Bubble Chart

Summary

The *Bubble Chart* is an X-Y scatterplot on which the value of a third and possibly fourth variable is shown by changing the size and/or color of the point symbols. It is one way to plot multivariate data in 2 dimensions.

Sample StatFolio: *bubblechart.sgp*

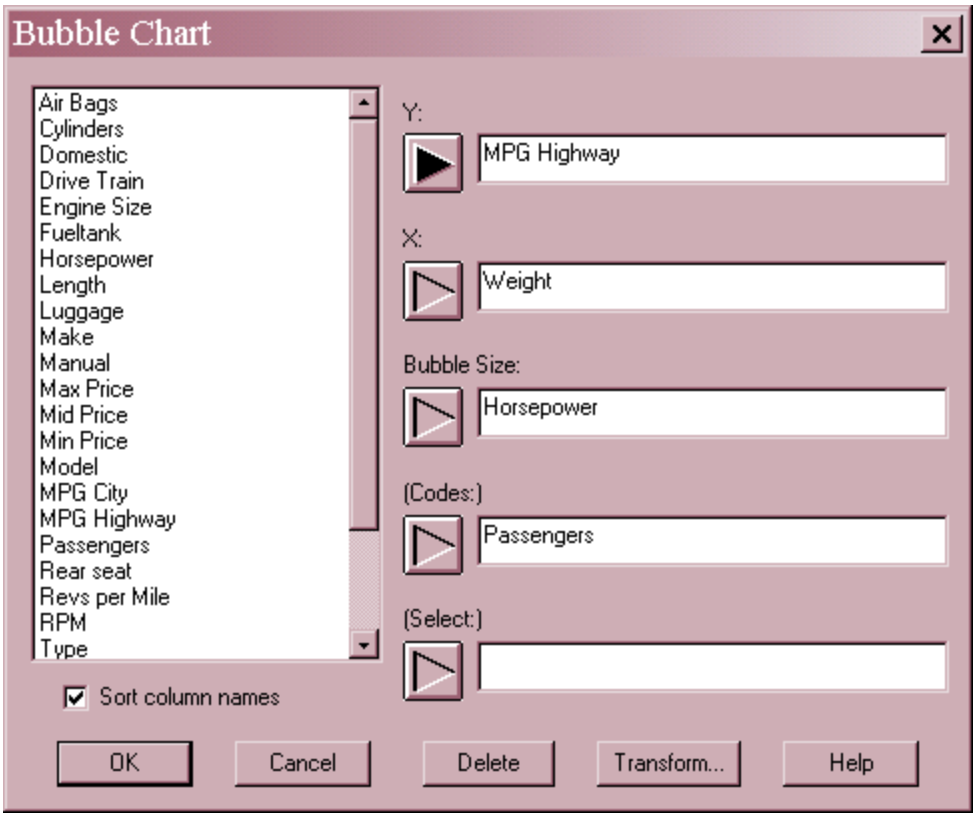
Sample Data:

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 6 columns from that file:

Make	*Model*	*MPG Highway*	*Weight*	*Horsepower*	*Passengers*
Acura	Integra	31	2705	140	5
Acura	Legend	25	3560	200	5
Audi	90	26	3375	172	5
Audi	100	26	3405	172	6
BMW	535i	30	3640	208	4
Buick	Century	31	2880	110	6
Buick	LeSabre	28	3470	170	6
Buick	Roadmaster	25	4105	180	6
Buick	Riviera	27	3495	170	5
Cadillac	DeVille	25	3620	200	6
Cadillac	Seville	25	3935	295	5
Chevrolet	Cavalier	36	2490	110	5

Data Input

The data to be analyzed consist of 3 numeric columns and an optional numeric or non-numeric column used to color code the points.



- **Y :** numeric column containing data to be plotted on the vertical axis.
- **X:** numeric column containing data to be plotted on the horizontal axis.
- **Bubble Size:** numeric column used to scale the size of the point symbols.
- **Codes:** numeric or non-numeric column used to determine the color of each point. Different colors are assigned to each unique value in this column.
- **Select:** subset selection.

Analysis Summary

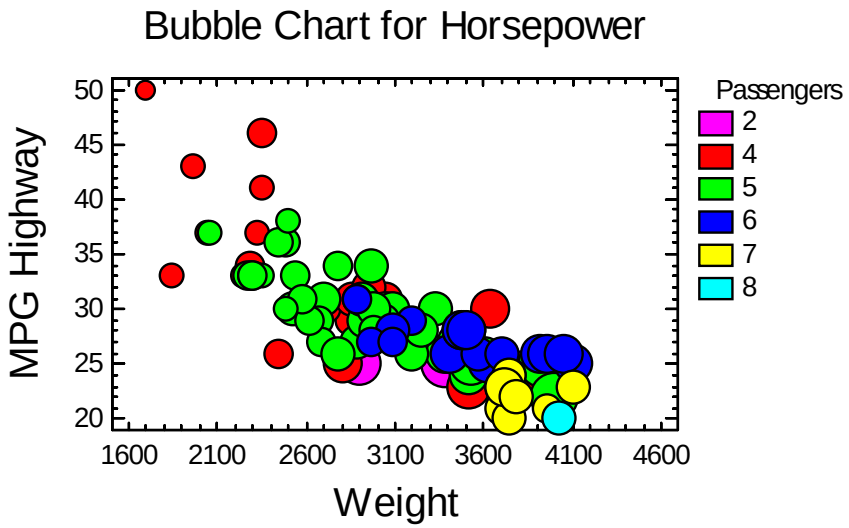
The *Analysis Summary* shows the number of observations in the data column.

Bubble Chart - MPG Highway vs. Weight
Y variable: MPG Highway (miles per gallon in highway driving)
X variable: Weight (pounds)
Bubble size: Horsepower (maximum)
Codes: Passengers (persons)
Number of observations: 93
Number of levels: 6

If *Codes* are specified, the number of different values (levels) in that column is also indicated.

Bubble Chart

This pane displays the bubble chart.

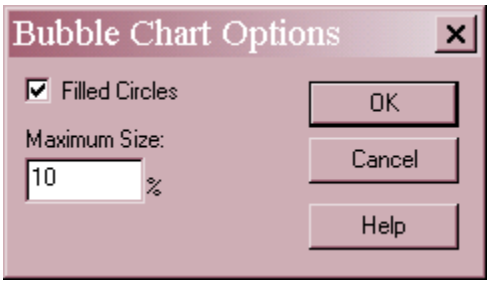


The plot is constructed in the following manner:

- The X and Y variables are used to position the points on the plot.
- The size of each point symbol is proportional to the value in the *Bubble Size* variable. If the range of values in the column is small relative to the average value, you may want to rescale the size by entering a STATGRAPHICS expression into the *Bubble Size* field that subtracts a constant value.
- The value of the *Codes* column is used to determine the point colors.

In the above plot, cars with larger horsepower are shown as larger bubbles.

Pane Options



- **Filled Circles**: if checked, the point symbols are solid colors. If not checked, they are unfilled circles.
- **Maximum Size**: The size of the points is scaled so that the largest bubble equals the specified percentage of the shorter axis.

Example – Scaling the bubble size

If the *Size* field on the data input dialog box is set to

Horsepower – 40

there will be more of a range in the bubble sizes.