Multivariate visualization for the interpretation of clustering results

Ursula Laa

Statistics Institute, University of Natural Resources and Life Sciences

Abstract

Clustering is a useful tool in exploratory data analysis, helping us with interpretation by introducing structure and presenting us with representative benchmark points to study and compare in more detail. However, the cluster assignment will depend on many choices made by an analyst: data preprocessing (e.g. scaling), clustering algorithm, distance metric, linkage, or the number of clusters in the final result. Often there is no one perfect solution, and exploring these settings interactively, paired with multivariate visualization of the results, will provide a better understanding of the structures in the data. In this talk I will show how this was implemented in an interactive Shiny app, pandemonium, that explores clustering outcomes in data from particle physics.