R SubtypeDiscovery

a data mining scenario designed to infer subtypes in data by cluster analysis

data preparation

scale data
adjust data
*select features
select samples
visualize for EDA

mixture model clustering

model based
clustering (R mclust)
repeat EM

model selection

model ranking
BIC quantile statistics
top N models

characterization and comparison

visualize heatmaps,
parallel coordinates
and series
cross compare models
*mean difference testing

subtypes validation

application specific
relevance tests
(risk ratio, X2, ...)

*in development