

# Interactive Graphical Displays on the Web

Webster West

*University of South Carolina, Department of Statistics*

*Columbia, SC USA*

*west@stat.sc.edu*

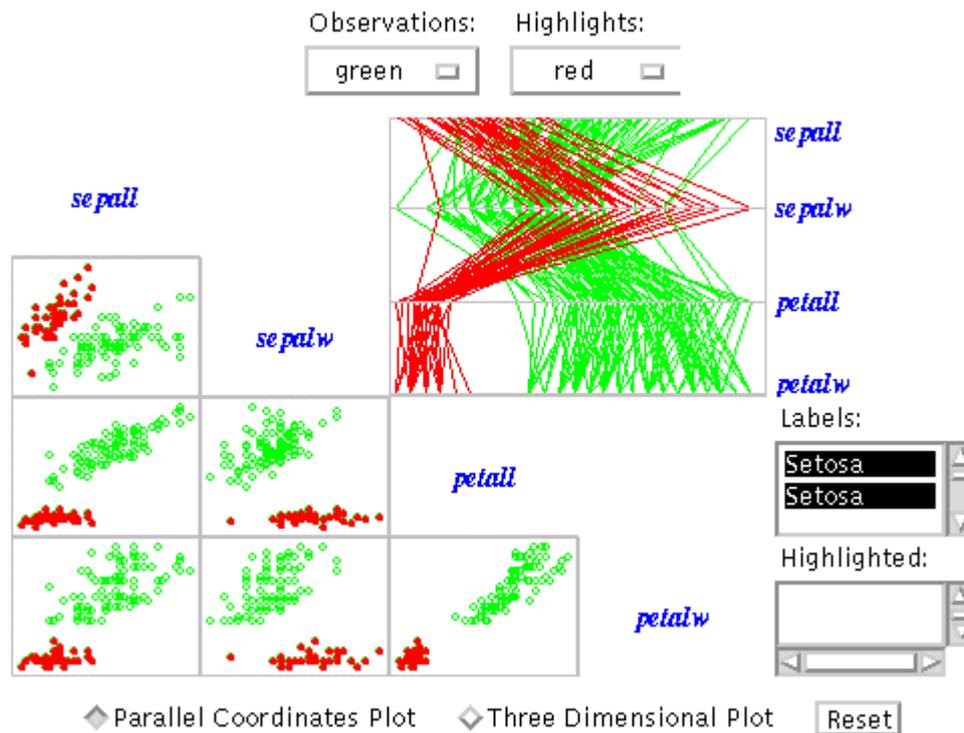
## 1. Introduction

The Internet has become a primary source for collecting and sharing data, but there are a limited number of web-based tools for analyzing data online. West and Ogden (1997) released WebStat (<http://www.stat.sc.edu/webstat/>) as a freely available basic analysis tool that runs as a Java applet within the context of a Web browser. There are also a number of commercial products developed such as Statlets (<http://www.sgcorp.com/>) and StatServer (<http://www.insightful.com/>) that have been developed. These tools generally offer standard tabular and graphical analysis options. The development of interactive graphics in a web-based environment presents a new set of technical problems.

## 2. A client side approach

Bradley and West (2001) presented a full suite of Java applets for interactive graphical analysis. These applets include interactive scatter plots in two and three dimensions, an interactive parallel coordinate plot, an interactive matrix of scatter plots and a graphical tool that allows the user to combine these components as shown in Figure 1. The data displayed in Figure 1 are Fisher's iris data (1936). This data set contains five variables: type of flower (type), sepal length (sepal), sepal width (sepalw), petal length (petal), and petal width (petalw). There are 150 total observations with 50 observations each of the Setosa, Versicolor and Virginica types. The applet allows for highlighting, brushing and labeling observations as well as spinning the three-dimensional scatter plot. Interested readers may use this applet as well as other members of the suite at <http://www.stat.sc.edu/~west/bradley/>.

Since Java applets run on the client side, an applet parameter within the HTML document must specify the raw data or a URL for the data. In either case, the data must be completely loaded onto the client before it can be analyzed. The time and cost associated with downloading large data sets to the client may limit the use of Java applets for interactive graphical analysis in some settings. This process is perhaps even more of a concern in situations where data security is vital. Of course, it is also possible to develop an interface that allows users to input data directly from the client without transmission over the Web, but this restricts the user to local data sources.



