The rgr Package

October 3, 2007

Type Package

Title The GSC Applied Geochemistry EDA Package

Version 1.0.3

Author Robert G. Garrett

Maintainer Robert G. Garrett <garrett@NRCan.gc.ca>

Depends akima, MASS

Description Geological Survey of Canada (GSC) R functions for exploratory data analysis with applied geochemical data, with special application to the estimation of background ranges to support both environmental studies and mineral exploration.

License GPL

URL http://gsc.nrcan.gc.ca/dir/index_e.php?id=4961

R topics documented:

`anova1` *Analysis of Variance (ANOVA)*

Description

Undertakes a random effects model Analysis of Variance (ANOVA) on a set of duplicate measurements to determine if the analytical, or combined sampling and analytical, (within) variability is significantly smaller than the variability between the duplicates.

Usage

```
anova1(x1, x2, name = deparse(substitute(x1)), log = FALSE)
```

Arguments

x1	a column vector from a matrix or data frame, x1[1], ..., x1[n].
x2	another column vector from a matrix or data frame, x2[1], ..., x2[n]. x1,x2 must be of identical length, n, where x2 is a duplicate measurement of x1.
name	a title can be displayed with the results, e.g., "name = Duplicate measurements of Magnetic Susceptibility". If this field is undefined the character string for x1 is used as a default.
log	if a logarithmic transformation of the data is required to meet homogeneity of variance considerations (i.e. severe heteroscedasticity) set log = TRUE. This is also advisable if the range of the observations exceeds 1.5 orders of magnitude.

Details

In field geochemical surveys the combined sampling and analytical variability is more important than analytical variability alone. If the at site (within) variability is not significantly smaller than the between duplicate sites variability it cannot be stated that there are statistically significant spatial patterns in the data, and they are likely not suitable for mapping. This may not mean that the data cannot be used to recognize individuals with above threshold or action level observations. However, under these conditions there also may be above threshold or action level instances that the survey data have failed to detect (Garrett, 1983).

A random effects ANOVA is undertaken, the ANOVA table is displayed, together with estimates of the variance components, i.e. how much of the total variability is between and within the duplicate measurements, and the USGS mapping reliability measures of V and Vm (Miesch et al., 1976). Additionally, the data are investigated through a two-way model following the procedure of Bolviken and Sinding-Larsen (1973).

If the data are as a single concatenated vector from a matrix or data frame as x1[1], ..., x1[n] followed by x[n+1], ..., x[2n], or alternated as x[1] and x[2] being a pair through to x[2*i+1] and x[2*i+2], for the i in 1:n duplicate pairs use function anova2.

Note

The script does not follow a standard computation of Mean Squares, but is based on a procedure developed after Garrett (1969) for use in the field in the 1970s when pocket calculators first had mean and standard deviation functions.

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see ltdl.fix.df.

Duplicate pairs x1,x2 containing any NAs are omitted from the calculations.

If a log transformation is undertaken and any less than or equal to zero values occur in the data the function will halt with a warning to that effect.

Author(s)

Robert G. Garrett

References

Bolviken, B. and Sinding-Larsen, R., 1973. Total error and other criteria in the interpretation of stream sediment data. In Geochemical Exploration 1972, Institution of Mining and Metallurgy, London, pp. 285-295.

Garrett, R.G., 1969. The determination of sampling and analytical errors in exploration geochemistry. Economic Geology, 64(4):568-569.

Garrett, R.G., 1983. Sampling methodology. In Chapter 4 of Handbook of Exploration Geochemistry, Vol. 2, Statistics and Data Analysis in Geochemical Prospecting (Ed. R.J. Howarth), Elsevier, pp. 83-110.

Miesch, A.T. et al., 1976. Geochemical survey of Missouri - methods of sampling, analysis and statistical reduction of data. U.S. Geological Survey Professional Paper 954A, 39 p.

See Also

anova2, ltdl.fix.df

Examples

```
## Make test data ms.data1 available
data(ms.data1)
attach(ms.data1)

## Undertake an ANOVA for duplicate measurements on rock samples
anova1(MS.1, MS.2, log=TRUE,
        name = "Duplicate measurements of Magnetic Susceptibility")

## Detach test data ms.data1
detach(ms.data1)
```

---

anova2 *Analysis of Variance (ANOVA), Alternate Input*

Description

Function to prepare data stored in alternate forms from that expected by function anova1 for its use. For further details see 'x' in Arguments below .

Usage

```
anova2(x, name = deparse(substitute(x)), log = FALSE, ifalt = FALSE)
```

Arguments

x	a column vector from a matrix or data frame, x[1], ..., x[2*n]. The default is that the first n members of the vector are the first measurements and the second n members are the duplicate measurements. If the measurements alternate, i.e. duplicate pair 1 measurement 1 followed by measurement 2, etc., set ifalt = TRUE.

name	a title can be displayed with the results, e.g., `name = "Duplicate measurements of Magnetic Susceptibility"`. If this field is undefined the character string for `x` is used as a default.
log	if a logarithmic transformation of the data is required to meet homogeneity of variance considerations (i.e. severe heteroscedasticity) set `log = TRUE`. This is also advisable if the range of the observations exceeds 1.5 orders of magnitude.
ifalt	set `ifalt = TRUE` to accommodate alternating sets of paired observations.

Details

For further details see anova1.

If the data are as n duplicate pairs, x1 and x2, use function anova1.

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see ltdl.fix.df.

Author(s)

Robert G. Garrett

See Also

anova1, ltdl.fix.df

Examples

```
## Make test data ms.data2 available
data(ms.data2)
attach(ms.data2)

## Undertake an ANOVA for duplicate measurements on rock samples
anova2(MS, log=TRUE,
        name = "Duplicate measurements of Magnetic Susceptibility")

## Detach test data ms.data2
detach(ms.data2)

## Make test data ms.data3 available
data(ms.data3)
attach(ms.data3)

## Undertake an ANOVA for duplicate measurements on rock samples
anova2(MS, log=TRUE, ifalt = TRUE,
        name = "Duplicate measurements of Magnetic Susceptibility")

## Detach test data ms.data3
detach(ms.data3)
```

`bwplot` *Plot Vertical Box-and-Whisker Plots*

Description

Plots a series of vertical box-and-whisker plots where the individual boxplots represent the data subdivided by the value of some factor. Optionally the y-axis may be scaled logarithmically. A variety of other plot options are available, see Details and Note below.

Usage

```
bwplot(x, by, log = FALSE, wend = 0.05, notch = TRUE, xlab = "",
        ylab = deparse(substitute(x)), ylim = NULL, main = "",
        label = NULL, plot.order = NULL, xpos = NA, width,
        space = 0.25, las = 1, cex = 1, adj = 0.5, add = FALSE,
        ssll = 1, colr = 8, pch = 3, ...)
```

Arguments

`x`	name of the variable to be plotted.
`by`	the name of the factor variable to be used to subdivide the data. See Details below for when `by` is undefined.
`log`	if it is required to display the data with logarithmic (y-axis) scaling, set `log = TRUE`.
`wend`	the locations of the whisker-ends have to be defined. By default these are at the 5th and 95th percentiles of the data. Setting `wend = 0.02` plots the whisker ends at the 2nd and 98th percentiles.
`notch`	determines if the boxplots are to be "notched" such that the notches indicate the 95% confidence intervals for the medians. The default is to notch the boxplots, to suppress the notches set `notch = FALSE`. See Details below.
`xlab`	a title for the x-axis, by default none is provided.
`ylab`	a title for the y-axis. It is often desirable to replace the default y-axis title of the input variable name text string with a more informative title, e.g., `ylab = "Cu (mg/kg) in <2 mm C-horizon soil"`.
`ylim`	defines the limits of the y-axis if the default limits based on the range of the data are unsatisfactory. It can be used to ensure the y-axis scaling in multiple sets of boxplots are the same to facilitate visual comparison.
`main`	a main title may be added optionally above the display by setting `main`, e.g., `main = "Kola Project, 1995"`.
`label`	provides an alternate set of labels for the boxplots along the x-axis. By default the character strings defining the factors are used. Thus, `label = c("Alt1", "Alt2", "Alt3")`.
`plot.order`	provides an alternate order for the boxplots. Thus, `plot.order = c(2, 1, 3)` will plot the 2nd ordered factor in the 1st position, the 1st in the 2nd, and the 3rd in its 3rd ordered postion, see Details and Examples below.

xpos	the locations along the x-axis for the individual vertical boxplots to be plotted. By default this is set to NA, which causes default equally spaced positions to be used, i.e. boxplot 1 plots at value 1 on the x-axis, boxplot 2 at value 2, etc., up to boxplot "n" at value "n". See Details below for defining xpos.
width	the width of the boxes, by default this is set to the minimum distance between all adjacent boxplots times the value of space. With the default values of xpos this results in a minimum difference of 1, and with the default of space = 0.25 the width is computed as 0.25. To specify different widths for all boxplots use, for example, width = c(0.3). See Details below for changing individual boxplot widths.
space	the space between the individual boxplots, by default this is 0.25 x-axis units.
las	controls whether the x-axis labels are written parallel to the x-axis, the default las = 1, or are written down from the x-axis by setting las = 2. See also, Details below.
cex	controls the size of the font used for the factor labels plotted along the x-axis. By default this is 1, however, if the labels are long it is sometimes necessary to use a smaller font, for example cex = 0.8 results in a font 80% of normal size.
adj	controls the justification of the x-axis labels. By default they are centred, adj = 0.5, to left justify them if the labels are written downwards set adj = 0.
add	permits the user to plot additional boxplots into an existing display. It is recommended that this option is left as add = FALSE.
ssll	determines the minimum data subset size for which a subset will be plotted. By default this is set to 1, which leads to only a plus sign being plotted, as the subset size increases additional features of the boxplot are displayed. If ssll results in subset boxplots not being plotted, a gap is left and the factor label is still plotted on the x-axis.
colr	by default the boxes are infilled in grey, colr = 8 . If no infill is required, set colr = 0. See display.lty for the range of available colours.
pch	by default the plotting symbol for the subset maxima and minima are set to a plus, pch = 3, alternate plotting symbols may be chosen from those displayed by display.marks.
...	further arguments to be passed to methods.

Details

There are two ways to execute this function. Firstly by defining x and by, and secondly by combining the two variables with the split function. See the first two examples below. The split function can be useful if the factors to use in the boxplot are to be generated at run-time, see the last example below. Note that when the split construct is used instead of by the whole split statement will be displayed as the default y-axis title. Also note that when using by the subsets are listed in the order that the factors are encountered in the data, but when using split the subsets are listed alphabetically. In either case they can be re-ordered using plot.order, see Examples.

In a box-and-whisker plot there are two special cases. When wend = 0 the whiskers extend to the observed minima and maxima that are not plotted with the plus symbol. When wend = 0.25 no

whiskers or the data minima and maxima are plotted, only the medians and boxes representing the span of the middle 50% of the data are displayed.

The `width` option can be used to define different widths for the individual boxplots. For example, the widths could be scaled to be proportional to the subset population sizes as some function of the square root (const * sqrt(n)) or logarithm (const * log10(n)) of those sizes (n). The constant, const, would need to be chosen so that on average the width of the individual boxes would be approximately 0.25, see Example below. It may be desirable for cosmetic purposes to adjust the positions of the boxes along the x-axis, this can be achieved by specifying xpos.

Long subset (factor) names can lead to display problems, changing the `las` parameter from its default of `las = 1` which plots subset labels parallel to the axis to `las = 2`, to plot perpendicular to the axis, can help. It may also help to use `label` and split the character string into two lines, e.g., by changing the string `"Granodiorite"` that was supplied to replace the coded factor variable `GRDR` to `"Grano-`
`ndiorite"`. If this, or setting `las = 2`, causes a conflict with the x-axis title, if one is needed, the title can be moved down a line by using `xlab = "`
`nLithological Units"`. In both cases the
`n` forces the following text to be placed on the next lower line.

If there are more than 7 labels (subsets) and no alternate labels are provided `las` is set to 2, otherwise some labels may fail to be displayed.

The notches in the boxplots indicate the 95% confidence intervals for the medians and can extend beyond the upper and lower limits of the boxes indicating the middle 50% of the data when subset population sizes are small. The confidence intervals are estimated using the binomial theorem. It can be argued that for small populations a normal approximation would be better. However, it was decided to remain with a non-parametric estimate to be consistent with the use of non-parametric statistics in this display.

Note

This function is based on a script shared by Doug Nychka on S-News, April 28, 1992.

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see `ltdl.fix.df`.

Any `NA`s in the data vector are removed prior to preparing the boxplots.

Author(s)

Robert G. Garrett and Douglas W. Nychka

See Also

`cat2list`, `ltdl.fix.df`

Examples

```
## Make test data kola.c available
data(kola.c)
attach(kola.c)
```

```
## Display a simple box-and-whisker plot
bwplot(Cu, by = COUNTRY)
bwplot(split(Cu,COUNTRY))

## Display a more appropriately labelled and scaled box-and-whisker plot
bwplot(Cu, by = COUNTRY, log = TRUE, xlab = "Country",
        ylab = "Cu (mg/kg) in <2 mm C-horizon soil")

## Display a west-to-east re-ordered plot using the full country names
bwplot(split(Cu, COUNTRY), log = TRUE,
        ylab = "Cu (mg/kg) in <2 mm C-horizon soil",
        label = c("Finland", "Norway", "Russia"),
        plot.order = c(2, 1, 3))

## Detach test data kola.c
detach(kola.c)

## Make test data kola.o available, setting a -9999, indicating a
## missing pH measurement, to NA
data(kola.o)
kola.o.fixed <- ltdl.fix.df(kola.o, coded = -9999)
attach(kola.o.fixed)

## Display relationship between pH in one pH unit intervals and Cu in
## O-horizon (humus) soil, extending the whiskers to the 2nd and 98th
## percentiles
bwplot(split(Cu,trunc(pH+0.5)), log=TRUE, wend = 0.02,
        xlab = "O-horizon soil pH to the nearest pH unit",
        ylab = "Cu (mg/kg) in <2 mm O-horizon soil")

## As above, but demonstrating the use of variable box widths and the
## suppression of 95% confidence interval notches.  The box widths are
## computed as (Log10(n)+0.1)/5, the 0.1 is added as one subset has a
## population of 1.
table(trunc(pH+0.5))
bwplot(split(Cu,trunc(pH+0.5)), log=TRUE, wend = 0.02, notch = FALSE,
        xlab = "O-horizon soil pH to the nearest pH unit,\nbox widths proportional to Log(su
        ylab = "Cu (mg/kg) in <2 mm O-horizon soil",
        width = c(0.26, 0.58, 0.24, 0.02))

## Detach test data kola.o.fixed
detach(kola.o.fixed)
```

`bwplot.by.var` *Plot Vertical Box-and-Whisker Plots for Variables*

Description

Plots a series of vertical box-and-whisker plots where the individual boxplots represent the data subdivided by variables. Optionally the y-axis may be scaled logarithmically. A variety of other plot options are available, see Details and Note below.

