Exploratory Visualization of Correlation Matrices

Shi-Tao Yeh, GlaxoSmithKline, King of Prussia, PA.

ABSTRACT

A correlation measures how two variables are related. A correlation matrix shows a table which lists the correlation coefficients of the columns and rows. This paper discusses the techniques for displaying correlation matrices and for depicting the patterns of relationships among variables.

The purposes of the techniques described in this paper are: a) to display schematic scatter plots with ellipses, Loess, contours G3D and correlation coefficients with *p*-values, b) to create interactive correlation matrix graphs, c) to illustrate the correlation matrix iteration process with the creation of matrix maps, and d) to display multiple correlation matrices with interactive features on one page.

The SAS® version 9.1 products used in this paper are SAS BASE®, SAS/STAT®, and SAS/GRAPH® on the PC Windows platform and on the UNIX environment.

INTRODUCTION

A correlation or correlation coefficient measures the strength and direction of a linear relationship between two variables. The symbol *r* is used to stand for the correlation coefficient. The range of *r* is between [-1, 1] with 1 being a perfect correlation and 0 being no correlation. If the correlation is negative, the association is a negative relationship, and vice versa. A correlation matrix is a square table showing the correlations between all pairs of variables. The diagonal of a correlation matrix always consists of ones. A correlation matrix is always symmetrical with the values in the lower left always being a mirror of the values in the upper right.

The SAS CORR Procedure computes different coefficients for different situations. The correlation statistics provided from the CORR Procedure are: 1) Pearson product-moment correlation, 2) Spearman rank-order correlation, 3) Kendall's tau-b coefficient, 4) Hoeffding's measure of dependence, and 5) Pearson and Kendall partial correlation.

Hypothetical clinical test data has been used throughout the paper for illustration purposes.

A sample correlation matrix output from PROC CORR is shown as follows:

```
                                 The CORR Procedure

                      Pearson Correlation Coefficients, N = 217
                             Prob > |r| under H0: Rho=0

              ALAT       ASAT      ALKPH     BILDIR     BILTOT        GGT       ALBT      PLATE        BUN      CREAT

ALAT       1.00000    0.78270    0.12310    0.10301    0.07760    0.59192    0.21205   -0.09545    0.02236    0.09348
                       <.0001     0.0703     0.1304     0.2550     <.0001     0.0017     0.1612     0.7433     0.1701

ASAT       0.78270    1.00000    0.22850    0.11923    0.07624    0.61278    0.12725   -0.12570    0.05644    0.06687
           <.0001                0.0007     0.0797     0.2634     <.0001     0.0613     0.0646     0.4081     0.3268

ALKPH      0.12310    0.22850    1.00000   -0.04537   -0.08527    0.43529   -0.12384    0.15226   -0.07451   -0.05148
           0.0703     0.0007                0.5061     0.2109     <.0001     0.0686     0.0249     0.2745     0.4505

BILDIR     0.10301    0.11923   -0.04537    1.00000    0.83221    0.02678    0.12648   -0.21804    0.04703    0.00979
           0.1304     0.0797     0.5061                <.0001     0.6948     0.0629     0.0012     0.4908     0.8859

BILTOT     0.07760    0.07624   -0.08527    0.83221    1.00000   -0.00988    0.20415   -0.18226    0.02247    0.02071
           0.2550     0.2634     0.2109     <.0001                0.8849     0.0025     0.0071     0.7421     0.7617

GGT        0.59192    0.61278    0.43529    0.02678   -0.00988    1.00000   -0.01820    0.11027   -0.10104    0.02266
           <.0001     <.0001     <.0001     0.6948     0.8849                0.7898     0.1052     0.1379     0.7399

ALBT       0.21205    0.12725   -0.12384    0.12648    0.20415   -0.01820    1.00000    0.00805   -0.21580   -0.06244
           0.0017     0.0613     0.0686     0.0629     0.0025     0.7898                0.9061     0.0014     0.3600

PLATE     -0.09545   -0.12570    0.15226   -0.21804   -0.18226    0.11027    0.00805    1.00000   -0.23588   -0.24897
           0.1612     0.0646     0.0249     0.0012     0.0071     0.1052     0.9061                0.0005     0.0002

BUN        0.02236    0.05644   -0.07451    0.04703    0.02247   -0.10104   -0.21580   -0.23588    1.00000    0.50508
           0.7433     0.4081     0.2745     0.4908     0.7421     0.1379     0.0014     0.0005                <.0001

CREAT      0.09348    0.06687   -0.05148    0.00979    0.02071    0.02266   -0.06244   -0.24897    0.50508    1.00000
           0.1701     0.3268     0.4505     0.8859     0.7617     0.7399     0.3600     0.0002     <.0001
```

Table 1. Sample Correlation Matrix Output

SAS 9.1 includes an experimental feature, ODS Statistical Graphics (ODS Graphics for short), which introduces more than 30 procedures with ODS Graphics in four SAS products. This new feature enables the procedures to create statistical graphs automatically as part of the ODS output.

With simple ODS Graphics statements, a procedure creates graphs that are commonly needed for data analysis. The default ODS Graphics provides informative graphs along with important statistics. You can customize ODS Graphics to suit your needs.

The CORR procedure produces two types of ODS Graphics: scatter plot matrix and scatter plot with prediction ellipses, as part of the ODS output. Figure 1 shows a sample scatter plot matrix output from the CORR procedure. For a relatively small number of variables, the scatter plot matrix provides an excellent visual representation of the relationships among variables.
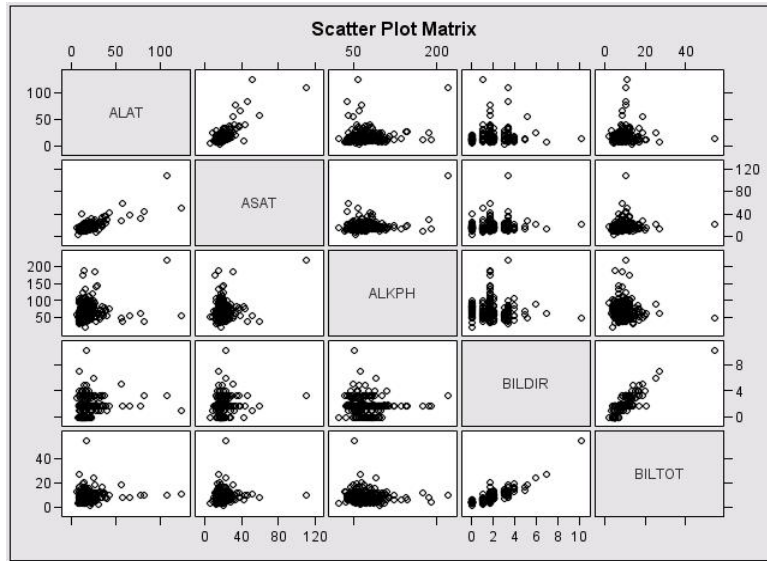
Figure 1. Sample Scatter Plot Matrix

Figure 2 shows a sample scatter plot output with ellipses from the CORR procedure. The ellipses are computed with specified confidence intervals. The magnitude of the correlation is shown in the shape of an ellipse.
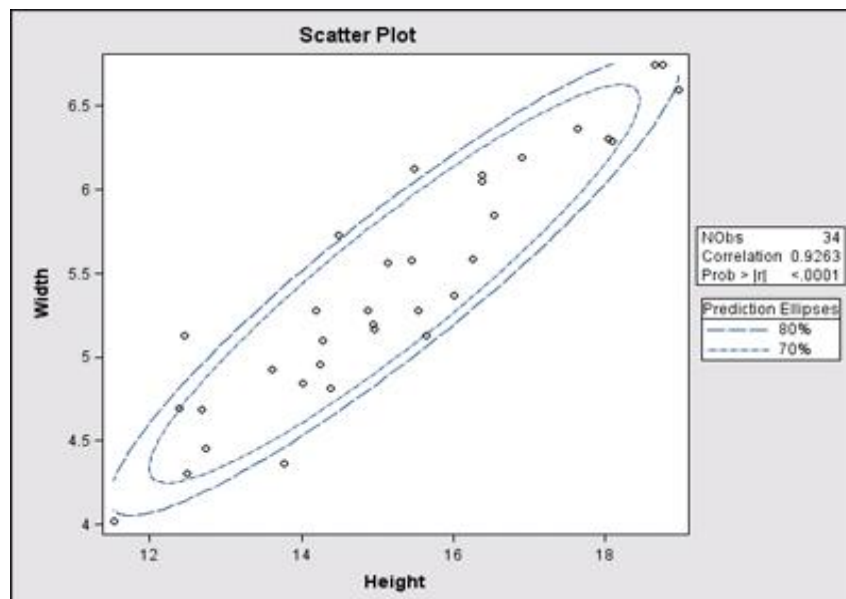

Figure 2. Sample Scatter Plot with Ellipses

The SAS statements required for creating ODS Graphics are simple and easy. The following code shows a sample statement for producing ODS Graphics.

```
ods html;
ods graphics on;

  proc corr data=lab nomiss noprint
     plots=scatter(ellipse=mean nmaxvar=2 alpha=.1 .3);
     var alat asat ;
  run;
ods graphics off;
ods html close;
run;
```

If there is a need to modify the ODS Graphics, it is difficult to accomplish this task. The appearance of the ODS Graphics is governed by the ODS Graphical Template Language (GTL). The modification of ODS Graphics can be made by three sources: 1) from the data set that is used for the procedure execution, 2) from changing the style template, and 3) from modification of the graph template. In most cases, it involves the modification of the ODS Template or the need to create a new ODS Graphics Template.

The correlation matrix graphical display techniques and designs have been developed in different application software packages and research papers. A correlation matrix can be displayed in a variety of forms. The correlation coefficient *r* has two distinctive characteristics: 1) coefficient value itself indicates the strength of the association, and 2) the sign of coefficient indicates the direction of the association. Symbols such as bars, shaded squares, ellipses, numbers, or special symbols are used to show the magnitude of the correlation value. Color is used to show the sign of correlation with blue representing the positive value and red for negative values.

The techniques and layouts described in this paper are:

a) schematic scatter plots with ellipses, Loess, univariate plots, bivariate plots and correlation coefficients with *p*-values
b) interactive correlation matrix graphs
c) visualization of correlation matrices in the iteration process
d) multiple correlation matrices displays with interactive features

Some of the specific ideas, techniques and layouts have been suggested before, and some are novel.

