

# Package ‘mvoutlier’

January 2, 2012

**Version** 1.9.4

**Date** 2011-11-07

**Title** Multivariate outlier detection based on robust methods

**Author** Peter Filzmoser <P.Filzmoser@tuwien.ac.at> and  
Moritz Gschwandtner <e0125439@student.tuwien.ac.at>

**Maintainer** Peter Filzmoser <P.Filzmoser@tuwien.ac.at>

**Depends** robustbase, compositions, robCompositions, R (>= 2.10)

**Description** This packages was made for multivariate outlier detection.

**License** GPL (>= 3)

**URL** <http://www.statistik.tuwien.ac.at/public/filz/>

**Repository** CRAN

**Date/Publication** 2011-11-07 14:05:30

## R topics documented:

aq.plot . . . . .	2
arw . . . . .	3
bhorizon . . . . .	5
bss.background . . . . .	7
bssbot . . . . .	8
bsstop . . . . .	10
chisq.plot . . . . .	12
chorizon . . . . .	13
color.plot . . . . .	17
cor.plot . . . . .	18
dd.plot . . . . .	19
humus . . . . .	20
kola.background . . . . .	22
map.plot . . . . .	23

moss . . . . .	24
mvoutlier.CoDa . . . . .	26
pbb . . . . .	28
pcout . . . . .	29
pkb . . . . .	31
plot.mvoutlierCoDa . . . . .	32
sign1 . . . . .	34
sign2 . . . . .	35
symbol.plot . . . . .	37
uni.plot . . . . .	38

<b>Index</b>	<b>41</b>
--------------	-----------

---

aq.plot	<i>Adjusted Quantile Plot</i>
---------	-------------------------------

---

## Description

The function `aq.plot` plots the ordered squared robust Mahalanobis distances of the observations against the empirical distribution function of the  $M\hat{D}^2_i$ . In addition the distribution function of  $\$chisq\_p\$$  is plotted as well as two vertical lines corresponding to the `chisq`-quantile specified in the argument list (default is 0.975) and the so-called adjusted quantile. Three additional graphics are created (the first showing the data, the second showing the outliers detected by the specified quantile of the  $\$chisq\_p\$$  distribution and the third showing these detected outliers by the adjusted quantile).

## Usage

```
aq.plot(x, delta=qchisq(0.975, df=ncol(x)), quan=1/2, alpha=0.05)
```

## Arguments

<code>x</code>	matrix or data.frame containing the data; has to be at least two-dimensional
<code>delta</code>	quantile of the chi-squared distribution with <code>ncol(x)</code> degrees of freedom. This quantile appears as cyan-colored vertical line in the plot.
<code>quan</code>	proportion of observations which are used for mcd estimations; has to be between 0.5 and 1, default ist 0.5
<code>alpha</code>	Maximum thresholding proportion (optional scalar, default: <code>alpha = 0.05</code> )

## Details

The function `aq.plot` plots the ordered squared robust Mahalanobis distances of the observations against the empirical distribution function of the  $M\hat{D}^2_i$ . The distance calculations are based on the MCD estimator.

For outlier detection two different methods are used. The first one marks observations as outliers if they exceed a certain quantile of the chi-squared distribution. The second is an adaptive procedure

searching for outliers specifically in the tails of the distribution, beginning at a certain chisq-quantile (see Filzmoser et al., 2005).

The function behaves differently depending on the dimension of the data. If the data is more than two-dimensional the data are projected on the first two robust principal components.

### Value

outliers            boolean vector of outliers

### Author(s)

Moritz Gschwandtner <<e0125439@student.tuwien.ac.at>>

Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

### References

P. Filzmoser, R.G. Garrett, and C. Reimann. Multivariate outlier detection in exploration geochemistry. *Computers & Geosciences*, 31:579-587, 2005.

### Examples

```
# create data:
set.seed(134)
x <- cbind(rnorm(80), rnorm(80), rnorm(80))
y <- cbind(rnorm(10, 5, 1), rnorm(10, 5, 1), rnorm(10, 5, 1))
z <- rbind(x,y)
# execute:
aq.plot(z, alpha=0.1)
```

### Description

Adaptive reweighted estimator for multivariate location and scatter with hard-rejection weights. The multivariate outliers are defined according to the supremum of the difference between the empirical distribution function of the robust Mahalanobis distance and the theoretical distribution function.

### Usage

```
arw(x, m0, c0, alpha, pcrit)
```

**Arguments**

x	Dataset (n x p)
m0	Initial location estimator (1 x p)
c0	Initial scatter estimator (p x p)
alpha	Maximum thresholding proportion (optional scalar, default: alpha = 0.025)
pcrit	Critical value obtained by simulations (optional scalar, default value obtained from simulations)

**Details**

At the basis of initial estimators of location and scatter, the function arw performs a reweighting step to adjust the threshold for outlier rejection. The critical value pcrit was obtained by simulations using the MCD estimator as initial robust covariance estimator. If a different estimator is used, pcrit should be changed and computed by simulations for the specific dimensions of the data x.

**Value**

m	Adaptive location estimator (p x 1)
c	Adaptive scatter estimator (p x p)
cn	Adaptive threshold ("adjusted quantile")
w	Weight vector (n x 1)

**Author(s)**

Moritz Gschwandtner <<e0125439@student.tuwien.ac.at>>

Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

**References**

P. Filzmoser, R.G. Garrett, and C. Reimann. Multivariate outlier detection in exploration geochemistry. *Computers & Geosciences*, 31:579-587, 2005.

**Examples**

```
x <- cbind(rnorm(100), rnorm(100))
arw(x, apply(x,2,mean), cov(x))
```

---

bhorizon	<i>B-horizon of the Kola Data</i>
----------	-----------------------------------

---

**Description**

The Kola data were collected in the Kola Project (1993-1998, Geological Surveys of Finland (GTK) and Norway (NGU) and Central Kola Expedition (CKE), Russia). More than 600 samples in five different layers were analysed, this dataset contains the B-horizon.

**Usage**

```
data(bhorizon)
```

**Format**

A data frame with 609 observations on the following 48 variables.

