Parallel Coordinates Plot

Summary

The **Parallel Coordinates Plot** is a multivariate visualization technique that can be very useful in identifying differences and similarities amongst observed cases when the number of dimensions is too large to use a standard scatterplot.

Sample StatFolio: *parallelcoords.sgp*

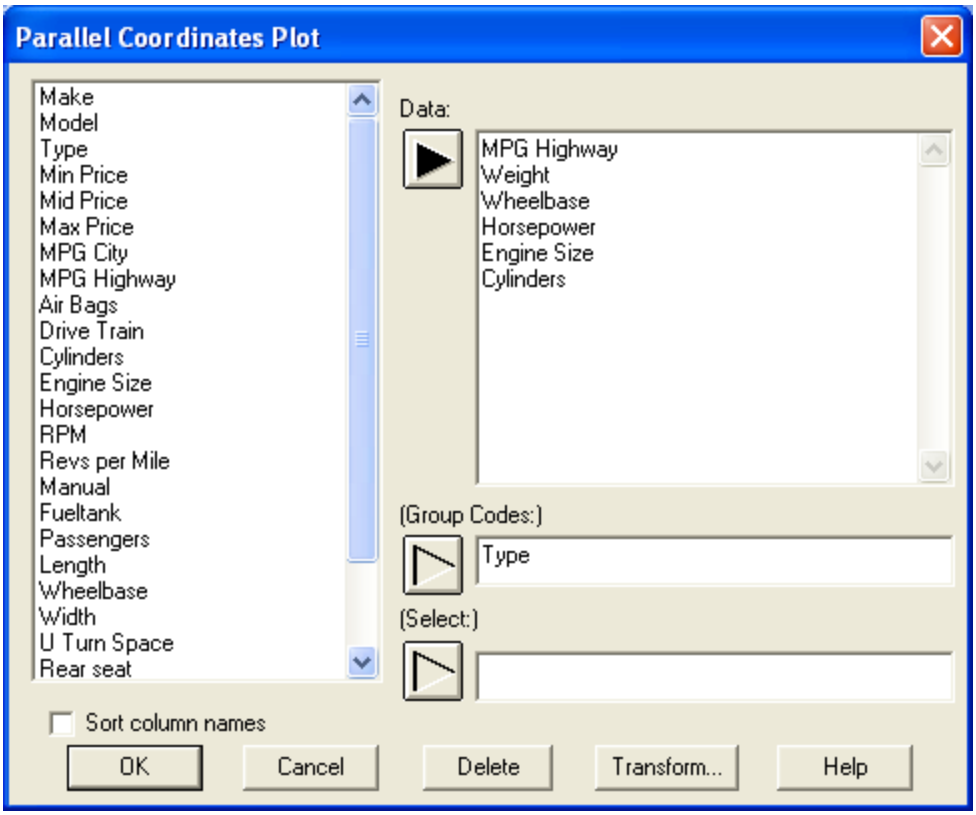
Sample Data:

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 the data in that file:

Make	*Model*	*MPG Highway*	*Weight*	*Wheelbase*	*Horsepower*	*Engine Size*	*Cylinders*
Acura	Integra	31	2705	102	140	1.8	4
Acura	Legend	25	3560	115	200	3.2	6
Audi	90	26	3375	102	172	2.8	6
Audi	100	26	3405	106	172	2.8	6
BMW	535i	30	3640	109	208	3.5	4
Buick	Century	31	2880	105	110	2.2	4
Buick	LeSabre	28	3470	111	170	3.8	6
Buick	Roadmaster	25	4105	116	180	5.7	6
Buick	Riviera	27	3495	108	170	3.8	6
Cadillac	DeVille	25	3620	114	200	4.9	8
Cadillac	Seville	25	3935	111	295	4.6	8
Chevrolet	Cavalier	36	2490	101	110	2.2	4
Chevrolet	Corsica	34	2785	103	110	2.2	4
Chevrolet	Camaro	28	3240	101	160	3.4	6
Chevrolet	Lumina	29	3195	108	110	2.2	4
Chevrolet	Lumina_APV	23	3715	110	170	3.8	6
Chevrolet	Astro	20	4025	111	165	4.3	6
Chevrolet	Caprice	26	3910	116	170	5.0	8
Chevrolet	Corvette	25	3380	96	300	5.7	8
Chrylser	Concorde	28	3515	113	153	3.3	6

Data Input

The data to be analyzed consist of 2 or more numeric columns and an optional column with group identifiers:



- **Data:** 2 or more numeric columns containing the data to be plotted.
- **Group Codes:** an optional column with levels to be used to identify groups of cases.
- **Select:** subset selection.

As an example, 6 variables have been selected. The type of vehicle will be used to identify the cases.

