

Package ‘corrr’

October 10, 2016

Type Package

Version 0.2.1

Title Correlations in R

Description A tool for exploring correlations.

It makes it possible to easily perform routine tasks when exploring correlation matrices such as ignoring the diagonal, focusing on the correlations of certain variables against others, or rearranging and visualising the matrix in terms of the strength of the correlations.

URL <https://github.com/drsimonj/corrr>

BugReports <https://github.com/drsimonj/corrr/issues>

Depends R (>= 3.3.1), dplyr (>= 0.5.0)

LazyData yes

License MIT + file LICENSE

RoxygenNote 5.0.1

Suggests testthat (>= 1.0.2), knitr (>= 1.13), rmarkdown (>= 0.9.6)

Imports magrittr (>= 1.5), tidyr (>= 0.5.1), ggplot2 (>= 2.1.0),
seriation (>= 1.2-0), lazyeval (>= 0.2.0), purrr (>= 0.2.2),
tibble (>= 1.1), ggrepel (>= 0.5)

VignetteBuilder knitr

NeedsCompilation no

Author Simon Jackson [aut, cre]

Maintainer Simon Jackson <drsimonjackson@gmail.com>

Repository CRAN

Date/Publication 2016-10-10 13:51:49

R topics documented:

as_matrix	2
correlate	3
fashion	4
first_col	5
focus	5
network_plot	6
pair_n	7
rearrange	7
rplot	8
shave	9
stretch	9

Index	11
--------------	-----------

as_matrix	<i>Convert cor_df to original matrix.</i>
-----------	---

Description

Convert a correlation data frame to original matrix format.

Usage

```
as_matrix(x, diagonal = 1)
```

Arguments

x	cor_df. See correlate .
diagonal	Value (typically numeric or NA) to set the diagonal to. Default = 1.

Value

Correlation matrix

Examples

```
x <- correlate(mtcars)
as_matrix(x)
```

correlate	<i>Correlation Data Frame</i>
-----------	-------------------------------

Description

An implementation of `stats::cor()`, which returns a correlation matrix in a specific format. See details below. Additional adjustment include the use of pairwise deletion by default.

Usage

```
correlate(x, y = NULL, use = "pairwise.complete.obs", method = "pearson")
```

Arguments

<code>x</code>	a numeric vector, matrix or data frame.
<code>y</code>	NULL (default) or a vector, matrix or data frame with compatible dimensions to <code>x</code> . The default is equivalent to <code>y = x</code> (but more efficient).
<code>use</code>	an optional character string giving a method for computing covariances in the presence of missing values. This must be (an abbreviation of) one of the strings "everything", "all.obs", "complete.obs", "na.or.complete", or "pairwise.complete.obs".
<code>method</code>	a character string indicating which correlation coefficient (or covariance) is to be computed. One of "pearson" (default), "kendall", or "spearman": can be abbreviated.

Details

This function returns a correlation matrix in the following format:

A tibble (see [tibble](#))

- An additional class, "cor_df"
- A "rowname" column
- Standardised variances (the matrix diagonal) set to missing values (NA) so they can be ignored in calculations.

The main feature is the use of the data frame. This is to make use of data frame manipulation packages like `dplyr` and `tidyr`.

Value

A correlation data frame (`cor_df`)

Examples

```
## Not run:
correlate(iris)

## End(Not run)

correlate(mtcars)
correlate(iris[-5])
```

fashion

Fashion a correlation data frame for printing.

Description

For the purpose of printing, convert a correlation data frame into a noquote matrix with the correlations cleanly formatted (leading zeros removed; spaced for signs) and the diagonal (or any NA) left blank.

Usage

```
fashion(x, decimals = 2, leading_zeros = FALSE, na_print = "")
```

Arguments

x	Scalar, vector, matrix or data frame.
decimals	Number of decimal places to display for numbers.
leading_zeros	Should leading zeros be displayed for decimals (e.g., 0.1)? If FALSE, they will be removed.
na_print	Character string indicating NA values in printed output

Value

noquote. Also a data frame if x is a matrix or data frame.

Examples

```
# Examples with correlate()
mtcars %>% correlate() %>% fashion()
mtcars %>% correlate() %>% fashion(decimals = 1)
mtcars %>% correlate() %>% fashion(leading_zeros = TRUE)
mtcars %>% correlate() %>% fashion(na_print = "*")

# But doesn't have to include correlate()
mtcars %>% fashion(decimals = 3)
c(0.234, 134.23, -.23, NA) %>% fashion(na_print = "X")
```

first_col	<i>Add a first column to a data.frame</i>
-----------	---

Description

Add a first column to a data.frame. This is most commonly used to append a rowname column to create a cor_df.

Usage

```
first_col(df, ..., var = "rowname")
```

Arguments

df	Data frame
...	Values to go into the column
var	Label for the column. Default is "rowname"

Examples

```
first_col(mtcars, 1:nrow(mtcars))
```

focus	<i>Focus on section of a correlation data frame.</i>
-------	--

Description

Convenience function to select a set of variables from a correlation matrix to keep as the columns, and exclude these or all other variables from the rows. This function will take a [correlate](#) correlation matrix, and expression(s) suited for dplyr::select(). The selected variables will remain in the columns, and these, or all other variables, will be excluded from the rows based on 'same. For a complete list of methods for using this function, see [select](#).

Usage

```
focus(x, ..., mirror = FALSE)
```

```
focus_(x, ..., .dots, mirror)
```

Arguments

x	cor_df. See correlate .
...	Comma separated list of unquoted expressions. You can treat variable names like they are positions. Use positive values to select variables; use negative values to drop variables.
mirror	Boolean. Whether to mirror the selected columns in the rows or not.
.dots	Use focus_ to do standard evaluations. See select .

