三· 圖形 Graphics & Plots

2D,3D 視覺化資料是分析 data 很重要的第一步，本章將對 S-Plus 所提供的統計繪圖工具以及編輯圖形的方法及技巧做一介紹。

1. 圖形的建立

1.1. 選擇 data 後按繪圖 button 或

1.2. Insert/Graph/Axes type, Plot type/OK

2. 視覺化資料 2D

2.1. Example:

sample size:15,

X~Poison(5)+Normal(0,1),

Y=0.5*X+Normal(0,1),
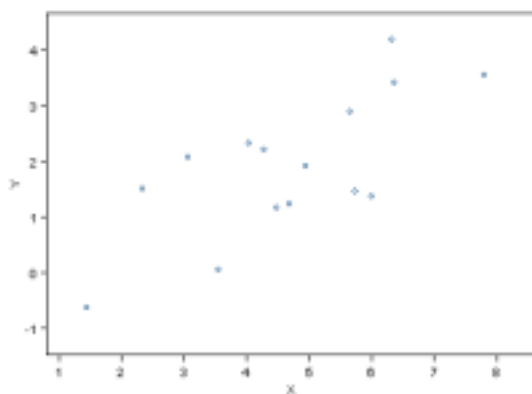
Z=1/X+1/Y+Normal(0,1)

C:categories(A,B,C) (Create categories 兩次以改變 level)

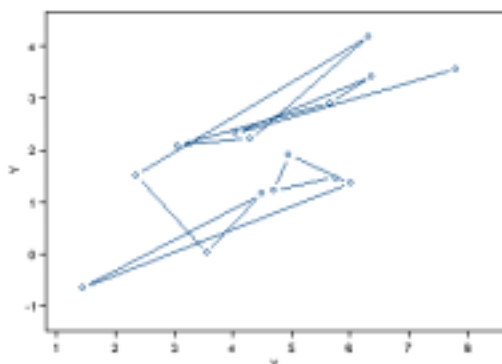| normal.0.1	poisson.5	x	y	z	c
0.39	6.00	6.39	3.59	0.83	c
-1.72	5.00	3.28	-0.08	-13.61	b
0.70	1.00	1.70	1.55	1.93	b
0.05	3.00	3.05	1.57	1.01	b
-0.38	7.00	6.62	2.93	0.11	c
-1.61	7.00	5.39	1.09	-0.50	c
-1.29	3.00	1.71	-0.44	-2.99	b

▼Scatter plot

▼Scatter plot (x,y,z) + Legend
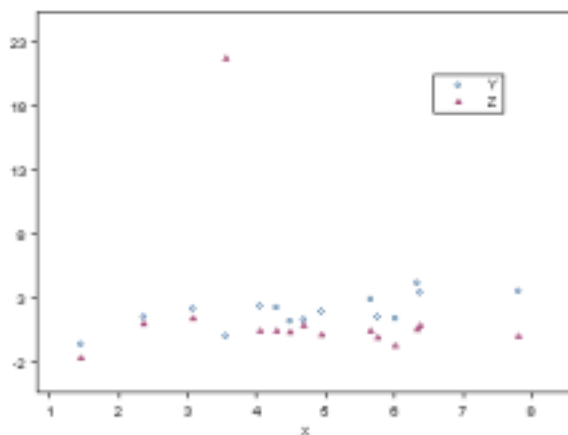
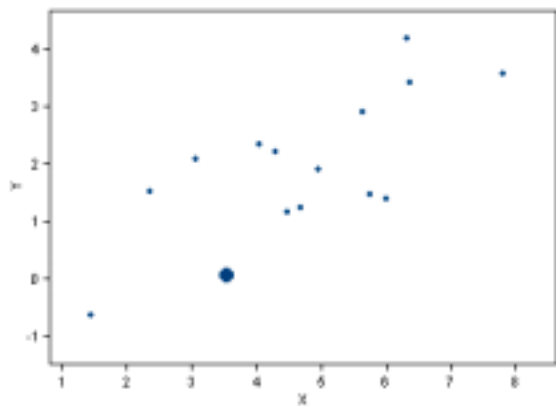▼Line plot with isolated points

線圖+資料點: 依觀察的順序連線。

symbol height: symbol 的 size

symbol freq: specify how frequency symbols are displayed on data points.
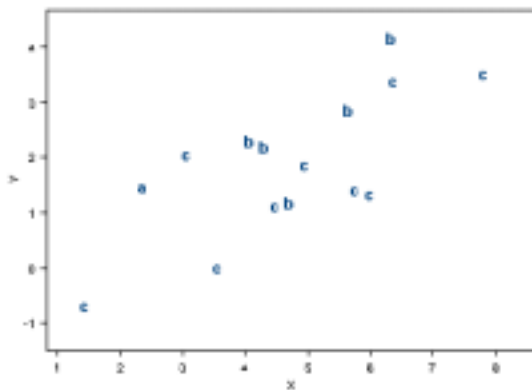
　ex:

　1. to place a symbol at each data point: 1

　2. every third point is plotted wth the symbol: 3

▼Bubble plot /Color plot /Bubble color plot

Bubble 和 color plots 是 scatter plots 的一種，利用符號的大小或顏色的變化來表示另一維度的資料。
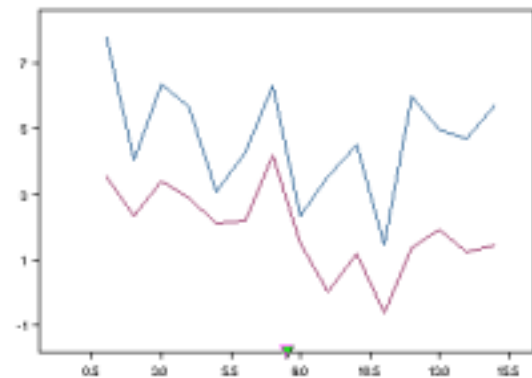
▼Line with text as symbols

X: double
Y: double
C: a,b,c
文字字串當成符號來畫圖，也是 scatter plot 的一種，適合用作觀察族群型的資料。
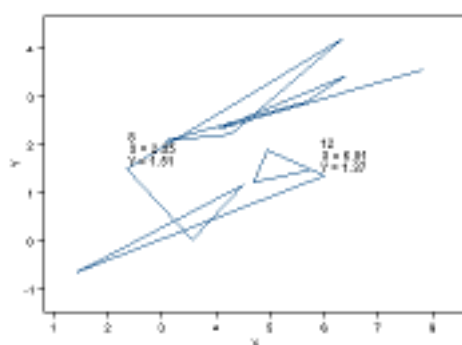
▼Y series lines

