Package 'GDAtools'

October 1, 2026

Type Package

Title Geometric Data Analysis

Version 2.3.1

Imports descriptio (>= 1.2), FactoMineR, ggplot2, ggrepel, rlang

Suggests TraMineR, sf, shiny, miniUI, esquisse, rclipboard, factoextra, ade4

Description Many tools for Geometric Data Analysis (Le Roux & Rouanet (2005) <doi:10.1007/1-4020-2236-0>), such as MCA variants (Specific Multiple Correspondence Analysis, Class Specific Analysis), many graphical and statistical aids to interpretation (structuring factors, concentration ellipses, inductive tests, bootstrap validation, etc.) and multiple-table analysis (Multiple Factor Analysis, between- and inter-class analysis, Principal Component Analysis and Correspondence Analysis with Instrumental Variables, etc.).

License GPL (>= 2)

Encoding UTF-8

URL https://framagit.org/nicolas-robette/GDAtools, https://nicolas-robette.frama.io/GDAtools/

BugReports https://framagit.org/nicolas-robette/GDAtools/-/work_items

LazyData true

NeedsCompilation no

Author Nicolas Robette [aut, cre]

Maintainer Nicolas Robette <nicolas.robette@uvsq.fr>

Repository CRAN

Date/Publication 2026-10-01 10:30:16 UTC

Contents

ahc.plots 3
angles.csa 4
barplot_contrib 5
bcMCA 7

bcPCA 8
bootvalid_supvars 10
bootvalid_variables 11
break_interaction 12
burt 14
coiMCA 15
coiPCA 16
conc.ellipse 17
contrib 19
csMCA 20
DA 22
DAQ 23
dichotom 25
dichotomixed 26
dimcontrib 27
dimdescr 28
dimeta2 30
dimtypicality 31
dist.chi2 32
flip.mca 32
getindexcat 33
ggadd_attractions 34
ggadd_chulls 35
ggadd_corr 37
ggadd_density 39
ggadd_ellipses 40
ggadd_interaction 42
ggadd_kellipses 43
ggadd_partial 45
ggadd_supind 47
ggadd_supvar 48
ggadd_supvars 49
ggaxis_variables 51
ggbootvalid_supvars 53
ggbootvalid_variables 55
ggcloud_indiv 57
ggcloud_variables 59
ggeta2_variables 62
ggsmoothed_supvar 63
gPCA 64
homog.test 66
ijunk 67
MCAiv 68
MCAoiv 69
medoids 71
modif.rate 72
multiMCA 73
Music 75

nsCA . . . 76
nsca.biplot . . . 77
PCAiv . . . 78
PCAoiv . . . 80
planecontrib . . . 81
plot.csMCA . . . 82
plot.multiMCA . . . 84
plot.speMCA . . . 85
plot.stMCA . . . 87
quadrant . . . 88
quasindep . . . 89
reshape_between . . . 90
rvcoef . . . 91
scaled.dev . . . 92
speMCA . . . 93
stMCA . . . 95
supind . . . 96
supvar . . . 97
supvars . . . 99
tabcontrib . . . 100
Taste . . . 102
textindsup . . . 103
textvarsup . . . 104
translate.logit . . . 105
wcMCA . . . 105
wcPCA . . . 107
wtable . . . 108

Index **110**

`ahc.plots` *Plots for Ascending Hierarchical Clustering*

Description

Draws various plots for Ascending Hierarchical Clustering results.

Usage

```
ahc.plots(ahc, distance = NULL, max.cl = 20, type = "dist")
```

Arguments

`ahc`	object of class `hclust` or `agnes`
`distance`	A dissimilarity matrix or a `dist` object. Only used if `type` is "inert" or "loss". Default is NULL.
`max.cl`	Integer. Maximum number of clusters taken into account in the plots.

type	Character string. If "dist" (default), the distance between agregated clusters is plotted. If "inert", it is the percentage of explained inertia (pseudo-R2). If "loss", it is the relative loss of explained inertia (pseudo-R2).

Details

The three kinds of plots proposed with this function are aimed at guiding in the choice of the number of clusters.

Author(s)

Nicolas Robette

See Also

dist.chi2

Examples

```
data(Taste)
# clustering of a subsample of the data
disjonctif <- dichotom(Taste[1:200, 1:11])
distance <- dist(disjonctif)
cah <- stats::hclust(distance, method = "ward.D2")
# distance between aggregated clusters
ahc.plots(cah, max.cl = 15, type = "dist")
# percentage of explained inertia
ahc.plots(cah, distance = distance, max.cl = 15, type = "inert")
# relative loss of explained inertia
ahc.plots(cah, distance = distance, max.cl = 15, type = "loss")
```

---

angles.csa *Cosine similarities and angles between CSA and MCA*

Description

Computes the cosines similarities and angles between the components of a CSA and those of a MCA.

Usage

```
angles.csa(rescsa, resmca, naxes = 3)
```

Arguments

rescsa	object of class csMCA
resmca	object of class MCA or speMCA
naxes	integer. Number of axes to retain. Default is 3.

Value

A list of matrices:

`cosines`	Cosine similarities between CSA and MCA dimensions
`angles.with.dims`	Angles between CSA and MCA dimensions
`angles.with.planes`	Angles between CSA dimensions and MCA planes
`best.planes`	Angles between each CSA dimension and its closest MCA plane

Author(s)

Lucas Page Pereira, Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

MCA, speMCA, csMCA

Examples

```
## Performs a specific MCA and a CSA on the Music example data set
## and computes cosine similarities and angles
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
resmca <- speMCA(Music[,1:5], excl = junk)
female <- Music$Gender=="Women"
rescsa <- csMCA(Music[,1:5], subcloud = female, excl = junk)
angles.csa(rescsa, resmca)
```

---

barplot_contrib *Bar plot of contributions*

Description

From MCA results, plots contributions to the axes.

Usage

```
barplot_contrib(resmca, dim = 1, which = "var",
  sort = FALSE, col = "tomato4", repel = FALSE)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
dim	the dimension to use. Default is 1.
which	If resmca is of class MCA, speMCA, csMCA or PCA, should be "var" to plot contributions of variables or "ind" to plot contributions of individuals. If resmca is of class CA, should be "row" to plot contributions of rows or "col" to plot contributions of columns. Default is "var".
sort	logical. If TRUE, bars are sorted by decreasing VIPs. Default is FALSE.
col	color of the bars
repel	logical. If TRUE, the names of the variables are repelled with geom_text_repel. Default is FALSE

Details

The contributions are multiplied by the sign of the coordinates, so that the plot shows on which side of the axis they contribute, which makes the interpretation easier.

Value

a ggplot2 object

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

speMCA, tabcontrib

Examples

```
# specific MCA on the Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# contributions of categories
barplot_contrib(mca)
```

bcMCA *Between-class MCA*

Description

Between-class MCA, also called Barycentric Discriminant Analysis

Usage

```
bcMCA(data, class, excl = NULL, row.w = NULL)
```

Arguments

data	data frame with only categorical variables, i.e. factors
class	factor specifying the class
excl	numeric vector indicating the indexes of the "junk" categories (default is NULL). See getindexcat or use ijunk interactive function to identify these indexes. It may also be a character vector of junk categories, specified in the form "namevariable.namecategory" (for instance "gender.male").
row.w	numeric vector of row weights. If NULL (default), a vector of 1 for uniform row weights is used.

Details

Between-class MCA is sometimes also called Barycentric Discriminant Analysis or Discriminant Correspondence Analysis. It consists in three steps : 1. Transformation of data into an indicator matrix (i.e. disjunctive table) 2. Computation of the barycenter of the transformed data for each category of class 3. Correspondence Analysis of the set of barycenters Between-class MCA can also be viewed as a special case of MCA with instrumental variables, with only one categorical instrumental variable.

Value

An object of class CA from FactoMineR package, with the indicator matrix of data as supplementary rows, and an additional item :

ratio	the between-class inertia percentage

Author(s)

Nicolas Robette

References

Abdi H., 2007, "Discriminant Correspondence Analysis", In: Neil Salkind (Ed.), *Encyclopedia of Measurement and Statistics*, Thousand Oaks (CA): Sage.

Bry X., 1996, *Analyses factorielles multiples*, Economica.

Lebart L., Morineau A. et Warwick K., 1984, *Multivariate Descriptive Statistical Analysis*, John Wiley and sons, New-York.)

See Also

DAQ, MCAiv

Examples

```
library(FactoMineR)
data(tea)
res <- bcMCA(tea[,1:18], tea$SPC)
# categories of class
plot(res, invisible = c("col", "row.sup"))
# Variables in tea data
plot(res, invisible = c("row", "row.sup"))
# between-class inertia percentage
res$ratio
```

---

bcPCA *Between-class Principal Component Analysis*

Description

Between-class Principal Component Analysis

Usage

```
bcPCA(data, class, row.w = NULL, scale.unit = TRUE, ncp = 5)
```

Arguments

data	data frame with only numeric variables
class	factor specifying the class
row.w	numeric vector of row weights. If NULL (default), a vector of 1 for uniform row weights is used.
scale.unit	logical. If TRUE (default) then data are scaled to unit variance.
ncp	number of dimensions kept in the results (by default 5)

Details

Between-class Principal Component Analysis consists in two steps : 1. Computation of the barycenter of `data` rows for each category of `class` 2. Principal Component Analysis of the set of barycenters

It is a quite similar to Linear Discriminant Analysis, but the metric is different.

It can be seen as a special case of PCA with instrumental variables, with only one categorical instrumental variable.

Value

An object of class `PCA` from `FactoMineR` package, with the original data as supplementary individuals, and an additional item :

`ratio`	the between-class inertia percentage

Author(s)

Nicolas Robette

References

Bry X., 1996, *Analyses factorielles multiples*, Economica.

Lebart L., Morineau A. et Warwick K., 1984, *Multivariate Descriptive Statistical Analysis*, John Wiley and sons, New-York.)

See Also

`PCAiv`, `DA`

Examples

```
library(FactoMineR)
data(decathlon)
points <- cut(decathlon$Points, c(7300, 7800, 8000, 8120, 8900), c("Q1","Q2","Q3","Q4"))
res <- bcPCA(decathlon[,1:10], points)
# categories of class
plot(res, choix = "ind", invisible = "ind.sup")
# variables in decathlon data
plot(res, choix = "var")
# between-class inertia percentage
res$ratio
```

`bootvalid_supvars` *Bootstrap validation (supplementary variables)*

Description

Bootstrap validation of MCA, through the computation of the coordinates of supplementary variables for bootstrap replications of the data.

Usage

```
bootvalid_supvars(resmca, vars = NULL, axes = c(1,2), K = 30)
```

Arguments

`resmca`	object created with `MCA`, `speMCA`, `csMCA`, `wcMCA` or `bcMCA` function
`vars`	a data frame of categorical supplementary variables. All these variables should be factors.
`axes`	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
`K`	integer. Number of bootstrap replications (default is 30).

Details

The bootstrap technique is used here as an internal and non-parametric validation procedure of the results of a multiple correspondence analysis. For supplementary variables, only "partial bootstrap" is possible. The partial bootstrap does not compute new MCAs: it projects bootstrap replications of the initial data as supplementary elements of the MCA (see references for more details).

Value

A data frame with the following elements :

`varcat`	Names of the active categories
`K`	Indexes of the bootstrap replications
`dim.x`	Bootstrap coordinates on the first selected axis
`dim.y`	Bootstrap coordinates on the second selected axis

Author(s)

Nicolas Robette

References

Lebart L. (2006). "Validation Techniques in Multiple Correspondence Analysis". In M. Greenacre et J. Blasius (eds), *Multiple Correspondence Analysis and related techniques*, Chapman and Hall/CRC, p.179-196.

Lebart L. (2007). "Which bootstrap for principal axes methods?". In P. Brito et al. (eds), *Selected Contributions in Data Analysis and Classification*, Springer, p.581-588.

See Also

ggbootvalid_supvars, bootvalid_variables

Examples

```
data(Taste)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA",
          "Comedy.NA", "Crime.NA", "Animation.NA", "SciFi.NA", "Love.NA",
          "Musical.NA")
resmca <- speMCA(Taste[,1:11], excl = junk)
supvars <- Taste[,c("Gender", "Age", "Educ")]
bv <- bootvalid_supvars(resmca, supvars, K = 5)
str(bv)
```

---

bootvalid_variables *Bootstrap validation (active variables)*

Description

Bootstrap validation of MCA, through the computation of the coordinates of active variables for bootstrap replications of the data.

Usage

```
bootvalid_variables(resmca, axes = c(1,2), type = "partial", K = 30)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA or bcMCA function
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
type	character string. Can be "partial", "total1", "total2" or "total3" (see details). Default is "partial".
K	integer. Number of bootstrap replications (default is 30).

Details

The bootstrap technique is used here as an internal and non-parametric validation procedure of the results of a multiple correspondence analysis. Following the work of Ludovic Lebart, several methods are proposed. The "total bootstrap" uses new MCAs computed from bootstrap replications of the initial data. In the type 1 total bootstrap (type = "total1"), the sign of the coordinates is corrected if necessary (the direction of the axes of an ACM being arbitrary). In type 2 (type = "total2"), the order of the axes and the sign of the coordinates are corrected if necessary. In type 3 (type = "total3"), a procrustean rotation is used to find the best superposition between the initial axes and the replicated axes. The "partial bootstrap"" (type = "partial") does not compute new MCAs: it projects bootstrap replications of the initial data as supplementary elements of the MCA. It gives a more optimistic view of the stability of the results than the total bootstrap. It also runs faster. See references for more details, pros and cons of the various types, etc.

Value

A data frame with the following elements :

varcat	Names of the active categories
K	Indexes of the bootstrap replications
dim.x	Bootstrap coordinates on the first selected axis
dim.y	Bootstrap coordinates on the second selected axis

Author(s)

Nicolas Robette

References

Lebart L. (2006). "Validation Techniques in Multiple Correspondence Analysis". In M. Greenacre et J. Blasius (eds), *Multiple Correspondence Analysis and related techniques*, Chapman and Hall/CRC, p.179-196.

Lebart L. (2007). "Which bootstrap for principal axes methods?". In P. Brito et al. (eds), *Selected Contributions in Data Analysis and Classification*, Springer, p.581-588.

See Also

ggbootvalid_variables, bootvalid_supvars

Examples

```
data(Taste)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA",
          "Comedy.NA", "Crime.NA", "Animation.NA", "SciFi.NA", "Love.NA",
          "Musical.NA")
resmca <- speMCA(Taste[,1:11], excl = junk)
bv <- bootvalid_variables(resmca, type = "partial", K = 5)
str(bv)
```

---

break_interaction *Additive Breakdowns of Variances*

Description

Computes three additive breakdown of variances for the interaction between two supplementary variables

Usage

```
break_interaction(resmca, v1, v2)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
v1	factor. The first categorical supplementary variable.
v2	factor. The second categorical supplementary variable.

Details

This function reproduces the approach developed in Le Roux & Rouanet (2010) in section 4.4, in particular table 4.5.

Value

A data frame

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

ggadd_interaction, ggadd_partial

Examples

```
# specific MCA of Taste example data set
data(Taste)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA",
          "Comedy.NA", "Crime.NA", "Animation.NA", "SciFi.NA", "Love.NA",
          "Musical.NA")
mca <- speMCA(Taste[,1:11], excl = junk)
# breakdowns of variance
# for the interaction between Gender and Age
break_interaction(mca, Taste$Gender, Taste$Age)
```

burt *Burt table*

Description

Computes a Burt table from a data frame composed of categorical variables.

Usage

```
burt(data)
```

Arguments

data	data frame with n rows (individuals) and p columns (categorical variables)

Details

A Burt table is a symmetric table that is used in correspondence analysis. It shows the frequencies for all combinations of categories of pairs of variables.

Value

Returns a square matrix. Its dimension is equal to the total number of categories in the data frame.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

dichotom

Examples

```
## Burt table of variables in columns 1 to 5
## in the Music example data set
data(Music)
burt(Music[,1:5])
```

coiMCA *Coinertia analysis between two groups of categorical variables*

Description

Coinertia analysis between two groups of categorical variables

Usage

```
coiMCA(Xa, Xb,
       excl.a = NULL, excl.b = NULL,
       row.w = NULL, ncp = 5)
```

Arguments

Xa	data frame with the first group of categorical variables
Xb	data frame with the second group of categorical variables
excl.a	numeric vector indicating the indexes of the "junk" categories in Xa (default is NULL). See getindexcat or use ijunk interactive function to identify these indexes. It may also be a character vector of junk categories, specified in the form "namevariable.namecategory" (for instance "gender.male").
excl.b	numeric vector indicating the indexes of the "junk" categories in Xb (default is NULL). See excl.a argument.
row.w	numeric vector of row weights. If NULL (default), a vector of 1 for uniform row weights is used.
ncp	number of dimensions kept in the results (by default 5)

Details

Coinertia analysis aims at capturing the structure common to two groups of variables. With groups of numerical variables, it is equivalent to Tucker's inter-battery analysis. With categorical data, it consists in the following steps : 1. Transformation of Xa and Xb into indicator matrices (i.e. disjunctive tables) Xad and Xbd 2. Computation of the covariance matrix t(Xad).Xbd 3. CA of the matrix

Value

An object of class CA from FactoMineR package, with an additional item :

RV	the RV coefficient between the two groups of variabels

Author(s)

Nicolas Robette

References

Tucker, L.R.. (1958) An inter-battery method of factor analysis. *Psychometrika*, 23-2, 111-136.

Dolédec, S. and Chessel, D. (1994) Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology*, 31, 277–294.

See Also

coiPCA, rvcoef

Examples