Value

A tbl or, if mirror = TRUE, a cor_df (see [correlate](#)).

Examples

```
x <- correlate(mtcars)
focus(x, mpg, cyl) # Focus on correlations of mpg and cyl with all other variables
focus(x, -disp, -mpg, mirror = TRUE) # Remove disp and mpg from columns and rows

x <- correlate(iris[-5])
focus(x, -matches("Sepal")) # Focus on correlations of non-Sepal
# variables with Sepal variables.
```

network_plot

Network plot of a correlation data frame

Description

Output a network plot of a correlation data frame in which variables that are more highly correlated appear closer together and are joined by stronger paths. Paths are also coloured by their sign (blue for positive and red for negative). The proximity of the points are determined using multidimensional clustering.

Usage

```
network_plot(rdf, min_cor = 0.3, legend = FALSE, colours = c("indianred2",
  "white", "skyblue1"), colors)
```

Arguments

rdf	Correlation data frame (cor_df) created with correlate .
min_cor	Number from 0 to 1 indicating the minimum value of correlations (in absolute terms) to plot.
legend	Boolean indicating whether a legend mapping the colours to the correlations should be displayed.
colours, colors	Vector of colours to use for n-colour gradient.

Examples

```
x <- correlate(mtcars)
network_plot(x)
network_plot(x, min_cor = .1)
network_plot(x, min_cor = .6)
network_plot(x, min_cor = .7, colors = c("red", "green"), legend = TRUE)
```

pair_n	<i>Number of pairwise complete cases.</i>
--------	---

Description

Compute the number of complete cases in a pairwise fashion for x (and y).

Usage

```
pair_n(x, y = NULL)
```

Arguments

x	a numeric vector, matrix or data frame.
y	NULL (default) or a vector, matrix or data frame with compatible dimensions to x. The default is equivalent to y = x (but more efficient).

Value

Matrix of pairwise sample sizes (number of complete cases).

Examples

```
pair_n(mtcars)
```

rearrange	<i>Re-arrange a correlation data frame</i>
-----------	--

Description

Re-arrange a correlation data frame to group highly correlated variables closer together.

Usage

```
rearrange(x, method = "PC", absolute = TRUE)
```

Arguments

x	cor_df. See correlate .
method	String specifying the arrangement (clustering) method. Clustering is achieved via seriate , which can be consulted for a complete list of clustering methods. Default = "PCA".
absolute	Boolean whether absolute values for the correlations should be used for clustering.

Value

cor_df. See [correlate](#).

Examples

```
x <- correlate(mtcars)

rearrange(x) # Default settings
rearrange(x, method = "HC") # Different seriation method
rearrange(x, absolute = FALSE) # Not using absolute values for arranging
```

rplot

Plot a correlation data frame.

Description

Plot a correlation data frame using `ggplot2`.

Usage

```
rplot(rdf, legend = FALSE, shape = 16, colours = c("indianred2", "white",
  "skyblue1"), print_cor = FALSE, colors)
```

Arguments

rdf	Correlation data frame (cor_df) created with correlate .
legend	Boolean indicating whether a legend mapping the colours to the correlations should be displayed.
shape	geom_point aesthetic.
colours, colors	Vector of colours to use for n-colour gradient.
print_cor	Boolean indicating whether the correlations should be printed over the shapes.

Value

Plots a correlation data frame

Examples

```
x <- correlate(mtcars)
rplot(x)

# Common use is following rearrange and shave
x <- rearrange(x, absolute = FALSE)
x <- shave(x)
rplot(x)
rplot(x, print_cor = TRUE)
rplot(x, shape = 20, colors = c("red", "green"), legend = TRUE)
```

shave	<i>Shave off upper/lower triangle.</i>
-------	--

Description

Convert the upper or lower triangle of a correlation data frame (`cor_df`) to missing values.

Usage

```
shave(x, upper = TRUE)
```

Arguments

`x` `cor_df`. See [correlate](#).
`upper` Boolean. If TRUE, set upper triangle to NA; lower triangle if FALSE.

Value

`cor_df`. See [correlate](#).

Examples

```
x <- correlate(mtcars)
shave(x) # Default; shave upper triangle
shave(x, upper = FALSE) # shave lower triangle
```

stretch	<i>Stretch correlation data frame into long format.</i>
---------	---

Description

`stretch` is a specified implementation of `tidyr::gather()` to be applied to a correlation data frame. It will gather the columns into a long-format data frame. The `rowname` column is handled automatically.

Usage

```
stretch(x, na.rm = FALSE)
```

Arguments

`x` `cor_df`. See [correlate](#).
`na.rm` Boolean. Whether rows with an NA correlation (originally the matrix diagonal) should be dropped? Will automatically be set to TRUE if `mirror` is FALSE.

Value

tbl with three columns (x and y variables, and their correlation)

Examples

```
x <- correlate(mtcars)
stretch(x) # Convert all to long format
stretch(x, na.rm = FALSE) # omit NAs (diagonal in this case)

x <- shave(x) # use shave to set upper triangle to NA and then...
stretch(x, na.rm = FALSE) # omit all NAs, therefore keeping each
# correlation only once.
```

Index

`as_matrix`, 2

`correlate`, 2, 3, 5–9

`fashion`, 4

`first_col`, 5

`focus`, 5

`focus_(focus)`, 5

`geom_point`, 8

`network_plot`, 6

`pair_n`, 7

`rearrange`, 7

`rplot`, 8

`select`, 5

`seriate`, 7

`shave`, 9

`stretch`, 9

`tibble`, 3