Usage

```
bwplot.by.var(xmat, log = FALSE, wend = 0.05, notch = FALSE,
        xlab = "Measured Variables", ylab = "Reported Values",
        main = "", label = NULL, plot.order = NULL, xpos = NA,
        las = 1, cex = 1, adj = 0.5, colr = 8, pch = 3, ...)
```

Arguments

xmat	the data matrix or data frame containing the data.
log	if it is required to display the data with logarithmic (y-axis) scaling, set log = TRUE.
wend	the locations of the whisker-ends has to be defined. By default these are at the 5th and 95th percentiles of the data. Setting wend = 0.02 plots the whisker ends at the 2nd and 98th percentiles. See Details below.
notch	determines if the boxplots are to be "notched" such that the notches indicate the 95% confidence intervals for the medians. The default is not to notch the boxplots, to have notches set notch = TRUE.
xlab	a title for the x-axis, by default xlab = "Measured Variables".
ylab	a title for the y-axis, by default ylab = "Reported Values".
main	a main title may be added optionally above the display by setting main, e.g., main = "Kola Project, 1995".
label	provides an alternate set of labels for the boxplots along the x-axis. By default the character strings defining the factors (variables) are used. Thus, label = c("Alt1", "Alt2", "Alt3").
plot.order	provides an alternate order for the boxplots. By default the boxplots are plotted in alphabetical order of the factor variables. Thus, plot.order = c(2, 1, 3) will plot the 2nd alphabetically ordered factor in the 1st position, the 1st in the 2nd, and the 3rd in its alphabetically 3rd ordered postion.
xpos	the locations along the x-axis for the individual vertical boxplots to be plotted. By default this is set to NA, which causes default equally spaced positions to be used, i.e. boxplot 1 plots at value 1 on the x-axis, boxplot 2 at value 2, etc., up to boxplot "n" at value "n". See Details below for defining xpos.
las	controls whether the x-axis labels are written parallel to the x-axis, the default las = 1, or are written down from the x-axis by setting las = 2. See also, Details below.
cex	controls the size of the font used for the factor labels plotted along the x-axis. By default this is 1, however, if the labels are long it is sometimes necessary to use a smaller font, for example cex = 0.8 results in a font 80% of normal size.
adj	controls the justification of the x-axis labels. By default they are centred, adj = 0.5, to left justify them if the labels are written downwards set adj = 0.
colr	by default the boxes are infilled in grey, colr = 8 . If no infill is required, set colr = 0. See display.lty for the range of available colours.

pch	by default the plotting symbol for the subset maxima and minima are set to a plus, pch = 3, alternate plotting symbols may be chosen from those displayed by display.marks.
...	further arguments to be passed to methods.

Details

There are two ways to provide data to this function. Firstly, if all the variables in a data frame are to be displayed, and there are no factor variables, the data frame name can be entered for xmat. However, if there are factor variables, or only a subset of the variables are to be displayed, the data are entered via the cbind construct, see Examples below.

In a box-and-whisker plot there are two special cases. When wend = 0 the whiskers extend to the observed minima and maxima that are not plotted with the plus symbol. When wend = 0.25 no whiskers or the data minima and maxima are plotted, only the medians and boxes representing the span of the middle 50% of the data are displayed.

Long variable names can lead to display problems, changing the las parameter from its default of las = 1 which plots subset labels parallel to the axis to las = 2, to plot perpendicular to the axis, can help. It may also help to use label and split the character string into two lines, e.g., by changing the string "Specific Conductivity" that was supplied to replace the variable name SC to "Specific
nConductivity". If this, or setting las = 2, causes a conflict with the x-axis title, if one is needed, the title can be moved down a line by using xlab = "
nPhysical soil properties". In both cases the
n forces the following text to be placed on the next lower line.

If there are more than 7 labels (variables) and no alternate labels are provided las is set to 2, otherwise some variable names may fail to be displayed.

The notches in the boxplots indicate the 95% confidence intervals for the medians and can extend beyond the upper and lower limits of the boxes indicating the middle 50% of the data when subset population sizes are small. The confidence intervals are estimated using the binomial theorem. It can be argued that for small populations a normal approximation would be better. However, it was decided to remain with a non-parametric estimate to be consistent with the use of non-parametric statistics in this display.

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see ltdl.fix.df.

Any NAs in the data vectors are removed prior to preparing the boxplots.

Author(s)

Robert G. Garrett

See Also

bwplot, var2fact, ltdl.fix.df

Examples

```
## Make test data kola.c available
data(kola.c)
attach(kola.c)

## Display a simple box-and-whisker plot for measured variables
bwplot.by.var(cbind(Co,Cu,Ni))

## Display a more appropriately labelled and scaled box-and-whisker plot
bwplot.by.var(cbind(Co,Cu,Ni), log = TRUE,
        ylab = "Levels (mg/kg) in < 2 mm C-horizon soil")

## Detach test data kola.c
detach(kola.c)

## Make test data ms.data1 available
data(ms.data1)

## Display variables in a data frame extending the whiskers to the
## 2nd and 98th percentiles of the data
bwplot.by.var(ms.data1, log=TRUE, wend = 0.02)
```

---

bxplot *Plot a Horizontal Boxplot or Box-and-Whisker Plot*

Description

Plots a single horizontal boxplot as part of the multi-panel display provided by function shape, the default is a Tukey boxplot, alternately a box-and-whisker plot may be displayed.

Usage

```
bxplot(xx, xlab = deparse(substitute(xx)), log = FALSE, ifbw = FALSE,
        wend = 0.05, xlim = NULL, main = " ", colr = 8)
```

Arguments

xx	name of the variable to be plotted.
xlab	a title for the x-axis. It is often desirable to replace the default x-axis title of the input variable name text string with a more informative title, e.g., xlab = "Cu (mg/kg) in <2 mm O-horizon soil".
log	if it is required to display the data with logarithmic (x-axis) scaling, set log = TRUE.
ifbw	the default is to plot a horizontal Tukey boxplot, if a box-and-whisker plot is required set ifbw = TRUE.

wend — if ifbw = TRUE the locations of the whisker-ends have to be defined. By default these are at the 5th and 95th percentiles of the data. Setting wend = 0.02 plots the whisker ends at the 2nd and 98th percentiles.

xlim — when used in the shape function, xlim is determined by gx.hist and used to ensure all four panels in shape have the same x-axis scaling. However when used stand-alone the limits may be user-defined by setting xlim, see Note below.

main — when used stand-alone a title may be added optionally above the plot by setting main, e.g., main = "Kola Project, 1995".

colr — by default the histogram and box are infilled in grey, colr = 8 . If no infill is required, set colr = 0. See display.lty for the range of available colours.

Details

The function can be used stand-alone, but as Tukey boxplots and box-and-whisker plots are usually used to compare the distributions of data subsets the functions tbplot (Tukey boxplots) and bwplot (box-and-whisker plots) are required for that purpose.

When the boxplot is displayed on a logarithmically scaled x-axis, the data are log transformed prior to the computation of the positions of the fences used in the Tukey boxplot to identify near and far outliers, plotted as plusses and circles, respectively.

In a box-and-whisker plot there are two special cases. When wend = 0 the whiskers extend to the observed minima and maxima that are not plotted with the plus symbol. When wend = 0.25 no whiskers or the data minimum and maximum are plotted, only the median and box representing the span of the middle 50 percent of the data are displayed.

Note

Any less than detection limit values represented by negative values, or zeros or numeric codes representing blanks in the data vector, must be removed prior to executing this function, see ltdl.fix.df.

Any NAs in the data vector are removed prior to displaying the plot.

If the default selection for xlim is inappropriate it can be set, e.g., xlim = c(0, 200) or c(2, 200). If the defined limits lie within the observed data range a truncated plot will be displayed. If this occurs the number of data points omitted is displayed below the total number of observations.

If it is desired to prepare a display of data falling within a defined part of the actual data range, then either a data subset can be prepared externally using the appropriate R syntax, or xx may be defined in the function call as, for example, Cu[Cu < some.value] which would remove the influence of one or more outliers having values greater than some.value. In this case the number of data values displayed will be the number that are <some.value.

Author(s)

Robert G. Garrett

References

Garrett, R.G., 1988. IDEAS - An Interactive Computer Graphics Tool to Assist the Exploration Geochemist. In Current Research Part F, Geological Survey of Canada Paper 88-1F, pp. 1-13 for a description of box-and-whisker plots.

See Also

shape, display.lty, ltdl.fix.df, remove.na

Examples

```
## Make test data available
data(kola.o)
attach(kola.o)

## Display a simple boxplot
bxplot(Cu)

## Display a more appropriately labelled and scaled boxplot
bxplot(Cu, xlab = "Cu (mg/kg) in <2 mm O-horizon soil", log = TRUE)

## Display a box-and-whisker plot with whiskers ending at the 2nd and
## 98th percentiles
bxplot(Cu, xlab = "Cu (mg/kg) in <2 mm O-horizon soil", ifbw = TRUE,
        wend = 0.02, log = TRUE)

## Detach test data
detach(kola.o)
```

---

caplot *Prepare a Concentration-Area (C-A) Plot*

Description

Displays a concentration-area (C-A) plot to assess whether the data are spatially multi-fractal (Cheng et al., 1994; Cheng and Agterberg, 1995) as a part of a four panel display. This procedure is useful for determining if multiple populations that are spatially dependent are present in a data set. It can be used to determine the practical limits, upper or lower bounds, of the influence of the biogeochemical processes behind the spatial distribution of the data. Optionally the data may be logarithmically transformed prior to interpolation, the points may be 'jittered' (see Arguments below), the size of the interpolated grid may be modified, and alternate colour schemes can be chosen for display of the interpolated data.

Usage

```
caplot(x, y, z, zname = deparse(substitute(z)),
        caname = deparse(substitute(z)), log =  FALSE, ifjit = FALSE,
        ifrev = FALSE, ngrid = 100, colr = topo.colors(16),
        xcoord = "Easting", ycoord = "Northing")
```

Arguments

x	name of the x-axis spatial coordinate, the eastings.
y	name of the y-axis spatial coordinate, the northings.
z	name of the variable to be processed and plotted.
zname	a title for the x-axes of the CPP (Cumulative Normal Percentage Probability) and C-A plot panels. It is often desirable to replace the default x-axis titles of the input variable name text string with a more informative title, e.g., `zname = "Cu (mg/kg) in <2 mm O-horizon soil"`.
caname	a title for the image of the interpolated data. It is often desirable to replace the default title of the input variable name text string with a more informative title, e.g., `caname = "Kola Project, 1995 nCu (mg/kg) in <2 mm O-horizon soil"`. For no title, set `caname = ""`.
log	if it is required to undertake the C-A plot interpolation following a logarithmic data transformation, set `log = TRUE`. This also results in the accompanying probability (CPP) plots being logarithmically scaled (x-axes).
ifjit	if there is a possibility that the data set contains multiple measurements at an identical spatial (x,y) location set `ifjit = TRUE`. The presence of multiple data at an identical location will cause the Akima (1996) interpolation function to fail.
ifrev	by default the empirical C-A function is plotted from highest value to lowest, `ifrev = FALSE`. As the C-A plot is a log-log display this provides greater detail for the highest values. The direction of accumulation can be key in detecting multi-fractal patterns, it is usually informative to also prepare a plot with `ifrev = TRUE`, i.e. accumaulation from lowest to highest values. To see a dramatic example of this, run the Examples below.
ngrid	by default `ngrid = 100`, this results in the data being interpolated into a 100 x 100 grid that extends between the data set's spatial extremes determined for the (x,y) spatial coordinates for the data. See Details below.
colr	by default the topo.colors(16) pallette is used to render the interpolated grid as an image. For alternative palettes see `colors`, and see Details below.
xcoord	a title for the x-axis, defaults to "Easting".
ycoord	a title for the y-axis, defaults to "Northing".

Details

The function creates a four panel display. The percentage cumulative probabity (CPP) plot of the data in the upper left, and the CPP plot of the interpolated data to be used in the C-A plot in the upper right. The lower left panel contains an image of the interpolated data, and the lower right the C-A plot.

Akima's (1978, 1996) interpolation function is used to obtain a linear interpolation between the spatial data values. If the data are positively skewed the use of a logarithmic data transformation, `log = TRUE`, is highly recommended. Following generation of the interpolated grid and prior to futher processing the interpolated grid values are clipped by the convex-hull of the spatial locations, therefore there is effectively no interpolation beyond the spatial extent, convex hull, of the data.

The use of the `topo.colors(16)` palette to display the image of the interpolated values leads to low values being plotted in blue, and as the interpolated values increase they take on green, yellow and orange colors. For a grey-scale display for black-and-white use set `colr = grey(0:8/8)`. This leads to lowest interpolated values being plotted in black and the highest in white, using `colr = grey(8:0/8)` reverses this, with the lowest values being plotted in white and the highest in black. In either case, if the values plotted in white occur at the study area boundary, i.e. at the convex hull, the difference between no data and white cannot be discerned.

For preparation of the C-A plot the ordered vector of interpolated values is used as a surrogate for the measurement of area greater than, or less than, a stated interpolated value. The cumulative percentage count of the interpolated values being plotted on the y-axis of the C-A plot. As noted above, it is both informative and important to display the C-A plot accumulated both upwards and downwards.

Note

This wrapper function was developed from a S-Plus function to prepare C-A plots using Akima's (1978, 1996) interpolation procedure written by Graeme Bonham-Carter, Geological Survey of Canada, in April 2004.

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see `ltdl.fix.df`.

Any sites with `NA`s in the (x,y,z) data vector are removed prior to spatial interpolation and preparation of the C-A plot.

In some R installations the generation of multi-panel displays and the use of function eqscplot from package MASS causes warning messages related to graphics parameters to be displayed on the current device. These may be suppressed by entering `options(warn = -1)` on the R command line, or that line may be included in a 'first' function prepared by the user that loads the rgr package, etc.

Author(s)

Robert G. Garrett and Graham F. Bonham-Carter

References

Akima, H. (1978). A Method of Bivariate Interpolation and Smooth Surface Fitting for Irregularly Distributed Data Points. ACM Transactions on Mathematical Software *4*, 148-164.

Akima, H. (1996). Algorithm 761: scattered-data surface fitting that has the accuracy of a cubic polynomial. ACM Transactions on Mathematical Software *22*, 362-371.

Cheng, Q. and Agterberg, F.P., 1995. Multifractal modeling and spatial point processes. Mathematical Geology 27(7):831-845.

Cheng, Q., Agterberg, F.P. and Ballantyne, S.B., 1994. The separation of geochemical anomalies from background by fractal methods. Journal of Geochemical Exploration 51(2):109-130.

See Also

`cnpplt`, `interp`, `colors`, `ltdl.fix.df`

Examples

```
## Make test data available
data(kola.o)
attach(kola.o)

## A default (uninformative) C-A plot
caplot(UTME/1000, UTMN/1000, Cu)

## Plot a more appropriately scaled (log transformed data) and
## titled display
caplot(UTME/1000, UTMN/1000, Cu, log = TRUE,
        zname = "Cu (mg/kg) in <2 mm O-horizon soil",
        caname = "Kola Project, 1995\nCu (mg/kg) in <2 mm O-horizon soil")

## Plot as above but with the C-A plot accumulation reversed
caplot(UTME/1000, UTMN/1000, Cu, log = TRUE, ifrev = TRUE,
        zname = "Cu (mg/kg) in <2 mm O-horizon soil",
        caname = "Kola Project, 1995\nCu (mg/kg) in <2 mm O-horizon soil")

## Detach test data
detach(kola.o)
```

---

cat2list *Divides Data into Subsets by Factor*

Description

Converts data into a list form where data are grouped together by factor. Achieves the same objective as the base function `split`.

Usage

```
cat2list(x, a)
```

Arguments

`x`	name of the data variable to be processed.
`a`	name of the factor variable by which the data are to be split.

Value

`data`	a list containing factors as columns and the values for those factors as rows. The order of the resulting groups, subsets, is the order in which the factor variable names were encountered in parameter 'a' passed to the function.

Note

This function is called by functions tbplot and bwplot to prepare Tukey boxplots and box-and-whisker plots, respectively. It is an integral part of the script shared by Doug Nychka on S-News, April 28, 1992. As such it may pre-date the time that split was added to the S-Plus library.

If by is undefined in the calling functions, tbplot and bwplot, the same result may be achieved by using the split(x, a) construct instead of stating x as the variable to be displayed as box-plots. In which case the data are grouped, subsetted, in alphabetical order of factor variable names.