```
data(Music)
# music tastes
Xa <- Music[,1:5]
# gender and age
Xb <- Music[,6:7]
# coinertia analysis
res <- coiMCA(Xa, Xb)
plot(res)
# RV coefficient
res$RV
```

---

coiPCA *Coinertia analysis between two groups of numerical variables*

Description

Coinertia analysis between two groups of numerical variables

Usage

```
coiPCA(Xa, Xb, row.w = NULL, ncp = 5)
```

Arguments

Xa	data frame with the first group of numerical variables
Xb	data frame with the second group of numerical variables
row.w	numeric vector of row weights. If NULL (default), a vector of 1 for uniform row weights is used.
ncp	number of dimensions kept in the results (by default 5)

Details

Coinertia analysis aims at capturing the structure common to two groups of variables. With groups of numerical variables, it is equivalent to Tucker's inter-battery analysis. It consists in the following steps : 1. Variables in Xa and Xb are centered and scaled 2. Computation of the covariance matrix t(Xa).Xb 3. PCA of the matrix

Value

An object of class PCA from FactoMineR package, with an additional item :

RV	the RV coefficient between the two groups of variabels

Author(s)

Nicolas Robette

References

Tucker, L.R. (1958) An inter-battery method of factor analysis. *Psychometrika*, 23-2, 111-136.

Dolédec, S. and Chessel, D. (1994) Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology*, 31, 277–294.

See Also

coiMCA, rvcoef

Examples

```
library(FactoMineR)
data(decathlon)
# variables of results for each sport
Xa <- decathlon[,1:10]
# rank and points variables
Xb <- decathlon[,11:12]
# coinertia analysis
res <- coiPCA(Xa, Xb)
# plot of variables in Xa
plot(res, choix = "ind")
# plot of variables in Xb
plot(res, choix = "var")
# RV coefficient
res$RV
```

---

conc.ellipse *Concentration ellipses*

Description

Adds concentration ellipses or other kinds of inertia ellipses to the cloud of individuals of a MCA.

Usage

```
conc.ellipse(resmca, var, sel = 1:nlevels(var), axes = c(1, 2),
 kappa = 2, col = rainbow(length(sel)), pcol = rainbow(length(sel)), pcex = 0.2,
 lty = 1, lwd = 1, tcex = 1, text.lab = TRUE)
```

Arguments

resmca	object of class MCA, speMCA, csMCA, multiMCA or stMCA
var	supplementary variable to plot
sel	numeric vector of indexes of the categories to plot (by default, ellipses are plotted for every categories)
axes	length 2 vector specifying the components to plot (default is c(1,2))
kappa	numeric. The kappa value (i.e. "index") of the inertia ellipses. By default, kappa = 2, which means that concentration ellipses are plotted.
col	vector of colors for the ellipses of plotted categories (by default, rainbow palette is used)
pcol	vector of colors for the points at the center of ellipses of plotted categories (by default, rainbow palette is used)
pcex	numerical value giving the amount by which points at the center of ellipses should be magnified (default is 0.2)
lty	line type for ellipses (default is 1)
lwd	line width for the ellipses (default is 1)
tcex	numerical value giving the amount by which labels at the center of ellipses should be magnified (default is 0.2)
text.lab	whether the labels at the center of ellipses should be displayed (default is TRUE)

Details

If kappa=2, ellipses are called "concentration" ellipses and, for a normally shaped subcloud, contain 86.47 percents of the points of the subcloud. If kappa=1, ellipses are "indicator" ellipses and contain 39.35 percents of the points of the subcloud. If kappa=1.177, ellipses are "median" ellipses and contain 50 percents of the points of the subcloud. This function has to be used after the cloud of individuals has been drawn.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

plot.speMCA, plot.csMCA, plot.multiMCA, plot.stMCA

Examples

```
## Performs specific MCA (excluding 'NA' categories) of 'Taste' example data set,
## plots the cloud of categories
## and adds concentration ellipses for gender variable
data(Taste)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA",
          "Comedy.NA", "Crime.NA", "Animation.NA", "SciFi.NA", "Love.NA",
          "Musical.NA")
mca <- speMCA(Taste[,1:11], excl = junk)
plot(mca, type = "i")
conc.ellipse(mca, Taste$Gender)

## Draws a blue concentration ellipse for men only
plot(mca, type = "i")
conc.ellipse(mca, Taste$Gender, sel = 1, col = "blue")
```

---

contrib *Contributions of active variables*

Description

From MCA results, computes contributions of categories and variables to the axes and the overall cloud.

Usage

```
contrib(resmca)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA or bcMCA function

Details

The contribution of a point to an axis depends both on the distance from the point to the origin point along the axis and on the weight of the point. The contributions of points to axes are the main aid to interpretation (see Le Roux and Rouanet, 2004 and 2010).

Value

A list of data frames:

ctr	Data frame with the contributions of categories to axes
var.ctr	Data frame with the contributions of variables to axes
ctr.cloud	Data frame with the contributions of categories to the overall cloud
vctr.cloud	Data frame with the contributions of variables to the overall cloud

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

speMCA, supvar, tabcontrib

Examples

```
# specific MCA on the Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# contributions of variables
contrib(mca)
```

---

csMCA *Class Specific Analysis*

Description

Performs a "class specific"" Multiple Correspondence Analysis, i.e. a variant of MCA consisting in analyzing a subcloud of individuals.

Usage

```
csMCA(data, subcloud = rep(TRUE, times = nrow(data)), excl = NULL, ncp = 5,
row.w = rep(1, times = nrow(data)))
```

Arguments

data	data frame with n rows (individuals) and p columns (categorical variables)
subcloud	a vector of logical values and length n. The subcloud of individuals analyzed with class specific MCA is made of the individuals with value TRUE.
excl	nnumeric vector indicating the indexes of the "junk" categories (default is NULL). See getindexcat or use ijunk interactive function to identify these indexes. It may also be a character vector of junk categories, specified in the form "namevariable.namecategory" (for instance "gender.male").
ncp	number of dimensions kept in the results (default is 5)
row.w	an optional numeric vector of row weights (by default, a vector of 1 for uniform row weights)

Details

This variant of MCA is used to study a subset of individuals with reference to the whole set of individuals, i.e. to determine the specific features of the subset. It consists in proceeding to the search of the principal axes of the subcloud associated with the subset of individuals (see references).

Value

An object of class `csMCA`, i.e. a list including:

`eig`	a list of vectors containing all the eigenvalues, the percentage of variance, the cumulative percentage of variance, the modified rates and the cumulative modified rates
`call`	a list with informations about input data
`ind`	a list of matrices containing the results for the individuals (coordinates, contributions)
`var`	a list of matrices containing all the results for the categories and variables (weights, coordinates, squared cosines, categories contributions to axes and cloud, test values (v.test), squared correlation ratio (eta2), variable contributions to axes and cloud

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

`ijunk`, `speMCA`

Examples

```
# class specific MCA of the subcloud of women
# from the Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
female <- Music$Gender=="Women"
mca <- csMCA(Music[,1:5],
             subcloud = female,
             excl = junk)
plot(mca)
```

DA *Discriminant Analysis*

Description

Descriptive discriminant analysis, aka "Analyse Factorielle Discriminante" for the French school of multivariate data analysis.

Usage

```
DA(data, class, row.w = NULL, type = "FR")
```

Arguments

data	data frame with only numeric variables
class	factor specifying the class
row.w	numeric vector of row weights. If NULL (default), a vector of 1 for uniform row weights is used.
type	If "FR" (default), the inverse of the total covariance matrix is used as metric. If "GB", it is the inverse of the within-class covariance matrix (Mahalanobis metric), which makes the results equivalent to linear discriminant analysis as implemented in lda function in MASS package.

Details

The results are the same with type "FR" or "GB", only the eigenvalues vary. With type="FR", these eigenvalues vary between 0 and 1 and can be interpreted as "discriminant power".

Value

An object of class PCA from FactoMineR package, with class as qualitative supplementary variable, and one additional item :

cor_ratio	correlation ratios between class and the discriminant factors

Note

The code is adapted from a script from Marie Chavent. See: https://marie-chavent.perso.math.cnrs.fr/teaching/

Author(s)

Marie Chavent, Nicolas Robette

References

Bry X., 1996, *Analyses factorielles multiples*, Economica.

Lebart L., Morineau A. et Warwick K., 1984, *Multivariate Descriptive Statistical Analysis*, John Wiley and sons, New-York.)

Saporta G., 2006, *Probabilités, analyses des données et statistique*, Editions Technip.

See Also

bcPCA, PCAiv

Examples

```
library(FactoMineR)
data(decathlon)
points <- cut(decathlon$Points, c(7300, 7800, 8000, 8120, 8900), c("Q1","Q2","Q3","Q4"))
res <- DA(decathlon[,1:10], points)
# plot of observations colored by class
plot(res, choix = "ind", invisible = "quali", habillage = res$call$quali.sup$numero)
# plot of class categories
plot(res, choix = "ind", invisible = "ind", col.quali = "darkblue")
# plot of variables
plot(res, choix = "varcor", invisible = "none")
```

DAQ *Discriminant Analysis of Qualitative Variables*

Description

Descriptive discriminant analysis (aka "Analyse Factorielle Discriminante" for the French school of multivariate data analysis) with qualitative variables.

Usage

```
DAQ(data, class, excl = NULL, row.w = NULL,
    type = "FR", select = TRUE)
```

Arguments

data	data frame with only categorical variables
class	factor specifying the class
excl	numeric vector indicating the indexes of the "junk" categories (default is NULL). See getindexcat or use ijunk interactive function to identify these indexes. It may also be a character vector of junk categories, specified in the form "namevariable.namecategory" (for instance "gender.male").
row.w	numeric vector of row weights. If NULL (default), a vector of 1 for uniform row weights is used.

type	character string. If "FR" (default), the inverse of the total covariance matrix is used as metric. If "GB", it is the inverse of the within-class covariance matrix (Mahalanobis metric), which makes the results equivalent to linear discriminant analysis as implemented in lda function in MASS package.
select	logical. If TRUE (default), only a selection of components of the MCA are used for the discriminant analysis step. The selected components are those corresponding to eigenvalues higher of equal to 1/Q, with Q the number of variables in data. If FALSE, all components are used.

Details

This approach is also known as "disqual" and was developed by G. Saporta (see references). It consists in two steps : 1. Multiple Correspondence Analysis of the data 2. Discriminant analysis of the components from the MCA

The results are the same with type "FR" or "GB", only the eigenvalues vary. With type="FR", these eigenvalues vary between 0 and 1 and can be interpreted as "discriminant power".

Value

An object of class PCA from FactoMineR package, with class as qualitative supplementary variable and the disjunctive table of data as quantitative supplementary variables, and two additional items :

cor_ratio	correlation ratios between class and the discriminant factors
mca	an object of class speMCA with the results of the MCA of the first step

Note

If there are NAs in data, these NAs will be automatically considered as junk categories. If one desires more flexibility, data should be recoded to add explicit factor levels for NAs and then excl option may be used to select the junk categories.

Author(s)

Nicolas Robette

References

Bry X., 1996, *Analyses factorielles multiples*, Economica.

Lebart L., Morineau A. et Warwick K., 1984, *Multivariate Descriptive Statistical Analysis*, John Wiley and sons, New-York.)

Saporta G., 1977, "Une méthode et un programme d'analyse discriminante sur variables qualitatives", *Premières Journées Internationales, Analyses des données et informatiques*, INRIA, Rocquencourt.

Saporta G., 2006, *Probabilités, analyses des données et statistique*, Editions Technip.

See Also

DA, bcMCA, MCAiv, speMCA

Examples

```
library(FactoMineR)
data(tea)
res <- DAQ(tea[,1:18], tea$SPC)
# plot of observations colored by class
plot(res, choix = "ind", invisible = "quali",
     label = "quali", habillage = res$call$quali.sup$numero)
# plot of class categories
plot(res, choix = "ind", invisible = "ind", col.quali = "black")
# plot of the variables in data
plot(res, choix = "var", invisible = "var")
# plot of the components of the MCA
plot(res, choix = "varcor", invisible = "quanti.sup")
```

---

dichotom *Dichotomizes the variables in a data frame*

Description

Dichotomizes the variables in a data frame exclusively composed of categorical variables, i.e. transforms the data into an indicator matrix (also known as disjunctive table)

Usage

```
dichotom(data, out = "numeric")
```

Arguments

data	data frame of categorical variables
out	character string defining the format for dichotomized variables in the output data frame. Format may be "numeric" (default) or "factor".

Value

Returns a data frame with dichotomized variables. The number of columns is equal to the total number of categories in the input data.

Author(s)

Nicolas Robette, Julien Barnier

Examples

```
## Dichotomizes Music example data frame
data(Music)
dic <- dichotom(Music[,1:5])
str(dic)
```

```
## with output variables in factor format
dic <- dichotom(Music[,1:5], out='factor')
str(dic)
```

`dichotomixed` *Dichotomizes the factor variables in a mixed format data frame*

Description

Dichotomizes the factor variables in a data frame composed of mixed format variables, i.e. transforms the factors into an indicator matrix (also known as disjunctive table) and keeps the numerical variables.

Usage

```
dichotomixed(data, out = "numeric")
```

Arguments

`data`	data frame of categorical and numerical variables
`out`	character string defining the format for dichotomized variables in the output data frame. Format may be "numeric" (default) or "factor".

Value

Returns a data frame with numerical variables and dichotomized factor variables

Author(s)

Nicolas Robette

Examples

```
## Dichotomizes Music example data frame
data(Music)
## recodes Age as numerical, for the sake of the example
Music$Age <- as.numeric(Music$Age)
## dichotomization
dic <- dichotomixed(Music)
str(dic)
```

dimcontrib *Description of the contributions to axes*

Description

Identifies the categories and individuals that contribute the most to each dimension obtained by a Multiple Correspondence Analysis.

Usage

```
dimcontrib(resmca, dim = c(1,2), best = TRUE)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA or bcMCA function
dim	numerical vector of the dimensions to describe (default is c(1,2))
best	logical. If FALSE, displays all the categories. If TRUE (default), displays only categories and individuals with contributions higher than average

Details

Contributions are sorted and assigned a positive or negative sign according to the corresponding categories or individuals coordinates, so as to facilitate interpretation.

Value

Returns a list with the following items :

var	a list of categories contributions to axes
ind	a list of individuals contributions to axes

Note

Contributions of individuals cannot be computed for objects created by wcMCA function.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

tabcontrib, dimdescr, dimeta2, dimtypicality

Examples

```
# specific MCA on Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# contributions to axes 1 and 2
dimcontrib(mca)
```

---

dimdescr *Description of the dimensions*

Description

Identifies the variables and the categories that are the most characteristic according to each dimension obtained by a MCA. It is inspired by dimdesc function in FactoMineR package (see Husson et al, 2010), but allows to analyze variants of MCA, such as specific MCA or class specific MCA.

Usage

```
dimdescr(resmca, vars = NULL, dim = c(1,2),
         limit = NULL, correlation = "pearson",
         na.rm.cat = FALSE, na.value.cat = "NA", na.rm.cont = FALSE,
         nperm = NULL, distrib = "asympt",
         shortlabs = TRUE)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
vars	data frame of variables to describes the MCA dimensions with. If NULL (default), the active variables of the MCA will be used.
dim	the dimensions which are described. Default is c(1,2)
limit	for the relationship between a dimension and a categorical variable, only associations (measured with point-biserial correlations) higher or equal to limit will be displayed. If NULL (default), they are all displayed.
correlation	character string. The type of correlation measure to be used between two numerical variables : "pearson" (default), "spearman" or "kendall".
na.rm.cat	logical, indicating whether NA values in the categorical variables should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the categorical variables (see na.value.cat argument).
na.value.cat	character string. Name of the level for NA category. Default is "NA". Only used if na.rm.cat = FALSE.

na.rm.cont	logical indicating whether NA values in the numerical variables should be silently removed before the computation proceeds. Default is FALSE.
nperm	numeric. Number of permutations for the permutation tests of independence. If NULL (default), no permutation test is performed.
distrib	the null distribution of permutation test of independence can be approximated by its asymptotic distribution ("asympt", default) or via Monte Carlo resampling ("approx").
shortlabs	logical. If TRUE (default), the data frame will have short column names, so that all columns can be displayed side by side on a laptop screen.

Details

See condesc.

Value

Returns a list of ncp lists including:

variables	associations between dimensions of the MCA and the variables in vars
categories	a data frame with categorical variables from vars and associations measured by correlation coefficients

Author(s)

Nicolas Robette

References

Husson, F., Le, S. and Pages, J. (2010). *Exploratory Multivariate Analysis by Example Using R*, Chapman and Hall.

See Also

condesc, dimcontrib, dimeta2, dimtypicality

Examples

```
# specific MCA on Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# description of the dimensions
dimdescr(mca, limit = 0.1, nperm = 10)
```

dimeta2 *Correlation ratios (aka eta-squared) of supplementary variables*

Description

Computes correlation ratios (also known as eta-squared) for a list of supplementary variables of a MCA.