The main contributions of this paper are to:
- introduce and visualize the correlation matrices with interactive features
- illustrate the correlation matrices in the iteration process
- display the new design of display layouts

The SAS version 9.1 products used in this paper are SAS BASE®, SAS/STAT®, and SAS/GRAPH® on the PC Windows® platform and on the UNIX environment.

I. SCATTER PLOT WITH ELLIPSES AND LOESS

Figures 1 and 2 display the scatter plot matrices and ellipses separately. This section describes several graphical display designs which combine the scatter plot matrices containing ellipses with Loess and other bivariate plots (such as contours and G3D) together.

3

The features of this display are:
1) the values in diagonal cells are always 1,
2) the diagonal cells are used for variable labels instead of 1,
3) the diagonal cells can be use to display other univariate statistics and univariate plots, such as histogram, cumulative density function, and kernel density function,
4) scatter plots are displayed in each off-diagonal cell with a 70% confidence ellipse in each off-diagonal cell,
5) a local regression line (Loess) is added to each lower off-diagonal cell,
6) a pair of correlation coefficients with p-value,
7) the bivariate contours or G3D are displayed in the upper off-diagonal cells.

Color is used to show the sign of correlation in data scatter plots with blue representing the positive value and red the negative value. The slope of Loess also indicates the sign of correlation with a positive slope for a positive correlation and a negative slope for a negative correlation. Table 1 provides the Pearson Correlation Coefficients and the p-value under the null hypothesis of a zero correlation. The color of the Loess line is used to show the significant p-value, with red being significant, and black not significant.

Figure 3 shows the first display layout design for a schematic scatter plot matrix with ellipses and Loess.
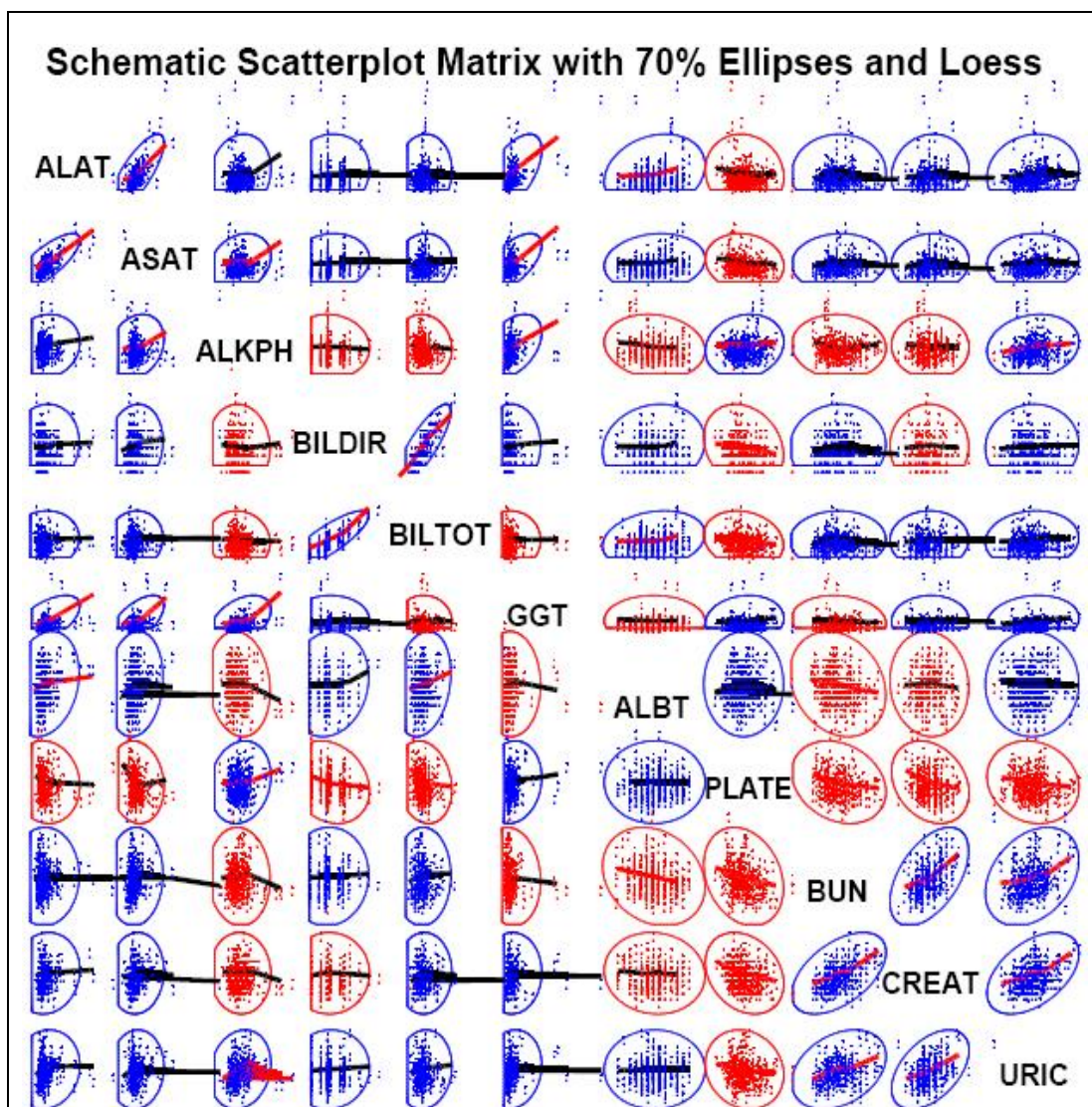

Figure 3. Schematic Scatter Plot with 70% Ellipses and Loess

The second design rearranges the matrix cell display. Since the correlation matrix is always a symmetrical matrix, the scatter plots, ellipses and Loess are only displayed in the lower triangular section of the matrix. The correlation coefficients and p-values are displayed in the upper portion of the matrix. Colors are used to represent the correlation within the scatter plots and *r* displays. The diagonal cells are displayed with histograms with sample size N and the mean value.

This design provides an informative presentation with a bivariate scatter plot, ellipse shapes and Loess for correlation magnitude, actual correlation *r* and *p*-values. It also provides a univariate histogram with the kernel density function, univariate statistics of sample size N, and the MEAN value in each diagonal cell.

Figure 4 shows a sample output of this layout. The slope of the Loess indicates the direction of the association and the ellipse shape reflects the magnitude of the association. The color of the Loess indicates whether the *p*-value is significant or not: red for significant and black for not. When the ellipse is shaped more like a circle, the association is weak, and the correlation coefficient is low.

Figure 5 is another design to replace coefficient numbers with bivariate contour lines. The procedure KDE is used to compute an estimation of density levels / percentiles and then to store the information in a SAS output data set. This output data set then is used by the SAS/GRAPH GCONTOUR procedure to produce the contours.

Correlation ordering is a technique to detect patterns of relations, trends, and anomalies. Figure 5 shows correlation ordering by the first eigenvector from Principal Component Analysis. The correlation ordering will be discussed in detail in the next section.
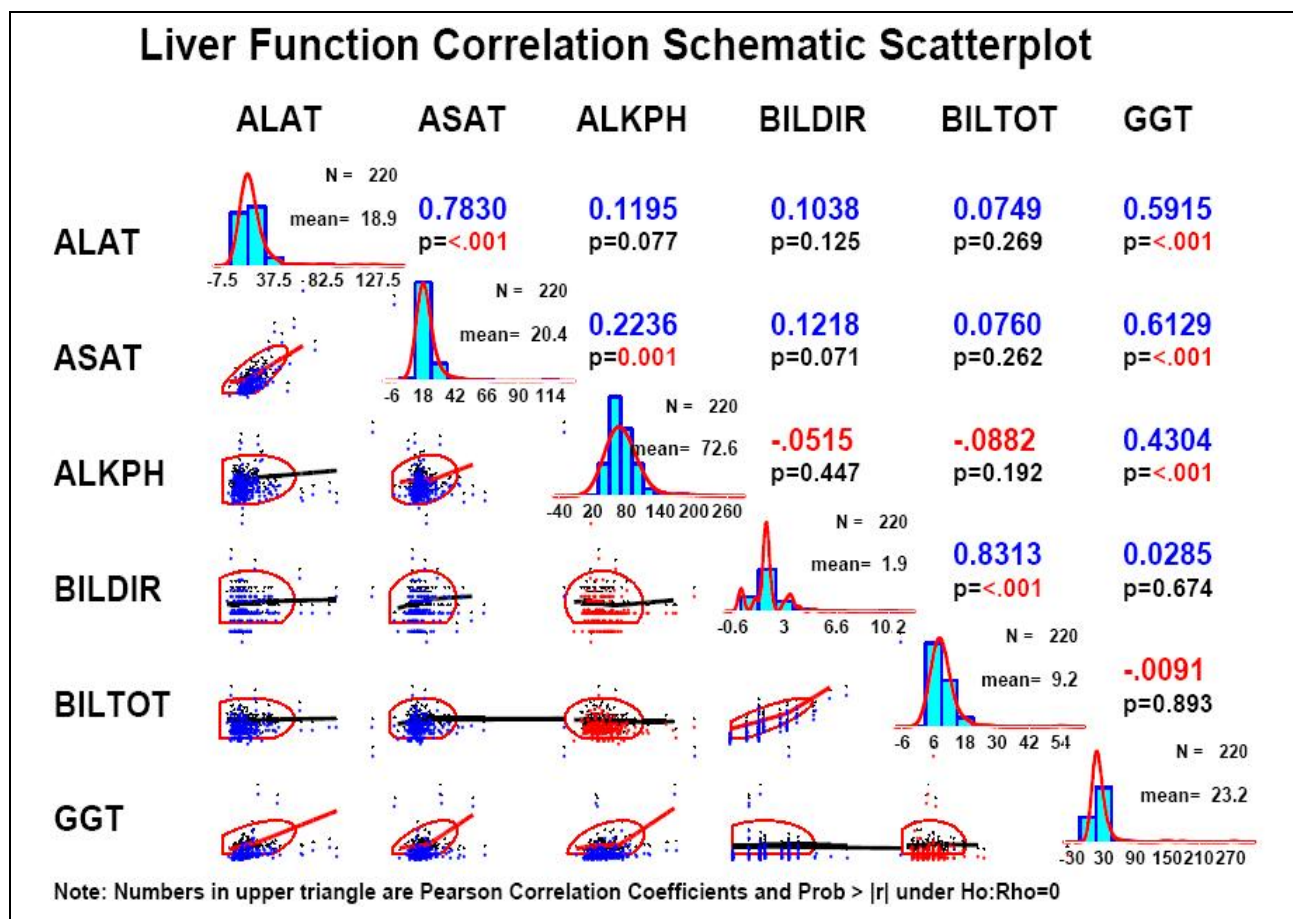