時間序列 Time series
選擇 y1, y2 對 x-axis 畫 line plots。
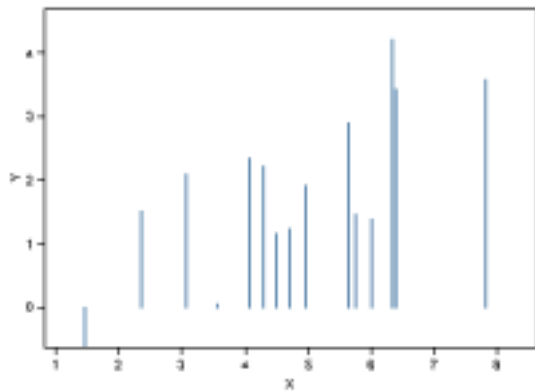Click x-axes 改變適合的 label type，如時間日期等。
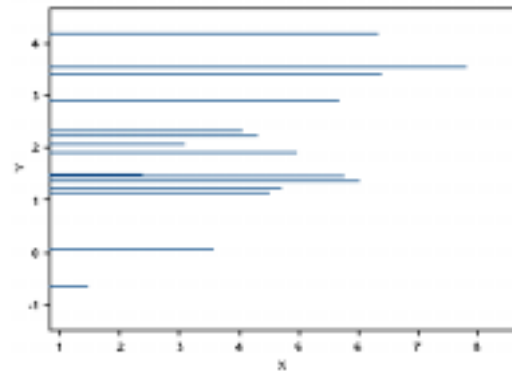
▼Line plots of xy pairs

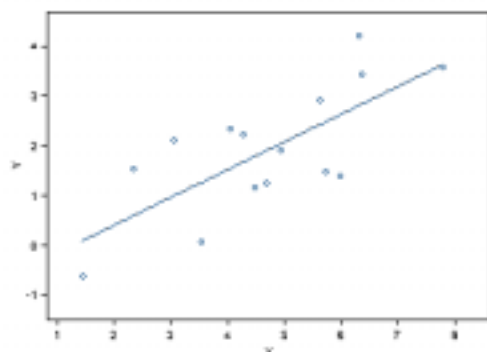An XY pairs line plot plot multiple sets of X and Y pairs on a common set of axes.
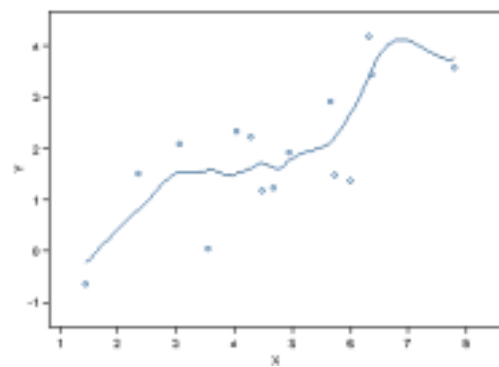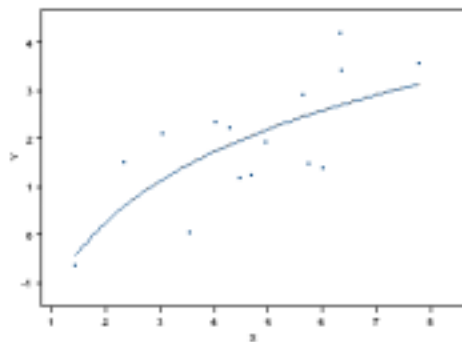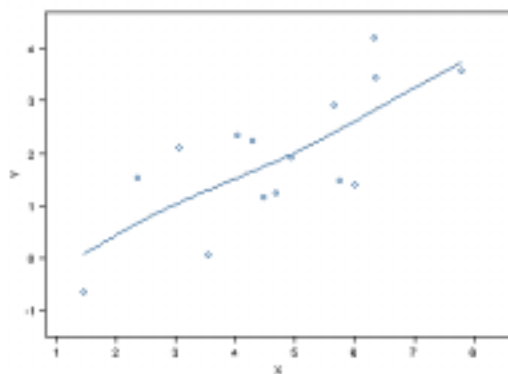X1,Y1,X2,Y2

▼Y zero density

▼Horiz. Density

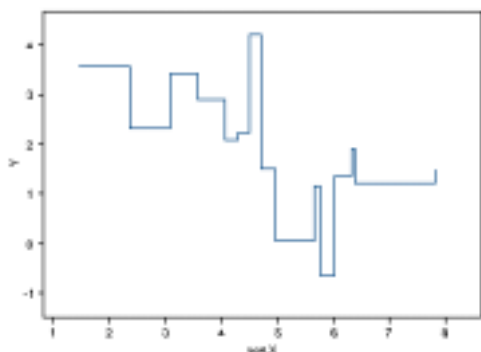▼Linear fit/ Polynomial curve fit /Exponential curve fit/ Power curve fit /Robust least trimmed square fit
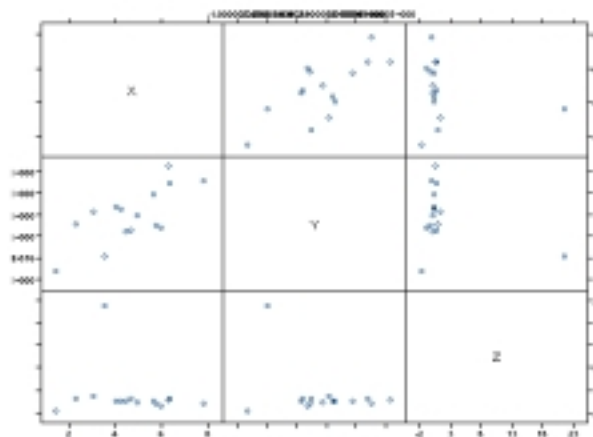
▼Log base e curve fit/ Log base 10 curve fit ▼Loess plot

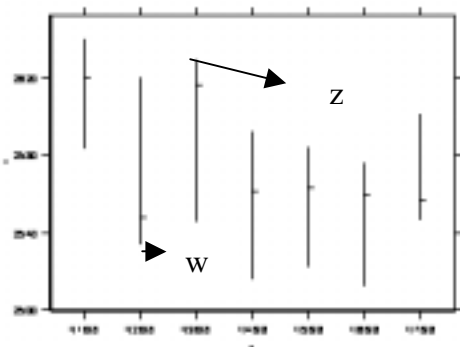▼Spline plot /Frieman super smoothing

▼Horiz. Step plot

▼Scatter plot matrix

▼High low plot

High-low plots typically display the daily, monthly, or yearly high and low values of a series, together with average or closing values, and perhaps the opening values. Meaningful high-low plots can thus include from three to five columns of data. The first column selected, containing the x data, is used to label the x-axis. Select the high and low data as the z- and w- columns. To create a high-low-open-close plot, select the open and/or close data as the y-column.

