

# The “Forgotten” JMP® Visualizations (Plus Some New Views in JMP® 9)

Sam Gardner, SAS Institute, Lafayette, IN, USA

## Abstract

JMP® has a rich set of visual displays that can help you see the information in your data. Many JMP® users, however, remain unfamiliar with the graphing options available for multivariate data visualization under the Graph Menu.

This paper will expose the features and value of several of these tools, including: Scatterplot3D, Scatterplot Matrix, Parallel Plot, Cell Plot, Bubble Plot, Variability Charts, and Tree Maps. In addition, some new visualization and charting features available in the JMP® version 9 will be shown.

## Introduction

This paper is structured around several graphing and charting tools available in JMP®. Many of the visualizations for multivariate data are unfamiliar (aka “forgotten”) or underutilized. This paper’s purpose is to introduce these visualization tools and to encourage their use. All charts were generated in JMP® version 9. All data used to create the charts in the examples below is available for download from the JMP File Exchange at <http://bit.ly/aulBvb> (search on the author’s name to find the correct download link).

## Scatterplot Matrix

A Scatterplot Matrix is a matrix of charts that show the two way relationships between a set of “Y” variables and “X” variables. To illustrate the Scatterplot Matrix, the example data set “Fitness.JMP” (available in the JMP® Sample Data). It contains the following variables:

- Name: individuals name
- Sex: (M)ale or (F)emale
- Weight: individuals weight in lbs.
- Oxy: blood oxygen content, measured after a period of running
- Runtime: length of time (minutes) running
- RunPulse: average heart rate during period of running
- RstPulse: resting heart rate
- MaxPulse: maximum heart rate during period of running

To see all of the two way “scatterplots” of the data (excluding the Name variable, which is simply a label in this data set), the Scatterplot Matrix can be launched from the JMP® menu (**Graph > Scatterplot Matrix**). The dialog with the variables selected is shown below, along with the resulting report window.

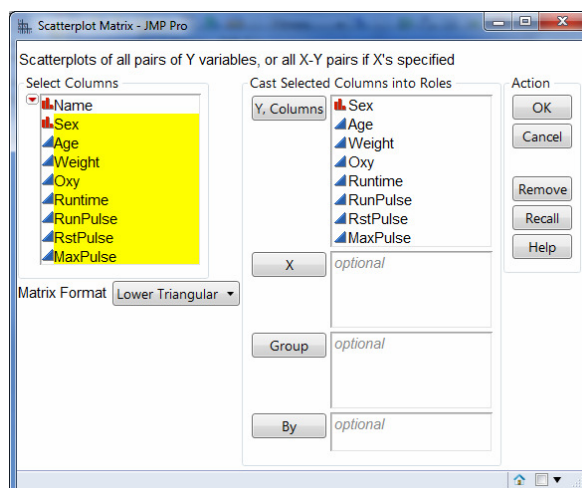
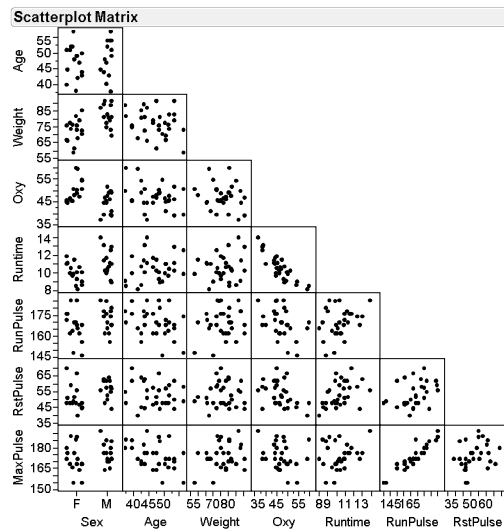


Figure 1: Scatterplot Matrix Launch Dialog



**Figure 2: Scatterplot Matrix on Fitness Data**

From the Scatterplot Matrix, we can see quickly which variables have strong two-way relationship: `MaxPulse` vs `RunPulse`, `Runtime` vs `Oxy`, `Oxy` vs `Sex`.

