

Package ‘corrgram’

July 3, 2015

Title Plot a Correlogram

Version 1.8

Date 2015-07-02

Type Package

Description Calculates correlation of variables and displays the results graphically. Included panel functions can display points, shading, ellipses, and correlation values with confidence intervals.

Imports graphics, grDevices, seriation, stats

Suggests knitr, Matrix

License GPL-2

LazyData yes

URL <https://github.com/kwstat/corrgram>

BugReports <https://github.com/kwstat/corrgram/issues/>

VignetteBuilder knitr

Author Kevin Wright [aut, cre]

Maintainer Kevin Wright <kw.stat@gmail.com>

NeedsCompilation no

Repository CRAN

Date/Publication 2015-07-03 05:44:50

R topics documented:

auto	2
baseball	3
corrgram	4
vote	7
Index	8

auto

Statistics of 1979 automobile models

Description

The data give the following statistics for 74 automobiles in the 1979 model year as sold in the US.

Usage

```
data(auto)
```

Format

A data frame with 74 observations on the following 14 variables.

Model Make and model of car.

Origin a factor with levels A E J

Price Price in dollars.

MPG Miles per gallon.

Rep78 Repair record for 1978 on 1 (worst) to 5 (best) scale.

Rep77 Repair record for 1978 on 1 to 5 scale.

Hroom Headroom in inches.

Rseat Rear seat clearance in inches.

Trunk Trunk volume in cubic feet.

Weight Weight in pounds.

Length Length in inches.

Turn Turning diameter in feet.

Displa Engine displacement in cubic inches.

Gratio Gear ratio for high gear.

Source

This data frame was created from <http://euclid.psych.yorku.ca/ftp/sas/sssg/data/auto.sas>.

References

Originally published in Chambers, Cleveland, Kleiner, and Tukey, *Graphical Methods for Data Analysis*, 1983, pages 352-355.

The data is from various sources, primarily *Consumer Reports*, April, 1979, and the United States government EPA statistics on fuel consumption.

Examples

```
corrgram(auto[, -c(1:2)])
```

baseball

Baseball Hitter's Data

Description

The data are for 322 Major League Baseball regular and substitute hitters in 1986.

Usage

```
data(baseball)
```

Format

A data frame with 322 observations on the following 22 variables.

Name The hitter/player's name

League Player's league (American/National) at the beginning of 1987

Team Player's team at the beginning of 1987

Position Player's position in 1986: 1B (first base), 2B (second base), 3B (third base), C (catcher), OF (outfield), DH (designated hitter), SS (short stop), UT (utility).

Atbat Number of times at bat in 1986

Hits Number of hits in 1986

Homer Number of home runs in 1986

Runs Number of runs in 1986

RBI Runs batted in during 1986

Walks Number of walks in 1986

Years Number of years in the major leagues

Atbatc Number of times at bat in his career

Hitsc Number of hits in career

Homer c Number of home runs in career

Runsc Number of runs in career

RBIc Number of Runs Batted In in career

Walksc Number of walks in career

Putouts Number of putouts in 1986

Assists Number of assists in 1986

Errors Number of errors in 1986

Salary Annual salary (in thousands) on opening day 1987

logSal Log of salary

Details

The levels of the player's positions have been collapsed to fewer levels for a simpler analysis. See the original data for the full list of positions.

Source

The version of the data used to create this data was found at <http://euclid.psych.yorku.ca/ftp/sas/sssg/data/baseball.sas>.

Michael Friendly analyzed the data in: **Corrgrams: Exploratory Displays for Correlation Matrices**, *The American Statistician*, Nov 2002, Vol 56.

References

The data was originally published for the 1988 ASA Statistical Graphics and Computing Data Exposition: <http://lib.stat.cmu.edu/data-expo/1988.html>.

The salary data were taken from Sports Illustrated, April 20, 1987. The salary of any player not included in that article is listed as an NA. The 1986 and career statistics were taken from The 1987 Baseball Encyclopedia Update published by Collier Books, Macmillan Publishing Company, New York.

Examples

```
vars2 <- c("Assists", "Atbat", "Errors", "Hits", "Homer", "logSal",
          "Putouts", "RBI", "Runs", "Walks", "Years")
corrgram(baseball[, vars2],
         lower.panel=panel.shade, upper.panel=panel.pie)
```

corrgram

Draw a correlogram

Description

The corrgram function produces a graphical display of a correlation matrix, called a correlogram. The cells of the matrix can be shaded or colored to show the correlation value.

