

Quality Graphics in Web-Based Applications for Summarizing Government Agency Data

Type: System demo
Paper as well?: No
Authors: Xiping Dai, Frank Hardisty and Alan MacEachren
Address: Department of Geography, 302 Walker Building
Pennsylvania State University, University Park, PA 16802-5011
814-865-9656, xpdai@psu.edu, 814-865-9656, Hardisty@geog.psu.edu
814-865-7491, MacEachren@psu.edu

High quality graphics aid federal government agencies in summarizing, analyzing, and disseminating federal statistical summaries, especially large highly complex summaries. The research efforts in Penn State Geography and the GeoVISTA Center (www.geovista.psu.edu), supported by a NSF Digital Government grant (#99883451), continue to develop web-based applications for exploring, analyzing, and presenting high dimensional geospatial data in ways that support exploration and dissemination. Progress has been achieved in the form of added visualization tools (applications and applets) in GeoVISTA *Studio*, a Java-based programming free visualization environment. GeoVISTA *Studio* is freely available to government agencies at www.geovistastudio.psu.edu, or at www.sourceforge.net. New visualization tools and functionalities added or enhanced this year include extensions to univariate and bivariate maps, a generic manipulable matrix, new parallel coordinate plot, a spreadsheet, conditioning and visual classification tools.

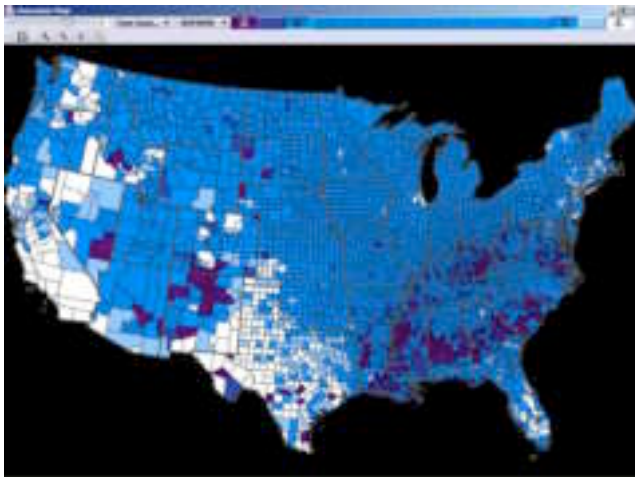


Figure 1 Map with color anchors showing change in proportion of U.S. population identifying as Hispanic, 1980 – 1990, highlighting counties with dramatic changes

Univariate and bivariate maps are designed to support exploratory analysis of geographically referenced data, such as population change data or health data. The univariate map (Figure 1) displays the spatial distribution of a demographic variable by a sequence of colors, with each color representing a class of the data set. Bivariate maps display two variables over geographic space by a bivariate color scheme. A novel color space manipulation method is used in the maps. The novel color method is a system of anchors for fixing points in color space, which can provide a flexible means of emphasizing parts of a data distribution.

A manipulable matrix is a generic component, which can accept a variety of representation forms as elements. Example elements are bivariate maps, scatter plots, a space-filling visualization, a treemap, and a histogram. The

matrix can display the same representation form on both sides of the diagonal as in a standard scatterplot matrix or can have different forms on either side as in the map and scatterplot matrix shown below (Figure 2). This is accomplished by simply connecting the matrix component with any desired representation forms in the *Studio* environment. The matrix developed allows users to select variables and generate a set of depictions, such as bivariate map and scatter plot, showing all pair wise variables. These depictions are organized as small components in the overall matrix and their positions can be directly manipulated by reordering their rows or columns position. The dynamic functions of the matrix allow the users to explore interesting combinations of variables in appropriate visualization format, and compare

different pairs of variables in maps or scatter plots or the same pair of variables in different types of visual tools. Using the example of a map and scatterplot matrix (Figure 2), the region above the diagonal displays a standard scatterplot matrix, in which the Y-axis is determined by the row and X-axis by the column that the scatterplot is in. The region below the diagonal contains a map matrix, in which position is geographic. Colors are assigned by the corresponding variables on the diagonal in the matrix.

The current Parallel coordinate plot (PCP) component is adapted from an open source component called Parvis. The PCP supports display of multiple variables at the same time by mapping an n-dimensional dataset to a two-dimensional space where variables are listed as parallel axes, and each observation is visualized as a polyline, connecting the points on axes, which are the observation's values on those axes.