Display options for the Scatterplot Matrix include:

- Show Points: Turn on/off display of points in plots (useful when displaying other plot elements)
- Points Jittered: Useful for categorical (nominal or ordinal) variables, applying some random scatter in the plotted points around the categorical levels
- Fit Line (new in JMP® 9): Display a simple linear regression line in each scatterplot
- Density Ellipses: display a bivariate normal density contour ellipse, based on the means and correlation of the pairwise variables
- Shaded Ellipses: display a shaded bivariate normal density contour ellipse
- Ellipses Coverage: change the contour level for the density ellipse
- Ellipses Transparency: adjust the transparency of the shaded contour ellipse
- Nonpar Density: display a non-parametric density contour (based on a bivariate kernel density estimate)
- Group By: choose a grouping variable for generation/display of graphic elements

To see how these display elements can be used, we first cluster the data using hierarchical clustering (**Analyze>Multivariate>Cluster**) on `Oxy`, `Runtime`, `RunPulse`, `RstPulse`, and `MaxPulse`. Choosing three clusters, we save the cluster identifier to the data table (variable name `Cluster`) and color the points by the `Cluster`. We recreate the Scatterplot Matrix, using the “Group By” option and choosing `Cluster` as the grouping variable. We then select “Shaded Ellipses” and the resulting Scatterplot Matrix is produced.

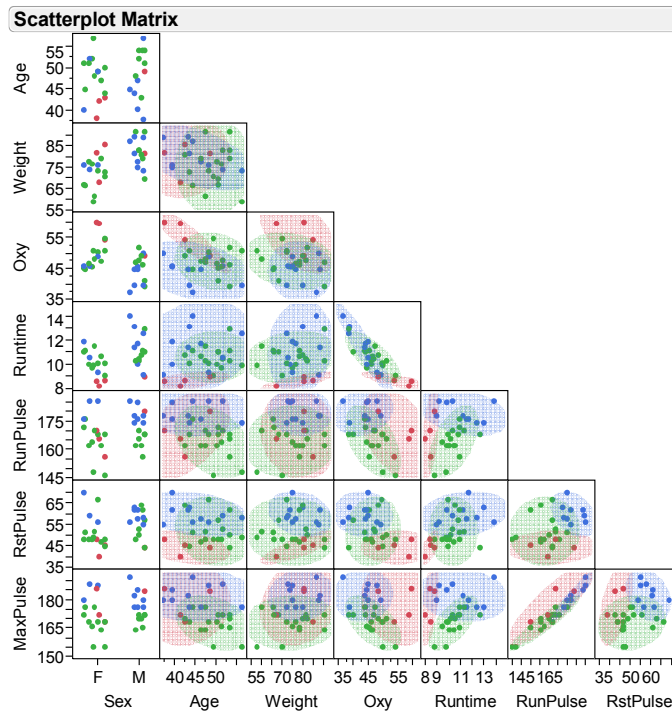


Figure 3: Scatterplot Matrix with Normal Density Contours by Cluster

### Scatterplot 3D

To visualize three dimensional relationships, the Scatterplot 3D platform can be utilized in JMP®. Using the same example as above, we launch the Scatterplot 3D platform, choosing the same variables as we did for the Scatterplot Matrix.

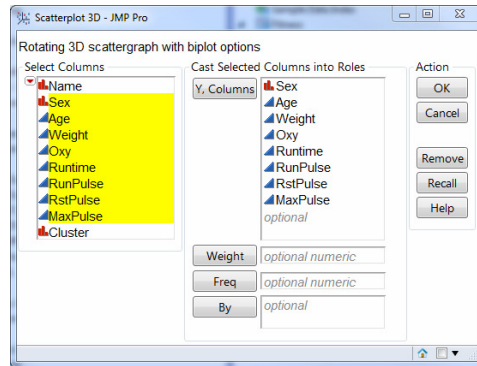


Figure 4: Scatterplot 3D Launch Dialog

The resulting plot is dynamic and rotatable. Static snapshots of the chart are shown below. The three variables displayed on the X, Y, and Z axis of the 3D plot are chosen using the drop down lists at the bottom of the chart. As in the Scatterplot Matrix, both multivariate normal and non-parametric density contours can be displayed on the plot, with or without a grouping variable. Additionally, the data can be transformed into a principle components coordinate space, and 3D biplot rays can be overlaid on the plot.