Figure 4.   Schematic Scatter Plot with 70% Ellipses, Loess, Histograms and Correlation Coefficients
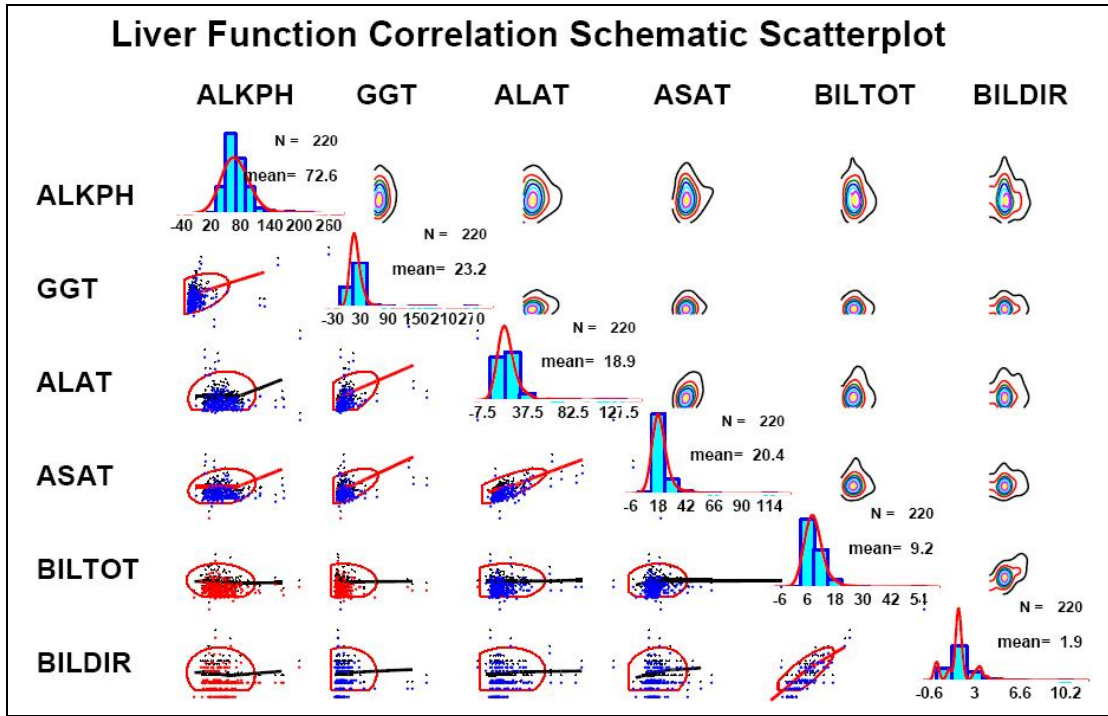
Figure 5. Sample of Schematic Scatter Plot with Contours

The other design is to replace contours by G3D plots. Again, the procedure KDE is used to compute the estimation of density levels / percentiles and store the information in a SAS output data set. This output data set then is used by the SAS/GRAPH G3D procedure to produce the G3D plots. Figure 6 shows the schematic scatter plots with G3D plots.
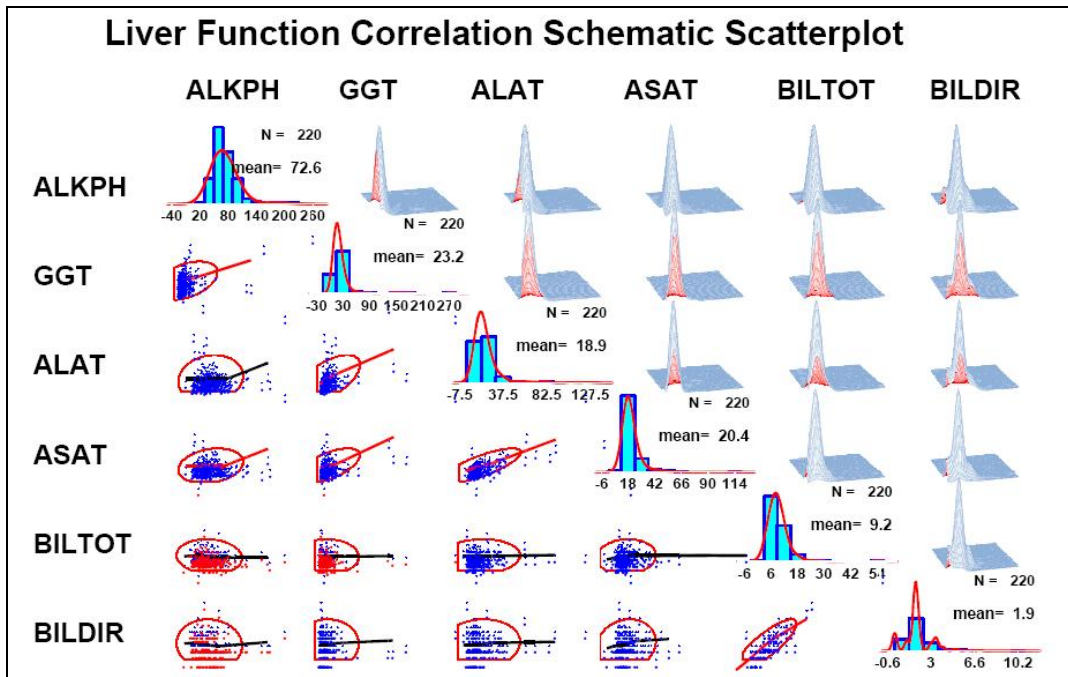

Figure 6. Sample of Schematic Scatter Plot with G3D Plots

In a clinical trial setting, the study protocol design usually involves multiple treatments for comparison. The next design overlays two treatment scatter plots on one matrix plot cell. Each treatment has its own scatter plot, ellipse and Loess. Blue represents treatments with Dose A and red for Dose B, respectively.

6

The cell in the upper triangular section of the matrix plot displays two pairs of correlation coefficients with the *p*-values, one pair for each treatment. The cumulative density function (CDF) and kernel density function are overlayed and displayed in the diagonal cells. Different colors of CDF and kernel density functions are used for different treatments. The purpose of this design is to overlay two treatments in one cell for easier comparison.
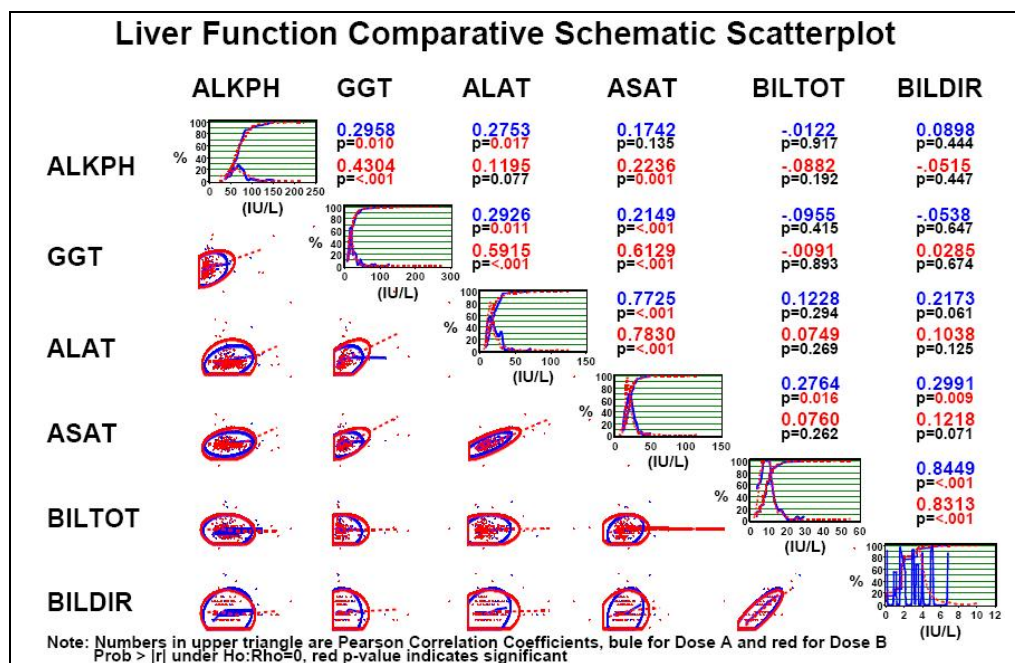

Figure 7. Sample of Comparative Schematic Scatter Plot with Overlaying 2 Treatments

II. INTERACTIVE CORRELATION MATRIX GRAPH

The interactive graphs allow for user (human) interactions. The users interact directly with the graphs, mainly through the user interfaces. The user interfaces are controlled with a mouse as well as onscreen buttons and menus.

The interactive graphs provide the following desirable features. The user can:
1) change graphical properties, including modification of titles and axes through the interactive menus,
2) use the data tips feature to include extra information on the graph,
3) link to subsequent graphs, maps or tables,
4) change the scale by zooming techniques. Zoomable interfaces come in two types:
   a) Geometric zooming where the scale is linear with the multiplier
   b) Semantic zooming where objects may change shape at different zoom levels
   Semantic zooming is a distortion technique that displays the object in a fisheye view. By moving around and changing the scale, the user can customize the data display.
5) produce slide show of static images

The SAS System implements several new technologies such as Java applets and ActiveX Control to produce the interactive graphs. SAS Graphics macro DS2CONST is selected for this application. SAS macro DS2CONST uses the Constellation Applet to produce a node-link constellation diagram. Each node can be linked to one or more nodes. Unlike the Treeview Applet, the Constellation Applet does not require a hierarchical relationship between the nodes. Although it can be used to display hierarchical relationships, the Constellation Applet does not automatically place the root node at the center of the display.

