at each component map (or maplet), we can examine the attraction area of each state.

Instead of showing the raw counts of companies moving between states, the matrix can also show any other set of values, e.g., the in/out difference for each pair of states or the percentage of relocated companies against the total number of companies in those two states. The data preprocessing step determines which set of values should be shown in the matrices.

6 Conclusion and Discussions

We introduced a geovisual analytic approach, which is a synergy of computational and visual methods. The approach makes it possible to derive complex patterns and gain insights from spatio-temporal and multivariate data sets. The implemented visualization system has at least two important advantages: 1) its effectiveness in detecting and visualizing geographic, temporal, and multivariate patterns in multiple ways (thus, it is not constrained by one perspective and is able to identify complex relationships across multiple spaces); and 2) its component-based design that provides flexibility in addressing a range of analysis questions or a variety of different data sets by allowing easy connection to other visual and computational methods.

One limitation of our approach is that small-sized spatial objects (e.g., states with a small area) are barely visible in maps, especially when the map matrix has many maps and makes each individual map very small. One solution to this problem is to use a cartogram approach

[15], [33]. Another limitation of our tools, in this initial analysis of the US company data set, is that we aggregated company statistics to state-level, thus are likely to miss patterns that span state boundaries as well as patterns that are geographically more localized. However, the variety of interesting patterns found at this coarse geographic resolution demonstrates the remarkable potential of the approach. Building on this start, we will extend our analysis environment to explore patterns at detailed geographic scales, e.g., county-level and point-level analysis.

By leveraging the strengths of both computational methods (e.g., clustering and ordering) and visual methods, our approach is able to address relatively much larger data sets than would be possible with visual methods alone. However, it remains a challenging problem when it comes to extremely large data sets (e.g., several gigabytes of data). In this case, our approach relies on data aggregation in the preprocessing stage to reduce the data size to a range that the system can handle. On the other hand, our framework can also replace current components with any other efficient methods if available and needed. Further usability studies are needed to empirically validate our approach and learn how actual users interact with the system.

Acknowledgments

References

1. Andrews DF. Plots of High-Dimensional Data. Biometrics. 1972; vol. 29:125–136.

2. Andrienko G, Andrienko N. Constructing Parallel Coordinates Plot for Problem Solving. Proc. First Int'l Symp. Smart Graphics. 2001:9–14.

3. Andrienko N, Andrienko G, Gatalsky P. Exploratory Spatio-Temporal Visualization: An Analytical Review. J. Visual Languages & Computing. 2003; vol. 14(no. 6):503–541.

4. Ankerst M, Berchtold S, Keim DA. Similarity Clustering of Dimensions for an Enhanced Visualization of Multidimensional Data. Proc. Conf. Information Visualization'98. 1998:52–60.

5. Bar-Joseph Z, Demaine ED, Gifford DK, Hamel AM, Jaakkola TS, Srebro N. K-Ary Clustering with Optimal Leaf Ordering for Gene Expression Data. Bioinformatics. 2003; vol. 19(no. 9):1070–1078. [PubMed: 12801867]

6. Bar-Joseph Z, Gifford DK, Jaakkola TS. Fast Optimal Leaf Ordering for Hierarchical Clustering. Bioinformatics. 2001; vol. 17 supplement 1:22–29.

7. Bertin, J. Semiology of Graphics. Diagrams, Networks, Maps. Madison, Wis: The Univ. of Wiscons; 1983. in Press

8. Bertin J. Matrix Theory of Graphics. Information Design J. 2001; vol. 10:5–19.

9. Bertini ED, Aquila L, Santucci G. SpringView: Cooperation of RadViz and Parallel Coordinates for View Optimization and Clutter Reduction. Proc. Third Int'l Conf. Coordinated and Multiple Views in Exploratory Visualization (CMV '05). 2005:22–29.

10. Brewer, CA. Color Use Guidelines for Mapping and Visualization. In: MacEachren, AM.; Taylor, DRF., editors. Visualization in Modern Cartography. Tarrytown, NY: Elsevier Science; 1994. p. 123-147.

11. Carr DB, White D, MacEachren AM. Conditioned Choropleth Maps and Hypothesis Generation. Annals of the Assoc. of Am. Geographers. 2005; vol. 95(no. 1):32–53.

12. Cook D. Visual Data Mining of Large, Multivariate Space-Time Data. Am. Geophysical Union, Fall Meeting 2001, abstract #NG41A-01. 2001:A1+.

13. Cook D, Buja A, Cabrera J, Hurley C. Grand Tour and Projection Pursuit. J. Computational and Graphical Statistis. 1995; vol. 4(no. 3):155–172.

14. Dibiase, D.; Reeves, C.; Krygier, J.; MacEachren, AM.; Weiss, MV.; Sloan, J.; Detweiller, M. Multivariate Display of Geographic Data: Applications in Earth System Science. In: MacEachren, AM.; Taylor, DRF., editors. Visualization in Modern Cartography. 1994. p. 287-312.

15. Dorling, D. Dorling, D.; Hearnshaw, H. Visualization and GIS. London: Belhaven Press; 1994. Cartograms for Visualizing Human Geography; p. 85-102.

16. Dykes J. Cartographic Visualization: Exploratory Spatial Data Analysis with Local Indicators of Spatial Association Using Tcl/ Tk and CDV'. The Statistician. 1998; vol. 47(no. 3):485–497.

17. Dykes JA, Mountain DM. Seeking Structure in Records of Spatio-Temporal Behavior: Visualization Issues, Efforts and Applications. Computational Statistics & Data Analysis. 2003; vol. 43(no. 4):581–603.

18. Friendly M, Kwan E. Effect Ordering for Data Displays. Computational Statistics & Data Analysis. 2003; vol. 43(no. 4):509–539.

19. Gahegan M. Scatterplots and Scenes: Visualization Techniques for Exploratory Spatial Analysis. Computers, Environment, and Urban Systems. 1998; vol. 22(no. 1):43–56.

20. Grinstein, G.; Cvek, U.; Derthick, M.; Trutschl, M. Proc. IEEE InfoVis 2005 Contest, Technology Data in the US. 2005. http:/ivpr.cs.uml.edu/infovis05

21. Grinstein G, Sieg JCJ, Smith S, Williams MG. Visualization for Knowledge Discovery. Int'l J. Intelligent Systems. 1992; vol. 7:637–648.

22. Guo D. Coordinating Computational and Visualization Approaches for Interactive Feature Selection and Multivariate Clustering. Information Visualization. 2003; vol. 2(no. 4):232–246.

23. Guo D, Gahegan M. Spatial Ordering and Encoding for Geographic Data Mining and Visualization. J. Intelligent Information Systems. in press.

24. Guo D, Gahegan M, MacEachren AM, Zhou B. Multivariate Analysis and Geovisualization with an Integrated Geographic Knowledge Discovery Approach. Cartography and Geographic Information Science. 2005; vol. 32(no. 2):113–132. [PubMed: 19960118]

25. Guo D, Peuquet D, Gahegan M. ICEAGE: Interactive Clustering and Exploration of Large and High-Dimensional Geodata. GeoInformatica. 2003; vol. 7(no. 3):229–253.

26. Harris, RL. Information Graphics: A Comprehensive Illustrated Reference. Oxford, UK: Oxford Press; 1999.

27. Inselberg A. The Plane with Parallel Coordinates. The Visual Computer. 1985; vol. 1:69–97.

28. Jain, AK.; Dubes, RC. Algorithms for Clustering Data. Engle-wood Cliffs, N.J: Prentice Hall; 1988. p. 320

29. Johansson J, Ljung P, Jern M, Cooper M. Revealing Structure within Clustered Parallel Coordinates Displays. Proc. IEEE Symp. Information Visualization. 2005:125–132.

