An Enhanced GIS Environment For Multivariate Exploration: a Linked Parallel Coordinate Plot Applied to Urban Greenway Use Survey Data

Robert M. Edsall
Department of Geography
Arizona State University
robedsall@asu.edu

Anna J. Roedler
Department of Geography
Arizona State University
ajroedler@yahoo.com

Abstract

In the winter of 2001-02, a study of use patterns of an urban greenway in Scottsdale, AZ, was carried out using on-site surveying. This study is designed to assess the influence of the greenways on the perception of the quality of life for users of the park. This paper will discuss methods of geovisualization and exploratory data analysis applied to the display and interpretation of the data that comes from these surveys. The representation of survey results poses challenges because the data are both multivariate and involve variables of several different measurement levels, including nominal-, ordinal-, and ratio-level. These variables include diverse but potentially related indices such as distance traveled, primary use type, age, and perception of safety. The character of the survey data makes visualization of the database both necessary and difficult: maps and graphs from ESDA and geovisualization that show many variables at a given time were required. For this task, a modification of a customized ArcView interface, enhanced with a parallel coordinate plot representation, was created. The parallel coordinate plot application in ArcView, heretofore used for visualization of physical (climate) and epidemiological (cancer mortality) data, proved insightful for the display of the social and cultural data contained within the surveys.

The overall project was inspired by a study conducted in Texas based on a user survey of three different urban greenways in Texas (Shafer et al. 2000). Through a three-day survey of trail users this group collect an astounding 1004 on-site surveys. After analysis of the survey results the three greenways were found to have very different use patterns. Although this study focused on how overall the three greenways were perceived to contribute to quality of life, there was little discussion of how the use of the greenway affected this perception. The results were mainly displayed using conventional tables of statistics. If the results could have been displayed not only by greenway, but also by use or by another variable, the researchers might have been able to gain insight about whether use patterns affected answers or whether it was another variable that had a strong correlation with certain quality of life answers. Our study is based on their methodological approach to greenway survey collection, but also incorporates the use of methods of geovisualization and ESDA to analyze the survey results.

The use of such methods in such a social science application, however, poses interesting and challenging issues. Geovisualization (and visualization in general) has traditionally been applied to the numerical data sets commonly found in the physical sciences (Orford et al. 1999). The adaptation of these techniques to handle qualitative data will be important for its application to the social sciences. Visualization methods in the social sciences often are directly adapted from those designed for quantitative data, as nominal or ranked data is converted, possibly arbitrarily, into quantitative ratio-level variables to fit methods of visualization (Dorling 1992). This is often the case in information visualization applications, where text and other apparently qualitative data types are converted and plotted into a similarity or possibility space based on some numerical index applied to each observation (Rose 1999).

The challenge, thus, is to devise a scheme to place values of a variable in some way that reflects their relative value to one another. This creates the danger of prompting false impressions of order in (non-ordered or arbitrarily ordered) qualitative data. Our paper will discuss attempts at tackling this challenge. One solution involves associating a nominal data set with a ordered numerical variable: for example, the survey asked participants which park feature they most frequently used. We decided that the park features would be ranked in terms of their mobility. That meant that fishing was given the lowest number of 1 (since it was the most dependent on a small area) and biking on paths was given the highest rank of 9. Another possible solution explored in the development of our application is the "mapping" of nominal data in an arbitrary order with a user-friendly interface element that allows – indeed encourages – the rearrangement of the values, thus emphasizing to a user the fluidity and subjectivity of the default ordering.

Aside from the inherent and important issues involved with visualizing nominal data, the survey results are also difficult to represent because of their multivariate nature. A simple and elegant solution to this problem is the parallel coordinate plot (PCP), first described by Inselberg (1985). On the parallel coordinate plot, observations are represented on a PCP as a series of unbroken line segments instead of as points (as in a scatter plot). These lines pass through parallel axes, each of which represents a different variable. Each line passes through an axis at a location that indicates the observation's value relative to all other values. The ends of the axis represent the maximum and minimum values of the axis variable for all observations under consideration. The result is a (possibly unique) multivariate *signature* for each observation, and a visual representation of relationships among many variables.

Using the PCP, interactions among variables can be quickly identified. Observations with similar data values across all variables will share similar signatures; clusters of like observations (survey respondents in this case) can thus be discerned. Two variables directly related to one another would appear on the PCP as two axes connected by a series of parallel (or at the least non-crossing) line segments. Conversely, an inverse relationship between two variables would be displayed as a series of line segments that cross each other between the axes. In a PCP representation of automobile specifications, gas mileage and vehicle weight is shown by Wegman (1990) to exhibit this relationship, and the PCP representation resembles, vaguely, a "bow tie" of intersecting line segments. Another advantage of the PCP is the easy visualization of distributions and characters of many variables at once; the PCP design resembles a series of connected histograms, where distributions that are normal or skewed, and variables that are continuous or discrete, can easily be distinguished.

Although this technique has been applied successfully to physical phenomena and to large data sets (MacEachren et al. 1999), its use in the social sciences seems to be lagging (Fotheringham 1999). One reason is that the PCP generally requires qualitative data to be subjectively ordered, since the order of the possible values on each axis has an effect on the resulting patterns of the lines and thus on the outcome of the visualization. The PCP has been modified by several developers to accomplish certain goals and represent certain types of data (Wegman 1990; Chang and Yang 1996; ; Fua et al. 1999; Edsall 1999), and this paper will represent another PCP modification that is designed to encourage creative thinking about the relationships among both qualitative and quantitative variables.

