Matrix Plot

statgraphics 18®
centurion

Revised: 10/11/2017

Summary

The **Matrix Plot** procedure creates a matrix of plots for 3 or more numeric variables. The diagonal of the matrix contains box-and-whisker plots for each variable. The off-diagonal positions contain 2-variable scatterplots for all pairs of variables. The procedure is very useful for obtaining an initial look at multivariate data. From the plot, one can often detect relationships amongst the variables, the presence of outliers, and other interesting features of the data.

Sample StatFolio: *matrixplot.sgp*

© 2017 by Statgraphics Technologies, Inc.

Sample Data

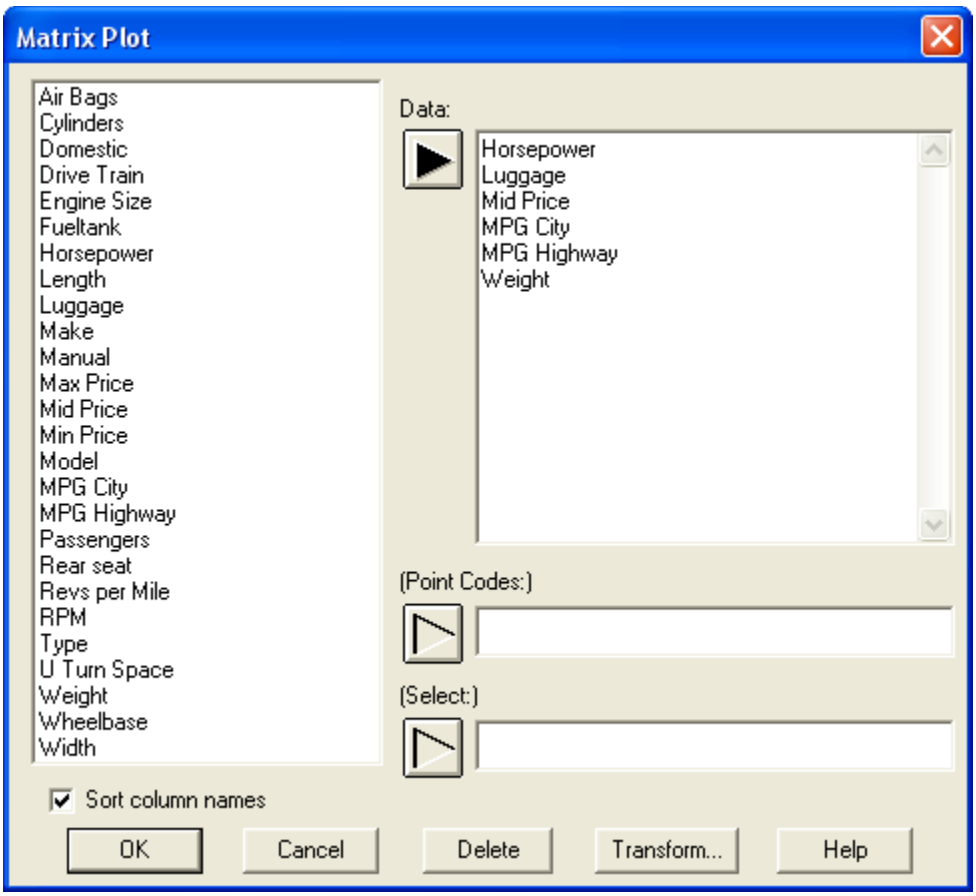
The file *93cars.sgd* contains information on 26 variables for $n = 93$ makes and models of automobiles, taken from Lock (1993). The table below shows a partial list of 8 columns from that file:

Make	*Model*	*MPG Highway*	*MPG City*	*Weight*	*Horsepower*	*Luggage Room*	*Midrange Price*
Acura	Integra	31	25	2705	140	5	15.9
Acura	Legend	25	18	3560	200	5	33.9
Audi	90	26	20	3375	172	5	29.1
Audi	100	26	19	3405	172	6	37.7
BMW	535i	30	22	3640	208	4	30
Buick	Century	31	22	2880	110	6	15.7
Buick	LeSabre	28	19	3470	170	6	20.8
Buick	Roadmaster	25	16	4105	180	6	23.7
Buick	Riviera	27	19	3495	170	5	26.3
Cadillac	DeVille	25	16	3620	200	6	34.7
Cadillac	Seville	25	16	3935	295	5	40.1
Chevrolet	Cavalier	36	25	2490	110	5	13.4

© 2017 by Statgraphics Technologies, Inc.