ID a numeric vector  
XC00 a numeric vector  
YC00 a numeric vector  
Ag a numeric vector  
Al a numeric vector  
Al\_XRF a numeric vector  
As a numeric vector  
Ba a numeric vector  
Be a numeric vector  
Bi a numeric vector  
Ca a numeric vector  
Ca\_XRF a numeric vector  
Cd a numeric vector  
Co a numeric vector  
Cr a numeric vector  
Cu a numeric vector  
EC a numeric vector  
Fe a numeric vector  
Fe\_XRF a numeric vector  
K a numeric vector  
K\_XRF a numeric vector  
LOI a numeric vector  
La a numeric vector

Li a numeric vector  
Mg a numeric vector  
Mg\_XRF a numeric vector  
Mn a numeric vector  
Mn\_XRF a numeric vector  
Mo a numeric vector  
Na a numeric vector  
Na\_XRF a numeric vector  
Ni a numeric vector  
P a numeric vector  
P\_XRF a numeric vector  
Pb a numeric vector  
S a numeric vector  
Sc a numeric vector  
Se a numeric vector  
Si a numeric vector  
Si\_XRF a numeric vector  
Sr a numeric vector  
Te a numeric vector  
Th a numeric vector  
Ti a numeric vector  
Ti\_XRF a numeric vector  
V a numeric vector  
Y a numeric vector  
Zn a numeric vector

### Source

Kola Project (1993-1998)

### References

Reimann C, Äyräs M, Chekushin V, Bogatyrev I, Boyd R, Caritat P de, Dutter R, Finne TE, Halleraker JH, Jæger Ø, Kashulina G, Lehto O, Niskavaara H, Pavlov V, Räsänen ML, Strand T, Volden T. Environmental Geochemical Atlas of the Central Barents Region. NGU-GTK-CKE Special Publication, Geological Survey of Norway, Trondheim, Norway, 1998.

### Examples

```
data(bhorizon)
# classical versus robust correlation
cor.plot(log(bhorizon[,"Al"]), log(bhorizon[,"Na"]))
```

---

bss.background	<i>Background map for the BSS project</i>
----------------	---

---

**Description**

Coordinates of the BSS data background map

**Usage**

```
data(bss.background)
```

**Format**

A data frame with 6093 observations on the following 2 variables.

V1 a numeric vector with the x-coordinates

V2 a numeric vector with the y-coordinates

**Details**

Is used by pbb()

**Source**

BSS project

**References**

Reimann C, Siewers U, Tarvainen T, Bityukova L, Eriksson J, Gilucis A, Gregorauskiene V, Lukashchuk VK, Matinian NN, Pasieczna A. Agricultural Soils in Northern Europe: A Geochemical Atlas. Geologisches Jahrbuch, Sonderhefte, Reihe D, Heft SD 5, Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, 2003.

**Examples**

```
data(bss.background)
pbb()
```

---

bssbot

*Bottom Layer of the BSS Data*

---

### Description

The BSS data were collected in agricultural soils from Northern Europe. from an area of about 1,800,000 km<sup>2</sup>. 769 samples on an irregular grid were taken in two different layers, the top layer (0-20cm) and the bottom layer. This dataset contains the bottom layer of the BSS data. It has 46 variables, including x and y coordinates.

### Usage

```
data(bssbot)
```

### Format

A data frame with 768 observations on the following 46 variables.

**ID** a numeric vector

**CNo** a numeric vector

**XCOO** x coordinates: a numeric vector

**YCOO** y coordinates: a numeric vector

**SiO2\\_B** a numeric vector

**TiO2\\_B** a numeric vector

**Al2O3\\_B** a numeric vector

**Fe2O3\\_B** a numeric vector

**MnO\\_B** a numeric vector

**MgO\\_B** a numeric vector

**CaO\\_B** a numeric vector

**Na2O\\_B** a numeric vector

**K2O\\_B** a numeric vector

**P2O5\\_B** a numeric vector

**SO3\\_B** a numeric vector

**Cl\\_B** a numeric vector

**F\\_B** a numeric vector

**LOI\\_B** a numeric vector

**As\\_B** a numeric vector

**Ba\\_B** a numeric vector

**Bi\\_B** a numeric vector

**Ce\\_B** a numeric vector

**Co\B** a numeric vector  
**Cr\B** a numeric vector  
**Cs\B** a numeric vector  
**Cu\B** a numeric vector  
**Ga\B** a numeric vector  
**Hf\B** a numeric vector  
**La\B** a numeric vector  
**Mo\B** a numeric vector  
**Nb\B** a numeric vector  
**Ni\B** a numeric vector  
**Pb\B** a numeric vector  
**Rb\B** a numeric vector  
**Sb\B** a numeric vector  
**Sc\B** a numeric vector  
**Sn\B** a numeric vector  
**Sr\B** a numeric vector  
**Ta\B** a numeric vector  
**Th\B** a numeric vector  
**U\B** a numeric vector  
**V\B** a numeric vector  
**W\B** a numeric vector  
**Y\B** a numeric vector  
**Zn\B** a numeric vector  
**Zr\B** a numeric vector

### Source

BSS Project in Northern Europe

### References

Reimann C, Siewers U, Tarvainen T, Bitjukova L, Eriksson J, Gilucis A, Gregorauskiene V, Lukashchuk VK, Matinian NN, Pasieczna A. Agricultural Soils in Northern Europe: A Geochemical Atlas. Geologisches Jahrbuch, Sonderhefte, Reihe D, Heft SD 5, Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, 2003.

### Examples

```
data(bssbot)
# classical versus robust correlation
cor.plot(log(bssbot[, "Al203_B"]), log(bssbot[, "Na20_B"]))
```

---

bsstop

*Top Layer of the BSS Data*

---

### Description

The BSS data were collected in agricultural soils from Northern Europe. from an area of about 1,800,000 km<sup>2</sup>. 769 samples on an irregular grid were taken in two different layers, the top layer (0-20cm) and the bottom layer. This dataset contains the top layer of the BSS data. It has 46 variables, including x and y coordinates.

### Usage

```
data(bsstop)
```

### Format

A data frame with 768 observations on the following 46 variables.