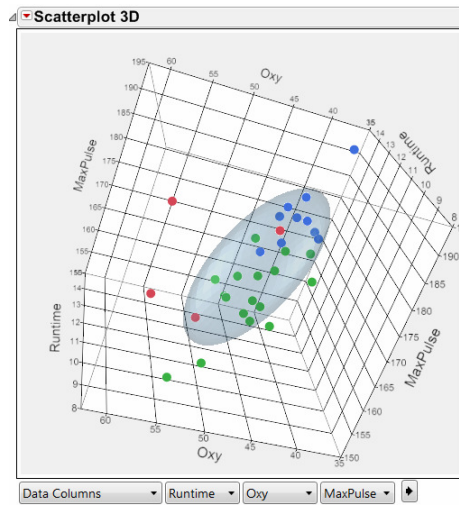


Figure 5: Scatterplot 3D with Normal Density Ellipse

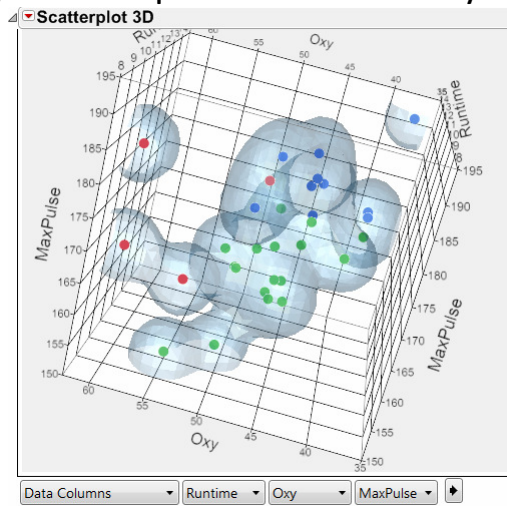


Figure 6: Scatterplot 3D with Non-parametric Density Contour

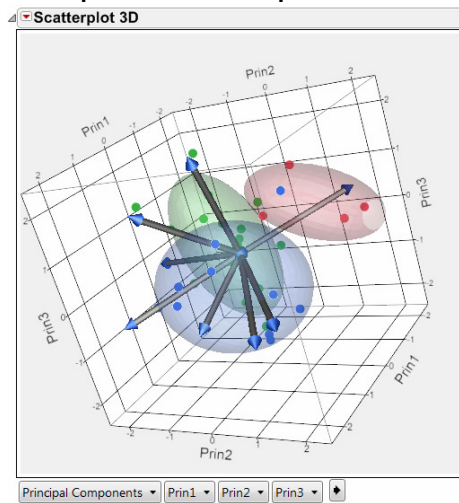
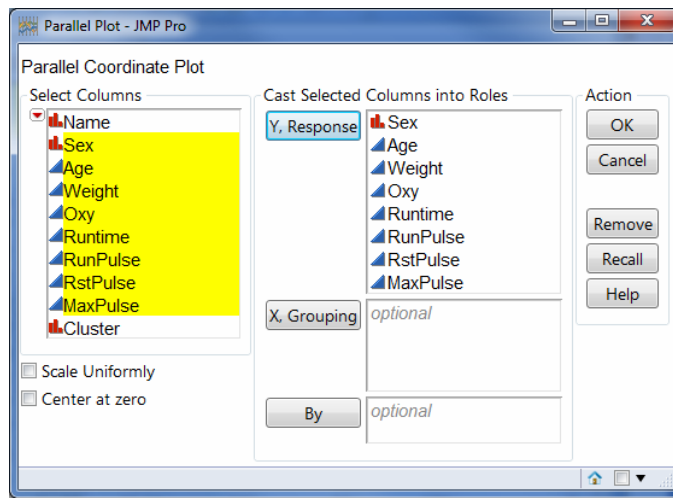


