

# Package ‘descriptio’

March 7, 2024

**Type** Package

**Title** Descriptive Statistical Analysis

**Version** 1.3

**Imports** MASS, ggplot2, rlang

**Suggests** rmarkdown, knitr, rmdformats, ggpattern, paletteer, GGally

**Description** Description of statistical associations between two variables : measures of local and global association between variables (phi, Cramér V, correlations, eta-squared, Goodman and Kruskal tau, permutation tests, etc.), multiple graphical representations of the associations between two variables (using 'ggplot2') and weighted statistics.

**License** GPL (>= 2)

**Encoding** UTF-8

**URL** <https://github.com/nicolas-robette/descriptio>,  
<https://nicolas-robette.github.io/descriptio/>

**BugReports** <https://github.com/nicolas-robette/descriptio/issues>

**LazyData** true

**RoxygenNote** 7.2.1

**NeedsCompilation** no

**Author** Nicolas Robette [aut, cre]

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

**Repository** CRAN

**Date/Publication** 2024-03-07 11:00:02 UTC

## R topics documented:

assoc.catcont . . . . .	2
assoc.catcont.by . . . . .	4
assoc.twocat . . . . .	5
assoc.twocat.by . . . . .	7
assoc.twocont . . . . .	10

assoc.twocont.by . . . . .	11
assoc.xx . . . . .	12
assoc.yx . . . . .	13
catdesc . . . . .	15
condesc . . . . .	17
darma . . . . .	18
ggassoc_assocplot . . . . .	20
ggassoc_bertin . . . . .	21
ggassoc_boxplot . . . . .	23
ggassoc_chiasmogram . . . . .	24
ggassoc_crosstab . . . . .	26
ggassoc_marimekko . . . . .	27
ggassoc_phiplot . . . . .	29
ggassoc_scatter . . . . .	31
Movies . . . . .	32
or.table . . . . .	32
pem.table . . . . .	33
phi.table . . . . .	35
profiles . . . . .	36
stat_twocat . . . . .	37
weighted.cor . . . . .	38
weighted.cor2 . . . . .	39
weighted.cov . . . . .	40
weighted.cov2 . . . . .	41
weighted.mad . . . . .	42
weighted.quantile . . . . .	43
weighted.sd . . . . .	44
weighted.table . . . . .	45
<b>Index</b>	<b>46</b>

---

assoc.catcont	<i>Measures the association between a categorical variable and a continuous variable</i>
---------------	--

---

## Description

Measures the association between a categorical variable and a continuous variable

## Usage

```
assoc.catcont(x, y, weights = NULL,
              na.rm.cat = FALSE, na.value.cat = "NA", na.rm.cont = FALSE,
              nperm = NULL, distrib = "asympt", digits = 3)
```

**Arguments**

x	the categorical variable (must be a factor)
y	the continuous variable (must be a numeric vector)
weights	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
na.rm.cat	logical, indicating whether NA values in the categorical variable (i.e. x) should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the categorical variable (see na.value.cat argument).
na.value.cat	character. 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 continuous variable (i.e. y) should be silently removed before the computation proceeds. Default is FALSE.
nperm	numeric. Number of permutations for the permutation test 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").
digits	integer. The number of digits (default is 3).

**Value**

A list with the following elements :

summary	summary statistics (mean, median, etc.) of the continuous variable for each level of the categorical variable
eta.squared	eta-squared between the two variables
permutation.pvalue	p-value from a permutation (i.e. non-parametric) test of independence
cor	point biserial correlation between the two variables, for each level of the categorical variable
cor.perm.pval	permutation p-value of the correlation between the two variables, for each level of the categorical variable
test.values	test-values as proposed by Lebart et al (1984)
test.values.pval	p-values corresponding to the test-values

**Author(s)**

Nicolas Robette

**References**

- Rakotomalala R., 'Comprendre la taille d'effet (effect size)', [[http://eric.univ-lyon2.fr/~ricco/cours/slides/effect\\_size.pdf](http://eric.univ-lyon2.fr/~ricco/cours/slides/effect_size.pdf)]
- Lebart L., Morineau A. and Warwick K., 1984, \*Multivariate Descriptive Statistical Analysis\*, John Wiley and sons, New-York.

**See Also**

[assoc.twocat](#), [assoc.twocont](#), [assoc.yx](#), [condesc](#), [catdesc](#), [darma](#)

**Examples**

```
data(Movies)
with(Movies, assoc.catcont(Country, Budget, nperm = 10))
```

---

assoc.catcont.by	<i>Measures the groupwise association between a categorical variable and a continuous variable</i>
------------------	--

---

**Description**

Measures the association between a categorical variable and a continuous variable, for each category of a group variable

**Usage**

```
assoc.catcont.by(x, y, by, weights = NULL,
                 na.rm.cat = FALSE, na.value.cat = "NA", na.rm.cont = FALSE,
                 nperm = NULL, distrib = "asympt", digits = 3)
```

**Arguments**

x	factor : the categorical variable
y	numeric vector : the continuous variable
by	factor : the group variable
weights	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
na.rm.cat	logical, indicating whether NA values in the categorical variable (i.e. x) should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the categorical variable (see na.value.cat argument).
na.value.cat	character. 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 continuous variable (i.e. y) should be silently removed before the computation proceeds. Default is FALSE.
nperm	numeric. Number of permutations for the permutation test 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").
digits	integer. The number of digits (default is 3).

**Value**

A list of items, one for each category of the group variable. Each item is a list with the following elements :

summary	summary statistics (mean, median, etc.) of the continuous variable for each level of the categorical variable
eta.squared	eta-squared between the two variables
permutation.pvalue	p-value from a permutation (i.e. non-parametric) test of independence
cor	point biserial correlation between the two variables, for each level of the categorical variable
cor.perm.pval	permutation p-value of the correlation between the two variables, for each level of the categorical variable
test.values	test-values as proposed by Lebart et al (1984)
test.values.pval	p-values corresponding to the test-values

**Author(s)**

Nicolas Robette

**References**

- Rakotomalala R., 'Comprendre la taille d'effet (effect size)', [[http://eric.univ-lyon2.fr/~ricco/cours/slides/effect\\_size.pdf](http://eric.univ-lyon2.fr/~ricco/cours/slides/effect_size.pdf)]  
 Lebart L., Morineau A. and Warwick K., 1984, \*Multivariate Descriptive Statistical Analysis\*, John Wiley and sons, New-York.

**See Also**

[assoc.catcont](#), [assoc.twocat](#), [assoc.twocont](#), [assoc.yx](#), [condesc](#), [catdesc](#), [darma](#)

**Examples**

```
data(Movies)
with(Movies, assoc.catcont.by(Country, Budget, ArtHouse, nperm = 10))
```

---

assoc.twocat	<i>Cross-tabulation and measures of association between two categorical variables</i>
--------------	---

---

**Description**

Cross-tabulation and measures of association between two categorical variables

**Usage**

```
assoc.twocat(x, y, weights = NULL, na.rm = FALSE, na.value = "NA",
             nperm = NULL, distrib = "asympt")
```

**Arguments**

x	the first categorical variable (must be a factor)
y	the second categorical variable (must be a factor)
weights	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the variables (see na.value argument).
na.value	character. Name of the level for NA category. Default is "NA". Only used if na.rm = FALSE.
nperm	numeric. Number of permutations for the permutation test 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).

**Value**

A list of lists with the following elements :

tables list :

freq	cross-tabulation frequencies
prop	percentages
rprop	row percentages
cprop	column percentages
expected	expected values

global list :

chi.squared	chi-squared value
cramer.v	Cramer's V between the two variables
permutation.pvalue	p-value from a permutation (i.e. non-parametric) test of independence
global.pem	global PEM
GK.tau.xy	Goodman and Kruskal tau (forward association, i.e. x is the predictor and y is the response)
GK.tau.yx	Goodman and Kruskal tau (backward association, i.e. y is the predictor and x is the respons)

local list :

std.residuals the table of standardized (i.e.Pearson) residuals.  
 adj.residuals the table of adjusted standardized residuals.  
 adj.res.pval the table of p-values of adjusted standardized residuals.  
 odds.ratios the table of odds ratios.  
 local.pem the table of local PEM  
 phi the table of the phi coefficients for each pair of levels  
 phi.perm.pval the table of permutation p-values for each pair of levels  
 gather : a data frame gathering informations, with one row per cell of the cross-tabulation.

