

Package ‘andrews’

February 5, 2023

Type Package

Title Various Andrews Curves

Version 1.1.1

Maintainer Sigbert Klinke <sigbert@hu-berlin.de>

Depends R (>= 2.10)

Description Visualisation of multidimensional data through different Andrews curves:
Andrews, D. F. (1972) Plots of High-Dimensional Data. Biometrics, 28(1), 125-136. <doi:10.2307/2528964>.

License GPL-3

URL <https://github.com/sigbertklinke/andrews> (development version)

Encoding UTF-8

LazyData true

Imports grDevices, graphics, primes

Suggests knitr, rmarkdown, robustbase, mclust

VignetteBuilder knitr

RoxygenNote 7.2.3

NeedsCompilation no

Author Jaroslav Myslivec [aut],
Sigbert Klinke [cre, ctb]

Repository CRAN

Date/Publication 2023-02-05 20:42:43 UTC

R topics documented:

andrews	2
andrews0	4
banknote	5
deftype	6
normalize	7
numarray	8

outlyingness	8
selectand	9
ZZZ	10

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

andrews	<i>Andrews curves</i>
---------	-----------------------

Description

Andrews curves for visualization of multidimensional data. For colouring the curves see the details. For differences between `andrews` and `andrews0` see the vignette("andrews"). With the same parameters called both functions should create the same plot. `type==5` is a modification of `type==3` and `type==6` is a modification of `type==4`.

Usage

```
andrews(
  df,
  type = 1,
  clr = NULL,
  step = 100,
  ymax = 10,
  alpha = NULL,
  palcol = NULL,
  lwd = 1,
  lty = "solid",
  ...
)
```

Arguments

<code>df</code>	data frame or an R object that can be converted into a data frame with <code>as.data.frame</code>
<code>type</code>	type of curve <ul style="list-style-type: none"> 1: $f(t) = x_1/\sqrt{2} + x_2 \sin(t) + x_3 \cos(t) + x_4 \sin(2t) + x_5 \cos(2t) + \dots$ 2: $f(t) = x_1 \sin(t) + x_2 \cos(t) + x_3 \sin(2t) + x_4 \cos(2t) + \dots$ 3: $f(t) = x_1 \cos(t) + x_2 \cos(\sqrt{2}t) + x_3 \cos(\sqrt{3}t) + \dots$ 4: $f(t) = 0.5^{p/2}x_1 + 0.5^{(p-1)/2}x_2(\sin(t) + \cos(t)) + 0.5^{(p-2)/2}x_3(\sin(t) - \cos(t)) + 0.5^{(p-3)/2}x_4(\sin(2t) + \cos(2t)) + 0.5^{(p-4)/2}x_5(\sin(2t) - \cos(2t)) + \dots$ with p the number of variables 5: $f(t) = x_1 \cos(\sqrt{p_0}t) + x_2 \cos(\sqrt{p_1}t) + x_3 \cos(\sqrt{p_2}t) + \dots$ with $p_0 = 1$ and p_i the i-th prime number 6: $f(t) = 1/\sqrt{2}(x_1 + x_2(\sin(t) + \cos(t)) + x_3(\sin(t) - \cos(t)) + x_4(\sin(2t) + \cos(2t)) + x_5(\sin(2t) - \cos(2t)) + \dots)$
<code>clr</code>	number/name of column in the data frame for color of curves

step	smoothness of curves
y _{max}	maximum of y coordinate
alpha	semi-transparent color ($0 \leq \alpha \leq 1$) which are supported only on some devices
palcol	a function which generates a set of colors, see details
lwd	line width, a positive number, defaulting to 1.
lty	line type, can either be specified as an integer (0=blank, 1=solid (default), 2=dashed, 3=dotted, 4=dotdash, 5=longdash, 6=twodash) or as one of the character strings "blank", "solid", "dashed", "dotted", "dotdash", "longdash", or "twodash", where "blank" uses 'invisible lines' (i.e., does not draw them).
...	further named parameters given to <code>graphics::plot.default()</code> except x, y, and type.

Details

If `clr` has length one then it is used as column number or column name for coloring the curves:

- If `df[,clr]` is numeric then `palcol` must be function which returns colors for values in the range $\in[0, 1]$ using normalized variable. The default is function `function(v) { hsv(0, 1, v) }`.
- Otherwise `df[,clr]` is converted to a factor and `palcol` must be a function which returns for each level a color. The parameter for `palcol` is the number of levels and the default is `grDevices::rainbow()`. If the length of `clr` is the number of rows of `df` then `clr` is interpreted as colors.

Andrews curves transform multidimensional data into curves. This package presents four types of curves.

