Package 'corrgram'

July 9, 2018

Title Plot a Correlogram

Version 1.13

Date 2018-07-06

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. See Friendly (2002) <doi:10.1198/000313002533>.

Imports graphics, grDevices, seriation, stats

Suggests gridBase, knitr, Matrix, psych, rmarkdown, sfsmisc, testthat

License GPL-3

LazyData yes

Encoding UTF-8

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

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

VignetteBuilder knitr

RoxygenNote 6.0.1

NeedsCompilation no

Author Kevin Wright [aut, cre] (<https://orcid.org/0000-0002-0617-8673>)

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

Repository CRAN

Date/Publication 2018-07-09 21:30:03 UTC

R topics documented:

---

auto   *Statistics of 1979 automobile models*

Description

Statistics for 74 automobiles in the 1979 model year as sold in the US.

Usage

    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.

Details

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

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.

Examples

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

---

baseball   *Baseball Hitter's Data*

Description

Data are for 322 Major Leaque Baseball regular and substitute hitters in 1986.

Usage

```
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=Outfild, 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

Homerc 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.

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.

Source

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

References

Michael Friendly (2002). Corrgrams: Exploratory Displays for Correlation Matrices, *The American Statistician*, Vol 56.

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, row1attop = TRUE, dir = "",
  gap = 0, abs = FALSE, col.regions = colorRampPalette(c("red", "salmon",
  "white", "royalblue", "navy")), cor.method = "pearson",
  outer.labels = NULL, ...)
```

Arguments

x	A *tall* data frame with one observation per row, or a correlation matrix.
type	Use 'data' or 'cor'/'corr' to explicitly specify that 'x' is data or a correlation matrix. Rarely needed.
order	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".
labels	Labels to use (instead of data frame variable names) for diagonal panels. If 'order' option is used, this vector of labels will be also be appropriately reordered by the function.
panel	Function used to plot the contents of each panel.
lower.panel, upper.panel	
	Separate panel functions used below/above the diagonal.
diag.panel, text.panel	
	Panel function used on the diagonal.
label.pos	Horizontal and vertical placement of label in diagonal panels.
label.srt	String rotation for diagonal labels.
cex.labels, font.labels	
	Graphics parameter for diagonal panels.
row1attop	TRUE for diagonal like "\", FALSE for diagonal like "/".
dir	Use dir="left" instead of 'row1attop'.
gap	Distance between panels.
abs	Use absolute value of correlations for clustering? Default FALSE.
col.regions	A *function* returning a vector of colors.
cor.method	Correlation method to use in panel functions. Default is 'pearson'. Alternatives: 'spearman', 'kendall'.
outer.labels	A list of the form 'list(bottom,left,top,right)', each component of which is a list of the form 'list(labels,cex,srt)'. This is used to add labels along the outside edges of the corrgram. Defaults: 'cex=1', 'srt=90' (bottom/top), 'srt=0' (left/right).
...	Additional arguments passed to plotting methods.

Details

Note: Use the 'col.regions' argument 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.fill, 'panel.bar, panel.ellipse, panel.conf. panel.cor.

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

The correlation matrix used for plotting is returned. The 'order' and 'abs' arguments affect the returned value.

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

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
# vignette, or 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, upper.panel=panel.cor)


# outer labels, all options, larger margins, xlab, ylab
labs=colnames(state.x77)
corrgram(state.x77, oma=c(7, 7, 2, 2),
         outer.labels=list(bottom=list(labels=labs,cex=1.5,srt=60),
                           left=list(labels=labs,cex=1.5,srt=30)))
mtext("Bottom", side=1, cex=2, line = -1.5, outer=TRUE, xpd=NA)
mtext("Left", side=2, cex=2, line = -1.5, outer=TRUE, xpd=NA)
```

---

vote   *Voting correlations*

Description

Voting correlations

Usage

```
vote
```

Format

A 12x12 matrix.

Details

These are the correlations of traits, where each trait is measured for 17 developed countries (Europe, US, Japan, Australia, New Zealand).

Source

Torben Iversen 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. Table A2.

References

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

Examples

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

Index