Author(s)

Douglas W. Nychka

---

cnpplt *Cumulative Normal Percentage Probability (CPP) Plot*

Description

Displays a cumulative normal percentage probability (CPP) plot, equivalent to a Q-Q plot, as has been traditionally used by physical scientists and engineers.

Usage

```
cnpplt(xx, xlab = deparse(substitute(xx)),
        ylab = "% Cumulative Probability", log = FALSE, xlim = NULL,
        main = " ", pch = 3, cex.axis = 1, ifqs = FALSE, ifshape = FALSE)
```

Arguments

xx	name of the variable to be plotted.
xlab	a title for the x-axis. It is often desirable to replace the default x-axis title of the input variable text string with a more informative title, e.g., xlab ="Cu (mg/kg) in <2 mm O-horizon soil".
ylab	a title for the y-axis, defaults to "% Cumulative Probability".
log	if it is required to display the data with logarithmic (x-axis) scaling, set log = TRUE.
xlim	when used in the shape function, xlim is determined by function gx.hist and used to ensure all four panels in shape have the same x-axis scaling. However when used stand-alone the limits may be user-defined by setting xlim, see Note below.
main	when used stand-alone a title may be added optionally above the plot by setting main, e.g., main = "Kola Project, 1995".
pch	by default the plotting symbol is set to a plus, pch = 3, alternate plotting symbols may be chosen from those displayed by display.marks.
cex.axis	if overplotting occurs in the y-axis labelling the size of the y-axis labels may be reduced by setting cex.axis to a number smaller than 1, e.g., 0.8.

ifqs — setting ifqs = TRUE results in horizontal and vertical dotted lines being plotted at the three central quartiles and their values, respectively.

ifshape — when used with function shape or caplot to plot into a panel set ifshape = TRUE to ensure only essential probability scale axis labels are displayed to avoid overplotting on the reduced size panel plot.

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see ltdl.fix.df.

Any NAs in the data vector are removed prior to displaying the plot.

If the default selection for xlim is inappropriate it can be set, e.g., xlim = c(0, 200) or c(2, 200). If the defined limits lie within the observed data range a truncated plot will be displayed. If this occurs the number of data points omitted is displayed below the total number of observations.

If it is desired to prepare a display of data falling within a defined part of the actual data range, then either a data subset can be prepared externally using the appropriate R syntax, or xx may be defined in the function call as, for example, Cu[Cu < some.value] which would remove the influence of one or more outliers having values greater than some.value. In this case the number of data values displayed will be the number that are <some.value.

Author(s)

Robert G. Garrett

See Also

display.marks, ltdl.fix.df, remove.na

Examples

```
## Make test data available
data(kola.o)
attach(kola.o)

## A stand-alone cumulative normal percentage probability plot
cnpplt(Cu)

## A more appropriately labelled and scaled cumulative normal percentage
## probability plot using a cross/x rather than a plus
cnpplt(Cu, xlab = "Cu (mg/kg) in <2 mm O-horizon soil", log = TRUE,
        pch = 4)

## Detach test data
detach(kola.o)
```

`cutter` *Function to Identify in Which Interval a Value Falls*

Description

Function to identify in which interval of a set of cut points, cuts, a value x falls within or beyond. The number of intervals is equal to the number of cut points plus 1. Values of x have to exceed the value of the cut point to be allocated to the higher interval.

Usage

```
cutter(x, cuts)
```

Arguments

`x`	name of the vector to be processed.
`cuts`	the vector of cut points.

Value

`xi`	a vector of the same length as x containing an integer between 1 and the number of cut points plus 1 indicating in which interval each value of x fell. Values <cut[1] have `xi` set to 1, and values >cut[highest] have `xi` set to the number of cut points plus 1.

Author(s)

Robert G. Garrett

Examples

```
## Make test data available
data(kola.c)
attach(kola.c)

## Cut the data into quartiles
xi <- cutter(Cu, quantile(Cu, probs = c(0.25, 0.5, 0.75)))

## Detach test data
detach(kola.c)
```

dftest *Check for the Existence of a Data Frame*

Description

A utility function to determine if a data frame is attached, or exists in the working directory. If the data frame exists the names of the variables are displayed.

Usage

```
dftest(dfname, x = NULL)
```

Arguments

dfname	name of a data frame.
x	setting x to an expected variable name in the data frame results in the variable's length, i.e., number of observations, being displayed if it is present.

Details

Based on a function shared on S-News.

Author(s)

Unkown

See Also

search, ls

Examples

```
## Make test data available
data(kola.o)

## Check that the data frame kola.o is available
dftest(kola.o)

## Check that kola.o is available and that the variable Cu is present
dftest(kola.o, x = Cu)
```

display.alts *Display Alt Codes*

Description

A utility function to display a selection of the Alt codes available from a PC keyboard.

Usage

```
display.alts()
```

Details

Based on a function shared on S-News.

Author(s)

Unknown

See Also

display.ascii.o, display.ascii.d

Examples

```
display.alts()
```

display.ascii.d *Display the ASCII Character Set by Decimal Number*

Description

A utility function to display the decimal numbers corresponding to the Windows Latin 1 Font.

Usage

```
display.ascii.d()
```

Details

Based on a function shared on S-News for octal numbers.

Author(s)

Robert G. Garrett

See Also

display.ascii.o, display.alts

Examples

```
display.ascii.d()
```

---

display.ascii.o *Display the Windows Latin 1 Font Octal Table*

Description

A utility function to display the octal numbers corresponding to the Windows Latin 1 Font.

Usage

```
display.ascii.o()
```

Details

The ASCII octal 'escape codes' are used to insert special characters in text strings for axis labelling, and titles etc., in graphical displays. For example the escape string
`265` results in the Greek letter mu, being displayed.

Based on a function shared on S-News.

Author(s)

Unknown

See Also

display.ascii.d, display.alts

Examples

```
display.ascii.o()
```

display.lty *Display Available Line Styles and Colour Codes*

Description

Displays the line styles and colours corresponding to `lty = 1` to `9` and `colr = 1` to `9`, respectively.

Usage

```
display.lty()
```

Details

Several rgr functions that plot boxes or polygons may have their default infill colour of grey, `colr = 8`, changed to an alternate colour, `colr = 1` to `7` or `9`. For no infill colour, set `colr` = 0.

Author(s)

Robert G. Garrett

display.marks *Display Available Plotting Marks*

Description

Displays the available plotting marks. By default the rgr functions use a plus sign, `pch = 3`, as the plotting symbol, alternate plotting marks may be selected from this display. For example, `pch = 1` results in an open circle, and `pch = 4` results in a cross/x.

Usage

```
display.marks()
```

Note

Function to display `pch` codes based on a script originally shared on S-News by Bill Venables, 1996/07, and modified by Shawn Boles, 1996/07/31.

Author(s)

Robert G. Garrett

display.rainbow *Display the Colours of the Rainbow(36) Pallette*

Description

Displays the available colours in the rainbow(36) pallette to support the selection of alternate colour schemes.

Usage

```
display.rainbow()
```

Author(s)

Robert G. Garrett

edamap *Plot a Map of Data using Proportional Symbols*

Description

Displays a simple map where the data are represented by open circles whose diameters are proportional to the value of the data at their spatial locations. The rate of change of symbol diameter with value and the absolute size of the symbols are defined by the user.

Usage

```
edamap(x, y, zz, p = 1, sfact = 1, zmin = NA, zmax = NA,
        xlab = "Easting", ylab = "Northing",
        zlab = deparse(substitute(zz)), main = "", tol = 0.04)
```

Arguments

x	name of the x-axis spatial coordinate, the eastings.
y	name of the y-axis spatial coordinate, the northings.
zz	name of the variable to be plotted.
p	a parameter that controls the rate of change of symbol diameter with changing value. A default of `p = 1` is provided that results in a linear rate of change. See Details below.
sfact	controls the absolute size of the plotted symbols, by default `sfact = 1`. Increasing sfact results in larger symbols.
zmin	a value below which all symbols will be plotted at the same minumum size. By default `zmin = NA` which results in the minimum value of the variable defining the minimum symbol size. See Details below.

zmax	a value above which all symbols will be plotted at the same maximum size. By default `zmax = NA` which results in the maximum value of the variable defining the maximum symbol size. See Details below.
xlab	a title for the x-axis, defaults to "Easting".
ylab	a title for the y-axis, defaults to "Northing".
zlab	by default, `zlab = deparse(substitute(zz))`, a map title is generated by appending the input variable name text string to "Proportional Symbol Map for". Alternative titles may be generated, see Details below.
main	an alternative map title, see Details below.
tol	a parameter used to ensure the area included within the neatline around the map is larger than the distribution of the points so that the plotted symbols fall within the neatline. By default `tol = 0.04`, if more clearance is required increase the value of `tol`.

Details

The symbol diameter is computed as a function of the value z to be plotted:
`diameter = dmin + (dmax - dmin) * {(z - zmin)/(zmax - zmin)}^p`
where dmin and dmax are defined as 0.1 and 1 units, so the symbol diameters range over an order of magnitude (and symbol areas over two); zmin and zmax are the observed range of the data, or the range over which the user wants the diameters to be computed; and p is a power defined by the user. The value of (z - zmin)/(zmax - zmin) is the value of z normalized, 0 - 1, to the range over which the symbol diameters are to be computed. After being raised to the power p, which will result in number in the range 0 to 1, this value is multiplied by the permissable range of diameters and added to the minimum diameter. This results in a diameter between 0.1 and 1 units that is proportional to the value of z.

A p value of `1` results in a linear rate of change. Values of p less than unity lead to a rapid intial rate of change with increasing value of z which is often suitable for displaying positively skewed data sets, see the example below. In contrast, values of p greater than unity result in an initial slow rate of change with increasing value of z which is often suitable for displaying negatively skewed data sets. Experimentation is usually necessary to obtain a satisfactory visual effect. See `syms.pfunc` for a graphic demonstrating the effect of varying the p parameter.

The user may choose to transform the variable to be plotted prior to determining symbol size etc., e.g. `log10(zz)`, to generate a logarithmic rate of symbol size change. See Example below.

If zmin or zmax are defined this has the effect of setting a minimum or maximum value of z, respectively, beyond which changes in the value of z do not result in changes in symbol diameter. This can be useful in limiting the effect of one or a few extreme outliers while still plotting them, they simply plot at the minimum or maximum symbol size and are not involved in the calculation of the range of z over which the diameter varies.

If `zlab` and `main` are undefined a default a map title is generated by appending the input variable name text string to "Proportional Symbol Map for ". If no map title is required set `zlab = ""`, and if some user defined map title is required it should be defined in `main`, e.g. `main = "Map Title Text"`.

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see `ltdl.fix.df`.

Any `NA`s in the data vector are removed prior to displaying the plot.

In some R installations the generation of multi-panel displays and the use of function eqscplot from package MASS causes warning messages related to graphics parameters to be displayed on the current device. These may be suppressed by entering `options(warn = -1)` on the R command line, or that line may be included in a 'first' function prepared by the user that loads the rgr package, etc.

Author(s)

Robert G. Garrett

See Also

`syms`, `syms.pfunc`, `ltdl.fix.df`, `remove.na`

Examples

```
## Make test data available
data(kola.o)
attach(kola.o)

## Plot a default symbol map
edamap(UTME, UTMN, Cu)

## Plot a differently symbol scaled and more appropriately labelled
## map
edamap(UTME/1000, UTMN/1000, Cu, p = 0.3, sfact = 2.0,
        xlab = "Kola Project UTM Eastings (km)",
        ylab = "Kola Project UTM Northings (km)" )

## Plot a map as above but where outliers above a value of 1000 are
## displayed with the same symbol
edamap(UTME/1000, UTMN/1000, Cu, p = 0.3, sfact = 2.0, zmax = 1000,
        xlab = "Kola Project UTM Eastings (km)",
        ylab = "Kola Project UTM Northings (km)" )

## plot a map where the symbols are logarithmically scaled
edamap(UTME/1000, UTMN/1000, log10(Cu), sfact = 2.0,
        xlab = "Kola Project UTM Eastings (km)",
        ylab = "Kola Project UTM Northings (km)" )

## Detach test data
detach(kola.o)
```

edamap7 *Plot a Symbol Map of Data Based on the Tukey Boxplot*

Description

Displays a simple map where the data are represented at their spatial locations by symbols indicating the value of the data in the context of a Tukey boxplot. Tukey boxplots divide data into 7 groups, the middle 50%, and three lower and higher groupings, see Details below. The computation of the fences used to subdivide the data may be carried out following a logarithmic transformation of the data. The colours of the symbols may be optionally changed.

Usage

```
edamap7(x, y, zz, sfact = 1, xlab = "Easting", ylab = "Northing",
        zlab = deparse(substitute(zz)), main = "", log = FALSE,
        ifgrey = FALSE, symcolr = NULL, tol = 0.04)
```

Arguments

x	name of the x-axis spatial coordinate, the eastings.
y	name of the y-axis spatial coordinate, the northings.
zz	name of the variable to be plotted.
sfact	controls the absolute size of the plotted symbols, by default `sfact = 1`. Increasing sfact results in larger symbols.
xlab	a title for the x-axis, defaults to "Easting".
ylab	a title for the y-axis, defaults to "Northing".
zlab	by default, 'zlab = deparse(substitute(zz))', a map title is generated by appending the input variable name text string to "EDA Tukey Boxplot Map for". Alternative titles may be generated, see Details below.
main	an alternative map title, see Details below.
log	if it is required to undertake the Tukey Boxplot computations after a logarithmic data transform, set `log = TRUE`.
ifgrey	set `ifgrey = TRUE` if a grey-scale map is required, see Details below.
symcolr	the default is a colour map and default colours are provided, deeper blues for lower values, green for the middle 50% of the data, and oranges and reds for higher values. A set of alternate symbol colours can be provided by defning `symcolr`, see Details below.
tol	a parameter used to ensure the area included within the neatline around the map is larger than the distribution of the points so that the plotted symbols fall within the neatline. By default `tol = 0.04`, if more clearance is required increase the value of `tol`.

Details

Tukey boxplots divide data into 7 groups, the middle 50%, and three lower and high groupings: within the whisker, near outliers and far outliers, respectively. Symbols for values below the first quartile (Q1) are plotted as increasingly larger circles, while symbols for values above the third quartile are plotted as increasingly larger squares. For the higher groupings, the whisker contains values `>Q3 and <(Q3 + 1.5 * HW)`, where `HW = (Q3 - Q1)`, the interquartile range; near outliers lie between `(Q3 + 1.5 * HW)` and `(Q3 + 3 * HW)`; and far outliers have values `>(Q3 + 3 * HW)`. For the lower groupings the group boundaries, fences, fall similarly spaced below Q1. The computation of the fences used to subdivide the data may be carried out following a logarithmic transformation of the data.

A summary table of the values of the symbol intervals, the number of values plotting as each symbol, and symbol shapes, sizes and colours is displayed on the current device.

If `zlab` and `main` are undefined a default a map title is generated by appending the input variable name text string to "EDA Tukey Boxplot Based Map for". If no map title is required set `zlab = ""`, and if some user defined map title is required it should be defined in `main`, e.g. `main = "Map Title Text"`.

If the grey-scale option is chosen the symbols are plotted 100% black for the far outliers, 85% black for the near outliers, 70% black for values within the whiskers, and 60% black for values falling within the middle 50% of the data.

The default colours, `symcolr = c(25, 22, 20, 13, 6, 4, 1)`, are selected from the `rainbow(36)` pallette, and alternate colour schemes need to be selected from the same palette. See `display.rainbow` for the available colours. It is essential that 7 colours be provided, e.g., `symcolr = c(27, 24, 22, 12, 5, 3, 36)`, if exactly 7 are not provided the default colours will be displayed.

