

Conditioned and Manipulable Matrix For Visual Exploration

Xiping Dai, Frank Hardisty

302 Walker Building,
University park, PA 16802,
(814) 865-9656,
xpdai@psu.edu,
hardisty@geog.psu.edu

<http://www.geovistastudio.psu.edu/GeoViz/jsp/index.jsp>

1. Introduction

As the size and complexity of federal statistical summaries increase and accumulate, the development of quality graphics for data visualization and exploration attracts increasing research effort. While some research has focused on new designs for visual presentation and exploration of highly multivariate data (Fua *et. al.*, 1999), a reconsideration of perceptual and cognitive factors leads us to another approach to improve users' ability to extract information based on traditional statistical graph and map representations. With effective interfaces and interaction techniques, users can maximize information obtained from familiar and easy-use exploratory statistical graphs.

User-centered interfaces and interaction design, as well as implementation of component-based traditional statistical graphs, for example the scatterplot matrix, are the focus of this paper. Based on a traditional matrix view of data, such as a scatterplot matrix, this paper enhances the matrix in two aspects. One is to support the direct manipulation of data representation in row/column position (Siirtola, 1999) and ease the selection of data to visualize using variable fields and observed data ranges. Users can easily select the variables or a range of observations they want to focus on and change the row/column position in the matrix dynamically via a graphical interface. The second enhancement to a "traditional" scatterplot matrix is to add the capability to *condition* the data depicted in the matrix on a related variable (conditioning is described below). Our goal is to offer an intuitive interface, through which users can visualize and manipulate any subsets within large geospatial datasets using familiar statistical tools and thus help analysts in partner U.S. government agencies to identify outliers, errors, spatial clusters and patterns, and possible multivariate correlations from both the users and datasets perspectives. Furthermore, using a component-based design for the matrix allows it to be coordinated with other software components, in order to enable more comprehensive and flexible information visualization (e.g., linked brushing) and database queries.

This project is part of the Digital Government Quality Graphics project (grant #99883451), which has a specified objective to develop and assess quality graphics for federal statistical summaries. The manipulable matrix introduced here is represented by a Java bean in a programming free visualization environment—GeoVISTA Studio (Gahegan *et. al.*, 2000; MacEachren *et. al.*, 2001), where the matrix can access coordination and collaborate with other visual exploratory statistic graphs or maps.

2. Manipulable Matrix

This research highlights human-centered and component-based design and focuses on the conditioned and manipulable matrix (part of the approach includes iterative usability testing that is addressed in a complementary paper by Suchan and Hardisty). As an exploratory visualization tool, this design method considers the usability of methods and tools developed visualizing particular datasets and particular ranges of observations in large volumes of statistical data.

The matrix developed allows users to select variables to display in the matrix directly (Figure 1), arrange the order of rows and columns by direct manipulation, and control the representation of observations in datasets based on any variables (displayed or not). The size of matrix changes according to the number of variables users select through the variable selection dialogue. For variables displayed, the user can set the upper and lower bounds of any scatter plot axis and the

resulting “focusing” action will be applied to all matrix elements that depict the variable acted upon. In addition, the users can exercise control through a conditioning component, modeled on ideas introduced by our Digital Government collaborator Dan Carr and colleagues (Carr *et. al.*, 2000). The conditioning component allows the user to limit visual analysis in the scatter plot matrix (or other display components) to a subset of data defined by a related variable (e.g., a scatter plot that depicts relationships among education, income, and mortality can be conditioned on proportion of individuals without health insurance, through a control that allows the user to apply a filter that depicts the primary variables on the scatter plots only if their values on the conditioning variable are within a range specified).

