GeoViKE: Support Knowledge Exploration with Geo-Visualization

Geng Sun[*]

Pennsylvania State University

Xiaolong Zhang[†]

Pennsylvania State University

ABSTRACT

This paper introduces GeoViKE (Geo-Visualization for Knowledge Exploration), a geo-visualization tool, to help people analyze, understand and explore technology trends in the United States based on the InfoVis 2005 Contest data set. GeoViKE integrates geo-visualization techniques, such as map using, dynamic query, and coordinated multiple views, and provides users with zooming tools to explore information at different levels of aggregation. The paper presents results of visualizing company moving, industry sales, industry distribution in the US, and industry composition in states. More details of this work are available online at http://zhang.ist.psu.edu/infovis2005contest.

1 INTRODUCTION

It is a challenge to understand large information structures. InfoVis 2005 Contest data set has about 500,000 entries of sales data of US technology companies from 1989 to 2003, and nearly 2-million entries of their products. Making sense of such a huge amount of information requires effective tools to organize data, search for useful information, and distill knowledge. Information visualization can help to address such issues[4].

In this paper, we present GeoViKE (Geo-Visualization for Knowledge Exploration), a visualization tool which integrates geo-visualization techniques into zoomable user interfaces. It can help users discover the distributions and patterns of information of interest at different geographic levels. This paper first introduces our data analysis technique, and then presents the strength of this work with some visualization results. Finally, we discuss some unaddressed issues of this work.

2 DATA ANALYSIS

We are interested in such information as company moving, industry sales, industry composition in individual states, and industry sales distributions in the US. To obtain relevant data, we first imported the InfoVis 2005 Contest data into Microsoft Access and then created various queries based on our needs.

2.1 Company Moving

A company is regarded as moving when its headquarters was in different states in consecutive years. Information collected from such a company includes its original state, its destination state, the year the moving happened, and its industry type, sales, and employee numbers before and after moving.

2.2 Industry Sales

Industry sales data was obtained by first classifying companies into different groups based on industry, year, and geographic location and then adding sales of all companies in each group together. To help better understand the sales distributions at different geographic levels, we grouped companies at three levels:

[*] email: gengsun@gmail.com

[†] email: lzhang@ist.psu.edu

region, state, and county. Six region groups were used: New England, Mid-Atlantic, South, Midwest, Southwest, and West. A company's zip code can be used to determine its county group. Information about zip codes and counties was from a commercial database[6].

2.3 Industry Composition

To understand the trends in individual states, all 18 industries were ranked based on their sales from 1989 to 2003. The percentages of the sales of the top three industries out of the total sales of all industries were calculated.

2.4 Industry Distribution

We also ranked 50 states in each industry based on each state's sales. The percentages of the sales of the top three states out of the total sales of all states were computed.

3 STRENGTH

Combining geo-visualization techniques and zooming tools can help users better discover and identify the major technology trends and relationships over time.

3.1 Tying Visual Presentation with Maps

In GeoViKE, visualization results are presented on the top of maps to help user link visual information with geographic attributes. For example, Figure 1a shows the most popular destination states of company moving. Each circle represents a state, and is positioned appropriately on a US map. Its size indicates the number of companies moving to that state. From this figure, users can easily tell which states companies were more likely to move into. Knowing California was the most popular destination, users can further examine the origins of those companies. Figure 1b shows cross-state moving as lines between two states. Line thickness represents the number of companies. It is simple to find where companies were from and how many.
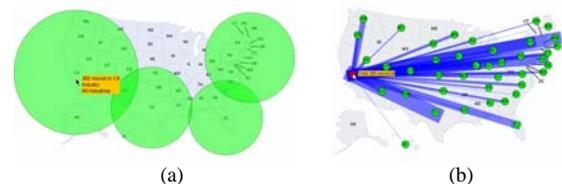

(a)   (b)
Figure 1: Visualizing Company Moving

3.2 Zoomable User Interface

GeoViKE is a zoomable user interface built upon Piccolo.Java[2], and allows users to observe important characteristics at different geographic levels. To understand industry sales data, for example, users can zoom in and out to find distribution patterns at different geographic levels, from region, state, to county. Figure 3 shows the distributions of sales in computer hardware and software industries at the levels of region, state, and county respectively. It is seen that sales concentrated in the West and Mid-Atlantic regions (Figure 2a); the industries clustered in the states on the west and east coasts (Figure 2b); and two counties dominated sales in Arizona (Figure 2c). Zooming tools help users explore information across different scales. Moving from Figure 3a to 3b,

users can tell that although the West region was strong in general, sales distribution was very uneven among states within it.

(a)   (b)   (c)

Figure 2: Sales Distributions at Three Geographic Levels

3.3 Coordinated Multiple Views

Multiple and coordinated views can help people explore the same set of information from different aspects[3]. GeoViKE has two windows for visualization and interactive control, which are coordinated and linked. When queries are modified in one window, information displayed in the other will change correspondingly. Also, interactive actions in one window will be reflected in the other. For example, in visualizing industry sales at the level of region, while one window renders regions with different colors in a map (Figure 2a), the other shows bar charts so that users can compare sales (Figure 3). Coordinated multiple views provide richer information and more flexible control.