Usage

```
dimeta2(resmca, vars, dim = c(1,2))
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
vars	a data frame of supplementary variables
dim	the axes for which eta2 are computed. Default is c(1,2)

Value

Returns a data frame with supplementary variables as rows and MCA axes as columns.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

dimdescr, dimcontrib, dimtypicality

Examples

```
# specific MCA on Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# correlation ratios
dimeta2(mca, Music[, c("Gender", "Age")])
```

`dimtypicality` *Typicality tests for supplementary variables*

Description

Computes typicality tests for a list of supplementary variables of a MCA.

Usage

```
dimtypicality(resmca, vars, dim = c(1,2), max.pval = 1)
```

Arguments

`resmca`	object created with `MCA`, `speMCA`, `csMCA`, `wcMCA`, `bcMCA`, `stMCA` or `multiMCA` function
`vars`	a data frame of supplementary variables
`dim`	the axes for which typicality tests are computed. Default is c(1,2)
`max.pval`	only categories with a p-value lower or equal to `max.pval` are displayed. If 1 (default), all categories are displayed

Value

Returns a list of data frames giving the typicality test statistics and p-values of the supplementary categories for the different axes.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

`dimdescr`, `dimeta2`, `dimcontrib`

Examples

```
# specific MCA on Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# typicality tests for gender and age
dimtypicality(mca, Music[, c("Gender", "Age")])
```

dist.chi2 *Chi-squared distance*

Description

Computes the chi-squared distance between the rows of a data frame of factors.

Usage

```
dist.chi2(X)
```

Arguments

X	data frame. All variables should be factors.

Details

This function is adapted from `chi2Dist` function in `ExPosition` package.

Value

A symmetrical matrix of distances

Author(s)

Nicolas Robette

Examples

```
data(Music)
d <- dist.chi2(Music[,1:5])
# a short piece of the distance matrix
d[1:3, 1:3]
```

flip.mca *Flips the coordinates*

Description

Flips the coordinates of the individuals and the categories on one or more dimensions of a MCA.

Usage

```
flip.mca(resmca, dim = 1)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
dim	numerical vector of the dimensions for which the coordinates are flipped. By default, only the first dimension is flipped

Value

Returns an object of the same class as resmca

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

ggcloud_variables, ggcloud_indiv

Examples

```
# MCA of Music example data set
data(Music)
mca <- speMCA(Music[,1:5])
ggcloud_variables(mca, legend = "none")
# Flips dimensions 1 and 2
flipped_mca <- flip.mca(mca, dim = c(1,2))
ggcloud_variables(flipped_mca, legend = "none")
```

---

getindexcat *Names of the categories in a data frame*

Description

Returns a vector of names corresponding the the categories in a data frame exclusively composed of categorical variables.

Usage

```
getindexcat(data)
```

Arguments

data	data frame of categorical variables

Details

This function may be useful prior to a specific MCA, to identify the indexes of the 'junk' categories to exclude.

Value

Returns a character vector with the names of the categories of the variables in the data frame

Author(s)

Nicolas Robette

See Also

ijunk, speMCA, csMCA

Examples

```
data(Music)
getindexcat(Music[,1:5])
mca <- speMCA(Music[,1:5], excl = c(3,6,9,12,15))
```

---

ggadd_attractions *Plot of attractions between categories*

Description

Adds attractions between categories, as measured by phi coefficients or percentages of maximum deviation (PEM), by plotting segments onto a MCA cloud of variables.

Usage

```
ggadd_attractions(p, resmca, axes = c(1,2), measure = "phi", min.asso = 0.3,
col.segment = "lightgray", col.text = "black", text.size = 3)
```

Arguments

p	ggplot2 object with the cloud of variables
resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).

measure	character string. The measure for attractions: "phi" (default) for phi coefficients, "pem" for percentages of maximum deviation (PEM).
min.asso	numerical value ranging from 0 to 1. The minimal attraction value for segments to be plotted. Default is 0.3.
col.segment	Character string with the color of the segments. Default is lightgray.
col.text	Character string with the color of the labels of the categories. Default is black.
text.size	Size of the labels of categories. Default is 3.

Value

a ggplot2 object

Author(s)

Nicolas Robette

References

Cibois, Philippe. *Les méthodes d'analyse d'enquêtes*. Nouvelle édition [en ligne]. Lyon: ENS Éditions, 2014. <http://books.openedition.org/enseditions/1443>

See Also

ggcloud_variables

Examples

```
# specific MCA on Taste example data set
data(Taste)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA",
          "Comedy.NA", "Crime.NA", "Animation.NA", "SciFi.NA", "Love.NA",
          "Musical.NA")
mca <- speMCA(Taste[,1:11], excl = junk)
# Plots attractions
p <- ggcloud_variables(mca, col="white", legend="none")
ggadd_attractions(p, mca, measure="phi", min.asso=0.1)
```

---

ggadd_chulls *Convex hulls for a categorical supplementary variable*

Description

Adds convex hulls for a categorical variable to a MCA cloud of individuals.

Usage

```
ggadd_chulls(p, resmca, var, sel = 1:nlevels(var), axes = c(1,2), prop = 1,
alpha = 0.2, label = TRUE, label.size = 5, legend = "right")
```

Arguments

p	ggplot2 object with the cloud of individuals
resmca	object of class MCA, speMCA, csMCA, stMCA or multiMCA
var	Factor. The categorical variable used to plot chulls.
sel	numeric vector of indexes of the categories to plot (by default, ellipses are plotted for every categories)
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
prop	proportion of all the points to be included in the hull (default is 1).
alpha	Numerical value from 0 to 1. Transparency of the polygon's fill. Default is O.2
label	Logical. Should the labels of the categories be plotted at the center of chulls ? Default is TRUE.
label.size	Size of the labels of the categories at the center of chulls. Default is 5.
legend	the position of legends ("none", "left", "right", "bottom", "top", or two-element numeric vector). Default is right.

Value

a ggplot2 object

Note

Chulls are colored according to the categories of the variable, using the default ggplot2 palette. The palette can be customized using any scale_color_* and scale_fill_* functions, such as scale_color_brewer() and scale_fill_brewer(), scale_color_grey() and scale_fill_grey(), or scale_color_manual() and scale_fill_manual().

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

ggcloud_indiv, ggadd_supvar, ggadd_supvars, ggadd_kellipses, ggadd_ellipses, ggadd_interaction, ggsmoothed_supvar, ggadd_corr, ggadd_density

Examples

```
# specific MCA of Taste example data set
data(Taste)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA",
          "Comedy.NA", "Crime.NA", "Animation.NA", "SciFi.NA", "Love.NA",
          "Musical.NA")
mca <- speMCA(Taste[,1:11], excl = junk)
# hierarchical clustering
# and partition of the individuals into 3 clusters
d <- dist(mca$ind$coord[, c(1,2)])
hca <- hclust(d, "ward.D2")
cluster <- factor(cutree(hca, 3))
# cloud of individuals
# with convex hulls for the clusters.
p <- ggcloud_indiv(mca, col = "black")
ggadd_chulls(p, mca, cluster)
```

---

ggadd_corr *Heatmap of under/over-representation of a supplementary variable*

Description

Adds a heatmap representing the correlation coefficients to a MCA cloud of individuals, for a numerical supplementary variable or one category of a categorical supplementary variable.

Usage

```
ggadd_corr(p, resmca, var, cat = levels(var)[1], axes = c(1,2),
xbins = 20, ybins = 20, min.n = 1, pal = "RdYlBu", limits = NULL, legend = "right")
```

Arguments

p	ggplot2 object with the cloud of variables
resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
var	factor or numerical vector. The supplementary variable used for the heatmap.
cat	character string. The category of var to plot (by default, the first level of var is plotted). Only used if var is a factor.
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
xbins	integer. Number of bins in the x axis. Default is 20.
ybins	integer. Number of bins in the y axis. Default is 20.
min.n	integer. Minimal number of points for a tile to be drawn. By default, every tiles are drawn.
pal	character string. Name of a (preferably diverging) palette from the RColorBrewer package. Default is "RdYlBu".

limits	numerical vector of length 2. Lower and upper limits of the correlation coefficients for the color scale. Should be centered around 0 for a better view of under/over-representations (for example c(-0.2,0.2)). By default, the maximal absolute value of the correlation coefficients is used.
legend	the position of legends ("none", "left", "right", "bottom", "top", or two-element numeric vector). Default is right.

Details

For each tile of the heatmap, a correlation coefficient is computed between the supplementary variable and the fact of belonging to the tile. This gives a view of the under/over-representation of the supplementary variable according to the position in the cloud of individuals.

Value

a ggplot2 object

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

ggcloud_variables, ggadd_supvar, ggadd_supvars, ggadd_kellipses, ggadd_ellipses, ggadd_interaction, ggsmoothed_supvar, ggadd_chulls, ggadd_density

Examples

```
# specific MCA of Taste example data set
data(Taste)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA",
          "Comedy.NA", "Crime.NA", "Animation.NA", "SciFi.NA", "Love.NA",
          "Musical.NA")
mca <- speMCA(Taste[,1:11], excl = junk)
# correlation heatmap for Age = 50+
p <- ggcloud_indiv(mca, col = "lightgrey")
ggadd_corr(p, mca, var = Taste$Age, cat = "50+", xbins = 10, ybins = 10)
```

ggadd_density *Density plot of a supplementary variable*

Description

For a given category of a supplementary variable, adds a layer representing the density of points to the cloud of individuals, either with contours or areas.

Usage

```
ggadd_density(p, resmca, var, cat = levels(var)[1], axes = c(1,2),
density = "contour", col.contour = "darkred", pal.area = "viridis",
alpha.area = 0.2, ellipse = FALSE)
```

Arguments

p	ggplot2 object with the cloud of variables
resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
var	factor or numerical vector. The supplementary variable to be plotted.
cat	character string. The category of var to plot (by default, the first level of var is plotted). Only used if var is a factor.
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
density	If "contour" (default), density is plotted with contours. If "area", density is plotted with areas.
col.contour	character string. The color of the contours.
pal.area	character string. The name of a viridis palette for areas.
alpha.area	numeric. Transparency of the areas. Default is 0.2.
ellipse	logical. If TRUE, a concentration ellipse is added.

Value

a ggplot2 object

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

ggcloud_indiv, ggadd_supvar, ggadd_supvars, ggadd_kellipses, ggadd_ellipses, ggadd_interaction, ggsmoothed_supvar, ggadd_chulls, ggadd_corr

Examples

```
# specific MCA of Taste example data set
data(Taste)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA",
          "Comedy.NA", "Crime.NA", "Animation.NA", "SciFi.NA", "Love.NA",
          "Musical.NA")
mca <- speMCA(Taste[,1:11], excl = junk)
p <- ggcloud_indiv(mca, col='lightgrey')
# density plot for Age = 50+ (with contours)
ggadd_density(p, mca, var = Taste$Age, cat = "50+")
# density plot for Age = 50+ (with contours)
ggadd_density(p, mca, var = Taste$Age, cat = "50+", density = "area")
```

---

ggadd_ellipses *Confidence ellipses*

Description

Adds confidence ellipses for a categorical variable to a MCA cloud of individuals

Usage

```
ggadd_ellipses(p, resmca, var, sel = 1:nlevels(var), axes = c(1,2),
level = 0.05, label = TRUE, label.size = 3, size = 0.5, points = TRUE,
legend = "right")
```

Arguments

p	ggplot2 object with the cloud of individuals
resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
var	Factor. The categorical variable used to plot ellipses.
sel	numeric vector of indexes of the categories to plot (by default, ellipses are plotted for every categories)
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
level	The level at which to draw an ellipse (see stat_ellipse). Default is 0.05, which means 95 percents confidence ellipses are plotted.
label	Logical. Should the labels of the categories be plotted at the center of ellipses ? Default is TRUE.
label.size	Size of the labels of the categories at the center of ellipses. Default is 3.

size	Size of the lines of the ellipses. Default is 0.5.
points	If TRUE (default), the points are coloured according to their subcloud.
legend	the position of legends ("none", "left", "right", "bottom", "top", or two-element numeric vector). Default is right.

Details

A confidence ellipse aims at measuring how the "true" mean point of a category differs from its observed mean point. This is achieved by constructing a confidence zone around the observed mean point. If we choose a conventional level alpha (e.g. 0.05), a (1 - alpha) (e.g. 95 percents) confidence zone is defined as the set of possible mean points that are not significantly different from the observed mean point.

Value

a ggplot2 object

Note

Ellipses are colored according to the categories of the variable, using the default ggplot2 palette. The palette can be customized using any scale_color_* function, such as scale_color_brewer(), scale_color_grey() or scale_color_manual().

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

ggcloud_indiv, ggadd_supvar, ggadd_supvars, ggadd_kellipses, ggadd_density, ggadd_interaction, ggsmoothed_supvar, ggadd_chulls, ggadd_corr

Examples

```
# specific MCA of Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# confidence ellipses for Age
p <- ggcloud_indiv(mca, col = "lightgrey")
ggadd_ellipses(p, mca, Music$Age)
```

ggadd_interaction *Plot of interactions between two categorical supplementary variables*

Description

Adds the interactions between two categorical supplementary variables to a MCA cloud of variables

Usage

```
ggadd_interaction(p, resmca, v1, v2, sel1 = 1:nlevels(v1), sel2 = 1:nlevels(v2),
axes = c(1,2), cloud = "v", textsize = 5, lines = TRUE, dashes = TRUE,
legend = "none", force = 1, max.overlaps = Inf)
```

Arguments

p	ggplot2 object with the cloud of variables or the cloud of individuals
resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
v1	factor. The first categorical supplementary variable.
v2	factor. The second categorical supplementary variable.
sel1	numeric vector of indexes of the categories of the first supplementary variable to be used in interaction. By default, every categories are used.
sel2	numeric vector of indexes of the categories of the second supplementary variable to be used in interaction. By default, every categories are used.
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
cloud	if "v" (default), the categories are plotted in the cloud of variables. If "i", the categories are plotted in the cloud of individuals.
textsize	size of the labels of categories. Default is 5.
lines	logical. Whether to add colored lines between the points of the categories of v1. Default is TRUE.
dashes	logical. Whether to add gray dashed lines between the points of the categories of v2. Default is TRUE.
legend	the position of legends ("none", "left", "right", "bottom", "top", or two-element numeric vector). Default is none.
force	force of repulsion between overlapping text labels. Defaults to 1. If 0, labels are not repelled at all.
max.overlaps	exclude text labels that overlap too many things. Defaults to Inf, which means no labels are excluded.

Value

a ggplot2 object

Note

Lines and labels are colored according to the first variable, using the default ggplot2 palette. The palette can be customized using any scale_color_* function, such as scale_color_brewer(), scale_color_grey() or scale_color_manual().

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

ggcloud_variables, ggadd_supvar, ggadd_supvars, ggadd_kellipses, ggadd_ellipses, ggadd_corr, ggsmoothed_supvar, ggadd_chulls, ggadd_density

Examples

```
# specific MCA of Taste example data set
data(Taste)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA",
          "Comedy.NA", "Crime.NA", "Animation.NA", "SciFi.NA", "Love.NA",
          "Musical.NA")
mca <- speMCA(Taste[,1:11], excl = junk)
# interaction between Gender and Age
p <- ggcloud_variables(mca, col = "lightgrey", shapes = FALSE)
ggadd_interaction(p, mca, Taste$Gender, Taste$Age)
```

---

ggadd_kellipses *Concentration ellipses and k-inertia ellipses*

Description

Adds concentration ellipses and other kinds of k-inertia ellipses for a categorical variable to a MCA cloud of individuals.

Usage

```
ggadd_kellipses(p, resmca, var, sel = 1:nlevels(var), axes = c(1,2),
kappa = 2, label = TRUE, label.size = 3, size = 0.5, points = TRUE,
legend = "right")
```

Arguments

p	ggplot2 object with the cloud of individuals
resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
var	Factor. The categorical variable used to plot ellipses.
sel	numeric vector of indexes of the categories to plot (by default, ellipses are plotted for every categories)
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
kappa	numeric. The kappa value (i.e. "index") of the inertia ellipses. By default, kappa = 2, which means that concentration ellipses are plotted.
label	Logical. Should the labels of the categories be plotted at the center of ellipses ? Default is TRUE.
label.size	Size of the labels of the categories at the center of ellipses. Default is 3.
size	Size of the lines of the ellipses. Default is 0.5.
points	If TRUE (default), the points are coloured according to their subcloud.
legend	the position of legends ("none", "left", "right", "bottom", "top", or two-element numeric vector). Default is right.

Details

If kappa=2, ellipses are called "concentration" ellipses and, for a normally shaped subcloud, contain 86.47 percents of the points of the subcloud. If kappa=1, ellipses are "indicator" ellipses and contain 39.35 percents of the points of the subcloud. If kappa=1.177, ellipses are "median" ellipses and contain 50 percents of the points of the subcloud. This function has to be used after the cloud of individuals has been drawn.

Value

a ggplot2 object

Note

Ellipses are colored according to the categories of the variable, using the default ggplot2 palette. The palette can be customized using any scale_color_* function, such as scale_color_brewer(), scale_color_grey() or scale_color_manual().

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

ggcloud_indiv, ggadd_supvar, ggadd_supvars, ggadd_ellipses, ggadd_density, ggadd_interaction, ggsmoothed_supvar, ggadd_chulls, ggadd_corr

Examples

```
# specific MCA of Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# concentration ellipses for Age
p <- ggcloud_indiv(mca, col = "lightgrey")
ggadd_ellipses(p, mca, Music$Age)
```

---

ggadd_partial *Main and partial effect of a supplementary variable*

Description

Adds the main and partial effects of a supplementary variable to a MCA cloud of individuals, with one or more supplementary partialled out

Usage

```
ggadd_partial(p, resmca, var, controls, excl = NULL,
axes = c(1,2), col = "black", textsize = 4, lines = TRUE, dashes = TRUE,
legend = "right", force = 1, max.overlaps = Inf)
```

Arguments

p	ggplot2 object with the cloud of variables or the cloud of individuals
resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
var	factor. The categorical supplementary variable.
controls	data frame of supplementary variables to be partialled out (i.e. control variables)
excl	character vector of categories from the var to exclude from the plot. If NULL (default), all the supplementary categories are plotted.
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
col	the color for the labels and lines. Default is "black".
textsize	size of the labels of categories. Default is 4.
lines	logical. Whether to add colored lines between the points of the categories of v1. Default is TRUE.
dashes	logical. Whether to add gray dashed lines between the points of the categories of v2. Default is TRUE.

legend	the position of legends ("none", "left", "right", "bottom", "top", or two-element numeric vector). Default is right.
force	force of repulsion between overlapping text labels. Defaults to 1. If 0, labels are not repelled at all.
max.overlaps	exclude text labels that overlap too many things. Defaults to Inf, which means no labels are excluded.