Figure 7: Scatterplot 3D with Principle Components and Biplot Rays

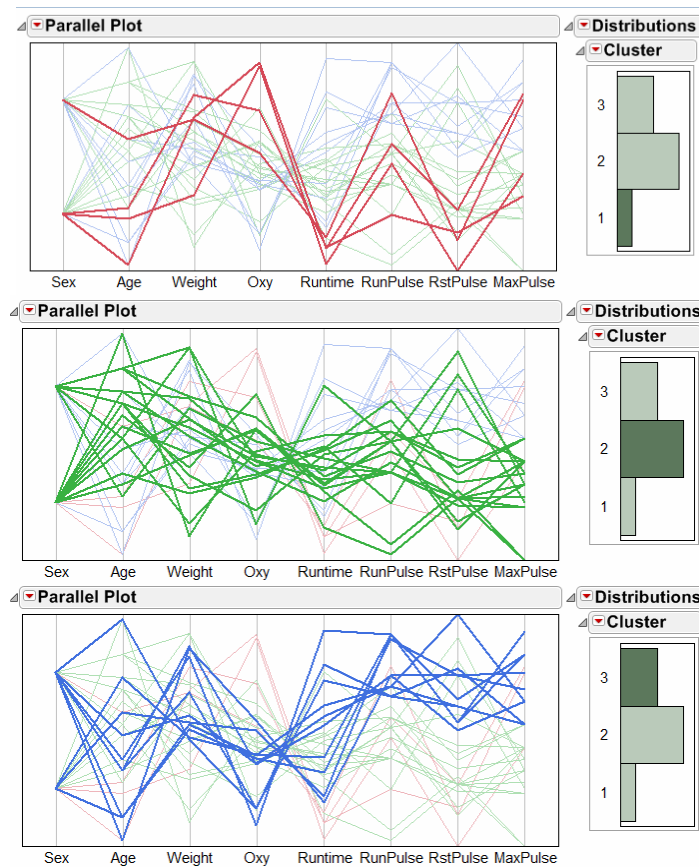
## Parallel Plot

Another multivariate visualization in JMP® is the Parallel Plot. We look at the same collection of variables as above.



**Figure 8: Parallel Plot Launch Dialog**

In the Parallel Plot, each variable is represented by a parallel line, and each row of data is displayed as a connected line segments across the parallel lines. Using the same example as above, one can see which individual variables are related to the clustering results by filtering on the cluster (in this case, visual filtering is achieved by creating a distribution/frequency chart of the cluster indicator, and then selecting the different cluster values in the frequency plot). From this display we see the cluster 1 is composed of primarily one Sex (in this case Male) that had a low Runtime and higher than average Oxy levels, while cluster 3 has lower than average Oxy levels and higher than average RunPulse, RstPulse, and MaxPulse.



**Figure 9: Parallel Plots of Fitness Data**

## Parallel Plot

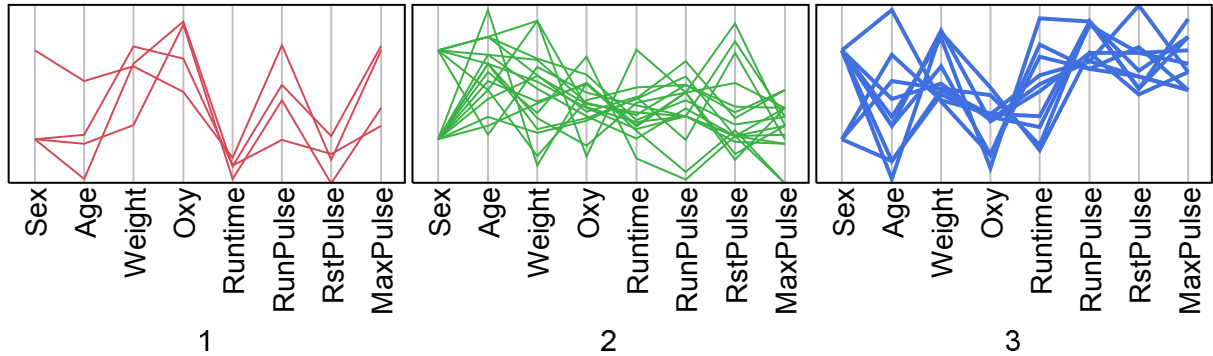


Figure 10: Parallel Plots of Fitness Data by Cluster

## Cell Plot

A cell plot is another way to visualize a multivariate data set, using color patterns to see correlations and groupings.