30. Kaski S, Venna J, Kohonen T. Coloring That Reveals Cluster Structures in Multivariate Data. Australian J. Intelligent Information Processing Systems. 2000; vol. 6:82–89.

31. Keim DA, Hao MC, Dayal U. Hierarchical Pixel Bar Charts. IEEE Trans. Visualization and Computer Graphics. 2002; vol. 8(no. 3):255–269.

32. Keim DA, Kriegel HP. Visualization Techniques for Mining Large Databases: A Comparison. IEEE Trans. Knowledge and Data Eng. 1996; vol. 8(no. 6)

33. Keim DA, North SC, Panse C, Schneidewind J. Efficient Cartogram Generation: A Comparison. Proc. IEEE Symp. Information Visualization. 2002:33–36.

34. Keim, DA.; Panse, C.; Sips, M. Information Visualization: Scope, Techniques and Opportunities for Geovisualization. In: Kraak, M-J., editor. Exploring Geovisualization. Amsterdam: Elsevier; 2005. p. 23-52.

35. Keim DA, Panse C, Sips M, North SC. Visual Data Mining in Large Geospatial Point Sets. IEEE Computer Graphics and Applications. 2004; vol. 24(no. 5):36–44. [PubMed: 15628099]

36. Kohonen, T. Self-Organizing Maps. third ed. Springer series in information sciences; 2001. p. 501

37. Kwan MP. Interactive Geovisualization of Activity-Travel Patterns Using Three-Dimensional Geographical Information Systems: A Methodological Exploration with a Large Data Set. Transportation Research Part C-Emerging Technologies. 2000; vol. 8:185–203.

38. MacEachren AM, Wachowicz M, Edsall R, Haug D, Masters R. Constructing Knowledge from Multivariate Spatiotemporal Data: Integrating Geographical Visualization with Knowledge Discovery in Database Methods. Int'l J. Geographical Information Science. 1999; vol. 13(no. 4): 311–334.

39. Mäkinen, E.; Siirtola, H. Theory and Application of Diagrams, Diagrams 2000, Lecture Notes in Artificial Intelligence 1889. Edinburgh, Scotland: Springer-Verlag; 2000. Reordering the Reorderable Matrix as an Algorithmic Problem; p. 453-467.

40. Monmonier M. Geographic Brushing: Enhancing Exploratory Analysis of the Scatterplot Matrix. Geographical Analysis. 1989; vol. 21(no. 1):81–84.

41. Park S, Bajaj C, Ihm I. Visualization of Very Large Oceanography Time-Varying Volume Data Sets. Proc. Int'l Conf. Conceptual Structures (ICCS '04). 2004:419–426.

42. Pickett, RM.; Grinstein, G.; Levkowitz, H.; Smith, S. Harnessing Preattentive Perceptual Processes in Visualization. In: Grinstein, G.; Levkowitz, H., editors. Perceptual Issues in Visualization. Springer; 1995. p. 33-45.

43. Qeli E, Wiechert W, Freisleben B. Visualizing Time-Varying Matrices Using Multidimensional Scaling and Reorderable Matrices. Proc. Eighth Int'l Conf. Information Visualization. 2004:561–567.

44. Shneiderman B, Wattenberg M. Ordered TreeMap Layouts. Proc. IEEE Symp. Information Visualization 2001 (INFOVIS). 2001

45. Siirtola H, Makinen E. Constructing and Reconstructing the Reorderable Matrix. Information Visualization. 2005; vol. 4:32–48.

46. Sutherland P, Rossini A, Lumley T, Lewin-Koh N, Dickerson J, Cox Z, Cook D. Orca: A Visualization Toolkit for High-Dimensional Data. J. Computational and Graphical Statistics. 2000; vol. 9(no. 3):509–529.

47. Swayne DF, Lang DT, Buja A, Cook D. GGobi: Evolving from Xgobi into an Extensible Framework for Interactive Data Visualization. Computational Statistics and Data Analysis. 2003; vol. 43(no. 4):423–444.

48. Theus, M. Statistical Data Exploration and Geographical Information Visualization. In: Dykes, J.; MacEachren, AM.; Kraak, M-J., editors. Exploring Geovisualization. Amsterdam: Elsevier; 2005. p. 127-142.

49. Thomas, JJ.; Cook, KA., editors. Illuminating the Path: The Research and Development Agenda for Visual Analytics. CS Press; 2005.

50. Ward MO. Finding Needles in Large-Scale Multivariate Data Haystacks. Computer Graphics and Applications. 2004; vol. 24(no. 5):16–19. [PubMed: 15628095]

51. Wilhelm A, Symanzik J, Wegman E. Visual Clustering and Classification: The Oronsay Particle Size Data Set Revisited. Computational Statistics. 1999; vol. 14(no. 1):109–146.

52. Wilkinson L. Permuting a Matrix to a Simple Pattern. Proc. Statistical and Computing Section of the Am. Statistical Assoc. 1979:409–412.

53. Williams M, Munzner T. Steerable, Progressive Multidimensional Scaling. Proc. IEEE Symp. Information Visualization. 2004:57–64.

54. Wittenbrink CM, Saxon E, Furman JJ, Pang A, Lodha S. Glyphs for Visualizing Uncertainty in Environmental Vector Fields. IEEE Trans. Visualization and Computer Graphics. 1995; vol. 2(no. 3):266–279.

55. Wong PC, Bergeron RD. Multivariate Visualization Using Metric Scaling. Proc. Eighth IEEE Visualization Conf. 1997:111–118.

56. Wong PC, Wong KK, Foote H, Thomas J. Global Visualization and Alignments of Whole Bacterial Genomes. IEEE Trans. Visualization and Computer Graphics. 2003; vol. 9(no. 3):361–377.

57. Zhang X, Pazner M. The Icon ImageMap Technique for Multivariate Geospatial Data Visualization: Approach and Software System. Cartography and Geographic Information Science. 2004; vol. 31(no. 1):29–41.

Biographies

Diansheng Guo received the BS (1996) degree in geography from the Peking University, the MS (1999) degree in GIS and cartography from the Chinese Academy of Sciences (CAS), and the PhD (2003) degree in geography from the Pennsylvania State University. He is an assistant professor in the Department of Geography, University of South Carolina. Dr. Guo served on the program committee of the Eighth and Ninth International Conferences on Information Visualization (IV04 & IV05). He has authored a number of papers in journals (including *Information Visualization, GeoInformatica*, and *Cartography and Geographic Information Science*) and conferences (including ACM GIS, IEEE InfoVis, and GIScience). His research interests include spatial data mining, spatio-temporal and high-dimensional visualization, and information theoretical approaches in data analysis. Dr. Guo is a member of the IEEE and the IEEE Computer Society.

Jin Chen received the BS (1995) degree from Beijing Institute of Technology and the MS (2000) degree from the University of Toledo, all in engineering. He is a PhD student in the Department of Geography and the GeoVISTA Center, Pennsylvania State University. He worked as an information system engineer at Chrysler Jeep Co, Ltd. (1995–1998) and at (China) Nokia Telecommunication Co, Ltd (1998–1999). He joined the GeoVISTA Center in 2002 as a research staff member. His research interests include information visualization, geovisualization, and OpenGIS development. His current research is supported by the National Cancer Institute and the Advanced Research and Development Activity (ARDA). He is a student member of the IEEE and the IEEE Computer Society.