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see `ltdl.fix.df`.

Any `NA`s in the data vector are removed prior to displaying the plot.

In some R installations the generation of multi-panel displays and the use of function eqscplot from package MASS causes warning messages related to graphics parameters to be displayed on the current device. These may be suppressed by entering `options(warn = -1)` on the R command line, or that line may be included in a 'first' function prepared by the user that loads the rgr package, etc.

Author(s)

Robert G. Garrett

See Also

`display.rainbow`, `ltdl.fix.df`, `remove.na`

Examples

```
## Make test data available
data(kola.o)
attach(kola.o)

## Plot a default symbol map
edamap7(UTME, UTMN, Cu)

## Plot a more appropriately scaled (log transformed data) and
## labelled map
edamap7(UTME/1000, UTMN/1000, Cu, log = TRUE,
        xlab = "Kola Project UTM Eastings (km)",
        ylab = "Kola Project UTM Northings (km)")

## Plot a grey-scale equivalent of the above map
edamap7(UTME/1000, UTMN/1000, Cu, log = TRUE, ifgrey = TRUE,
        xlab = "Kola Project UTM Eastings (km)",
        ylab = "Kola Project UTM Northings (km)")

## Plot the same map with an alternate colour scheme
edamap7(UTME/1000, UTMN/1000, Cu, log = TRUE,
        xlab = "Kola Project UTM Eastings (km)",
        ylab = "Kola Project UTM Northings (km)",
        symcolr = c(27, 24, 22, 12, 5, 3, 36))

## Detach test data
detach(kola.o)
```

---

edamap8 *Plot a Symbol Map of Data Based on their Percentiles*

Description

Displays a simple map where the data are represented at their spatial locations by symbols indicating within which group defined by the data's 2nd, 5th, 25th, 50th, 75th, 95th and 98th percentiles plotted a data value falls. The colours of the symbols may be optionally changed.

Usage

```
edamap8(x, y, zz, sfact = 1, xlab = "Easting", ylab = "Northing",
        zlab = deparse(substitute(zz)), main = "", ifgrey = FALSE,
        symcolr = NULL, tol = 0.04)
```

Arguments

x	name of the x-axis spatial coordinate, the eastings.
y	name of the y-axis spatial coordinate, the northings.
zz	name of the variable to be plotted.

sfact	controls the absolute size of the plotted symbols, by default sfact = 1. Increasing sfact results in larger symbols.
xlab	a title for the x-axis, defaults to "Easting".
ylab	a title for the y-axis, defaults to "Northing".
zlab	by default, 'zlab = deparse(substitute(zz))', a map title is generated by appending the input variable name text string to "EDA Percentile Based Map for". Alternative titles may be generated, see Details below.
main	an alternative map title, see Details below.
ifgrey	set ifgrey = TRUE if a grey-scale map is required, see Details below.
symcolr	the default is a colour map and default colours are provided, deeper blues for lower values, green for the middle 50% of the data, and oranges and reds for higher values. A set of alternate symbol colours can be provided by defining symcol, see Details below.
tol	a parameter used to ensure the area included within the neatline around the map is larger than the distribution of the points so that the plotted symbols fall within the neatline. By default tol = 0.04, if more clearance is required increase the value of tol.

Details

The selected percentiles, 2nd, 5th, 25th, 50th, 75th, 95th and 98th, divide the data into 8 groups. Values below the median are represented by increasingly larger deeper blue circles below the 25th percentile (Q1), and values above the 75th percentile (Q3) by increasingly larger orange and red squares. The mid 50% of the data are represented by green symbols, circles for the median (Q2) to Q1, and squares for the median (Q2) to Q3.

A summary table of the values of the symbol intervals, the number of values plotting as each symbol, and symbol shapes, sizes and colours is displayed on the current device.

If zlab and main are undefined a default a map title is generated by appending the input variable name text string to "EDA Percentile Based Map for". If no map title is required set zlab = "", and if some user defined map title is required it should be defined in main, e.g. main = "Map Title Text".

If the grey-scale option is chosen the symbols are plotted 100% black for the far outliers, 85% black for the near outliers, 70% black for values within the whiskers, and 60% black for values falling within the middle 50% of the data.

The default colours, symcolr = c(25, 22, 20, 13, 13, 6, 4, 1), are selected from the rainbow(36) pallette, and alternate colour schemes need to be selected from the same palette. See display.rainbow for the available colours. It is essential that 8 colours be provided, e.g., symcolr = c(27, 24, 22, 12, 12, 5, 3, 36), if exactly 8 are not provided the default colours will be displayed.

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see ltdl.fix.df.

Any NAs in the data vector are removed prior to displaying the plot.

In some R installations the generation of multi-panel displays and the use of function eqscplot from package MASS causes warning messages related to graphics parameters to be displayed on the current device. These may be suppressed by entering `options(warn = -1)` on the R command line, or that line may be included in a 'first' function prepared by the user that loads the rgr package, etc.

Author(s)

Robert G. Garrett

See Also

display.rainbow, ltdl.fix.df, remove.na

Examples

```
## Make test data available
data(kola.o)
attach(kola.o)

## Plot a default symbol map
edamap8(UTME, UTMN, Cu)

## Plot a more appropriately labelled map
edamap8(UTME/1000, UTMN/1000, Cu,
        xlab = "Kola Project UTM Eastings (km)",
        ylab = "Kola Project UTM Northings (km)")

## Plot a grey-scale equivalent of the above map
edamap8(UTME/1000, UTMN/1000, Cu, ifgrey = TRUE,
        xlab = "Kola Project UTM Eastings (km)",
        ylab = "Kola Project UTM Northings (km)")

## Plot the same map with an alternate colour scheme
edamap8(UTME/1000, UTMN/1000, Cu,
        xlab = "Kola Project UTM Eastings (km)",
        ylab = "Kola Project UTM Northings (km)",
        symcolr = c(27, 24, 22, 12,12, 5, 3, 36))

## Detach test data
detach(kola.o)
```

---

fences *Generate and Display Fence Values*

Description

Function to generate fence values to support the selection of the upper and lower bounds of background variability, i.e. threshold(s) or action levels, when an obvious graphical solution is not visually recognizable.

Usage

```
fences(xx, display = TRUE)
```

Arguments

`xx`	name of the variable to be processed.
`display`	the default is to display the tabular output on the current device, i.e. `display = TRUE`. However, when the function is used in conjunction with `fences.summary` then `display = FALSE` in order to suppress output to the current device as it will be saved to a text file for subsequent use/editing and reference.

Details

The fence values are computed by several procedures both with and without a logarithic data transformation, together with the 98th percentile of the data for display. These computations are based on results returned from function `gx.stats`. Fences are computed following Tukey's boxplot procedure, as median +/- 2 * MAD (Median Absolute Deviation), and mean +/- 2 * SD (Standard Deviation), see Reimann et al. (2005). It is essential that these estimates are viewed in the context of the graphical distributional displays, e.g., `shape` and its graphical components, `gx.hist`, `gx.ecdf`, `cnpplt` and `bxplot`, and if spatial coordinates for the sample sites are available `edamap7`, `edamap8` and `caplot`. The final selection of a range for background or the selection of a threshold level needs to take the statistical and spatial distributions of the data into account. It is also necessary to be aware that it might be appropriate to have more than one background range/threshold in an area (Reimann and Garrett, 2005). The presence of relevant information in the data frame may permit the data to be subset on the basis of criteria using the `tbplot` and `bwplot` functions. If these indicate that the medians and middle 50%s of the data are visibly different, multiple background ranges may be advisable.

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see `ltdl.fix.df`.

Any `NA`s in the data vector are removed prior to computing the fences.

Author(s)

Robert G. Garrett

References

Reimann, C. and Garrett, R.G., 2005. Geochemical background - Concept and reality. Science of the Total Environment, 350(1-3):12-27.

Reimann, C., Filzmoser, P. and Garrett, R.G., 2005. Background and threshold: critical comparison of methods of determination. Science of the Total Environment, 346(1-3):1-16.

See Also

gx.stats, fences.summary, ltdl.fix.df, remove.na

Examples

```
## Make test data available
data(kola.o)
attach(kola.o)

## Display the fences computed for Cu
fences(Cu)

## Detach test data
detach(kola.o)
```

---

fences.summary *Generate and Save Fence Values for Data Subsets*

Description

Function to generate fences and save the values in the R working directory for subsets of the data for a variable when the data can be subdivided by some criterion (factor) such as EcoRegion, Province, physical sample parent material, etc. The function supports the selection of the upper and lower bounds of background variability, and threshold(s) or action levels, when obvious graphical solutions are not visually recognizable.

Usage

```
fences.summary(group, x, file = NULL)
```

Arguments

group	the name of the factor variable by which the data are to be subset.
x	name of the variable to be processed.
file	the first part of the file name identifying the data source for saving the function output in the R working directory, see Details below.

Details

The fence values are computed by several procedures both with and without a logarithic data transformation, together with the 98th percentile of the data for display. These computations are based on results returned from function `gx.stats`. Fences are computed following Tukey's boxplot procedure, as median +/- 2 * MAD (Median Absolute Deviation), and mean +/- 2 * SD (Standard Deviation), see Reimann et al. (2005). It is essential that these estimates are viewed in the context of the graphical distributional displays, e.g., `shape` and its graphical components, `gx.hist`, `gx.ecdf`, `cnpplt` and `bxplot`, and if spatial coordinates for the sample sites are available `edamap7`, `edamap8` and `caplot`. The final selection of a range for background or the selection of a threshold level needs to take the statistical and spatial distributions of the data into account. It is also necessary to be aware that it might be appropriate to have more than one background range/threshold in an area (Reimann and Garrett, 2005). The presence of relevant information in the data frame may permit the data to be subset on the basis of criteria using the `tbplot` and `bwplot` functions. If these indicate that the medians and middle 50%s of the data are visibly different, multiple background ranges may be advisable.

`file` contains the first part of the file name identifying the data source for the output file to be saved in the R working directory, see Note below. The function concatenates the working directory name with `file`, _, `group` as a character string, _, `x` as a character string, and `_fences.txt` to be used as the file name.

Output to the current device is suppressed. The output file can be inspected with a text viewer, and column spacings edited for cosmetic purposes with an ASCII editor of the user's choice.

Note

To set the R working directory, use at the R command line, for example, `setwd("C:\R\WDn")` which will result in all saved output being placed in that folder.

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see `ltdl.fix.df`.

Any `NA`s in the data vector are removed prior to computing the fences.

The function `fences` is employed to compute the statistical fence estimates.

Author(s)

Robert G. Garrett

References

Reimann, C. and Garrett, R.G., 2005. Geochemical background - Concept and reality. Science of the Total Environment, 350(1-3):12-27.

Reimann, C., Filzmoser, P. and Garrett, R.G., 2005. Background and threshold: critical comparison of methods of determination. Science of the Total Environment, 346(1-3):1-16.

See Also

`fences`, `ltdl.fix.df`, `remove.na`

Examples

```
## Make test data available
data(kola.c)
attach(kola.c)

## Saves the file kola_c_COUNTRY_Cu_fences.txt for later use
## in the R working directory.
fences.summary(COUNTRY, Cu, file = "kola_c")

## Detach test data
detach(kola.c)
```

---

fix.test *Test Data for Function ltdl.fix.df*

Description

A set of test data to demonstrate how negative values are changed to half their positive value. Optionally numeric coded values representing missing data and/or zero values may be replaced by NAs.

Usage

```
fix.test
```

Format

A data frame containing 15 rows and 4 columns (2 factors and 2 numeric).

---

framework.stats *Compute Framework/Subset Summary Statistics*

Description

Function to compute summary statistics for use with function framework.summary.

Usage

```
framework.stats(xx)
```

Arguments

xx	name of the variable to be processed.

Details

The function computes summary statistics consisting of the count of valid data, the number of NAs, the minimum, 2nd, 5th, 10th, 25th (Q1), 50th (median), 75th (Q3), 90th, 95th and 98th percentiles and the maximum. The 95% confidence interval for the median is computed via the binomial theorem. In addition the Median Absolute Deviation (MAD) and Inter-Quartile Standard Deviation (IQSD) are computed as robust estimates of the standard deviation. Finally, the mean, standard deviation and coefficient of variation as a percentage are computed.

Value

table	a 20-element table is returned, see below.
[1]	the data/subset (sample) size, N.
[2]	number of NAs encountered in the input vector, NNA.
[3:13]	the data minimum, 2nd, 5th, 10th, 25th (Q1), 50th (median), 75th (Q3), 90th, 95th and 98th percentiles and the maximum.
[14:15]	the lower and upper 95% confidence bounds for the median.
[16]	the Median Absolute Deviation (MAD).
[17]	the Inter-Quartile Standard Deviation (IQSD).
[18]	the data (sample) Mean.
[19]	the data (sample) Standard Deviation (SD).
[20]	the Coefficient of Variation as a percentage (CV%).

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see ltdl.fix.df.

Any NAs in the data vector are counted and then removed prior to computing the summary statistics.

Author(s)

Robert G. Garrett

See Also

ltdl.fix.df, remove.na

Examples

```
## Make test data available
data(kola.c)
attach(kola.c)

## Computes summary statistics for the Cu data
fs <- framework.stats(Cu)
fs
```

```
## Computes summary statistics for Finnish subset of the Cu data
fs <- framework.stats(Cu[COUNTRY == "FIN"])
fs

## Detach test data
detach(kola.c)
```

`framework.summary` *Generate and Save Framework/Subset Summary Statistics*

Description

Function to generate 'framework' or subset summary statistics and save them as a .csv file in the R working directory. The file can be directly imported into a spreadsheet, e.g., MS Excel, for inspection, or into other software, e.g., a Geographical Information System (GIS) where the spatial information concerning the 'framework' units is available, e.g., ecoclassification units.

Usage

```
framework.summary(group, x, file = NULL)
```

Arguments

`group`	the name of the factor variable by which the data are to be subset.
`x`	name of the variable to be processed.
`file`	the first part of the file name identifying the data source for saving the function output in the R working directory, see Details below.

Details

`file` contains the first part of the file name identifying the data source for the output file to be saved in the R working directory, see Note below. The function concatenates the working directory name with `file`, _, `group` as a character string, _, `x` as a character string, and `_summary.csv` to be used as the file name.

Output to the current device is suppressed. The output file can be inspected with spread sheet software or a viewer of the user's choice.

Note

To set the R working directory, use at the R command line, for example, `setwd("C:\R\WDn")` which will result in all saved output being placed in that folder.

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see `ltdl.fix.df`.

Any `NA`s in the data vector are counted and then removed prior to computing the summary statistics.

The function `framework.stats` is employed to compute the summary statistics.

Author(s)

Robert G. Garrett

See Also

framework.stats, ltdl.fix.df, remove.na

Examples

```
## Make test data available
data(kola.c)
attach(kola.c)

## Saves the file kola_c_COUNTRY_Cu_summary.csv for later use
## in the R working directory.
framework.summary(COUNTRY, Cu, file = "kola_c")

## Detach test data
detach(kola.c)
```

---

gx.ecdf *Emprical Cumulative Distribution Function (ECDF)*

Description

Displays an empirical cumulative distribution function (ECDF) plot with a zero-to-one linear y-scale as part of the multi-panel display provided by shape. The function may also be used stand-alone.

