Space-Time-Attribute Analysis and Visualization of U.S. Company Data

Jin Chen[1]
GeoVISTA Center
Pennsylvania State University

Diansheng Guo[2]
Department of Geography
University of South Carolina

Alan M. MacEachren[1]
GeoVISTA Center
Pennsylvania State University

ABSTRACT:

This research integrates computational, visual, and cartographic methods to develop geo-visual analytic strategies through which analysts can detect and explore multivariate, spatio-temporal patterns. The paper introduces a new form of geographic small multiple display along with a novel integration of computational and visual methods, for clustering and sorting large, multivariate data sets and for exploring spatio-temporal patterns in these data. The methods are applied to a data set containing time series, geographically-referenced data for companies in the U.S.

CR Categories: H.2.8 [Database Management]: Database Applications—*data mining, data and knowledge visualization;* I.5.3 [Pattern Recognition]: Clustering—*algorithms, similarity measures;* I.5.3 [Pattern Recognition]: Implementation—*interactive systems*; I.6.9 [Simulation, Modeling, and Visualization]: Visualization—*information visualization, multivariate visualization.*

Keywords: geovisualization, EDA, SOM, parallel coordinates, small multiples, visual-computational analysis

1 INTRODUCTION

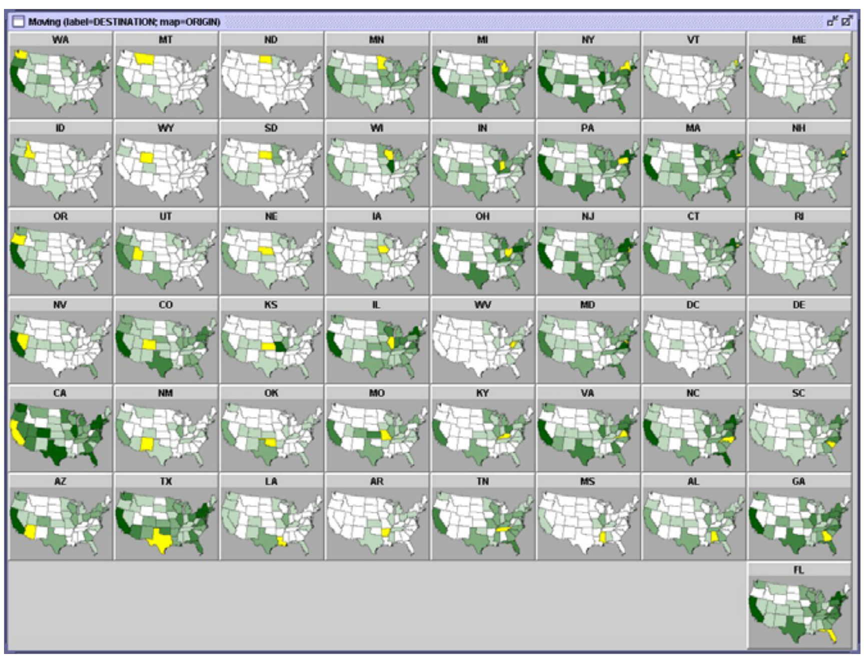
The research reported here focuses on developing geo-visual analytic methods that support analysis of large, multivariate, spatio-temporal data sets. This contribution leverages GeoVISTA *Studio* as a component sharing and application building environment. Specifically, we introduce representative examples from a set of novel, component-based data exploration and analysis tools that integrate computational, visual, and cartographic approaches. We demonstrate the application of these tools to analysis of the changing characteristics of U.S. industries.

The integrated approach to geo-visual analytic methods and tools we present is able to: (1) perform multivariate analysis (including time-series analysis) with the SOM; (2) encode the SOM result with colors derived from the ColorBrewerPlus component, which produces a 2D diverging-diverging color scheme; (3) visualize the data in a hierarchical data matrix view; (4) visualize the multivariate patterns with a modified Parallel Coordinate Plot (PCP) display and a map matrix; and (5) support human interactions to explore and examine patterns. The research shows that such integrated methods (computational and visual) can mitigate each other's weakness and collaboratively support discovery and analysis of complex space-time-attribute patterns, in an effective and efficient way.