### Note

The adjusted standardized residuals are strictly equivalent to test-values for nominal variables as proposed by Lebart et al (1984).

### Author(s)

Nicolas Robette

### References

Agresti, A. (2007). *An Introduction to Categorical Data Analysis*, 2nd ed. New York: John Wiley & Sons.

Rakotomalala R., *Comprendre la taille d'effet (effect size)*, [http://eric.univ-lyon2.fr/~ricco/cours/slides/effect\\_size.pdf](http://eric.univ-lyon2.fr/~ricco/cours/slides/effect_size.pdf)

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

### See Also

[assoc.catcont](#), [assoc.twocont](#), [assoc.yx](#), [condesc](#), [catdesc](#), [darma](#)

### Examples

```
data(Movies)
assoc.twocat(Movies$Country, Movies$ArtHouse, nperm=100)
```

---

assoc.twocat.by *Groupwise cross-tabulation and measures of association between two categorical variables*

---

### Description

Cross-tabulation and measures of association between two categorical variables, for each category of a group variable

**Usage**

```
assoc.twocat.by(x, y, by, weights = NULL, na.rm = FALSE, na.value = "NA",
               nperm = NULL, distrib = "asympt")
```

**Arguments**

x	factor : the first categorical variable
y	factor : the second categorical variable
by	factor : the group variable
weights	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the variables (see na.value argument).
na.value	character. Name of the level for NA category. Default is "NA". Only used if na.rm = FALSE.
nperm	numeric. Number of permutations for the permutation test 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).

**Value**

A list of items, one for each category of the group variable. Each item is a list of lists with the following elements :

tables list :

freq	cross-tabulation frequencies
prop	percentages
rprop	row percentages
cprop	column percentages
expected	expected values

global list :

chi.squared	chi-squared value
cramer.v	Cramer's V between the two variables
permutation.pvalue	p-value from a permutation (i.e. non-parametric) test of independence
global.pem	global PEM
GK.tau.xy	Goodman and Kruskal tau (forward association, i.e. x is the predictor and y is the response)
GK.tau.yx	Goodman and Kruskal tau (backward association, i.e. y is the predictor and x is the respons)



local list :

std.residuals the table of standardized (i.e.Pearson) residuals.  
adj.residuals the table of adjusted standardized residuals.  
adj.res.pval the table of p-values of adjusted standardized residuals.  
odds.ratios the table of odds ratios.  
local.pem the table of local PEM  
phi the table of the phi coefficients for each pair of levels  
phi.perm.pval the table of permutation p-values for each pair of levels

gather : a data frame gathering informations, with one row per cell of the cross-tabulation.

### Note

The adjusted standardized residuals are strictly equivalent to test-values for nominal variables as proposed by Lebart et al (1984).

### Author(s)

Nicolas Robette

### References

Agresti, A. (2007). *An Introduction to Categorical Data Analysis*, 2nd ed. New York: John Wiley & Sons.

Rakotomalala R., *Comprendre la taille d'effet (effect size)*, [http://eric.univ-lyon2.fr/~ricco/cours/slides/effect\\_size.pdf](http://eric.univ-lyon2.fr/~ricco/cours/slides/effect_size.pdf)

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

### See Also

[assoc.twocat](#), [assoc.catcont](#), [assoc.twocont](#), [assoc.yx](#), [condesc](#), [catdesc](#), [darma](#)

### Examples

```
data(Movies)
assoc.twocat.by(Movies$Country, Movies$ArtHouse, Movies$Festival, nperm=100)
```

---

 assoc.twocont

*Measures the association between two continuous variables*


---

**Description**

Measures the association between two continuous variables with Pearson, Spearman and Kendall correlations.

**Usage**

```
assoc.twocont(x, y, weights = NULL, na.rm = FALSE,
              nperm = NULL, distrib = "asympt")
```

**Arguments**

x	a continuous variable (must be a numeric vector)
y	a continuous variable (must be a numeric vector)
weights	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. Default is FALSE.
nperm	numeric. Number of permutations for the permutation test 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").

**Value**

A data frame with Pearson, Spearman and Kendall correlations. The correlation value is in the first row and a p-value from a permutation (so non parametric) test of independence is in the second row.

**Author(s)**

Nicolas Robette

**See Also**

[assoc.twocat](#), [assoc.catcont](#), [assoc.yx](#), [condesc](#), [catdesc](#), [darma](#)

**Examples**

```
## Hollander & Wolfe (1973), p. 187f.
## Assessment of tuna quality. We compare the Hunter L measure of
## lightness to the averages of consumer panel scores (recoded as
## integer values from 1 to 6 and averaged over 80 such values) in
## 9 lots of canned tuna.
```

```
x <- c(44.4, 45.9, 41.9, 53.3, 44.7, 44.1, 50.7, 45.2, 60.1)
y <- c( 2.6,  3.1,  2.5,  5.0,  3.6,  4.0,  5.2,  2.8,  3.8)
assoc.twocont(x,y,nperm=100)
```

---

assoc.twocont.by      *Measures the groupwise association between two continuous variables*

---

### Description

Measures the association between two continuous variables with Pearson, Spearman and Kendall correlations, for each category of a group variable.

### Usage

```
assoc.twocont.by(x, y, by, weights = NULL, na.rm = FALSE,
                 nperm = NULL, distrib = "asympt")
```

### Arguments

x	numeric vector : a continuous variable
y	numeric vector : a continuous variable
by	factor : the group variable
weights	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. Default is FALSE.
nperm	numeric. Number of permutations for the permutation test 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").

### Value

A list of items, one for each category of the groupe variable. Each item is a data frame with Pearson, Spearman and Kendall correlations. The correlation value is in the first row and a p-value from a permutation (so non parametric) test of independence is in the second row.

### Author(s)

Nicolas Robette

### See Also

[assoc.twocont](#), [assoc.twocat](#), [assoc.catcont](#), [assoc.yx](#), [condesc](#), [catdesc](#), [darma](#)

## Examples

```
## Hollander & Wolfe (1973), p. 187f.
## Assessment of tuna quality. We compare the Hunter L measure of
## lightness to the averages of consumer panel scores (recoded as
## integer values from 1 to 6 and averaged over 80 such values) in
## 9 lots of canned tuna.
x <- c(44.4, 45.9, 41.9, 53.3, 44.7, 44.1, 50.7, 45.2, 60.1)
y <- c( 2.6,  3.1,  2.5,  5.0,  3.6,  4.0,  5.2,  2.8,  3.8)
group <- factor(c("A","B","C","C","B","A","A","C","B"))
assoc.twocont.by(x,y,group,nperm=100)
```

---

assoc.xx

*Bivariate association measures between pairs of variables.*

---

## Description

Computes bivariate association measures between every pairs of variables from a data frame.

## Usage

```
assoc.xx(x, weights = NULL, correlation = "kendall",
na.rm.cat = FALSE, na.value.cat = "NA", na.rm.cont = FALSE,
nperm = NULL, distrib = "asympt", dec = c(3,3))
```

## Arguments

x	the data frame of variables
weights	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
correlation	character. The type of measure of correlation measure to use between two continuous variables: "pearson", "spearman" or "kendall" (default).
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. 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 continuous variables should be silently removed before the computation proceeds. Default is FALSE.
nperm	numeric. Number of permutations for the permutation test 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").
dec	vector of 2 integers for number of decimals. The first value if for association measures, the second for permutation p-values. Default is c(3,3).

**Details**

The function computes an association measure : Pearson's, Spearman's or Kendall's correlation for pairs of numeric variables, Cramer's V for pairs of factors and eta-squared for pairs numeric-factor. It can also compute the p-value of a permutation test of association for each pair of variables.