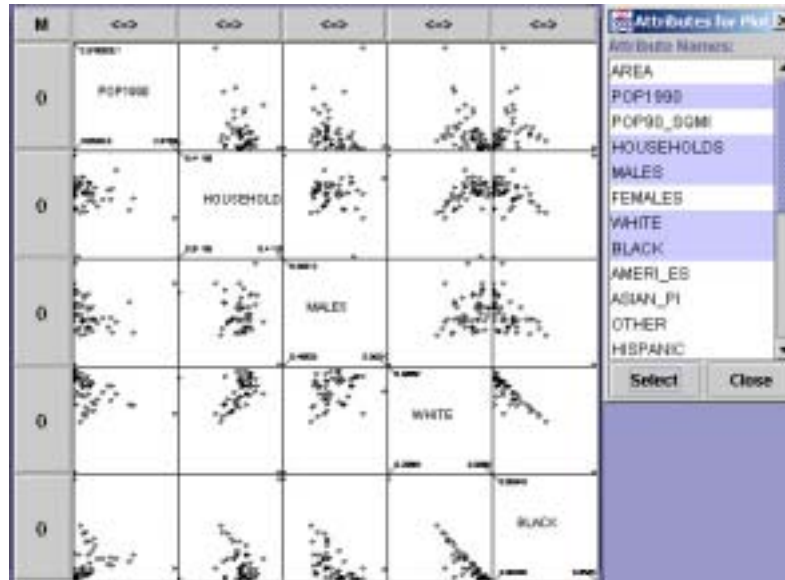


Figure 1. Variable selection dialog. Selected variables will be displayed in the matrix.

As noted above, the matrix can also be coordinated with other views on the data. This is accomplished through a “coordination bean.” The matrix is made “coordinator-aware” through a “stub” component that makes use of the coordinator bean to communicate with other independent components, such as a map, a parallel coordinate plot (PCP), and a table browser. Dynamic coordination among components enables a dynamic and comprehensive analysis of geospatial statistical data. The concept of dynamic coordination between independent visual exploratory components extends the traditional idea of linked brushing (Becker and Cleveland, 1987), and realized in this study as selection and focusing (Figure 2).

Figure 2 illustrates an application example using a conditioned manipulable scatterplot matrix with dynamic linked with corresponding 2D map. The scatterplot matrix on the left side displays five variables selected by users. The names of variables are located on the diagonal elements of the matrix. The other elements in the matrix are scatterplots with horizontal and vertical axis corresponding to variables in the diagonal boxes. The order of variables or scatterplots displayed in the matrix can be changed dynamically by using cursor to drag rows or columns to new positions. The default color displaying observation dots is black. The blue color in the scatterplot matrices and map indicates that those observations have been selected (brushed). The selection action on any coordinator-aware component, a scatterplot in the matrix here, is passed into other scatterplots in the matrix and other components through the coordination bean described above.

Conditioning is used here to limit scatter plot displays to data elements for which the value for the conditioning variable fits criteria specified through a conditioning dialogue (Figure 2), which is connected to the matrix. In the example, we change the condition range from 0-37.9 to 10-37.9 based on unemployment rate. This focused observation depicts the relationship among smoking, obesity, and several demographic variables for the subset of places having relatively high unemployment.

The linked brushing implemented in the matrix design enables dynamic interaction between different components, such as map, PCP, table browser, in the connected system. In this example, the

selection information was passed into map component and the counties with high smoking rates were highlighted. A spatial pattern of relatively high smoking rates based on county level is revealed.