**ID** a numeric vector

**CNo** a numeric vector

**XCOO** x coordinates: a numeric vector

**YCOO** y coordinates: a numeric vector

**SiO2\\_T** a numeric vector

**TiO2\\_T** a numeric vector

**Al2O3\\_T** a numeric vector

**Fe2O3\\_T** a numeric vector

**MnO\\_T** a numeric vector

**MgO\\_T** a numeric vector

**CaO\\_T** a numeric vector

**Na2O\\_T** a numeric vector

**K2O\\_T** a numeric vector

**P2O5\\_T** a numeric vector

**SO3\\_T** a numeric vector

**Cl\\_T** a numeric vector

**F\\_T** a numeric vector

**LOI\\_T** a numeric vector

**As\\_T** a numeric vector

**Ba\\_T** a numeric vector

**Bi\\_T** a numeric vector

**Ce\\_T** a numeric vector

**Co**\\_T a numeric vector  
**Cr**\\_T a numeric vector  
**Cs**\\_T a numeric vector  
**Cu**\\_T a numeric vector  
**Ga**\\_T a numeric vector  
**Hf**\\_T a numeric vector  
**La**\\_T a numeric vector  
**Mo**\\_T a numeric vector  
**Nb**\\_T a numeric vector  
**Ni**\\_T a numeric vector  
**Pb**\\_T a numeric vector  
**Rb**\\_T a numeric vector  
**Sb**\\_T a numeric vector  
**Sc**\\_T a numeric vector  
**Sn**\\_T a numeric vector  
**Sr**\\_T a numeric vector  
**Ta**\\_T a numeric vector  
**Th**\\_T a numeric vector  
**U**\\_T a numeric vector  
**V**\\_T a numeric vector  
**W**\\_T a numeric vector  
**Y**\\_T a numeric vector  
**Zn**\\_T a numeric vector  
**Zr**\\_T a numeric vector

### Source

BSS Project in Northern Europe

### References

Reimann C, Siewers U, Tarvainen T, Bityukova L, Eriksson J, Gilucis A, Gregorauskiene V, Lukashchuk VK, Matinian NN, Pasiieczna A. Agricultural Soils in Northern Europe: A Geochemical Atlas. Geologisches Jahrbuch, Sonderhefte, Reihe D, Heft SD 5, Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, 2003.

### Examples

```
data(bsstop)
# classical versus robust correlation
cor.plot(log(bsstop[, "Al203_T"]), log(bsstop[, "Na20_T"]))
```

---

`chisq.plot`*Chi-Square Plot*

---

### Description

The function `chisq.plot` plots the ordered robust mahalanobis distances of the data against the quantiles of the Chi-squared distribution. By user interaction this plotting is iterated each time leaving out the observation with the greatest distance.

### Usage

```
chisq.plot(x, quan=1/2, ask=TRUE, ...)
```

### Arguments

<code>x</code>	matrix or <code>data.frame</code> containing the data
<code>quan</code>	amount of observations which are used for mcd estimations. has to be between 0.5 and 1, default ist 0.5
<code>ask</code>	logical. specifies whether user interacton is allowed or not. default is TRUE
<code>...</code>	additional graphical parameters

### Details

The function `chisq.plot` plots the ordered robust mahalanobis distances of the data against the quantiles of the Chi-squared distribution. If the data is normal distributed these values should approximately correspond to each other, so outliers can be detected visually. By user interaction this procedure is repeated, each time leaving out the observation with the greatest distance (the number of the observation is printed on the console). This method can be seen as an iterative deletion of outliers until a straight line appears.

### Value

`outliers` indices of the outliers that are removed by left-click on the plotting device.

### Author(s)

Moritz Gschwandtner <<e0125439@student.tuwien.ac.at>>

Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

### References

R.G. Garrett (1989). The chi-square plot: a tools for multivariate outlier recognition. *Journal of Geochemical Exploration*, 32 (1/3), 319-341.

**Examples**

```
data(humus)
res <-chisq.plot(log(humus[,c("Co", "Cu", "Ni")]))
res$outliers # these are the potential outliers
```

---

chorizon

*C-horizon of the Kola Data*

---

**Description**

The Kola Data were collected in the Kola Project (1993-1998, Geological Surveys of Finland (GTK) and Norway (NGU) and Central Kola Expedition (CKE), Russia). More than 600 samples in five different layers were analysed, this dataset contains the C-horizon.

**Usage**

```
data(chorizon)
```

**Format**

A data frame with 606 observations on the following 110 variables.

ID a numeric vector

XCO0 a numeric vector

YCO0 a numeric vector

Ag a numeric vector

Ag\_INAA a numeric vector

Al a numeric vector

Al203 a numeric vector

As a numeric vector

As\_INAA a numeric vector

Au\_INAA a numeric vector

B a numeric vector

Ba a numeric vector

Ba\_INAA a numeric vector

Be a numeric vector

Bi a numeric vector

Br\_IC a numeric vector

Br\_INAA a numeric vector

Ca a numeric vector

Ca\_INAA a numeric vector

CaO a numeric vector  
Cd a numeric vector  
Ce\_INAA a numeric vector  
Cl\_IC a numeric vector  
Co a numeric vector  
Co\_INAA a numeric vector  
EC a numeric vector  
Cr a numeric vector  
Cr\_INAA a numeric vector  
Cs\_INAA a numeric vector  
Cu a numeric vector  
Eu\_INAA a numeric vector  
F\_IC a numeric vector  
Fe a numeric vector  
Fe\_INAA a numeric vector  
Fe203 a numeric vector  
Hf\_INAA a numeric vector  
Hg a numeric vector  
Hg\_INAA a numeric vector  
Ir\_INAA a numeric vector  
K a numeric vector  
K20 a numeric vector  
La a numeric vector  
La\_INAA a numeric vector  
Li a numeric vector  
LOI a numeric vector  
Lu\_INAA a numeric vector  
wt\_INAA a numeric vector  
Mg a numeric vector  
MgO a numeric vector  
Mn a numeric vector  
MnO a numeric vector  
Mo a numeric vector  
Mo\_INAA a numeric vector  
Na a numeric vector  
Na\_INAA a numeric vector  
Na20 a numeric vector

Nd\_INAA a numeric vector  
Ni a numeric vector  
Ni\_INAA a numeric vector  
NO3\_IC a numeric vector  
P a numeric vector  
P205 a numeric vector  
Pb a numeric vector  
pH a numeric vector  
PO4\_IC a numeric vector  
Rb a numeric vector  
S a numeric vector  
Sb a numeric vector  
Sb\_INAA a numeric vector  
Sc a numeric vector  
Sc\_INAA a numeric vector  
Se a numeric vector  
Se\_INAA a numeric vector  
Si a numeric vector  
SiO2 a numeric vector  
Sm\_INAA a numeric vector  
Sn\_INAA a numeric vector  
SO4\_IC a numeric vector  
Sr a numeric vector  
Sr\_INAA a numeric vector  
SUM\_XRF a numeric vector  
Ta\_INAA a numeric vector  
Tb\_INAA a numeric vector  
Te a numeric vector  
Th a numeric vector  
Th\_INAA a numeric vector  
Ti a numeric vector  
TiO2 a numeric vector  
U\_INAA a numeric vector  
V a numeric vector  
W\_INAA a numeric vector  
Y a numeric vector  
Yb\_INAA a numeric vector