▼Candle stick plot
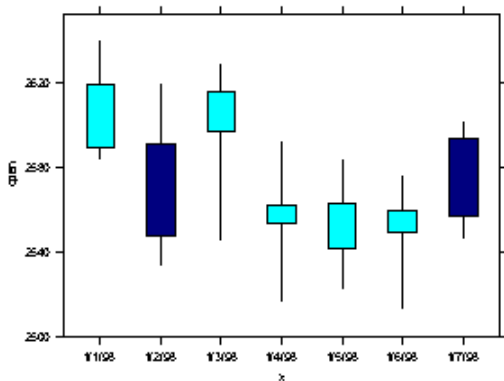
A candlestick plot is a variation on the High-Low-Open-Close plot that displays the difference between the Open and Close as a filled rectangle. The color of the rectangle shows whether the difference is positive or negative. The order of data columns selected is important. The first column selected, containing the x data, is used to label the x-axis. Select the open and close data as the y columns. High and low data are optional and correspond to the z- and w-columns.
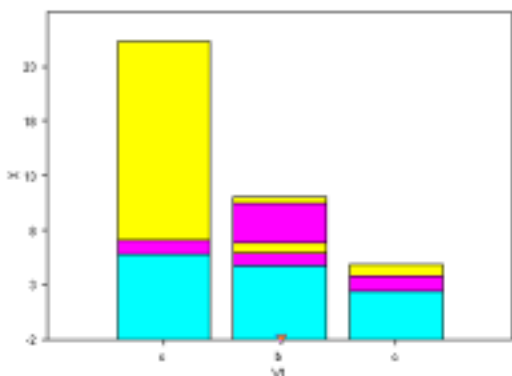
▼Bar with base at zero /Bar with base at y min.
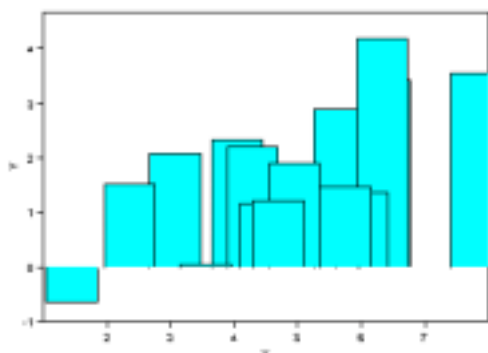
▼Grouped bar /Horiz. bar /Grouped horiz. Bar /Stacked horiz. Bar
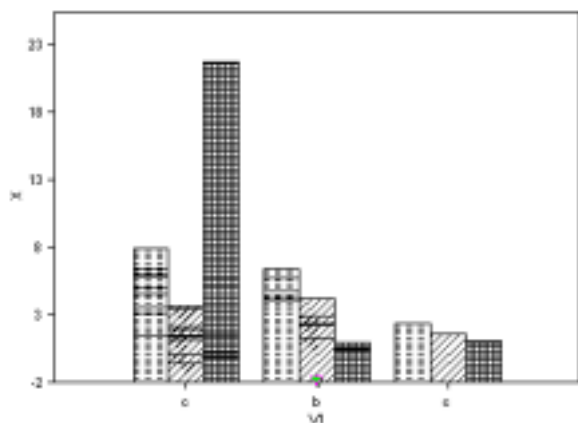
A grouped bar plot displays data as clusters of bars. The X values are the labels. Multiple y columns determine the bar heights.
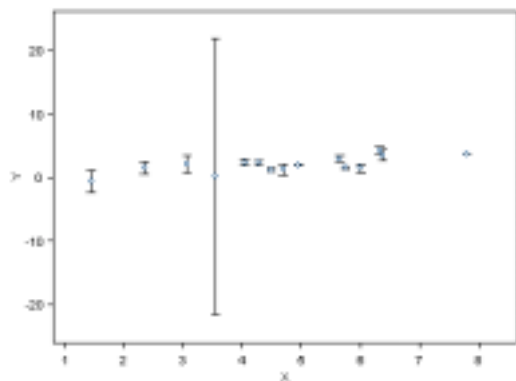　先選取 label 當 X, 其它 columns 當 Y
right-click x 軸, by bar 改變設定
double-click on x 改變 label。

▼Stacked bar

Stacked bar plot 以累積上去的 bars 來表示 data，X 的值是 labels， 多重的 Y columns 表示 bar segment 的高。

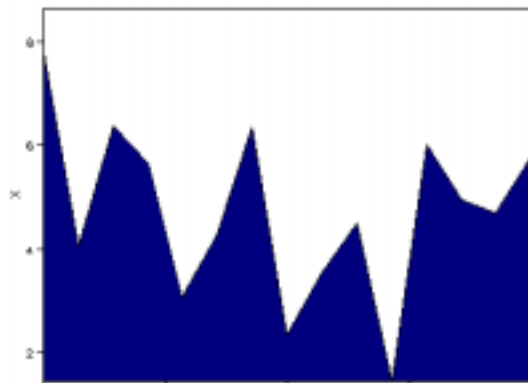▼Bar with error /Error bar-vert./ Error bar-horiz. /Error bar-both direction

X: double
Y: double
Z: double (error 項)
Error bar plots display a range of error around plotted data points. The x values determine the positions of the bars along the x-axis.
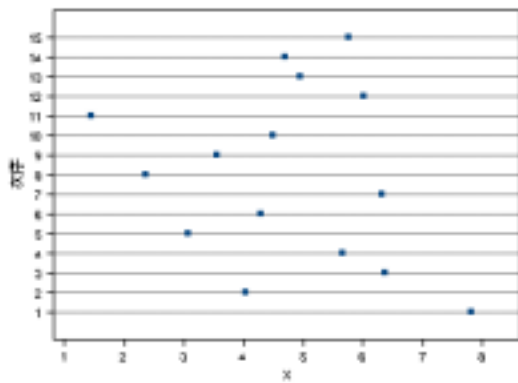
▼Area chart
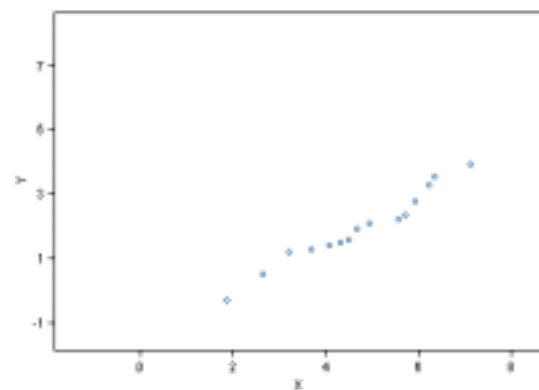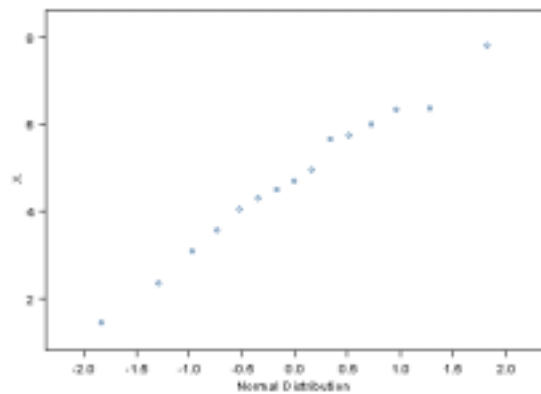
X: dates
Y1,Y2,..:double
Area charts are useful for showing how each series in a set of data affects the whole over time
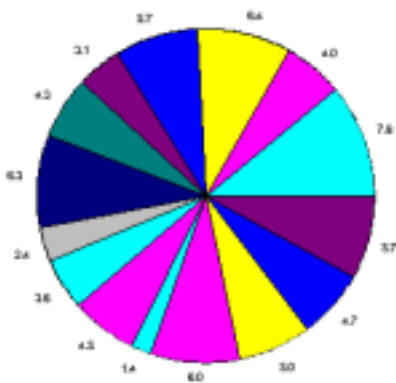
▼Dot plot