Value

nothing

Author(s)

Sigbert Klinke sigbert@hu-berlin.de, Jaroslav Myslivec jaroslav.myslivec@upce.cz

References

- Andrews, D. F. (1972) Plots of High-Dimensional Data. *Biometrics*, vol. 28, no. 1, pp. 125-136.
- Khattree, R., Naik, D. N. (2002) Andrews Plots for Multivariate Data: Some New Suggestions and Applications. *Journal of Statistical Planning and Inference*, vol. 100, no. 2, pp. 411-425.

Examples

```
data(iris)
op <- par(mfrow=c(1,2))
andrews0(iris,clr=5,ymax=3)
andrews(iris,clr=5,ymax=3)
par(op)
andrews(iris,type=4,clr=5,ymax=NA)
```

andrews0

*Andrews curves***Description**

Andrews curves for visualization of multidimensional data. For differences between `andrews` and `andrews2` see the `vignette("andrews")`. For colouring the curves see the details.

Usage

```
andrews0(
  df,
  type = 1,
  clr = NULL,
  step = 100,
  ymax = 10,
  main = NULL,
  sub = NULL
)
```

Arguments

<code>df</code>	data frame
<code>type</code>	type of curve <ul style="list-style-type: none"> 1: $f(t) = x_1/\sqrt{2} + x_2 \sin(t) + x_3 \cos(t) + x_4 \sin(2t) + x_5 \cos(2t) + \dots$ 2: $f(t) = x_1 \sin(t) + x_2 \cos(t) + x_3 \sin(2t) + x_4 \cos(2t) + \dots$ 3: $f(t) = 0.5^{p/2}x_1 + 0.5^{(p-1)/2}x_2(\sin(t) + \cos(t)) + 0.5^{(p-2)/2}x_3(\sin(t) - \cos(t)) + 0.5^{(p-3)/2}x_4(\sin(2t) + \cos(2t)) + 0.5^{(p-6)/2}x_5(\sin(2t) - \cos(2t)) + \dots$ with p the number of variables 4: $f(t) = 1/\sqrt{2}(x_1 + x_2(\sin(t) + \cos(t)) + x_3(\sin(t) - \cos(t)) + x_4(\sin(2t) + \cos(2t)) + x_5(\sin(2t) - \cos(2t)) + \dots)$
<code>clr</code>	number/name of column in the data frame for color of curves
<code>step</code>	smoothness of curves
<code>ymax</code>	maximum of y coordinate.
<code>main</code>	main title for the plot
<code>sub</code>	sub title for the plot

Details

Andrews curves transform multidimensional data into curves. This package presents four types of curves

If `df[,clr]` is numeric then `hsv(1,1,v)` with the normalized values (on $[0, 1]$) of `df[,clr]` is used. Otherwise the number of unique values in `nuv <- unique(df[,clr])` is used in connection with `rainbow(nuv)`.

Value

nothing

Author(s)

Jaroslav Myslivec jaroslav.myslivec@upce.cz

References

- Andrews, D. F. (1972) Plots of High-Dimensional Data. *Biometrics*, vol. 28, no. 1, pp. 125-136.
- Khatree, R., Naik, D. N. (2002) Andrews Plots for Multivariate Data: Some New Suggestions and Applications. *Journal of Statistical Planning and Inference*, vol. 100, no. 2, pp. 411-425.

Examples

```
data(iris)
andrews0(iris,clr=5,ymax=3)
andrews0(iris,type=4,clr=5,ymax=2)
```

banknote

Swiss banknotes data

Description

The data set contains six measurements made on 100 genuine and 100 counterfeit old-Swiss 1000-franc bank notes. The data frame and the documentation is a copy of [mclust::banknote](#).

Usage

```
banknote
```

Format

A data frame with 200 rows and 7 columns:

Status the status of the banknote: genuine or counterfeit

Length Length of bill (mm)

Left Width of left edge (mm)

Right Width of right edge (mm)
Bottom Bottom margin width (mm)
Top Top margin width (mm)
Diagonal Length of diagonal (mm)

Source

Flury, B. and Riedwyl, H. (1988). Multivariate Statistics: A practical approach. London: Chapman & Hall, Tables 1.1 and 1.2, pp. 5-8.

deftype	<i>deftype</i>
---------	----------------

Description

Defines a function which can be used as basis for Andrews curves $f_i(t) = \sum_{j=1}^p x_{ij} f_j(t)$.