Zn a numeric vector  
Zn\_INAA a numeric vector  
ELEV a numeric vector  
COUN a numeric vector  
ASP a numeric vector  
TOPC a numeric vector  
LITO a numeric vector  
Al\_XRF a numeric vector  
Ca\_XRF a numeric vector  
Fe\_XRF a numeric vector  
K\_XRF a numeric vector  
Mg\_XRF a numeric vector  
Mn\_XRF a numeric vector  
Na\_XRF a numeric vector  
P\_XRF a numeric vector  
Si\_XRF a numeric vector  
Ti\_XRF a numeric vector

### Source

Kola Project (1993-1998)

### References

Reimann C, Äyräs M, Chekushin V, Bogatyrev I, Boyd R, Caritat P de, Dutter R, Finne TE, Halleraker JH, Jæger Ø, Kashulina G, Lehto O, Niskavaara H, Pavlov V, Räsänen ML, Strand T, Volden T. Environmental Geochemical Atlas of the Central Barents Region. NGU-GTK-CKE Special Publication, Geological Survey of Norway, Trondheim, Norway, 1998.

### Examples

```
data(chorizon)
# classical versus robust correlation
cor.plot(log(chorizon[,"Al"]), log(chorizon[,"Na"]))
```

---

`color.plot`*Color Plot*

---

### Description

The function `color.plot` plots the (two-dimensional) data using different symbols according to the robust mahalanobis distance based on the mcd estimator with adjustment and using different colors according to the euclidean distances of the observations.

### Usage

```
color.plot(x, quan=1/2, alpha=0.025, ...)
```

### Arguments

<code>x</code>	two dimensional matrix or data.frame containing the data.
<code>quan</code>	amount of observations which are used for mcd estimations. has to be between 0.5 and 1, default ist 0.5
<code>alpha</code>	amount of observations used for calculating the adjusted quantile (see function <code>arw</code> ).
<code>...</code>	additional graphical parameters

### Details

The function `color.plot` plots the (two-dimensional) data using different symbols (see function `symbol.plot`) according to the robust mahalanobis distance based on the mcd estimator with adjustment and using different colors according to the euclidean distances of the observations. Blue is typical for a little distance, whereas red is the opposite. In addition four ellipsoids are drawn, on which mahalanobis distances are constant. These constant values correspond to the 25%, 50%, 75% and adjusted quantiles (see function `arw`) of the chi-square distribution (see Filzmoser et al., 2005).

### Value

<code>outliers</code>	boolean vector of outliers
<code>md</code>	robust mahalanobis distances of the data
<code>euclidean</code>	euclidean distances of the observations according to the minimum of the data.

### Author(s)

Moritz Gschwandtner <<e0125439@student.tuwien.ac.at>>

Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

### References

P. Filzmoser, R.G. Garrett, and C. Reimann. Multivariate outlier detection in exploration geochemistry. *Computers & Geosciences*, 31:579-587, 2005.

**See Also**

[symbol.plot](#), [dd.plot](#), [arw](#)

**Examples**

```
# create data:
x <- cbind(rnorm(100), rnorm(100))
y <- cbind(rnorm(10, 5, 1), rnorm(10, 5, 1))
z <- rbind(x,y)
# execute:
color.plot(z, quan=0.75)
```

---

cor.plot

*Correlation Plot: robust versus classical bivariate correlation*

---

**Description**

The function `cor.plot` plots the (two-dimensional) data and adds two correlation ellipsoids, based on classical and robust estimation of location and scatter. Robust estimation can be thought of as estimating the mean and covariance of the 'good' part of the data.

**Usage**

```
cor.plot(x, y, quan=1/2, alpha=0.025, ...)
```

**Arguments**

<code>x</code>	vector to be plotted against <code>y</code> and of which the correlation with <code>y</code> is calculated.
<code>y</code>	vector to be plotted against <code>x</code> and of which the correlation with <code>x</code> is calculated.
<code>quan</code>	amount of observations which are used for mcd estimations. has to be between 0.5 and 1, default ist 0.5
<code>alpha</code>	Determines the size of the ellipsoids. An observation will be outside of the ellipsoid if its mahalanobis distance exceeds the 1-alpha quantile of the chi-squared distribution.
<code>...</code>	additional graphical parameters

**Value**

<code>cor.cla</code>	correlation between <code>x</code> and <code>y</code> based on classical estimation of location and scatter
<code>cor.rob</code>	correlation between <code>x</code> and <code>y</code> based on robust estimation of location and scatter

**Author(s)**

Moritz Gschwandtner <<e0125439@student.tuwien.ac.at>>  
 Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

**See Also**[covMcd](#)**Examples**

```
# create data:
x <- cbind(rnorm(100), rnorm(100))
y <- cbind(rnorm(10, 3, 1), rnorm(10, 3, 1))
z <- rbind(x,y)
# execute:
cor.plot(z[,1], z[,2])
```

dd.plot

*Distance-Distance Plot***Description**

The function `dd.plot` plots the classical mahalanobis distance of the data against the robust mahalanobis distance based on the mcd estimator. Different symbols (see function `symbol.plot`) and colours (see function `color.plot`) are used depending on the mahalanobis and euclidean distance of the observations (see Filzmoser et al., 2005).

**Usage**

```
dd.plot(x, quan=1/2, alpha=0.025, ...)
```

**Arguments**

<code>x</code>	matrix or data frame containing the data
<code>quan</code>	amount of observations which are used for mcd estimations. has to be between 0.5 and 1, default ist 0.5
<code>alpha</code>	amount of observations used for calculating the adjusted quantile (see function <code>arw</code> ).
<code>...</code>	additional graphical parameters

**Value**

<code>outliers</code>	boolean vector of outliers
<code>md.cla</code>	mahalanobis distances of the observations based on classical estimators of location and scatter.
<code>md.rob</code>	mahalanobis distances of the observations based on robust estimators of location and scatter (mcd).

**Author(s)**

Moritz Gschwandtner <<e0125439@student.tuwien.ac.at>>

Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

## References

P. Filzmoser, R.G. Garrett, and C. Reimann. Multivariate outlier detection in exploration geochemistry. *Computers & Geosciences*, 31:579-587, 2005.

