Package 'FactoClass'

January 31, 2024

Version 1.2.9

Date 2024-01-24

Title Combination of Factorial Methods and Cluster Analysis

Author Campo Elias Pardo <cepardot@unal.edu.co>,
Pedro Cesar del Campo <pcdelcampon@unal.edu.co> and
Camilo Jose Torres <cjtorresj@unal.edu.co>,
with the contributions from.
Ivan Diaz <ildiazm@unal.edu.co>,
Mauricio Sadinle <msadinleg@unal.edu.co>,
Jhonathan Medina <jmedinau@unal.edu.co>.

Maintainer Campo Elias Pardo <cepardot@unal.edu.co>

Description Some functions of 'ade4' and 'stats' are combined in order to obtain
a partition of the rows of a data table, with columns representing
variables of scales: quantitative, qualitative or frequency.
First, a principal axes method is performed and then, a combination
of Ward agglomerative hierarchical classification and K-means is
performed, using some of the first coordinates obtained from the
previous principal axes method.
In order to permit different weights of the elements to be clustered,
the function 'kmeansW', programmed in C++, is included.
It is a modification of 'kmeans'. Some graphical functions include
the option: 'gg=FALSE'. When 'gg=TRUE', they use the 'ggplot2'
and 'ggrepel' packages to avoid the super-position of the labels.

Depends R (>= 2.10), ade4,ggplot2,ggrepel,xtable,scatterplot3d

Imports KernSmooth

License GPL (>= 2)

Encoding UTF-8

NeedsCompilation yes

Repository CRAN

Date/Publication 2024-01-31 14:40:12 UTC

R topics documented:

`addgrids3d` *Add grids to a scatterplot3d (modified)*

Description

The goal of this function is to add grids on an existing plot created using the package scatterplot3d

Usage

```
addgrids3d(x, y = NULL, z = NULL, grid = TRUE, col.grid = "grey",
  lty.grid = par("lty"), lab = par("lab"), lab.z = mean(lab[1:2]),
  scale.y = 1, angle = 40, xlim = NULL, ylim = NULL, zlim = NULL)
```

Arguments

x, y, z	numeric vectors specifying the x, y, z coordinates of points. x can be a matrix or a data frame containing 3 columns corresponding to the x, y and z coordinates. In this case the arguments y and z are optional
grid	specifies the facet(s) of the plot on which grids should be drawn. Possible values are the combination of "xy", "xz" or "yz". Example: grid = c("xy", "yz"). The default value is TRUE to add grids only on xy facet.
col.grid, lty.grid	color and line type to be used for grids
lab	a numerical vector of the form c(x, y, len). The values of x and y give the (approximate) number of tickmarks on the x and y axes.
lab.z	the same as lab, but for z axis
scale.y	of y axis related to x- and z axis
angle	angle between x and y axis
xlim	the x limits (min, max) of the plot
ylim	the y limits (min, max) of the plot
zlim	the z limits (min, max) of the plot.

Note

Users who want to extend an existing scatterplot3d graphic with the function addgrids3d, should consider to set the arguments scale.y, angle, ..., to the value used in scatterplot3d.

Author(s)

Alboukadel Kassambara <alboukadel.kassambara@gmail.com>

References

http://www.sthda.com

Examples

```
library(scatterplot3d)
data(iris)
scatterplot3d(iris[, 1:3], pch = 16, grid=TRUE, box=FALSE)
addgrids3d(iris[, 1:3], grid = c("xy", "xz", "yz"))
```

admi *Admitted students to the seven careers of the Science Faculty*

Description

Score obtained by each of the 445 students admitted to the seven careers of the Facultad de Ciencias of the Universidad Nacional de Colombia Bogota to the first semester of 2013, and some socio demographic information:

carr a factor with the careers as its levels

mate,cien,soci,text,imag score achieved in each of the areas of the admission exam

exam total score of the admission exam

gene gender of the admitted

estr socioeconomic stratum in 3 categories

orig geographic origin of the admitted

edad age of the admitted in categories

niLE if the admitted requires nivelation in language

niMa if the admitted requires nivelation in mathematics

estr socioeconomic stratum in 7 categories

age age of the admitted in years

Usage

```
data(admi)
```

Format

Object of class `data.frame` with 445 rows and 15 columns.

Source

SIA: Academic Information System

References

C.E. Pardo (2015). Estadística descriptiva multivariada. Universidad Nacional de Colombia. Facultad de Ciencias.

Bogota *Localities by Stratums in Bogota City*

Description

Contingency Table that indicates the number of blocks of Bogota, in localities by stratums (DAPD 1997, p.77).

Usage

```
data(Bogota)
```

Format

Object whit class `data.frame` of 19 rows and 7 columns.

Source

DAPD (1997), Population, stratification and socioeconomic aspects of Bogota

References

C.E. Pardo y J.E. Ortiz (2004). Analisis multivariado de datos en R. Simposio de Estadistica, Cartagena Colombia.

cafe *Cofee cups*

Description

Results of the mesure of some properties of twelve coffe cups

Usage

```
data(cafe)
```

Format

Object of class `data.frame` with 12 rows and 16 columns.

Source

R. Duarte and M. Suarez and E. Moreno and P. Ortiz (1996). An\'alisis multivariado por componentes principales, de caf\'es tostados y molidos adulterados con cereales. Cenicaf\'e, 478(2):65-76

References

C.E. Pardo (2023). Estadistica descriptiva multivariada. Universidad Nacional de Colombia. Facultad de Ciencias.