Usage

```
deftype(index = NULL, FUN = NULL, xlim = c(-pi, pi))
```

Arguments

index	index/name of the function
FUN	function of the form function(n, t) {...}
xlim	default range for displaying curves (default: c(-pi, pi))

Value

either a list of all functions or a single function

Examples

```
# define a new andrews curve, just with sine curves
deftype("sine", function(n, t) {
  n <- as.integer(if (n<1) 1 else n)
  m <- matrix(NA, nrow=length(t), ncol=n)
  for (i in 1:n) m[,i] <- sin(i*t)
  m
})
andrews(iris, "sine")
# query
deftype()
deftype("sine")
```

normalize	<i>Normalization</i>
-----------	----------------------

Description

Normalization of a variable:

- `type==1`: `ar` normalized into $[0, 1]$,
- `type==2`: `ar` is standardized,
- otherwise no normalization is done.

Usage

```
normalize(ar, type = 1)
```

Arguments

<code>ar</code>	numeric variable.
<code>type</code>	integer: type of normalization (default: 1)

Details

Normalization of variable: $ar \leftarrow (ar - \min(ar)) / (\max(ar) - \min(ar))$

Value

Returns normalized variable.

Author(s)

Jaroslav Myslivec jaroslav.myslivec@upce.cz, Sigbert Klinke sigbert@hu-berlin.de

Examples

```
normalize(iris[,1])
```

numarray

Numeric array

Description

Extracts numeric array from data frame.

Usage

```
numarray(df)
```

Arguments

df data frame.

Details

Extracts numeric array from data frame.

Value

Returns numeric array.

Author(s)

Jaroslav Myslivec jaroslav.myslivec@upce.cz, Sigbert Klinke sigbert@hu-berlin.de

Examples

```
numarray(iris)
```

outlyingness

outlyingness

Description

Computes the Stahel-Donoho outlyingness. If type is any of the available types by [andrews\(\)](#) then the projection vectors are generated along the andrews curves. Otherwise step random directions will be used. Note that the projection vectors are always normalized to length one.

Usage

```
outlyingness(x, type = 1, step = 100, xlim = NULL, normalize = 1)
```


Arguments

x	data frame
type	type of curve, see andrews()
step	step smoothness of curves
xlim	the x limits (x1, x2)
normalize	type of normalization, see normalize()

Value

the Stahel-Donoho outlyingness

References

- Stahel, W. (1981), Robuste Schätzungen: infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen, PhD thesis, ETH Zürich.
- Donoho, D. (1982), Breakdown properties of multivariate location estimators, Ph.D. Qualifying paper, Dept. Statistics, Harvard University, Boston.

Examples

```
# use projection vectors from the Andrews curve
sdo <- outlyingness(iris)
col <- gray(1-sdo/max(sdo))
andrews(iris, clr=col, ymax=NA)
# use 1000 random projection vectors
sdo <- outlyingness(iris, type=0, step=1000)
col <- gray(1-sdo/max(sdo))
andrews(iris, clr=col, ymax=NA)
# use 1000 random projection vectors with adjusted outlyingness
library("robustbase")
x <- numarray(iris)
x <- scale(x, center=apply(x, 2, min), scale=apply(x, 2, max)-apply(x, 2, min))
sdo <- adjOutlyingness(x, ndir=1000, only.outlyingness=TRUE)
col <- gray(1-sdo/max(sdo))
andrews(as.data.frame(x), clr=col, ymax=NA)
```

selectand

Selecting in Andrews curves

Description

Selecting object utility in Andrews curves

Usage

```
selectand(df, type = 1, step = 100, ncol = 0, from = 0, to = 1, col = 2)
```

Arguments

df	data frame.
type	type of curve.
step	smoothness of curves.
ncol	number of column in data frame for selection.
from	from value.
to	to value.
col	color of selected objects.

Details

Define which objects will be selected (colored) in Andrews curves.

Value

Nothing

Author(s)

Jaroslav Myslivec jaroslav.myslivec@upce.cz

Examples

```
data(iris)
andrews(iris, clr=5, ymax=3)
selectand(iris, ncol=1, from=5, to=5.5, col=1)
```

zzz

Comparison

Description

Creates and displays a temporary PDF file with different diagrams comparing `andrews` and `andrews0` plots.

Usage

```
zzz()
```

Value

nothing

Examples

```
if (interactive()) zzz()
```

Index

* **datasets**

banknote, 5

* **hplot**

andrews, 2

andrews0, 4

selectand, 9

andrews, 2

andrews(), 8, 9

andrews0, 4

banknote, 5

deftype, 6

graphics::plot.default(), 3

grDevices::rainbow(), 3

mclust::banknote, 5

normalize, 7

normalize(), 9

numarray, 8

outlyingness, 8

selectand, 9

zzz, 10