Value

a ggplot2 object

Note

The partial effects of the supplementary variable are computed with the Average Marginal Effects of a linear regression, with individual coordinates as dependent variable, and the supplementary and control variables as independent variables.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

ggcloud_variables, ggadd_supvar, ggadd_supvars, ggadd_interaction

Examples

```
# specific MCA of Taste example data set
data(Taste)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA",
          "Comedy.NA", "Crime.NA", "Animation.NA", "SciFi.NA", "Love.NA",
          "Musical.NA")
mca <- speMCA(Taste[,1:11], excl = junk)
# effect of education,
# with age partialled out (partial effect) or not (main effect)
p <- ggcloud_indiv(mca, col = "grey95")
ggadd_partial(p, mca, Taste$Educ, Taste$Age)
```

ggadd_supind *Plot of supplementary individuals*

Description

Adds supplementary individuals to a MCA cloud of the individuals

Usage

```
ggadd_supind(p, resmca, dfsup, axes = c(1,2),
col = "black", textsize = 5, pointsize = 2)
```

Arguments

p	ggplot2 object with the cloud of individuals.
resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
dfsup	data frame with the supplementary individuals. It must have the same factors as the data frame used as input for the initial MCA.
axes	numeric vector of length 2, specifying the dimensions (axes) to plot (default is c(1,2))
col	color for the labels and points of the individuals (default is black)
textsize	Size of the labels of the individuals. Default is 5.
pointsize	Size of the points of the individuals. If NULL, only labels are plotted. Default is 2.

Details

The function uses the row names of dfsup as labels for the individuals.

Author(s)

Nicolas Robette

See Also

supind, ggcloud_indiv

Examples

```
# specific MCA of Music example data set
data(Music)
rownames(Music) <- paste0("i", 1:nrow(Music))
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# adds individuals 1, 20 and 300 as supplementary individuals
```

```
# onto the cloud of individuals
p <- ggcloud_indiv(mca, col = "lightgrey")
ggadd_supind(p, mca, Music[c(1,20,300), 1:5])
```

---

ggadd_supvar *Plot of a categorical supplementary variable*

Description

Adds a categorical supplementary variable to a MCA cloud of variables.

Usage

```
ggadd_supvar(p, resmca, var, sel = 1:nlevels(var), axes = c(1,2),
col = "black", shape = 1, prop = NULL, textsize = 3, shapesize = 6,
segment = FALSE, vname = NULL)
```

Arguments

p	ggplot2 object with the cloud of variables
resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
var	Factor. The categorical supplementary variable. It does not need to have been used at the MCA step.
sel	Numeric vector of indexes of the categories of the supplementary variable to be added to the plot. By default, labels are plotted for every categories.
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
col	Character. Color of the shapes and labels of the categories. Default is black.
shape	Symbol to be used in addition the the labels of categories (default is 1). If NULL, only labels are plotted.
prop	If NULL, the size of the labels (if shape=NULL) or the shapes (otherwise) is constant. If 'n', the size is proportional the the weights of categories; if 'vtest1', the size is proportional to the test values of the categories on the first dimension of the plot; if 'vtest2', the size is proportional to the test values of the categories on the second dimension of the plot; if 'cos1', the size is proportional to the cosines of the categories on the first dimension of the plot; if 'cos2', the size is proportional to the cosines of the categories on the second dimension of the plot; if 'cos12', the size is proportional to the total cosines of the categories on the two dimensions of the plot.
textsize	Size of the labels of categories if shape is not NULL, or if shape=NULL and prop=NULL. Default is 3.
shapesize	Size of the shapes if prop=NULL, maximum size of the shapes in other cases. Default is 6.
segment	Logical. Should one add lines between categories ? Default is FALSE.
vname	A character string to be used as a prefix for the labels of the categories. If NULL (default), no prefix is added.

Value

a ggplot2 object

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

ggcloud_variables, ggadd_supvars, ggadd_ellipses, ggadd_kellipses, ggadd_density, ggadd_interaction, ggsmoothed_supvar, ggadd_chulls, ggadd_corr

Examples

```
# specific MCA of Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# adds Age as a supplementary variable
# onto the cloud of variables
p <- ggcloud_variables(mca, col = "lightgrey", shapes = FALSE)
ggadd_supvar(p, mca, Music$Age, segment = TRUE)
```

---

ggadd_supvars *Plot of categorical supplementary variables*

Description

Adds categorical supplementary variables to a MCA cloud of variables.

Usage

```
ggadd_supvars(p, resmca, vars, excl = NULL, points = "all", min.cos2 = 0.1,
axes = c(1,2), col = NULL,
shapes = FALSE, prop = NULL, textsize = 3, shapesize = 6,
vlab = TRUE, vname = NULL,
force = 1, max.overlaps = Inf)
```

Arguments

p	ggplot2 object with the cloud of variables
resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
vars	A data frame of categorical supplementary variables. All these variables should be factors.
excl	character vector of supplementary categories to exclude from the plot, specified in the form "namevariable.namecategory" (for instance "Gender.Men"). If NULL (default), all the supplementary categories are plotted.
points	character string. If 'all' all categories are plotted (default); if 'besth' only those with a minimum squared cosine on horizontal axis are plotted; if 'bestv' only those with a minimum squared cosine on vertical axis are plotted; if 'besthv' only those with a minimum squared cosine on horizontal or vertical axis are plotted; if 'best' only those with a minimum squared cosine on the plane are plotted.
min.cos2	numerical value. The minimal squared cosine if 'points' argument is different from 'all'. Default
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
col	character string. Color name for the labels (and the shapes if shapes=TRUE) of the categories. If NULL, the default palette of ggplot2 is used, with one color per variable.
shapes	Logical. If TRUE, symbols are used in addition to the labels of categories. Default is FALSE.
prop	If NULL, the size of the labels (if shapes=FALSE), or of the labels and the shapes (if shapes=TRUE) is constant. If 'n', the size is proportional the the weights of categories; if 'vtest1', the size is proportional to the test values of the categories on the first dimension of the plot; if 'vtest2', the size is proportional to the test values of the categories on the second dimension of the plot; if 'cos1', the size is proportional to the cosines of the categories on the first dimension of the plot; if 'cos2', the size is proportional to the cosines of the categories on the second dimension of the plot; if 'cos12', the size is proportional to the total cosines of the categories on the two dimensions of the plot.
textsize	Size of the labels of categories if shapes is TRUE, or if shapes is FALSE and prop is NULL. Default is 3.
shapesize	Size of the shapes if prop=NULL, maximum size of the shapes in other cases. Default is 6.
vlab	Logical. If TRUE (default), the variable name is added as a prefix for the labels of the categories.
vname	deprecated, use vlab instead
force	Force of repulsion between overlapping text labels. Defaults to 1. If 0, labels are not repelled at all.
max.overlaps	Exclude text labels that overlap too many things. Defaults to Inf, which means no labels are excluded.

Value

a ggplot2 object

Note

Shapes and labels are colored according to the categories of the variable, using the default ggplot2 palette. The palette can be customized using any scale_color_* function, such as scale_color_brewer(), scale_color_grey() or scale_color_manual().

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

ggcloud_variables, ggadd_supvar, ggadd_ellipses, ggadd_kellipses, ggadd_density, ggadd_interaction, ggsmoothed_supvar, ggadd_chulls, ggadd_corr

Examples

```
# specific MCA of Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# adds several supplementary variables
# onto the cloud of variables
p <- ggcloud_variables(mca, col = "lightgrey", shapes = FALSE)
ggadd_supvars(p, mca, Music[, c("Gender","Age")])
# the same, excluding men
ggadd_supvars(p, mca, Music[, c("Gender","Age")], excl = "Gender.Men")
# the same, keeping only categories
# with cos2 >= 0.001 for dimension 1
ggadd_supvars(p, mca, Music[, c("Gender","Age")], points = "besth", min.cos2 = 0.001)
```

---

ggaxis_variables *Plot of variables on a single axis*

Description

Plots variables on a single axis of a Multiple Correspondence Analysis. Variables can be active or supplementary.

Usage

```
ggaxis_variables(resmca, var = NULL, axis = 1,
min.ctr = NULL, prop = NULL,
underline = FALSE, col = NULL, vlab = TRUE,
force = 1, max.overlaps = Inf)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
var	If NULL (default), all the active variables of the MCA are plotted. If a character string, the named active variables of the MCA is plotted. If a factor, it is plotted as a supplementary variable.
axis	numeric value. The MCA axis to plot. Default is 1.
min.ctr	If NULL (default), all the categories are displayed. If "best", only the categories that contribute more than the average (i.e. 100 / number of categories) are displayed. If a numerical value between 0 and 100, only categories that contribute more than min.ctr are displayed.
prop	If NULL (default), the size of the labels is constant. If "freq", the size is proportional to the weights of categories. If "ctr", it's proportional to the contributions of categories (only used for active variables). If "cos2", it's proportional to the squared cosines of the categories. If "pval", it's proportional to 1 minus the p-values of typicality tests (only used for supplementary variables). If "cor", it's proportional to the point biserial correlation of the categories (only used for supplementary variables).
underline	logical. If TRUE, the labels of the categories with contributions above average are underlined. Default is FALSE. Only used for active variables.
col	character string. Color name for the labels of the categories. If NULL and var=NULL, the default palette of ggplot2 is used, with one color per variable. If NULL and var is not NULL, labels are black.
vlab	Logical. Should the variable names be used as a prefix for the labels of the categories. Default is TRUE.
force	Force of repulsion between overlapping text labels. Defaults to 1. If 0, labels are not repelled at all.
max.overlaps	Exclude text labels that overlap too many things. Defaults to Inf, which means no labels are excluded.

Value

a ggplot2 object

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

ggcloud_variables

Examples

```
# specific MCA of Taste example data set
data(Taste)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA",
          "Comedy.NA", "Crime.NA", "Animation.NA", "SciFi.NA", "Love.NA",
          "Musical.NA")
mca <- speMCA(Taste[,1:11], excl = junk)
# plots all the active categories on axis 1
ggaxis_variables(mca)
# the same with other plotting options
ggaxis_variables(mca, prop = "freq", underline = TRUE, col = "black")
# plots Active variable Classical on axis 1
ggaxis_variables(mca, var = "Classical", axis = 1, prop = "ctr", underline = TRUE)
# plots supplementary variable Educ on axis 1
ggaxis_variables(mca, var = Taste$Educ, axis = 1, prop = "pval")
```

---

ggbootvalid_supvars *Ellipses of bootstrap validation (supplementary variables)*

Description

Ellipses for bootstrap validation of MCA, through the computation of the coordinates of supplementary variables for bootstrap replications of the data.

Usage

```
ggbootvalid_supvars(resmca, vars = NULL, axes = c(1,2), K = 30,
                    ellipse = "norm", level = 0.95,
                    col = NULL, active = FALSE, legend = "right",
                    aspect = TRUE)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA or bcMCA function
vars	A data frame of categorical supplementary variables. All these variables should be factors.

axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
K	integer. Number of bootstrap replications (default is 30).
ellipse	character string. The type of ellipse. The default "norm" assumes a multivariate normal distribution, "t" assumes a multivariate t-distribution, and "euclid" draws a circle with the radius equal to level, representing the euclidean distance from the center.
level	numerical value. The level at which to draw an ellipse, or, if ellipse="euclid", the radius of the circle to be drawn.
col	Character string. Color name for the ellipses and labels of the categories. If NULL (default), the default ggplot2 palette is used, with one color per variable.
active	logical. If TRUE, the labels of active variables are added to the plot in lightgray. Default is FALSE.
legend	the position of legends ("none", "left", "right", "bottom", "top", or two-element numeric vector). Default is right.
aspect	Logical. TRUE (default) ensures that one unit on the x-axis is the same length as one unit on the y-axis.

Details

The bootstrap technique is used here as an internal (and non-parametric) validation procedure of the results of a multiple correspondence analysis. For supplementary variables, only partial bootstrap is possible. The partial bootstrap does not compute new MCAs: it projects bootstrap replications of the initial data as supplementary elements of the MCA. See references for more details.

The default parameters for ellipses assume a multivariate normal distribution drawn at level 0.95.

Value

a ggplot2 object

Note

If col argument is NULL, ellipses and labels are colored according to the variables, using the default ggplot2 palette. The palette can be customized using any scale_color_* function, such as scale_color_brewer(), scale_color_grey() or scale_color_manual().

Author(s)

Nicolas Robette

References

Lebart L. (2006). "Validation Techniques in Multiple Correspondence Analysis". In M. Greenacre et J. Blasius (eds), *Multiple Correspondence Analysis and related techniques*, Chapman and Hall/CRC, p.179-196.

Lebart L. (2007). "Which bootstrap for principal axes methods?". In P. Brito et al. (eds), *Selected Contributions in Data Analysis and Classification*, Springer, p.581-588.

See Also

bootvalid_supvars, ggbootvalid_variables

Examples

```
# specific MCA of Taste example data set
data(Taste)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA",
          "Comedy.NA", "Crime.NA", "Animation.NA", "SciFi.NA", "Love.NA",
          "Musical.NA")
mca <- speMCA(Taste[,1:11], excl = junk)
# bootstrap validation ellipses
# for three supplementary variables
sup <- Taste[,c("Gender", "Age", "Educ")]
ggbootvalid_supvars(mca, sup)
```

---

ggbootvalid_variables *Ellipses of bootstrap validation (active variables)*

Description

Ellipses for bootstrap validation of MCA, through the computation of the coordinates of active variables for bootstrap replications of the data.

Usage

```
ggbootvalid_variables(resmca, axes = c(1,2), type = "partial", K = 30,
                      ellipse = "norm", level = 0.95,
                      col = NULL, legend = "right", aspect = TRUE)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA or bcMCA function
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
type	character string. Can be "partial", "total1", "total2" or "total3" (see details). Default is "partial".
K	integer. Number of bootstrap replications (default is 30).
ellipse	character string. The type of ellipse. The default "norm" assumes a multivariate normal distribution, "t" assumes a multivariate t-distribution, and "euclid" draws a circle with the radius equal to level, representing the euclidean distance from the center.
level	numerical value. The level at which to draw an ellipse, or, if ellipse="euclid", the radius of the circle to be drawn.
col	Character string. Color name for the ellipses and labels of the categories. If NULL (default), the default ggplot2 palette is used, with one color per variable.

legend	the position of legends ("none", "left", "right", "bottom", "top", or two-element numeric vector). Default is right.
aspect	Logical. TRUE (default) ensures that one unit on the x-axis is the same length as one unit on the y-axis.

Details

The bootstrap technique is used here as an internal (and non-parametric) validation procedure of the results of a multiple correspondence analysis. Following the work of Lebart, several methods are proposed. The total bootstrap uses new MCAs computed from bootstrap replications of the initial data. In the type 1 bootstrap (type = "total1"), the sign of the coordinates is corrected if necessary (the direction of the axes of an ACM being arbitrary). In type 2 (type = "total2"), the order of the axes and the sign of the coordinates are corrected if necessary. In type 3 (type = "total3"), a procrustean rotation is used to find the best superposition between the initial axes and the replicated axes. The partial bootstrap (type = "partial") does not compute new MCAs: it projects bootstrap replications of the initial data as supplementary elements of the MCA. It gives a more optimistic view of the stability of the results than the total bootstrap. It is also faster. See references for more details, pros and cons of the various types, etc.

The default parameters for ellipses assume a multivariate normal distribution drawn at level 0.95.

Value

a ggplot2 object

Note

If col argument is NULL, ellipses and labels are colored according to the variables, using the default ggplot2 palette. The palette can be customized using any scale_color_* function, such as scale_color_brewer(), scale_color_grey() or scale_color_manual().

Author(s)

Nicolas Robette

References

Lebart L. (2006). "Validation Techniques in Multiple Correspondence Analysis". In M. Greenacre et J. Blasius (eds), *Multiple Correspondence Analysis and related techniques*, Chapman and Hall/CRC, p.179-196.

Lebart L. (2007). "Which bootstrap for principal axes methods?". In P. Brito et al. (eds), *Selected Contributions in Data Analysis and Classification*, Springer, p.581-588.

See Also

bootvalid_variables , ggbootvalid_supvars

Examples

```
# specific MCA of Taste example data set
data(Taste)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA",
          "Comedy.NA", "Crime.NA", "Animation.NA", "SciFi.NA", "Love.NA",
          "Musical.NA")
mca <- speMCA(Taste[,1:11], excl = junk)
# bootstrap validation ellipses for active variables
ggbootvalid_variables(mca, type = "partial", K = 5)
```

---

ggcloud_indiv *Plot of the cloud of individuals*

Description

Plots a Multiple Correspondence Analysis cloud of individuals.

Usage

```
ggcloud_indiv(resmca, type = "i", points = "all", axes = c(1,2),
col = "dodgerblue4", point.size = 0.5, alpha = 0.6,
repel = FALSE, text.size = 2,
density = NULL, col.contour = "darkred", hex.bins = 50,
hex.pal = "viridis", aspect = TRUE)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
type	If 'i', points are plotted. If 'inames', labels of individuals are plotted.
points	character string. If 'all' all points are plotted (default). If 'besth' only those who contribute most to horizontal axis are plotted. If 'bestv' only those who contribute most to vertical axis are plotted. If 'besthv' only those who contribute most to horizontal or vertical axis are plotted. If 'best' only those who contribute most to the plane are plotted.
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
col	If a factor, points or labels are colored according to their category regarding this factor. If a string with color name, every points or labels have the same color. Default is "dodgerblue4".
point.size	Size of the points of individuals. Default is 0.5.
alpha	Transparency of the points or labels of individuals. Default is 0.6.
repel	Logical. When type="inames"", should labels of individuals be repeled ? Default is FALSE.
text.size	Size of the labels of individuals. Default is 2.

density	If NULL (default), no density layer is added. If "contour", density is plotted with contours. If "hex", density is plotted with hexagon bins.
col.contour	character string. The color of the contours. Only used if density="contour".
hex.bins	integer. The number of bins in both vertical and horizontal directions. Only used if density="hex".
hex.pal	character string. The name of a viridis palette for hexagon bins. Only used if density="hex".
aspect	Logical. TRUE (default) ensures that one unit on the x-axis is the same length as one unit on the y-axis.