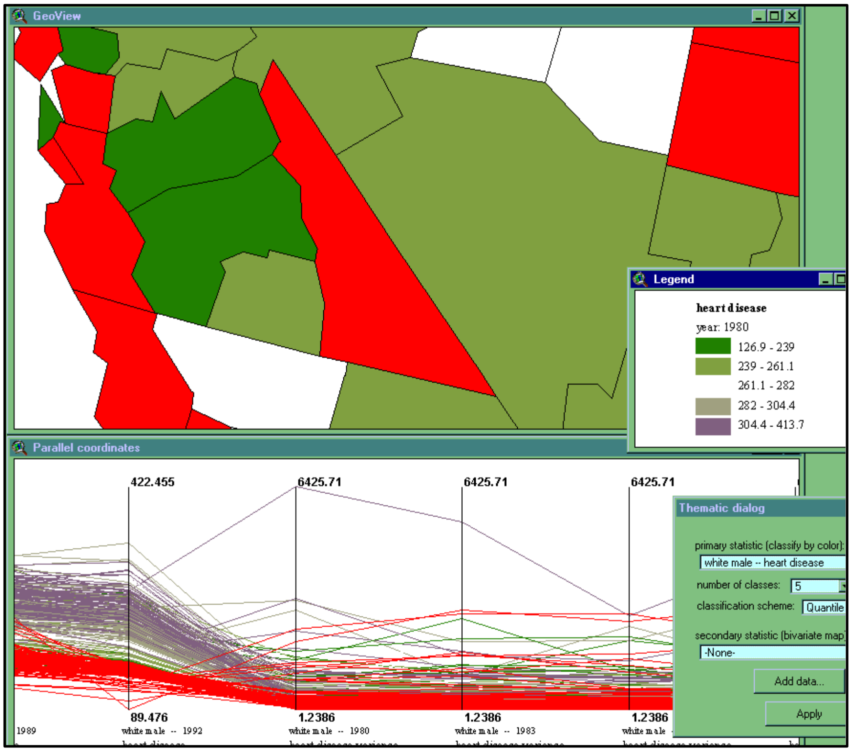

Figure 1. The modified ArcView interface on which the tool described here is based (from Edsall 2002).

The environment to be presented is an enhanced version of ArcView's interface. The enhanced version is itself a modification of an environment developed for the visualization of health statistics (Figure 1, Edsall 2002). Geographic information systems have been linked to exploratory data analysis software in the past (Cook, Majure et al. 1996). This version was created using ArcView's internal interface development language. The GIS environment was selected for its facility to link newly constructed graphical representations to maps, clearly a priority in geovisualization applications. In this application, each participant's trace in the PCP will be linked to the corresponding location of their interview and/or the recreational area being used depicted on the survey map. A variety of interactive elements, including but not limited to those that are standard in GIS interfaces (like a pan and zoom), were added to the PCP (and where appropriate, linked to the map) to facilitate the visual analysis of the data. These elements will be described in detail in the full paper.

By adapting methods of visualization to fit qualitative and quantitative data collected in the social sciences it is hoped that more effective analysis can take place. Through the example of visualizing survey results with the parallel coordinate plot we will show how visualization can be used with data collected to describe human behavior and perceptions.

References

Chang, D. H. and S. J. Yang (1996). "Dynamic Parallel Coordinate Plot and its Usage." *Journal of the Korean Society of Applied Statistics* 9 : 45-52.

Cook, D., J. J. Majure, J. Symanzik and N. Cressie (1996). "Dynamic Graphics in a GIS: Exploring and Analyzing Multivariate Spatial Data Using Linked Software." *Computational Statistics* 11 : 467-480.

Dorling, D. 1992. Visualizing People in Time and Space, *Environment and Planning B: Planning and Design*. 19: 613-637.

Edsall, R. (1999). Development of Interactive Tools for the Exploration of Large Geographic Databases. *Proceedings of the 19th International Cartographic Conference*, Ottawa, University of Victoria. 34B.

------------ (2002) The parallel coordinate plot in action. *Computational Statistics and Data Analysis* (in press).

Fotheringham, S. 1999. Trends in qualitative methods III: stressing the visual. *Progress in Human Geography* 23(4): 597-606.

Fua, Y.-H., M. O. Ward and E. A. Rudensteiner (1999). Hierarchical parallel coordinates for exploration of large datasets. *Proceedings of the IEEE Symposium of Information Visualization*, San Francisco, CA, IEEE Computer Society.

Inselberg, A. (1985). "The Plane with Parallel Coordinates." *The Visual Computer* 1 : 69-97.

MacEachren, A., Wachowicz, M., Edsall, R., and Haug, D., 1999. Constructing knowledge from multivariable integrating geographical visualization with knowledge discovery in database methods, *International Journal of Geographical Information Science*. 13 (4): 311-334

Orford, S., Harris, R., and Dorling, D. 1999. Geography: Information Visualization in the Social Sciences., A state-of-the-art-review. *Social Science Computer Review*. 17(3): 289-304.

Rose, S. (1999). The sunflower visualization metaphor, a new paradigm for dimensional compression. *Proceedings of the IEEE Symposium on Information Visualization (InfoVis '99)*, San Francisco, CA, IEEE Computer Society

Shafer, C.S., Koo Lee, B., and Turner, S., 2000. A tale of three greenway trails: user perception related to quality of life. *Landscape and Urban Planning* 49: 163-178.

Wegman, J. J. (1990). "Hyperdimensional Data Analysis Using Parallel Coordinates." *Journal of the American Statistical Association* 85 (411): 664-675.