Usage

```
gx.ecdf(xx, xlab = deparse(substitute(xx)),
        ylab = "Empirical Cumulative Distribution Function",
        log = FALSE, xlim = NULL, main = " ", pch = 3, ifqs = FALSE)
```

Arguments

xx	name of the variable to be plotted.
xlab	a title for the x-axis. It is often desirable to replace the default x-axis title of the input variable name text string with a more informative title, e.g., xlab = "Cu (mg/kg) in <2 mm O-horizon soil".
ylab	a title for the y-axis, defaults to "Emprical Cumulative Distribution Function".
log	if it is required to display the data with logarithmic (x-axis) scaling, set log = TRUE.
xlim	when used in the shape function, xlim is determined by gx.hist and used to ensure all four panels in shape have the same x-axis scaling. However, when used stand-alone the limits may be user-defined by setting xlim, see Note below.

main	when used stand-alone a title may be added optionally above the plot by setting main, e.g., main = "Kola Project, 1995".
pch	by default the plotting symbol is set to a plus, pch = 3, alternate plotting symbols may be chosen from those displayed by display.marks.
ifqs	setting ifqs = TRUE results in horizontal and vertical dotted lines being plotted at the three central quartiles and their values, respectively.

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codesrepresenting blanks in the data vector, must be removed prior to executing this function, see ltdl.fix.df.

Any NAs in the data vector are removed prior to displaying the plot.

Although the cumulative normal percentage probability (CPP) plot is often the preferred method for displaying the cumulative data distribution as it provides greater detail for inspection in the tails of the data, the ECDF is particularly useful for studying the central parts of data distributions as it has not been compressed to make room for the scale expansion in the tails of a cumulative normal percentage probability (CPP) plot.

If the default selection for xlim is inappropriate it can be set, e.g., xlim = c(0, 200) or c(2, 200). If the defined limits lie within the observed data range a truncated plot will be displayed. If this occurs the number of data points omitted is displayed below the total number of observations.

If it is desired to prepare a display of data falling within a defined part of the actual data range, then either a data subset can be prepared externally using the appropriate R syntax, or xx may be defined in the function call as, for example, Cu[Cu < some.value] which would remove the influence of one or more outliers having values greater than some.value. In this case the number of data values displayed will be the number that are <some.value.

Author(s)

Robert G. Garrett

See Also

display.marks, ltdl.fix.df, remove.na

Examples

```
## Make test data available
data(kola.o)
attach(kola.o)

## Plot a simple ECDF
gx.ecdf(Cu)

## Plot an ECDF with more appropriate labelling and with the quartiles
## indicated
gx.ecdf(Cu , xlab = "Cu (mg/kg) in <2 mm O-horizon soil", log = TRUE,
ifqs = TRUE)
```

```
## Detach test data
detach(kola.o)
```

---

gx.hist *Plot a Histogram*

Description

Plots a histogram for a data set, the user has options for defining the axis and main titles, the x-axis limits, arithmetic or logarithmic x-axis scaling, the number of bins the data are displayed in, and the colour of the infill.

Usage

```
gx.hist(xx, xlab = deparse(substitute(xx)),
        ylab = "Number of Observations", log = FALSE, xlim = NULL,
        main = "", nclass = "Scott", colr = 8, ifnright = TRUE)
```

Arguments

xx	name of the variable to be plotted
xlab	a title for the x-axis. It is often desirable to replace the default x-axis title of the input variable name text string with a more informative title, e.g., `xlab = "Cu (mg/kg) in <2 mm O-horizon soil"`.
ylab	a default y-axis title of `"Number of Observations"` is provided, this may be changed, e.g., `ylab = "Counts"`.
log	if it is required to display the data with logarithmic (x-axis) scaling, set `log = TRUE`.
xlim	default limits of the x-axis are determined in the function for use in other panel plots of function `shape`. However when used stand-alone the limits may be user-defined by setting `xlim`, see Note below.
main	when used stand-alone a title may be added optionally above the plot by setting `main`, e.g., `main = "Kola Project, 1995"`.
nclass	the default procedure for preparing the histogram is to use the Scott (1979) rule. This usually provides an informative histogram, other optional rules are `nclass = "sturges"` or `nclass = "fd"`; the later standing for Freedman-Diaconis (1981), a rule that is resistant to the presence of outliers in the data.
colr	by default the histogram is infilled in grey, `colr = 8`. If no infill is required, set `colr = 0`. See function `display.lty` for the range of available colours.
ifnright	controls where the sample size is plotted in the histogram display, by default this in the upper right corner of the plot. If the data distribution is such that the upper left corner would be preferable, set `ifnright = FALSE`.

Value

xlim — A two element vector containing the actual minimum [1] and maximum [2] x-axis limits used in the histogram display are returned. These are use in function shape to ensure all panels have the same x-axis limits.

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see ltdl.fix.df.

Any NAs in the data vector are removed prior to displaying the plots.

If the default selection for xlim is inappropriate it can be set, e.g., xlim = c(0, 200) or c(2, 200). If the defined limits lie within the observed data range a truncated plot will be displayed. If this occurs the number of data points omitted is displayed below the total number of observations.

If it is desired to prepare a display of data falling within a defined part of the actual data range, then either a data subset can be prepared externally using the appropriate R syntax, or xx may be defined in the function call as, for example, Cu[Cu < some.value] which would remove the influence of one or more outliers having values greater than some.value. In this case the number of data values displayed will be the number that are <some.value.

Author(s)

Robert G. Garrett

References

Venables, W.N. and Ripley, B.D., 2001. Modern Applied Statistics with S-Plus, 3rd Edition, Springer - see pp. 119 for a description of histogram bin selection computations.

See Also

display.lty, ltdl.fix.df, remove.na

Examples

```
## Make test data available
data(kola.o)
attach(kola.o)

## Generates an initial display to have a first look at the data and
## decide how best to proceed
gx.hist(Cu)

## Provides a more appropriate initial display
gx.hist(Cu, xlab = "Cu (mg/kg) in <2 mm O-horizon soil", log = TRUE)

## Causes the Friedman-Diaconis rule to be used to select the number
## of histogram bins
shape(Cu, xlab = "Cu (mg/kg) in <2 mm O-horizon soil", log = TRUE,
```

```
        nclass = "fd")

## Detach test data
detach(kola.o)
```

`gx.stats` *Compute Summary Statistics*

Description

Function to compute summary statistics for a 'one-page' report and display in `inset`. Function may be used stand-alone.

Usage

```
gx.stats(xx, xlab = deparse(substitute(xx)), display = TRUE)
```

Arguments

`xx`	name of the variable to be processed.
`xlab`	a title for the table. It is often desirable to replace the default table title of the input variable name text string with a more informative title, e.g., `xlab = "Cu (mg/kg) in <2 mm O-horizon soil"`.
`display`	if `display = TRUE` the summary statistics are displayed on the current device. If `display = FALSE` output is suppressed.

Details

The summary statistics comprise the data minimum, maximum and percentile values, robust estimates of standard deviation, the Median Absolute Deviation (MAD) and the Inter Quartile Standard Deviation (IQSD), and the mean, variance, standard deviation (SD) and coefficient of variation (CV%).

Value

`table`	the computed summary statistics to be used in function `inset`. The list returned, `table`, is a 26-element vector, see below.
`[1:10]`	the minimum value, and the 1st, 2nd, 5th, 10th, 20th, 25th (Q1), 30th, 40 and 50th(Q2) percentiles
`[11:19]`	the 60th, 70th, 75th(Q3), 80th 90th, 95th, 98th and 99th percentiles and the maximum value
`[20]`	the sample size, N
`[21]`	the Median Absolute Deviation (MAD)
`[22]`	The Inter-Quartile Standard Deviation (IQSD)
`[23]`	the data (sample) Mean

[24]	the data (sample) Variance
[25]	the data (sample) Standard Deviation (SD)
[26]	the Coefficient of Variation as a percentage (CV%)

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see ltdl.fix.df.

Any NAs in the data vector are removed prior to computation.

Author(s)

Robert G. Garrett

See Also

ltdl.fix.df, remove.na

Examples

```
## Make test data available
data(kola.o)
attach(kola.o)

## Generates an initial display
gx.stats(Cu)

## Provides a more appropriate labelled display
gx.stats(Cu, xlab = "Cu (mg/kg) in <2 mm O-horizon soil")

## Detach test data
detach(kola.o)
```

---

gx.subset *Extracts a Subset of Rows from a Dataframe*

Description

The function extracts a subset of rows from a dataframe and returns the subset as a new dataframe based on the criterion provided by the user. Unused factor names are dropped.

Usage

```
gx.subset(dfname, subset = TRUE)
```

Arguments

dfname	name of the dataframe from which rows are to be extracted.
subset	the criterion for selecting the subset (rows).

Details

The subset criterion can be 'complex' and be a combination of conditions, see Examples below.

Value

data	a dataframe only containing the rows of the input dataframe where the criterion is met.

Note

This function is based on a script shared by Bill Venables on S-News, October 10, 1997. As such it may pre-date the time that subset was added to the S-Plus library. It is simple to use and has been retained.

Author(s)

William N. Venables

Examples

```
## Make test data available
data(kola.c)

## Make a subset of the data for Finland
finland.c <- gx.subset(kola.c, COUNTRY == "FIN")

## Make a subset of the data for rock type, LITHO, 82 occurring
## in Russia. Note that both COUNTRY and LITHO are factor variables
russia.82 <- gx.subset(kola.c, COUNTRY == "RUS" & LITHO == 82)

## Make a subset of the data for Cu exceeding 50(ppm) in Norway
norway.cugt50 <- gx.subset(kola.c, COUNTRY == "NOR" & Cu >50)
```

---

inset *An EDA Graphical and Statistical Summary*

Description

Plots a two panel graphical distributional summary for a data set, comprising a histogram and a cumulative normal percentage probability (CPP) plot, together with a table of selected percentiles of the data and summary statistics between them. Optionally the EDA graphics may be plotted with logarithmic scaling.

Usage

```
inset(xx, xlab = deparse(substitute(xx)), log = FALSE, xlim = NULL,
          nclass = NULL, ifnright = TRUE, ...)
```

Arguments

xx	name of the variable to be plotted.
xlab	a title for the x-axis. It is often desirable to replace the default x-axis title of the input variable name text string with a more informative title, e.g., `xlab = "Cu (mg/kg) in <2 mm O-horizon soil"`.
log	if it is required to display the data with logarithmic (x-axis) scaling, set `log = TRUE`.
xlim	default limits of the x-axis are determined in the function. However when used stand-alone the limits may be user-defined by setting `xlim`, see Note below.
nclass	the default procedure for preparing the histogram is to use the Scott (1979) rule. This usually provides an informative histogram, other optional rules are `nclass = "sturges"` or `nclass = "fd"`; the later standing for Freedman-Diaconis (1981), a rule that is resistant to the presence of outliers in the data.
ifnright	controls where the sample size is plotted in the histogram display, by default this in the upper right corner of the plot. If the data distribution is such that the upper left corner would be preferable, set `ifnright = FALSE`.
...	further arguments to be passed to or from methods. For example, by default individual data points in the ECDF and CPP plots are marked by a plus sign, `pch = 3`, if a cross or open circle is desired, then set `pch = 4` or `pch = 1`, respectively. See `display.marks` for all available symbols. Adding `ifqs = TRUE` results in horizontal and vertical dotted lines being plotted at the three central quartiles and their values, respectively, in the CPP plot.

Details

A histogram is displayed on the left, and a cumulative normal percentage probability plot on the right. Between the two is a table of simple summary statistics, computed in `gx.stats`, including minimum, maximum and percentile values, robust estimates of standard deviation, and the mean, standard deviation and coefficient of variation. The plots may be displayed with logarithmic axes, however, the summary statistics are not computed with a logarithmic transform.

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see `ltdl.fix.df`.

Any NAs in the data vector are removed prior to displaying the plot.

If the default selection for `xlim` is inappropriate it can be set, e.g., `xlim = c(0, 200)` or `c(2, 200)`. If the defined limits lie within the observed data range a truncated plot will be displayed. If this occurs the number of data points omitted is displayed below the total number of observations.

The purpose of this function is to prepare publication quality graphics (.wmf) files that can be included in reports or used as inset statistical summaries for maps. If a series of these are to be prepared the function `inset.exporter` can be used to advantage as it saves a graphics file as part of its procedure.

In some instances if the graphics window has been resized the last line(s) of the table may not be displayed. Resizing the window to be smaller will display the whole table. If the whole table is not visible it will not be saved properly to the graphics file in `inset.exporter`. Once as a complete graphics file the image may be resized in the receiving document.

In some R installations the generation of multi-panel displays and the use of function eqscplot from package MASS causes warning messages related to graphics parameters to be displayed on the current device. These may be suppressed by entering `options(warn = -1)` on the R command line, or that line may be included in a 'first' function prepared by the user that loads the rgr package, etc.

Author(s)

Robert G. Garrett

References

Venables, W.N. and Ripley, B.D., 2001. Modern Applied Statistsis with S-Plus, 3rd Edition, Springer - see pp. 119 for a description of histogram bin selection computations.

See Also

`gx.hist`, `cnpplt`, `gx.stats`, `inset.exporter`, `ltdl.fix.df`, `remove.na`

Examples

```
## Make test data available
data(kola.o)
attach(kola.o)

## Generates an initial display
inset(Cu)

## Provides a more appropriate display for pubication
inset(Cu, xlab = "Cu (mg/kg) in <2 mm O-horizon soil", log = TRUE)

## Detach test data
detach("kola.o")
```

---

inset.exporter *Saves an EDA Graphical and Statistical Summary*

Description

Saves the output from inset as a graphics file in the R working directory for use in report or map preparation.

Usage

```
inset.exporter(x, xlab = deparse(substitute(x)), log = FALSE, xlim = NULL,
        nclass = NULL, ifnright = TRUE, file = NULL, gtype = "wmf", ...)
```

Arguments

x	name of the variable to be plotted.
xlab	a label for the x-axis. It is often desirable to replace the default x-axis label of the input variable name text string with a more informative label, e.g., `xlab = "Cu (mg/kg) in <2 mm O-horizon soil"`.
log	if it is required to display the data with logarithmic (x-axis) scaling, set `log = TRUE`.
xlim	default limits of the x-axis are determined in the function. However when used stand-alone the limits may be user-defined by setting `xlim`, see Note below.
nclass	the default procedure for preparing the histogram is to use the Scott (1979) rule. This usually provides an informative histogram, other optional rules are `nclass = "sturges"` or `nclass = "fd"`; the later standing for Freedman-Diaconis (1981), a rule that is resistant to the presence of outliers in the data.
ifnright	controls where the sample size is plotted in the histogram display, by default this in the upper right corner of the plot. If the data distribution is such that the upper left corner would be preferable, set `ifnright = FALSE`.
file	the first part of the file name identifying the data source for saving the function output in the R working directory, see Details below.
gtype	the format of the graphics file to be saved. By default `gtype = "wmf"` for a Windows metafile. Other alternatives are `gtype = "jpg"` for a jpeg file, `gtype = "png"` for a portable graphics file, `gtype = "ps"` for a postscript file, or `gtype = "pdf"` for a pdf file.
...	further arguments to be passed to methods. For example, by default individual data points in the ECDF and CPP plots are marked by a plus sign, `pch = 3`, if a cross or open circle is desired, then set `pch = 4` or `pch = 1`, respectively. See `display.marks` for all available symbols. Adding `ifqs = TRUE` results in horizontal and vertical dotted lines being plotted at the three central quartiles and their values, respectively, in the CPP plot.

Details

See `inset` for details concerning the `inset` parameters.

`file` contains the first part of the file name identifying the data source for the output file to be saved in the R working directory, see Note below. The function concatenates the working directory name with `file`, _, `x` as a character string, and `_inset`. Subsequently the suffix `gtype` is appended and the file saved in the R working directory.

Note

To set the R working directory, use at the R command line, for example, `setwd("C:\R\WDn")` which will result in all saved output being placed in that folder.