Details

Sometimes the dots are too many and overlap. It is then difficult to get an accurate idea of the distribution of the cloud of individuals. The density argument allows you to add an additional layer to represent the density of points in the plane, in the form of contours or hexagonal areas.

Value

a ggplot2 object

Note

If col argument is a factor, points or labels are colored according to the categories of the factor, using the default ggplot2 palette. The palette can be customized using any scale_color_* function, such as scale_color_brewer(), scale_color_grey() or scale_color_manual().

Author(s)

Anton Perdoncin, Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

ggcloud_variables

Examples

```
# specific MCA of Taste example data set
data(Taste)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA",
          "Comedy.NA", "Crime.NA", "Animation.NA", "SciFi.NA", "Love.NA",
          "Musical.NA")
mca <- speMCA(Taste[,1:11], excl = junk)
```

```
# cloud of individuals
ggcloud_indiv(mca)
# points are colored according to gender
ggcloud_indiv(mca, col=Taste$Gender)
# a density layer of contours is added
ggcloud_indiv(mca, density = "contour")
# a density layer of hexagon bins is added
ggcloud_indiv(mca, density = "hex", hex.bin = 10)
```

`ggcloud_variables` *Plot of the cloud of variables*

Description

Plots a Multiple Correspondence Analysis cloud of variables.

Usage

```
ggcloud_variables(resmca, axes = c(1,2), points = "all",
min.ctr = NULL, max.pval = 0.01, face = "pp",
shapes = TRUE, prop = NULL, textsize = 3, shapesize = 3,
col = NULL, col.by.group = TRUE, alpha = 1,
segment.alpha = 0.5, vlab = TRUE, sep = ".", legend = "right",
force = 1, max.overlaps = Inf, aspect = TRUE)
```

Arguments

`resmca`	object created with `MCA`, `speMCA`, `csMCA`, `wcMCA`, `bcMCA`, `stMCA` or `multiMCA` function
`axes`	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
`points`	character string. If 'all' all categories are plotted (default); if 'besth' only those who contribute most to horizontal axis are plotted; if 'bestv' only those who contribute most to vertical axis are plotted; if 'besthv' only those who contribute most to horizontal or vertical axis are plotted; if 'best' only those who contribute most to the plane are plotted.
`min.ctr`	Numerical value between 0 and 100. The minimum contribution (in percent) for a category to be displayed if the `points` argument is equal to "best", "besth" or "bestv" and `resmca` is of type `MCA`, `speMCA` or `csMCA`. If NULL (default), only the categories that contribute more than the average (i.e. 100 / number of modalities) are displayed.
`max.pval`	Numerical value between 0 and 100. The maximal p-value derived from test-values for a category to be displayed if the `points` argument is equal to "best", "besth" or "bestv" and `resmca` is of type `stMCA` or `multiMCA`.

face	character string. Changes the face of the category labels when their contribution is greater than min.ctr. The first letter refers to the first represented axis, the second letter to the second. "p" is for plain text, "u" for underlined, "i" for italic and "b" for bold. For example, "ui" means that the labels of the most contributing categories on the first axis will be underlined and the labels of the most contributing categories on the second axis will be italicized. By default ("pp"), no font face change is made.
shapes	Logical. Should shapes be plotted for categories (in addition to labels) ? Default is TRUE.
prop	If NULL, the size of the labels (if shapes=FALSE) or the shapes (if shapes=TRUE) is constant. If 'n', the size is proportional the the weights of categories; if 'ctr1', the size is proportional to the contributions of the categories on the first dimension of the plot; if 'ctr2', the size is proportional to the contributions of the categories on the second dimension of the plot; if 'ctr12', the size is proportional to the contributions of the categories on the plane ; if 'ctr.cloud', the size is proportional to the total contributions of the categories on the whole cloud; if 'cos1', the size is proportional to the quality of representation (squared cosines) of the categories on the first dimension of the plot; if 'cos2', the size is proportional to the quality of representation of the categories on the second dimension of the plot; if 'cos12', the size is proportional to the quality of representation of the categories on the plane; if 'vtest1', the size is proportional to the test-values of the categories on the first dimension of the plot; if 'vtest2', the size is proportional to the test-values of the categories on the second dimension of the plot.
textsize	Size of the labels of categories if shapes=TRUE, or if shapes=FALSE and prop=NULL. Default is 3.
shapesize	Size if the shapes of categories if shapes=TRUE and prop=FALSE. Default is 3.
col	Character string. Color name for the shapes and labels of the categories. If NULL (default), the default ggplot2 palette is used, with one color per variable.
col.by.group	Logical. If resmca is of type multimCA, categories are colored by group from the MFA if TRUE (default) and by variable if FALSE.
alpha	Transparency of the shapes and labels of categories. Default is 1.
segment.alpha	Transparency of the line segment beside labels of categories. Default is 0.5.
vlab	Logical. Should the variable names be used as a prefix for the labels of the categories. Default is TRUE.
sep	Character string used as a separator if vlab=TRUE.
legend	the position of legends ("none", "left", "right", "bottom", "top", or two-element numeric vector). Default is right.
force	Force of repulsion between overlapping text labels. Defaults to 1. If 0, labels are not repelled at all.
max.overlaps	Exclude text labels that overlap too many things. Defaults to Inf, which means no labels are excluded.
aspect	Logical. TRUE (default) ensures that one unit on the x-axis is the same length as one unit on the y-axis.

Value

a ggplot2 object

Note

If col argument is NULL, shapes or labels are colored according to the variables, using the default ggplot2 palette. The palette can be customized using any scale_color_* function, such as scale_color_brewer(), scale_color_grey() or scale_color_manual().

If resmca is of type stMCA or multiMCA and points is not equal to "all", test-values are used instead of contributions (which are not available for these MCA variants) to select the most important categories ; if points is equal to best, only categories with high test-values for horizontal axis or vertical axis are plotted.

Author(s)

Anton Perdoncin, Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

ggcloud_indiv

Examples

```
# specific MCA of Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# cloud of variables
ggcloud_variables(mca)
# cloud of variables with only categories contributing the most
ggcloud_variables(mca, points = "best", prop = "n")
# cloud of variables with other plotting options
ggcloud_variables(mca, shapes = FALSE, legend = "none",
col = "black", face = "ui")
```

ggeta2_variables *eta-squared plot*

Description

Plots the eta-squared (squared correlation ratios) of the active variables of a MCA.

Usage

```
ggeta2_variables(resmca, axes = c(1,2))
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).

Details

This plot was proposed by Escofier and Pagès (2008) under the name "carré des liaisons", i.e. square of relationships, using correlation ratios to measure these relationships. Eta-squared (i.e. correlation ratio) is a measure of global association between a continuous variable and a categorical variable : it measures the share of variance of the continuous variables "explained" by the categorical variable. Here, it is used to plot the association between the active variables and the axes of the MCA cloud.

Value

a ggplot2 object

Author(s)

Nicolas Robette

References

Escofier B. and Pagès J., 2008, *Analyses factorielles simples et multiples*, Dunod.

See Also

ggcloud_variables, ggadd_attractions

Examples

```
data(Music)
junk <- c("FrenchPop.NA","Rap.NA","Jazz.NA","Classical.NA","Rock.NA")
mca <- speMCA(Music[,1:5], excl = junk)
ggeta2_variables(mca)
```

ggsmoothed_supvar *Plots the density a supplementary variable*

Description

Plots the density of a supplementary variable in a MCA space, using a grid, smoothing and interpolation (via inverse distance weighting.)

Usage

```
ggsmoothed_supvar(resmca, var, cat, axes = c(1,2),
                  center = FALSE, scale = FALSE,
                  nc = c(20, 20), power = 2,
                  limits = NULL, pal = "RdBu")
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
var	factor or numeric vector. The supplementary variable to be plotted.
cat	character string. If var is a factor, the name of the level of the supplementary variable to be plotted.
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).
center	logical. Whether the supplementary variable should be centered or not. Default is FALSE.
scale	logical. Whether the supplementary variable should be scaled to unit variance or not. Default is FALSE.
nc	integer vector of length 2. Number of grid cells in x and y direction (columns, rows).
power	numerical value. The power to use in weight calculation for inverse distance weighting. Default is 2.
limits	numerical vector of length 2. Lower and upper limit of the scale for the supplementary variable.
pal	character string. Name of a (preferably diverging) palette from the RColorBrewer package. Default is "RdBu".

Details

The construction of the plot takes place in several steps. First, the two-dimensional MCA space is cut into a grid of hexagonal cells. Then, for each cell, the average value of the supplementary variable is calculated for the observations located in that cell (if the variable is numerical), or the proportion of observations belonging to the category studied (if the variable is categorical). The results are interpolated and smoothed to make the plot easier to read, using the *inverse distance weighting* technique, which is very common in spatial analysis.

The supplementary variable can be centered beforehand, to represent deviations from the mean (for a numerical variable) or from the mean proportion (for a categorical variable). It can also be scaled to measure deviations in numbers of standard deviations, which can be useful for comparing the results of several supplementary variables.

Value

a ggplot2 object

Author(s)

Nicolas Robette

References

Shepard, Donald (1968). "A two-dimensional interpolation function for irregularly-spaced data". *Proceedings of the 1968 ACM National Conference*. pp. 517–524. doi:10.1145/800186.810616

See Also

ggadd_supvar, ggadd_supvars, ggadd_kellipses, ggadd_ellipses, ggadd_interaction, ggadd_corr, ggadd_chulls, ggadd_density

Examples

```
# specific MCA of Taste example data set
data(Taste)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA",
          "Comedy.NA", "Crime.NA", "Animation.NA", "SciFi.NA", "Love.NA",
          "Musical.NA")
mca <- speMCA(Taste[,1:11], excl = junk)
# density plot for Educ = "High"
ggsmoothed_supvar(mca, Taste$Educ, "High")
# centered and scaled density plot for Age
ggsmoothed_supvar(mca, as.numeric(Taste$Age), center = TRUE, scale = TRUE)
```

---

gPCA *Generalized Principal Component Analysis*

Description

Generalized Principal Component Analysis

Usage

```
gPCA(X, row.w = NULL, col.w = NULL, center = FALSE, scale = FALSE, tol = 1e-07)
```

Arguments

X	data frame of active variables
row.w	numeric vector of row weights. If NULL (default), a vector of 1 for uniform row weights is used.
col.w	numeric vector of column weights. If NULL (default), a vector of 1 for uniform column weights is used.
center	logical. If TRUE, variables are centered (default is FALSE).
scale	logical. If TRUE, variables are scaled to unit variance (default is FALSE).
tol	a tolerance threshold for null eigenvalues (a value less than tol times the first one is considered as null)

Details

Generalized PCA is basically a PCA with the possibility to specify row weights (i.e. "masses") and variable weights (i.e. the "metric"), and to choose whether to center and scale the variables. This flexibility makes it the building block of many variants of PCA, such as Correspondence Analysis and Multiple Correspondence Analysis.

Generalized PCA is also known as "biweighted PCA", "duality diagram" or "generalized singular value decomposition".

Value

An object of class PCA from FactoMineR package

Author(s)

Nicolas Robette

References

Bry X., 1995, *Analyses factorielles simples*, Economica.

Escofier B. and Pagès J., *Analyses factorielles simples et multiples*, Dunod (2008).

Escoufier, Y. (1987) The duality diagram : a means of better practical applications In *Development in numerical ecology*, Legendre, P. & Legendre, L. (Eds.) NATO advanced Institute, Serie G. Springer Verlag, Berlin, 139–156.

Examples

```
library(FactoMineR)
data(decathlon)
res <- gPCA(decathlon[,1:10], center = TRUE, scale = TRUE)
plot(res, choix = "var")
```

homog.test *Homogeneity test for a categorical supplementary variable*

Description

From MCA results, computes a homogeneity test between categories of a supplementary variable, i.e. characterizes the homogeneity of several subclouds.

Usage

```
homog.test(resmca, var, dim = c(1,2))
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
var	the categorical supplementary variable. It does not need to have been used at the MCA step.
dim	the axes which are described. Default is c(1,2)

Value

Returns a list of lists, one for each selected dimension in the MCA. Each list has 2 elements :

test.stat	The square matrix of test statistics
p.values	The square matrix of p-values

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

supvar, supvars, dimtypicality

Examples

```
# specific MCA of Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# homogeneity test for variable Age
homog.test(mca, Music$Age)
```

---

ijunk *App for junk categories of specific MCA*

Description

This function launches a shiny app to define interactively the junk categories before a specific MCA.

Usage

```
ijunk(data, init_junk = NULL)
```

Arguments

data	data frame of categorical variables to be used as active in a specific MCA
init_junk	optional vector of junk categories. Can be a numeric vector indicating the indexes of the junk categories or a character vector of junk categories, specified in the form "namevariable.namecategory" (for instance "gender.male"). Default is NULL.

Details

Once the selection of junk categories is interactively done, the function provides the code to use in a script. It also offer the opportunity to select a set of junk categories at once by writing the common suffix of these categories.

Value

A character vector of junk categories

Author(s)

Nicolas Robette

See Also

speMCA, csMCA, getindexcat

Examples

```
## Not run:
data(Music)
ijunk(Music[,1:5])
# or
junk <- ijunk(Music[,1:5])
# To update an existing vector of junk categories
junk <- ijunk(Music[,1:5], init_junk = c("Rock.NA", "Rap.NA"))
# and then
mca <- speMCA(Music[,1:5], excl = junk)

## End(Not run)
```

MCAiv *Multiple Correspondence Analysis with Instrumental Variables*

Description

Multiple Correspondence Analysis with Instrumental Variables

Usage

```
MCAiv(Y, X, excl = NULL, row.w = NULL, ncp = 5)
```

Arguments

Y	data frame with only factors
X	data frame of instrumental variables, which can be numeric or factors. It must have the same number of rows as Y.
excl	numeric vector indicating the indexes of the "junk" categories (default is NULL). See getindexcat or use ijunk interactive function to identify these indexes. It may also be a character vector of junk categories, specified in the form "namevariable.namecategory" (for instance "gender.male").
row.w	Numeric vector of row weights. If NULL (default), a vector of 1 for uniform row weights is used.
ncp	number of dimensions kept in the results (by default 5)

Details

Multiple Correspondence Analysis with Instrumental Variables consists in three steps : 1. Specific MCA of Y, keeping all the dimensions of the space 2. Computation of one linear regression for each dimension in the specific MCA, with individual coordinates as response and all variables in X as explanatory variables. 3. Principal Component Analysis of the set of predicted values from the regressions in 2.

Multiple Correspondence Analysis with Instrumental Variables is also known as "Canonical Correspondence Analysis" or "Constrained Correspondence Analysis".

Value

An object of class PCA from FactoMineR package, with Y and X as supplementary variables, and an additional item :

ratio	the share of inertia explained by the instrumental variables

.

Note

If there are NAs in Y, these NAs will be automatically considered as junk categories. If one desires more flexibility, Y should be recoded to add explicit factor levels for NAs and then excl option may be used to select the junk categories.

Author(s)

Nicolas Robette

References

Bry X., 1996, *Analyses factorielles multiples*, Economica.

Lebart L., Morineau A. et Warwick K., 1984, *Multivariate Descriptive Statistical Analysis*, John Wiley and sons, New-York.)

See Also

bcMCA, DAQ, bcPCA, DA, PCAiv

Examples

```
library(FactoMineR)
data(tea)
# MCAIV of tea data
# with age, sex, SPC and Sport as instrumental variables
mcaiv <- MCAiv(tea[,1:18], tea[,19:22])
mcaiv$ratio
plot(mcaiv, choix = "ind", invisible = "ind", col.quali = "black")
```

MCAoiv *Multiple Correspondence Analysis with Orthogonal Instrumental Variables*

Description

Multiple Correspondence Analysis with Orthogonal Instrumental Variables

Usage

```
MCAoiv(X, Z, excl = NULL, row.w = NULL, ncp = 5)
```

Arguments

X	data frame with only factors
Z	data frame of instrumental variables, which can be numeric or factors. It must have the same number of rows as X.
excl	numeric vector indicating the indexes of the "junk" categories (default is NULL). See getindexcat or use ijunk interactive function to identify these indexes. It may also be a character vector of junk categories, specified in the form "namevariable.namecategory" (for instance "gender.male").
row.w	Numeric vector of row weights. If NULL (default), a vector of 1 for uniform row weights is used.
ncp	number of dimensions kept in the results (by default 5)

Details

Multiple Correspondence Analysis with Orthogonal Instrumental Variables consists in three steps : 1. Specific MCA of Y, keeping all the dimensions of the space 2. Computation of one linear regression for each dimension in the specific MCA, with individual coordinates as response and all variables in X as explanatory variables. 3. Principal Component Analysis of the set of residuals from the regressions in 2.

Value

An object of class PCA from FactoMineR package, with X as supplementary variables, and an additional item :

ratio	the share of inertia not explained by the instrumental variables

.

Note

If there are NAs in Y, these NAs will be automatically considered as junk categories. If one desires more flexibility, Y should be recoded to add explicit factor levels for NAs and then excl option may be used to select the junk categories.

Author(s)

Nicolas Robette

References

Bry X., 1996, *Analyses factorielles multiples*, Economica.

Lebart L., Morineau A. et Warwick K., 1984, *Multivariate Descriptive Statistical Analysis*, John Wiley and sons, New-York.)

See Also

wcMCA, wcPCA, PCAoiv

Examples