## See Also

[symbol.plot](#), [color.plot](#), [arw](#), [covPlot](#)

## Examples

```
# create data:
x <- cbind(rnorm(100), rnorm(100))
y <- cbind(rnorm(10, 3, 1), rnorm(10, 3, 1))
z <- rbind(x,y)
# execute:
dd.plot(z)
#
# Identify multivariate outliers for Co-Cu-Ni in humus layer of Kola data:
data(humus)
dd.plot(log(humus[,c("Co", "Cu", "Ni")]))
```

---

humus

*Humus Layer (O-horizon) of the Kola Data*

---

## Description

The Kola Data were collected in the Kola Project (1993-1998, Geological Surveys of Finland (GTK) and Norway (NGU) and Central Kola Expedition (CKE), Russia). More than 600 samples in five different layers were analysed, this dataset contains the humus layer.

## Usage

```
data(humus)
```

## Format

A data frame with 617 observations on the following 44 variables.

ID a numeric vector  
XC00 a numeric vector  
YC00 a numeric vector  
Ag a numeric vector  
Al a numeric vector  
As a numeric vector  
B a numeric vector  
Ba a numeric vector

Be a numeric vector  
Bi a numeric vector  
Ca a numeric vector  
Cd a numeric vector  
Co a numeric vector  
Cr a numeric vector  
Cu a numeric vector  
Fe a numeric vector  
Hg a numeric vector  
K a numeric vector  
La a numeric vector  
Mg a numeric vector  
Mn a numeric vector  
Mo a numeric vector  
Na a numeric vector  
Ni a numeric vector  
P a numeric vector  
Pb a numeric vector  
Rb a numeric vector  
S a numeric vector  
Sb a numeric vector  
Sc a numeric vector  
Si a numeric vector  
Sr a numeric vector  
Th a numeric vector  
Tl a numeric vector  
U a numeric vector  
V a numeric vector  
Y a numeric vector  
Zn a numeric vector  
C a numeric vector  
H a numeric vector  
N a numeric vector  
LOI a numeric vector  
pH a numeric vector  
Cond a numeric vector

**Source**

Kola Project (1993-1998)

**References**

Reimann C, Äyräs M, Chekushin V, Bogatyrev I, Boyd R, Caritat P de, Dutter R, Finne TE, Halleraker JH, Jæger Ø, Kashulina G, Lehto O, Niskavaara H, Pavlov V, Räsänen ML, Strand T, Volden T. Environmental Geochemical Atlas of the Central Barents Region. NGU-GTK-CKE Special Publication, Geological Survey of Norway, Trondheim, Norway, 1998.

**Examples**

```
data(humus)
# classical versus robust correlation:
cor.plot(log(humus[, "Al"]), log(humus[, "Na"]))
```

---

kola.background	<i>Background map for the Kola project</i>
-----------------	--

---

**Description**

Coordinates of the Kola background map

**Usage**

```
data(kola.background)
```

**Format**

The format is: List of 4 \$ boundary: 'data.frame': 50 obs. of 2 variables: ..\$ V1: num [1:50] 388650 388160 386587 384035 383029 ... ..\$ V2: num [1:50] 7892400 7881248 7847303 7790797 7769214 ... \$ coast : 'data.frame': 6259 obs. of 2 variables: ..\$ V1: num [1:6259] 438431 439102 439102 439643 439643 ... ..\$ V2: num [1:6259] 7895619 7896495 7896495 7895800 7895542 ... \$ borders : 'data.frame': 504 obs. of 2 variables: ..\$ V1: num [1:504] 417575 417704 418890 420308 422731 ... ..\$ V2: num [1:504] 7612984 7612984 7613293 7614530 7615972 ... \$ lakes : 'data.frame': 6003 obs. of 2 variables: ..\$ V1: num [1:6003] 547972 546915 NA 547972 547172 ... ..\$ V2: num [1:6003] 7815109 7815599 NA 7815109 7813873 ...

**Details**

Is used by map.plot()

**Source**

Kola Project (1993-1998)

## References

Reimann C, Äyräs M, Chekushin V, Bogatyrev I, Boyd R, Caritat P de, Dutter R, Finne TE, Halleraker JH, Jæger Ø, Kashulina G, Lehto O, Niskavaara H, Pavlov V, Räisänen ML, Strand T, Volden T. Environmental Geochemical Atlas of the Central Barents Region. NGU-GTK-CKE Special Publication, Geological Survey of Norway, Trondheim, Norway, 1998.

## Examples

```
example(map.plot)
```

---

map.plot	<i>Plot Multivariate Outliers in a Map</i>
----------	--

---

## Description

The function map.plot creates a map using geographical (x,y)-coordinates. This is thought for spatially dependent data of which coordinates are available. Multivariate outliers are marked.

## Usage

```
map.plot(coord, data, quan=1/2, alpha=0.025, symb=FALSE, plotmap=TRUE, map="kola.background", which.ma
```

## Arguments

coord	(x,y)-coordinates of the data
data	matrix or data.frame containing the data.
quan	amount of observations which are used for mcd estimations. has to be between 0.5 and 1, default ist 0.5
alpha	amount of observations used for calculating the adjusted quantile (see function arw).
symb	logical for plotting special symbols (see details).
plotmap	logical for plotting the background map.
map	see plot.kola.background()
which.map	see plot.kola.background()
map.col	see plot.kola.background()
map.lwd	see plot.kola.background()
...	additional graphical parameters

## Details

The function map.plot shows multivariate outliers in a map. If symb=FALSE (default), only two colors and no special symbols are used to mark multivariate outliers (the outliers are marked red). If symb=TRUE different symbols and colors are used. The symbols (cross means big value, circle means little value) are selected according to the robust mahalanobis distance based on the adjusted mcd estimator (see function symbol.plot) Different colors (red means big value, blue means little value) according to the euclidean distances of the observations (see function color.plot) are used. For details see Filzmoser et al. (2005).

**Value**

outliers	boolean vector of outliers
md	robust mahalanobis distances of the data
euclidean	(only if symb=TRUE) euclidean distances of the observations according to the minimum of the data.

**Author(s)**

Moritz Gschwandtner <<e0125439@student.tuwien.ac.at>>

Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

**References**

P. Filzmoser, R.G. Garrett, and C. Reimann. Multivariate outlier detection in exploration geochemistry. *Computers & Geosciences*, 31:579-587, 2005.

**See Also**

[symbol.plot](#), [color.plot](#), [arw](#)

**Examples**

```
data(humus) # Load humus data
xy <- humus[,c("XC00", "YC00")] # X and Y Coordinates
myhumus <- log(humus[, c("As", "Cd", "Co", "Cu", "Mg", "Pb", "Zn")])
map.plot(xy, myhumus, symb=TRUE)
```

---

moss

*Moss Layer of the Kola Data*

---

**Description**

The Kola Data were collected in the Kola Project (1993-1998, Geological Surveys of Finland (GTK) and Norway (NGU) and Central Kola Expedition (CKE), Russia). More than 600 samples in five different layers were analysed, this dataset contains the moss layer.