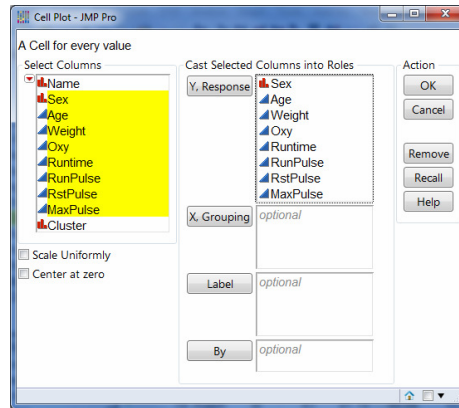


Figure 11: Cell Plot Launch Dialog

The default color scale for the cell plot is blue-to-gray-to-red (low-to-high), with each variable having a separate scale based on the range of values in the column. Similar color patterns between columns indicate a positive correlation; while opposite color patterns indicate a negative correlation. Examining the *Oxy* and *Runtime* columns we see that, overall, they are negatively correlated. However, taking into account the clustering that was performed earlier, strong correlations between those columns are only seen in clusters 1 and 3.

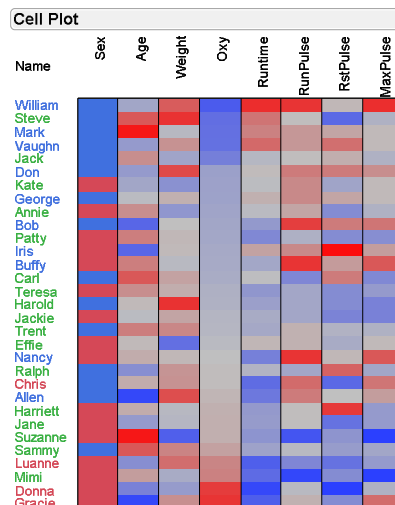


Figure 12: Cell Plot for Fitness Data

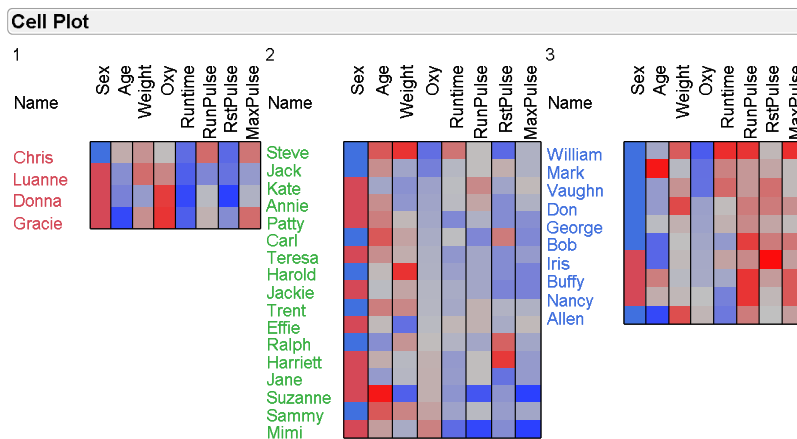


Figure 13: Cell Plot for Fitness Data by Cluster

## Variability Chart

A Variability Chart is a useful tool for looking at the variability in one continuous response and how that variability relates to several categorical factors. Consider the situation of a high throughput chemical screening process, where compounds being screened are placed in 96 well plates (an 8x12 array of small wells that hold small amounts of material).

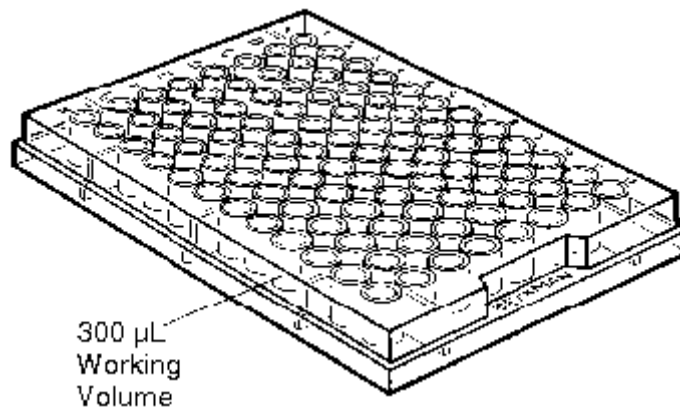


Figure 14: 96 Well Plate

To monitor the performance of the screening system, six control plates (with the same control compound in each of the 96 wells) are assayed throughout each screening run. Variability within the plate (row by row variation and within row variation), as well as variation between plates is the main concern for this system. These sources of variability can be seen easily using a Variability Chart. Using the example data set "Variation in Plates.JMP", we look at the response variable Y1 and its variability vs. Plate and Row.