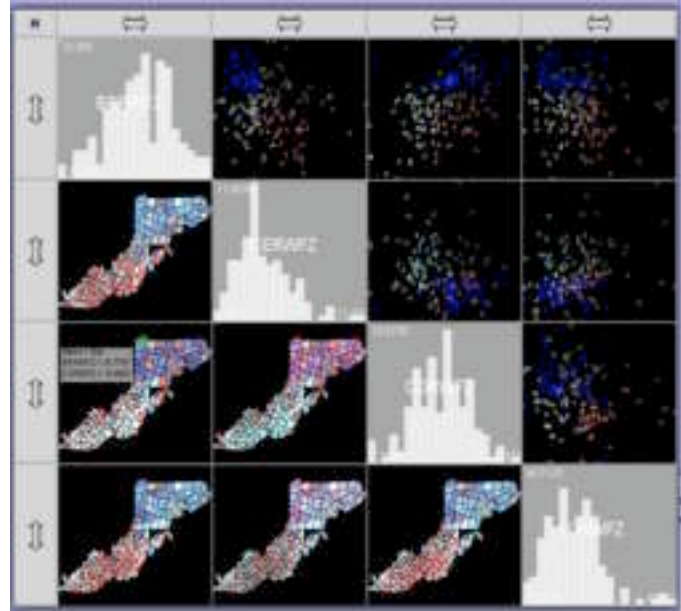


Figure 2 Map and scatterplot matrix of cervical, breast, and colorectal cancer data for white women, 1970-1994, in NAACCR gold states within the Appalachian Cancer Network

The *Studio* components described above are “coordinator-aware” and make use a coordinator bean to communicate with each other. The concept of dynamic coordination between independent visual exploratory components extends the traditional idea of linked brushing, and realized in this study as selection, classification and focusing and enables a dynamic and comprehensive analysis of geospatial statistical data. For example, if a user is examining a set of variables using a scatterplot and map matrix, selecting a set of observations, with high values in one variable (breast cancer), in a scatterplot, the locations of selections will be highlighted in all of the windows in the matrix, and spatial patterns of these observations can be depicted in map elements (Figure 2). Another example is that the PCP and map can link through the coordinator bean to share selection, classification and color symbolization (Figure 3). This allows the analyst to see relationships between geographic patterns and attribute patterns, across multiple attributes. In the map (Figure 3) below, some counties in north east Pennsylvania have been selected. These counties are then highlighted in the PCP, which depicts the multivariate signature in attribute space of these counties.

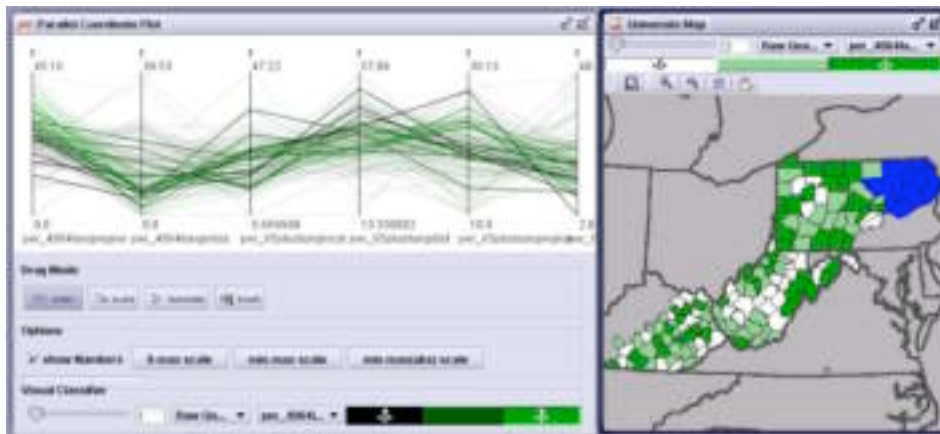


Figure 3 Parallel coordinate plot and linked map showing diagnosis stage for lung cancer in for counties in NAACCR gold states within the Appalachian Cancer Network

As noted above, a separate coordinator component has been developed for flexibly coordinating software components in any way, and providing a user interface for coordination. The software automatically discovers what types of events (actions), such as selection and classification, each component is trying to broadcast and to receive. It then registers the components with each other. A graphical user interface allows the user to customize how the components are connected without doing any programming.

Conditioning and visual classifier components have been developed to support data exploration and analysis. Conditioning is used to limit visual displays to data elements for which the value for the conditioning variable fits criteria specified through a conditioning dialogue. The visual classifier is designed to allow users to directly manipulate the classification methods and color scheme used in displaying data via various visual components. The conditioning and classifier, through the coordinator bean, make coordinated components share the conditioning and classification information, including classification types and color symbolization.

The web-based tools developed in *Studio* for the Digital Government program are designed for users with varied skill levels. Analysts in agencies who are interested in visual exploration of geospatial data can use the various visualization components and exploratory tools in *Studio*. Web start applications are distributed freely online for customized usage. Some of our tools are available to general public or users with no experience, through online as Java Applets, which can be found at www.geovista.psu.edu. A wide range of applications and applets developed with GeoVISTA *Studio* will be demonstrated at the System Demonstration session during the dg.o2003 conference.