`centroids` *Centroids of the Classes of a Partition*

Description

It evaluates the centroids of a partition with the weights in `rw`

Usage

```
centroids(df,cl,rw=rep(1/nrow(df),nrow(df)))
```

Arguments

`df`	object of class `data.frame`, with the data of variables or coordinates
`cl`	vector indicating the cluster of each element
`rw`	weight of the rows of df, by default the same

Value

Object of class `list` with the following:

`centroids`	class centroids
`weights`	class weights
`cr`	correlation ratios

Author(s)

Campo Elias Pardo <cepardot@unal.edu.co>

Examples

```
data(iris)
centroids(iris[,-5],iris[,5])
```

`chisq.carac` *Chisquare tests of a qualitative variable by several qualitative variables*

Description

Chisqure tests are performed for the contingency tables crossing a qualitative variable named `cl` and the qualitative variables present in columns from `df`

Usage

```
chisq.carac(df,cl,thr=2,decr=TRUE)
```

Arguments

`df`	`data.frame`, with factors contain the categories of the qualitave variables
`cl`	factor indicating the category of each subject
`thr`	threshold of test value, if decr=TRUE, only the rows where `tval` >= `thr` are returned
`decr`	if decr=TRUE the rows are returned in decreasing order

Value

Matrix with the following columns:

`chi2`	chisquare statistic
`dfr`	degree of fredom of chisquare densidity
`pval`	p value
`tval`	quantil `qnorm(pval,lower.tail=FALSE`
`phi2`	`phi2=chi2/n`

Author(s)

Campo Elias Pardo <cepardot@unal.edu.co>

Examples

```
data(DogBreeds)
round(chisq.carac(DogBreeds[,-7],DogBreeds[,7]),3)
round(chisq.carac(DogBreeds[,-7],DogBreeds[,7],decr=FALSE),3)
```

cluster.carac *Cluster Characterization by Variables*

Description

It makes the characterization of the classes or cluster considering the variables in tabla. These variables can be quantitative, qualitative or frequencies.

Usage

```
cluster.carac( tabla,class,tipo.v="d",v.lim= 2,dn=3,dm=3,neg=TRUE)
```

Arguments

tabla	object data.frame with variables of characterization, the variables must be of a single type (quantitative, qualitative or frequencies)
class	vector that determines the partition of the table
tipo.v	type of variables: quantitative("continuas"), qualitative ("nominales") or frequencies("frecuencia")
v.lim	test value to show the variable or category like characteristic.
dn	number of decimal digits for the p and test values.
dm	number of decimal digits for the means.
neg	if neg=TRUE, the variables or categories with negative test values are showed.

Details

For nominal or frecuency variables it compares the percentage of the categories within each class with the global percentage. For continuous variables it compares the average within each class with the general average. Categories and variables are ordered within each class by the test values and it shows only those that pass the threshold v.lim.

Value

Object of class list. It has the characterization of each class or cluster.

Author(s)

Pedro Cesar del Campo <pcdelcampon@unal.edu.co>, Campo Elias Pardo <cepardot@unal.edu.co>, Mauricio Sadinle <msadinleg@unal.edu.co>

References

Lebart, L. and Morineau, A. and Piron, M. (1995) Statisitique exploratoire multidimensionnelle, Paris.

Examples

```
data(DogBreeds)
DB.act <- DogBreeds[-7]  # active variables
DB.function <- subset(DogBreeds,select=7)
cluster.carac(DB.act,DB.function,"ca",2.0)  #  nominal variables


data(iris)
iris.act <- Fac.Num(iris)$numeric
class <- Fac.Num(iris)$factor
cluster.carac(iris.act,class,"co",2.0)  #  continuous variables

# frequency variables
data(DogBreeds)
attach(DogBreeds)
weig<-table(FUNC,WEIG)
weig<-data.frame(weig[,1],weig[,2],weig[,3])
cluster.carac(weig,  row.names(weig), "fr", 2) # frequency variables
detach(DogBreeds)
```

ColorAdjective *Associations between colors and adjectives.*

Description

A group of students from Nanterre University (Paris X) were presented with a list of eleve colours: blue, yellow, red, white, pink, brown, purple, grey, black, green and orange. Each person in the group was asked to describe each color with one ore more adjectives. A final list of 89 adjectives were associates with eleven colors.

Usage

```
data(ColorAdjective)
```

Format

Object of class data.frame with 89 rows and 11 columns.

Source

Jambu, M. and Lebeaux M.O. Cluster Analysis and Data Analysis. North-Holland. Amsterdam 1983.

References

Fine, J. (1996), *Iniciacion a los analisis de datos multidimensionales a partir de ejemplos*, Notas de curso, Montevideo

DogBreeds *Dog Breeds*

Description

Table that describes 27 dog breeds considering their size, weight, speed, intelligence, affectivity, aggressiveness and function.

Usage

```
data(DogBreeds)
```

Format

Object of class data.frame with 27 rows and 7 columns with the following description:

	VARIABLE	CATEGORIES		
[,1]	Size(SIZE)	Small(sma)	Mediun(med)	Large(lar)
[,2]	Weight(WEIG)	lightweight(lig)	Mediun(med)	Heavy(hea)
[,3]	Speed(SPEE)	Low(low)	Mediun(med)	High(hig)
[,4]	Intelligence(INTE)	Low(low)	Mediun(med)	High(hig)
[,5]	Affectivity(AFFE)	Low(low)	High(hig)	
[,6]	aggressiveness(AGGR)	Low(low)	High(hig)	
[,7]	function(FUNC)	Company(com)	Hunt(hun)	Utility(uti)

Source

Fine, J. (1996), 'Iniciacion a los analisis de datos multidimensionales a partir de ejemplos', Notas de clase, Montevideo.