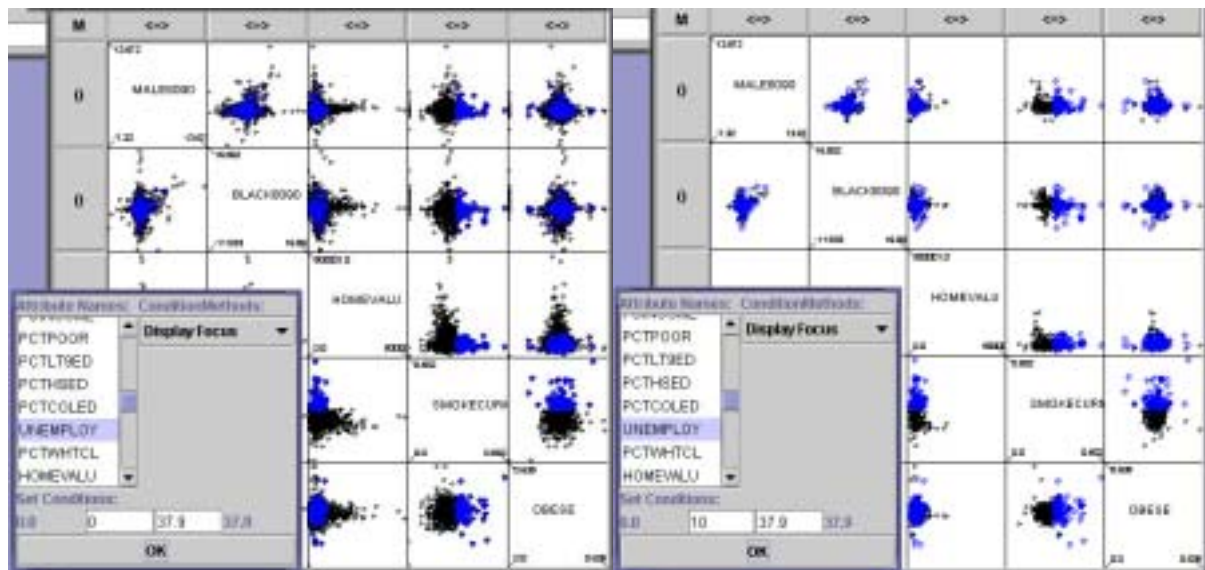
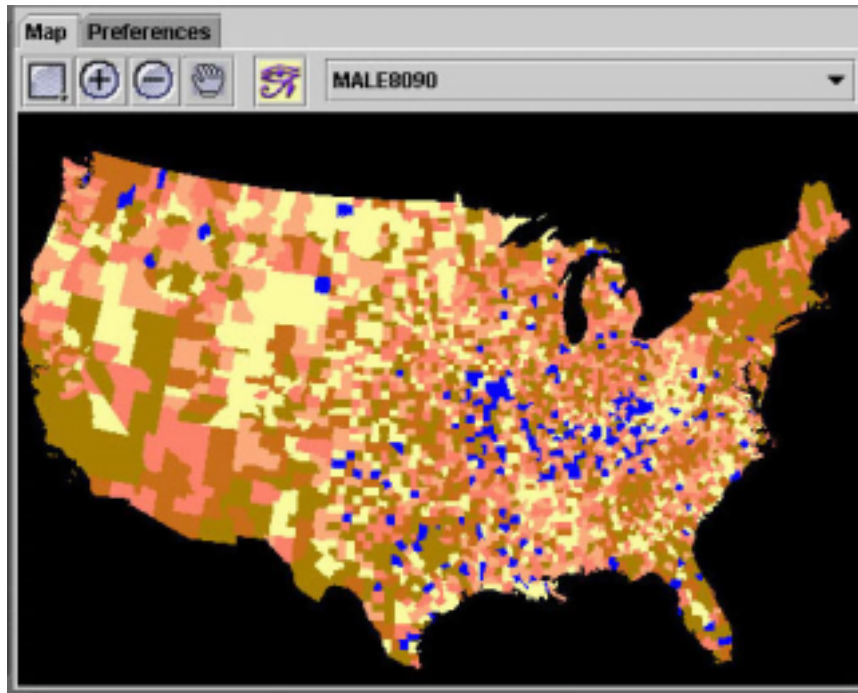


Figure 2. Selection shared by scatterplot matrix and map, and focus look at the data based on unemployed rate variable.