**Value**

A table with the following elements :

```
measure      : name of the association measure
association   : value of the association measure
permutation.pvalue
               : p-value from the permutation test
```

**Author(s)**

Nicolas Robette

**See Also**

[darma](#), [assoc.twocat](#), [assoc.twocont](#), [assoc.catcont](#), [condesc](#), [catdesc](#), [assoc.yx](#)

**Examples**

```
data(iris)
iris2 = iris
iris2$Species = factor(iris$Species == "versicolor")
assoc.xx(iris2, nperm = 10)
```

---

assoc.yx

*Bivariate association measures between a response and predictor variables.*

---

**Description**

Computes bivariate association measures between a response and predictor variables (and, optionally, between every pairs of predictor variables.)

**Usage**

```
assoc.yx(y, x, weights = NULL, xx = TRUE, correlation = "kendall",
na.rm.cat = FALSE, na.value.cat = "NA", na.rm.cont = FALSE,
nperm = NULL, distrib = "asympt", dec = c(3,3))
```

**Arguments**

<code>y</code>	the response variable
<code>x</code>	the predictor variables
<code>weights</code>	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
<code>xx</code>	whether the association measures should be computed for couples of predictor variables (default) or not. With a lot of predictors, consider setting <code>xx</code> to FALSE (for reasons of computation time).
<code>correlation</code>	character. The type of measure of correlation measure to use between two continuous variables : "pearson", "spearman" or "kendall" (default).
<code>na.rm.cat</code>	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 <code>na.value.cat</code> argument).
<code>na.value.cat</code>	character. Name of the level for NA category. Default is "NA". Only used if <code>na.rm.cat = FALSE</code> .
<code>na.rm.cont</code>	logical, indicating whether NA values in the continuous variables should be silently removed before the computation proceeds. Default is FALSE.
<code>nperm</code>	numeric. Number of permutations for the permutation test of independence. If NULL (default), no permutation test is performed.
<code>distrib</code>	the null distribution of permutation test of independence can be approximated by its asymptotic distribution ("asympt", default) or via Monte Carlo resampling ("approx").
<code>dec</code>	vector of 2 integers for number of decimals. The first value is for association measures, the second for permutation p-values. Default is <code>c(3,3)</code> .

**Details**

The function computes an association measure : Pearson's, Spearman's or Kendall's correlation for pairs of numeric variables, Cramer's V for pairs of factors and eta-squared for pairs numeric-factor. It can also compute the p-value of a permutation test of association for each pair of variables.

**Value**

A list of the following items :

<code>YX</code>	: a table with the association measures between the response and predictor variables
<code>XX</code>	: a table with the association measures between every pairs of predictor variables

In each table :

<code>measure</code>	: name of the association measure
<code>association</code>	: value of the association measure
<code>permutation.pvalue</code>	: p-value from the permutation test

**Author(s)**

Nicolas Robette

**See Also**[darma](#), [assoc.twocat](#), [assoc.twocont](#), [assoc.catcont](#), [condesc](#), [catdesc](#)**Examples**

```
data(iris)
iris2 = iris
iris2$Species = factor(iris$Species == "versicolor")
assoc.yx(iris2$Species, iris2[,1:4], nperm=10)
```

---

catdesc	<i>Measures the association between a categorical variable and some continuous and/or categorical variables</i>
---------	---

---

**Description**

Measures the association between a categorical variable and some continuous and/or categorical variables

**Usage**

```
catdesc(y, x, weights = NULL,
na.rm.cat = FALSE, na.value.cat = "NA", na.rm.cont = FALSE,
measure = "phi", limit = NULL, correlation = "kendall", robust = TRUE,
nperm = NULL, distrib = "asympt", digits = 2)
```

**Arguments**

y	the categorical variable to describe (must be a factor)
x	a data frame with continuous and/or categorical variables
weights	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
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. 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 continuous variables should be silently removed before the computation proceeds. Default is FALSE.

measure	character. The measure of local association between categories of categorical variables. Can be "phi" for phi coefficient (default), "or" for odds ratios, "std.residuals" for standardized (i.e. Pearson) residuals, "adj.residuals" for adjusted standardized residuals or "pem" for local percentages of maximum deviation from independence.
limit	for the relationship between y and a categorical variable, only associations higher or equal to limit will be displayed. If NULL (default), they are all displayed.
correlation	character. The type of measure of correlation measure to use between two continuous variables : "pearson", "spearman" or "kendall" (default).
robust	logical. If TRUE (default), median and mad are used instead of mean and standard deviation.
nperm	numeric. Number of permutations for the permutation test 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").
digits	numeric. Number of digits for mean, median, standard deviation and mad. Default is 2.

**Value**

A list of the following items :

variables	associations between y and the variables in x
bylevel	a list with one element for each level of y

Each element in bylevel has the following items :

categories	a data frame with categorical variables from x and local associations
continuous.var	a data frame with continuous variables from x and associations measured by correlation coefficients

**Note**

If nperm is not NULL, permutation tests of independence are computed and the p-values from these tests are provided.

**Author(s)**

Nicolas Robette

**References**

Rakotomalala R., 'Comprendre la taille d'effet (effect size)', [[http://eric.univ-lyon2.fr/~ricco/cours/slides/effect\\_size.pdf](http://eric.univ-lyon2.fr/~ricco/cours/slides/effect_size.pdf)]

**See Also**

[catdes](#), [condesc](#), [assoc.yx](#), [darma](#)



**Examples**

```
data(Movies)
catdesc(Movies$ArtHouse, Movies[,c("Budget", "Genre", "Country")])
```

---

condesc	<i>Measures the association between a continuous variable and some continuous and/or categorical variables</i>
---------	--

---

**Description**

Measures the association between a continuous variable and some continuous and/or categorical variables

**Usage**

```
condesc(y, x, weights = NULL,
na.rm.cat = FALSE, na.value.cat = "NA", na.rm.cont = FALSE,
limit = NULL, correlation = "kendall", robust = TRUE,
nperm = NULL, distrib = "asympt", digits = 2)
```

**Arguments**

y	the continuous variable to describe
x	a data frame with continuous and/or categorical variables
weights	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
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. 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 continuous variables should be silently removed before the computation proceeds. Default is FALSE.
limit	for the relationship between y and a category of a categorical variable, only associations (point-biserial correlations) higher or equal to limit will be displayed. If NULL (default), they are all displayed.
correlation	character. The type of correlation measure to use between two continuous variables : "pearson", "spearman" or "kendall" (default).
robust	logical. If TRUE (default), meadian and mad are used instead of mean and standard deviation.
nperm	numeric. Number of permutations for the permutation test 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").
digits	numeric. Number of digits for mean, median, standard deviation and mad. Default is 2.

**Value**

A list of the following items :

variables	associations between y and the variables in x
categories	a data frame with categorical variables from x and associations measured by point biserial correlation.

**Note**

If nperm is not NULL, permutation tests of independence are computed and the p-values from these tests are provided.

**Author(s)**

Nicolas Robette

**References**

Rakotomalala R., 'Comprendre la taille d'effet (effect size)', [[http://eric.univ-lyon2.fr/~ricco/cours/slides/effect\\_size.pdf](http://eric.univ-lyon2.fr/~ricco/cours/slides/effect_size.pdf)]

**See Also**

[condes](#), [catdesc](#), [assoc.yx](#), [darma](#)

**Examples**

```
data(Movies)
condesc(Movies$BoxOffice, Movies[,c("Budget", "Genre", "Country")])
```

---

darma

*Describes Associations as in a Regression Model Analysis.*

---

**Description**

Computes bivariate association measures between a response and predictor variables, producing a summary looking like a regression analysis.

**Usage**

```
darma(y, x, weights = NULL, target = 1,
      na.rm.cat = FALSE, na.value.cat = "NA", na.rm.cont = FALSE,
      correlation = "kendall",
      nperm = NULL, distrib = "asympt", dec = c(1,3,3))
```

**Arguments**

<code>y</code>	the response variable
<code>x</code>	the predictor variables
<code>weights</code>	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
<code>target</code>	rank or name of the category of interest when <code>y</code> is categorical
<code>na.rm.cat</code>	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 <code>na.value.cat</code> argument).
<code>na.value.cat</code>	character. Name of the level for NA category. Default is "NA". Only used if <code>na.rm.cat = FALSE</code> .
<code>na.rm.cont</code>	logical, indicating whether NA values in the continuous variables should be silently removed before the computation proceeds. Default is FALSE.
<code>correlation</code>	character. The type of measure of correlation measure to use between two continuous variables : "pearson", "spearman" or "kendall" (default).
<code>nperm</code>	numeric. Number of permutations for the permutation test of independence. If NULL (default), no permutation test is performed.
<code>distrib</code>	the null distribution of permutation test of independence can be approximated by its asymptotic distribution ("asympt", default) or via Monte Carlo resampling ("approx").
<code>dec</code>	vector of 3 integers for number of decimals. The first value if for percents or medians, the second for association measures, the third for permutation p-values. Default is <code>c(1,3,3)</code> .