References

Brefort, A.(1982), 'Letude des races canines a partir de leurs caracteristiques qualitatives', HEC - Jouy en Josas

dudi.tex *LaTeX Tables of Coordinates and Aids to Interpretation of Principal Axis Methods*

Description

Coordinates and aids of interpretation are wrote in tabular environment of LaTeX inside a Table

Usage

```
dudi.tex(dudi,job="",aidsC=TRUE,aidsR=TRUE,append=TRUE)
latex(obj,job="latex",tit="",lab="",append=TRUE,dec=1)
```

Arguments

dudi	an object of class dudi
job	a name to identify files and outputs
aidsC	if it is TRUE the coordinates and aids of interpretation of the columns are printed
aidsR	if it is TRUE the coordinates and aids of interpretation of the rows are printed
append	if it is TRUE LaTeX outputs are appended on the file
obj	object to export to LaTeX
tit	title of the table
lab	label for crossed references of LaTeX table
dec	number of decimal digits

Details

latex function is used to builp up a table. The aids of interpretation are obtained with inertia.dudi of ade4. A file is wrote in the work directory (job.txt) with the following tables:

tvalp eigenvalues

c1 eigenvectors

co column coordinates

col.abs column contributions in percentage

col.rel quality of the representation of columns in percentage

col.cum accumulated quality of the representation of columns in percentage/100

li row coordinates

row.abs row contributions in percent

row.rel quality of the representation of rows in percentage

row.cum accumulated quality of the representation of rows in percentage/100

Author(s)

Campo Elias PARDO <cepardot@unal.edu.co>

Examples

```
data(Bogota)
coa1 <- dudi.coa(Bogota[,2:7], scannf = FALSE)
# In order to create a file: Bogota.tex in LaTeX
# dudi.tex(coa1,job="Bogota")
```

Fac.Num *Division of qualitative and quantitative variables*

Description

An object of class data.frame is divided into a list with two tables, one with quantitative variables and the other with qualitative variables.

Usage

```
Fac.Num(tabla)
```

Arguments

tabla	object of class 'data.frame'

Value

It returns one list with one or two objects of class data.frame with the following characteristics:

factor	table with the qualitative variables
numeric	table with the quantitative variables

Author(s)

Pedro Cesar Del Campo <pcdelcampon@unal.edu.co>

Examples

```
data(DogBreeds)
Fac.Num(DogBreeds)

data(iris)
Fac.Num(iris)
```

FactoClass *Combination of Factorial Methods and Cluster Analysis*

Description

Performs the factorial analysis of the data and a cluster analysis using the nfcl first factorial coordinates

Usage

```
FactoClass( dfact, metodo, dfilu = NULL , nf = 2, nfcl = 10, k.clust = 3,
            scanFC = TRUE , n.max = 5000 , n.clus = 1000 ,sign = 2.0,
            conso=TRUE , n.indi = 25,row.w = rep(1, nrow(dfact)) )
## S3 method for class 'FactoClass'
print(x, ...)
analisis.clus(X,W)
```

Arguments

dfact	object of class data.frame, with the data of active variables.
metodo	function of ade4 for ade4 factorial analysis, dudi.pca,Principal Component Analysis; dudi.coa, Correspondence Analysis; witwit.coa, Internal Correspondence Analysis; dudi.acm, Multiple Correspondence Analysis ...
dfilu	ilustrative variables (default NULL)
nf	number of axes to use into the factorial analysis (default 2)
nfcl	number of axes to use in the classification (default 10)
k.clust	number of classes to work (default 3)
scanFC	if is TRUE, it asks in the console the values nf, nfcl y k.clust
n.max	when rowname(dfact)>=n.max, k-means is performed previous to hierarchical clustering (default 5000)
n.clus	when rowname(fact)>=n.max, the previous k-means is performed with *n.clus* groups (default 1000)
sign	threshold test value to show the characteristic variables and modalities
conso	when conso is TRUE, the process of consolidating the classification is performed (default TRUE)
n.indi	number of indices to draw in the histogram (default 25)
row.w	vector containing the row weights if metodo<>dudi.coa
x	object of class FactoClass
...	further arguments passed to or from other methods
X	coordinates of the elements of a class
W	weights of the elements of a class

Details

Lebart et al. (1995) present a strategy to analyze a data table using multivariate methods, consisting of an intial factorial analysis according to the nature of the compiled data, followed by the performance of mixed clustering. The mixed clustering combines hierarchic clustering using the Ward's method with K-means clustering. Finally a partition of the data set and the characterization of each one of the classes is obtained, according to the active and illustrative variables, being quantitative, qualitative or frequency.

FactoClass is a function that connects procedures of the package ade4 to perform the analysis factorial of the data and from stats for the cluster analysis.

The function analisis.clus calculates the geometric characteristics of each class: size, inertia, weight and square distance to the origin.