```
library(FactoMineR)
data(tea)
mcaoiv <- MCAoiv(tea[,1:18], tea[,19:22])
mcaoiv$ratio
plot(mcaoiv, choix = "ind", invisible = "ind", col.quali = "black")
```

`medoids` *Medoids of clusters*

Description

Computes the medoids of a cluster solution.

Usage

```
medoids(D, cl)
```

Arguments

`D`	square distance matrix (n rows * n columns, i.e. n individuals) or `dist` object
`cl`	vector with the clustering solution (its length should be n)

Details

A medoid is a representative object of a cluster whose average dissimilarity to all the objects in the cluster is minimal. Medoids are always members of the data set (contrary to means or centroids).

Value

Returns a numeric vector with the indexes of medoids.

Author(s)

Nicolas Robette

References

Kaufman, L. and Rousseeuw, P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.

Anja Struyf, Mia Hubert & Peter J. Rousseeuw (1996). "Clustering in an Object-Oriented Environment". *Journal of Statistical Software*.

See Also

`dist`, `hclust`, `cutree` , `pam`

Examples

```
# hierarchical clustering of the Music example data set,
# partition into 3 groups
# and then computation of the medoids.
data(Music)
temp <- dichotom(Music[,1:5])
d <- dist(temp)
clus <- cutree(hclust(d), 3)
medoids(d, clus)
```

`modif.rate` *Benzecri's modified rates of variance*

Description

Computes Benzecri's modified rates of variance of a multiple correspondence analysis.

Usage

```
modif.rate(resmca)
```

Arguments

`resmca`	object of class `MCA`, `speMCA`, `csMCA`, `stMCA` or `multiMCA`

Details

As MCA clouds often have a high dimensionality, the variance rates of the first principle axes may be quite low, which makes them hard to interpret. Benzecri (1992, p.412) proposed to use *modified rates* to better appreciate the relative importance of the principal axes.

Value

Returns a list of two data frames. The first one is called `raw` and has 3 variables:

`eigen`	eigen values
`rate`	rates
`cum.rate`	cumulative rates

The second one is called `modif` and has 2 variables:

`mrate`	modified rates
`cum.mrate`	cumulative modified rates

Author(s)

Nicolas Robette

References

Benzecri J.P., *Correspondence analysis handbook*, New-York: Dekker (1992).

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

MCA, speMCA, csMCA

Examples

```
# MCA of Music' example data set
data(Music)
mca <- speMCA(Music[,1:5])
# modified rates of variance
modif.rate(mca)
```

`multiMCA` *Multiple Factor Analysis*

Description

Performs Multiple Factor Analysis, drawing on the work of Escofier and Pages (1994). It allows the use of MCA variants (e.g. specific MCA or class specific MCA) as inputs.

Usage

```
multiMCA(l_mca, ncp = 5, compute.rv = FALSE)
```

Arguments

`l_mca`	a list of objects of class `MCA`, `speMCA` or `csMCA`
`ncp`	number of dimensions kept in the results (default is 5)
`compute.rv`	whether RV coefficients should be computed or not (default is FALSE, which makes the function execute faster)

Details

This function binds individual coordinates from every MCA in `l_mca` argument, weights them by the first eigenvalue, and the resulting data frame is used as input for Principal Component Analysis (PCA).

Value

Returns an object of class `multiMCA`, i.e. a list:

`eig`	a list of numeric vector for eigenvalues, percentage of variance and cumulative percentage of variance
`var`	a list of matrices with results for input MCAs components (coordinates, correlations between variables and axes, squared cosines, contributions)
`ind`	a list of matrices with results for individuals (coordinates, squared cosines, contributions)
`call`	a list with informations about input data
`VAR`	a list of matrices with results for categories and variables in the input MCAs (coordinates, squared cosines, test-values, variances)
`my.mca`	lists the content of the objects in `l_mca` argument
`RV`	a matrix of RV coefficients

Author(s)

Nicolas Robette

References

Escofier, B. and Pages, J. (1994) "Multiple Factor Analysis (AFMULT package)". *Computational Statistics and Data Analysis*, 18, 121-140.

See Also

`plot.multiMCA`, `speMCA`, `csMCA`

Examples

```
data(Taste)
# specific MCA on music variables of Taste example data set
mca1 <- speMCA(Taste[,1:5], excl = c(3,6,9,12,15))
# specific MCA on movie variables of Taste example data set
mca2 <- speMCA(Taste[,6:11], excl = c(3,6,9,12,15,18))
# Multiple Factor Analysis of the two sets of variables
mfa <- multiMCA(list(mca1,mca2))
plot.multiMCA(mfa)
```

Music *Music (data)*

Description

The data concerns tastes for music of a set of 500 individuals. It contains 5 variables of likes for music genres (french pop, rap, rock, jazz and classical), 2 variables about music listening and 2 additional variables (gender and age).

Usage

```
data(Music)
```

Format

A data frame with 500 observations and the following 7 variables:

`FrenchPop` factor with levels `No`, `Yes`, `NA`

`Rap` factor with levels `No`, `Yes`, `NA`

`Rock` factor with levels `No`, `Yes`, `NA`

`Jazz` factor with levels `No`, `Yes`, `NA`

`Classical` factor with levels `No`, `Yes`, `NA`

`Gender` factor with levels `Men`, `Women`

`Age` factor with levels `15-24`, `25-49`, `50+`

`OnlyMus` factor with levels `Daily`, `Often`, `Rare`, `Never`, indicating how often one only listens to music.

`Daily` is a factor with levels `No`, `Yes` indicating if one listens to music every day.

Details

`NA` stands for "not available"

Examples

```
data(Music)
str(Music)
```

nsCA *Nonsymmetric Correspondence Analysis*

Description

Nonsymmetric correspondence analysis, for analysing contingency tables with a dependence structure

Usage

```
nsCA(X, ncp = 5, row.sup = NULL,
     col.sup = NULL, quanti.sup = NULL, quali.sup = NULL,
     graph = FALSE, axes = c(1,2), row.w = NULL)
```

Arguments

X	a data frame or a table with n rows and p columns, i.e. a contingency table. Predictor variable should be in rows and response variable in columns.
ncp	number of dimensions kept in the results (by default 5)
row.sup	a vector indicating the indexes of the supplementary rows
col.sup	a vector indicating the indexes of the supplementary columns
quanti.sup	a vector indicating the indexes of the supplementary continuous variables
quali.sup	a vector indicating the indexes of the categorical supplementary variables
graph	boolean, if TRUE a graph is displayed
axes	a length 2 vector specifying the components to plot
row.w	an optional row weights (by default, a vector of 1 and each row has a weight equals to its margin); the weights are given only for the active rows

Details

When dealing with a contingency table with a dependence structure, i.e. when the role of the two variables is not symmetrical but, on the contrary, one can be considered as predicting the other, nonsymmetric correspondence analysis (NSCA) can be used to represent the predictive structure in the table and to assess the predictive power of the predictor variable.

Technically, NSCA is very similar to the standard CA, the main difference being that the columns of the contingency table are not weighted by their rarity (i.e. the inverse of the marginal frequencies).

Value

An object of class CA from FactoMineR package, with an additional item :

GK.tau	Goodman and Kruskal tau

Note

The code is adapted from the CA function in FactoMineR package.

Author(s)

Nicolas Robette

References

Kroonenberg P.M. and Lombardo R., 1999, "Nonsymmetric Correspondence Analysis: A Tool for Analysing Contingency Tables with a Dependence Structure", *Multivariate Behavioral Research*, 34 (3), 367-396.

See Also

nsca.biplot

Examples

```
data(Music)
# The combination of Gender and Age is the predictor variable
# "Focused" listening to music is the response variable
tab <- with(Music, table(interaction(Gender, Age), OnlyMus))
nsca <- nsCA(tab)
nsca.biplot(nsca)
# Goodman and Kruskal tau
nsca$GK.tau
```

---

nsca.biplot *Biplot for Nonsymmetric Correspondence Analysis*

Description

Biplot for Nonsymmetric correspondence analysis, for analysing contingency tables with a dependence structure

Usage

```
nsca.biplot(nsca, axes = c(1,2))
```

Arguments

nsca	an object of class CA created by nsCA() function
axes	numeric vector of length 2, specifying the components (axes) to plot. Default is c(1,2).

Details

The biplots of an NSCA reflect the dependency structure of the contingency table and thus should not be interpreted as the planes of a standard CA. A first principle is that the graph displays the centred row profiles. A second principle is that the relationships between rows and columns are contained in their inner products : the rows are depicted as vectors, also called biplot axes, and the columns are projected on these vectors. If some columns have projections on the row vector far away from the origin, then the row has a comparatively large increase in predictability, and its profile deviates considerably from the marginal one, especially for that column.

For more detailed interpretational guidelines, see Kroonenberg and Lombardo (1999, pp.377-378).

Value

a ggplot2 object

Author(s)

Nicolas Robette

References

Kroonenberg P.M. and Lombardo R., 1999, "Nonsymmetric Correspondence Analysis: A Tool for Analysing Contingency Tables with a Dependence Structure", *Multivariate Behavioral Research*, 34 (3), 367-396.

See Also

nsCA

Examples

```
data(Music)
# The combination of Gender and Age is the predictor variable
# "Focused" listening to music is the response variable
tab <- with(Music, table(interaction(Gender, Age), OnlyMus))
nsca <- nsCA(tab)
nsca.biplot(nsca)
# Goodman and Kruskal tau
nsca$GK.tau
```

PCAiv *Principal Component Analysis with Instrumental Variables*

Description

Principal Component Analysis with Instrumental Variables

Usage

```
PCAiv(Y, X, row.w = NULL, ncp = 5)
```

Arguments

Y	data frame with only numeric variables
X	data frame of instrumental variables, which can be numeric or factors. It must have the same number of rows as Y.
row.w	Numeric vector of row weights. If NULL (default), a vector of 1 for uniform row weights is used.
ncp	number of dimensions kept in the results (by default 5)

Details

Principal Component Analysis with Instrumental Variables consists in two steps : 1. Computation of one linear regression for each variable in Y, with this variable as response and all variables in X as explanatory variables. 2. Principal Component Analysis of the set of predicted values from the regressions in 1 ("Y hat").

Principal Component Analysis with Instrumental Variables is also known as "redundancy analysis"

Value

An object of class PCA from FactoMineR package, with X as supplementary variables, and an additional item :

ratio	the share of inertia explained by the instrumental variables

.

Author(s)

Nicolas Robette

References

Bry X., 1996, *Analyses factorielles multiples*, Economica.

Lebart L., Morineau A. et Warwick K., 1984, *Multivariate Descriptive Statistical Analysis*, John Wiley and sons, New-York.)

See Also

bcPCA, DA, bcMCA, DAQ, MCAiv

Examples

```
library(FactoMineR)
data(decathlon)
# PCAiv of decathlon data set
# with Points and Competition as instrumental variables
pcaiv <- PCAiv(decathlon[,1:10], decathlon[,12:13])
pcaiv$ratio
# plot of \code{Y} variables + quantitative instrumental variables (here Points)
plot(pcaiv, choix = "var")
# plot of qualitative instrumental variables (here Competition)
plot(pcaiv, choix = "ind", invisible = "ind", col.quali = "black")
```

PCAoiv *Principal Component Analysis with Orthogonal Instrumental Variables*

Description

Principal Component Analysis with Orthogonal Instrumental Variables

Usage

```
PCAoiv(X, Z, row.w = NULL, ncp = 5)
```

Arguments

X	data frame with only numeric variables
Z	data frame of instrumental variables to be "partialled out"", which can be numeric or factors. It must have the same number of rows as X.
row.w	Numeric vector of row weights. If NULL (default), a vector of 1 for uniform row weights is used.
ncp	number of dimensions kept in the results (by default 5)

Details

Principal Component Analysis with Orthogonal Instrumental Variables consists in two steps : 1. Computation of one linear regression for each variable in X, with this variable as response and all variables in Z as explanatory variables. 2. Principal Component Analysis of the set of residuals from the regressions in 1.

Value

An object of class PCA from FactoMineR package, and an additional item :

ratio	the share of inertia not explained by the instrumental variables

.

Author(s)

Nicolas Robette

References

Bry X., 1996, *Analyses factorielles multiples*, Economica.

Lebart L., Morineau A. et Warwick K., 1984, *Multivariate Descriptive Statistical Analysis*, John Wiley and sons, New-York.)

See Also

wcPCA, wcMCA, MCAoiv

Examples

```
library(FactoMineR)
data(decathlon)
pcaoiv <- PCAoiv(decathlon[,1:10], decathlon[,12:13])
plot(pcaoiv, choix = "var", invisible = "quanti.sup")
```

---

planecontrib *Contributions to a plane*

Description

For a given plane of a MCA, computes contributions and squared cosines of the active variables and categories and of the active individuals.

Usage

```
planecontrib(resmca, axes = c(1,2))
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA or bcMCA function
axes	numeric vector of length 2, specifying the axes forming the plane to describe. Default is c(1,2).

Value

A list of two lists. The first deals with variables :

ctr	vector of contributions of the active categories to the plane
cos2	vector of squared cosines of the active categories in the plane
vctr	vector of contributions of the active variables to the plane

The second deals with observations :

ctr	vector of contributions of the observations to the plane
cos2	vector of squared cosines of the observations in the plane

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

dimcontrib, tabcontrib

Examples

```
data(Music)
junk <- c("FrenchPop.NA","Rap.NA","Jazz.NA","Classical.NA","Rock.NA")
mca <- speMCA(Music[,1:5], excl = junk)
co <- planecontrib(mca)
co$var
```

---

plot.csMCA *Plot of class specific MCA*

Description

Plots a class specific Multiple Correspondence Analysis (resulting from csMCA function), i.e. the clouds of individuals or categories.

Usage

```
## S3 method for class 'csMCA'
plot(x, type = "v", axes = 1:2, points = "all",
col = "dodgerblue4", app = 0, ...)
```

Arguments

x	object of class csMCA
type	character string: 'v' to plot the categories (default), 'i' to plot individuals' points, 'inames' to plot individuals' names
axes	numeric vector of length 2, specifying the components (axes) to plot (c(1,2) is default)
points	character string. If 'all' all points are plotted (default); if 'besth' only those who contribute most to horizontal axis are plotted; if 'bestv' only those who contribute most to vertical axis are plotted; if 'besthv' only those who contribute most to horizontal or vertical axis are plotted.

col	color for the points of the individuals or for the labels of the categories (default is 'dodgerblue4')
app	numerical value. If 0 (default), only the labels of the categories are plotted and their size is constant; if 1, only the labels are plotted and their size is proportional to the weights of the categories; if 2, points (triangles) and labels are plotted, and points size is proportional to the weight of the categories.
...	further arguments passed to or from other methods, such as cex, cex.main, ...

Details

A category is considered to be one of the most contributing to a given axis if its contribution is higher than the average contribution, i.e. 100 divided by the total number of categories.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

csMCA, textvarsup, conc.ellipse

Examples

```
# class specific MCA on Music example data set
# ignoring every NA values categories
# and focusing on the subset of women,
data(Music)
female <- Music$Gender=="Women"
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- csMCA(Music[,1:5], subcloud = female, excl = junk)
# cloud of categories
plot(mca)
# cloud of most contributing categories
plot(mca,axes=c(2,3), points = "besthv", col = "darkred", app = 1)
```

plot.multiMCA *Plot of Multiple Factor Analysis*

Description

Plots Multiple Factor Analysis data, resulting from multiMCA function.

Usage

```
## S3 method for class 'multiMCA'
plot(x, type = "v", axes = c(1, 2), points = "all", threshold = 2.58,
groups = 1:x$call$ngroups, col = rainbow(x$call$ngroups), app = 0, ...)
```

Arguments

x	object of class multiMCA
type	character string: 'v' to plot the categories (default), 'i' to plot individuals' points, 'inames' to plot individuals' names
axes	numeric vector of length 2, specifying the components (axes) to plot (c(1,2) is default)
points	character string. If 'all' all points are plotted (default); if 'besth' only those who are the most correlated to horizontal axis are plotted; if 'bestv' only those who are the most correlated to vertical axis are plotted; if 'best' only those who are the most coorelated to horizontal or vertical axis are plotted.
threshold	numeric value. V-test minimal value for the selection of plotted categories.
groups	numeric vector specifying the groups of categories to plot. By default, every groups of categories will be plotted
col	a color for the points of the individuals or a vector of colors for the labels of the groups of categories (by default, rainbow palette is used)
app	numerical value. If 0 (default), only the labels of the categories are plotted and their size is constant; if 1, only the labels are plotted and their size is proportional to the weights of the categories; if 2, points (triangles) and labels are plotted, and points size is proportional to the weight of the categories.
...	further arguments passed to or from other methods, such as cex, cex.main, ...

Details

A category is considered to be one of the most correlated to a given axis if its test-value is higher then 2.58 (which corresponds to a 0.05 threshold).

Author(s)

Nicolas Robette

References

Escofier, B. and Pages, J. (1994) "Multiple Factor Analysis (AFMULT package)". *Computational Statistics and Data Analysis*, 18, 121-140.