X: double
y: categories
Dot plots plot independent data against categorical dependent data using gridlines to mark the dependent levels
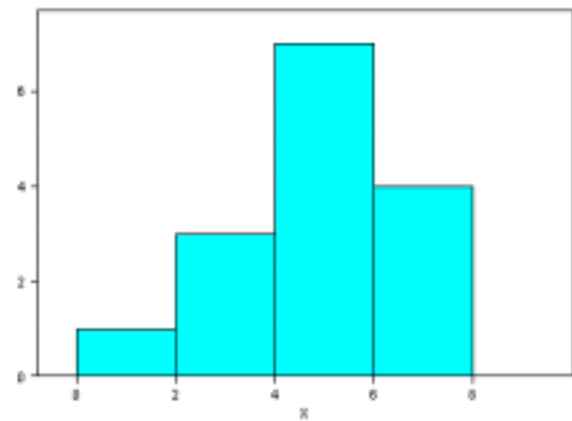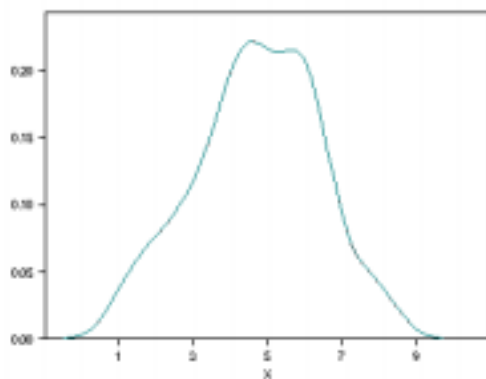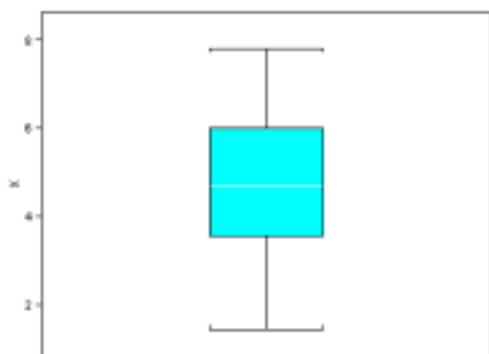
▼QQ plot

▼Pie

▼Histogram
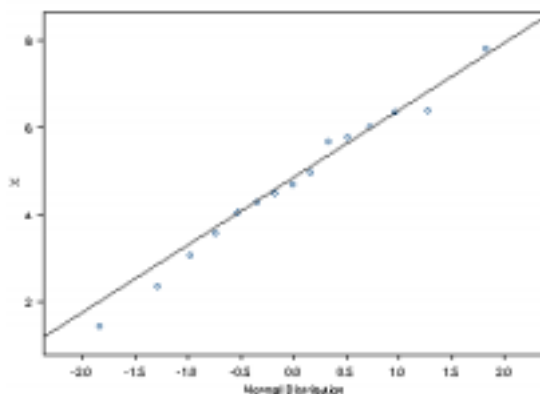

▼density /Histogram with density line

▼Box plot

Box plots display the location of
 ‧the median,
 ‧the upper and lower quartiles,
 ‧outer fences that indicate the extent of the
　data beyond the quartiles
of the distribution of one-dimension data.
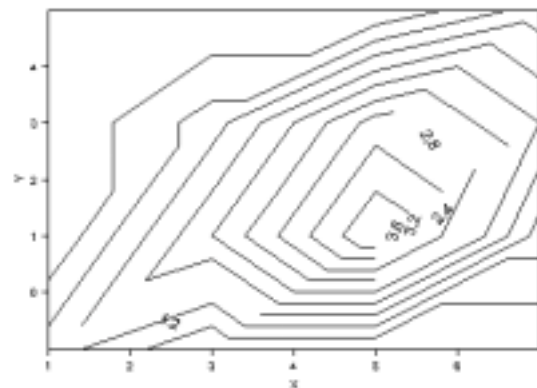
▼QQ normal with line

QQplot 是用來檢測一維 data 是否為 Normal
分配最初步的工具。
Compares your data with the quantiles of a
cumulative probability distribution function,
or compares the distributions of two data
series
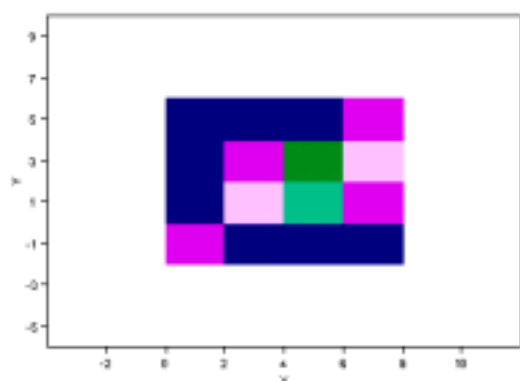選 data/滑鼠右鍵/可選 distribution

▼Contour plot /Filled contour

2D contour/levels plots are representations of
three-dimensional data in a two-dimensional
plane. Each contour line represents a level or
height from the corresponding
three-dimensional surface. Filled contour
plots use color between contour lines to
differentiate between the levels.
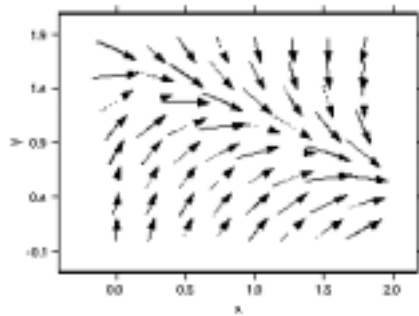
▼Levels plot