The Constellation Applet supports the following interactive features. It allows:

1) zooming/panning or fish-eye distortion on the portion of the diagram that is in the center of the display,
2) an embedded scroll bar to subset the links and nodes,
3) click and drag the diagram to change the portion of the diagram,
4) a pop-up menu that has several functions, such as highlighting specific links and searching for specific nodes.
5) pop-up data tips for links and nodes.
6) hotspot links to Web sites or files.

The design of the correlation matrix display focuses on the following features:
1) It converts the table to a graphical display with blank diagonal cells,
2) The size of the bar symbol in each cell indicates the magnitude of the association,
3) Color is used to encode the sign of the correlation - blue for positive and red for negative. Scaled color is also used to show the magnitude of the correlation. The more intense (darker) the color is, the higher correlation value.

The interactive features for this application are: 1) It provides variable labels as pop-up data tips for variable nodes, 2) It also provides *r* as pop-up data tips for bars (links), 3) The embedded scroll bar allows the subset of nodes and links. Figure 8 shows the constellation display of the correlation matrix.

Figure 8. Constellation Display of Correlation Matrix
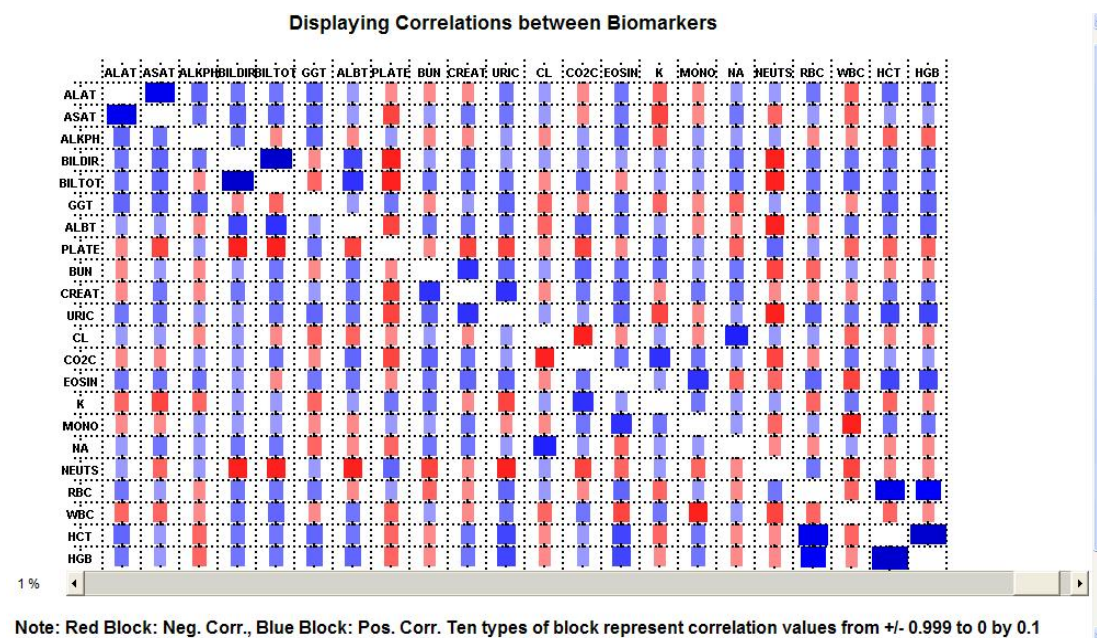
The pop-up data tip is an interactive feature that is enabled when the user moves the mouse curser over a graphical element triggering a text box to display. You can decide on the content to be shown inside the text box. In this application, the correlation table cells are converted to link bar text boxes, and variable labels have been converted to node text boxes.
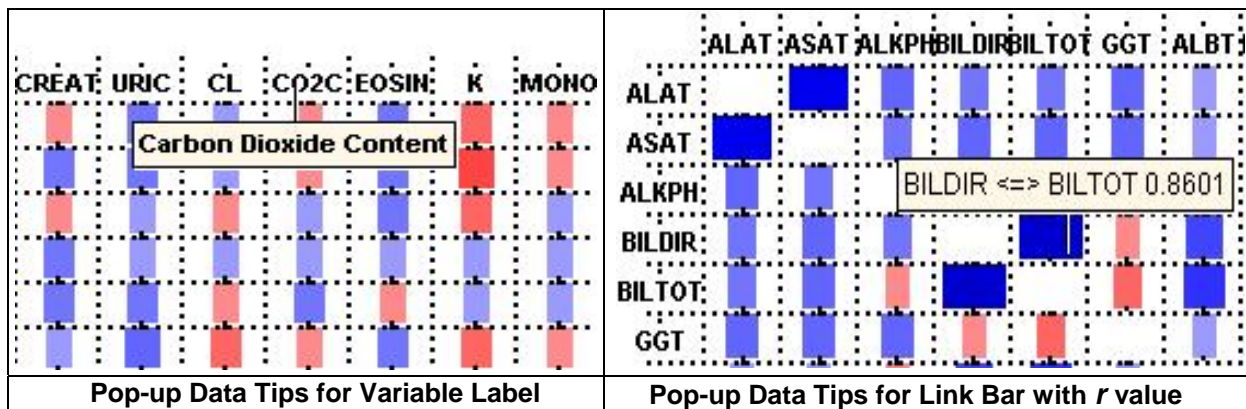
| Pop-up Data Tips for Variable Label | Pop-up Data Tips for Link Bar with *r* value |

Figure 9. Illustration of Interactive Pop-up Data Tips

The other useful interactive feature is the embedded scroll bar for subsetting nodes and links. In this application, the absolute correlation values are grouped by increments of 0.1. When the viewer slides the scroll bar from right to left, the smallest absolute value groups will disappear from the display. This data reduction feature allows the viewer to focus the associations that are significant. Figure 10 shows the process of data reduction by sliding the scroll bar from right to left.
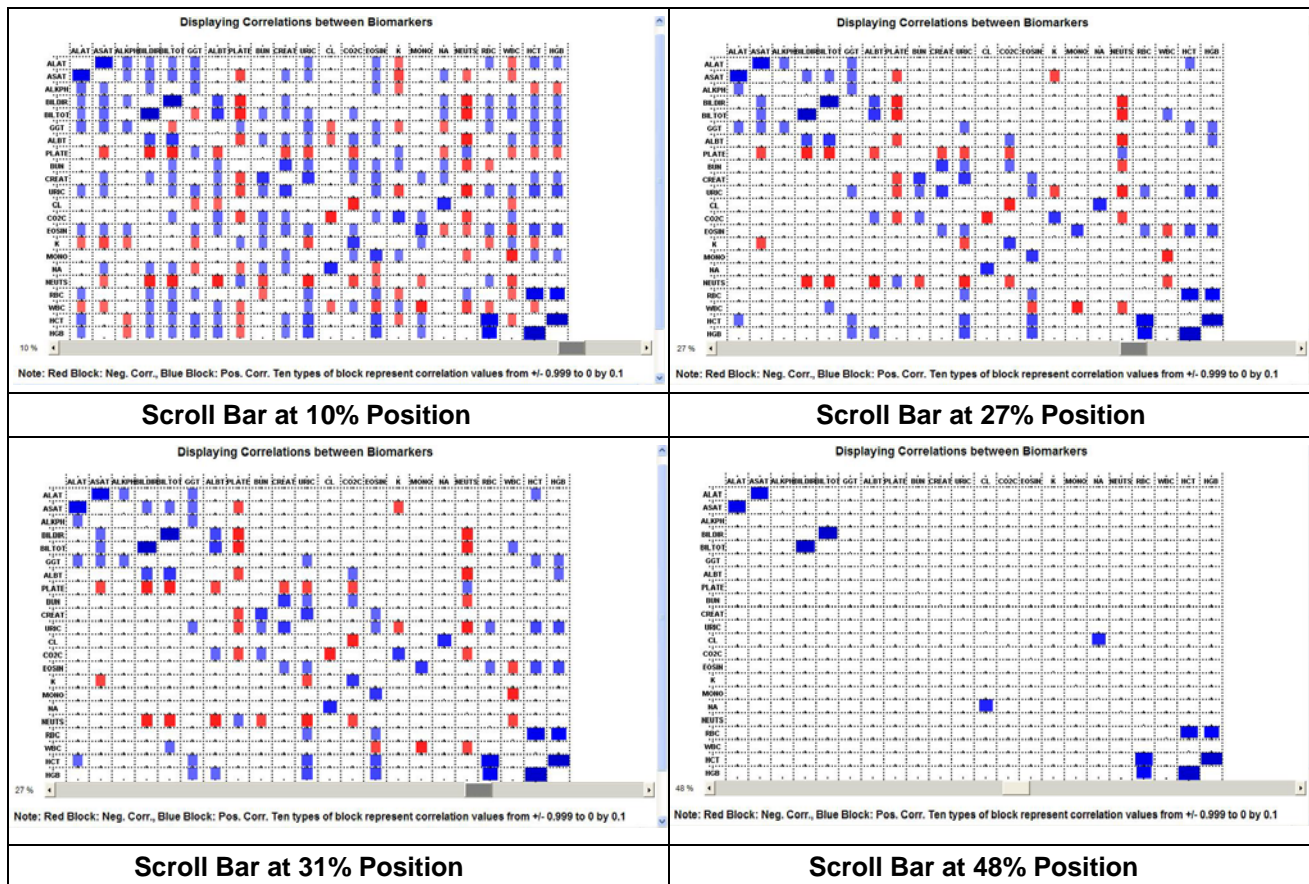

| Scroll Bar at 10% Position | Scroll Bar at 27% Position |
| Scroll Bar at 31% Position | Scroll Bar at 48% Position |

Figure 10. Using Scroll Bar for Subsetting Nodes and Links

The unordered data usually shows a random pattern. Good data ordering can rearrange the display, and relational patterns can be revealed. In addition, it will be easier to detect patterns of relationships, similarities, trends, and data irregularities. The purpose of correlation ordering is to arrange the variables with similarities in a way that simplifies the pattern of relations among variables. Friendly and Kwan (2002) proposed a general principle called "effect-ordered data display" which states with data display

(table or graph), unordered factors or variables should be ordered according to what we wish to show or see.