See Also

multiMCA, textvarsup, speMCA, csMCA

Examples

```
# specific MCA on music variables of Taste example data set
## another one on movie variables of 'Taste' example data set,
## and then a Multiple Factor Analysis and plots the results.
data(Taste)
# specific MCA on music variables of Taste example data set
mca1 <- speMCA(Taste[,1:5], excl = c(3,6,9,12,15))
# specific MCA on movie variables of Taste example data set
mca2 <- speMCA(Taste[,6:11], excl = c(3,6,9,12,15,18))
# Multiple Factor Analysis
mfa <- multiMCA(list(mca1,mca2))
# plot
plot.multiMCA(mfa, col = c("darkred", "darkblue"))
# plot of the second set of variables (movie)
plot.multiMCA(mfa, groups = 2, app = 1)
```

---

plot.speMCA *Plot of specific MCA*

Description

Plots a specific Multiple Correspondence Analysis (resulting from speMCA function), i.e. the clouds of individuals or categories.

Usage

```
## S3 method for class 'speMCA'
plot(x, type = "v", axes = c(1,2), points = "all", col = "dodgerblue4", app = 0, ...)
```

Arguments

x	object of class speMCA
type	character string: 'v' to plot the categories (default), 'i' to plot individuals' points, 'inames' to plot individuals' names
axes	numeric vector of length 2, specifying the components (axes) to plot (c(1,2) is default)

points	character string. If 'all' all points are plotted (default); if 'besth' only those who contribute most to horizontal axis are plotted; if 'bestv' only those who contribute most to vertical axis are plotted; if 'besthv' only those who contribute most to horizontal or vertical axis are plotted; if 'best' only those who contribute most to the plane are plotted.
col	color for the points of the individuals or for the labels of the categories (default is 'dodgerblue4')
app	numerical value. If 0 (default), only the labels of the categories are plotted and their size is constant; if 1, only the labels are plotted and their size is proportional to the weights of the categories; if 2, points (triangles) and labels are plotted, and points size is proportional to the weight of the categories.
...	further arguments passed to or from other methods, such as cex, cex.main, ...

Details

A category is considered to be one of the most contributing to a given axis if its contribution is higher than the average contribution, i.e. 100 divided by the total number of categories.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

speMCA, textvarsup, conc.ellipse

Examples

```
# specific MCA of Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# cloud of categories
plot(mca)
```

plot.stMCA *Plot of standardized MCA*

Description

Plots a standardized Multiple Correspondence Analysis (resulting from stMCA function), i.e. the clouds of individuals or categories.

Usage

```
## S3 method for class 'stMCA'
plot(x, type = "v", axes = 1:2, points = "all", threshold = 2.58, groups=NULL,
                            col = "dodgerblue4", app = 0, ...)
```

Arguments

x	object of class stMCA
type	character string: 'v' to plot the categories (default), 'i' to plot individuals' points, 'inames' to plot individuals' names
axes	numeric vector of length 2, specifying the components (axes) to plot (c(1,2) is default)
points	character string. If 'all' all points are plotted (default); if 'besth' only those who are the most correlated to horizontal axis are plotted; if 'bestv' only those who are the most correlated to vertical axis are plotted; if 'best' only those who are the most coorelated to horizontal or vertical axis are plotted.
threshold	numeric value. V-test minimal value for the selection of plotted categories.
groups	only if x$call$input.mca = 'multiMCA', i.e. if the MCA standardized to x object was a multiMCA object. Numeric vector specifying the groups of categories to plot. By default, every groups of categories will be plotted
col	color for the points of the individuals or for the labels of the categories (default is 'dodgerblue4')
app	numerical value. If 0 (default), only the labels of the categories are plotted and their size is constant; if 1, only the labels are plotted and their size is proportional to the weights of the categories; if 2, points (triangles) and labels are plotted, and points size is proportional to the weight of the categories.
...	further arguments passed to or from other methods, such as cex, cex.main, ...

Details

A category is considered to be one of the most correlated to a given axis if its test-value is higher then 2.58 (which corresponds to a 0.05 threshold).

Author(s)

Nicolas Robette

References

Bry X., Robette N., Roueff O., 2016, « A dialogue of the deaf in the statistical theater? Adressing structural effects within a geometric data analysis framework », *Quality & Quantity*, 50(3), pp 1009–1020 [https://link.springer.com/article/10.1007/s11135-015-0187-z]

See Also

stMCA, textvarsup, conc.ellipse

Examples

```
# standardized MCA of Music example data set
# controlling for age
## and then draws the cloud of categories.
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
stmca <- stMCA(mca, control = list(Music$Age))
# cloud of categories
plot(stmca)
# cloud of categories on dimensions 2 and 3
plot(stmca, axes = c(2,3), points = "best", col = "darkred", app = 1)
```

---

quadrant *Quadrant of active individuals*

Description

Computes the quadrant of active individuals from a MCA.

Usage

```
quadrant(resmca, dim = c(1,2))
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
dim	dimensions of the space (default is c(1,2))

Value

Returns a factor with four levels : upper_left, lower_left, upper_right, lower_right

Author(s)

Nicolas Robette

See Also

speMCA, csMCA

Examples

```
# specific MCA of Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# distribution of the quadrants
table(quadrant(mca, c(1,2)))
```

---

quasindep *Quasi-correspondence analysis*

Description

Transforms a symmetrical contingency table so that it can be used for quasi-correspondence analysis, also called correspondence analysis of incomplete contingency table.

Usage

```
quasindep(tab, order = 3, tol = 1e-6)
```

Arguments

tab	a symmetric table or matrix
order	numeric value. Order of reconstitution of the quasi-independence data. Default is 3.
tol	numeric value. The tolerance threshold to be considered for convergence to null during iteration process. Default is 1e-6.

Details

In order to carry out a "quasi-correspondence analysis", also called "correspondence analysis of incomplete table", the principle is to stop analyzing the differences between the observed data and the situation of independence between the variable in rows and the variable in columns, as it is the case in the classical correspondence analysis, and to consider the differences between the data and a situation of quasi-independence, i.e. independence for some cells of the table only. In the most common situation, it is therefore a matter of applying the independence hypothesis to the off-diagonal cells only and replacing the diagonal with values that do not influence the analysis. Such values are obtained in an iterative way by replacing the numbers of the cells of the diagonal by their third order reconstruction, then by recalculating the correspondence analysis until convergence is reached. The algorithm used is developed in van der Heijden (1992: 11-12).

Value

An object of the same class and dimensions as `tab` : the quasi-independence data to be analyzed with Correspondence Analysis.

Note

This function is adapted from Milan Bouchet-Valat's script in the supplementary material of his article indicated in the reference section.

References

De Leeuw J et van der Heijden PGM (1985) *Quasi-Correspondence Analysis*. Leiden: University of Leiden.

Van der Heijden PGM (1992) Three Approaches to Study the Departure from Quasi-independence. *Statistica Applicata* 4: 465-80.

Bouchet-Valat M (2015) L'analyse statistique des tables de contingence carrées - L'homogamie socioprofessionnelle en France - I, L'analyse des correspondances *Bulletin de Méthodologie Sociologique* 125: 65–88. <doi:10.1177/0759106314555655>

Examples

```
## Not run:
tab <- matrix(c(165,49,70,100,48,223,
                6,201,226,212,90,216,
                4,96,446,214,72,77,
                5,84,305,317,126,188,
                3,52,151,190,110,189,
                17,234,310,601,309,1222),
                nrow = 6, ncol = 6, byrow = TRUE)
newtab <- quasindep(tab)

## End(Not run)
```

`reshape_between` *Reshapes objects created with bcMCA()*

Description

reshapes objects created with 'bcMCA()' so that they can be used with other functions from the package.

Usage

```
reshape_between(bcmca)
```

Arguments

`bcmca`	object created with `bcMCA` function

Value

Returns an object of class bcMCA

Author(s)

Nicolas Robette

References

Abdi H., 2007, "Discriminant Correspondence Analysis", In: Neil Salkind (Ed.), *Encyclopedia of Measurement and Statistics*, Thousand Oaks (CA): Sage.

Bry X., 1996, *Analyses factorielles multiples*, Economica.

Lebart L., Morineau A. et Warwick K., 1984, *Multivariate Descriptive Statistical Analysis*, John Wiley and sons, New-York.)

See Also

bcMCA

Examples

```
data(tea, package = "FactoMineR")
res <- bcMCA(tea[,1:18], tea$SPC)
res_ok <- reshape_between(res)
ggcloud_variables(res_ok)
```

---

rvcoef *RV coefficient*

Description

Computes the RV coefficient between two groups of numerical variables.

Usage

```
rvcoef(Xa, Xb, row.w = NULL)
```

Arguments

Xa	data frame with the first group of numerical variables
Xb	data frame with the second group of numerical variables
row.w	numeric vector of row weights. If NULL (default), a vector of 1 for uniform row weights is used.

Details

Xa and Xb should have the same number of rows.

Value

numerical value : the RV coefficient

Author(s)

Nicolas Robette

References

Escouffier, Y. (1973) Le traitement des variables vectorielles. *Biometrics* 29 751–760.

See Also

coiPCA, coiMCA, multiMCA

Examples

```
# RV coefficient between decathlon results by sport
# and Rank and Points
library(FactoMineR)
data(decathlon)
Xa <- decathlon[,1:10]
Xb <- decathlon[,11:12]
str(Xa)
str(Xb)
rvcoef(Xa, Xb)
```

---

scaled.dev *Scaled deviations for a categorical supplementary variable*

Description

From MCA results, computes scaled deviations between categories for a categorical supplementary variable.

Usage

```
scaled.dev(resmca, var)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
var	the categorical supplementary variable. It does not need to have been used at the MCA step.

Value

Returns a list with one matrix for each dimension of the MCA. Each matrix is filled with scaled deviations between the categories of the supplementary variable, for a given dimension.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

supvar, supvars, ggadd_supvar, ggadd_supvars, textvarsup, supind

Examples

```
# specific MCA of Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# computes scaled deviations for Age supplementary variable
scaled.dev(mca,Music$Age)
```

---

speMCA *specific MCA*

Description

Performs a specific Multiple Correspondence Analysis, i.e. a variant of MCA that allows to treat undesirable categories as passive categories.

Usage

```
speMCA(data, excl = NULL, ncp = 5, row.w = NULL)
```

Arguments

data	data frame with n rows (individuals) and p columns (categorical variables)
excl	numeric vector indicating the indexes of the "junk" categories (default is NULL). See getindexcat or use ijunk interactive function to identify these indexes. It may also be a character vector of junk categories, specified in the form "namevariable.namecategory" (for instance "gender.male").

ncp	number of dimensions kept in the results (default is 5)
row.w	an optional numeric vector of row weights. If NULL (default), a vector of 1 for uniform row weights)

Details

Undesirable (i.e. "junk") categories may be of several kinds: infrequent categories (say, <5 percents), heterogeneous categories (e.g. "others") or uninterpretable categories (e.g. "not available"). In these cases, specific MCA may be useful to ignore these categories for the determination of distances between individuals (see references).

If there are NAs in data, these NAs will be automatically considered as junk categories. If one desires more flexibility, data should be recoded to add explicit factor levels for NAs and then excl option may be used to select the junk categories.

Value

Returns an object of class speMCA, i.e. a list including:

eig	a list of vectors containing all the eigenvalues, the percentage of variance, the cumulative percentage of variance, the modified rates and the cumulative modified rates
call	a list with informations about input data
ind	a list of matrices containing the results for the individuals (coordinates, contributions, squared cosines and total distances)
var	a list of matrices containing all the results for the categories and variables (weights, coordinates, squared cosines, categories contributions to axes and cloud, test values (v.test), squared correlation ratio (eta2), variable contributions to axes and cloud, total distances

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

getindexcat, ijunk, plot.speMCA, ggcloud_indiv, ggcloud_variables, csMCA

Examples

```
# specific MCA of Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# This is equivalent to :
mca <- speMCA(Music[,1:5], excl = c(3,6,9,12,15))
```

---

stMCA *Standardized MCA*

Description

Performs a standardized Multiple Correspondence Analysis, i.e it takes MCA results and forces all the dimensions to be orthogonal to a supplementary "control" variable.

Usage

```
stMCA(resmca, control)
```

Arguments

resmca	an object of class MCA, speMCA, csMCA or multiMCA
control	a list of control variables

Details

Standardized MCA unfolds in several steps. 1. First, for each dimension of an input MCA, individual coordinates are used as dependent variable in a linear regression model and the 'control' variable is included as covariate in the same model. 2. The residuals from every models are retained and bound together. The resulting data frame is composed of continuous variables and its number of columns is equal to the number of dimensions in the input MCA. 3. Lastly, this data frame is used as input in a Principal Component Analysis.

It is exactly equivalent to MCA with one orthogonal instrumental variable (see MCAoiv)

Value

Returns an object of class stMCA. This object will be similar to resmca argument, still it does not comprehend modified rates, categories contributions and variables contributions.

Author(s)

Nicolas Robette

References

Bry X., Robette N., Roueff O., 2016, « A dialogue of the deaf in the statistical theater? Adressing structural effects within a geometric data analysis framework », *Quality & Quantity*, 50(3), pp 1009–1020 [https://link.springer.com/article/10.1007/s11135-015-0187-z]

See Also

plot.stMCA

Examples

```
# standardized MCA of Music example data set
# controlling for age
## and then draws the cloud of categories.
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
stmca <- stMCA(mca, control = list(Music$Age))
```

---

supind *Statistics for supplementary individuals*

Description

From MCA results, computes statistics (coordinates, squared cosines) for supplementary individuals.

Usage

```
supind(resmca, supdata)

indsup(resmca, supdata)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA or bcMCA function
supdata	data frame with the supplementary individuals. It must have the same factors as the data frame used as input for the initial MCA.

Value

Returns a list with the following items :

coord	matrix of individuals coordinates
cos2	matrix of individuals squared cosines

Note

indsup is softly deprecated. Please use supind instead.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

ggadd_supind,textindsup, supvar, supvars

Examples

```
# specific MCA of Music example data set
# excluding the first two observations
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[3:nrow(Music),1:5], excl = junk)
# computes coordinates and squared cosines
# of the first two (supplementary) observations
supind(mca,Music[1:2,1:5])
```

---

supvar *Statistics for a categorical supplementary variable*

Description

From MCA results, computes statistics (weights, coordinates, contributions, test-values, variances) for a categorical supplementary variable.

Usage

```
supvar(resmca, var)

varsup(resmca, var)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
var	the categorical supplementary variable. It does not need to have been used at the MCA step.

Value

Returns a list:

weight	numeric vector of categories weights
coord	data frame of categories coordinates
cos2	data frame of categories squared cosines
var	data frame of categories within variances, variance between and within categories and variable squared correlation ratio (eta2)
typic	data frame of categories typicality test statistics
pval	data frame of categories p-values from typicality test statistics
cor	data frame of categories correlation coefficients

Note

varsup is softly deprecated. Please use supvar instead.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

supvars, ggadd_supvar, ggadd_supvars, textvarsup, supind

Examples

```
# specific MCA of Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# computes statistics for Age supplementary variable
supvar(mca,Music$Age)
```

supvars *Statistics for categorical supplementary variables*

Description

From MCA results, computes statistics (weights, coordinates, squared cosines, contributions, test-values, variances) for categorical supplementary variables.

Usage

```
supvars(resmca, vars)

varsups(resmca, vars)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA, bcMCA, stMCA or multiMCA function
vars	A data frame of categorical supplementary variables. All these variables should be factors.

Value

Returns a list with the following items :

weight	numeric vector of categories weights
coord	data frame of categories coordinates
cos2	data frame of categories squared cosines
var	a list of data frames of categories within variances, variance between and within categories and variable square correlation ratio (eta2)
typic	data frame of categories typicality test statistics
pval	data frame of categories p-values from typicality test statistics
cor	data frame of categories correlation coefficients

Note

varsups is softly deprecated. Please use supvars instead.

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

supvar, ggadd_supvar, ggadd_supvars, textvarsup, supind

Examples

```
# specific MCA of Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# computes statistics for Gender and Age supplementary variables
supvars(mca, Music[, c("Gender","Age")])
```

---

tabcontrib *Table with the main contributions of categories to an axis*

Description

Identifies the categories that contribute the most to a given dimension of a Multiple Correspondence Analysis and organizes these informations into a fancy table.

Usage

```
tabcontrib(resmca, dim = 1,
best = TRUE, limit = NULL,
dec = 2, shortlabs = FALSE)
```

Arguments

resmca	object created with MCA, speMCA, csMCA, wcMCA or bcMCA function
dim	dimension to describe (default is 1st dimension)
best	if FALSE, displays all the categories; if TRUE (default), displays only categories which contribute the most (see limit argument below)
limit	numerical value between 0 and 100. If best = TRUE (see above), only categories with percentage of contribution higher or equal to limit are displayed. If best = TRUE and limit = NULL (default), only categories with contributions higher or equal to average are displayed.
dec	integer. The number of decimals for the results (default is 2)
shortlabs	logical. If TRUE, the data frame will have short column names, so that all columns can be displayed side by side on a laptop screen. Default is FALSE (long explicit column names).

Value

A data frame with the following contributions.:

`Variable` names of the variables

`Category` names of the categories

`Weight` weights of the categories

`Quality of representation`
quality of representation (squared cosine) of the categories on the axis

`Contribution (left)`
contributions of the categories located on one side of the axis

`Contribution (right)`
contributions of the categories located on the other side of the axis

`Total contribution`
contributions summed by variable

`Cumulated contribution`
cumulated sum of the contributions

`Contribution of deviation`
for each variable, contribution of the deviation between the barycenter of the categories located on one side of the axis and the barycenter of those located on the other side

`Proportion to variable`
contribution of deviation expressed as a proportion of the contribution of the variable

Author(s)

Nicolas Robette

References

Le Roux B. and Rouanet H., *Multiple Correspondence Analysis*, SAGE, Series: Quantitative Applications in the Social Sciences, Volume 163, CA:Thousand Oaks (2010).

Le Roux B. and Rouanet H., *Geometric Data Analysis: From Correspondence Analysis to Stuctured Data Analysis*, Kluwer Academic Publishers, Dordrecht (June 2004).

See Also

`dimcontrib`, `dimdescr`, `dimeta2`, `dimtypicality`

Examples