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see `ltdl.fix.df`.

Any NAs in the data vector are removed prior to displaying and saving the plot.

If the default selection for `xlim` is inappropriate it can be set, e.g., `xlim = c(0, 200)` or `c(2, 200)`. If the defined limits lie within the observed data range a truncated plot will be displayed. If this occurs the number of data points omitted is displayed below the total number of observations.

In some instances if the graphics window has been resized the last line(s) of the table may not be displayed. Resizing the window to be smaller will display the whole table. If the whole table is not visible it will not be saved properly to the graphics file in `inset.exporter`. Once as a complete graphics file the image may be resized in the receiving document.

In some R installations the generation of multi-panel displays and the use of function eqscplot from package MASS causes warning messages related to graphics parameters to be displayed on the current device. These may be suppressed by entering `options(warn = -1)` on the R command line, or that line may be included in a 'first' function prepared by the user that loads the rgr package, etc.

Author(s)

Robert G. Garrett

See Also

`inset`, `ltdl.fix.df`

Examples

`kola.c` *Kola Project C-horizon Soil Data*

Description

These data arise from an ecogeochemical survey undertaken by the Central Kola Expedition of Russia (CKE), the Geological Survey of Finland (GTK) and the Norwegian Geological Survey (NGU). In 1995 a variety of soil and biological materials were collected from almost 700 sites lying between the Arctic Circle and the Barents Sea, and Longitudes 35.5 and 40.0 East. This specific data set is for C-horizon soils found at 606 of the sites visited. The data consist of an integer identifier, Universal Transverse Mercator (m) eastings and northings coordinates, the country the site was located in as a 3 character string, the lithology of the underlying bedrock as an integer code, 36 chemical measurements (total or near-total geochemical analyses), and soil pH for the <2 mm fraction of the C-horizon soils. The data reflect the natural geochemical variations in the parent material of the overlying soils. Further details concerning the project, methods of sampling and analysis can be found in Reimann et al. (1998) and the numerous papers published by the co-authors in international scientific journals.

Usage

```
kola.c
```

Format

A data frame containing 44 observations for 617 sites.

Source

These data are a subset of the full Kola C-horizon data set available from: http://doi.pangaea.de/10.1594/PANGAEA.56227

However, it should be noted that the spatial coordinates are recorded as Latitudes and Longitudes in the full data set.

References

Reimann, C., Ayras, M., Chekushin, V., Bogatyrev, I., Boyd, R., de Caritat, P., Dutter, R., Finne, T.E., Halleraker, J.H., Jager, O., Kashulina, G., Niskavaara, H., Pavlov, V., Raisanen, M.L., Strand, T. and Volden, T., 1998. A geochemical atlas of the central parts of the Barents Region. Geological Survey of Norway (NGU), Trondheim, Norway. ISBN 82-7385-176-1. 745 p.

---

kola.o *Kola Project O-horizon Soil Data*

Description

These data arise from an ecogeochemical survey undertaken by the Central Kola Expedition of Russia (CKE), the Geological Survey of Finland (GTK) and the Norwegian Geological Survey (NGU). In 1995 a variety of soil and biological materials were collected from almost 700 sites lying between the Arctic Circle and the Barents Sea, and Longitudes 35.5 and 40.0 East. This specific data set is for O-horizon soils found at 617 of the sites visited. The data consist of an integer identifier, Universal Transverse Mercator (m) eastings and northings coordinates, 38 chemical measurements (total or near-total geochemical analyses), Loss on Ignition, soil pH and specific conductivity for the <2 mm fraction of the O-horizon (humus) soils. The data reflect both natural biogeochemical variations and the presence of heavy industry. Further details concerning the project, and methods of sampling and analysis can be found in Reimann et al. (1998) and the numerous papers published by the co-authors in international scientific journals.

Usage

```
kola.o
```

Format

A data frame containing 44 observations for 617 sites.

Source

These data are the same as in the R package mvoutlier. However, note that the names of the spatial coordinates have been changed from XCOO and YCOO to UTME and UTMN, respectively, and COND (specific conductivity) to SC.

The full data set is available from: http://doi.pangaea.de/10.1594/PANGAEA.56279

However, it should be noted that this is a superset containing all geochemical analyses and the spatial coordinates are recorded as Latitudes and Longitudes in the full data set.

References

Reimann, C., Ayras, M., Chekushin, V., Bogatyrev, I., Boyd, R., de Caritat, P., Dutter, R., Finne, T.E., Halleraker, J.H., Jager, O., Kashulina, G., Niskavaara, H., Pavlov, V., Raisanen, M.L., Strand, T. and Volden, T., 1998. A geochemical atlas of the central parts of the Barents Region. Geological Survey of Norway (NGU), Trondheim, Norway. ISBN 82-7385-176-1. 745 p.

---

ltdl.fix *Replace Negative Values Representing Less Than Detects for a Vector*

Description

Function to process a vector to replace negative values representing less than detects (<value) with positive half that value. This permits processing of these effectively categorical data as real numbers and their display on logarithmically scaled axes. In addition, some software packages replace blank fields that should be interpreted as NAs, i.e. no information, with zeros. The facility is provided to replace any zero values with NAs. In other instances data files have been built using an integer code, e.g., -9999, to indicate 'no data', i.e. the equivalent of NAs. The facility is provided to replace any so coded values with NAs.

A report of the changes made is displayed on the current device.

For processing data matrices or dataframes, see ltdl.fix.df.

Usage

```
ltdl.fix(x, zero2na = FALSE, coded = NA)
```

Arguments

x	name of the vector to be processed.
zero2na	to replace any zero values with NAs, set zero2na = TRUE.
coded	to replace any numeric coded values, e.g., -9999 with NAs, set coded = -9999.

Value

A vector identical to that input but where any negative values have been replaced by half their positive values, and optionally any zero or numeric coded values have been replaced by NAs.

Note

If data are being accessed through an ODBC link to a database, rather than from a dataframe that can be processed by `ltdl.fix.df`, it may be important to run this function on the retrieved vector prior to any subsequent processing. The necessity for such vector processing can be ascertained using the range function, e.g., `range(na.omit(x))`, where x is the variable name, to determine the presence of any negative values. The presence of any `NA`s in the vector will return `NA`s in the `range` function without the `na.omit`, i.e. `range(x)`.

Great care needs to be taken when processing data where a large proportion of the data are less than detects (<value). In such cases parametric statistics have limited value, and can be missleading. Records should be kept of variables containing <values, and the fixed replacement values changed in tables for reports to the appropriate <values. Thus, in tables of percentiles the <value should replace the fixed value computed from absolute(-value)/2. Various rules have been proposed as to how many less than detects treated in this way can be tolerated before means, variances, etc. become biassed and of little value. Less than 5% in a large data set is usually tolerable, with greater than 10% concern increases, and with greater than 20% alternate procedures for processing the data should be sought.

Author(s)

Robert G. Garrett

See Also

`ltdl.fix.df`

Examples

```
## Replace any missing data coded as -9999 with NAs and any remaining
## negative values representing less than detects with Abs(value)/2
data(fix.test)
x <- fix.test[, 3]
x
x.fixed <- ltdl.fix(x, coded = -9999)
x.fixed

## As above, and replace any zero values with NAs
x.fixed <- ltdl.fix(x, coded = -9999, zero2na = TRUE)
x.fixed

## Make test data kola.o available, setting a -9999, indicating a
## missing pH measurement, to NA
data(kola.o)
attach(kola.o)
pH.fixed <- ltdl.fix(pH, coded = -9999)

## Display relationship between pH in one pH unit intervals and Cu in
## O-horizon (humus) soil, extending the whiskers to the 2nd and 98th
## percentiles, finally removing the temporary data vector pH.fixed
bwplot(split(Cu,trunc(pH.fixed+0.5)), log=TRUE, wend = 0.02,
        xlab = "O-horizon soil pH to the nearest pH unit",
```

```
        ylab = "Cu (mg/kg) in < 2 mm O-horizon soil")
rm(pH.fixed)

## Or directly
bwplot(split(Cu,trunc(ltdl.fix(pH, coded = -9999)+0.5)), log=TRUE,
        wend = 0.02, xlab = "O-horizon soil pH to the nearest pH unit",
        ylab = "Cu (mg/kg) in < 2 mm O-horizon soil")

## Detach test data kola.o
detach(kola.o)
```

`ltdl.fix.df` *Replace Negative Values Representing Less Than Detects for a Data Frame*

Description

Function to process a matrix or dataframe to replace negative values representing less than detects (<value) with positive half that value. This permits processing of these effectively categorical data as real numbers and their display on logarithmically scaled axes. In addition, some software packages replace blank fields that should be interpreted as `NA`s, i.e. no information, with zeros. The facility is provided to replace any zero values with `NA`s. In other instances data files have been built using an integer code, e.g., -9999, to indicate 'no data', i.e. the equivalent of `NA`s. The facility is provided to replace any so coded values with `NA`s. Any factor variables in the input matrix or data frame are passed to the output matrix or data frame.

If a single vector is to be processed, use `ltdl.fix`

A report of the changes made is displayed on the current device.

Usage

```
ltdl.fix.df(x, zero2na = FALSE, coded = NA)
```

Arguments

`x`	name of the matrix or dataframe to be processed.
`zero2na`	to replace any zero values with `NA`s, set `zero2na = TRUE`.
`coded`	to replace any numeric coded values, e.g., -9999 with `NA`s, set `coded = -9999`.

Value

A matrix or dataframe identical to that input but where any negative values have been replaced by half their positive values, and optionally any zero values or numeric coded values have been replaced by `NA`s.

Note

Great care needs to be taken when processing data where a large proportion of the data are less than detects (<value). In such cases parametric statistics have limited value, and can be missleading. Records should be kept of variables containing <values, and the fixed replacement values changed in tables for reports to the appropriate <values. Thus, in tables of percentiles the <value should replace the fixed value computed from absolute(-value)/2. Various rules have been proposed as to how many less than detects treated in this way can be tolerated before means, variances, etc. become biassed and of little value. Less than 5% in a large data set is usually tolerable, with greater than 10% concern increases, and with greater than 20% alternate procedures for processing the data should be sought.

Author(s)

Robert G. Garrett and David Lorenz

See Also

ltdl.fix

Examples

```
## Replace any missing data coded as -9999 with NAs and any remaining
## negative values representing less than detects with Abs(value)/2
data(fix.test)
fix.test
fix.test.fixed <- ltdl.fix.df(fix.test, coded = -9999)
fix.test.fixed

## As above, and replace any zero values with NAs
fix.test.fixed <- ltdl.fix.df(fix.test, coded = -9999, zero2na = TRUE)
fix.test.fixed
```

---

ms.data1 *Measurement Variability Test Data*

Description

These magnetic susceptibility data were used by Stanley (2003) to demonstrate the Thompson-Howarth (Thompson and Howarth, 1973 & 1978) procedure for estimating analytical variability. They are used in the rgr package examples for the duplicate analysis ANOVA and Thompson-Howarth plot functions, anova1 and thplot1, respectively. See also Garrett and Grunsky (2003).

Usage

```
ms.data1
```

Format

A data frame containing 2 measurements of magnetic susceptibility for each of 16 rock samples in 16 records.

Source

Stanley (2003), see below.

References

Garrett, R.G. and Grunsky, E.C., 2003. S and R functions for the display of Thompson-Howarth plots. Computers & Geosciences, 29(2):239-242.

Stanley, C.R., 2003. THPLOT.M: a MATLAB function to implement generalized Thompson-Howarth error analysis using replicate data. Computers & Geosciences, 29(2):225-237.

Thompson, M. and Howarth, R.J, 1973. The rapid estimation and control of precision by duplicate determinations. The Analyst, 98(1164):153-160.

Thompson, M. and Howarth, R.J, 1978. A new approach to the estimation of analytical precision. Journal of Geochemical Exploration, 9(1):23-30.

---

ms.data2 *Measurement Variability Test Data*

Description

These magnetic susceptibility data were used by Stanley (2003) to demonstrate the Thompson-Howarth (Thompson and Howarth, 1973 & 1978) procedure for estimating analytical variability. They are used in the rgr package examples for the duplicate analysis ANOVA and Thompson-Howarth plot functions, `anova2` and `thplot2`, respectively, with `ifalt = FALSE`. See also Garrett and Grunsky (2003).

Usage

```
ms.data2
```

Format

A data frame containing 2 measurements of magnetic susceptibility for each of 16 rock samples in 32 records. The measurements for the original analyses are in records 1 to 16, and the duplicate measurements are in records 17 to 32 in the same order.

Source

Stanley (2003), see below.

References

Garrett, R.G. and Grunsky, E.C., 2003. S and R functions for the display of Thompson-Howarth plots. Computers & Geosciences, 29(2):239-242.

Stanley, C.R., 2003. THPLOT.M: a MATLAB function to implement generalized Thompson-Howarth error analysis using replicate data. Computers & Geosciences, 29(2):225-237.

Thompson, M. and Howarth, R.J, 1973. The rapid estimation and control of precision by duplicate determinations. The Analyst, 98(1164):153-160.

Thompson, M. and Howarth, R.J, 1978. A new approach to the estimation of analytical precision. Journal of Geochemical Exploration, 9(1):23-30.

---

ms.data3 *Measurement Variability Test Data*

Description

These magnetic susceptibility data were used by Stanley (2003) to demonstrate the Thompson-Howarth (Thompson and Howarth, 1973 & 1978) procedure for estimating analytical variability. They are used in the rgr package examples for the duplicate analysis ANOVA and Thompson-Howarth plot functions, `anova2` and `thplot2`, respectively, with `ifalt = TRUE`. See also Garrett and Grunsky (2003).

Usage

```
ms.data3
```

Format

A data frame containing 2 measurements of magnetic susceptibility for each of 16 rock samples in 32 records. The measurements for the original and duplicate analyses alternate. So the first duplicate pair are in records 1 and 2, and the last in records 31 and 32.

Source

Stanley (2003), see below.

References

Garrett, R.G. and Grunsky, E.C., 2003. S and R functions for the display of Thompson-Howarth plots. Computers & Geosciences, 29(2):239-242.

Stanley, C.R., 2003. THPLOT.M: a MATLAB function to implement generalized Thompson-Howarth error analysis using replicate data. Computers & Geosciences, 29(2):225-237.

Thompson, M. and Howarth, R.J, 1973. The rapid estimation and control of precision by duplicate determinations. The Analyst, 98(1164):153-160.

Thompson, M. and Howarth, R.J, 1978. A new approach to the estimation of analytical precision. Journal of Geochemical Exploration, 9(1):23-30.

remove.na *Remove and Count NAs*

Description

Function to remove rows containing NAs from a data vector or matrix. Also counts the number of rows remaining, the number of rows deleted, and in the case of a matrix the number of columns. The results are returned in a list for subsequent processing in the calling function.

Usage

```
remove.na(xx)
```

Arguments

xx	name of the vector or matrix to be processed.

Details

This function is called by many of the procedures in the rgr package. If one or more NAs are found the user is informed of how many. In general a dataframe will have been cleared of any <values represented by negative values or zeros prior to executing the procedure calling this function, see ltdl.fix.df, or ltdl.fix if a single vector is being processed.

Value

x	a data vector or matrix containing the rows of xx without NAs.
n	the length of x.
m	the number of columns in the matrix xx, if xx is a vector the value 1 is returned.
nna	the number of rows removed from xx.