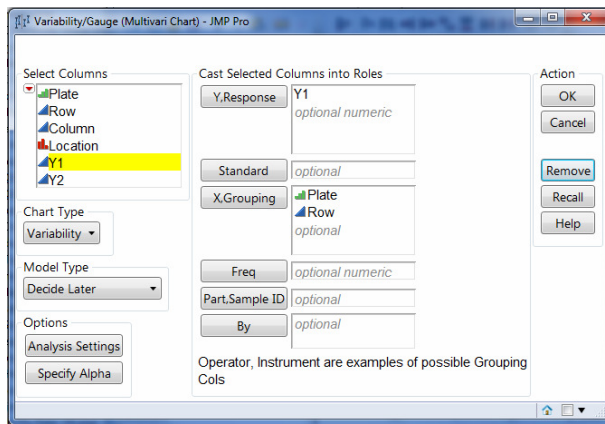


Figure 15: Variability/Gauge Chart Launch Dialog

The resulting chart is shown below. This shows that there is some systematic variability between rows, in that the odd numbered and even number rows appear to have a different mean value. However, plate to plate variability and within row variability appears to be consistent.

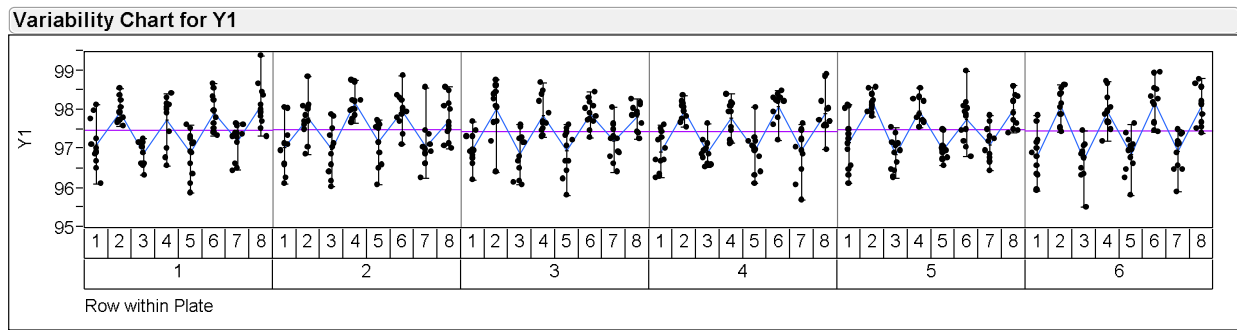


Figure 16: Variability Chart for 96 Well Plate Data

## Tree Map

A Tree Map is useful when the data of interest has categorical variables with many levels, and potentially some descriptive continuous variables. The focus of the tree map is on the categories. Traditional visualizations like frequency charts and pie charts suffer from clutter and a lack of compactness in the view of the data.

Consider the example data set “Cities.JMP”. This contains air quality data for 52 cities in the US, along with population, weather, and geographic reference information.

If we wish to see how OZONE levels relate to each city, taking into account the population of each city, a tree map gives a nice, compact display.