For impression in LaTeX format see FactoClass.tex

To draw factorial planes with cluster see plotFactoClass

Value

object of class FactoClass with the following:

dudi	object of class dudi from ade4 with the specifications of the factorial analysis
nfcl	number of axes selected for the classification
k	number of classes
indices	table of indices obtained through WARD method
cor.clus	coordinates of the clusters
clus.summ	summary of the clusters
cluster	vector indicating the cluster of each element
carac.cate	cluster characterization by qualitative variables
carac.cont	cluster characterization by quantitative variables
carac.frec	cluster characterization by frequency active variables

Author(s)

Pedro Cesar del Campo <pcdelcampon@unal.edu.co>, Campo Elias Pardo <cepardot@unal.edu.co>, Ivan Diaz <ildiazm@unal.edu.co>, Mauricio Sadinle <msadinleg@unal.edu.co>

References

Lebart, L. and Morineau, A. and Piron, M. (1995) Statisitique exploratoire multidimensionnelle, Paris.

Examples

```
# Cluster analysis with Correspondence Analysis
data(ColorAdjective)
FC.col <-FactoClass(ColorAdjective, dudi.coa)
6
10
5

FC.col

FC.col$dudi


# Cluster analysis with Multiple Correspondence Analysis
data(DogBreeds)
```

```
DB.act <- DogBreeds[-7]  # active variables
DB.ilu <- DogBreeds[7]   # ilustrative variables

FC.db <-FactoClass( DB.act, dudi.acm, k.clust = 4,
                      scanFC = FALSE, dfilu = DB.ilu, nfcl = 10)

FC.db

FC.db$clus.summ
FC.db$indices
```

FactoClass.tex *Table of Coordinates, Aids of Interpretation of the Principal Axes and Cluster Analysis in LaTeX format.*

Description

The coordinates, aids of interpretation and results of cluster analysis of an object of class FactoClass are written in tables for edition in LaTeX format and written in a file.

Usage

```
FactoClass.tex(FC,job="",append=TRUE, dir = getwd(), p.clust = FALSE )

## S3 method for class 'FactoClass.tex'
print(x, ...)

latexDF(obj, job="latex" ,tit="" ,lab="" ,append=TRUE ,dec=1,
                      dir = getwd() , to.print = TRUE )
roundDF(tabla,dec=1)
```

Arguments

FC	object of class FactoClass.
job	A name to identify the exit.
append	if is 'TRUE' the exit in LaTeX format is added to the file.
dir	name of the directory in which the file is kept.
p.clust	the value of this parameter is 'TRUE' or 'FALSE' to print or not the cluster of each element.
tabla	object of class 'data frame'.
dec	number of decimal.
x	object of class FactoClass.tex

...	further arguments passed to or from other methods
obj	object of class data.frame.
tit	title of the table in LaTeX format.
lab	label of the table in LaTeX format.
to.print	if it is 'TRUE' the table is also printed in the console.

Details

This function helps with the construction of tables in *LaTeX* format. Besides, it allows a easy reading of the generated results by FactoClass. The function latexDF is an entrance to xtable and turns an object of class data.frame a table in LaTeX format.

Value

object of class FactoClass.tex with the following characteristics:

tvalp	eigenvalues * 1000.
c1	eigenvectors.
co	coordinates of the columns.
col.abs	contribution of each column to the inertia of the axis (percentage).
col.rel	quality of representation of each column (percentage).
col.cum	quality of representation of each column accumulated in the subspace (percentage).
li	coordinates of the rows.
row.abs	contribution of each rows to the inertia of the axis (percentage).
row.rel	quality of representation of each rows (percentage).
row.cum	quality of representation of each rows accumulated in the subspace (percentage).
indices	table of indices of level generated by the Ward cluster analysis.
cor.clus	coordinates of the center of gravity of each cluster.
clus.summ	summary of the cluster.
carac.cate	cluster characterization by qualitative variables.
carac.cont	cluster characterization by quantitative variables.
cluster	vector indicating the cluster of each element.

Author(s)

Pedro Cesar del Campo <pcdelcampon@unal.edu.co>, Campo Elias Pardo <cepardot@unal.edu.co>

Examples

```
data(DogBreeds)
DB.act <- DogBreeds[-7]  # active variables
DB.ilu <- DogBreeds[7]   # illustrative variables
# MCA
FaCl <- FactoClass( DB.act, dudi.acm,
                    scanFC = FALSE, dfilu = DB.ilu, nfcl = 10, k.clust = 4 )
# In order to create a file in LaTeX format
# FactoClass.tex(FaCl,job="DogBreeds1", append=TRUE)
# FactoClass.tex(FaCl,job="DogBreeds", append=TRUE , p.clust = TRUE)
```

---

icfes08 *Departmenst by Levels of Schools in Colombia*

Description

Contingency Table that classificaes the schools of Colombia by departments and level of the schools agree with the performance of its students.

Usage

```
data(icfes08)
```

Format

Object whit class `data.frame` of 29 rows and 12 columns.

Source

ICFES Colombia

References

C.E. Pardo, M. B\'ecue and J.E. Ortiz (2013). Correspondence Analysis of Contingency Tables with Subpartitions on Rowsand Columns. Revista Colombiana de Estad\'istica, 36(1):115-144.

---

kmeansW *K-means with Weights of the Elements*

Description

It is a modification of kmeans Hartigan-Wong algorithm to consider the weight of the elements to classify.

Usage

```
kmeansW(x, centers, weight = rep(1,nrow(x)),
        iter.max = 10, nstart = 1)
```

Arguments

x	A numeric vector, matrix or data frame.
centers	Either the number of clusters or a set of initial (distinct) cluster centres. If a number, a random set of (distinct) rows in x is chosen as the initial centres.
weight	weight of the elements of x. by default the same.
iter.max	The maximum number of iterations allowed.
nstart	If centers is a number, how many random sets should be chosen?

