# Visualization and Benchmarking of Cluster Solutions

Friedrich Leisch

Universität für Bodenkultur Wien

Centroid-based partitioning cluster analysis is a popular method for segmenting data into more homogeneous subgroups. Visualization can help tremendously to understand the positions of these subgroups relative to each other in higher dimensional spaces and to assess the quality of partitions. In this talk we present several improvements on existing cluster displays using neighborhood graphs with edge weights based on cluster separation and convex hulls of inner and outer cluster regions. Using symbols or complete high-level plots in the nodes of the graph help to understand the association of background variables and clusters.

Visualizations can also help to assess the stability of cluster solutions or compare different partitions. By sampling from the empirical distribution of a given data set and clustering each of these bootstrap samples, a sample of iid observations of complete partitions is derived. Interesting characteristics of these partitions can then be examined using standard techniques from exploratory data analysis and inferential statistics, most importantly standard statistical test procedures. To demonstrate the usefulness in practice, the theoretical results are applied to benchmark studies based on artificial and real world data.