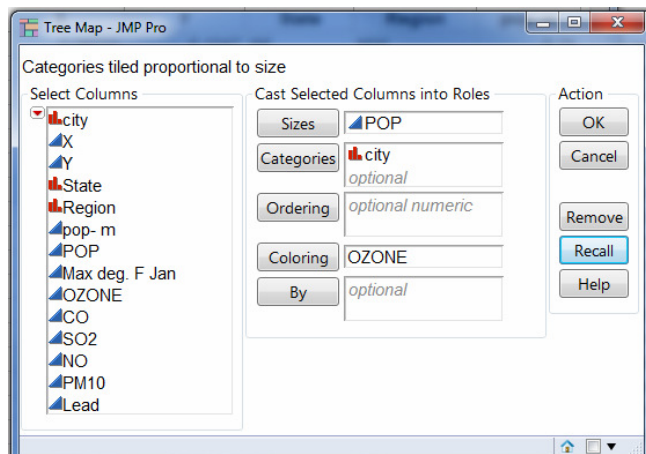
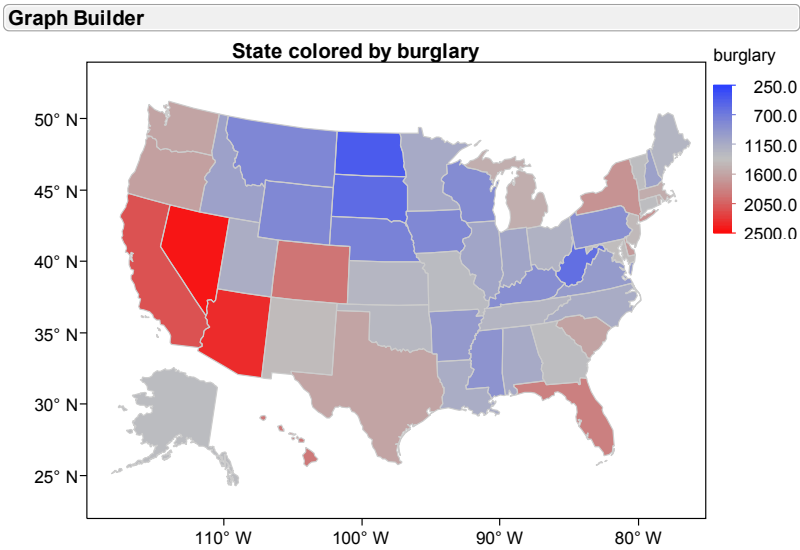


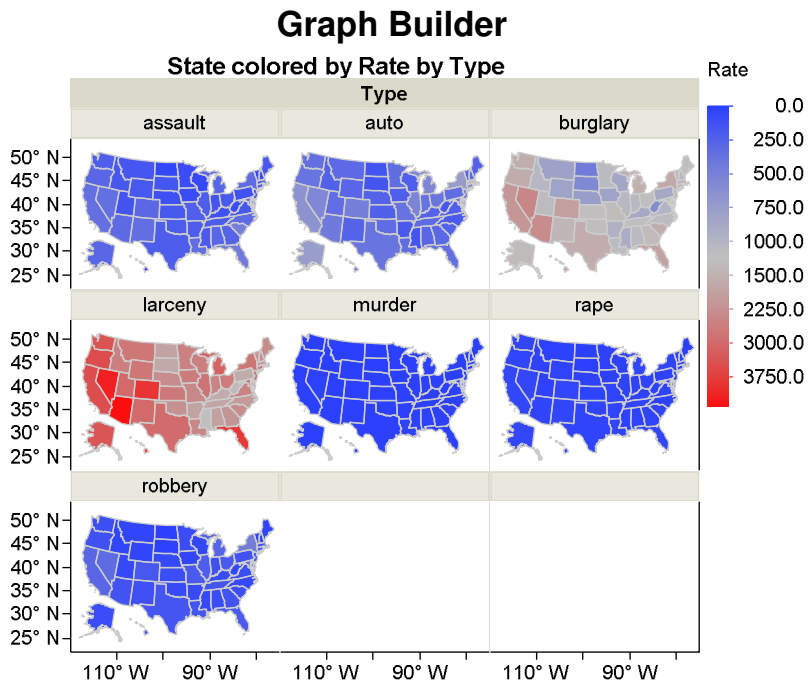
Figure 17: Tree Map Launch Dialog







**Figure 20: Graph Builder of Burglary Rate by State (with Shapes)**



**Figure 21: Crime Rates by State Using Shapes and Coloring**

## Conclusion

JMP® has many valuable graphs that allow you to visualize multivariate data. Many of these graphs are not well known, but exposure to these graphs in practical applications can increase the value of JMP® to the JMP® user. This paper strives to introduce several of the JMP® visualizations that are often “forgotten”. The reader is encouraged to use these charts to help see the patterns in their own data and to communicate their results more effectively.

## Contact Information

Your comments and questions are valued and encouraged. Contact the author at:

Sam Gardner  
SAS Institute/JMP Division  
Phone: 765.589.3696  
E-mail: [Sam.Gardner@JMP.com](mailto:Sam.Gardner@JMP.com)  
Web: [www.jmp.com](http://www.jmp.com)

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 trademarks of their respective companies.