Details

With the 'Hartigan-Wong' algorithm, this function performs the *K-means* clustering diminishing inertia intra classes. In this version the Fortran code kmnsW.f was changed by C++ code kmeanw.cc programed by Camilo Jose Torres, modifing C code programed by Burkardt.

Value

object of class `FactoClass.tex` with the following characteristics:

cluster	vector indicating the cluster of each element.
...	

Author(s)

Camilo Jose Torres <cjtorresj@unal.edu.co>, Campo Elias Pardo <cepardot@unal.edu.co>

References

Hartigan, J. A. and Wong, M. A. (1979). A K-means clustering algorithm. *Applied Statistics* **28**, 100–108.

Burkardt, J. (2008). ASA136 The K-Means Algorithm. https://people.sc.fsu.edu/~jburkardt/cpp_src/asa136/asa136.html

Examples

```
data(Bogota)
ac.bog <- Bogota[-1]
il.bog <- Bogota[ 1]

acs <- dudi.coa( ac.bog, nf=6 , scannf = FALSE )

kmeansW( acs$li, 7, acs$lw )
```

`list.to.data` *list to data.frame*

Description

Modification of an object of class `list` into an object of class `data.frame`.

Usage

```
list.to.data(lista,nvar="clasif")
```

Arguments

`lista`	`list` that contains several `data.frame` of the same structure.
`nvar`	(Optional) Name of the new variable that considers the partition given by the elements of the list.

Details

This function turns an object of class `list` into an object of class `data.frame`, this function is used internally to create objects of class data.frame to make tables in *LaTeX* format.

Value

Object of class data.frame.

Author(s)

Pedro Cesar Del Campo <pcdelcampon@unal.edu.co>

Examples

```
A <- data.frame(r1=rnorm(5),r2=rnorm(5))
B <- data.frame(r1=rnorm(15),r2=rnorm(15))

LL <- list(A=A,B=B)
LL
list.to.data(LL)
```

plot.dudi *Factorial Planes from Objects of Class dudi*

Description

It plots factorial planes from objects of class dudi

Usage

```
## S3 method for class 'dudi'
plot(x,ex=1,ey=2,xlim=NULL,ylim=NULL,main=NULL,rotx=FALSE,
      roty=FALSE,roweti=row.names(dudi$li),
      coleti=row.names(dudi$co),axislabel=TRUE,font.col="plain",
      font.row="plain",col.row="black",col.col="blue",
      alpha.col=1,alpha.row=1,cex=0.8,cex.row=0.8,cex.col=0.8,
      all.point=TRUE,Trow=TRUE,Tcol=TRUE,cframe=1.2,ucal=0,
      cex.global=1,infaxes="out",gg=FALSE,...)
sutil.grid(cgrid,scale=TRUE)
```

Arguments

x	object of type dudi
ex	number indentifying the factor to be used as horizontal axis. Default 1
ey	number indentifying the factor to be used as vertical axis. Default 2
xlim	the x limits (x1, x2) of the plot
ylim	the y limits of the plot
main	graphic title
rotx	TRUE if you want change the sign of the horizontal coordinates. Default FALSE
roty	TRUE if you want change the sign of the vertical coordinates. Default FALSE
roweti	selected row points for the graphic. Default all points
coleti	selected column points for the graphic. Default all points
font.row	type of font for row labels. Default "plain"
font.col	type of font for column labels. Default "plain"
axislabel	if it is TRUE the axis information is written
col.row	color for row points and row labels. Default "black"
col.col	color for column points and column labels. Default "blue"
alpha.row	transparency for row points and row labels. Default cex.ilu=1
alpha.col	transparency for column points and column labels. Default cex.ilu=1
cex	global scale for the labels. Default cex=0.8
cex.row	scale for row points and row labels. Default cex.row=0.8
cex.col	scale for column points and column labels. Default cex.col=0.8

all.point	If if is TRUE, all points are outlined. Default all.point=TRUE
Trow	if it is TRUE the row points are outlined. Default TRUE
Tcol	if it is TRUE the column points are outlined. Default TRUE
cframe	scale for graphic limits
ucal	quality representation threshold (percentage) in the plane . Default ucal=0
cex.global	scale for the label sizes
infaxes	place to put the axes information: "out","in","no". Default infaxes="out". If infaxes="out" the graphic is similar to FactoMineR graphics, otherwise the style is similar to the one in ade4, without axes information when infaxes="no"
gg	If TRUE the version ggplot ggrepel is perfomance. Default FALSE
...	further arguments passed to or from other methods
cgrid	internal parameter
scale	internal

Details

Plot the selected factorial plane. sutil.grid is used by plot.dudi

Value

It graphs the factorial plane x,y using $co, $li of a "dudi" object. If ucal > 0, the function inertia.dudi is used to calculate the quality of representation on the plane

Author(s)

Campo Elias Pardo <cepardot@unal.edu.co> and Jhonathan Medina <jmedinau@unal.edu.co>