Alan M. MacEachren received the BS (1974) degree from Ohio University and the MS (1976) and PhD (1979) degrees from the University of Kansas in 1979, all in geography. He is a professor of geography and director of the Geo VISTA Center (www.GeoVISTA.psu.edu) at Pennsylvania State University (Penn State). He held faculty positions at Virginia Tech and the University of Colorado before joining Penn State in 1985 (where he was awarded professor rank in 1992 and named the E. Willard and Ruby S. Miller Professor of Geography in 2004. Dr. MacEachren served as chair of the International Cartographic Association Commission on Visualization and Virtual Environments (1999–2005) and was named honorary fellow of that organization in 2005. He was also a member of the National Research Council Computer Science and Telecommunications Board Committee on the Intersections between Geospatial Information and Information

Technology, and an associate editor of *Information Visualization* and of the National Visualization and Analytics Center R&D Agenda panel. His research interests include geovisualization, geocollaboration, interfaces to geospatial information technologies, human spatial cognition as it relates to use of those technologies, human-centered systems, and user-centered design. His current research is supported by the US National Science Foundation, the National Institutes of Health, Centers for Disease Control, the Disruptive Technologies Office, and the US Air Force. Dr. MacEachren is author of *How Maps Work: Representation, Visualization, and Design* (Guilford Press, 1995) and *Some Truth with Maps* (Association of American Geographers, 1994), and is coeditor of several additional books (including *Exploring Geovisualization* (Elsevier, 2005)) and journal special issues (including *Research Challenges in Geovisualization*, a special issue of cartography and geographic information science, January 2001, vol. 28, no. 1 and a forthcoming issue of *IEEE Computer Graphics and Applications* theme issue on geovisualization). He is a member of the IEEE and the IEEE Computer Society.