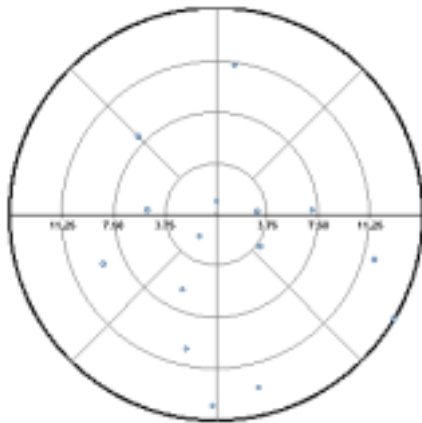X: x-axis
Y: y-axis
Z: levels (用顏色表示程度)

▼Vector plot

X,Y,Z: angle values
W: magnitude values
Vector plots display the direction and velocity
of flow at positions in the xy-plane. You can
also use vector plots to draw any group of
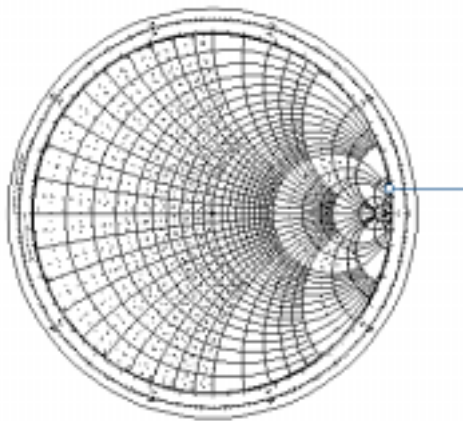arrows using data in a data set.

▼Polar scatter

X: radii
Y: angles
Polar plots display data in polar coordinates.
Specify radii for the x values and angles for
the y values.

▼Smith chart

Smith plots, which are drawn in polar
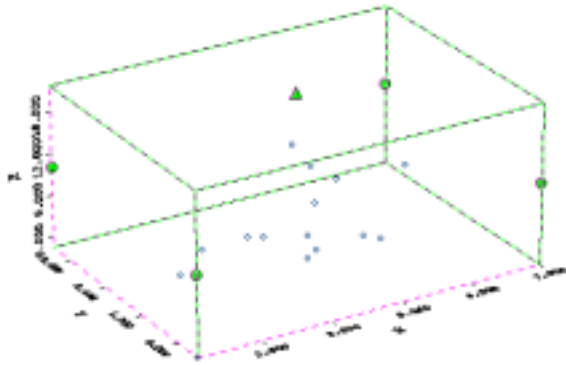coordinates, are often used in microwave
engineering to show impedance
characteristics.

3‧視覺化資料 3D

例：　Data set 同上例

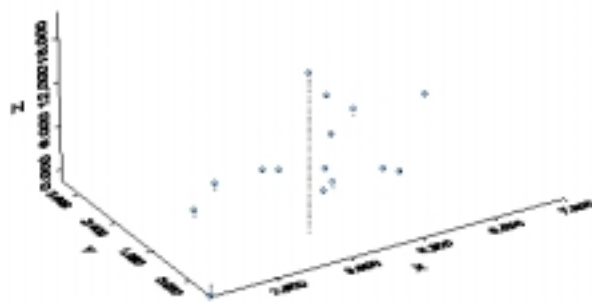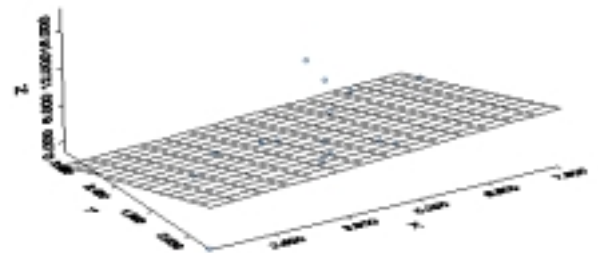▼3D scatter plot

▼3D line plot/3D line with scatter plot
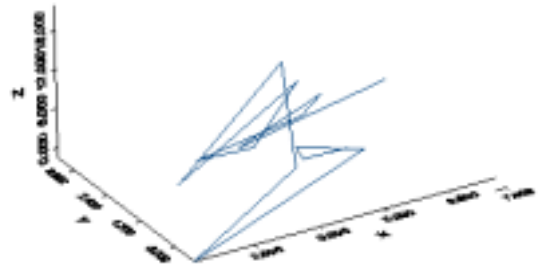
▼3D scatter with drop line