Examples

```
data(Bogota)
ca <- dudi.coa(Bogota[,2:7],scannf=FALSE,nf=4)
# with ggplot2 and ggrepel
plot(ca,gg=TRUE)
dev.new()
# ade4 style
plot.dudi(ca,ex=3,ey=4,ucal=0.2,all.point=FALSE,infaxes="in")
```

plotcc *Correlation circle from coordinates*

Description

It plots Correlation circle from a coordinate table

Usage

```
plotcc(x,ex=1,ey=2,cex.label=4.5,col.label="black",font.label="bold",col.arrow="black",
fullcircle=TRUE,y=NULL)
```

Arguments

x	matrix or data.frame with coordinates
ex	the component like horizontal axis
ey	the component like vertical axis
cex.label	size of the variable labels. Default 4.5
col.label	color of the variable labels. Default black
font.label	font of the variable labels from fontface of ggplot2. Defult bold
col.arrow	color of the arrows. Default black
fullcircle	if it is TRUE (default), the circle is complete
y	internal

Details

Plot the selected factorial plane as a correlation circle for the variables from a normed PCA.

Value

It graphs the factorial plane ex,ey using a data.frame or matrix x with axis coordinates.

Author(s)

Jhonathan Medina <jmedinau@unal.edu.co> and Campo Elias Pardo <cepardot@unal.edu.co>

Examples

```
data(admi)
pca <- dudi.pca(admi[,2:6],scannf=FALSE,nf=2)
# fullcircle
plotcc(pca$co)
# no fullcircle
plotcc(pca$co,fullcircle=FALSE)
```

plotct *Row and Column Profiles of a Contingency Table*

Description

It plots barplot profiles of rows or columns from a contingency table including marginal profiles

Usage

```
plotct(x,profiles="both",legend.text=TRUE,tables=FALSE,nd=1,... )
```

Arguments

x	contingency table
profiles	select profiles: "both" file and column profiles in two graph devices, "row" only row profiles, "col" only column profiles
legend.text	if it is TRUE a box with legends is included at the right
tables	logical, if TRUE tables with marginals are returned
nd	number of decimals to profiles as percentages
...	further arguments passed to or from other methods

Details

Plot row profiles in horizontal form and columns profiles in vertical form

Value

if tables=TRUE, object of class `list` with the following:

ct	contingengy table with row and column marginals
perR	row profile with marginal, in percent
perC	column profile with marginal, in percent

Author(s)

Camilo Jose Torres <cjtorresj@unal.edu.co> , Campo Elias Pardo <cepardot@unal.edu.co>

Examples

```
mycolors<-colors()[c(1,26,32,37,52,57,68,73,74,81,82,84,88,100)]
data(Bogota)
plotct(Bogota[,2:7],col=mycolors)
# return tables with marginals
tabs <- plotct(Bogota[,2:7],col=mycolors,tables=TRUE,nd=0)
```

plotFactoClass *Factorial Planes Showing the Classes*

Description

For objects of class FactoClass it graphs a factorial plane showing the center of gravity of the cluster, and identifying with colors the cluster to which each element belongs.

Usage

```
plotFactoClass(FC,x=1,y=2,xlim=NULL,ylim=NULL,rotx=FALSE,roty=FALSE,
                roweti=row.names(dudi$li),coleti=row.names(dudi$co),
                titre=NULL,axislabel=TRUE,col.row=1:FC$k,
                col.col="blue",cex=0.8,cex.row=0.8,cex.col=0.8,
                all.point=TRUE,Trow=TRUE,Tcol=TRUE,cframe=1.2,ucal=0,
                cex.global=1,infaxes="out",
                nclus=paste("cl", 1:FC$k, sep=""),
                cex.clu=cex.row,cstar=1,gg=FALSE)
```

Arguments

FC	object of class FactoClass.
x	number indentifying the factor to be used as horizontal axis. Default x=1
y	number indentifying the factor to be used as vertical axis. Default y=2
xlim	the x limits (x1, x2) of the plot
ylim	the y limits of the plot
rotx	TRUE if you want change the sign of the horizontal coordinates (default FALSE).
roty	TRUE if you want change the sign of the vertical coordinates (default FALSE).
roweti	selected row points for the graphic. Default all points.
coleti	selected column points for the graphic. Default all points.
titre	graphics title.
axislabel	if it is TRUE the axis information is written.
col.row	color for row points and row labels. Default 1:FC$k.
col.col	color for column points and column labels. Default "grey55".
cex	global scale for the labels. Default cex=0.8.
cex.row	scale for row points and row labels. Default cex.row=0.8.
cex.col	scale for column points and column labels. Default cex.col=0.8.
cex.clu	scale for cluster points and cluster labels. (default cex.row).
all.point	if if is TRUE, all points are outlined. Default all.point=TRUE.
Trow	if it is TRUE the row points are outlined. Default TRUE.

Tcol	if it is TRUE the column points are outlined. Default TRUE.
nclus	labels for the clusters (default cl1, cl2, ...
cframe	scale for graphics limits
ucal	quality Representation Threshold in the plane. Default ucal=0
cex.global	scale for the label sizes
infaxes	place to put the axes information: "out","in","no". Default infaxes="out". If infaxes="out" the graphic is similar to FactoMineR graphics, otherwise the style is similar to the one in ade4, without axes information when infaxes="no"
cstar	length of the rays between the centroids of the classes and their points
gg	If TRUE the version ggplot ggrepel is perfomance. Default FALSE

Details

It draws the factorial plane with the clusters. Only for objects FactoClass see FactoClass. The factorial plane is drawn with planfac and the classes are projected with s.class of ade4

Value

It draws the factorial plane x, y using $co, $li of the object of class FactoClass. If ucal > 0, the function inertia.dudi is used to calculate the quality of representation in the plane.