**Details**

The function computes association measures (phi, correlation coefficient, Kendall's correlation) between the variable of interest and the other variables. It can also compute the p-values permutation tests.

**Value**

A data frame

**Author(s)**

Nicolas Robette

**See Also**

[assoc.yx](#), [assoc.twocat](#), [assoc.twocont](#), [assoc.catcont](#), [condesc](#), [catdesc](#)

**Examples**

```
data(iris)
iris2 = iris
iris2$Species = factor(iris$Species == "versicolor")
darma(iris2$Species, iris2[,1:4], target=2, nperm=100)
```

---

ggassoc\_assocplot      *Association plot*

---

**Description**

For a cross-tabulation, plots measures of local association with bars of varying height and width, using ggplot2.

**Usage**

```
ggassoc_assocplot(data, mapping, measure = "std.residuals",
                  limits = NULL, sort = "none",
                  na.rm = FALSE, na.value = "NA",
                  colors = NULL, direction = 1, legend = "right")
```

**Arguments**

data	dataset to use for plot
mapping	aesthetics being used. x and y are required, weight can also be specified.
measure	character. The measure of association used to fill the rectangles. Can be "phi" for phi coefficient, "or" for odds ratios, "std.residuals" (default) for standardized (i.e. Pearson) residuals, "adj.residuals" for adjusted standardized residuals or "pem" for local percentages of maximum deviation from independence.
limits	a numeric vector of length two providing limits of the scale. If NULL (default), the limits are automatically adjusted to the data.
sort	character. If "both", rows and columns are sorted according to the first factor of a correspondence analysis of the contingency table. If "x", only rows are sorted. If "y", only columns are sorted. If "none" (default), no sorting is done.
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the variables (see na.value argument).
na.value	character. Name of the level for NA category. Default is "NA". Only used if na.rm = FALSE.
colors	vector of colors that will be interpolated to produce a color gradient. If NULL (default), the "Temps" palette from rcartocolors package is used.
direction	Sets the order of colours in the scale. If 1, the default, colours are as output by RColorBrewer::brewer.pal(). If -1, the order of colours is reversed.
legend	the position of legend ("none", "left", "right", "bottom", "top"). If "none", no legend is displayed.

## Details

The measure of local association measures how much each combination of categories of  $x$  and  $y$  is over/under-represented.

The bars vary in width according to the square root of the expected frequency. They vary in height and color shading according to the measure of association. If the measure chosen is "std.residuals" (Pearson's residuals), as in the original association plot from Cohen and Friendly, the area of the bars is proportional to the difference in observed and expected frequencies.

This function can be used as a high-level plot with `ggduo` and `ggpairs` functions of the `GGally` package.

## Value

a `ggplot` object

## Author(s)

Nicolas Robette

## References

Cohen, A. (1980), On the graphical display of the significant components in a two-way contingency table. *Communications in Statistics—Theory and Methods*, 9, 1025–1041. doi:10.1080/03610928008827940.

Friendly, M. (1992), Graphical methods for categorical data. *SAS User Group International Conference Proceedings*, 17, 190–200. <http://datavis.ca/papers/sugi/sugi17.pdf>

## See Also

[assoc.twocat](#), [phi.table](#), [catdesc](#), [assoc.yx](#), [darma](#), [ggassoc\\_crosstab](#), [ggpairs](#)

## Examples

```
data(Movies)
ggassoc_assocplot(data=Movies, mapping=ggplot2::aes(Country, Genre))
```

---

ggassoc\_bertin

*Bar plot of a crosstabulation inspired by Bertin*

---

## Description

For a cross-tabulation, plots bars for the conditional percentages of variable  $y$  according to variable  $x$ , using `ggplot2`. The general display is inspired by Bertin's plots.

## Usage

```
ggassoc_bertin(data, mapping, prop.width = FALSE,
sort = "none", add.gray = FALSE, add.rprop = FALSE,
na.rm = FALSE, na.value = "NA")
```

**Arguments**

data	dataset to use for plot
mapping	aesthetics being used. x and y are required, weight can also be specified.
prop.width	logical. If TRUE, the width of the bars is proportional to the margin percentages of variable x.
sort	character. If "both", rows and columns are sorted according to the first factor of a correspondence analysis of the contingency table. If "x", only variable x is sorted. If "y", only variable y is sorted. If "none" (default), no sorting is done.
add.gray	logical. If FALSE (default), only white and black are used to fill the bars. If TRUE, gray is used additionally to fill the part of the bars corresponding to margin percentages of variable y.
add.rprop	logical. If TRUE, row percentages are displayed on top of the bars. Default is FALSE.
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the variables (see na.value argument).
na.value	character. Name of the level for NA category. Default is "NA". Only used if na.rm = FALSE.

**Details**

The height of the bars is proportional to the conditional frequency of variable y. The bars are filled in black if the conditional frequency is higher than the marginal frequency; otherwise it's filled in white.

This graphical representation is inspired by the principles of Jacques Bertin and the online AMADO tool (<https://paris-timemachine.huma-num.fr/amado/main.html>).

Note : It does not allow faceting.

**Value**

a ggplot object

**Author(s)**

Nicolas Robette

**References**

J. Bertin: La graphique et le traitement graphique de l'information. Flammarion: Paris 1977.

**See Also**

[assoc.twocat](#), [phi.table](#), [catdesc](#), [ggassoc\\_crosstab](#), [ggassoc\\_assocplot](#), [ggassoc\\_phiplot](#), [ggassoc\\_chiasmogram](#)

**Examples**

```
data(Movies)
ggassoc_bertin(Movies, ggplot2::aes(x = Country, y = Genre))
ggassoc_bertin(Movies, ggplot2::aes(x = Country, y = Genre),
  sort = "both", prop.width = TRUE, add.gray = 3, add.rprop = TRUE)
```

---

ggassoc\_boxplot      *Boxplots with violins*

---

**Description**

Displays of boxplot and combines it with a violin plot, using ggplot2.

**Usage**

```
ggassoc_boxplot(data, mapping,
  na.rm.cat = FALSE, na.value.cat = "NA", na.rm.cont = FALSE,
  axes.labs = TRUE, ticks.labs = TRUE, text.size = 3,
  sort = FALSE, box = TRUE, notch = FALSE, violin = TRUE)
```

**Arguments**

data	dataset to use for plot
mapping	aesthetic being used. It must specify x and y.
na.rm.cat	logical, indicating whether NA values in the categorical variable (i.e. x) should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the categorical variable (see na.value.cat argument).
na.value.cat	character. Name of the level for NA category. Default is "NA". Only used if na.rm = FALSE.
na.rm.cont	logical, indicating whether NA values in the continuous variable (i.e. y) should be silently removed before the computation proceeds. Default is FALSE.
axes.labs	Whether to display the labels of the axes, i.e. the names of x and y. Default is TRUE.
ticks.labs	Whether to display the labels of the categories of x and y. Default is TRUE.
text.size	Size of the association measure. If NULL, the text is not added to the plot.
sort	logical. If TRUE, the levels of the categorical variable are reordered according to the conditional medians, so that boxplots are sorted. Default is FALSE.
box	Whether to draw boxplots. Default is TRUE.
notch	If FALSE (default) make a standard box plot. If TRUE, make a notched box plot. Notches are used to compare groups; if the notches of two boxes do not overlap, this suggests that the medians are significantly different.
violin	Whether to draw a violin plot. Default is TRUE.

**Details**

Eta-squared measure of global association between x and y is displayed in upper-left corner of the plot.

This function can be used as a high-level plot with `ggduo` and `ggpairs` functions of the `GGally` package.