Data Input

The data to be analyzed consist of 2 or more numeric columns containing $n = 2$ or more observations.



- **Data :** 2 or more numeric columns containing the data to be plotted.
- **Point codes**: optional numeric or non-numeric column used to code the points.
- **Select:** subset selection.

Analysis Summary

The *Analysis Summary* shows the names of the data columns and an indication of how missing values have been handled.

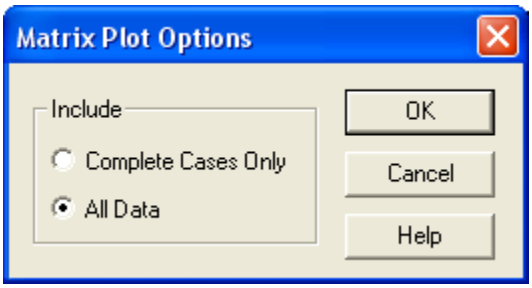
Matrix Plot
Data variables:
Horsepower (maximum)
Luggage (cu. ft.)
Mid Price (average of min and max prices in $1,000)
MPG City (miles per gallon in city driving)
MPG Highway (miles per gallon in highway driving)
Weight (pounds)

© 2017 by Statgraphics Technologies, Inc.

All available data are shown in the plot.

The *Analysis Options* dialog box controls whether all cases containing missing values in one or more columns are excluded from all of the plots ("casewise exclusion"), or whether all data is used wherever possible.

Analysis Options

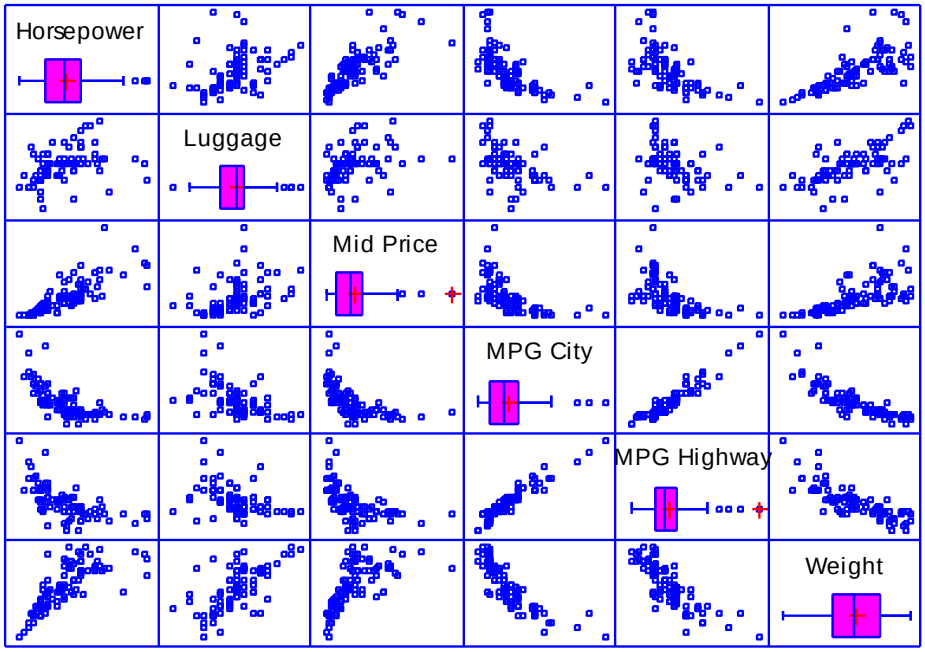


- **Complete Cases Only**: exclude from all plots any row in which one or more of the input data columns contains a missing value. In this case, all box-and-whisker plots and 2-variable scatterplots will be based on the same number of rows.
- **All Data**: use all data wherever possible (the default). In this case, the box-and-whisker plots will contain all non-missing data for the indicated column and the 2-variable Scatterplots will display all rows in which neither of the variables being plotted are is missing.