Author(s)

Campo Elias Pardo <cepardot@unal.edu.co> Pedro Cesar del Campo <pcdelcampon@unal.edu.co>,

Examples

```
data(Bogota)
Bog.act <- Bogota[-1]
Bog.ilu <- Bogota[ 1]

FC.Bogota<-FactoClass(Bog.act, dudi.coa,Bog.ilu,nf=2,nfcl=5,k.clust=5,scanFC=FALSE)

plotFactoClass(FC.Bogota,titre="First Factorial Plane from the SCA of Bogota's Blocks",
     col.row=c("maroon2","orchid4","darkgoldenrod2","dark red","aquamarine4"))
```

---

plotfp *Factorial Planes from Coordinates*

Description

It plots factorial planes from a coordinate table

Usage

```
plotfp(co,x=1,y=2,eig=NULL,cal=NULL,ucal=0,xlim=NULL,ylim=NULL,main=NULL,rotx=FALSE,
     roty=FALSE,eti=row.names(co),axislabel=TRUE,col.row="black",cex=0.8,cex.row=0.8,
       all.point=TRUE,cframe=1.2,cex.global=1,infaxes="out",asp=1,gg=FALSE)
```

Arguments

co	matrix or data.frame with coordinates
x	the component like horizontal axis
y	the component like vertical axis
eig	numeric with the eigenvalues
cal	matrix or data.frame with the square cosinus
ucal	quality representation threshold (percentage) in the plane . Default ucal=0
xlim	the x limits (x1, x2) of the plot
ylim	the y limits of the plot
main	graphic title
rotx	TRUE if you want change the sign of the horizontal coordinates. Default FALSE
roty	TRUE if you want change the sign of the vertical coordinates. Default FALSE
eti	selected row points for the graphic. Default all points
axislabel	if it is TRUE the axis information is written
col.row	color for row points and row labels. Default "black"
cex	global scale for the labels. Default cex=0.8
cex.row	scale for row points and row labels. Default cex.row=0.8
all.point	If if is TRUE, all points are outlined. Default all.point=TRUE
cframe	scale for graphic limits
cex.global	scale for the label sizes
infaxes	place to put the axes information: "out","in","no". Default infaxes="out". If infaxes="out" the graphic is similar to FactoMineR graphics, otherwise the style is similar to the one in ade4, without axes information when infaxes="no"
asp	the y/x aspect ratio
gg	If TRUE the version ggplot ggrepel is perfomance. Default FALSE

Details

Plot the selected factorial plane.

Value

It graphs the factorial plane x,y using co and optional information of eigenvalues and representation quality of the points. If ucal > 0, only the points with the quality of representation on the plane bigger than ucal are pointed

Author(s)

Campo Elias Pardo <cepardot@unal.edu.co> and Jhonathan Medina <jmedinau@unal.edu.co>

Examples

```
data(Bogota)
ca <- dudi.coa(Bogota[,2:7],scannf=FALSE,nf=2)
# ade4 style
plotfp(ca$li,eig=ca$eig,main="First Factorial Plane",infaxes="in")
# with ggplot2 and ggrepel
plotfp(ca$li,eig=ca$eig,main="First Factorial Plane",gg=TRUE)
```

---

plotpairs *Modified pairs plot*

Description

Modified pairs plot: marginal kernel densities in diagonal, bivariated kernel densities in triangular superior; and scatter bivariate plots in triangular inferior

Usage

```
plotpairs(X,maxg=5,cex=1)
```

Arguments

X	matrix or data.frame of numeric colums
maxg	maximum number of variables to plot
cex	size of the points in dispersion diagrams

Details

Plot row profiles in horizontal form and columns profiles in vertical form

Value

The function does not return values

Author(s)

Campo Elias Pardo <cepardot@unal.edu.co>

Examples

```
data(iris)
plotpairs(iris[,-5])
```

stableclus *Stable clusters for cluster analysis*

Description

Performs Stable Cluster Algorithm for cluster analysis, using factorial coordinates from a dudi object

Usage

```
stableclus(dudi,part,k.clust,ff.clus=NULL,bplot=TRUE,kmns=FALSE)
```

Arguments

dudi	A dudi object, result of a previous factorial analysis using ade4 or FactoClass
part	Number of partitions
k.clust	Number of clusters in each partition
ff.clus	Number of clusters for the final output, if NULL it asks in the console (Default NULL)
bplot	if TRUE, prints frequencies barplot of each cluster in the product partition (Default TRUE)
kmns	if TRUE, the process of consolidating the classification is performed (Default FALSE)

Details

Diday (1972) (cited by Lebart et al. (2006)) presented a method for cluster analysis in an attempt to solve one of the inconvinients with the *kmeans* algorithm, which is convergence to local optims. Stable clusters are built by performing different partitions (using kmeansW algorithmn), each one with different initial points. The groups are then formed by selecting the individuals belonging to the same cluster in every partion.

Value

object of class stableclus with the following characteristics:

cluster	vector indicating the cluster of each element.

...

Author(s)

Carlos Andres Arias <caariasr@unal.edu.co>, Campo Elias Pardo <cepardot@unal.edu.co>

References

Arias, C. A.; Zarate, D.C. and Pardo C.E. (2009), 'Implementacion del metodo de grupos estables en el paquete FactoClass de R', in: XIX Simposio Colombiano de Estadistica. Estadisticas Oficiales Medellin Colombia, Julio 16 al 20 de 2009 Universidad Nacional de Colombia. Bogota.