Author(s)

Robert G. Garrett

See Also

ltdl.fix.df

Examples

```
## remove NAs
xx <- c(15, 39, 18, 16, NA, 53)
temp.x <- remove.na(xx)
x <- temp.x$x[1:temp.x$n]

## to recover the other values returned
n <- temp.x$n
```

```
m <- temp.x$m
nna <- temp.x$nna

## to remove NA replacing a -9999 in kola.o
data(kola.o)
kola.o.fixed <- ltdl.fix.df(kola.o, coded = -9999)
temp.x <- remove.na(kola.o.fixed$pH)
x <- temp.x$x[1:temp.x$n]
```

`rgr-package` *The GSC (Geological Survey of Canada) Applied Geochemistry EDA Package*

Description

R functions for Exploratory Data Analysis with applied geochemical data

Details

The functions in this package are used to support the display and analysis of applied geochemical survey data. Particularly in the context of estimating the ranges of background variation due to natural phenomena and the identification of outliers that may be due to natural processes or anthropogenic contamination.

Author(s)

Robert G. Garrett

Maintainer: Robert G. Garrett <garrett@NRCan.gc.ca>

References

Reimann, C., Filzmoser, P. and Garrett, R.G., 2005. Background and threshold: critical comparison of methods of determination. Science of the Total Environment, 346(1/3):1-16.

Reimann, C. and Garrett, R.G., 2005. Geochemical background - Concept and reality. Science of the Total Environment, 350(1/3):12-27.

Venables, W.N. and Ripley, B.D., 2001. Modern Applied Statistics with S-Plus, 3rd Edition, Springer.

shape *An EDA Graphical Summary*

Description

Plots a simple four panel graphical distributional summary for a data set, comprising a histogram, a horizontal Tukey boxplot or box-and-whisker plot, an empirical cumulative distribution function (ECDF), and a cumulative normal percentage probability (CPP) plot. The plots in all four panels will have identical x-axis scaling. Optionally the EDA graphics may be plotted with logarithmic scaling.

Usage

```
shape(xx, xlab = deparse(substitute(xx)), log = FALSE, xlim = NULL,
        nclass = "scott", ifbw = FALSE, wend = 0.05, colr = 8,
        ifnright = TRUE, ...)
```

Arguments

xx	name of the variable to be plotted.
xlab	a title for the x-axes. It is often desirable to replace the default x-axis title of the input variable name text string with a more informative title, e.g., `xlab = "Cu (mg/kg) in <2 mm O-horizon soil"`.
log	if it is required to display the data with logarithmic (x-axis) scaling, set `log = TRUE`.
xlim	is determined by `gx.hist` and used to ensure all four panels in this function have the same x-axis scaling. `xlim` may be defined, see Note below.
nclass	the default procedure for preparing the histogram is to use the Scott (1979) rule. This usually provides an informative histogram, other optional rules are `nclass = "sturges"` or `nclass = "fd"`; the later standing for Freedman-Diaconis (1981), a rule that is resistant to the presence of outliers in the data.
ifbw	the default is to plot a horizontal Tukey boxplot, if a box-and-whisker plot is required set `ifbw = TRUE`.
wend	if `ifbw = TRUE` the locations of the whisker-ends have to be defined. By default these are at the 5th and 95th percentiles of the data, setting `wend = 0.02` plots the whisker ends at the 2nd and 98th percentiles.
colr	by default the histogram and box are infilled in grey, `colr = 8`. If no infill is required, set `colr = 0`. See `display.lty` for the range of available colours.
ifnright	controls where the sample size is plotted in the histogram display, by default this in the upper right corner of the plot. If the data distribution is such that the upper left corner would be preferable, set `ifnright = FALSE`.

`...` further arguments to be passed to methods. For example, by default individual data points in the ECDF and CPP plots are marked by a plus sign, `pch = 3`, if a cross or open circle is desired, then set `pch = 4` or `pch = 1`, respectively. See `display.marks` for all available symbols. Adding `ifqs = TRUE` results in horizontal and vertical dotted lines being plotted at the three central quartiles and their values, respectively, in the ECDF and CPP plots.

Details

A histogram is displayed upper left, an ECDF is displayed below it (lower left). To the right of the histogram a horizontal Tukey boxplot (default) or box-and-whisker plot (option) is displayed (upper right). In the lower right quadrant a cumulative normal percentage probability (CPP) plot is displayed.

In a box-and-whisker plot there are two special cases. When `wend = 0` the whiskers extend to the observed minima and maxima that are not plotted with the plus symbol. When `wend = 0.25` no whiskers or the data minimum and maximum are plotted, only the median and box representing the span of the middle 50 percent of the data are displayed.

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see `ltdl.fix.df`.

Any `NA`s in the data vector are removed prior to displaying the plots.

If the default selection for `xlim` is inappropriate it can be set, e.g., `xlim = c(0, 200)` or `c(2, 200)`. If the defined limits lie within the observed data range truncated plots will be displayed. If this occurs the number of data points omitted is displayed below the total number of observations in the various panels.

If it is desired to prepare a display of data falling within a defined part of the actual data range, then either a data subset can be prepared externally using the appropriate R syntax, or `xx` may be defined in the function call as, for example, `Cu[Cu < some.value]` which would remove the influence of one or more outliers having values greater than `some.value`. In this case the number of data values displayed will be the number that are `<some.value`.

In some R installations the generation of multi-panel displays and the use of function eqscplot from package MASS causes warning messages related to graphics parameters to be displayed on the current device. These may be suppressed by entering `options(warn = -1)` on the R command line, or that line may be included in a 'first' function prepared by the user that loads the rgr package, etc.

Author(s)

Robert G. Garrett

References

Venables, W.N. and Ripley, B.D., 2001. Modern Applied Statistsis with S-Plus, 3rd Edition, Springer - see pp. 119 for a description of histogram bin selection computations.

Garrett, R.G., 1988. IDEAS - An Interactive Computer Graphics Tool to Assist the Exploration Geochemist. In Current Research Part F, Geological Survey of Canada Paper 88-1F, pp. 1-13 for a description of box-and-whisker plots.

See Also

gx.hist, bxplot, gx.ecdf, cnpplt, remove.na, display.lty, display.marks, ltdl.fix.df, inset

Examples

```
## Make test data available
data(kola.o)
attach(kola.o)

## Generates an initial display to have a first look at the data and
## decide how best to proceed
shape(Cu)

## Provides a more appropriate initial display and indicates the
## quartiles
shape(Cu, xlab = "Cu (mg/kg) in <2 mm O-horizon soil", log = TRUE,
        ifqs = TRUE)

## Causes the Friedman-Diaconis rule to be used to select the number of
## histogram bins and changes the ECDF and CPP plotting symbols to a
## cross/x
shape(Cu, xlab = "Cu (mg/kg) in <2 mm O-horizon soil", log = TRUE,
        nclass = "fd", pch = 4)

## Replaces the Tukey boxplot with a box-and-whisker plot where the
## whiskers extend to the 10th and 90th percentiles and the minimum
## and maximum observed values are marked with a plus sign.
shape(Cu, xlab = "Cu (mg/kg) in <2 mm O-horizon soil", log = TRUE,
        ifbw =TRUE, wend = 0.1)

## Detach test data
detach(kola.o)
```

---

syms *Function to Compute the Diameters of Proportional Symbols*

Description

This function computes the diameters of the open circles to be plotted in a map or other display.

Usage

```
syms(z, zrange = c(NA, NA), p = 1)
```

Arguments

z	name of the variable to be plotted for which diameters are to be computed.
zrange	The mimimum and maximum values of z to be used as the lower and upper limits, respectively, for the computed symbol diameters.
p	a parameter that controls the rate of change of symbol diameter with changing value. A default of `p = 1` is provided that results in a linear rate of change. See Details below.

Details

The symbol diameter is computed as a function of the value z to be plotted:
`diameter = dmin + (dmax - dmin) * {(z - zmin)/(zmax - zmin)}^p`
where dmin and dmax are defined as 0.1 and 1 units, so the symbol diameters range over an order of magnitude (and symbol areas over two); zmin and zmax are the observed range of the data, or the range over which the user wants the diameters to be computed; and p is a power defined by the user. The value of `(z - zmin)/(zmax - zmin)` is the value of z normalized, 0 - 1, to the range over which the symbol diameters are to be computed. After being raised to the power p, which will result in number in the range 0 to 1, this value is multiplied by the permissable range of diameters and added to the minimum diameter. This results in a diameter between 0.1 and 1 units that is proportional to the value of z.

A p value of `1` results in a linear rate of change. Values of p less than unity lead to a rapid intial rate of change with increasing value of z which is often suitable for displaying negatively skewed data sets, see the example below. In contrast, values of p greater than unity result in an initial slow rate of change with increasing value of z which is often suitable for displaying positively skewed data sets. Experimentation is usually necessary to obtain a satisfactory visual effect. See `syms.pfunc` for a graphic demonstrating the effect of varying the p parameter.

If zmin or zmax are defined this has the effect of setting a minimum or maximum value of z, respectively, beyond which changes in the value of z do not result in changes in symbol diameter. This can be useful in limiting the effect of one or a few extreme outliers while still plotting them, they simply plot at the minimum or maximum symbol size and are not involved in the calculation of the range of z over which the diameter varies.

Author(s)

Robert G. Garrett

See Also

`syms.pfunc`

Examples

```
## Make test data available
data(kola.o)
attach(kola.o)

## Compute default symbol diameters
circle.diam <- syms(Cu, p = 0.3)
```

```
circle.diam

## Compute symbol diameters holding all symbols for values greater
## than 1000 to the same size
circle.diam <- syms(Cu, zrange = c(NA, 1000), p = 0.3)
circle.diam

## Detach test data
detach(kola.o)
```

---

syms.pfunc *Function to Demonstrate the Effect of Different Values of p*

Description

This function displays a plot demonstrating the effect of varying the value of p, for a range of p values from 0.2 to 5, on the 0 to 1 normalized values of a variable in order to compute corresponding circular symbol diameters.

Usage

```
syms.pfunc()
```

Author(s)

Robert G. Garrett

---

tbplot *Plot Vertical Tukey Boxplots*

Description

Plots a series of vertical Tukey boxplots where the individual boxplots represent the data subdivided by the value of some factor. Optionally the y-axis may be scaled logarithmically and the values of the Tukey fences used to identify near and far outliers may also be optionally based on the logarithmically transformed data. A variety of other plot options are available, see Details and Note below.

Usage

```
tbplot(x, by, log = FALSE, logx = FALSE, notch = TRUE, xlab = "",
        ylab = deparse(substitute(x)), ylim = NULL, main = "",
        label = NULL, plot.order = NULL, xpos = NA, width, space = 0.25,
        las = 1, cex = 1, adj = 0.5, add = FALSE, ssll = 1, colr = 8,
        ...)
```

Arguments

x	name of the variable to be plotted.
by	the name of the factor variable to be used to subdivide the data. See Details below for when by is undefined.
log	if it is required to display the data with logarithmic (y-axis) scaling, set log = TRUE.
logx	if the position of the Tukey boxplot fences are to be computed on the basis of log transformed data set logx = TRUE. For general usage, if log = TRUE then set logx = TRUE.
notch	determines if the boxplots are to be "notched" such that the notches indicate the 95% confidence intervals for the medians. The default is to notch the boxplets, to suppress the notches set notch = FALSE. See Details below.
xlab	a title for the x-axis, by default none is provided.
ylab	a title for the y-axis. It is often desirable to replace the default y-axis title of the input variable name text string with a more informative title, e.g., ylab = "Cu (mg/kg) in <2 mm C-horizon soil".
ylim	defines the limits of the y-axis if the default limits based on the range of the data are unsatisfactory. It can be used to ensure the y-axis scaling in multiple sets of boxplots are the same to facilitate visual comparison.
main	a main title may be added optionally above the display by setting main, e.g., main = "Kola Project, 1995".
label	provides an alternate set of labels for the boxplots along the x-axis. By default the character strings defining the factors are used. Thus, label = c("Alt1", "Alt2", "Alt3").
plot.order	provides an alternate order for the boxplots. Thus, plot.order = c(2, 1, 3) will plot the 2nd ordered factor in the 1st position, the 1st in the 2nd, and the 3rd in its 3rd ordered postion, see Details and Examples below.
xpos	the locations along the x-axis for the individual vertical boxplots to be plotted. By default this is set to NA, which causes default equally spaced positions to be used, i.e. boxplot 1 plots at value 1 on the x-axis, boxplot 2 at value 2, etc., up to boxplot "n" at value "n". See Details below for defining xpos.
width	the width of the boxes, by default this is set to the minimum distance between all adjacent boxplots times the value of space. With the default values of xpos this results in a minimum difference of 1, and with the default of space = 0.25 the width is computed as 0.25. To specify different widths for all boxplots use, for example, width = c(0.3). See Details below for changing individual boxplot widths.
space	the space between the individual boxplots, by default this is 0.25 x-axis units.
las	controls whether the x-axis labels are written parallel to the x-axis, the default las = 1, or are written down from the x-axis by setting las = 2. See also, Details below.
cex	controls the size of the font used for the factor labels plotted along the x-axis. By default this is 1, however, if the labels are long it is sometimes necessary to use a smaller font, for example cex = 0.8 results in a font 80% of normal size.

adj	controls the justification of the x-axis labels. By default they are centred, adj = 0.5, to left justify them if the labels are written downwards at an angle set adj = 0.
add	permits the user to plot additional boxplots into an existing display. It is recommended that this option is left as add = FALSE.
ssll	determines the minimum data subset size for which a subset will be plotted. By default this is set to 1, which leads to only a circle with a median bar being plotted, as the subset size increases additional features of the boxplot are displayed. If ssll results in subset boxplots not being plotted, a gap is left and the factor label is still plotted on the x-axis.
colr	by default the boxes are infilled in grey, colr = 8 . If no infill is required, set colr = 0. See display.lty for the range of available colours.
...	further arguments to be passed to methods.

Details

There are two ways to execute this function. Firstly by defining x and by, and secondly by combining the two variables with the split function. See the first two examples below. The split function can be useful if the factors to use in the boxplot are to be generated at run-time, see the last example below. Note that when the split construct is used instead of by the whole split statement will be displayed as the default y-axis title. Also note that when using by the subsets are listed in the order that the factors are encountered in the data, but when using split the subsets are listed alphabetically. In either case they can be re-ordered using plot.order, see Examples.

The width option can be used to define different widths for the individual boxplots. For example, the widths could be scaled to be proportional to the subset population sizes as some function of the square root (const * sqrt(n)) or logarithm (const * log10(n)) of those sizes (n). The constant, const, would need to be chosen so that on average the width of the individual boxes would be approximately 0.25, see Example below. It may be desirable for cosmetic purposes to adjust the positions of the boxes along the x-axis, this can be achieved by specifying xpos.

Long subset (factor) names can lead to display problems, changing the las parameter from its default of las = 1 which plots subset labels parallel to the axis to las = 2, to plot perpendicular to the axis, can help. It may also help to use label and split the character string into two lines, e.g., by changing the string "Granodiorite" that was supplied to replace the coded factor variable GRDR to "Grano-
ndiorite". If this, or setting las = 2, causes a conflict with the x-axis title, if one is needed, the title can be moved down a line by using xlab = "
nLithological Units". In both cases the
n forces the following text to be placed on the next lower line.

If there are more than 7 labels (subsets) and no alternate labels are provided las is set to 2, otherwise some labels may fail to be displayed.

The notches in the boxplots indicate the 95% confidence intervals for the medians and can extend beyond the upper and lower limits of the boxes indicating the middle 50% of the data when subset population sizes are small. The confidence intervals are estimated using the binomial theorem. It can be argued that for small populations a normal approximation would be better. However, it was decided to remain with a non-parametric estimate despite the fact that the calculation of the Tukey fence values involves normality assumptions.

Note

This function is based on a script shared by Doug Nychka on S-News, April 28, 1992.

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see `ltdl.fix.df`.

Any `NA`s in the data vector are removed prior to preparing the boxplots.