▼3D regression scatter/ 3D regression plot

▼Coarse surface /Filled coarse grid surface
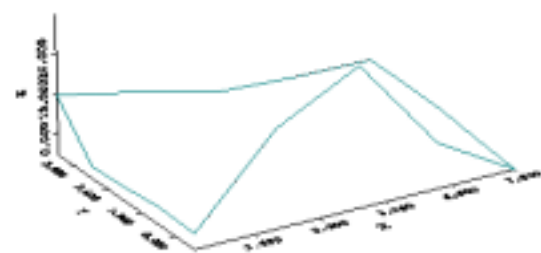
▼Data grid surface /Filled data grid surface
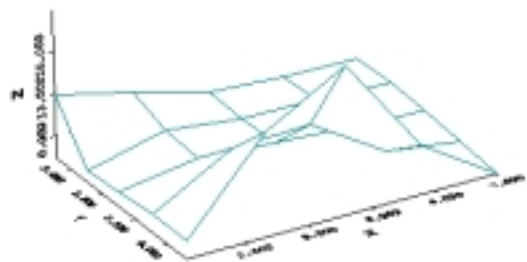
▼Spline fine grid surface /Filled spline fine grid
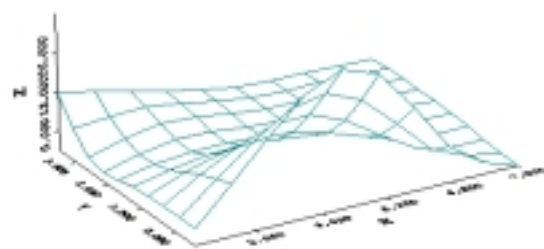
▼XY/XZ/YZ plane at Z/Y/X maximum
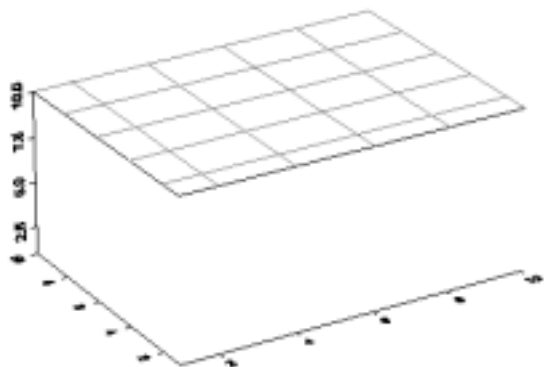
▼XY/XZ/YZ plane at Z/Y/X minimum
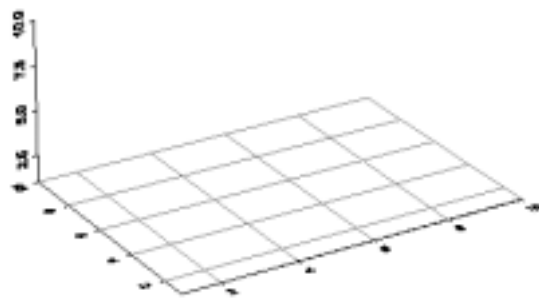
▼2/4/6 panel rotation
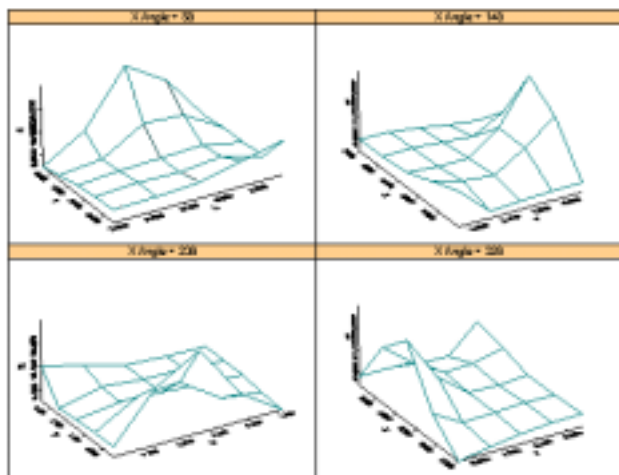
▼Panels conditioning on X/Y/Z
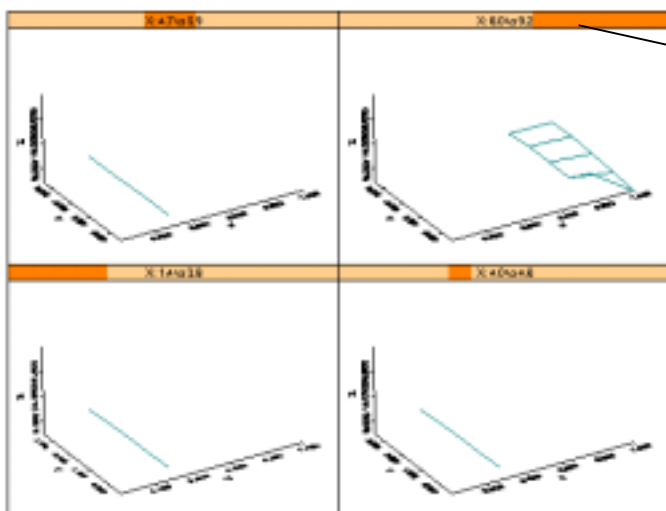
X 值的範圍

▼No conditioning/4/6 panel conditioning
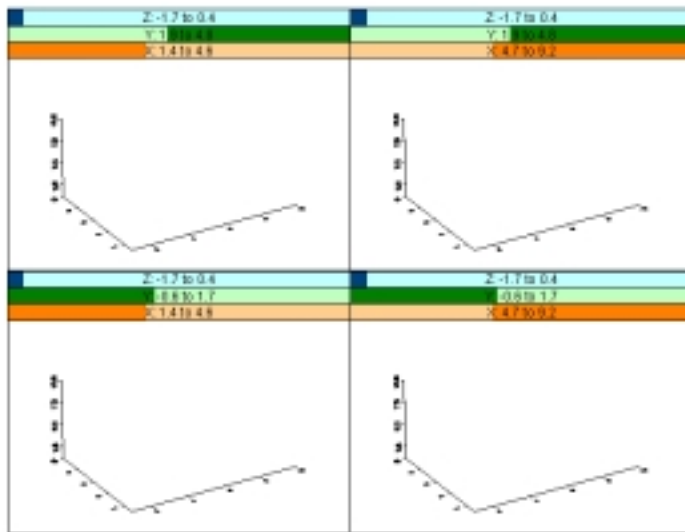
▼8/16/32 color draping surface
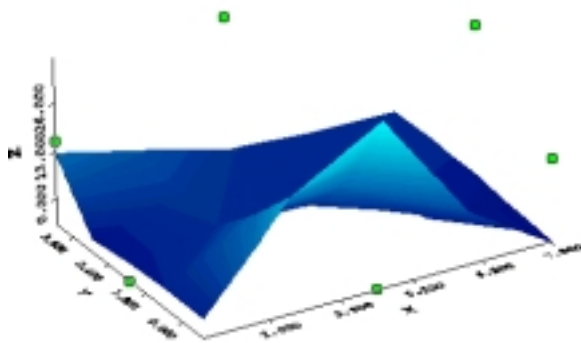
▼3D bar chart
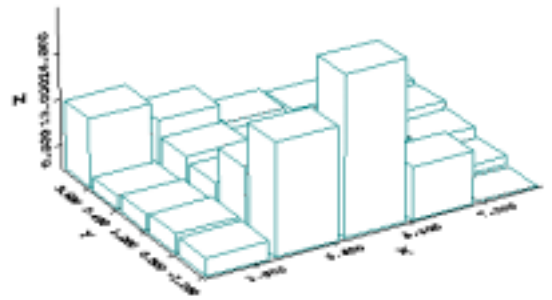
▼3D contour plot/3D filled contour
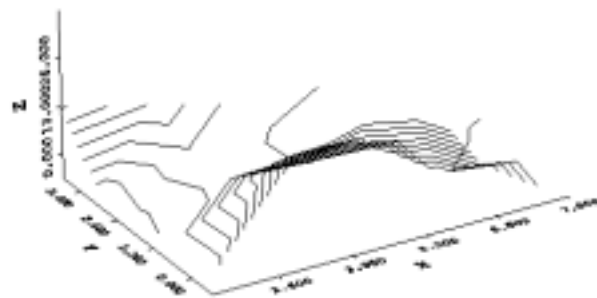
3D contour plots are identical to 2D contour plots except that the contour lines are drawn in three-dimensional space.