Figure 3: Coordinated Bar Charts in the Second Window

3.4 Dynamic Query

Dynamic query can improve the presentation of visualization and provide more focused results [1]. For example, Figure 4a shows which states companies moved into and how many. To reduce the clusters, users can dynamically adjust the minimum size of visible circles to simplify the view (Figure 4b) or set up a query criterion, such as finding companies with better sales after moving, to narrow down the results (Figure 4c). Dynamic query allows a better control on the visual presentation and needed information.
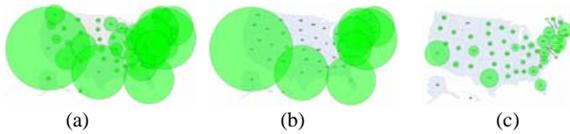

(a)   (b)   (c)

Figure 4: Dynamic Query

3.5 Various Visualization Techniques for Exploration

GeoViKE integrates various visualization techniques for information exploration. Pie charts are used to show industry composition in individual states. Figure 5a shows two states with similar industry composition. Users can easily see that California and Texas had the same top three industries, but their No. 1 industry was different. GeoViKE uses geometry size to represent magnitude information. Figure 5b visualizes the top three states in pharmaceuticals. In addition to seeing the industry clustering in northeast states, users can also compare the sales among three states based on the size of circles.

GeoViKE provides parallel coordinates in visualizing multi-dimensional data[5]. Figure 6a shows industry sales from 1991 to 2003. Two interesting characteristics can be found: computer hardware and telecommunication industries peaked in 1993, but it is the non high-tech sector that consistently outperformed all other industries from 1991 to 2003, except in 1993. Examining the data of sales per employee (Figure 6b), we can find that although the peaks of hardware and telecommunication industries in 1993 are still visible, the non

high-tech sector is not seen better than others. (The energy industry actually has the best record here.) This may imply a market boom in the industries of hardware and telecommunication in 1993 and then a quick correction thereafter, and a much larger work force in the non high-tech sector than others.
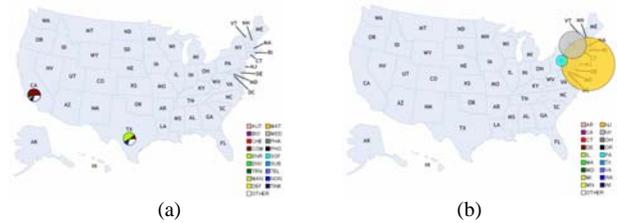

(a)   (b)

Figure 5: Pie Charts in Visualizing Industry Composition
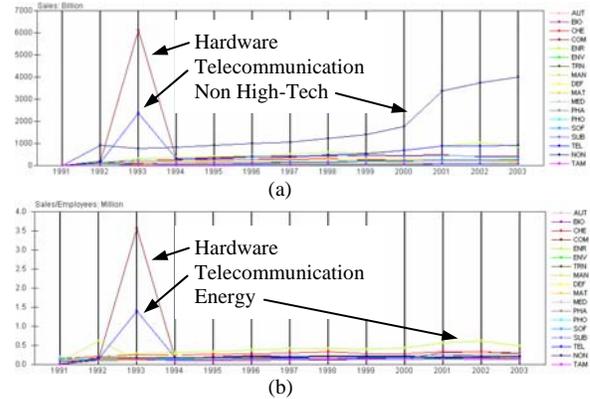

(a)

(b)

Figure 6: Industry Sales: a) Total Sales; b) Sales /Employee

4 CONCLUSIONS

We presented a zoomable, geo-visualization tool to analyze and understand the InfoVis 2005 Contest data set. Using only GeoViKE, we answered all contest questions. More importantly, GeoViKE allows users to explore the huge data set and discover interesting patterns and relationships, such as the similarity of state industry composition.

GeoViKE primarily focuses on visualizing numeric data. Text data, such as company products, has not been addressed. Also, GeoViKE uses color to code category variables, such as state and industry. Users may find it difficult to distinguish color schemes.

5 ACKNOWLEDGEMENTS

REFERERNCE

[1] Ahlberg, C., et al. (1992) Dynamic Queries for Information Exploration: An Implementation and Evaluation. In Proceedings of CHI 92, 619-626.

[2] Bederson, B., et al. (2003) Toolkit Design for Interactive Structured Graphics, IEEE Transactions on Software Engineering, 30(8), 536-46.

[3] Brodbeck, D., & Girardin, L. (2003) Design Study: Using Multiple Coordinated Views to Analyze Geo-Referenced High Dimensional datasets. In Proceedings of IEEE CMV 2003.

[4] Card, S. et al. (1999) Readings in Information Visualization. Morgan Kaufman Publishers, San Francisco, CA.

[5] Inselberg, A., & Dimsdale, B. (1990) Parallel Coordinates: A Tool for Visualizing Multi-Dimensional Geometry. In Proceedings of the 1st Conference on Visualization '90, 361–78.

[6] http://www.melissadata.com/index.htm