In the two sections below, we introduce our schematic map matrix (Map^2) and a representative, more complex combination of integrated visual and computational tools. For more details on these methods and tools and on their application, see: www.geovista.psu.edu/resources/infovis2005/.

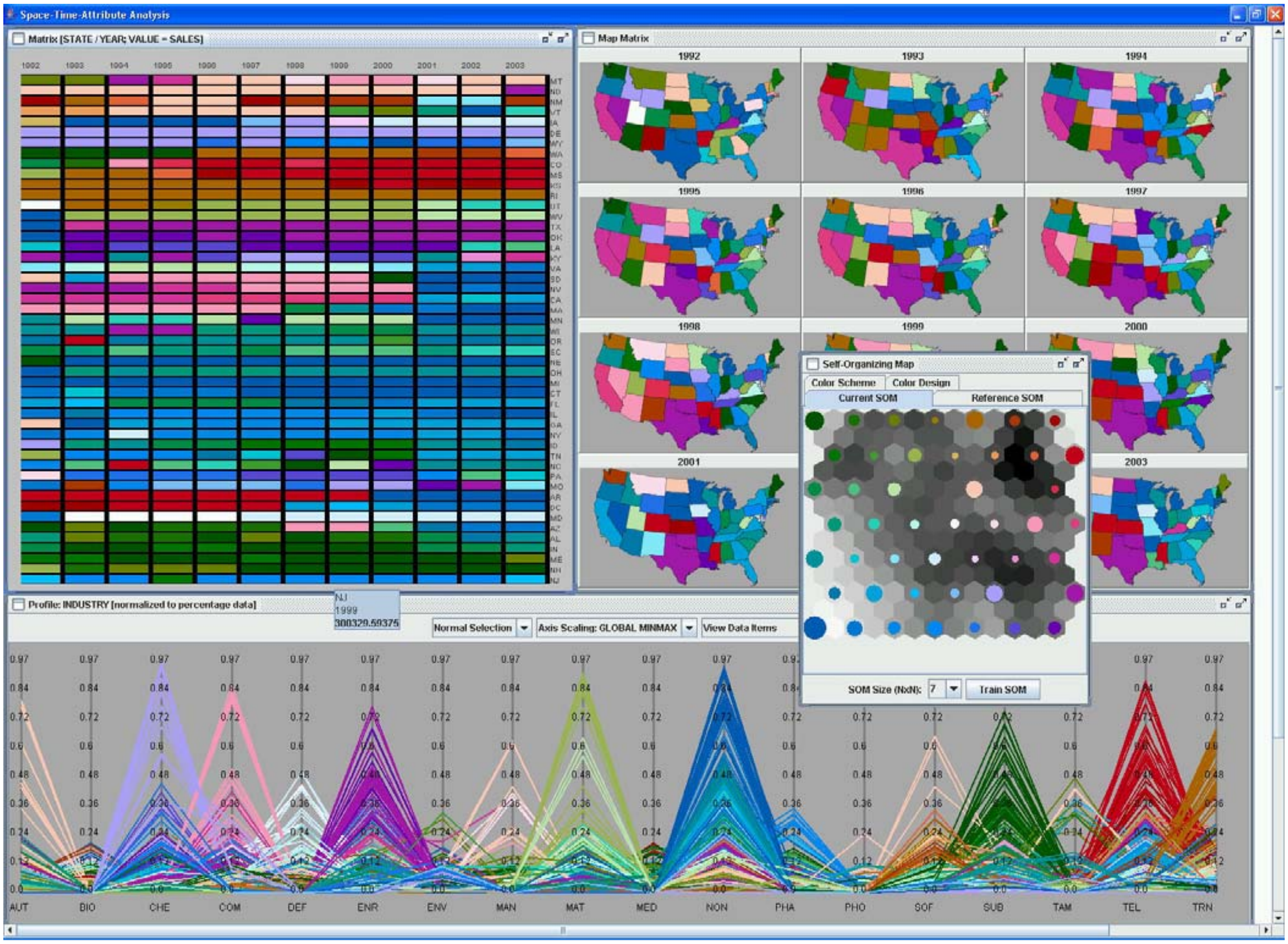
2 SCHEMATIC MAP MATRIX (MAP^2)