**Value**

a `ggplot` object

**Author(s)**

Nicolas Robette

**See Also**

[assoc.catcont](#), [condesc](#), [assoc.yx](#), [darma](#), [ggpairs](#)

**Examples**

```
data(Movies)
ggassoc_boxplot(Movies, mapping = ggplot2::aes(x = Critics, y = ArtHouse))
```

---

`ggassoc_chiasmogram` *Plots counts and associations of a crosstabulation*

---

**Description**

For a cross-tabulation, plots the number of observations by using rectangles with proportional areas, and the phi measures of association between the categories with a diverging gradient of colour, using `ggplot2`.

**Usage**

```
ggassoc_chiasmogram(data, mapping, measure = "phi",
  limits = NULL, sort = "none",
  na.rm = FALSE, na.value = "NA",
  colors = NULL, direction = 1)
```

**Arguments**

<code>data</code>	dataset to use for plot
<code>mapping</code>	aesthetics being used. <code>x</code> and <code>y</code> are required, <code>weight</code> can also be specified.
<code>measure</code>	character. The measure of association used for filling the rectangles. Can be "phi" for phi coefficient (default), "or" for odds ratios, "residuals" for Pearson residuals, "std.residuals" for standardized Pearson residuals or "pem" for local percentages of maximum deviation from independence.



limits	a numeric vector of length two providing limits of the scale. If NULL (default), the limits are automatically adjusted to the data.
sort	character. If "both", rows and columns are sorted according to the first factor of a correspondence analysis of the contingency table. If "x", only rows are sorted. If "y", only columns are sorted. If "none" (default), no sorting is done.
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the variables (see na.value argument).
na.value	character. Name of the level for NA category. Default is "NA". Only used if na.rm = FALSE.
colors	vector of colors that will be interpolated to produce a color gradient. If NULL (default), the "Temps" palette from rcartocolors package is used.
direction	Sets the order of colours in the scale. If 1, the default, colours are as output by RColorBrewer::brewer.pal(). If -1, the order of colours is reversed.

### Details

The height of the rectangles is proportional to the marginal frequency of the row variable ; their width is proportional to the marginal frequency of the column variable. So the area of the rectangles is proportional to the expected frequency.

The rectangles are filled according to a measure of local association, which measures how much each combination of categories of x and y is over/under-represented.

This function can be used as a high-level plot with ggduo and ggpairs functions of the GGally package.

Note : It does not allow faceting.

### Value

a ggplot object

### Author(s)

Nicolas Robette

### References

Bozon Michel, Héran François. La découverte du conjoint. II. Les scènes de rencontre dans l'espace social. *Population*, 43(1), 1988, pp. 121-150.

### See Also

[assoc.twocat](#), [phi.table](#), [catdesc](#), [assoc.yx](#), [darma](#), [ggassoc\\_phiplot](#), [ggpairs](#)

### Examples

```
data(Movies)
ggassoc_chiasmogram(data=Movies, mapping=ggplot2::aes(Genre, Country))
```

---

ggassoc\_crosstab      *Proportional area plot*

---

### Description

For a cross-tabulation, plots the observed (or expected) frequencies by using rectangles with proportional areas, and the measures of local association between the categories with a diverging gradient of colour, using ggplot2.

### Usage

```
ggassoc_crosstab(data, mapping, size = "freq", max.size = 20,
  measure = "phi", limits = NULL, sort = "none",
  na.rm = FALSE, na.value = "NA",
  colors = NULL, direction = 1, legend = "right")
```

### Arguments

data	dataset to use for plot
mapping	aesthetics being used. x and y are required, weight can also be specified.
size	character. If "freq" (default), areas are proportional to observed frequencies. If "expected", they are proportional to expected frequencies.
max.size	numeric value, specifying the maximum size of the squares. Default is 20.
measure	character. The measure of association used for filling the rectangles. Can be "phi" for phi coefficient (default), "or" for odds ratios, "std.residuals" for standardized residuals, "adj.residuals" for adjusted standardized residuals or "pem" for local percentages of maximum deviation from independence.
limits	a numeric vector of length two providing limits of the scale. If NULL (default), the limits are automatically adjusted to the data.
sort	character. If "both", rows and columns are sorted according to the first factor of a correspondence analysis of the contingency table. If "x", only rows are sorted. If "y", only columns are sorted. If "none" (default), no sorting is done.
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the variables (see na.value argument).
na.value	character. Name of the level for NA category. Default is "NA". Only used if na.rm = FALSE.
colors	vector of colors that will be interpolated to produce a color gradient. If NULL (default), the "Temps" palette from rcartocolors package is used.
direction	Sets the order of colours in the scale. If 1, the default, colours are as output by RColorBrewer::brewer.pal(). If -1, the order of colours is reversed.
legend	the position of legend ("none", "left", "right", "bottom", "top"). If "none", no legend is displayed.

## Details

The measure of local association measures how much each combination of categories of x and y is over/under-represented.

The areas of the rectangles are proportional to observed or expected frequencies. Their color shading varies according to the measure of association.

This function can be used as a high-level plot with `ggduo` and `ggpairs` functions of the `GGally` package.

## Value

a `ggplot` object

## Author(s)

Nicolas Robette

## See Also

[assoc.twocat](#), [phi.table](#), [catdesc](#), [assoc.yx](#), [darma](#), [ggassoc\\_phiplot](#), [ggpairs](#)

## Examples

```
data(Movies)
ggassoc_crosstab(data=Movies, mapping=ggplot2::aes(Genre, Country))
```

---

`ggassoc_marimekko`      *Marimekko plot*

---

## Description

For a cross-tabulation, plots a marimekko chart (also called mosaic plot), using `ggplot2`.

## Usage

```
ggassoc_marimekko(data, mapping, type = "classic",
  measure = "phi", limits = NULL,
  na.rm = FALSE, na.value = "NA",
  palette = NULL, colors = NULL, direction = 1,
  linecolor = "gray60", linewidth = 0.1,
  sort = "none", legend = "right")
```

**Arguments**

<code>data</code>	dataset to use for plot
<code>mapping</code>	aesthetics being used. <code>x</code> and <code>y</code> are required, <code>weight</code> can also be specified.
<code>type</code>	character. If "classic" (default), a simple marimekko chart is plotted, with no use of local associations. If type is "shades", tiles are shaded according to the local associations between categories. If type is "patterns", tiles are filled with patterns, and the density of patterns is proportional to the absolute level of local association between categories.
<code>measure</code>	character. The measure of association used for filling (if type is "shades") or patterning (if type is "patterns") the tiles. Can be "phi" for phi coefficient, "or" for odds ratios, "std.residuals" (default) for standardized (i.e. Pearson) residuals, "adj.residuals" for adjusted standardized residuals or "pem" for local percentages of maximum deviation from independence.
<code>limits</code>	a numeric vector of length two providing limits of the scale. If NULL (default), the limits are automatically adjusted to the data. Only used for type "shades".
<code>na.rm</code>	logical, indicating whether NA values should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the variables (see <code>na.value</code> argument).
<code>na.value</code>	character. Name of the level for NA category. Default is "NA". Only used if <code>na.rm = FALSE</code> .
<code>palette</code>	A character vector of color codes. The number of colors should be equal or higher than the number of categories in <code>y</code> . If NULL (default), the "Tableau" palette from <code>ggthemes</code> package is used. Only used for types "classic" and "patterns".
<code>colors</code>	vector of colors that will be interpolated to produce a color gradient. If NULL (default), the "Temps" palette from <code>rcartocolors</code> package is used. Only used for type "shades".
<code>direction</code>	Sets the order of colours in the scale. If 1, the default, colours are as output by <code>RColorBrewer::brewer.pal()</code> . If -1, the order of colours is reversed.
<code>linecolor</code>	character. Color of the contour lines of the tiles. Default is gray60.
<code>linewidth</code>	numeric. Width of the contour lines of the tiles. Default is 0.1.
<code>sort</code>	character. If "both", rows and columns are sorted according to the first factor of a correspondence analysis of the contingency table. If "x", only rows are sorted. If "y", only columns are sorted. If "none" (default), no sorting is done.
<code>legend</code>	the position of legend ("none", "left", "right", "bottom", "top"). If "none", no legend is displayed.

**Details**

The measure of local association measures how much each combination of categories of `x` and `y` is over/under-represented.