*Figure 1*

### 3. A client-server approach

An alternative to loading the data on the client is to construct the graphical output on the server. For example, one might construct graphical output as JPEG image on the server and then transmit this image to the client. A plot of several thousand observations may only be a few kilobytes depending on the plot's dimensions. The problem with this approach is that the resulting image will not be interactive. In order to produce an interactive result, a mixture of server and client side tools must be used. As an example, consider an interactive scatter plot that allows users to identify observations by drawing a rectangle around selected points. When constructing the graphic on the server, one must record the pixel location of the origin other graphical parameters that allow for transforming image coordinates back to the original data scale. The image along with the parameters is passed to a Java applet that is displayed on the client. This applet displays the image and provides for the desired interactivity of drawing rectangles, but the applet does not require the raw data. After selecting observations, the user presses an identify button on the HTML page which submits a request to a server side program which delivers a new HTML document that identifies the selected observations. As part of the request, the applet provides the range of data values to be identified. An example of this type of interactive graphic for Fisher's Iris data is shown in Figure 2. For this example, the range selected by the applet is sent to PHP (<http://www.php.net/>) document on the server that then interacts with a MySQL (<http://www.mysql.com/>) database where the data is stored to determine the selected types of flowers. The identifying information is displayed in a new browser window shown in the forefront of Figure 2. The selected observations in the lower left hand portion of the scatter plot are the 50

observations of the Setosa type. Interested users may reach this example at <http://dostat.stat.sc.edu/isi2001/>.

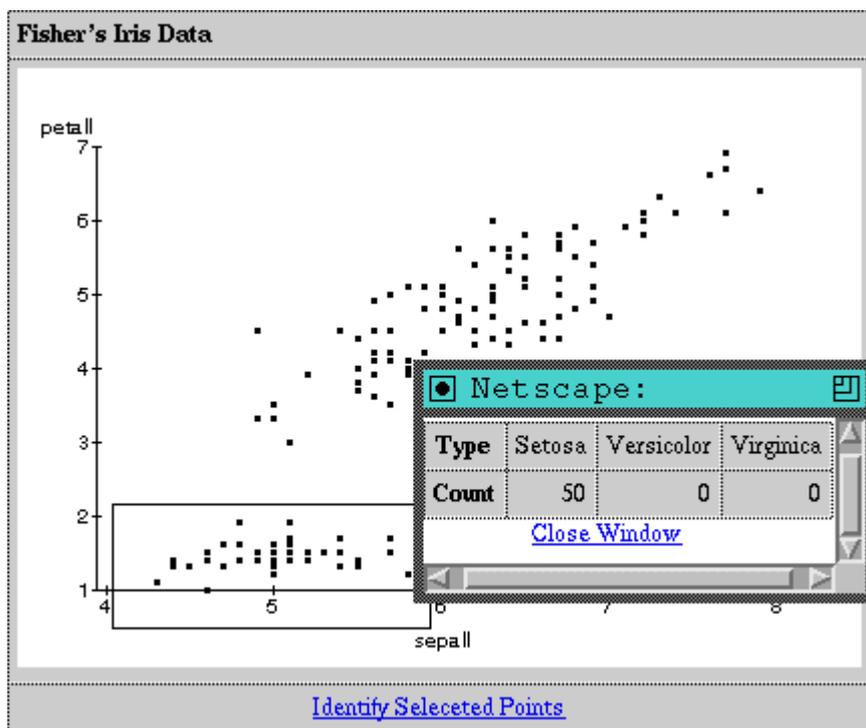


Figure 2

#### 4. Discussion

The need for Web-based interactive graphics will only increase, as the Web itself becomes more of a data warehouse. Applications cover a wide range of fields in business and research. This new medium, however, presents a unique set of challenges to the developer. The proper approach for delivering interactive graphics depends largely on the size of the data set as well as any security concerns associated with the data. For small data sets, a client side Java approach is probably desirable in most situations. For larger data sets, the client-server approach may be more suitable. More sophisticated types of interactivity such as highlighting points in connected graphs may also be possible using this approach, but this will require more complex server side programming.

#### REFERENCES

- Bradley, J. W. and West, R. W. (2001) "Interactive Java Tools for Exploring High Dimensional Data," *Journal of Statistical Software* 6:1.
- Fisher, R. A. (1936). "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, 7(II), 179-188.
- West, R. W. and Ogden, R. T. (1997) "Statistical Analysis with WebStat, a Java Applet for the World Wide Web," *Journal of Statistical Software* 2:3.