Author(s)

Robert G. Garrett and Douglas W. Nychka

See Also

`cat2list`, `ltdl.fix.df`

Examples

```
## Make test data kola.c available
data(kola.c)
attach(kola.c)

## Display a simple Tukey boxplot
tbplot(Cu, by = COUNTRY)
tbplot(split(Cu,COUNTRY))

## Display a more appropriately labelled and scaled boxplot
tbplot(Cu, by = COUNTRY, log =  TRUE, logx = TRUE, xlab = "Country",
        ylab = "Ni (mg/kg) in <2 mm C-horizon soil")

## Display a west-to-east re-ordered plot using the full country names
tbplot(split(Cu, COUNTRY), log = TRUE, logx = TRUE,
        ylab = "Ni (mg/kg) in <2 mm C-horizon soil",
        label = c("Finland", "Norway", "Russia"),
        plot.order = c(2, 1, 3))

## Detach test data kola.c
detach(kola.c)

## Make test data kola.o available, setting a -9999, indicating a
## missing pH measurement, to NA
data(kola.o)
kola.o.fixed <- ltdl.fix.df(kola.o, coded = -9999)
attach(kola.o.fixed)

## Display relationship between pH in one pH unit intervals and Cu in
## O-horizon (humus) soil
tbplot(split(Cu,trunc(pH+0.5)), log=TRUE, logx = TRUE,
        xlab = "O-horizon soil pH to the nearest pH unit",
        ylab = "Cu (mg/kg) in <2 mm O-horizon soil")

## As above, but demonstrating the use of variable box widths and the
```

```
## suppression of 95% confidence interval notches.  The box widths are
## computed as (Log10(n)+0.1)/5, the 0.1 is added as one subset has a
## population of 1.
table(trunc(pH+0.5))
tbplot(split(Cu,trunc(pH+0.5)), log=TRUE, logx = TRUE, notch = FALSE,
        xlab = "O-horizon soil pH to the nearest pH unit,\nbox widths proportional to Log(su
        ylab = "Cu (mg/kg) in <2 mm O-horizon soil",
        width = c(0.26, 0.58, 0.24, 0.02))

## Detach test data kola.o.fixed
detach(kola.o.fixed)
```

`tbplot.by.var` *Plot Vertical Tukey Boxplots for Variables*

Description

Plots a series of vertical Tukey boxplots where the individual boxplots represent the data subdivided by variables. Optionally the y-axis may be scaled logarithmically. A variety of other plot options are available, see Details and Note below.

Usage

```
tbplot.by.var(xmat, log = FALSE, logx = FALSE, notch = FALSE,
        xlab = "Measured Variables", ylab = "Reported Values",
        main = "", label = NULL, plot.order = NULL, xpos = NA,
        las = 1, cex = 1, adj = 0.5, colr = 8, ...)
```

Arguments

`xmat`	the data matrix or data frame containing the data.
`log`	if it is required to display the data with logarithmic (y-axis) scaling, set `log = TRUE`.
`logx`	if the positions of the Tukey boxplot fences are to be computed on the basis of log transformed data set `logx = TRUE`. For general usage, if `log = TRUE` then set `logx = TRUE`.
`notch`	determines if the boxplots are to be "notched" such that the notches indicate the 95% confidence intervals for the medians. The default is not to notch the boxplots, to have notches set `notch = TRUE`.
`xlab`	a title for the x-axis, by default `xlab = "Measured Variables"`.
`ylab`	a title for the y-axis, by default `ylab = "Reported Values"`.
`main`	a main title may be added optionally above the display by setting `main`, e.g., `main = "Kola Project, 1995"`.
`label`	provides an alternate set of labels for the boxplots along the x-axis. By default the character strings defining the factors (variables) are used. Thus, `label = c("Alt1", "Alt2", "Alt3")`.

plot.order	provides an alternate order for the boxplots. By default the boxplot are plotted in alphabetical order of the factor variables. Thus, plot.order = c(2, 1, 3) will plot the 2nd alphabetically ordered factor in the 1st position, the 1st in the 2nd, and the 3rd in its alphabetically 3rd ordered postion.
xpos	the locations along the x-axis for the individual vertical boxplots to be plotted. By default this is set to NA, which causes default equally spaced positions to be used, i.e. boxplot 1 plots at value 1 on the x-axis, boxplot 2 at value 2, etc., up to boxplot "n" at value "n". See Details below for defining xpos.
las	controls whether the x-axis labels are written parallel to the x-axis, the default las = 1, or are written down from the x-axis by setting las = 2. See also, Details below.
cex	controls the size of the font used for the factor labels plotted along the x-axis. By default this is 1, however, if the labels are long it is sometimes necessary to use a smaller font, for example cex = 0.8 results in a font 80% of normal size.
adj	controls the justification of the x-axis labels. By default they are centred, adj = 0.5, to left justify them if the labels are written downwards set adj = 0.
colr	by default the boxes are infilled in grey, colr = 8 . If no infill is required, set colr = 0. See display.lty for the range of available colours.
...	further arguments to be passed to methods.

Details

There are two ways to provide data to this function. Firstly, if all the variables in a data frame are to be displayed, and there are no factor variables, the data frame name can be entered for xmat. However, if there are factor variables, or only a subset of the variables are to be displayed, the data are entered via the cbind construct, see Examples below.

Long variable names can lead to display problems, changing the las parameter from its default of las = 1 which plots subset labels parallel to the axis to las = 2, to plot perpendicular to the axis, can help. It may also help to use label and split the character string into two lines, e.g., by changing the string "Specific Conductivity" that was supplied to replace the variable name SC to "Specific
nConductivity". If this, or setting las = 2, causes a conflict with the x-axis title, if one is needed, the title can be moved down a line by using xlab = "
nPhysical soil properties". In both cases the
n forces the following text to be placed on the next lower line.

If there are more than 7 labels (variables) and no alternate labels are provided las is set to 2, otherwise some variable names may fail to be displayed.

The notches in the boxplots indicate the 95% confidence intervals for the medians and can extend beyond the upper and lower limits of the boxes indicating the middle 50% of the data when subset population sizes are small. The confidence intervals are estimated using the binomial theorem. It can be argued that for small populations a normal approximation would be better. However, it was decided to remain with a non-parametric estimate despite the fact that the calculation of the Tukey fence values involves normality assumptions.

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see ltdl.fix.df.

Any NAs in the data vectors are removed prior to preparing the boxplots.

Author(s)

Robert G. Garrett

See Also

tbplot, var2fact, ltdl.fix.df

Examples

```
## Make test data kola.c available
data(kola.c)
attach(kola.c)

## Display a simple Tukey boxplot for measured variables
tbplot.by.var(cbind(Co,Cu,Ni))

## Display a more appropriately labelled and scaled Tukey boxplot
tbplot.by.var(cbind(Co,Cu,Ni), log = TRUE, logx = TRUE,
        ylab = "Levels (mg/kg) in <2 mm C-horizon soil")

## Detach test data kola.c
detach(kola.c)

## Make test data ms.data1 available
data(ms.data1)

## Display variables in a data frame
tbplot.by.var(ms.data1, log=TRUE, logx = TRUE)
```

---

thplot1 *Display a Thompson-Howarth Plot of Duplicate Measurements*

Description

Function displays a Thompson-Howarth (1973 & 1978) plot for a set of duplicate measurements to visually inspect them as a part of the QA/QC process. By inputting a target precision the data may be visually checked to determine if they meet that criterion.

Usage

```
thplot1(x1, x2, name = "", ifzero = 0.01, xlow = NA, xhih = NA,
        yhih = NA, rsd = 5, ptile = 95, main = "")
```

Arguments

x1	a column vector from a matrix or data frame, x1[1], ..., x1[n].
x2	another column vector from a matrix or data frame, x2[1], ..., x2[n]. x1,x2 must be of identical length, n, where x2 is a duplicate measurement of x1.
name	a title can be displayed with the results, e.g., "name = Magnetic Susceptibility". If this field is undefined the character string for x1 is used as a default.
ifzero	as the Thompson-Howarth plot is log-scaled values of zero cannot be displayed, therefore the parameter ifzero has to be specified. A suitable choice is a value one order of magnitude lower than the value of the detection limit. A default value of ifzero = 0.01 units is provided, corresponding to a detection limit of 0.1 units.
xlow	if is desired to produce plots with consistent scaling this may be achieved by defining xlow, xhih and yhih, the ylow value is set equal to ifzero. Enter an appropriate value of xlow to ensure all data are displayed on all plots.
xhih	enter an appropriate value of xhih to ensure all data are displayed on all plots.
yhih	enter an appropriate value of yhih to ensure all data are displayed on all plots.
rsd	to assist in QA/QC inspection a target precision may be defined as a RSD%, a default of rsd = 5 is provided. See comments concerning RSD in Details below.
ptile	defines the confidence interval for a line to be drawn on the plot above which only 100 - ptile% of the points should plot if the defined target RSD is being met. A default of ptile = 95 is provided. The function counts the number of points falling 'out of limits' and reports the probability that this number would have fallen 'out of limits' by chance alone.
main	a title may be added optionally above the display, e.g., main = "Stanley (2003) Test Data".

Details

This function expects the RSD% as a measure of measurement repeatability (precision), which is more familiar to the current generation of applied geochemists, rather than the precision at the 2 Standard Deviation level. The necessary calculations to conform with the Thompson and Howarth procedure are made internally.

Duplicate pairs containing any NAs are omitted from the calculations.

If the data are as a single concatenated vector from a matrix or data frame as x[1], ..., x[n] followed by x[n+1], ..., x[2n], or alternated as x[1] and x[2] being a pair through to x[2*i+1] and x[2*i+2], for the i in 1:n duplicate pairs use function thplot2.

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see ltdl.fix.df.

Duplicate pairs x1,x2 containing any NAs are omitted from the calculations.

This script was published by Garrett and Grunsky (2003)

Author(s)

Robert G. Garrett

References

Garrett, R.G. & Grunsky, E.C., 2003. S and R functions to display Thompson-Howarth plots. Computers & Geosciences 29(2):239-242.

Stanley, C.R., 2003. THPLOT.M: A MATLAB function to implement generalized Thompson-Howarth error analysis using replicate data. Computers & Geosciences 29(2):225-237.

Thompson, M. and Howarth, R.J, 1973. The rapid estimation and control of precision by duplicate determinations. The Analyst, 98(1164):153-160.

Thompson, M. and Howarth, R.J., 1978. A new approach to the estimation of analytical precision. Journal of Geochemical Exploration, 9(1):23-30.

See Also

thplot2, ltdl.fix.df

Examples

```
## Make the Stanley (2003) test data available
data(ms.data1)
attach(ms.data1)

# Display the default plot
thplot1(MS.1, MS.2, name = "Magnetic Susceptibility",
        main = "Stanley (2003) Test Data")

# Display a Thompson-Howarth plot for a RSD of 7.5% and a draw the limit for a
# confidence interval of 90%
thplot1(MS.1, MS.2, name = "Magnetic Susceptibility", rsd = 7.5, ptile = 90,
        main = "Stanley (2003) Test Data")

## Detach test data
detach(ms.data1)
```

---

thplot2 *Display a Thompson-Howarth Plot of Duplicate Measurements, Alternate Input*

Description

Function to prepare data stored in alternate forms from that expected by function thplot1 for its use. For further details see 'x' in Arguments below.

Usage

```
thplot2(x, name = deparse(substitute(x)), ifzero = 0.01, xlow = NA,
        xhih = NA, yhih = NA, rsd = 5, ptile = 95, main = "",
        ifalt = FALSE)
```

Arguments

x	a column vector from a matrix or data frame, `x[1], ..., x[2*n]`. The default is that the first n members of the vector are the first measurements and the second n members are the duplicate measurements. If the measurements alternate, i.e. duplicate pair 1 measurement 1 followed by measurement 2, etc., set `ifalt = TRUE`.
name	a title can be displayed with the results, e.g., `name = "Magnetic Susceptibility"`. If this field is undefined the character string for `x` is used as a default.
ifzero	as the Thompson-Howarth plot is log-scaled values of zero cannot be displayed, therefore the parameter `ifzero` has to be specified. A suitable choice is a value one order of magnitude lower than the value of the detection limit. A default value of `ifzero = 0.01` units is provided, corresponding to a detection limit of 0.1 units.
xlow	if is desired to produce plots with consistent scaling this may be achieved by defining `xlow`, `xhih` and `yhih`, the `ylow`, the `ylow` value is set equal to `ifzero`. Enter an appropriate value of `xlow` to ensure all data are displayed on all plots.
xhih	enter an appropriate value of `xhih` to ensure all data are displayed on all plots.
yhih	enter an appropriate value of `yhih` to ensure all data are displayed on all plots.
rsd	to assist in QA/QC inspection a target precision may be defined as a RSD%, a default of `rsd = 5` is provided. See comments concerning RSD in details below.
ptile	defines the confidence interval for a line to be drawn on the plot above which only 100 - ptile% of the points should plot if the defined target RSD is being met. A default of `ptile = 95` is provided. The function counts the number of points falling 'out of limits' and reports the probability that this number would have fallen 'out of limits' by chance alone.
main	a title may be added optionally above the display, e.g., `main = "Stanley (2003) Test Data"`
ifalt	set `ifalt = TRUE` to accommodate alternating sets of paired observations.

Details

This function expects the RSD% as a measure of measurement repeatability (precision), which is more familiar to the current generation of applied geochemists, rather than the precision at the 2 Standard Deviation level. The necessary calculations to conform with the Thompson and Howarth procedure are made internally.

For further details see `thplot1`.

If the data are as n duplicate pairs, x1 and x2, use function `thplot1`.

Note

Any less than detection limit values represented by negative values, or zeros or other numeric codes representing blanks in the data vector, must be removed prior to executing this function, see `ltdl.fix.df`.

Author(s)

Robert G. Garrett

See Also

`thplot1`, `ltdl.fix.df`

Examples

```
## Make test data ms.data2 available
data(ms.data2)
attach(ms.data2)

## Display the default plot
thplot2(MS, name = "Magnetic Susceptibility",
        main = "Stanley (2003) Test Data")

## Detach test data ms.data2
detach(ms.data2)

## Make test data ms.data3 available
data(ms.data3)
attach(ms.data3)

# Display a Thompson-Howarth plot for a RSD of 7.5% and draw
# the limit for a confidence interval of 90%
thplot2(MS, name = "Magnetic Susceptibility", rsd = 7.5, ptile = 90,
        main = "Stanley (2003) Test Data", ifalt = TRUE)

## Detach test data ms.data3
detach(ms.data3)
```

`var2fact` *Rearranges Data for Variables as Factors*

Description

Rearranges data from a matrix or dataframe into a matrix where data are tagged by their variables names as factors. Used to concatenate data for display with functions `tbplot.by.var` and `bwplot.by.var`.

Usage

```
var2fact(xmat)
```

Arguments

xmat	name of the n x m data matrix or dataframe to be processed.

Details

If the data for only some of the variables available in an attached matrix or dataframe are to be processed use the cbind construct. Thus, `temp.mat <- cbind(varname1, varname3, varname6, varname8)`.

Value

xx	a `n * m` by 2 matrix where each of the `n * m` rows contains a value that is paired with its variable name as a factor, see Note below.

Note

The m variables for n cases results in a `n * m` by 2 matrix, where `[1:n, 1]` contains the variable name for value `[1]` and `[1:n, 2]` contains the values for the n rows in the first column of xmat. Then rows `[n+1:2n, 1]` contain the variable name for value `[2]` and `[n+1:2n, 2]` contain the values for n rows in the second column, and so on.

Author(s)

Robert G. Garrett

Examples

```
## Display, convert data frame and display the result
data(ms.data1)
ms.data1
temp <- var2fact(ms.data1)
temp
```

Index