Cluster Analysis

Average Linkage Clustering:

Patterns: Humans are good at discerning subtle patterns but also imagining them (Carl Sagan).

Bivariate: Two variables.

Multivariate: Many variables.

Scatterplot Matrix: Symmetric grid of bivariate scatterplots. Works well when there are not many variables.

Principal Component Analysis: When there are many variables, scatterplot matrixes must be adapted with a technique to preserve their multivariate structure. PCA transforms the variables into uncorrelated variables for better visualizing in lower dimensional space.

Multidimensional Scaling: Representing similarities or differences using measure of distance.

Swadesh: Set of most basic words that can be expanded into similar word groupings.

Mixed Mode Data: Data with different measurement types (categorical, ordinal, continuous).

Clustering programs: SPSS, Stata, R (cluster, clusterSim, proxy, daisy), SAS.