Lebart, L. (2015), 'DtmVic: Data and Text Mining - Visualization, Inference, Classification. Exploratory statistical processing of complex data sets comprising both numerical and textual data.', Web. http://www.dtmvic.com/

Lebart, L., Morineau, A., Lambert, T. and Pleuvret, P. (1999), *SPAD*. Syst?me Pour L'Analyse des Don?es, Paris.

Lebart, L., Piron, M. and Morineau, A. (2006), Statisitique exploratoire multidimensionnelle. Visualisation et inference en fouilles de donnees, 4 edn, Dunod, Paris.

Examples

```
 data(ColorAdjective)
 FCcol <-FactoClass(ColorAdjective, dudi.coa,nf=6,nfcl=10,k.clust=7,scanFC = FALSE)
 acs <- FCcol$dudi
# stableclus(acs,3,3,4,TRUE,TRUE)
```

`supqual` *Projection of Qualitative Variables in PCA and MCA*

Description

It returns the coordinates and aids to the interpretation when one or more qualitative variables are projected as ilustrative in PCA or MCA

Usage

```
supqual(du,qual)
```

Arguments

`du`	a object of class "pca" or "acm" ("dudi") obtained with `dudi.pca` or `dudi.acm` of package `ade4`
`qual`	a data.frame of qualitative variables as factors

Value

object of class `list` with the following:

`wcat`	weight of the categories in PCA case
`ncat`	frequency of the categories in MCA case
`dis2`	square distance to the origin from the complete space

coor	factorial coordinates
tv	test values
cos2	square cosinus
scr	relation of correaltion

Author(s)

Campo Elias Pardo <cepardot@unal.edu.co>

Examples

```
# in PCA
data(admi)
Y<-admi[,2:6]
pcaY<-dudi.pca(Y,scannf=FALSE)
Yqual<-admi[,c(1,8)]
supqual(pcaY,Yqual)
# in MCA
Y<-admi[,c(8,11,9,10)]
mcaY<-dudi.acm(Y,scannf=FALSE)
supqual(mcaY,admi[,c(1,13)])
```

Vietnam *Student opinions about the Vietnam War*

Description

The newspaper of the students of the University of Chapel Hill (North Carolina) conducted a survey of student opinions about the Vietnam War in May 1967. Responses were classified by sex, year in the program and one of four opinions:

A defeat power of North Vietnam by widespread bombing and land invasion

B follow the present policy

C withdraw troops to strong points and open negotiations on elections involving the Viet Cong

D immediate withdrawal of all U.S. troops

Usage

```
data(Vietnam)
```

Format

The 3147 consulted students were classified considering the sex, year of study and chosen strategy, originating a contingency table of 10 rows: M1 to M5 and F1 to F5 (the years of education are from 1 to 5 and sexes are male (M) and female (F)) and 4 columns A, B, C and D.

Source

Fine, J. (1996), 'Iniciacion a los analisis de datos multidimensionales a partir de ejemplos', Notes of course, Montevideo

---

`ward.cluster` *Hierarchic Classification by Ward's Method*

Description

Performs the classification by Ward's method from the matrix of Euclidean distances.

Usage

```
ward.cluster(dista, peso = NULL , plots = TRUE, h.clust = 2, n.indi = 25 )
```

Arguments

`dista`	matrix of Euclidean distances (class(dista)=="dist").
`peso`	(Optional) weight of the individuals, by default equal weights
`plots`	it makes dendrogram and histogram of the Ward's method
`h.clust`	if it is '0' returns a object of class `hclust` and a table of level indices, if it is '1' returns a object of class `hclust`, if it is '2' returns a table of level indices.
`n.indi`	number of indices to draw in the histogram (default 25).

Details

It is an entrance to the function `h.clus` to obtain the results of the procedure presented in Lebart et al. (1995). Initially the matrix of distances of Ward of the elements to classify is calculated:

The Ward's distance between two elements to classify i and l is given by:

$$W(i,l) = (m_i * m_l)/(m_i + m_i) * dist(i,l)^2$$

where m_i y m_l are the weights and $dist(i,l)$ is the Euclidean distance between them.

Value

It returns an object of class hclust and a table of level indices (depending of h.clust). If plots = TRUE it draws the indices of level and the dendrogram.

Author(s)

Pedro Cesar del Campo <pcdelcampon@unal.edu.co>, Campo Elias Pardo <cepardot@unal.edu.co>

References

Lebart, L. and Morineau, A. and Piron, M. (1995) *Statisitique exploratoire multidimensionnelle*, Paris.

Examples

```
data(ardeche)
ca <- dudi.coa(ardeche$tab,scannf=FALSE,nf=4)

 ward.cluster( dista= dist(ca$li), peso=ca$lw )

 dev.new()
 HW <- ward.cluster( dista= dist(ca$li), peso=ca$lw ,h.clust = 1)
 plot(HW)
 rect.hclust(HW, k=4, border="red")
```

`Whisky` *Whisky example*

Description

Data frame with five features of 35 whisky brands:

price in Frace Francs

malt proportion in percentage

type by malt proportion: low, medium, pure

aging in years

taste mean score of a taste panel

Usage

```
data(Whisky)
```

Source

Fine, J. (1996), 'Iniciacion a los analisis de datos multidimensionales a partir de ejemplos', Notes of course, Montevideo

Index