|  |  |
| --- | --- |
| VisuClust {VisuClust} | R Documentation |

**VisuClust**

**Description**

The VisuClust package provides functions to display the result of clusterings, if they are either disjoint (every observation belongs to one cluster) or fuzzy (for each observation i and cluster k there is a value u[i,k](i=1,...,nPoints; k=1,...,nClusters)), which states the membership of the i.th observation for the cluster k, where sum(u[i,k],k=1,...,nClusters) = 1) .

A disjoint clustering can be created with known hierarchical methods, if the desired number of clusters is provided, or by using k-means-clustering (for example [hclust](http://127.0.0.1:20022/library/VisuClust/help/hclust), [kmeans](http://127.0.0.1:20022/library/VisuClust/help/kmeans) from the package cluster). A fuzzy clustering can be created for example with fanny from the package cluster.

The basis of the visualization is Sammon's method Nonlinear Mapping (NLM), which displays all observations in a plane so that the distances in the multivariate space are approximated best (see sammon from the package MASS). The clusters in the NLM diagram can be indicated by coloring, drawing lines between similar point pairs, and other features.

**Author(s)**

Michael Sieger <michael.sieger@student.hswt.de>   
Georg Ohmayer <georg.ohmayer@hswt.de>

**References**

Will be published in the GIL conference transcript 2012:   
Visualisierung von Clustern in multivariaten Daten unter Einsatz von R   
Georg Ohmayer, Michael Sieger

|  |  |
| --- | --- |
| LinkageMap {VisuClust} | R Documentation |

**Displays a Linkage Map**

**Description**

A Linkage Map is a 2D Scatter Plot that displays the result of Sammon's Nonlinear Mapping. A number of ranges (>= 1) for the distance values d[i,j] can be defined. Each point pair whose distance is inside of the defined range is connected with a line.

The ranges are defined by thresholds t[k](1<=k<=3) in the following way   
range[1]: d[i,j] <= t[1]   
range[k]: t[k-1] < d[i,j] <= t[k]

The ranges can be changed dynamically at runtime by changing the thresholds t[k] with sliders.

A estimation of the density function for the distances d[i,j] (for example with kernel density estimation, see function density) is helpful for finding the ideal thresholds t[k] (=local minima in the density function).

**Usage**

LinkageMap(xSammon, dist, lineTypes=c("solid","dotted", "dashed"), lineColors=c("red","green","blue"), lineWidths=c(1,1,1),labels = NULL, cluster = NULL, maxValue=0.5, legendDigits = 2, xlab = "", ylab = "", main = "")

**Arguments**

|  |  |
| --- | --- |
| xSammon | A matrix with size (nPoints,2) that contains the projected points from Sammon's Nonlinear Mapping. |
| dist | A distance matrix for the given data (to compute with [dist](http://127.0.0.1:20022/library/VisuClust/help/dist)). |
| lineTypes | An array of line types. The size must match with line colors and line widths. The various line types are described in [par](http://127.0.0.1:20022/library/VisuClust/help/par) |
| lineColors | An array of colors. The size must match with line types and line widths. see [colors](http://127.0.0.1:20022/library/VisuClust/help/colors) |
| lineWidths | An array of line widths. The size must match with line colors and line types. See [par](http://127.0.0.1:20022/library/VisuClust/help/par) |
| labels | A string-array with labels for the points (size: nPoints). |
| cluster | A array containing cluster memberships of the points. The point membership will be indicated with different colors. The array size must match with nPoints. |
| maxValue | maxValue\*max(d[i,j]) is the maximum value that can be adjusted with the sliders. |
| legendDigits | The number of fractional digits to be displayed in the legend. |
| xlab | A title for the x axis (as described in plot). |
| ylab | A title for the y axis (as described in plot). |
| main | An overall title for the plot (as described in plot). |

**Author(s)**

Michael Sieger <michael.sieger@student.hswt.de>   
Georg Ohmayer <georg.ohmayer@hswt.de>

**Examples**

library(MASS)

data("MilchSmall")

M <- scale(MilchSmall[3:6])

D <- dist(M)

S <- sammon(D)

K <- kmeans(D,center=5)

LinkageMap(S$points, D, cluster=K$cluster, labels=MilchSmall[,2])

|  |  |
| --- | --- |
| FuzzyPlot {VisuClust} | R Documentation |

**Displays a Fuzzy Plot**

**Description**

A Fuzzy Plot is a 2D Scatter Plot that displays the result of Sammon's Nonlinear Mapping together with a fuzzy clustering. The values u[i,k] (i=1,...,nPoints; k=1,...,nPoints), which states the membership of the i.th observation for the cluster k, can be computed for example with fanny from the package cluster.