**Usage**

```
data(moss)
```

**Format**

A data frame with 598 observations on the following 34 variables.

ID a numeric vector

XC00 a numeric vector

YC00 a numeric vector

Ag a numeric vector

Al a numeric vector

As a numeric vector

B a numeric vector

Ba a numeric vector

Bi a numeric vector

Ca a numeric vector

Cd a numeric vector

Co a numeric vector

Cr a numeric vector

Cu a numeric vector

Fe a numeric vector

Hg a numeric vector

K a numeric vector

Mg a numeric vector

Mn a numeric vector

Mo a numeric vector

Na a numeric vector

Ni a numeric vector

P a numeric vector

Pb a numeric vector

Rb a numeric vector

S a numeric vector

Sb a numeric vector

Si a numeric vector

Sr a numeric vector

Th a numeric vector

Tl a numeric vector

U a numeric vector

V a numeric vector

Zn a numeric vector

**Source**

Kola Project (1993-1998)

**References**

Reimann C, Åyräs M, Chekushin V, Bogatyrev I, Boyd R, Caritat P de, Dutter R, Finne TE, Halleraker JH, Jæger Ø, Kashulina G, Lehto O, Niskavaara H, Pavlov V, Räsänen ML, Strand T, Volden T. Environmental Geochemical Atlas of the Central Barents Region. NGU-GTK-CKE Special Publication, Geological Survey of Norway, Trondheim, Norway, 1998.

**Examples**

```
data(moss)
# classical versus robust correlation:
cor.plot(log(moss[,"Al"]), log(moss[,"Na"]))
```

---

mvoutlier.CoDa

*Interpreting multivariate outliers of CoDa*


---

**Description**

Computes the basis information for plot functions supporting the interpretation of multivariate outliers in case of compositional data.

**Usage**

```
mvoutlier.CoDa(x, quan = 0.75, alpha = 0.025, col.quantile = c(0, 0.05, 0.1, 0.5, 0.9, 0.95, 1), symb.pch
```

**Arguments**

x	data set (matrix or data frame) containing the raw untransformed compositional data
quan	quantity of data used for robust estimation; between 0.5 and 1
alpha	maximum threshold for adaptive outlier detection
col.quantile	quantiles of an average concentration defining the colors
symb.pch	plotting character for symbols
symb.cex	plotting size for symbols
adaptive	if TRUE then the adaptive method for the outlier threshold is used

## Details

In a first step, the raw compositional data set is transformed by the isometric logratio (ilr) transformation to the usual Euclidean space. Then adaptive outlier detection is performed: Starting from a quantile  $1-\alpha$  of the chi-square distribution, one looks for the supremum of the differences between the chi-square distribution and the empirical distribution of the squared Mahalanobis distances. The latter are derived from the MCD estimator using the proportion  $q$  of the data. The supremum is the outlier cutoff, and certain colors and symbols for the outliers are computed: The colors should reflect the magnitude of the median element concentration of the observations, which is done by computing for each observation along the single ilr variables the distances to the medians. The median of all distances determines the color (or grey scale): a high value, resulting in a red (or dark) symbol, means that most univariate parts have higher values than the average, and a low value (blue or light symbol) refers to an observation with mainly low values. The symbols are according to the cut-points from the quantiles 0.25, 0.5, 0.75, and the outlier cutoff of the squared Mahalanobis distances.

## Value

<code>ilrvariables</code>	the ilr transformed data matrix
<code>outliers</code>	TRUE/FALSE vector; TRUE refers to outlier
<code>pcaobj</code>	object from PCA
<code>colcol</code>	vector with the colors
<code>colbw</code>	vector with the grey scales
<code>pchvec</code>	vector with plotting symbols
<code>cexvec</code>	vector with sizes of plot symbols

## Author(s)

Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

## References

P. Filzmoser, K. Hron, and C. Reimann. Interpretation of multivariate outliers for compositional data. Submitted to Computers and Geosciences.

## See Also

[plot.mvoutlierCoDa](#), [arw](#), [map.plot](#), [uni.plot](#)

## Examples

```
data(humus)
d <- humus[,c("As", "Cd", "Co", "Cu", "Mg", "Pb", "Zn")]
res <- mvoutlier.CoDa(d)
str(res)
```

---

pbb

*BSS background Plot*

---

### Description

Plots the BSS background map

### Usage

```
pbb(map = "bss.background", add.plot = FALSE, ...)
```

### Arguments

map	List of coordinates. For the correct format see also <code>help(kola.background)</code>
add.plot	logical. If true background is added to an existing plot
...	additional plot parameters, see <code>help(par)</code>

### Details

The list of coordinates is plotted as a polygon line.

### Value

The plot is produced on the graphical device.

### Author(s)

Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

### References

Reimann C, Siewers U, Tarvainen T, Bityukova L, Eriksson J, Gilucis A, Gregorauskiene V, Lukashchuk VK, Matinian NN, Pasieczna A. Agricultural Soils in Northern Europe: A Geochemical Atlas. Geologisches Jahrbuch, Sonderhefte, Reihe D, Heft SD 5, Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, 2003.

### See Also

See also [pkb](#)

### Examples

```
data(bss.background)
data(bsstop)
plot(bsstop$XC00, bsstop$YC00, col="red", pch=3)
pbb(add=TRUE)
```

---

pcout

*PCOut Method for Outlier Identification in High Dimensions*


---

### Description

Fast algorithm for identifying multivariate outliers in high-dimensional and/or large datasets, using the algorithm of Filzmoser, Maronna, and Werner (CSDA, 2007).

### Usage

```
pcout(x, makeplot = FALSE, explvar = 0.99, crit.M1 = 1/3, crit.c1 = 2.5, crit.M2 = 1/4, crit.c2 = 0.99, cs
```

### Arguments

x	a numeric matrix or data frame which provides the data for outlier detection
makeplot	a logical value indicating whether a diagnostic plot should be generated (default to FALSE)
explvar	a numeric value between 0 and 1 indicating how much variance should be covered by the robust PCs (default to 0.99)
crit.M1	a numeric value between 0 and 1 indicating the quantile to be used as lower boundary for location outlier detection (default to 1/3)
crit.c1	a positive numeric value used for determining the upper boundary for location outlier detection (default to 2.5)
crit.M2	a numeric value between 0 and 1 indicating the quantile to be used as lower boundary for scatter outlier detection (default to 1/4)
crit.c2	a numeric value between 0 and 1 indicating the quantile to be used as upper boundary for scatter outlier detection (default to 0.99)
cs	a numeric value indicating the scaling constant for combined location and scatter weights (default to 0.25)
outbound	a numeric value between 0 and 1 indicating the outlier boundary for defining values as final outliers (default to 0.25)
...	additional plot parameters, see help(par)

### Details

Based on the robustly sphered data, semi-robust principal components are computed which are needed for determining distances for each observation. Separate weights for location and scatter outliers are computed based on these distances. The combined weights are used for outlier identification.