```
# specific MCA of Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# main contributions on axis 1
tabcontrib(mca, 1)
# main contributions on axis 2
tabcontrib(mca, 2)
```

Taste *Taste (data)*

Description

The data concerns tastes for music and movies of a set of 2000 individuals. It contains 5 variables of likes for music genres (french pop, rap, rock, jazz and classical), 6 variables of likes for movie genres (comedy, crime, animation, science fiction, love, musical) and 2 additional variables (gender and age).

Usage

```
data(Taste)
```

Format

A data frame with 2000 observations and the following 13 variables:

`FrenchPop` factor with levels `No`, `Yes`, `NA`

`Rap` factor with levels `No`, `Yes`, `NA`

`Rock` factor with levels `No`, `Yes`, `NA`

`Jazz` factor with levels `No`, `Yes`, `NA`

`Classical` factor with levels `No`, `Yes`, `NA`

`Comedy` factor with levels `No`, `Yes`, `NA`

`Crime` factor with levels `No`, `Yes`, `NA`

`Animation` factor with levels `No`, `Yes`, `NA`

`SciFi` factor with levels `No`, `Yes`, `NA`

`Love` factor with levels `No`, `Yes`, `NA`

`Musical` factor with levels `No`, `Yes`, `NA`

`Gender` factor with levels `Men`, `Women`

`Age` factor with levels `15-24`, `25-49`, `50+`

`Educ` factor with levels `none`, `low`, `medium`, `high`

Details

`NA` stands for "not available"

Examples

```
data(Taste)
str(Taste)
```

textindsup *Plot of supplementary individuals*

Description

Adds supplementary individuals to a MCA cloud of the individuals.

Usage

```
textindsup(resmca, supdata, axes = c(1, 2), col = "darkred")
```

Arguments

resmca	object of class MCA, speMCA, or csMCA
supdata	data frame with the supplementary individuals. It must have the same factors as the data frame used as input for the initial MCA.
axes	numeric vector of length 2, specifying the dimensions (axes) to plot (default is c(1,2))
col	color for the labels of the categories (default is "darkred")

Author(s)

Nicolas Robette

See Also

supind, plot.speMCA, plot.csMCA

Examples

```
# specific MCA of Music example data set
# excluding the first two observations
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[3:nrow(Music), 1:5], excl = junk)
# cloud of active individuals
# with the two supplementary individuals
plot(mca, type = "i")
textindsup(mca, Music[1:2, 1:5])
```

textvarsup *Plot of a categorical supplementary variable*

Description

Adds a categorical supplementary variable to a MCA cloud of categories.

Usage

```
textvarsup(resmca, var, sel = 1:nlevels(var), axes = c(1, 2),
           col = "black", app = 0, vname = NULL)
```

Arguments

resmca	object of class MCA, speMCA, csMCA, stMCA or multiMCA
var	the categorical supplementary variable. It does not need to have been used at the MCA step.
sel	numeric vector of indexes of the categories of the supplementary variable to be added to the plot (by default, labels are plotted for every categories)
axes	numeric vector of length 2, specifying the dimensions (axes) to plot (default is c(1,2))
col	color for the labels of the categories (default is black)
app	numerical value. If 0 (default), only the labels are plotted and their size is constant; if 1, only the labels are plotted and their size is proportional to the weights of the categories; if 2, points (triangles) and labels are plotted, and points size is proportional to the weight of the categories.
vname	a character string to be used as a prefix for the labels of the categories (null by default)

Author(s)

Nicolas Robette

See Also

supvar, supvars, plot.speMCA, plot.csMCA

Examples

```
# specific MCA of Music example data set
data(Music)
junk <- c("FrenchPop.NA", "Rap.NA", "Rock.NA", "Jazz.NA", "Classical.NA")
mca <- speMCA(Music[,1:5], excl = junk)
# cloud of categories
# with Gender and Age supplementary variables
plot(mca, col = "gray")
```

```
textvarsup(mca, Music$Gender,col = "darkred")
textvarsup(mca, Music$Age, sel = c(1,3), col = "orange",
           vname = "age", app = 1)
```

`translate.logit` *Deprecated function*

Description

This function has been moved to the `translate.logit` package.

Usage

```
translate.logit(...)
```

Arguments

`...`	arguments are ignored

`wcMCA` *Within-class MCA*

Description

Within-class MCA, also called conditional MCA

Usage

```
wcMCA(data, class, excl = NULL, row.w = NULL, ncp = 5)
```

Arguments

`data`	data frame with only categorical variables, i.e. factors
`class`	factor specifying the class
`excl`	numeric vector indicating the indexes of the "junk" categories (default is NULL). See `getindexcat` or use `ijunk` interactive function to identify these indexes. It may also be a character vector of junk categories, specified in the form "namevariable.namecategory" (for instance "gender.male").
`row.w`	numeric vector of row weights. If NULL (default), a vector of 1 for uniform row weights is used.
`ncp`	number of dimensions kept in the results (by default 5)

Details

Within-class Multiple Correspondence Analysis is a MCA where the active categories are centered on the mean of their class (i.e. conditional frequencies) instead of the overall mean (i.e. marginal frequencies).

It is also known as "conditional MCA" and can be seen as a special case of MCA on orthogonal instrumental variables, with only one (categorical) instrumental variable.

Value

An object of class `speMCA`, with an additional item :

`ratio`	the within-class inertia percentage

.

Note

The code is adapted from `speMCA` function.

As in `speMCA`, if there are NAs in `data`, these NAs will be automatically considered as junk categories. If one desires more flexibility, `data` should be recoded to add explicit factor levels for NAs and then `excl` option may be used to select the junk categories.

Author(s)

Nicolas Robette

References

Escofier B., 1990, Analyse des correspondances multiples conditionnelle, *La revue de Modulad*, 5, 13-28.

Lebart L., Morineau A. et Warwick K., 1984, *Multivariate Descriptive Statistical Analysis*, John Wiley and sons, New-York.)

See Also

`MCAoiv`, `wcPCA`, `PCAoiv`

Examples

```
# within-class analysis of tea data
# with SPC as class
library(FactoMineR)
data(tea)
res <- wcMCA(tea[,1:18], tea$SPC)
res$ratio
ggcloud_variables(res)
```

wcPCA *Within-class Principal Component Analysis*

Description

Within-class Principal Component Analysis

Usage

```
wcPCA(X, class, scale.unit = TRUE, ncp = 5, ind.sup = NULL, quanti.sup = NULL,
          quali.sup = NULL, row.w = NULL, col.w = NULL, graph = FALSE,
          axes = c(1, 2))
```

Arguments

X	a data frame with *n* rows (individuals) and *p* columns (numeric variables)
class	factor specifying the class
scale.unit	a boolean, if TRUE (default) then data are scaled to unit variance
ncp	number of dimensions kept in the results (by default 5)
ind.sup	a vector indicating the indexes of the supplementary individuals
quanti.sup	a vector indicating the indexes of the quantitative supplementary variables
quali.sup	a vector indicating the indexes of the categorical supplementary variables
row.w	an optional row weights (by default, a vector of 1 for uniform row weights); the weights are given only for the active individuals
col.w	an optional column weights (by default, uniform column weights); the weights are given only for the active variables
graph	boolean, if TRUE a graph is displayed. Default is FALSE.
axes	a length 2 vector specifying the components to plot

Details

Within-class Principal Component Analysis is a PCA where the active variables are centered on the mean of their class instead of the overall mean.

It is a "conditional" PCA and can be seen as a special case of PCA with orthogonal instrumental variables, with only one (categorical) instrumental variable.

Value

An object of class PCA from FactoMineR package, with an additional item :

ratio	the within-class inertia percentage

.

Note

The code is adapted from PCA function from FactoMineR package.

Author(s)

Nicolas Robette

References

Escofier B., 1990, Analyse des correspondances multiples conditionnelle, *La revue de Modulad*, 5, 13-28.

Lebart L., Morineau A. et Warwick K., 1984, *Multivariate Descriptive Statistical Analysis*, John Wiley and sons, New-York.)

See Also

PCAoiv, wcMCA, MCAoiv

Examples

```
# within-class analysis of decathlon data
# with quatiles of points as class
library(FactoMineR)
data(decathlon)
points <- cut(decathlon$Points, c(7300, 7800, 8000, 8120, 8900), c("Q1","Q2","Q3","Q4"))
res <- wcPCA(decathlon[,1:10], points)
plot(res, choix = "var")
```

---

wtable *Deprecated functions*

Description

These functions have been moved to the descriptio package. You may check its documentation here : https://nicolas-robette.github.io/descriptio/

Usage

```
wtable(...)

pem(...)

phi.table(...)

assoc.twocont(...)

assoc.twocat(...)
```

```
assoc.catcont(...)

assoc.yx(...)

darma(...)

catdesc(...)

condesc(...)

ggassoc_phiplot(...)

ggassoc_boxplot(...)

ggassoc_scatter(...)

ggassoc_crosstab(...)
```

Arguments

...	arguments are ignored

Index

∗ **aplot**
conc.ellipse, 17
ggadd_attractions, 34
ggadd_chulls, 35
ggadd_corr, 37
ggadd_density, 39
ggadd_ellipses, 40
ggadd_interaction, 42
ggadd_kellipses, 43
ggadd_partial, 45
ggadd_supind, 47
ggadd_supvar, 48
ggadd_supvars, 49
ggaxis_variables, 51
ggbootvalid_supvars, 53
ggbootvalid_variables, 55
ggcloud_indiv, 57
ggcloud_variables, 59
ggeta2_variables, 62
ggsmoothed_supvar, 63
nsca.biplot, 77
plot.csMCA, 82
plot.multiMCA, 84
plot.speMCA, 85
plot.stMCA, 87
textindsup, 103
textvarsup, 104
∗ **datasets**
Music, 75
Taste, 102
∗ **misc**
ahc.plots, 3
angles.csa, 4
barplot_contrib, 5
bootvalid_supvars, 10
bootvalid_variables, 11
break_interaction, 12
burt, 14
conc.ellipse, 17
contrib, 19
dichotom, 25
dichotomixed, 26
dimcontrib, 27
dimdescr, 28
dimeta2, 30
dimtypicality, 31
dist.chi2, 32
flip.mca, 32
getindexcat, 33
homog.test, 66
ijunk, 67
medoids, 71
modif.rate, 72
planecontrib, 81
quadrant, 88
quasindep, 89
reshape_between, 90
rvcoef, 91
scaled.dev, 92
supind, 96
supvar, 97
supvars, 99
tabcontrib, 100
∗ **multivariate**
ahc.plots, 3
angles.csa, 4
barplot_contrib, 5
bcMCA, 7
bcPCA, 8
bootvalid_supvars, 10
bootvalid_variables, 11
break_interaction, 12
burt, 14
coiMCA, 15
coiPCA, 16
conc.ellipse, 17
contrib, 19
csMCA, 20

DA, 22
DAQ, 23
dichotom, 25
dichotomixed, 26
dimcontrib, 27
dimdescr, 28
dimeta2, 30
dimtypicality, 31
dist.chi2, 32
flip.mca, 32
getindexcat, 33
ggadd_attractions, 34
ggadd_chulls, 35
ggadd_corr, 37
ggadd_density, 39
ggadd_ellipses, 40
ggadd_interaction, 42
ggadd_kellipses, 43
ggadd_partial, 45
ggadd_supind, 47
ggadd_supvar, 48
ggadd_supvars, 49
ggaxis_variables, 51
ggbootvalid_supvars, 53
ggbootvalid_variables, 55
ggcloud_indiv, 57
ggcloud_variables, 59
ggeta2_variables, 62
ggsmoothed_supvar, 63
gPCA, 64
homog.test, 66
ijunk, 67
MCAiv, 68
MCAoiv, 69
medoids, 71
modif.rate, 72
multiMCA, 73
nsCA, 76
nsca.biplot, 77
PCAiv, 78
PCAoiv, 80
planecontrib, 81
plot.csMCA, 82
plot.multiMCA, 84
plot.speMCA, 85
plot.stMCA, 87
quadrant, 88
quasindep, 89
reshape_between, 90
rvcoef, 91
scaled.dev, 92
speMCA, 93
stMCA, 95
supind, 96
supvar, 97
supvars, 99
tabcontrib, 100
textindsup, 103
textvarsup, 104
wcMCA, 105
wcPCA, 107

ahc.plots, 3
angles.csa, 4
assoc.catcont *(*wtable*)*, 108
assoc.twocat *(*wtable*)*, 108
assoc.twocont *(*wtable*)*, 108
assoc.yx *(*wtable*)*, 108

barplot_contrib, 5
bcMCA, 7, *24*, *69*, *79*, *91*
bcPCA, 8, *23*, *69*, *79*
bootvalid_supvars, 10, *12*, *55*
bootvalid_variables, *11*, 11, *56*
break_interaction, 12
burt, 14

catdesc *(*wtable*)*, 108
coiMCA, 15, *17*, *92*
coiPCA, *16*, 16, *92*
conc.ellipse, 17, *83*, *86*, *88*
condesc, *29*
condesc *(*wtable*)*, 108
contrib, 19
csMCA, *5*, 20, *34*, *67*, *73, 74*, *82, 83*, *85*, *89*, *94*
cutree, *71*

DA, *9*, 22, *24*, *69*, *79*
DAQ, *8*, 23, *69*, *79*
darma *(*wtable*)*, 108
dichotom, *14*, 25
dichotomixed, 26
dimcontrib, 27, *29–31*, *82*, *101*
dimdesc, *28*
dimdescr, *28*, 28, *30, 31*, *101*
dimeta2, *28, 29*, 30, *31*, *101*
dimtypicality, *28–30*, 31, *66*, *101*

dist, *71*
dist.chi2, *4*, 32

flip.mca, 32

getindexcat, *7*, *15*, *20*, *23*, 33, *67, 68*, *70*, *93, 94*, *105*
ggadd_attractions, 34, *62*
ggadd_chulls, 35, *38*, *40, 41*, *43*, *45*, *49*, *51*, *64*
ggadd_corr, *36*, 37, *40, 41*, *43*, *45*, *49*, *51*, *64*
ggadd_density, *36*, *38*, 39, *41*, *43*, *45*, *49*, *51*, *64*
ggadd_ellipses, *36*, *38*, *40*, 40, *43*, *45*, *49*, *51*, *64*
ggadd_interaction, *13*, *36*, *38*, *40, 41*, 42, *45, 46*, *49*, *51*, *64*
ggadd_kellipses, *36*, *38*, *40, 41*, *43*, 43, *49*, *51*, *64*
ggadd_partial, *13*, 45
ggadd_supind, 47, *97*
ggadd_supvar, *36*, *38*, *40, 41*, *43*, *45, 46*, 48, *51*, *64*, *93*, *98*, *100*
ggadd_supvars, *36*, *38*, *40, 41*, *43*, *45, 46*, *49*, 49, *64*, *93*, *98*, *100*
ggassoc_boxplot *(*wtable*)*, 108
ggassoc_crosstab *(*wtable*)*, 108
ggassoc_phiplot *(*wtable*)*, 108
ggassoc_scatter *(*wtable*)*, 108
ggaxis_variables, 51
ggbootvalid_supvars, *11*, 53, *56*
ggbootvalid_variables, *12*, *55*, 55
ggcloud_indiv, *33*, *36*, *40, 41*, *45*, *47*, 57, *61*, *94*
ggcloud_variables, *33*, *35*, *38*, *43*, *46*, *49*, *51*, *53*, *58*, 59, *62*, *94*
ggeta2_variables, 62
ggsmoothed_supvar, *36*, *38*, *40, 41*, *43*, *45*, *49*, *51*, 63
gPCA, 64

hclust, *71*
homog.test, 66

ijunk, *7*, *15*, *20, 21*, *23*, *34*, 67, *68*, *70*, *93, 94*, *105*
indsup *(*supind*)*, 96

MCA, *5*, *73*
MCAiv, *8*, *24*, 68, *79*
MCAoiv, 69, *81*, *95*, *106*, *108*
medoids, 71
modif.rate, 72
multiMCA, 73, *84, 85*, *92*
Music, 75

nsCA, 76, *78*
nsca.biplot, *77*, 77

pam, *71*
PCAiv, *9*, *23*, *69*, 78
PCAoiv, *70*, 80, *106*, *108*
pem *(*wtable*)*, 108
phi.table *(*wtable*)*, 108
planecontrib, 81
plot.csMCA, *18*, 82, *103, 104*
plot.multiMCA, *18*, *74*, 84
plot.speMCA, *18*, 85, *94*, *103, 104*
plot.stMCA, *18*, 87, *96*

quadrant, 88
quasindep, 89

reshape_between, 90
rvcoef, *16, 17*, 91

scaled.dev, 92
speMCA, *5, 6*, *20, 21*, *24*, *34*, *67*, *73, 74*, *85, 86*, *89*, 93
stMCA, *87, 88*, 95
supind, *47*, *93*, 96, *98*, *100*, *103*
supvar, *20*, *66*, *93*, *97*, 97, *100*, *104*
supvars, *66*, *93*, *97, 98*, 99, *104*

tabcontrib, *6*, *20*, *28*, *82*, 100
Taste, 102
textindsup, *97*, 103
textvarsup, *83*, *85, 86*, *88*, *93*, *98*, *100*, 104
translate.logit, 105

varsup *(*supvar*)*, 97
varsups *(*supvars*)*, 99

wcMCA, *70*, *81*, 105, *108*
wcPCA, *70*, *81*, *106*, 107
wtable, 108