# Scatter Plots

A scatter plot (Chambers 1983) reveals relationships or association between two variables. **The relationship between two variables is called *correlation*.** A scatter plot usually consists of a large body of data. The closer the data points come when plotted to making a straight line, the higher the correlation between the two variables, or the stronger the relationship. If the data points make a straight line going from the origin out to high x- and y-values, then the variables are said to have a **positive correlation**. If the line goes from a high-value on the y-axis down to a high-value on the x-axis, the variables have a **negative correlation**.
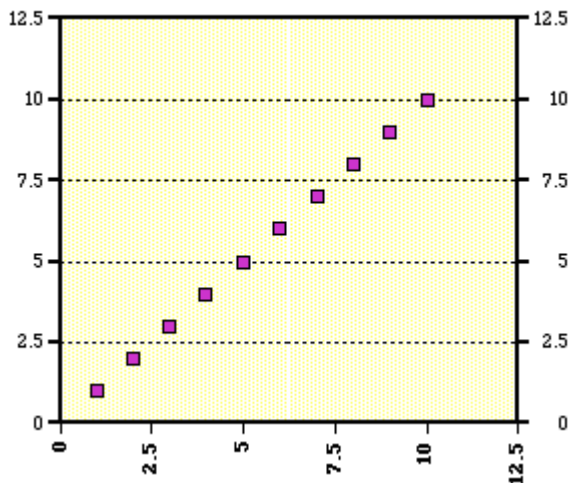
No relationship: If there is absolutely no correlation present the value given is 0.
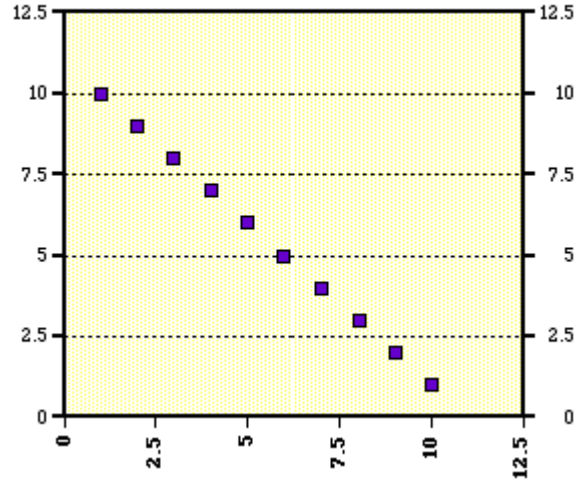