**Value**

wfinal01	0/1 vector with final weights for each observation; weight 0 indicates potential multivariate outliers.
wfinal	numeric vector with final weights for each observation; small values indicate potential multivariate outliers.
wloc	numeric vector with weights for each observation; small values indicate potential location outliers.
wscat	numeric vector with weights for each observation; small values indicate potential scatter outliers.
x.dist1	numeric vector with distances for location outlier detection.
x.dist2	numeric vector with distances for scatter outlier detection.
M1	upper boundary for assigning weight 1 in location outlier detection.
const1	lower boundary for assigning weight 0 in location outlier detection.
M2	upper boundary for assigning weight 1 in scatter outlier detection.
const2	lower boundary for assigning weight 0 in scatter outlier detection.

**Author(s)**

Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

**References**

P. Filzmoser, R. Maronna, M. Werner. Outlier identification in high dimensions, *Computational Statistics and Data Analysis*, 52, 1694-1711, 2008.

**See Also**

[sign1](#), [sign2](#)

**Examples**

```
# geochemical data from northern Europe
data(bsstop)
x=bsstop[,5:14]
# identify multivariate outliers
x.out=pcout(x,makeplot=FALSE)
# visualize multivariate outliers in the map
op <- par(mfrow=c(1,2))
data(bss.background)
pbb(asp=1)
points(bsstop$XC00,bsstop$YC00,pch=16,col=x.out$wfinal01+2)
title("Outlier detection based on pcout")
legend("topleft",legend=c("potential outliers","regular observations"),pch=16,col=c(2,3))

# compare with outlier detection based on MCD:
require(robustbase)
```

```
x.mcd=covMcd(x)
pbb(asp=1)
points(bsstop$XC00,bsstop$YC00,pch=16,col=x.mcd$mcd.wt+2)
title("Outlier detection based on MCD")
legend("topleft",legend=c("potential outliers","regular observations"),pch=16,col=c(2,3))
par(op)
```

pkb

*Kola background Plot***Description**

Plots the Kola background map

**Usage**

```
pkb(map = "kola.background", which.map = c(1, 2, 3, 4), map.col = c(5, 1, 3, 4), map.lwd = c(2, 1, 2, 1),
```

**Arguments**

map	List of coordinates. For the correct format see also <code>help(kola.background)</code>
which.map	which==1 ... plot project boundary \# which==2 ... plot coast line \# which==3 ... plot country borders \# which==4 ... plot lakes and rivers
map.col	Map colors to be used
map.lwd	Defines linestyle of the background
add.plot	logical. if true background is added to an existing plot
...	additional plot parameters, see <code>help(par)</code>

**Details**

Is used by `map.plot()`

**Author(s)**

Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

**References**

Reimann C, Åyräs M, Chekushin V, Bogatyrev I, Boyd R, Caritat P de, Dutter R, Finne TE, Halleraker JH, Jæger Ø, Kashulina G, Lehto O, Niskavaara H, Pavlov V, Räisänen ML, Strand T, Volden T. Environmental Geochemical Atlas of the Central Barents Region. NGU-GTK-CKE Special Publication, Geological Survey of Norway, Trondheim, Norway, 1998.

**Examples**

```
example(map.plot)
```

---

plot.mvoutlierCoDa      *Plots for interpreting multivariate outliers of CoDa*

---

### Description

Plots the computed information by `mvoutlier.CoDa` for supporting the interpretation of multivariate outliers in case of compositional data.

### Usage

```
## S3 method for class 'mvoutlierCoDa'
plot(x, ..., which = c("biplot", "map", "uni", "parallel"), choice = 1:2, coord = NULL, map = NULL, onlyout
```

### Arguments

<code>x</code>	resulting object from function <code>mvoutlier.CoDa</code>
<code>...</code>	further plotting arguments
<code>which</code>	type of plot that should be made
<code>choice</code>	select the pair of PCs used for the biplot
<code>coord</code>	coordinates for the presentation in a map
<code>map</code>	coordinates for the background map; see details below
<code>onlyout</code>	if TRUE only the outliers are shown in the plot
<code>bw</code>	if TRUE symbols will be in grey scale rather than in color
<code>symp</code>	if TRUE special symbols are used according to outlyingness
<code>symptxt</code>	if TRUE text labels are used for plotting
<code>col</code>	define colors to be used for outliers and non-outliers
<code>pch</code>	define plotting symbols to be used for outliers and non-outliers
<code>obj.cex</code>	define symbol size for outliers and non-outliers
<code>transp</code>	define transparency for parallel coordinate plot

### Details

The function `mvoutlier.CoDa` prepares the information needed for this plot function: In a first step, the raw compositional data set is transformed by the isometric logratio (ilr) transformation to the usual Euclidean space. Then adaptive outlier detection is performed: Starting from a quantile  $1-\alpha$  of the chisquare distribution, one looks for the supremum of the differences between the chisquare distribution and the empirical distribution of the squared Mahalanobis distances. The latter are derived from the MCD estimator using the proportion  $quan$  of the data. The supremum is the outlier cutoff, and certain colors and symbols for the outliers are computed: The colors should reflect the magnitude of the median element concentration of the observations, which is done by computing for each observation along the single ilr variables the distances to the medians. The median of all distances determines the color (or grey scale): a high value, resulting in a red (or dark) symbol, means that most univariate parts have higher values than the average, and a low value

(blue or light symbol) refers to an observation with mainly low values. The symbols are according to the cut-points from the quantiles 0.25, 0.5, 0.75, and the outlier cutoff of the squared Mahalanobis distances. This plot function then allows to visualize the information.

The optional background map for the representation of the outliers in a map can be included using the argument `map`. This should consist of one or more polygons representing the geographical x- and y-coordinates of the background map. Of course, this map should be represented in the same coordinate system as the coordinates for the sample locations provided by `coord`. The structure of `map` is as follows: It should consist of 2 columns, one for the x-, one for the y-coordinates. If a polygon ends, a row with 2 entries NA should follow. At the end two NA rows are needed. See also examples below.