Beyond a matrix filled by scatter plots, our design separates the matrix from the element components in matrix. The separation of matrix and its components, which can be seen in the visual programming design box (Figure 3), enable the potential flexibility of constructing other visual exploratory tools and combining statistical analysis with geographical representations, such as maps. A short-term goal is to connect the map component with the matrix and create a map matrix (as initially suggested by Bertin (1981), which will enable multivariate geospatial analysis.

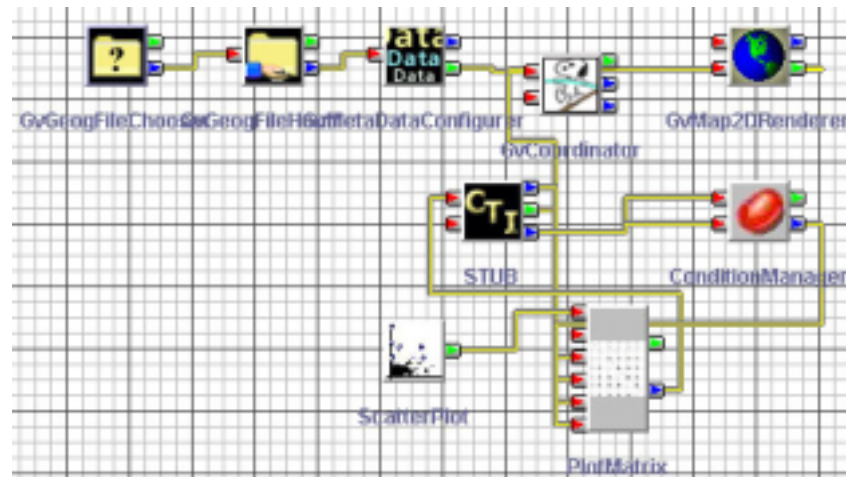


Figure 3. Design for connecting scatterplot with matrix, and connecting matrix with coordinator and conditioning components.

3. Conclusion

The user-centered and component-based matrix described here will help analysts in our eight partner agencies (and others) to work with large and increasing volume of geospatially referenced statistical data (we are coordinating directly with colleagues at the National Cancer Institute and the Census Bureau on the user-centered design aspect of the tool development discussed here). The matrix is a flexible and usable tool for multivariate geospatial data exploration and increases the potential for visual analysis of the data. In further work, we will extend the functions of the matrix and develop more exploration components, which can be displayed in the matrix. More working examples and functions are available from <http://www.geovista.psu.edu/grants/dg-qg/pubs.html>.

4. Reference

- [1] Becker, R.A. and Cleveland, W.S., 1987. Brushing Scatterplots. *Technometrics*, 29: 127-142.
- [2] Bertin, J. *Graphics and Graphic Information Processing*. Walter de Gruyter, Berlin, 1981.
- [3] Carr, D., Wallin, J.F. and Carr, A., 2000. Two new templates for epidemiology applications: linked micromap plots and conditioned choropleth maps. *Statistics in Medicine*, 19: 2521-2538.
- [4] Fua, Y., Ward, O. M., and Rundersteiner, A. E., 1999. Hierarchical parallel coordinates for exploration of large datasets. *Proceedings: Information Visualization*, 1999.
- [5] Gahegan, M., Takatsuka, M., Wheeler, M. and Hardisty, F., 2000. *GeoVISTA Studio: A Geocomputational WorkBench*, *Proceedings: GeoComputation 2000*, University of Greenwich, Medway Campus, UK, Aug. 2000.
- [6] MacEachren, A.M. et al., 2001. Supporting visual integration and analysis of geospatially-referenced statistics through web-deployable, cross-platform tools, *Proceeding, dg.o.2001*, National Conference for Digital Government Research, Los Angeles, CA, pp. 17-24.
- [7] Siirtola, H., Interaction with the Reorderable Matrix. In E. Banissi, F. Khosrowshahi, M. Sarfraz, E. Tatham, and A. Ursyn, editors, *Information Visualization IV '99*, pages 272 -277. *Proceedings International Conference on Information Visualization*, IEEE Computer Society, July 1999.