The other ordering principle is to place the variables onto the eigen-space for optimal unidimensional order or multiple dimensional order. This matrix iteration or seriation technique is based on the decomposition of a matrix to eigenvectors and eigenvalues of the correlation matrix **R**. This iteration process was first introduced by McQuitty (1968). The matrix iteration process is stated as follows:

Given a proximity *pxp* matrix **D**, a sequence of simple Pearson correlation matrices is iteratively generated from **D**,

$$\mathbf{R} = (\ \mathbf{R}^{(1)}, \mathbf{R}^{(2)}, \ldots)\quad where \quad \mathbf{R}^{(n)} = \Phi\,(\mathbf{R}^{(n-1)}),\ \ n > 1 \quad and \quad \mathbf{R}^{(1)} = \Phi\,(\mathbf{D}).$$

The sequence of iterated correlation matrices will converge to a limiting matrix $\mathbf{R}^{(\infty)}$ in which all elements are 1 or -1. This limiting matrix $\mathbf{R}^{(\infty)}$ is in a convergence stage and partitions the p objects into two disjoint groups.
In mathematics, convergence describes limit behavior, particularly of an infinite sequence or series toward some limit.

This theorem has two components. 1) the Pearson correlation matrix is used as entry for the iteration process, and 2) the iteration process will reach a convergence stage with a limiting matrix of $\mathbf{R}^{(\infty)}$. Exploratory visualization is a good approach to understand such a process. Figure 8 is a representation of a Pearson correlation matrix. It is used as an entry for the iteration process. Figure 11 shows $\mathbf{R}^{(1)}$ to $\mathbf{R}^{(5)}$ from left to right in the first row, and $\mathbf{R}^{(6)}$ to $\mathbf{R}^{(10)}$ from left to right in the second row. It takes 10 iterations to reach the convergence stage. The limiting matrix $\mathbf{R}^{(10)}$ contains solid colors of blue and red with values of +1 and -1 in the matrix cells. $\mathbf{R}^{(10)}$ is also a representation of a correlation matrix map. The seriation cuts the correlation map into several blocks. Blocks for proximity matrices could be categorized as within-group blocks on the map diagonal and between-group blocks out of the diagonal.
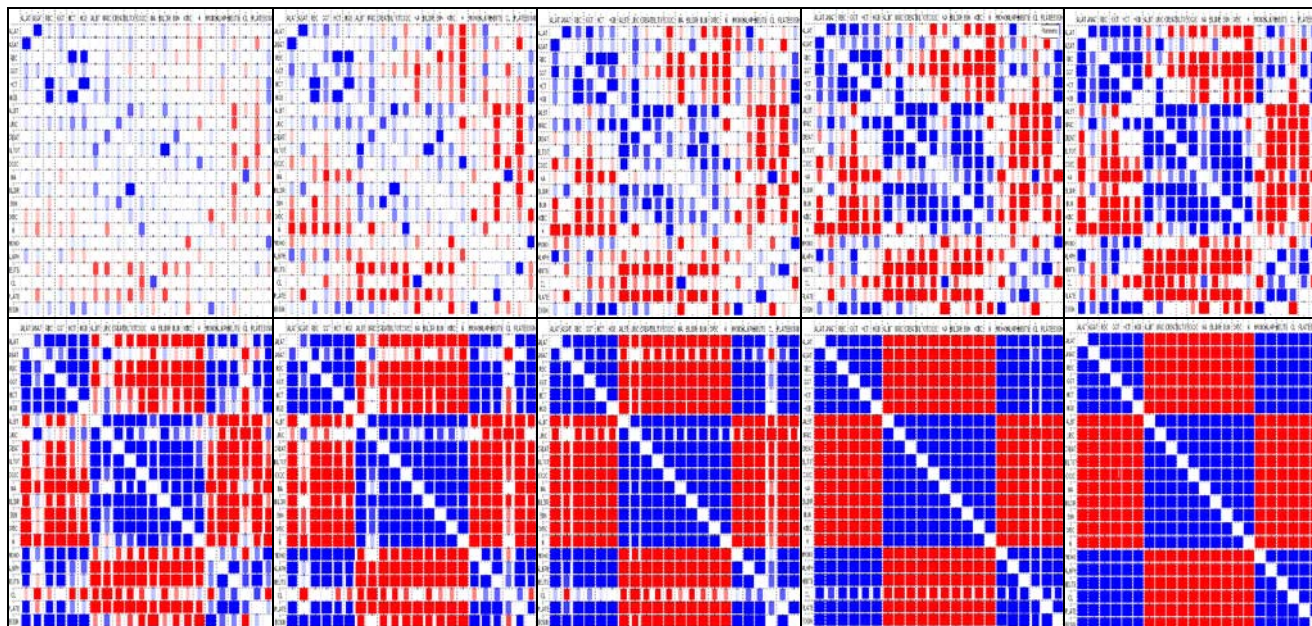

Figure 11. Pearson Correlation Matrix Iteration Process

We can use hierarchical clustering to observe the seriation and variables. The SAS procedure DISTANCE is used to produce a proximity matrix. The procedures CLUSTER and TREE are used to produce the hierarchical clustering tree. Table 2 shows a sample output of the proximity matrix. Figure 12 shows the proximity matrix with clustering after reordering.

NAME	ALAT	ASAT	ALKPH	BILDIR	BILTOT	GGT	ALBT	PLATE	BUN	CREAT	URIC	CL	CO2C
ALAT	0.00000	.	.	.	.	.	.	.	.	.	.	.	.
ASAT	0.47692	0.00000	.	.	.	.	.	.	.	.	.	.	.
ALKPH	0.84386	0.90873	0.00000	.	.	.	.	.	.	.	.	.	.
BILDIR	0.87514	0.83721	0.92964	0.00000	.	.	.	.	.	.	.	.	.
BILTOT	0.92726	0.85063	0.98485	0.38415	0.00000	.	.	.	.	.	.	.	.
GGT	0.84108	0.88607	0.83914	1.02654	1.04664	0.00000	.	.	.	.	.	.	.
ALBT	0.96722	0.95989	1.01805	0.81142	0.75397	1.00196	0.00000	.	.	.	.	.	.
PLATE	1.01503	1.10297	0.99202	1.16185	1.18119	0.91663	1.14156	0.00000	.	.	.	.	.
BUN	1.02363	0.96569	1.00360	0.99424	0.91105	1.01769	0.90696	1.04914	0.00000	.	.	.	.
CREAT	1.01553	0.93825	0.99289	0.93774	0.88850	0.99733	0.92845	1.13020	0.77112	0.00000	.	.	.
URIC	0.93873	0.92009	0.98778	1.01187	0.96145	0.88881	0.91576	1.11426	0.84499	0.78036	0.00000	.	.
CL	1.00805	0.99156	1.03522	1.01092	1.04855	1.08925	1.08473	1.02191	1.00549	1.01384	0.98175	0.00000	.
CO2C	0.99782	1.00074	0.98646	0.94814	0.93992	1.01875	0.87998	1.09686	0.88425	0.92004	0.99661	1.17188	0.00000

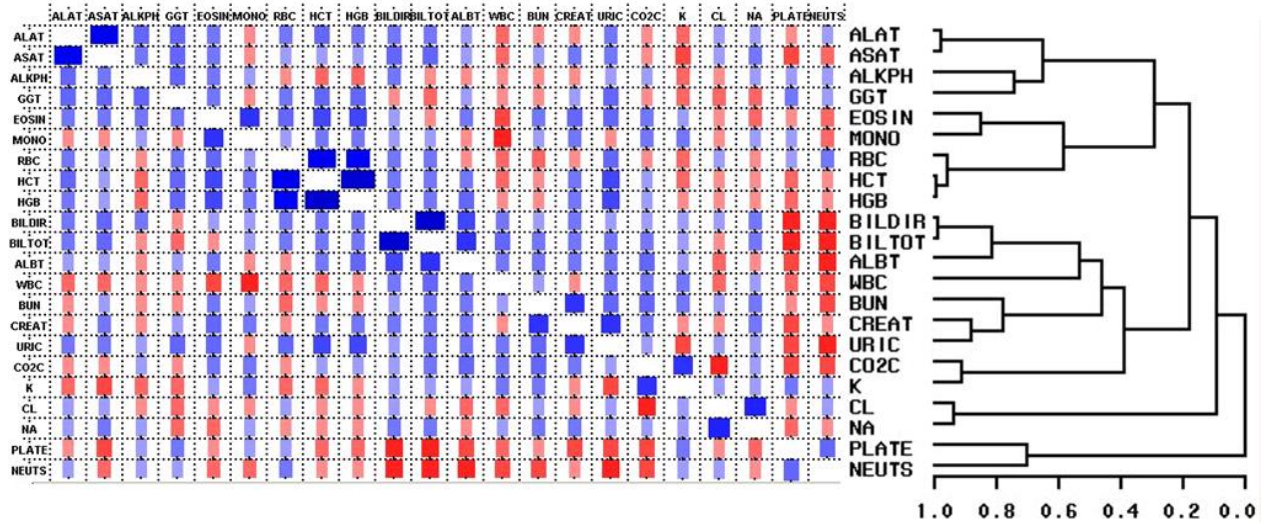**Table 2. A Sample Proximity Matrix Output from PROC DISTANCE**

Figure 12. Pearson Correlation Matrix with Hierarchical Clustering Tree

The same iteration process can be observed in an eigen-space. In a visualization context, convergent series provides the visual plots for observing the iteration process. When a series reaches a convergence stage, the current matrix plot is identical to the next sequence. The plot of a convergent matrix is also a limiting matrix plot.