Ke Liao received the BS (1998) degree from the Lanzhou University, China, and the MS (2002) degree in geography from the East China Normal University. She also holds an MS (2004) degree in geography from Northern Illinois University. She is currently a PhD student in the Department of Geography, University of South Carolina. Her research interests include geographic visualization, spatial data mining, and exploratory spatial analysis. She is a student member of the Association of American Geographers.
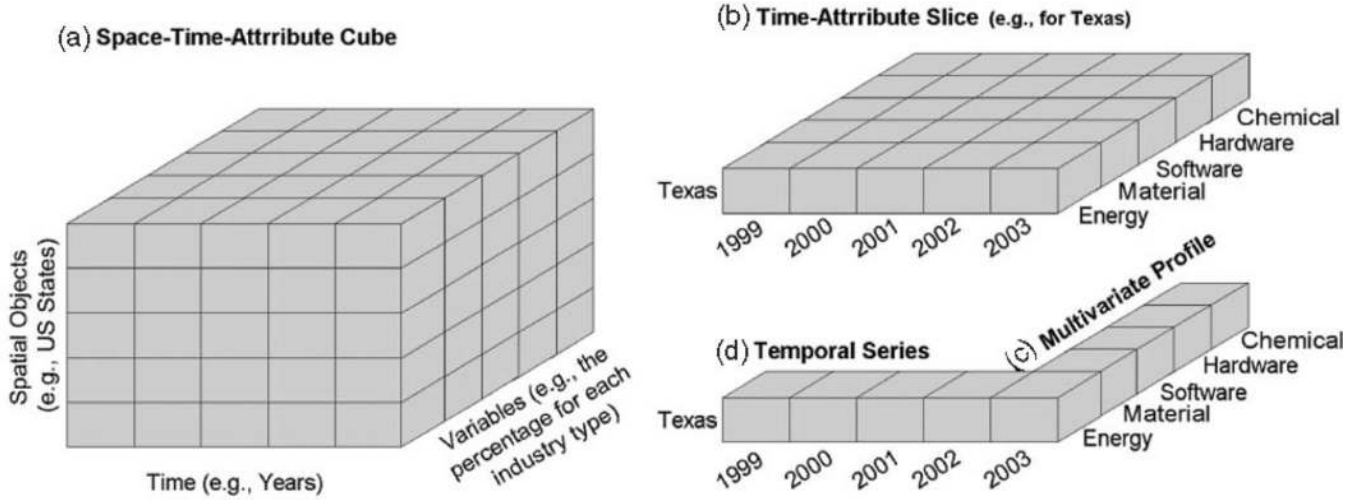
Fig. 1.
(a) The spatio-temporal and multivariate data cube, which can be decomposed into (b) a set of time-attribute slices (one for each state), or (c) a set of multivariate profiles (one for each state/year combination), or (d) a set of time series (one for each state/variable combination).
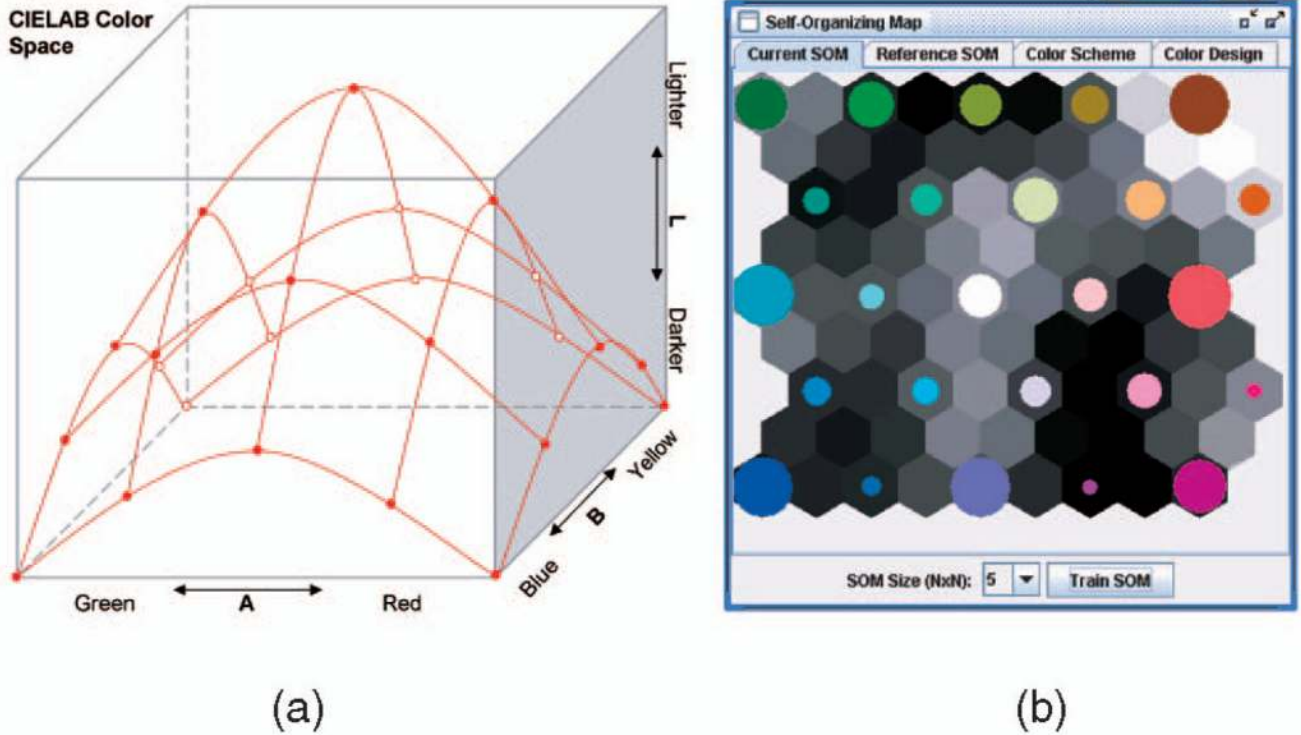
Fig. 2.
(a) The two-dimensional color model and (b) the color-encoded SOM. The 2D array of colors are derived from the 2D color model, which horizontally rotates the bell-shaped mesh 25 degrees clockwise and then samples a color at each knot on the mesh. See [24] for details on the color design method and interface. Each circle in the SOM represents a nonempty node and the area is linearly scaled and proportional to the number of data items contained in that cluster.
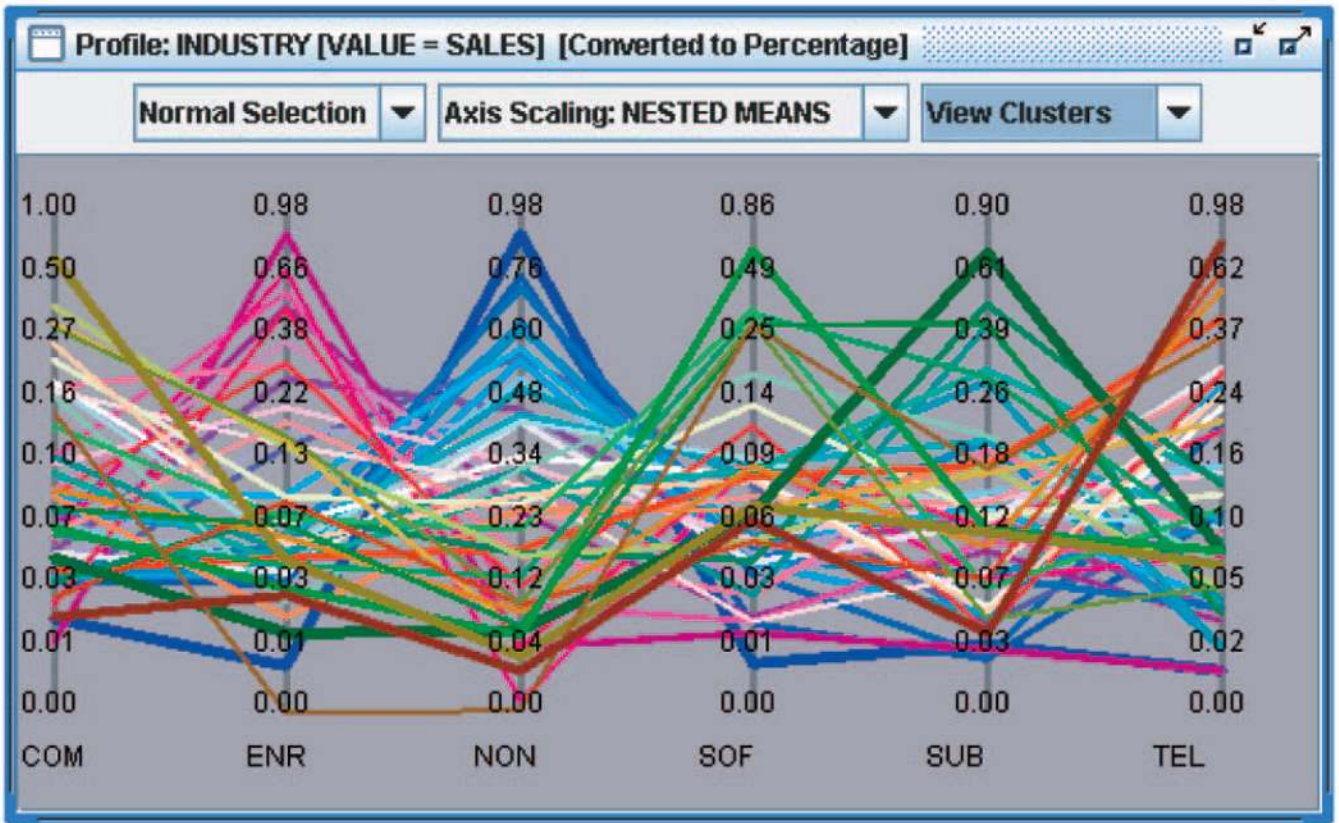
Fig. 3.
Each string in the PCP is a cluster of multivariate profiles (e.g., industry compositions). Each string is positioned using the cluster mean vector and colored by the SOM (see Fig. 5 (bottom-right)). Each axis is scaled using the nested-means method, which always puts the mean value at the center of each axis. Each variable is the sales value for an industry for a state/year as a percentage of the total sales of all six selected industries for that state/year. The thickness of each string represents the number of data items contained in that cluster.
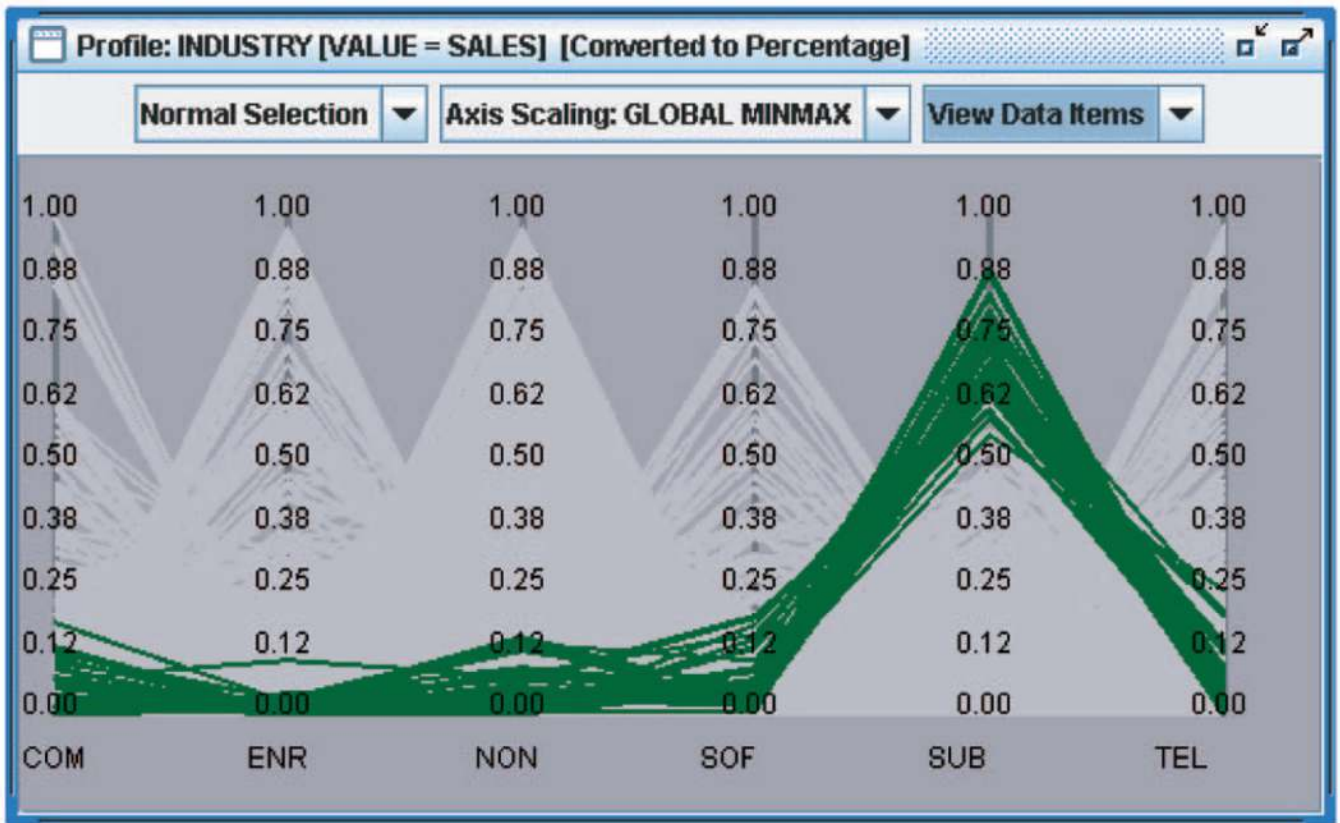
Fig. 4.
The dark green cluster that has the highest percentage for SUB industry in the PCP above (Fig. 3) is selected and shown at the data item level. We now see all the data items contained in that cluster and notice details such as the one for which contributions from the energy industry are higher than for others in this cluster. The scaling of each axis is also changed to global min-max.
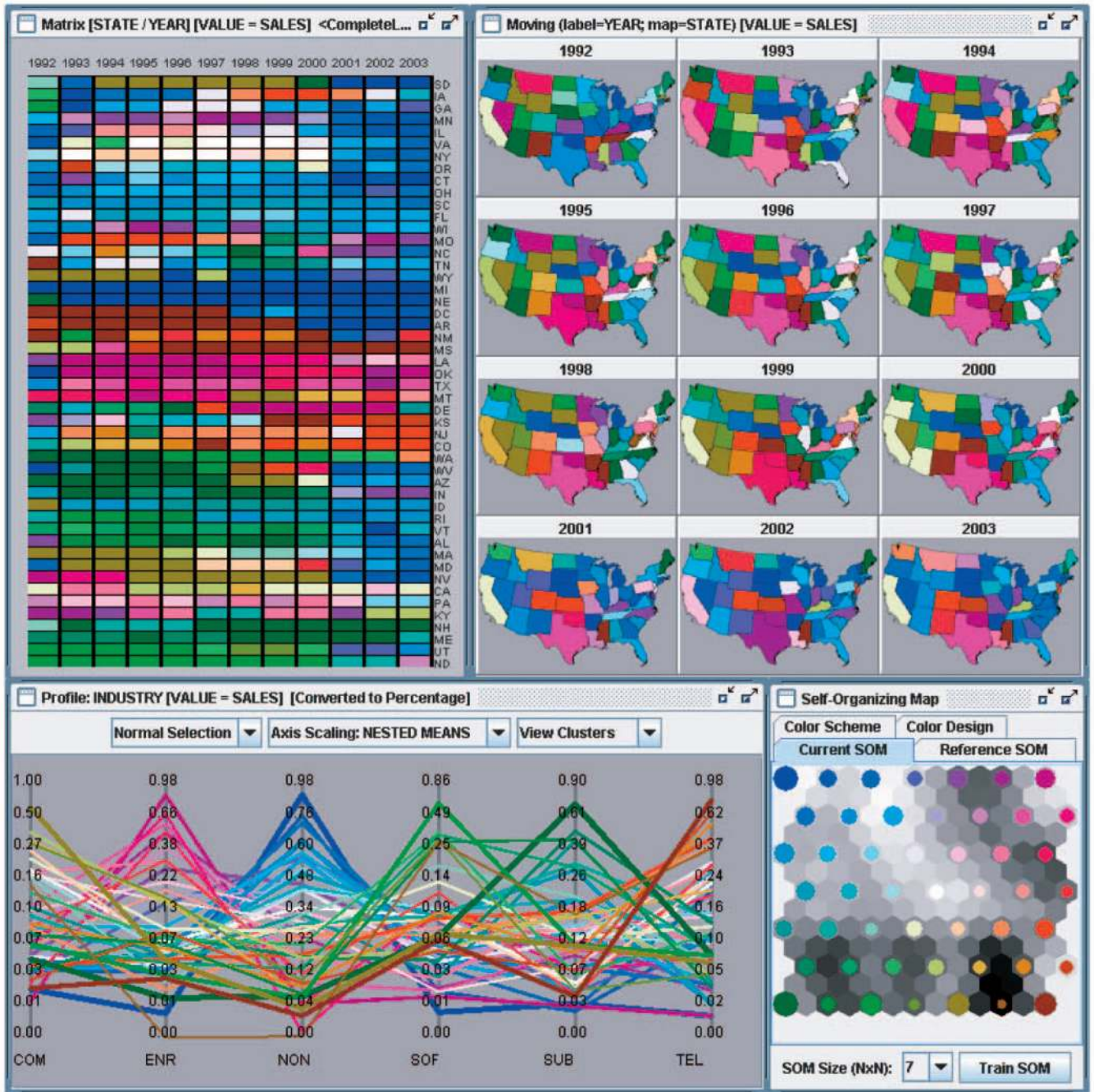
Fig. 5.