This function can be used as a high-level plot with `ggduo` and `ggpairs` functions of the `GGally` package.

Note : It does not allow faceting.

**Value**

a ggplot object

**Author(s)**

Nicolas Robette

**References**

Hartigan, J.A., and Kleiner, B. (1984), "A mosaic of television ratings". *The American Statistician*, 38, 32–35.

Friendly, M. (1994), "Mosaic displays for multi-way contingency tables". *Journal of the American Statistical Association*, 89, 190–200.

**See Also**

[assoc.twocat](#), [phi.table](#), [catdesc](#), [assoc.yx](#), [darma](#), [ggassoc\\_crosstab](#), [ggpairs](#)

**Examples**

```
data(Movies)
ggassoc_marimekko(data=Movies, mapping=ggplot2::aes(Genre, Country))
ggassoc_marimekko(data=Movies, mapping=ggplot2::aes(Genre, Country), type = "patterns")
ggassoc_marimekko(data=Movies, mapping=ggplot2::aes(Genre, Country), type = "shades")
```

---

ggassoc\_phiplot

*Bar plot of measures of local association of a crosstabulation*

---

**Description**

For a cross-tabulation, plots the measures of local association with bars of varying height, using ggplot2.

**Usage**

```
ggassoc_phiplot(data, mapping, measure = "phi",
                limit = NULL, sort = "none",
                na.rm = FALSE, na.value = "NA")
```

**Arguments**

data	dataset to use for plot
mapping	aesthetics being used. x and y are required, weight can also be specified.
measure	character. The measure of association used for filling the rectangles. Can be "phi" for phi coefficient (default), "or" for odds ratios, "std.residuals" for standardized residuals, "adj.residuals" for adjusted standardized residuals or "pem" for local percentages of maximum deviation from independence.

<code>limit</code>	numeric value, specifying the upper limit of the scale for the height of the bars, i.e. for the measures of association (the lower limit is set to 0-limit). It corresponds to the maximum absolute value of association one wants to represent in the plot. If NULL (default), the limit is automatically adjusted to the data.
<code>sort</code>	character. If "both", rows and columns are sorted according to the first factor of a correspondence analysis of the contingency table. If "x", only rows are sorted. If "y", only columns are sorted. If "none" (default), no sorting is done.
<code>na.rm</code>	logical, indicating whether NA values should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the variables (see <code>na.value</code> argument).
<code>na.value</code>	character. Name of the level for NA category. Default is "NA". Only used if <code>na.rm = FALSE</code> .

### Details

The measure of association measures how much each combination of categories of `x` and `y` is over/under-represented. The bars vary in width according to the number of observations in the categories of the column variable. They vary in height according to the measure of association. Bars are black if the association is positive and white if it is negative.

The genuine version of this plot (see Cibois, 2004) uses the measure of association called "pem", i.e. the local percentages of maximum deviation from independence.

This function can be used as a high-level plot with `ggduo` and `ggpairs` functions of the `GGally` package.

### Value

a `ggplot` object

### Author(s)

Nicolas Robette

### References

Cibois Philippe, 2004, *Les écarts à l'indépendance. Techniques simples pour analyser des données d'enquêtes*, Collection "Méthodes quantitatives pour les sciences sociales"

### See Also

[assoc.twocat](#), [phi.table](#), [catdesc](#), [assoc.yx](#), [darma](#), [ggassoc\\_crosstab](#), [ggpairs](#)

### Examples

```
data(Movies)
ggassoc_phiplot(data=Movies, mapping=ggplot2::aes(Country, Genre))
```

---

ggassoc\_scatter      *Scatter plot with a smoothing line*

---

### Description

Displays of scatter plot and adds a smoothing line, using ggplot2.

### Usage

```
ggassoc_scatter(data, mapping, na.rm = FALSE,  
axes.labs = TRUE, ticks.labs = TRUE, text.size = 3)
```

### Arguments

data	dataset to use for plot
mapping	aesthetic being used. It must specify x and y.
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. Default is FALSE.
axes.labs	Whether to display the labels of the axes, i.e. the names of x and y. Default is TRUE.
ticks.labs	Whether to display the labels of the categories of x and y. Default is TRUE.
text.size	Size of the association measure. If NULL, the text is not added to the plot.

### Details

Kendall's tau rank correlation between x and y is displayed in upper-left corner of the plot.

Smoothing is performed with gam.

This function can be used as a high-level plot with ggduo and ggpairs functions of the GGally package.

### Value

a ggplot object

### Author(s)

Nicolas Robette

### See Also

[assoc.twocont](#), [condesc](#), [assoc.yx](#), [darma](#), [ggpairs](#)

### Examples

```
data(Movies)  
ggassoc_scatter(Movies, mapping = ggplot2::aes(x = Budget, y = Critics))
```

---

Movies

*Movies (data)*

---

### Description

The data concerns a sample of 1000 Movies which were on screens in France and come of their characteristics.

### Usage

```
data(Movies)
```

### Format

A data frame with 1000 observations and the following 7 variables:

Budget numeric vector of movie budgets

Genre is a factor with 9 levels

Country is a factor with 4 level. Country of origin of the movie.

ArtHouse is a factor with levels No, Yes. Whether the movie had the "Art House" label.

Festival is a factor with levels No, Yes. Whether the movie was selected in Cannes, Berlin or Venise film festivals.

Critics numeric vector of average ratings from intellectual criticism.

BoxOffice numeric vector of number of admissions.

### Examples

```
data(Movies)
str(Movies)
```

---

or.table

*Computes the odds ratios for every cells of a contingency table*

---

### Description

Computes the odds ratio for every cells of the cross-tabulation between two categorical variables

### Usage

```
or.table(x, y, weights = NULL, na.rm = FALSE, na.value = "NA", digits = 3)
```



**Arguments**

x	the first categorical variable
y	the second categorical variable
weights	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the variables (see na.value argument).
na.value	character. Name of the level for NA category. Default is "NA". Only used if na.rm = FALSE.
digits	integer. The number of digits (default is 3). If NULL, the results are not rounded.

**Value**

A table with the odds ratios

**Author(s)**

Nicolas Robette

**See Also**

[assoc.twocat](#), [assoc.catcont](#), [condesc](#), [catdesc](#)

**Examples**

```
data(Movies)
or.table(Movies$Country, Movies$ArtHouse)
```

---

pem.table	<i>Computes the local and global Percentages of Maximum Deviation from Independence (pem)</i>
-----------	---

---

**Description**

Computes the local and global Percentages of Maximum Deviation from Independence (pem) of a contingency table.

**Usage**

```
pem.table(x, y, weights = NULL, sort = FALSE, na.rm = FALSE, na.value = "NA", digits = 1)
```

### Arguments

x	the first categorical variable
y	the second categorical variable
weights	an optional numeric vector of weights (by default, a vector of 1 for uniform weights)
sort	logical. Whether rows and columns are sorted according to a correspondence analysis or not (default is FALSE).
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the variables (see na.value argument).
na.value	character. Name of the level for NA category. Default is "NA". Only used if na.rm = FALSE.
digits	integer. The number of digits (default is 1). If NULL, the results are not rounded.

### Details

The Percentage of Maximum Deviation from Independence (pem) is an association measure for contingency tables and also provides attraction (resp. repulsion) measures in each cell of the crosstabulation (see Cibois, 1993). It is an alternative to khi2, Cramer's V coefficient, etc.

### Value

Returns a list:

peml	Table with local percentages of maximum deviation from independence
pemg	Numeric value, i.e. the global percentage of maximum deviation from independence

### Author(s)

Nicolas Robette

### References

Cibois P., 1993, Le pem, pourcentage de l'ecart maximum : un indice de liaison entre modalites d'un tableau de contingence, *Bulletin de methodologie sociologique*, n40, p.43-63. <https://cibois.pagesperso-orange.fr/bms93.pdf>

### See Also

[table](#), [chisq.test](#), [phi.table](#), [assocstats](#)

### Examples

```
data(Movies)
pem.table(Movies$Country, Movies$ArtHouse)
```

---

`phi.table`*Computes the phi coefficient for every cells of a contingency table*

---

**Description**

Computes the phi coefficient for every cells of the cross-tabulation between two categorical variables

**Usage**

```
phi.table(x, y, weights = NULL, na.rm = FALSE, na.value = "NA", digits = 3)
```

**Arguments**

<code>x</code>	the first categorical variable
<code>y</code>	the second categorical variable
<code>weights</code>	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
<code>na.rm</code>	logical, indicating whether NA values should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the variables (see <code>na.value</code> argument).
<code>na.value</code>	character. Name of the level for NA category. Default is "NA". Only used if <code>na.rm = FALSE</code> .
<code>digits</code>	integer. The number of digits (default is 3). If NULL, the results are not rounded.