Figure 13 shows the Pearson correlation matrix iteration process where the eigenvectors are plotted in an eigen-space. The sequence reaches a convergent stage at the 9th iteration.
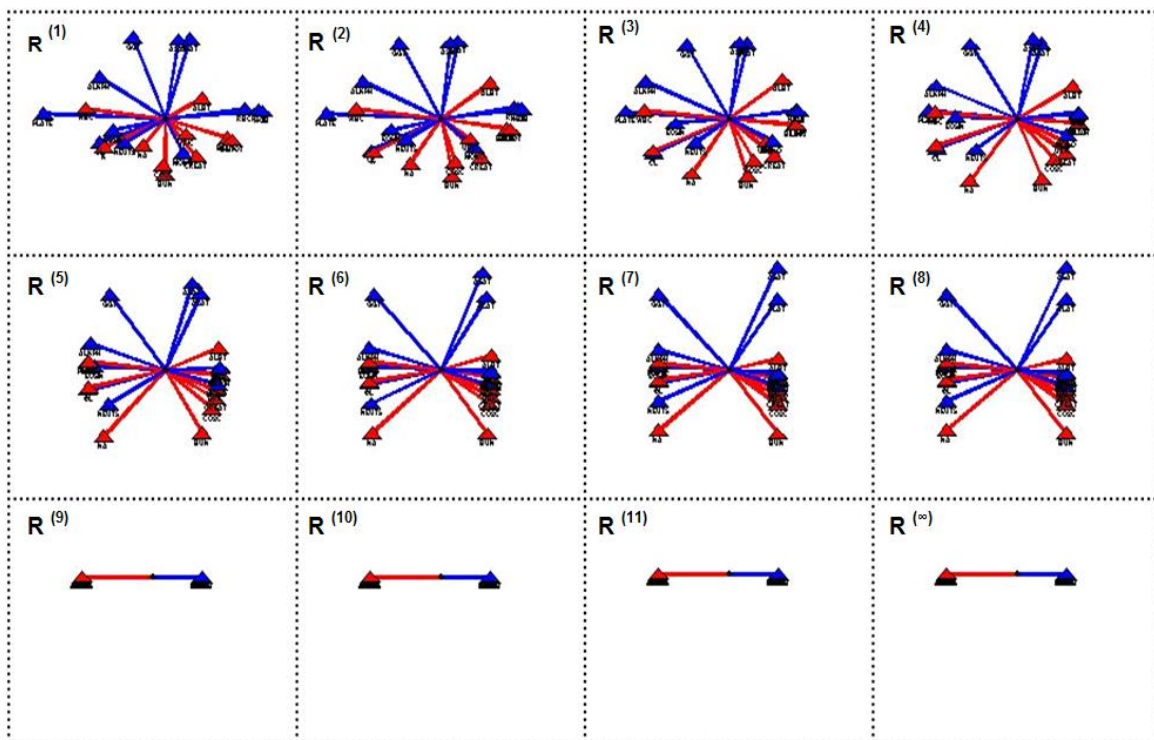
Figure 13. Pearson Correlation Matrix Converging Sequence in a Eigen-Space

We have used dozens of different Pearson correlation matrices as entries for this matrix iteration process. In our experiments, all iterations can reach the convergence stage in 9 to 12 iterations. What happens to other correlation matrices as entries for the iteration process? Several Spearman rank-order, Kendall's Tau-b and Hoeffding's coefficient matrices from PROC CORR are used as entries for these experiments. Figures 13, 14, 16 and 17 are produced by calling the SAS/GRAPH DS2CONST macro. Figure 14 uses the Spearman correlation matrix as an entry matrix for the iteration process.
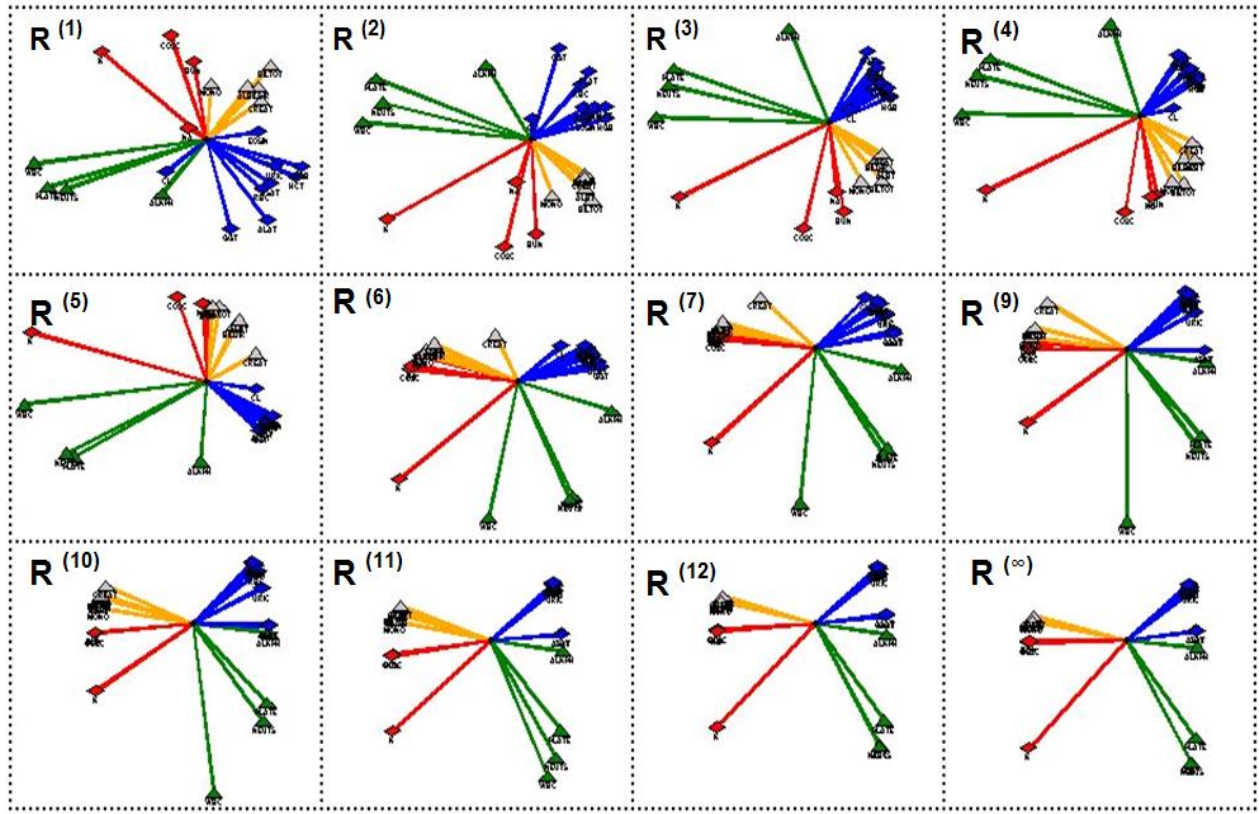
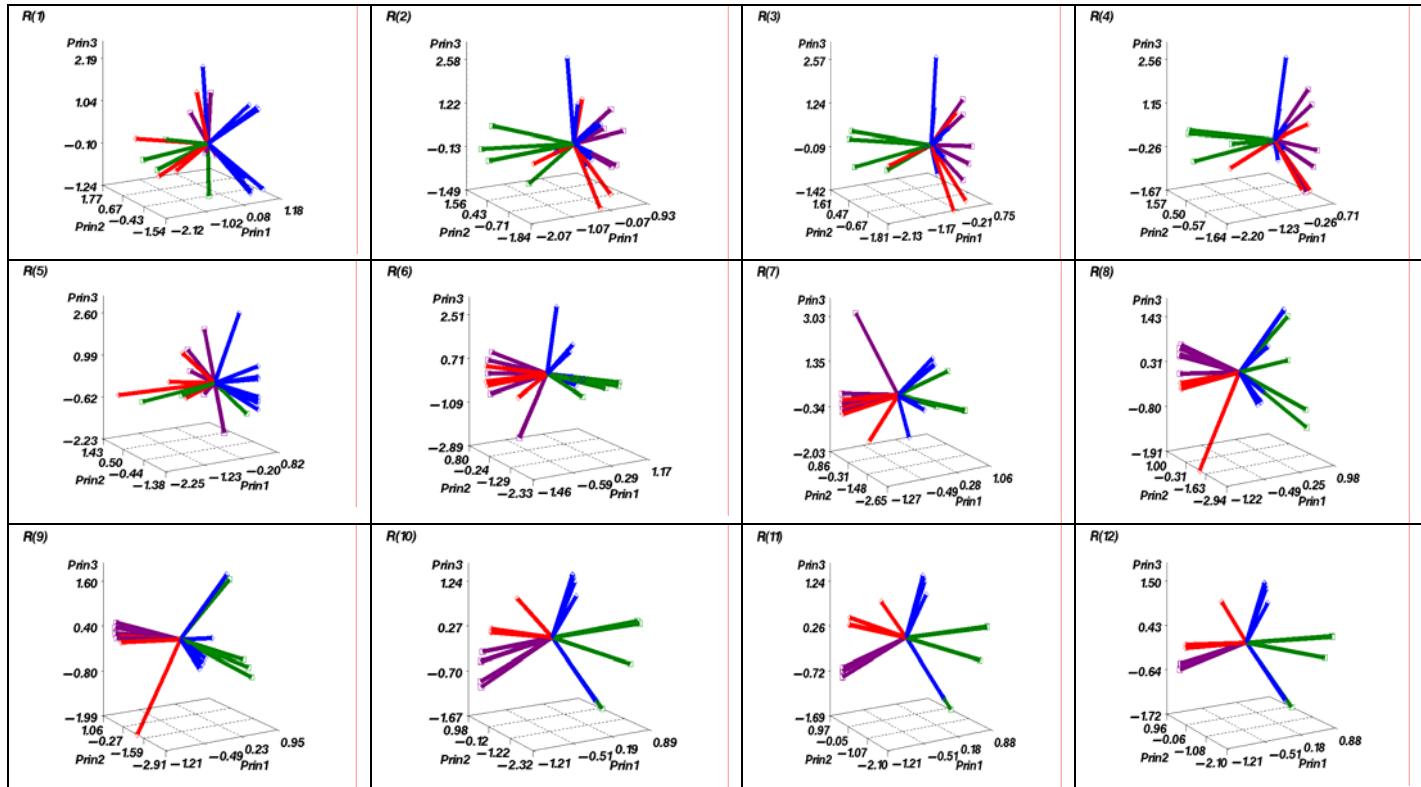**Figure 14. Spearman Correlation Matrix Converging Sequence in a Two-Dimensional Eigen-Space**

Figure 15. Spearman Correlation Matrix Converging Sequence in a 3-Dimensional Eigen-Space