No Correlation
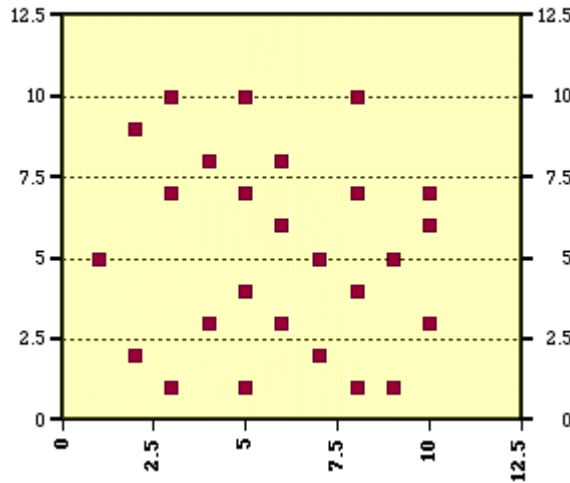
Perfect linear correlation: A perfect positive correlation is given the value of 1. A perfect negative correlation is given the value of -1.
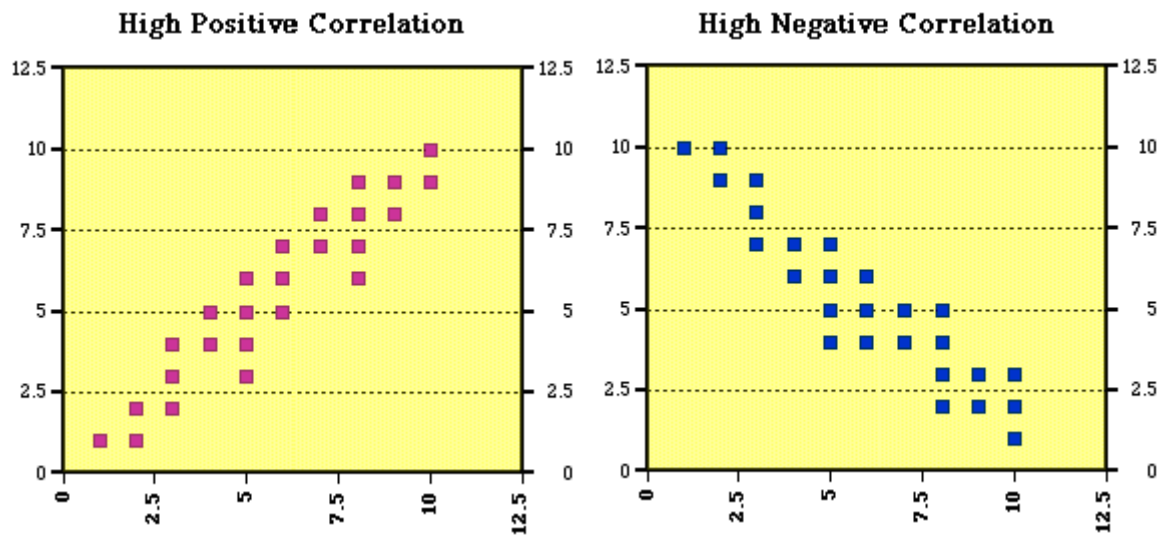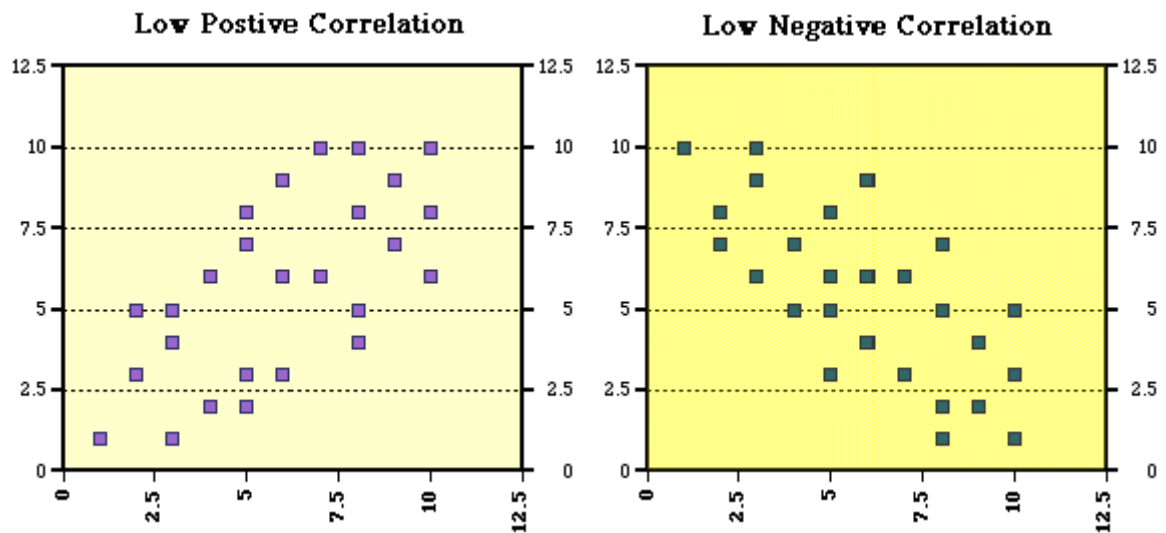

Perfect Positive Correlation

Perfect Negative Correlation

Strong linear correlation: The closer the number is to 1 or -1, the stronger the correlation, or the stronger the relationship between the variables.


Weak linear correlation: The closer the number is to 0, the weaker the correlation.

Scatter plot matrix: Given a set of variables, the scatter plot matrix contains all **the pair-wise scatter plots of the variables** on a single page in a matrix format. The example generated by XmdvTool shows a 4x4 scatter plot matrix of the variables medhvalue (median house value), rooms (# of rooms), bedrooms (# of bedrooms), and households (# of households).

Top row in the graph shows 1. the scatter plot of medhvalue and rooms, 2. the scatter plot of medhvalue and bedrooms, and 3. the scatter plot of medhvalue and households. The second row in the graph shows 4. the scatter plot of rooms and bedrooms and 5. the scatter plot of rooms and households. The third row in the graph shows 6. the scatter plot of bedrooms and households.