(Top-left) The reorderable matrix, (top-right) map matrix, (bottom-left) multivariate
"legend"-PCP, and (bottom-right) the SOM. The states (on the vertical dimension in the
reorderable matrix) are ordered with complete-linkage ordering (see next section). The
meaning for colors, which represent industry compositions, can be interpreted from the PCP.
In this snapshot, each axis in the PCP is scaled using nested-means and each string
represents a SOM cluster. The thickness of each string is linearly scaled to show the cluster
size.
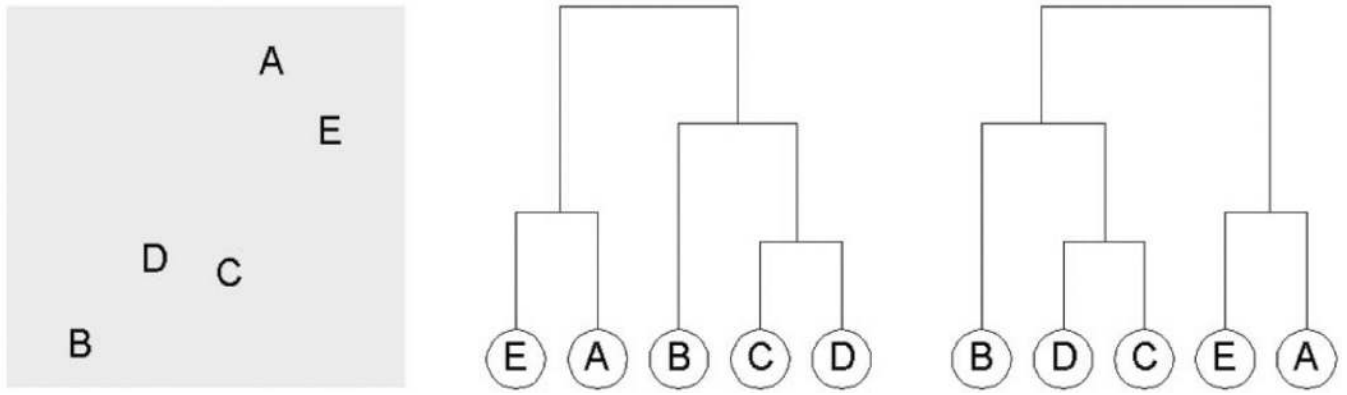
Fig. 6.
The two dendrograms represent the same cluster hierarchy of the five points but the right dendrogram gives a better ordering.