Analysis Summary

The *Analysis Summary* shows the number of rows with complete data and summary statistics for those rows:

Parallel Coordinates Plot

Data variables:
- MPG Highway (miles per gallon in highway driving)
- Weight (pounds)
- Wheelbase (inches)
- Horsepower (maximum)
- Engine Size (liters)
- Cylinders

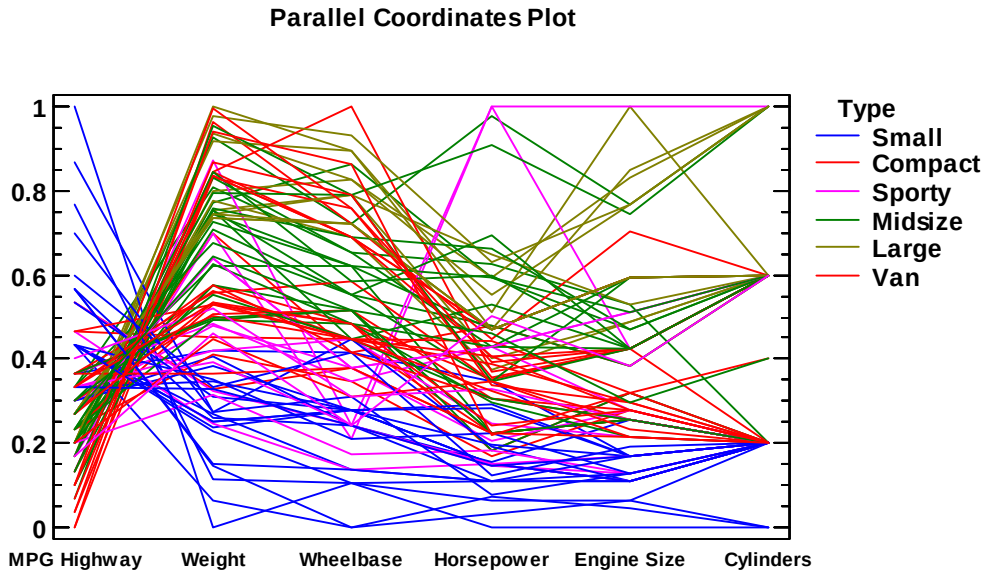
Number of complete cases: 92

	Sample mean	*Standard deviation*	*Minimum*	*Maximum*
MPG Highway	29.1304	5.34362	20.0	50.0
Weight	3074.84	592.832	1695.0	4105.0
Wheelbase	104.033	6.8057	90.0	119.0
Horsepower	142.62	51.3413	55.0	300.0
Engine Size	2.68261	1.03304	1.0	5.7
Cylinders	4.96739	1.30469	3.0	8.0

There are 92 rows with data for all of the variables.

Parallel Coordinates Plot

The *Parallel Coordinates Plot* draws one line for each row with complete data:



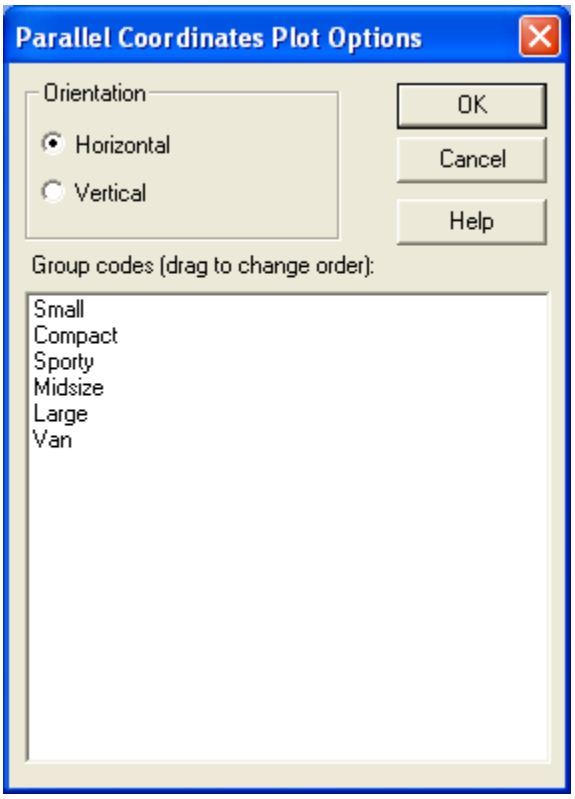
The line for the *i-th* row connects the standardized values of each variable in that row, where the standardized value equals the observed value minus the sample mean of that variable divided by the sample range:

$$X'_{ij} = \frac{X_{ij} - \min_j}{\max_j - \min_j} \tag{1}$$

If a group code variable is supplied, its values will be used to color the lines. In many cases, differences between groups of cases can be seen. For example, the plot above shows a large amount of clustering by type of vehicle. Small cars (shown as blue lines) tend to have a large values of *MPG Highway* but small values of all the other variables. You can also see some unusual *Sporty* cars that don't follow the standard pattern. If you click on a line with the left mouse button, the row number corresponding to that line will be displayed on the analysis toolbar.

Analysis Options

The *Analysis Options* dialog box allows you to change the orientation of the plot and the order of the group codes:



- **Orientation:** the direction of the lines on the plot.
- **Group Codes:** the order of the group codes in the legend block. You may drag level codes to change their order.