This scatter plot matrix shows that rooms, bedrooms, and households are highly correlated. Note that when there is a high correlation between A and B and between A and C, there is a high correlation between B and C.
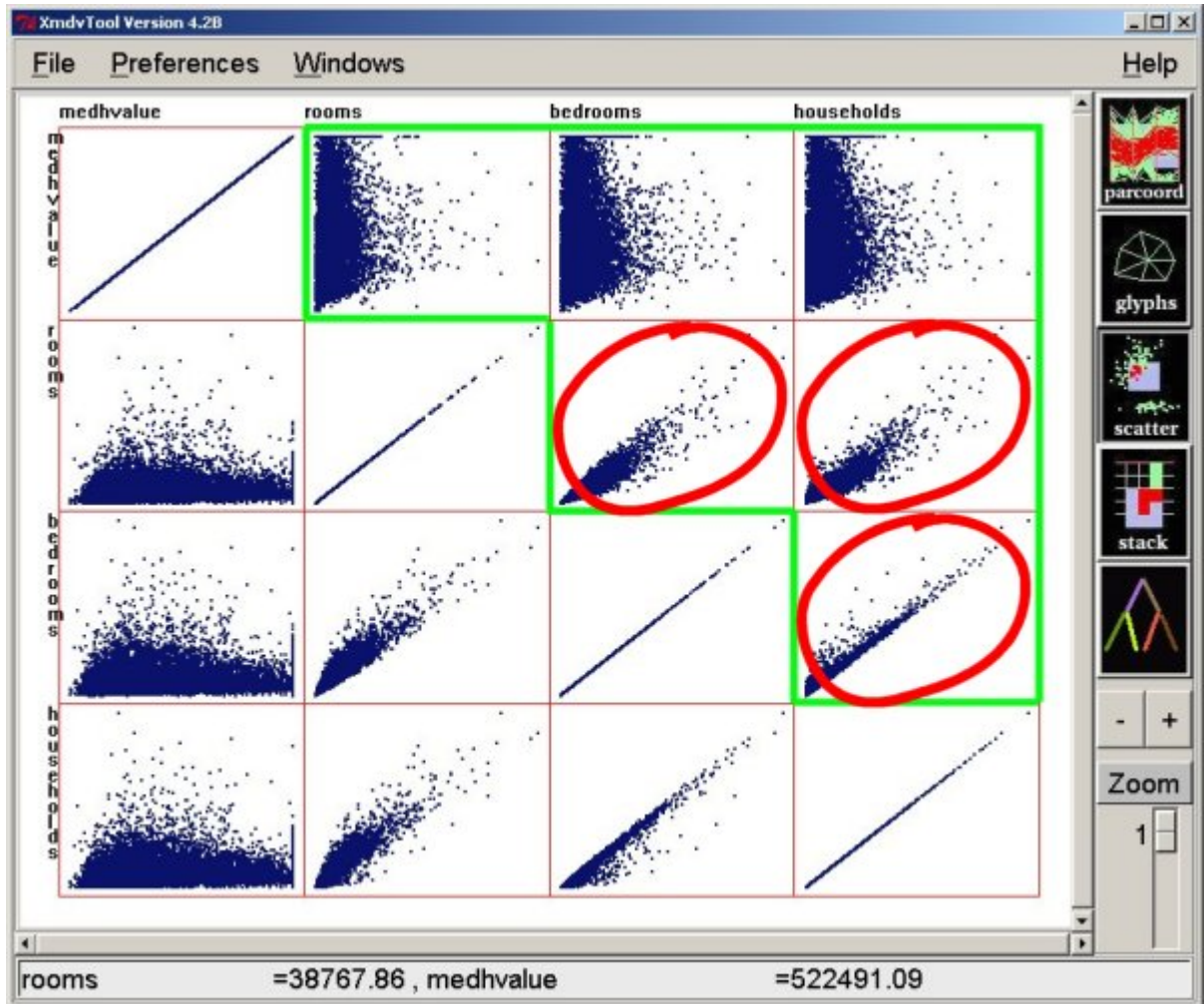
Reference:

Chambers, John, William Cleveland, Beat Kleiner, and Paul Tukey, (1983), *Graphical Methods for Data Analysis*, Wadsworth.

http://www.mste.uiuc.edu/courses/ci330ms/youtsey/scatterinfo.html

http://www.itl.nist.gov/div898/handbook/eda/section3/scatterp.htm

http://gsbapp2.uchicago.edu/sas/sashtml/insight/chap5/sect3.htm

<-Training