4. Trellis Graphs
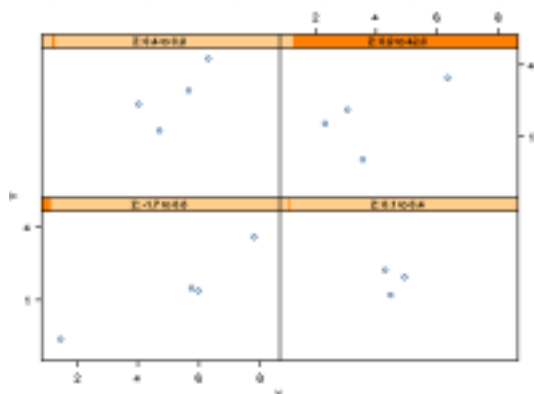
Trellis graphs let you view relationships between different variables in a data set through conditioning. A series of panels is displayed, with each panel containing a subset of the data divided into intervals of a conditioning variable.

1. 2D condition: X,Y,Z

2. 3D condition: X,Y,Z,W

5. More Graphs

5.1 Multipanel Graph
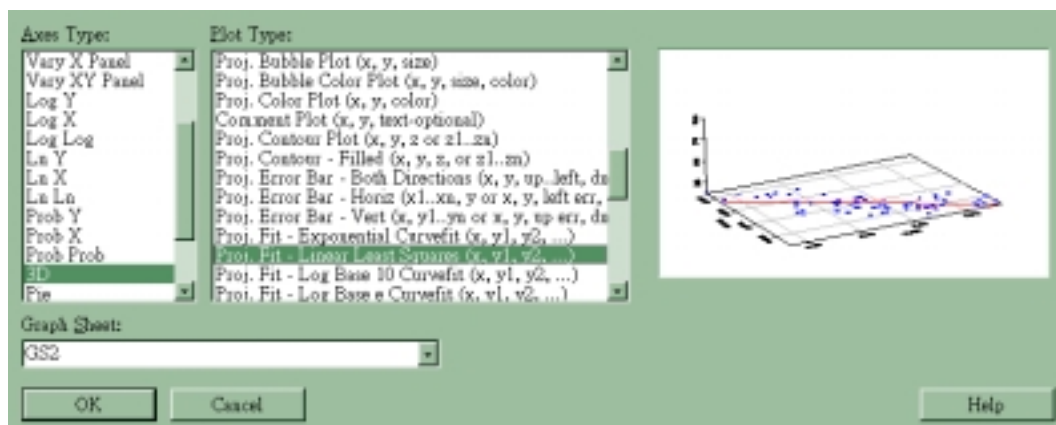
5.2 Brush and Spin (Dynamic Graphics)
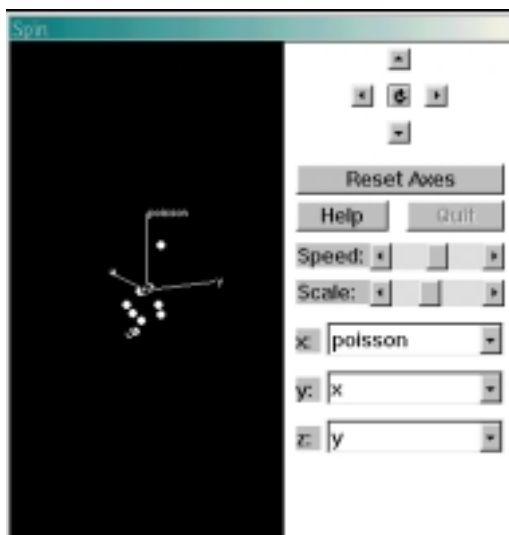
6. 圖形編輯與處理
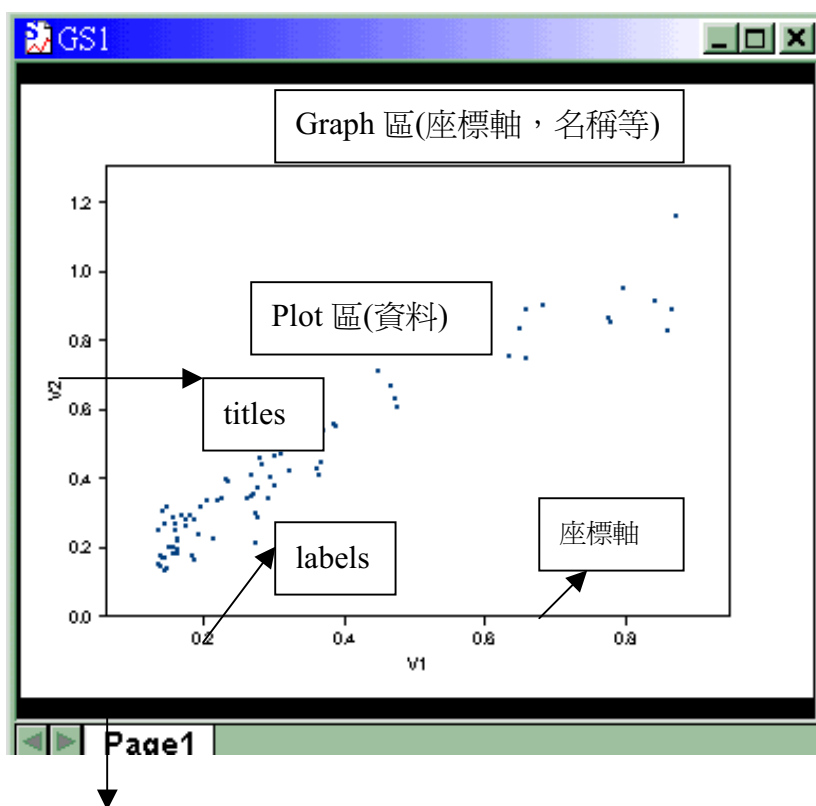
Graph sheet (可包含多個 graph)

6.0. Graph Tools 圖形工具列

- Select Tool: 選擇工作模式
- Label Point: 標 data point 的值
- Select Data Points: 選擇 data point (有和 Data set 做 Link)
- Crop Graph: 拖曳滑鼠選取欲細看的 data point (Zoom Out)
- Auto Scale Axes: 自動 scale 座標軸
- Pan Up/Right/Left/Down: 經 Crop 後，可以按 Pan 來上下左右來察看圖形
- Extract Panel: 從許多 panels 中選取一個 panel 來重畫
- Show All Panels: 秀出全部 panels
- No/4/9 Panels: 要畫 Panels 的個數
- Plots in Separate panels: Place each plot in separate panel
- Separate Panels with Varying Y/X/Both Axes:　Place each plot in separate panel varying Y/X/Both Axes
- Lower/Upper/Both X/Y Axes: 拖曳按鈕，增加座標軸