If the slider is on the last position, all cluster are displayed together. The color and the symbol indicates the nearest crisp clustering then (max u[i,k], k=1,...,nClusters). The color intensity displays the probability of membership to the nearest cluster.

A single cluster can be selected too by setting the slider in position 1 to nClusters. The color intensity and the label size (if labels are set) shows the probability of membership to the selected cluster.

**Usage**

FuzzyPlot(xSammon, probs, clusterColors=rainbow(dim(probs)[2]), clusterSymbols=rep(21,dim(probs)[2]), labels=NULL, labelSize=c(0.6, 1.0), xlab="", ylab="", main="", enableLegend=TRUE, cex=1.4)

**Arguments**

|  |  |
| --- | --- |
| xSammon | A matrix with size (nPoints,2) that contains the projected points from Sammon's Nonlinear Mapping. See sammon (package MASS). |
| probs | A matrix with size (nPoints, nClusters) that contains the probabilities of membership for each cluster. See the membership argument of [fanny](http://127.0.0.1:20022/library/VisuClust/help/fanny) (package cluster) |
| clusterColors | A vector of size nClusters that contains the colors for the clusters. One should only choose colors with a very high intensity. See [colors](http://127.0.0.1:20022/library/VisuClust/help/colors) |
| clusterSymbols | A vector of size nClusters that contains the symbols for the clusters. Not all symbols can be filled. If this is wanted, one should use the symbols 15-20. See [points](http://127.0.0.1:20022/library/VisuClust/help/points) |
| labels | A string-array with labels for the points (size: nPoints). |
| labelSize | The size of the labels in the form c(min, max). |
| xlab | A title for the x axis (as described in [plot](http://127.0.0.1:20022/library/VisuClust/help/plot)). |
| ylab | A title for the y axis (as described in [plot](http://127.0.0.1:20022/library/VisuClust/help/plot)). |
| main | An overall title for the plot (as described in plot) |
| enableLegend | Switch legend on/off. |
| cex | A numerical value giving the amount by which plotting text and symbols should be magnified relative to the default (as described in par). |

**Author(s)**

Michael Sieger <michael.sieger@student.hswt.de>   
Georg Ohmayer <georg.ohmayer@hswt.de>

**Examples**

library(cluster)

library(MASS)

# The example data

data("MilchSmall")

M <- scale(MilchSmall[3:6])

D <- dist(M)

S <- sammon(D)

F <- fanny(data, 6)

FuzzyPlot(S$points, F$membership, labels=MilchSmall[,2], clusterColors=c("red", "green", "magenta", "yellow", "blue", "black"))

|  |  |
| --- | --- |
| MilchBig {VisuClust} | R Documentation |

* 1. **Milk components of 50 mammals**
     1. **Description**

This dataset contains the fat, protein, lactose and ash of different milks.

* + 1. **Usage**

MilchBig

**Author(s)**  
Georg Ohmayer <georg.ohmayer@hswt.de>, Herbert Seiler

**References**

Ohmayer G., Seiler H. (1985): Numerische Gruppierung und graphische Darstellung von Daten: Ein Methodenvergleich; EDV in Medizin und Biologie, Vol. 2, S. 65-73

|  |  |
| --- | --- |
| MilchSmall {VisuClust} | R Documentation |

* 1. **Milk components of 20 mammals**
     1. **Description**

This dataset contains the fat, protein, lactose and ash of different milks. This is a reduced version (20 instead of 50 observations).

* + 1. **Usage**

MilchSmall

**Author(s)**  
Georg Ohmayer <georg.ohmayer@hswt.de>, Herbert Seiler