Scatterplot Matrix

The *Scatterplot Matrix* shows a rectangular matrix of plots.

© 2017 by Statgraphics Technologies, Inc.



Each data variable defines a row and a column. For example, variable 1 (*Horsepower*) is shown in the first row and first column, variable 2 (*Luggage*) is shown in the second row and second column, etc.

The *i*-th diagonal position displays a box-and-whisker plot for variable *i*. These plots contain:

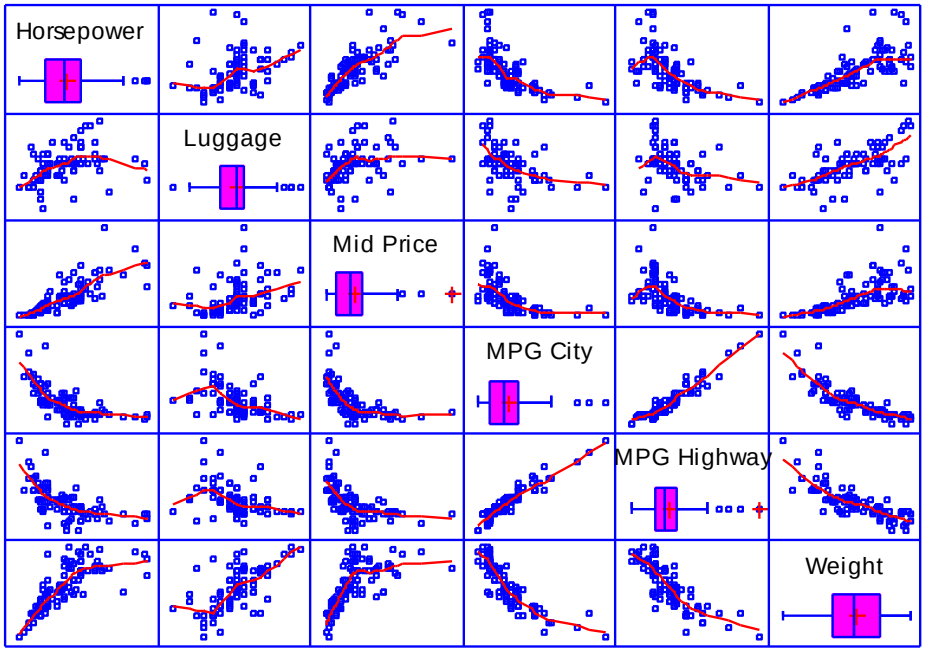
- A central box covering the middle half of the data values for that variable.
- A vertical line at the sample median.
- A plus sign at the sample mean.
- Whiskers extending from the box to the largest and smallest data values, unless one or more data values are classified as "outside points". In such cases, the whiskers extend only to the most extreme values that are not classified as outside points.
- Point symbols for all points that are more than 1.5 times the sample interquartile range above or below the box ("outside" points).
- Point symbols with superimposed plus signs for all points that are more than 3 times the sample interquartile range above or below the box ("far outside" points).

Certain aspects of the box-and-whisker plots can be suppressed using the *Pane Options* dialog box.

The scatterplots in the off-diagonal positions show observed pairs of variables. The scatterplot in row *i*, column *j* displays variable *i* on the vertical axis and variable *j* on the horizontal axis. In the matrix, every pair of variables is thus plotted twice, once with the first variable on the X axis and once with that variable on the Y axis.