13

Figure 16 uses the Kendall's Tau-b correlation matrix as an entry matrix for the iteration process. The iteration process reaches a convergence stage at $\mathbf{R}^{(12)}$.
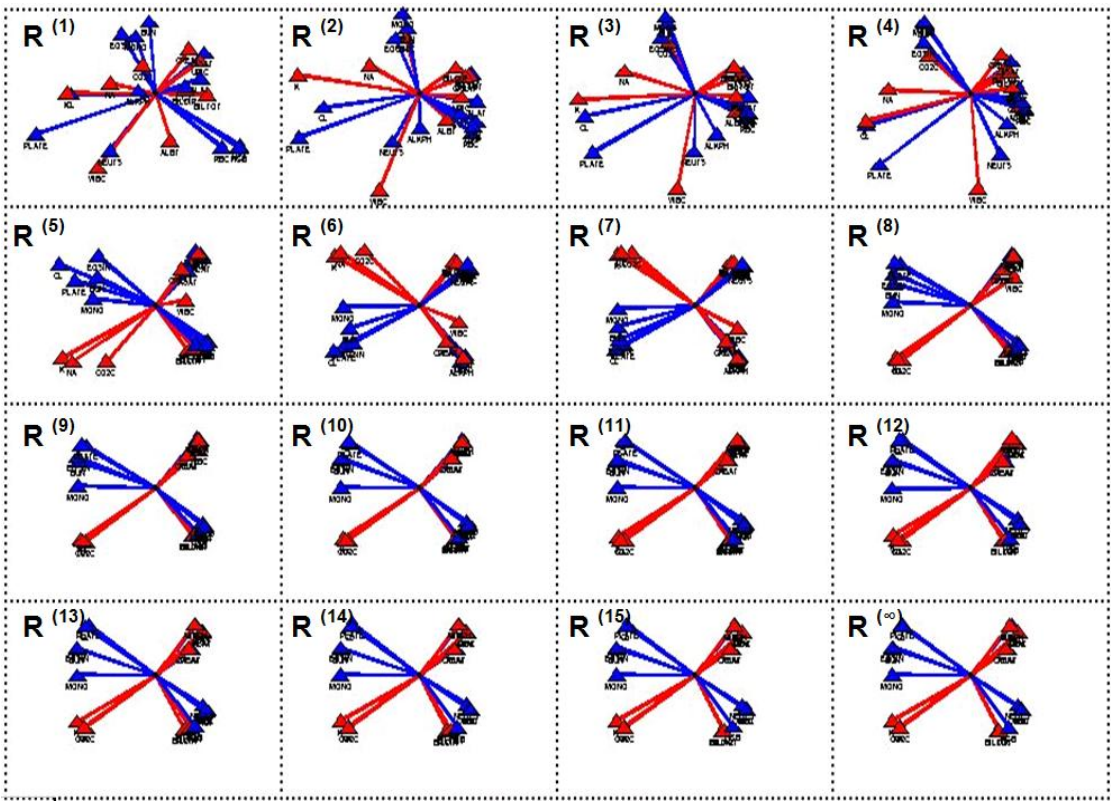

Figure 16. Kendall's Tau-b Matrix Converging Sequence in a Two-Dimensional Eigen-Space
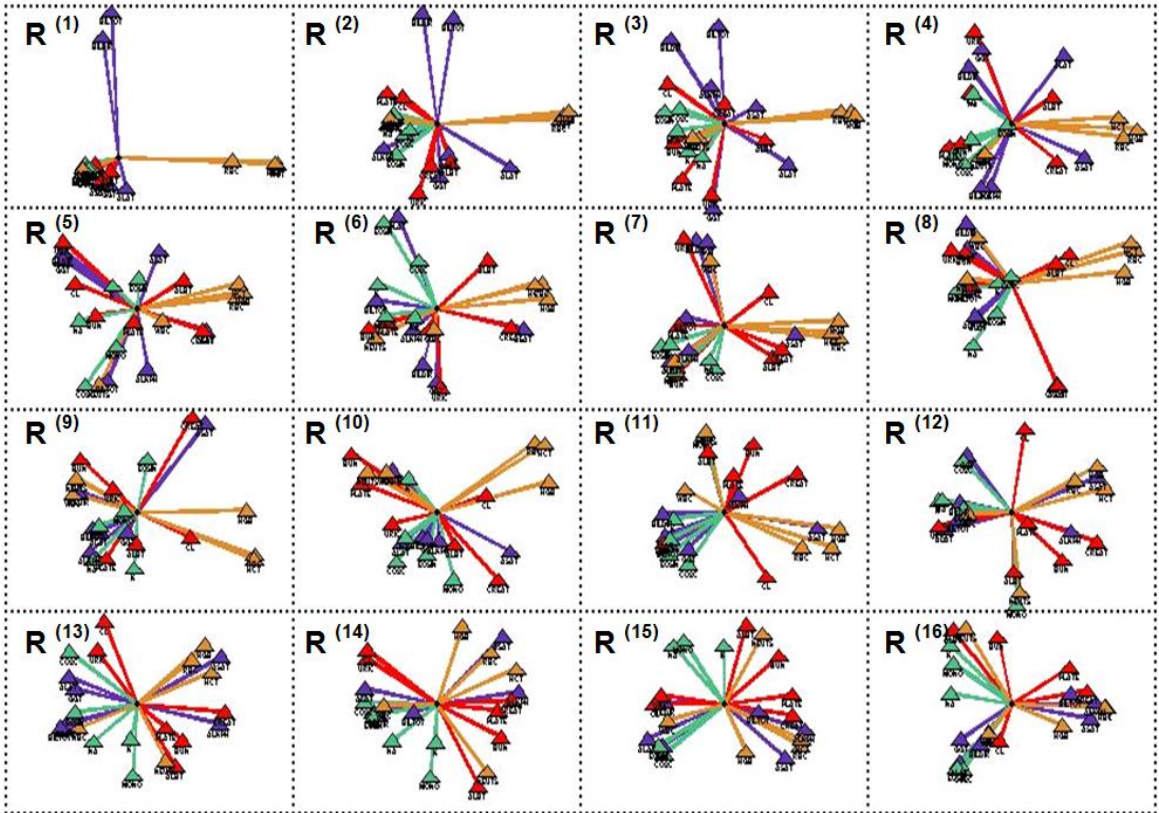

Figure 17. Hoeffding's Coefficient Matrix Iteration Process in a Two-Dimensional Eigen-Space

Figure 17 shows that the iteration does not converge. The Hoeffding's coefficient matrix is not a proximity matrix and prevents the seriation from converging.

III. DISPLAY OF MULTIPLE MATRICES WITH INTERACTIVE FEATURES

A clinical study usually involves multiple treatments and multiple time points. A desirable graph will display multiple correlation matrices on one page for easier comparison. This section discusses the display of multiple matrices with interactive features. Again, the DS2CONST macro is selected for this application. The display panel is divided into four sub-panels. Each sub-panel contains one matrix display. The clinical laboratory test parameters are the nodes, and the correlation coefficients are the links, in the constellation representation. The colors of the link lines represent the directions of the correlation - blue for positive and red for negative. The thickness of the line indicates the magnitude of the association. The interactive features in this application are data tips and an embedded scroll bar for subsetting nodes and links.

Figure 18 shows the display of correlation matrices from two treatments and two visits in four sub-panels. Figure 19 illustrates the interactive features of data tips for nodes and links.
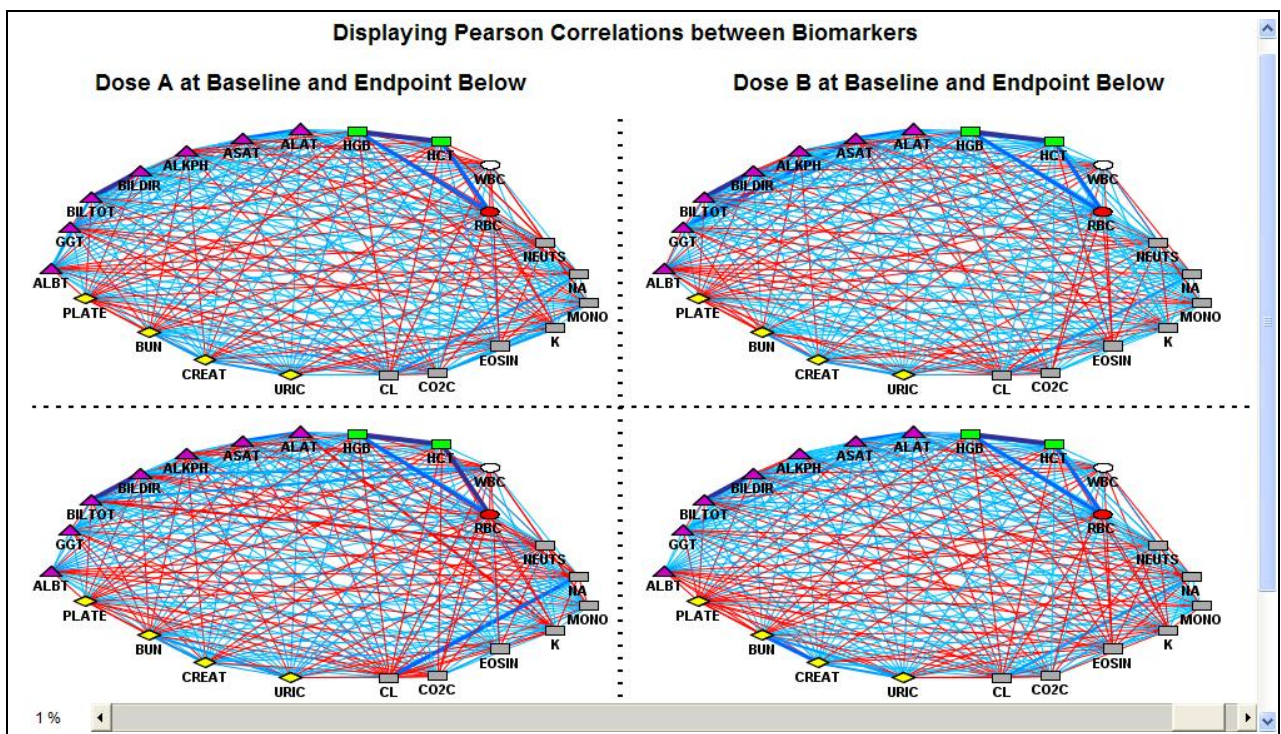