To address the unique challenges in analyzing and visualizing those companies that moved (from an ORIGIN state to a DESTINATION state), we developed a novel map matrix (Map^2) – see the figure below. 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. For example, the top-left map shows all the companies that moved to WA. The color represents the total number of jobs involved. These individual maps are ordered into an abstract map layout in which 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). The layout also tries to maintain the neighborhood relationship among states. Thus, this layout could be considered as a form of discontiguous cartogram. In each component map, the destination state is signified in yellow and its



name is shown as a label above that component map. When view the above map matrix as a single (abstract) map, we can see where most of the relocated companies (and jobs) moved. When look at each component map (or maplet), we can examine the attraction area of each state.

[1]GeoVISTA Center, 302 Walker, Penn State University, University Park, PA 16803; [jxc93; maceachren]@psu.edu
[2]Department of Geography, University of South Carolina, 709 Bull St., Room 127, Columbia, SC 29208; guod@sc.edu



3 VISUAL-COMPUTATIONAL ANALYSIS

The tool set we have developed supports integration of the Map^2 tool, or a standard small multiple map matrix, into applications containing a mix of computational and visual analysis components. In the example, SOM clustering groups industry mixes for space-time records of 18 industries (SOM output is in the foreground window). Circle size represents number of observations in a cluster. A diverging-diverging color scheme is applied to the cluster results and these colors are consistent across all views—the same color represents the same meaning, and similar colors represent similar multivariate data items [1]. The data matrix view (top-left in the image above) integrates an ordering algorithm [2], which can order the data entries in the matrix to effectively present space-time "regions" of similar multivariate patterns.

An advantage of our integrated approach is that, even without human interaction (e.g., brushing and focusing), we can perceive a holistic view of the multivariate spatial patterns by visually comparing several displays (a data matrix, a PCP, a map matrix, and a SOM). Thus, the approach inherently supports overview plus detail analysis. For example, from the image above we can easily spot states (across years) that have similar industry makeup and perceive how that makeup varies over time and space. The tools also provide a powerful capability for human interaction with objects in the dynamically linked views; this capability is illustrated in the accompanying video.

4 DISCUSSION

Our integrated analysis environment 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 it creates the potential to identify complex relationships in multivariate space); and (2) its flexibility in addressing a range of analysis questions dealing with space, time, and attributes. The combination of computational and visual methods makes it possible to derive meaning from a much larger data set than would be possible with visual methods alone.

A limitation of our tools, in this initial analysis of the industrial benchmark data set, is that we aggregated company statistics to state-level, thus are likely to miss some patterns that span state boundaries. However, the variety of interesting patterns found at this coarse geographic resolution demonstrate the 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.

REFERENCES

[1] Guo, D., M. Gahegan, A.M. MacEachren, and B. Zhou (2005). "Multivariate Analysis and Geovisualization with an Integrated Geographic Knowledge Discovery Approach". Cartography and Geographic Information Science. 32(2), pp. 113-132.

[2] Guo, D. (2003). "Coordinating Computational and Visualization Approaches for Interactive Feature Selection and Multivariate Clustering". Information Visualization, 2(4): 232-246.

Acknowledgement: This research was supported by an ARDA GI2-Vis grant and by grant CA95949 from the National Cancer Institute.