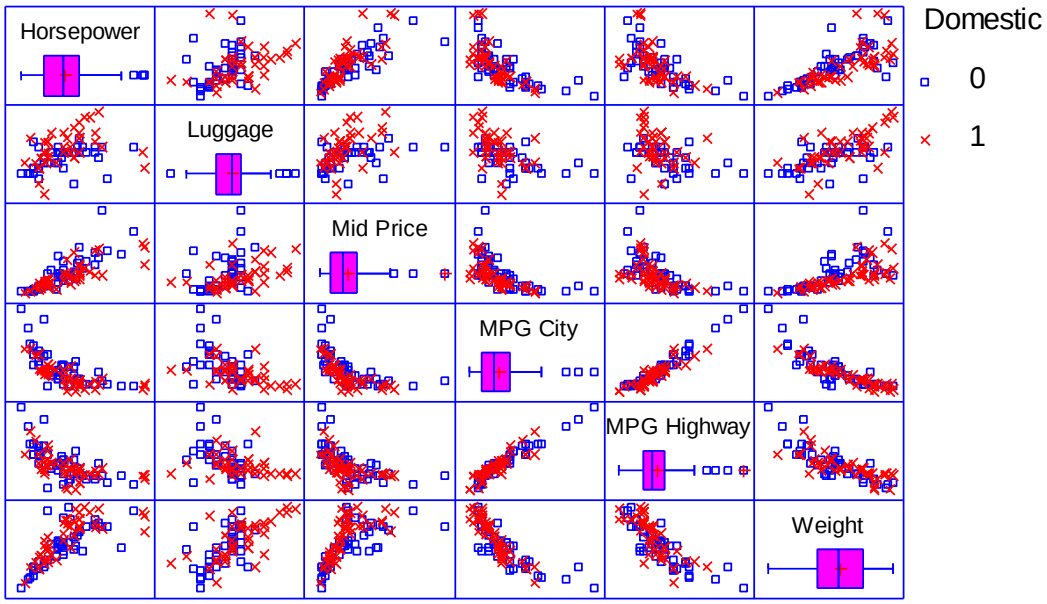
It is sometimes useful to smooth the scatterplots by pushing the *Smooth/Rotate* button on the analysis toolbar. The plot below uses the default *Robust LOWESS* smoother:

© 2017 by Statgraphics Technologies, Inc.



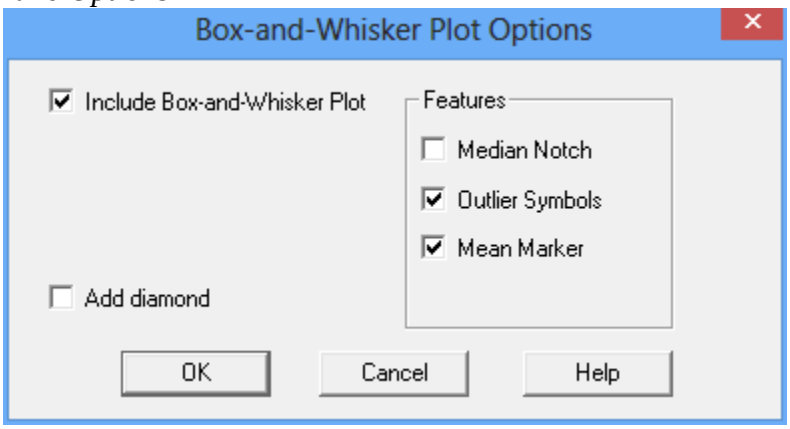
It is now easier to judge the relationships that exist amongst the variables.

The *Point Codes* field on the data input dialog box can also be used to code the point symbol color and type based upon the value of an additional column.



© 2017 by Statgraphics Technologies, Inc.

Pane Options



- **Include box-and-whisker plot**: whether to include box-and-whisker plots in the display.
- **Direction**: direction of the box-and-whisker plot (disabled in this procedure).
- **Features**: the box-and-whisker plots may include a notch to indicate a 95% confidence interval for the median, outlier symbols to indicate the presence of outside points, and/or a plus sign to indicate the location of the sample mean.
- **Add diamond:** if selected, a diamond will be added to the plot showing a 100(1-α)% confidence interval for the mean at the default system confidence level.

© 2017 by Statgraphics Technologies, Inc.