6.1. 選擇並 highlight 資料點 ：

6.2. Line fit with selected points deleted

    a. select points/change the cursor back

    b. Format/Exclude Selected Points

    c. Format/Include All Points (回復)

6.3. Multiple Plots on a single graph

    plot 可以全是同一 type 或不同 type 的組合，

    但這些 plot 必須有相同 type 的座標軸

    a. 選擇 graph region

    b. 選擇要 plot 的 data

    c. 按 Shift + plot palette button
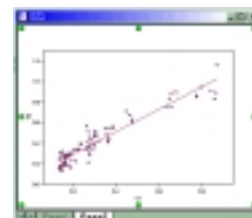
6.4. Multiple graphs on a graph sheet

    a. 在 graph sheet 上不用選按圖形

    b. 按 Shift 畫圖
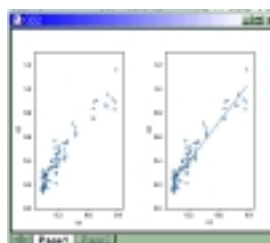
6.5. 在 graph sheet 上加一新頁

    a. 滑鼠指到 graph sheet 右下角 Page 1，按滑鼠右鍵選 Insert Page

6.6. 加主標題，子標題，座標軸名稱

    a. Insert/Titles/Main

    b. Insert/Titles/Subtitle

    c. click 座標點/Insert/Titles/Axis

6.7. 在圖上加 Legend，文字，符號

    a. Insert/Legend 或按

    b. Insert/Text

    c. Insert/Annotation 或按

6.8. 座標 labels 的位置

    a. 選擇座標 labels

    b. Drag the labels inside or outside the axis by dragging the triangular selection knob

6.9. 在圖上加時間

    Annotation\Date Stamp Tool button

6.10. 在圖上加 Curve fit equation

    a. Curve fit plot/選 data

    b. Insert/Curve Fit Equation

6.11. 在圖上加 Confidence Bounds

    a. 在圖一選任一 data point

　　b. Format/Select Curve Fitting Plot: By Conf Bound

　　c: 選合適的 Levels, Line Attributes: style, color and weight/OK

6.12. 座標軸的 rescale

　　a. 滑鼠點選 x-axis

　　b. green triangle 表示選了 axis labels，green knob 表示選了 x-axis line}

　　c. Double-click the x-axis line/Range/Axis Range, Tick Range/OK

6.13. Embedding and Extracting Data in Graph sheet

(only the variables actually displayed in the graph are embedded)

　　**Embedding Data** : 1. Graph/Embed Data　2.File/Save/**.sgr

　　ps1. changes to the original data set are not reflected in the Graph Sheet

　　ps2. can't use the Select Data Point Graph Tool if data are embedded

　　Extracting Data : Graph/Extract Data/Enter a name/OK

6.14. 連結 S-plus plots to data (when the data are likely to change)

　　OLE: Object Linking and Embedding

　　by default, data are linked to plots with automatic updatings.

　　a. Select the data in the source application

　　b. Copy the data to the clipboard

　　c. With S-plus plot selected and in focus

　　d. Edit/Paste

　　編輯連結: Edit/Links/Automatic

　　改變連結: Edit/Links/Change the name/OK

6.15. 匯出圖形到一個指定格式的檔案

　　File/Export Graph/Save as Type/File Name/Save
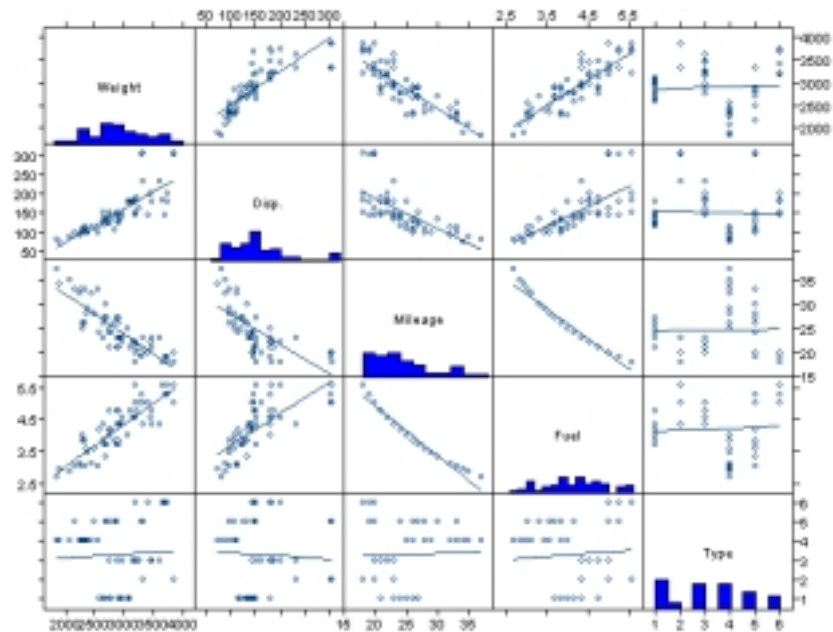
7. 例題

7.1. 2D+Legend

dataset: exfuel.sdd

a. click weight, then

　　Ctrl-click Mileage and Type

b. 2D plot button

c. Auto Legend
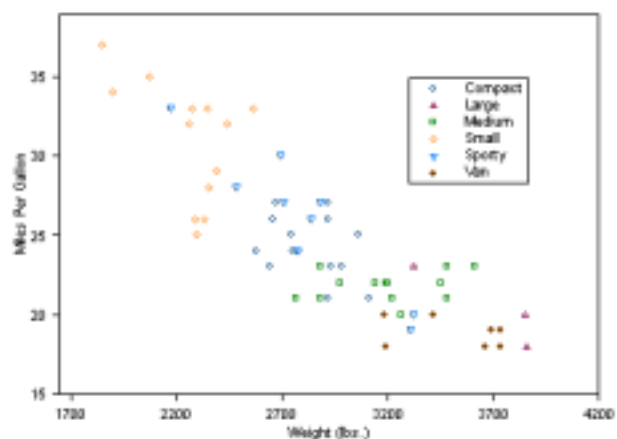
7.2. Scatterplot Matrix + Histogram + Line fit

a. select all data

b. Scatter Matrix button

c. right-click any data point: Line/Histogram/Draw Histogram/OK

d. right-click any data point: Smooth/Smoothing type/OK

7.3 Bubble Color Plots + Color Scale Legend

data set: exsensor

a. 選前四個 columns of the data

b. 按 Bubble Color button

c. V1 為 x-axis, V2 為 y-axis, The size of the symbols will vary with the value of V3, The symbol color will vary with the value of V4

d. 按 Color Scale Legend, add a legend relating the V4 value to the color of each point