Fig. 7.
A comparison of two ordering methods. The matrix on the left is in an alphabetical ordering of US states. The matrix in the middle is ordered using the single-linkage ordering, while the matrix on the right is ordered using the complete-linkage ordering. The dissimilarity between two rows (i.e., states) is the Euclidean distance between their time-attribute slices, each of which is a 2D array of numerical values.
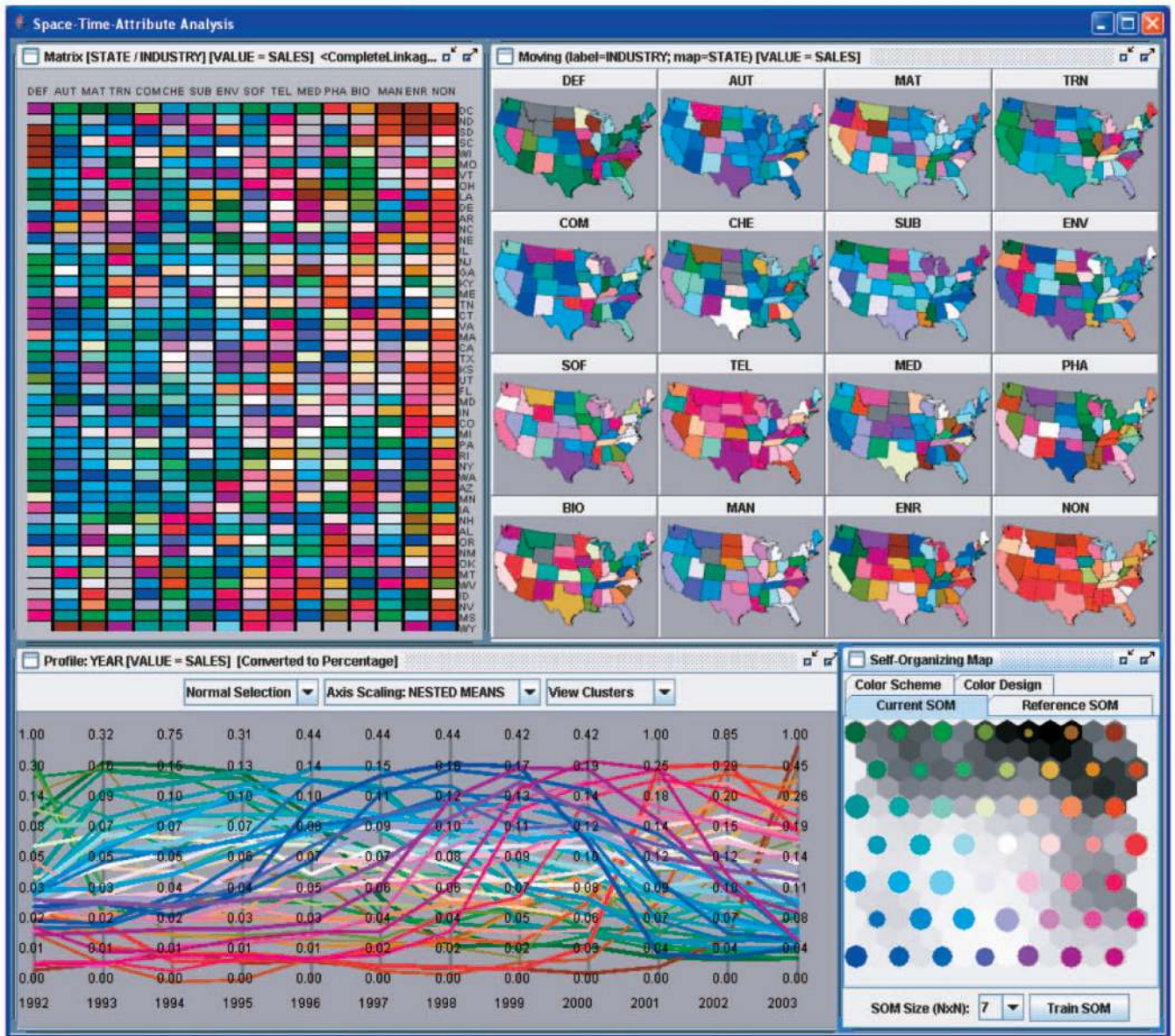
Fig. 8.
Spatial distribution and industry-by-industry variation of temporal trends. Both the rows and columns are ordered using the complete-linkage ordering method. The maps in the map matrix also follow the same order as that of the columns in the reorderable matrix.
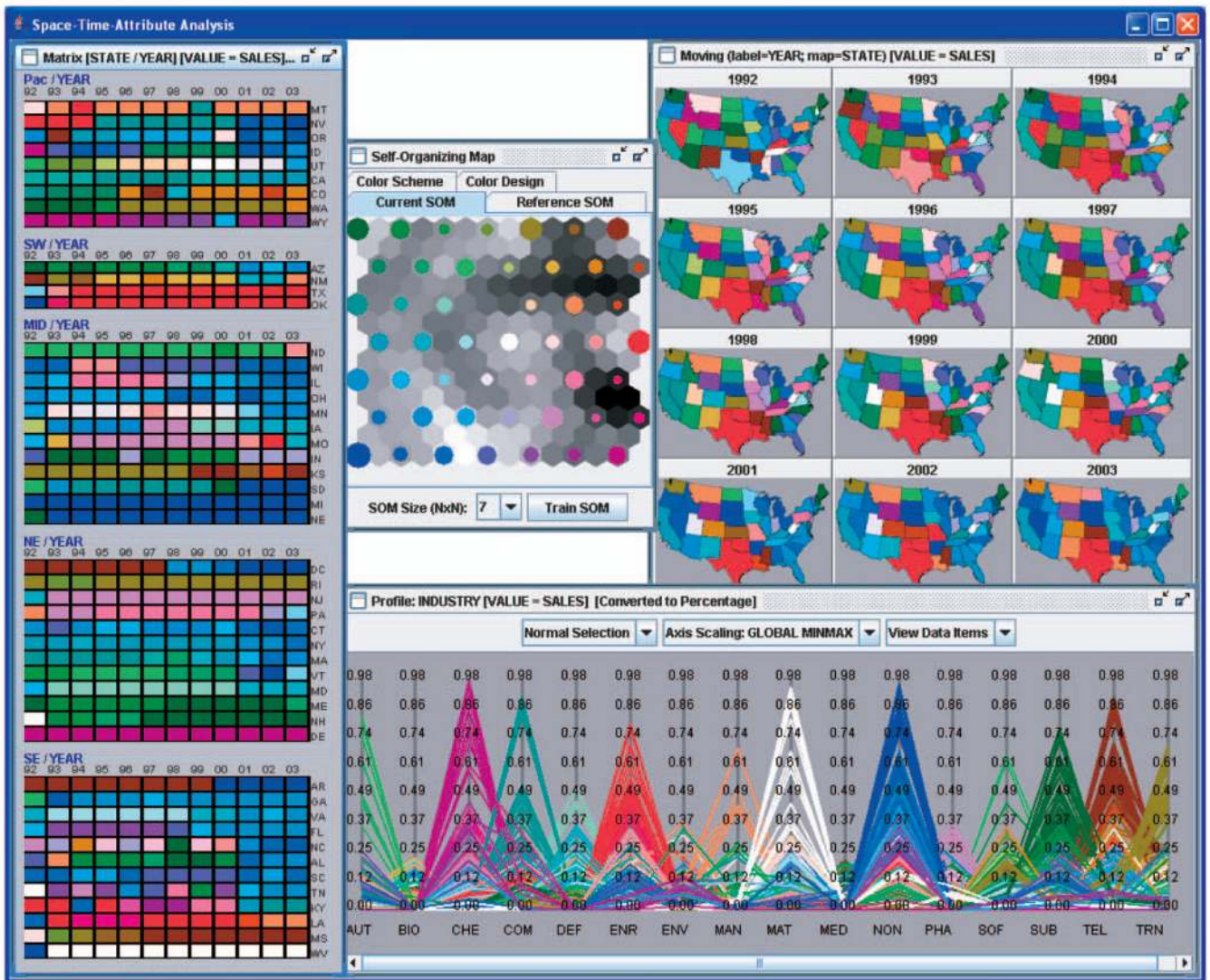
Fig. 9.
This is an overview of spatio-temporal and multivariate patterns of sales in the company data, which includes 16 industry types for 49 states (including DC) and 12 years. The states (on the vertical dimension in the reorderable matrix) are grouped into geographic regions and the states in each region are ordered with the complete-linkage ordering method.
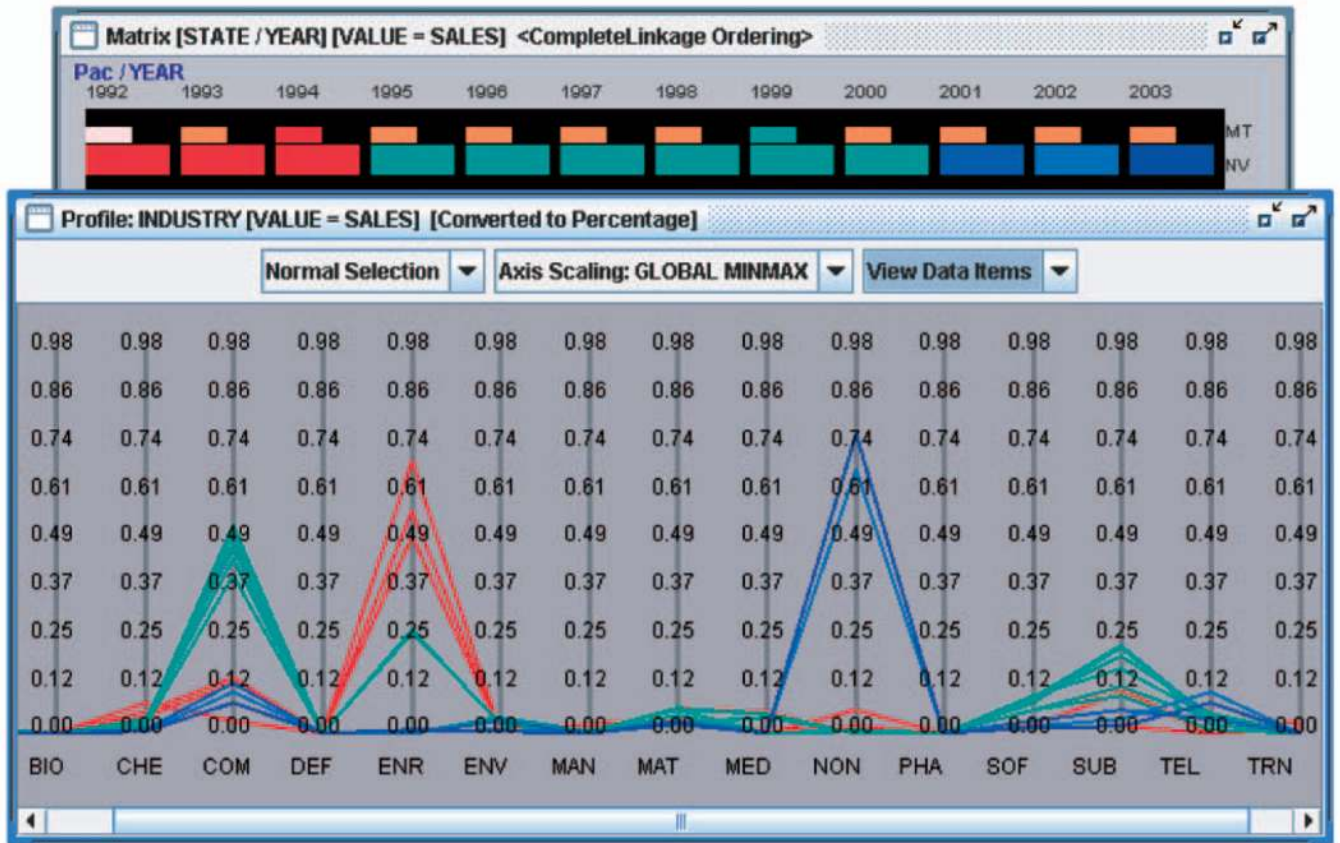
Fig. 10.
The row for Nevada (NV) is selected in the reorderable matrix to examine the change of
Nevada's industry composition over the 12 years. The PCP shows the selection at the data
item level and, thus, has 12 strings (one for each year). Each axis is scaled using the global
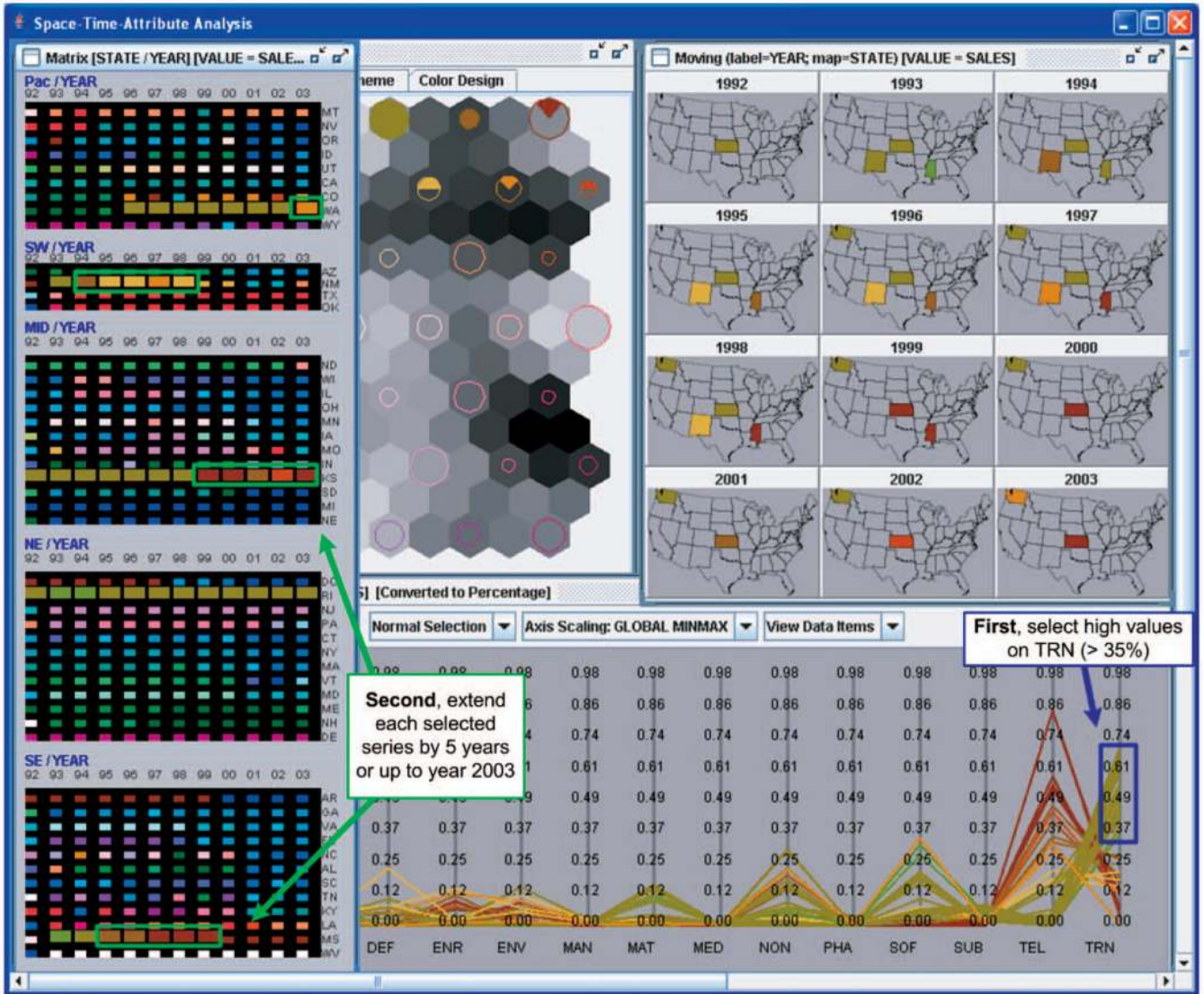min-max method.