**Value**

A table with the phi coefficients

**Author(s)**

Nicolas Robette

**References**

Rakotomalala R., 'Comprendre la taille d'effet (effect size)', [http://eric.univ-lyon2.fr/~ricco/cours/slides/effect\\_size.pdf](http://eric.univ-lyon2.fr/~ricco/cours/slides/effect_size.pdf)

**See Also**

[assoc.twocat](#), [assoc.catcont](#), [condesc](#), [catdesc](#)

**Examples**

```
data(Movies)
phi.table(Movies$Country, Movies$ArtHouse)
```

---

`profiles`*Profiles by level of a categorical variable*

---

**Description**

Computes profiles (frequencies or percentages) for subgroups of observations defined by the levels of a categorical variable.

**Usage**

```
profiles(X, y, weights = NULL, stat = "cprop",
         mar = TRUE, digits = 1)
```

**Arguments**

<code>X</code>	data frame. The variables which are described in the profiles. There should be only factors.
<code>y</code>	factor. The categorical variable which defines subgroups of observations whose profiles will be computed.
<code>weights</code>	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
<code>stat</code>	character. Whether to compute frequencies ("freq"), percentages ("prop"), row percentages ("rprop") or column percentages ("cprop", default).
<code>mar</code>	logical, indicating whether to compute margins. Default is TRUE.
<code>digits</code>	numeric. Number of digits. Default is 1.

**Value**

A data frame with profiles in columns

**Author(s)**

Nicolas Robette

**See Also**

[catdesc](#), [assoc.yx](#), [darma](#), [assoc.twocat](#), [assoc.twocat.by](#)

**Examples**

```
data(Movies)
profiles(Movies[,c(2,4,5)], Movies$Country)
```

---

stat_twocat	<i>Cross-tabulation statistics for ggplot2</i>
-------------	--

---

### Description

Computes statistics of a cross-tabulation using `assoc.twocat` function.

### Usage

```
stat_twocat(mapping = NULL,
            data = NULL,
            geom = "point",
            position = "identity",
            ...,
            show.legend = NA,
            inherit.aes = TRUE)
```

### Arguments

mapping	Set of aesthetic mappings created by <code>aes()</code> . If specified and <code>inherit.aes = TRUE</code> (the default), it is combined with the default mapping at the top level of the plot. You must supply mapping if there is no plot mapping.
data	The data to be displayed in this layer. There are three options: If <code>NULL</code> , the default, the data is inherited from the plot data as specified in the call to <code>ggplot()</code> . A <code>data.frame</code> , or other object, will override the plot data. All objects will be fortified to produce a data frame. See <code>fortify()</code> for which variables will be created. A function will be called with a single argument, the plot data. The return value must be a <code>data.frame</code> , and will be used as the layer data. A function can be created from a formula (e.g. <code>~ head(.x, 10)</code> ).
geom	Override the default connection with <code>ggplot2::geom_point()</code> .
position	Position adjustment, either as a string naming the adjustment (e.g. <code>"jitter"</code> to use <code>position_jitter</code> ), or the result of a call to a position adjustment function. Use the latter if you need to change the settings of the adjustment.
...	Other arguments passed on to <code>layer()</code> . These are often aesthetics, used to set an aesthetic to a fixed value, like <code>colour = "red"</code> or <code>size = 3</code> . They may also be parameters to the paired <code>geom/stat</code> .
show.legend	logical. Should this layer be included in the legends? <code>NA</code> , the default, includes if any aesthetics are mapped. <code>FALSE</code> never includes, and <code>TRUE</code> always includes. It can also be a named logical vector to finely select the aesthetics to display.
inherit.aes	If <code>FALSE</code> , overrides the default aesthetics, rather than combining with them. This is most useful for helper functions that define both data and aesthetics and shouldn't inherit behaviour from the default plot specification, e.g. <code>borders()</code> .

### Value

A `ggplot2` plot with the added statistic.

**Author(s)**

Nicolas Robette

---

`weighted.cor`*Weighted correlation*

---

**Description**

Computes the weighted correlation between two distributions. This can be Pearson, Spearman or Kendall correlation.

**Usage**

```
weighted.cor(x, y, weights = NULL, method = "pearson", na.rm = FALSE)
```

**Arguments**

<code>x</code>	numeric vector
<code>y</code>	numeric vector
<code>weights</code>	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
<code>method</code>	a character string indicating which correlation coefficient is to be computed. One of "pearson" (default), "kendall", or "spearman".
<code>na.rm</code>	logical, indicating whether NA values should be silently removed before the computation proceeds. Default is FALSE.

**Value**

a length-one numeric vector

**Author(s)**

Nicolas Robette

**See Also**[weighted.sd](#), [weighted.cor2](#)**Examples**

```
data(Movies)
weighted.cor(Movies$Critics, Movies$BoxOffice, weights = rep(c(.8,1.2), 500))
weighted.cor(Movies$Critics, Movies$BoxOffice, weights = rep(c(.8,1.2), 500), method = "spearman")
```

---

weighted.cor2	<i>Weighted correlations</i>
---------------	------------------------------

---

**Description**

Computes a matrix of weighted correlations between the columns of *x* and the columns of *y*. This can be Pearson, Spearman or Kendall correlation.

**Usage**

```
weighted.cor2(x, y = NULL, weights = NULL, method = "pearson", na.rm = FALSE)
```

**Arguments**

<i>x</i>	a data frame of numeric vectors
<i>y</i>	an optional data frame of numeric vectors. Default is NULL, which means that correlations between the columns of <i>x</i> are computed.
<i>weights</i>	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
<i>method</i>	a character string indicating which correlation coefficient is to be computed. One of "pearson" (default), "kendall", or "spearman".
<i>na.rm</i>	logical, indicating whether NA values should be silently removed before the computation proceeds. Default is FALSE.

**Value**

a matrix of correlations

**Author(s)**

Nicolas Robette

**See Also**

[weighted.cor](#)

**Examples**

```
data(Movies)
weighted.cor2(Movies[,c("Budget", "Critics", "BoxOffice")], weights = rep(c(.8,1.2), 500))
```

---

weighted.cov	<i>Weighted covariance</i>
--------------	----------------------------

---

**Description**

Computes the weighted covariance between two distributions.

**Usage**

```
weighted.cov(x, y, weights = NULL, na.rm = FALSE)
```

**Arguments**

x	numeric vector
y	numeric vector
weights	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. Default is FALSE.

**Value**

a length-one numeric vector

**Author(s)**

Nicolas Robette

**See Also**

[weighted.sd](#), [weighted.cor](#), [weighted.cov2](#)

**Examples**

```
data(Movies)
weighted.cov(Movies$Critics, Movies$BoxOffice, weights = rep(c(.8,1.2), 500))
```



---

weighted.cov2	<i>Weighted covariances</i>
---------------	-----------------------------

---

**Description**

Computes a matrix of weighted covariances between the columns of `x` and the columns of `y`.

**Usage**

```
weighted.cov2(x, y = NULL, weights = NULL, na.rm = FALSE)
```

**Arguments**

<code>x</code>	a data frame of numeric vectors
<code>y</code>	an optional data frame of numeric vectors. Default is <code>NULL</code> , which means that covariances between the columns of <code>x</code> are computed.
<code>weights</code>	numeric vector of weights. If <code>NULL</code> (default), uniform weights (i.e. all equal to 1) are used.
<code>na.rm</code>	logical, indicating whether NA values should be silently removed before the computation proceeds. Default is <code>FALSE</code> .