Figure 18. Multiple Correlation Matrices Display

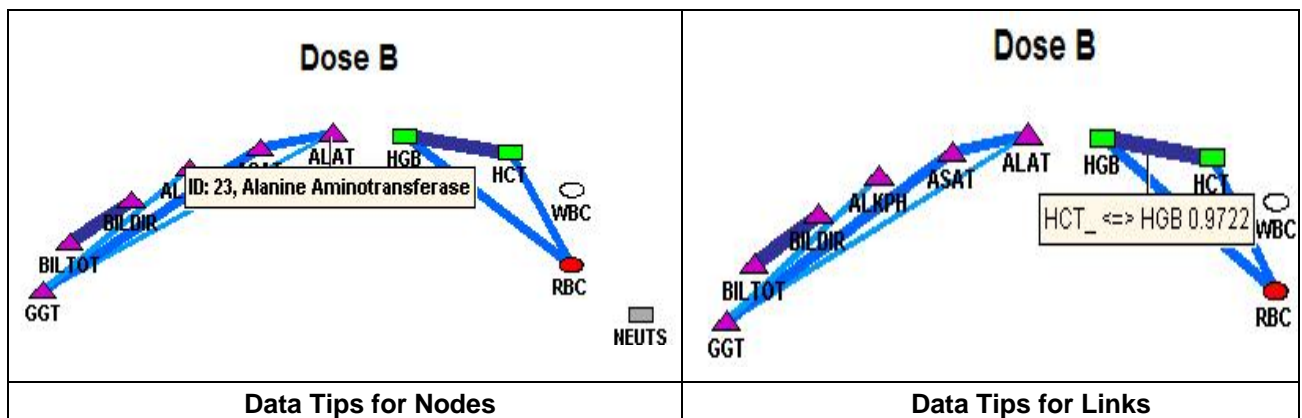| Data Tips for Nodes | Data Tips for Links |

15

Figure 19. Interactive Features of Data Tips for Nodes and Links

Figure 20 shows the interactive features with the viewer's control of the scroll bar interaction. This interactive feature allows the viewer to remove the weak associations from the display. The viewer is able to focus on the stronger associations among the variables.
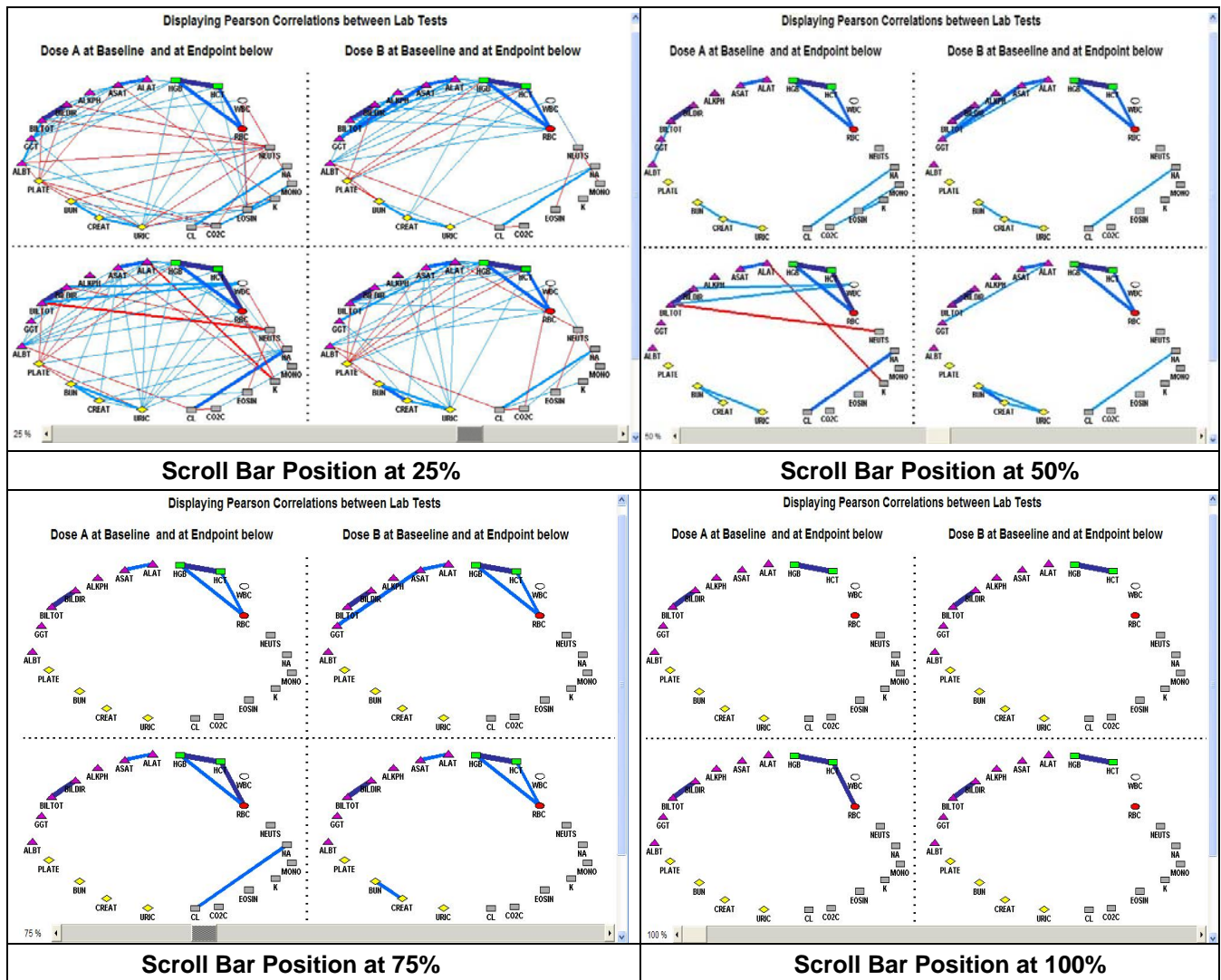

Figure 20. Using Scroll Bar for Subsetting Nodes and Links

CONCLUSION

Correlation matrices are commonly used for multivariate techniques and data analyses. A visual display of a correlation matrix helps in depicting the patterns of relations among variables. The SAS system procedures used in this paper are: CORR, UNIVARIATE, LOESS, DISTANCE, CLUSTER, TREE, KDE, PRINCOMP, CONELIP macro, G3D, GCONTOUR, GPLOT. GREPLAY, graphic DS2CONST macro, and ODS Graphics. The computations and display power from the SAS System are fully utilized in this paper for the display design tasks.

The interactive features are very desirable and enable the user to: 1) change graph properties without rerunning the job, 2) include extra information on the output, and 3) control the subsetting nodes and links allowing the viewer to focus on stronger association nodes.

The matrix visualization techniques and methods discussed in this paper could be effectively used to visualize data patterns, clusters, classifications, and data reduction, as well as to explore large datasets.

TRADEMARKS

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. [®] indicates USA registration.
Other brand and product names are registered trademarks or trademarks of their respective companies.

REFERENCES

[1] Bilker, W.B., C.M. Brensinger, R.C. Gur, M. Herlin.(2005): *A SAS Macro for Visually Displaying Correlations between Brain Regions Using fMRI Data* , Proceedings of the 30th SAS Users Group International Conference*,* Paper 151-30, April 2005

[2] Chen, C.H. (1996) *The properties and applications of the convergence of correlation matrices.* Proceedings of the Statistical Computing Section, pp. 49-54, Alexandria, VA, 1996, American Statistical Association.

[3] Chen, C.H. (1999) *Extensions of generalized association plots (GAP)*, Proceedings of the Statistical Computing Section, pp. 111-116, Alexandria, VA, 1999, American Statistical Association.

[4] Chen, C.H. (2002) *Generalized Association Plots: Information Visualization Via Iteratively Generated Correlation Matrices*, Statistica Sinica 12, pp-7-29, Academia Sinica, Taipei, Taiwan. 2002

[5] Friendly, M., (1991): *SAS System for Statistical Graphics*, SAS Institute, Cary NC.

[6] Friendly, M., (2001): *Visualizing Categorical Data* , SAS Institute, Cary NC.

[7] Friendly, M., (2002): *Corrgrams: Exploratory displays for correlation matrices*, The American Statistician, V1.5, Aug. 2002. pp 1-18

[8] Friendly, M and E. Kwan, W.B., (2002): *Effect ordering for data displays*, Computational Statistics and Data Analysis, 37, 2002 .

[9] Koziol, J.A., Alexander, J.E., Bauer, L.O., Kuperman, S., Morzorati, S., O'Connor, S.J., Rohrbangh, J., Porjesz, B., Begleiter, H., and Polich,J. (1997): *A Graphical Technique for Displaying Correlation Matrices* , The American Statistician, Vol. 51, No.4, (Nov. 1997), pp 301-304

[10] Legat, D., (2005): *Matrix Visualization,* WDS'05 Proceedings of Contributed Papers, Part 1, pp 126-131, 2005.

[11] McQuitty, L. L, (1968): *Multiple clusters, types, and dimensions from iterative intercolumnar correlational analysis*, Multivariate Behavioral Research 3, pp 465-477.

[12] SAS Institute Inc.,(2003): *SAS[®] OnlineDoc 9.* , Cary NC. http://support.sas.com/91doc/docMainpage.jsp .

[13] SAS Institute Inc.,(2004): *Base SAS[®] 9.1 Procedure Guide* , Cary NC. SAS Institute Inc.

[14] SAS Institute Inc.,(2004): *SAS/GRAPH[®] 9.1 Reference, Volumes 1 and 2* , Cary NC. SAS Institute Inc.

[15] SAS Institute Inc.,(2004): *SAS/STAT® 9.1 User's Guide* , Cary NC. SAS Institute Inc.

[16] Wishart, D., (1999): *ClustanGraphics3 Interactive Graphics for Cluster Analysis.* Published in*: Classification in the Information Age,* Gaul, W., and Locarek-Junge, H., (Eds), Springer 1999, pp-268-275

[17] Yeh, S.T.(2005): *Using Interactive Treeview Diagrams for Clinical Adverse Experiences Records* , Proceedings of the North East SAS Users Group 2005 Conference*,* Paper pos8   , September 2005.

[18] Yeh, S.T.(2005): *Using RADAR Chart to Display Clinical Data* , Proceedings of the North East SAS Users Group 2005 Conference*,* Paper pos5, September 2005.

[19] Yeh, S.T.(2005): *Customizing ODS Statistical Graphics* , Proceedings of the 30th SAS Users Group International Conference*,* Paper 183-30, April 2005.

[20] Yeh, S.T.(2006): *Interactive Graphs from the SAS System* , Proceedings of the 31th SAS Users Group International Conference*,* Paper 181-31, March 2006.

CONTACT INFORMATION

Shi-Tao Yeh, Ph. D.
(610) 787-3856 (W)
E-mail: Shi-Tao.Yeh@gsk.com