Fig. 11.
This is a union of several selections made in both the PCP and the reorderable matrix. The purpose is to examine how states that were once dominated by the transportation industry (TRN) shifted to other industries.
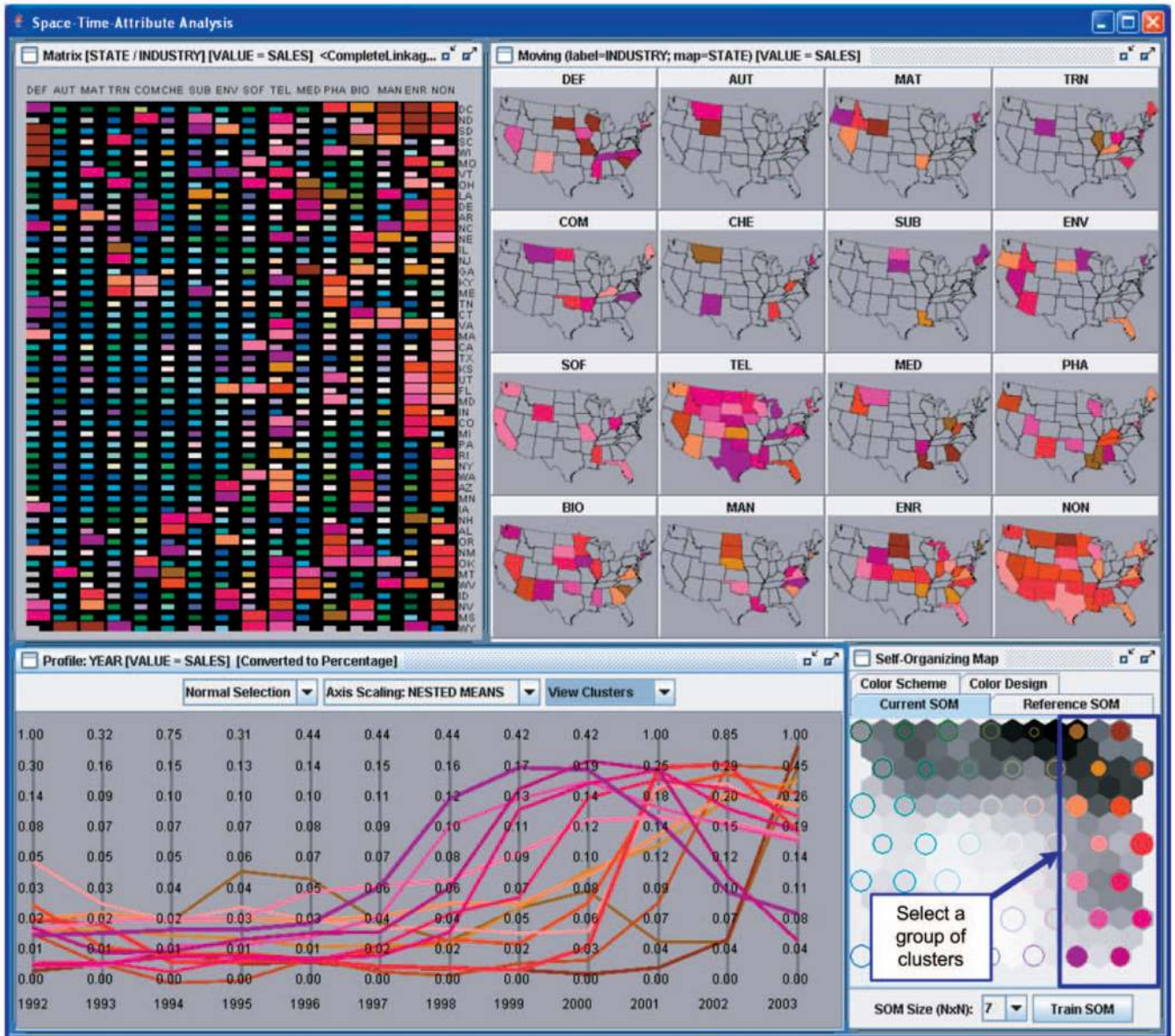
Fig. 12.
A selection of clusters made in the SOM view. See Fig. 8 for the overview of patterns. As interpreted from the PCP, purple clusters represent temporal trends that had their peaks during the late 1990's and 2000, while red and brown clusters represent a rapid growth since 2001.