Usage

```
corrgram(x, type=NULL, order = FALSE, labels, panel=panel.shade,
        lower.panel=panel, upper.panel=panel, diag.panel=NULL,
        text.panel=textPanel, label.pos=c(0.5, 0.5), label.srt=0, cex.labels=NULL,
        font.labels=1, row1atop=TRUE, dir="", gap=0, abs=FALSE,
        col.regions=colorRampPalette(c("red", "salmon", "white", "royalblue", "navy")),
        cor.method="pearson",
        ...)
```

Arguments

<code>x</code>	A <i>tall</i> data frame with one observation per row, or a correlation matrix.
<code>type</code>	Use 'data' or 'cor'/'corr' to explicitly specify that 'x' is data or a correlation matrix. Rarely needed.
<code>order</code>	Should variables be re-ordered? Use TRUE/"PCA" for PCA-based re-ordering. Options from the 'seriate' package include "OLO" for optimal leaf ordering, "GW", and "HC".
<code>labels</code>	Labels to use (instead of data frame variable names) for diagonal panels
<code>panel</code>	Function used to plot the contents of each panel
<code>lower.panel</code> , <code>upper.panel</code>	Separate panel functions used below/above the diagonal
<code>diag.panel</code> , <code>text.panel</code>	Panel function used on the diagonal
<code>label.pos</code>	Horizontal and vertical placement of label in diagonal panels
<code>label.srt</code>	String rotation for diagonal labels
<code>cex.labels</code> , <code>font.labels</code>	Graphics parameter for diagonal panels
<code>rowlaptop</code>	TRUE for diagonal like "\ ", FALSE for diagonal like " / ".
<code>dir</code>	Use <code>dir="left"</code> instead of 'rowlaptop'
<code>gap</code>	Distance between panels
<code>abs</code>	Use absolute value of correlations for clustering? Default FALSE
<code>col.regions</code>	A <i>function</i> returning a vector of colors
<code>cor.method</code>	Correlation method to use in panel functions. Default is 'pearson'. Alternatives: 'spearman', 'kendall'
<code>...</code>	Additional arguments passed to plotting methods.

Details

Note: Use the 'col.regions' argument to specify colors. Earlier versions used a function 'col.corrgram' to specify colors.

Non-numeric columns in the data will be ignored.

The off-diagonal panels are specified with `panel.pts`, `panel.pie`, `panel.shade`, `panel.bar`, `panel.ellipse`, `panel.conf`.

Diagonal panels are specified with `panel.txt`, `panel.minmax`, `panel.density`.

Use a NULL panel to omit drawing the panel.

This function is basically a modification of the `pairs.default` function with the use of customized panel functions.

The `panel.conf` function uses `cor.test` and calculates pearson correlations. Confidence intervals are not available in `cor.test` for other methods (kendall, spearman).

You can create your own panel functions by starting with one of the included panel functions and making suitable modifications. Note that because of the way the panel functions are called inside the main function, your custom panel function must include the arguments shown in the `panel.pts` function, even if the custom panel function does not use those arguments!

TODO: legend, grid graphics version.

Value

No value is returned. A plot is created.

Author(s)

Kevin Wright

References

Friendly, Michael. 2002. Corrgrams: Exploratory Displays for Correlation Matrices. *The American Statistician*, 56, 316–324. <http://datavis.ca/papers/corrgram.pdf>

A SAS macro by Michael Friendly is at <http://datavis.ca/sasmac/corrgram.html>.

D. J. Murdoch and E. D. Chow. 1996. A Graphical Display of Large Correlation Matrices. *The American Statistician*, 50, 178-180.

Examples

```
# To reproduce the figures in Michael Friendly's paper, see the file
# 'friendly.r' in this package's test directory.

# Demonstrate density panel, correlation confidence panel
corrgram(iris, lower.panel=panel.pts, upper.panel=panel.conf,
         diag.panel=panel.density)

# Demonstrate panel.shade, panel.pie, principal component ordering
vars2 <- c("Assists", "Atbat", "Errors", "Hits", "Homer", "logSal",
          "Putouts", "RBI", "Runs", "Walks", "Years")
corrgram(baseball[vars2], order=TRUE, main="Baseball data PC2/PC1 order",
         lower.panel=panel.shade, upper.panel=panel.pie)

# CAUTION: The latticeExtra package also has a 'panel.ellipse' function
# that clashes with the same-named function in corrgram. In order to use
# the right one, the example below uses 'lower.panel=corrgram::panel.ellipse'.
# If you do not have latticeExtra loaded, you can just use
# 'lower.panel=panel.ellipse'.

# Demonstrate panel.bar, panel.ellipse, panel.minmax, col.regions
corrgram(auto, order=TRUE, main="Auto data (PC order)",
         lower.panel=corrgram::panel.ellipse, upper.panel=panel.bar, diag.panel=panel.minmax,
         col.regions=colorRampPalette(c("darkgoldenrod4", "burlywood1", "darkkhaki", "darkgreen")))

# 'vote' is a correlation matrix, not a data frame
corrgram(vote, order=TRUE)
```

vote

Voting correlations

Description

Voting correlations

Usage

```
data(vote)
```

Format

A 12x12 matrix.

Details

No details.

Source

Iversen, Torben, and David Soskice. 2006. Electoral institutions and the politics of coalitions: Why some democracies redistribute more than others. *American Political Science Review*, 100, 165-81.

References

Using Graphs Instead of Tables. http://tables2graphs.com/doku.php?id=03_descriptive_statistics#a_correlation_matrix

Examples

```
corrgram(vote, order=TRUE)
```

Index

*Topic **hplot**
corrgram, 4

auto, 2

baseball, 3

corrgram, 4

panel.bar (corrgram), 4

panel.conf (corrgram), 4

panel.density (corrgram), 4

panel.ellipse (corrgram), 4

panel.minmax (corrgram), 4

panel.pie (corrgram), 4

panel.pts (corrgram), 4

panel.shade (corrgram), 4

panel.txt (corrgram), 4

vote, 7