**Value**

a matrix of covariances

**Author(s)**

Nicolas Robette

**See Also**

[weighted.cov](#)

**Examples**

```
data(Movies)
weighted.cov2(Movies[,c("Budget", "Critics", "BoxOffice")], weights = rep(c(.8,1.2), 500))
```

---

weighted.mad	<i>Weighted median absolute deviation to median</i>
--------------	---

---

## Description

Computes the weighted median absolute deviation to median (aka MAD) of a distribution.

## Usage

```
weighted.mad(x, weights = NULL, na.rm = FALSE)
```

## Arguments

x	numeric vector
weights	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. Default is FALSE.

## Value

a length-one numeric vector

## Author(s)

Nicolas Robette

## See Also

[weighted.quantile](#)

## Examples

```
data(Movies)
weighted.mad(Movies$Critics, weights = rep(c(.8,1.2), 500))
```

---

weighted.quantile      *Weighted quantiles*

---

**Description**

Computes the weighted quantiles of a distribution.

**Usage**

```
weighted.quantile(x, weights = NULL, probs = seq(0, 1, 0.25),  
                 na.rm = FALSE, names = FALSE)
```

**Arguments**

x	numeric vector whose sample quantiles are wanted
weights	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
probs	numeric vector of probabilities with values in [0,1]
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. Default is FALSE.
names	logical. if TRUE, the result has a names attribute. Default is FALSE.

**Value**

A numeric vector of the same length as probs argument.

**Note**

This function is taken from <https://stackoverflow.com/questions/2748725/is-there-a-weighted-median-function>

**See Also**

[weighted.mad](#)

**Examples**

```
data(Movies)  
weighted.quantile(Movies$Critics, weights = rep(c(.8,1.2), 500), names = TRUE)
```

---

weighted.sd	<i>Weighted standard deviation</i>
-------------	------------------------------------

---

**Description**

Computes the weighted standard deviation of a distribution.

**Usage**

```
weighted.sd(x, weights = NULL, na.rm = FALSE)
```

**Arguments**

x	numeric vector
weights	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. Default is FALSE.

**Value**

a length-one numeric vector

**Author(s)**

Nicolas Robette

**See Also**

[weighted.cor](#)

**Examples**

```
data(Movies)
weighted.sd(Movies$Critics, weights = rep(c(.8,1.2), 500))
```

---

weighted.table	<i>Computes a (possibly weighted) contingency table</i>
----------------	---

---

**Description**

Computes a contingency table from one or two vectors, with the possibility of specifying weights.

**Usage**

```
weighted.table(x, y = NULL, weights = NULL, stat = "freq",  
              mar = FALSE, na.rm = FALSE, na.value = "NA", digits = 1)
```

**Arguments**

x	an object which can be interpreted as factor
y	an optional object which can be interpreted as factor
weights	numeric vector of weights. If NULL (default), uniform weights (i.e. all equal to 1) are used.
stat	character. Whether to compute a contingency table ("freq", default), percentages ("prop"), row percentages ("rprop") or column percentages ("cprop").
mar	logical, indicating whether to compute margins. Default is FALSE.
na.rm	logical, indicating whether NA values should be silently removed before the computation proceeds. If FALSE (default), an additional level is added to the variables (see na.value argument).
na.value	character. Name of the level for NA category. Default is "NA". Only used if na.rm = FALSE.
digits	integer indicating the number of decimal places (default is 1)

**Value**

Returns a contingency table.

**Author(s)**

Nicolas Robette

**See Also**

[table](#), [assoc.twocat](#)

**Examples**

```
data(Movies)  
weighted.table(Movies$Country, Movies$ArtHouse)
```

# Index

## \* **aplot**

- ggassoc\_assocplot, 20
- ggassoc\_bertin, 21
- ggassoc\_boxplot, 23
- ggassoc\_chiasmogram, 24
- ggassoc\_crosstab, 26
- ggassoc\_marimekko, 27
- ggassoc\_phiplot, 29
- ggassoc\_scatter, 31

## \* **datasets**

- Movies, 32

## \* **misc**

- assoc.catcont, 2
- assoc.catcont.by, 4
- assoc.twocat, 5
- assoc.twocat.by, 7
- assoc.twocont, 10
- assoc.twocont.by, 11
- catdesc, 15
- condesc, 17
- or.table, 32
- pem.table, 33
- phi.table, 35
- profiles, 36
- weighted.cor, 38
- weighted.cor2, 39
- weighted.cov, 40
- weighted.cov2, 41
- weighted.mad, 42
- weighted.quantile, 43
- weighted.sd, 44
- weighted.table, 45

## \* **multivariate**

- assoc.catcont, 2
- assoc.catcont.by, 4
- assoc.twocat, 5
- assoc.twocat.by, 7
- assoc.twocont, 10
- assoc.twocont.by, 11

- catdesc, 15
- condesc, 17
- ggassoc\_assocplot, 20
- ggassoc\_bertin, 21
- ggassoc\_boxplot, 23
- ggassoc\_chiasmogram, 24
- ggassoc\_crosstab, 26
- ggassoc\_marimekko, 27
- ggassoc\_phiplot, 29
- ggassoc\_scatter, 31
- or.table, 32
- pem.table, 33
- phi.table, 35
- profiles, 36
- weighted.cor, 38
- weighted.cor2, 39
- weighted.cov, 40
- weighted.cov2, 41
- weighted.mad, 42
- weighted.quantile, 43
- weighted.sd, 44
- weighted.table, 45

aes(), 37

assoc.catcont, 2, 5, 7, 9–11, 13, 15, 19, 24, 33, 35

assoc.catcont.by, 4

assoc.twocat, 4, 5, 5, 9–11, 13, 15, 19, 21, 22, 25, 27, 29, 30, 33, 35, 36, 45

assoc.twocat.by, 7, 36

assoc.twocont, 4, 5, 7, 9, 10, 11, 13, 15, 19, 31

assoc.twocont.by, 11

assoc.xx, 12

assoc.yx, 4, 5, 7, 9–11, 13, 13, 16, 18, 19, 21, 24, 25, 27, 29–31, 36

assocstats, 34

borders(), 37

catdes, [16](#)  
catdesc, [4](#), [5](#), [7](#), [9–11](#), [13](#), [15](#), [15](#), [18](#), [19](#), [21](#),  
[22](#), [25](#), [27](#), [29](#), [30](#), [33](#), [35](#), [36](#)  
chisq.test, [34](#)  
condes, [18](#)  
condesc, [4](#), [5](#), [7](#), [9–11](#), [13](#), [15](#), [16](#), [17](#), [19](#), [24](#),  
[31](#), [33](#), [35](#)  
  
darma, [4](#), [5](#), [7](#), [9–11](#), [13](#), [15](#), [16](#), [18](#), [18](#), [21](#), [24](#),  
[25](#), [27](#), [29–31](#), [36](#)  
  
fortify(), [37](#)  
  
ggassoc\_assocplot, [20](#), [22](#)  
ggassoc\_bertin, [21](#)  
ggassoc\_boxplot, [23](#)  
ggassoc\_chiasmogram, [22](#), [24](#)  
ggassoc\_crosstab, [21](#), [22](#), [26](#), [29](#), [30](#)  
ggassoc\_marimekko, [27](#)  
ggassoc\_phiplot, [22](#), [25](#), [27](#), [29](#)  
ggassoc\_scatter, [31](#)  
ggpairs, [21](#), [24](#), [25](#), [27](#), [29–31](#)  
ggplot(), [37](#)  
ggplot2::geom\_point(), [37](#)  
  
layer(), [37](#)  
  
Movies, [32](#)  
  
or.table, [32](#)  
  
pem.table, [33](#)  
phi.table, [21](#), [22](#), [25](#), [27](#), [29](#), [30](#), [34](#), [35](#)  
profiles, [36](#)  
  
stat\_twocat, [37](#)  
StatTwocat(stat\_twocat), [37](#)  
  
table, [34](#), [45](#)  
  
weighted.cor, [38](#), [39](#), [40](#), [44](#)  
weighted.cor2, [38](#), [39](#)  
weighted.cov, [40](#), [41](#)  
weighted.cov2, [40](#), [41](#)  
weighted.mad, [42](#), [43](#)  
weighted.quantile, [42](#), [43](#)  
weighted.sd, [38](#), [40](#), [44](#)  
weighted.table, [45](#)