Fig. 13.
Using the reorderable matrix to visualize spatial interactions, e.g., companies that relocated
from one state to another. Origin states are on the rows and destination states are on the
columns. The color (from a 5-class classification) of each cell represents the number of
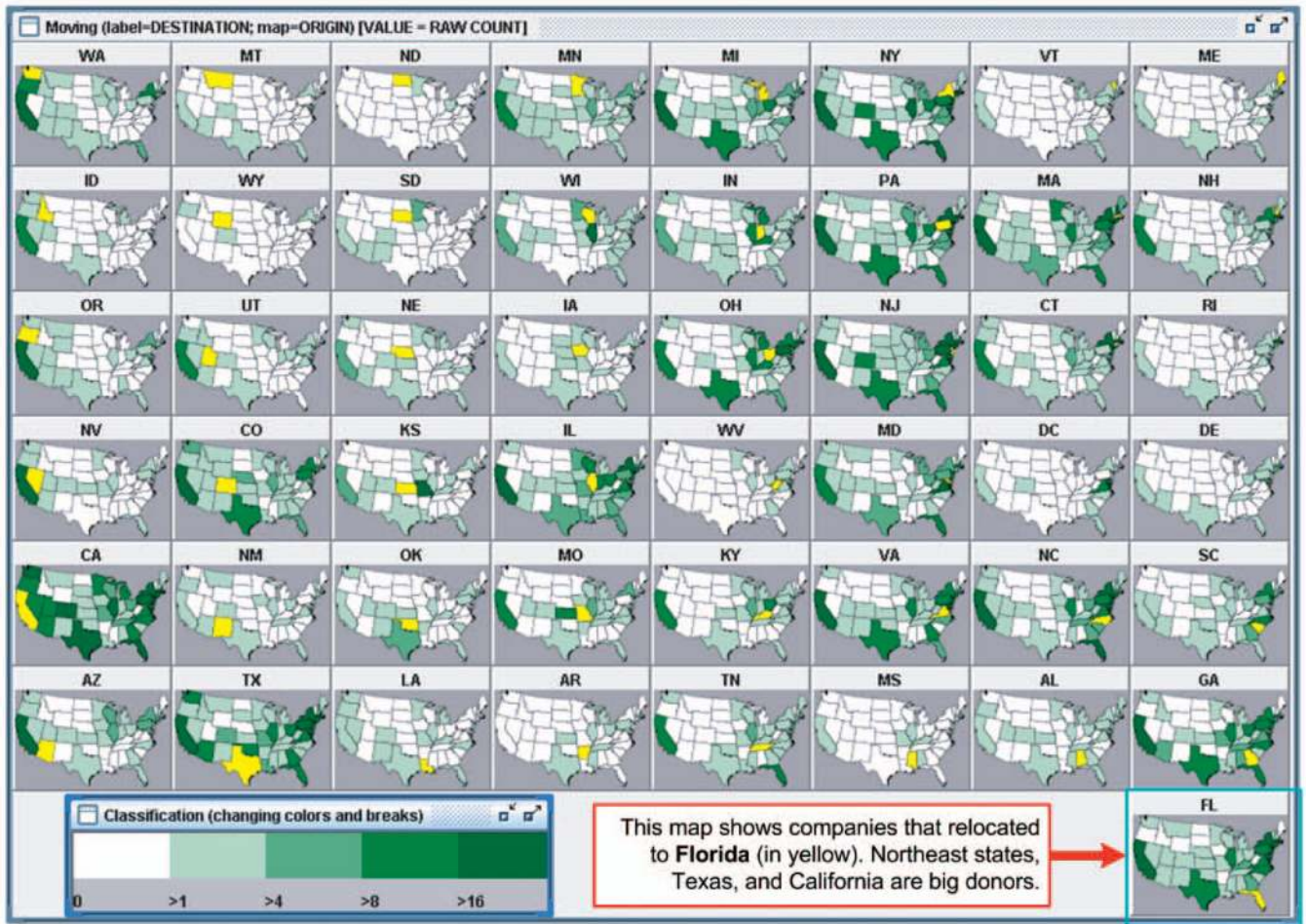companies relocated from the row state to the column state.

Fig. 14.
The Map^2 matrix to visualize spatial interactions. The overall view is a schematic "map" that contains multiple maplets. Each small map in the matrix represents all the companies that moved from all other states into a specific state, labeled above and highlighted in yellow.