### Value

A plot is drawn.

### Author(s)

Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

### References

P. Filzmoser, K. Hron, and C. Reimann. Interpretation of multivariate outliers for compositional data. Submitted to Computers and Geosciences.

### See Also

[mvoutlier.CoDa](#), [arw](#), [map.plot](#), [uni.plot](#)

### Examples

```
data(humus)
el=c("As", "Cd", "Co", "Cu", "Mg", "Pb", "Zn")
dsel <- humus[,el]
data(kola.background) # contains different information (coast, borders, etc.)
coo <- rbind(kola.background$coast, kola.background$boundary, kola.background$borders)
XY <- humus[,c("XC00", "YC00")]
set.seed(123)
res <- mvoutlier.CoDa(dsel)

par(ask=TRUE)
### Parallel coordinate plot:
## show for all observations (transp is only useful when generating e.g. a pdf):
# plot(res, onlyout=FALSE, bw=TRUE, which="parallel", symb=FALSE, symbtxt=FALSE, transp=0.3)
## show only outliers with special colors and labels in the margins:
plot(res, onlyout=TRUE, bw=FALSE, which="parallel", symb=TRUE, symbtxt=TRUE, transp=0.3)

### Biplot:
## show all data points, outliers are in different color and have different symbol:
# plot(res, onlyout=FALSE, which="biplot", bw=FALSE, symb=FALSE, symbtxt=FALSE)
## show only the outliers with special symbols and colors:
```

```

plot(res,onlyout=TRUE,which="biplot",bw=FALSE,symb=TRUE,symbtxt=TRUE)

### Map:
## show all data points, outliers are in different color and have different symbol:
# plot(res,coord=XY,map=coo,onlyout=FALSE,which="map",bw=FALSE,symb=FALSE,symbtxt=FALSE)
## show only the outliers with special symbols and colors:
plot(res,coord=XY,map=coo,onlyout=TRUE,which="map",bw=FALSE,symb=TRUE,symbtxt=TRUE)

### Univariate scatterplot:
## show all data points, outliers are in different color and have different symbol:
# plot(res,onlyout=FALSE,which="uni",symb=FALSE,symbtxt=FALSE)
## show only the outliers with special symbols and colors:
plot(res,onlyout=TRUE,which="uni",symb=TRUE,symbtxt=TRUE)

```

sign1

*Sign Method for Outlier Identification in High Dimensions - Simple Version*

## Description

Fast algorithm for identifying multivariate outliers in high-dimensional and/or large datasets, using spatial signs, see Filzmoser, Maronna, and Werner (CSDA, 2007). The computation of the distances is based on Mahalanobis distances.

## Usage

```
sign1(x, makeplot = FALSE, qcrit = 0.975, ...)
```

## Arguments

x	a numeric matrix or data frame which provides the data for outlier detection
makeplot	a logical value indicating whether a diagnostic plot should be generated (default to FALSE)
qcrit	a numeric value between 0 and 1 indicating the quantile to be used as critical value for outlier detection (default to 0.975)
...	additional plot parameters, see help(par)

## Details

Based on the robustly sphered and normed data, robust principal components are computed. These are used for computing the covariance matrix which is the basis for Mahalanobis distances. A critical value from the chi-square distribution is then used as outlier cutoff.

## Value

wfinal01	0/1 vector with final weights for each observation; weight 0 indicates potential multivariate outliers.
x.dist	numeric vector with distances used for outlier detection.
const	outlier cutoff value.

**Author(s)**

Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

**References**

P. Filzmoser, R. Maronna, M. Werner. Outlier identification in high dimensions, *Computational Statistics and Data Analysis*, 52, 1694-1711, 2008.

N. Locantore, J. Marron, D. Simpson, N. Tripoli, J. Zhang, and K. Cohen (1999). Robust principal components for functional data, *Test* 8, 1-73.

**See Also**

[pcout](#), [sign2](#)

**Examples**

```
# geochemical data from northern Europe
data(bsstop)
x=bsstop[,5:14]
# identify multivariate outliers
x.out=sign1(x,makeplot=FALSE)
# visualize multivariate outliers in the map
op <- par(mfrow=c(1,2))
data(bss.background)
pbb(asp=1)
points(bsstop$XC00,bsstop$YC00,pch=16,col=x.out$wfinal01+2)
title("Outlier detection based on signout")
legend("topleft",legend=c("potential outliers","regular observations"),pch=16,col=c(2,3))

# compare with outlier detection based on MCD:
require(robustbase)
x.mcd=covMcd(x)
pbb(asp=1)
points(bsstop$XC00,bsstop$YC00,pch=16,col=x.mcd$mcd.wt+2)
title("Outlier detection based on MCD")
legend("topleft",legend=c("potential outliers","regular observations"),pch=16,col=c(2,3))
par(op)
```

---

sign2

*Sign Method for Outlier Identification in High Dimensions - Sophisticated Version*

---

**Description**

Fast algorithm for identifying multivariate outliers in high-dimensional and/or large datasets, using spatial signs, see Filzmoser, Maronna, and Werner (CSDA, 2007). The computation of the distances is based on principal components.

**Usage**

```
sign2(x, makeplot = FALSE, explvar = 0.99, qcrit = 0.975, ...)
```

**Arguments**

x	a numeric matrix or data frame which provides the data for outlier detection
makeplot	a logical value indicating whether a diagnostic plot should be generated (default to FALSE)
explvar	a numeric value between 0 and 1 indicating how much variance should be covered by the robust PCs (default to 0.99)
qcrit	a numeric value between 0 and 1 indicating the quantile to be used as critical value for outlier detection (default to 0.975)
...	additional plot parameters, see help(par)

**Details**

Based on the robustly sphered and normed data, robust principal components are computed which are needed for determining distances for each observation. The distances are transformed to approach chi-square distribution, and a critical value is then used as outlier cutoff.

**Value**

wfinal01	0/1 vector with final weights for each observation; weight 0 indicates potential multivariate outliers.
x.dist	numeric vector with distances used for outlier detection.
const	outlier cutoff value.

**Author(s)**

Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

**References**

P. Filzmoser, R. Maronna, M. Werner. Outlier identification in high dimensions, *Computational Statistics and Data Analysis*, 52, 1694–1711, 2008.

N. Locantore, J. Marron, D. Simpson, N. Tripoli, J. Zhang, and K. Cohen. Robust principal components for functional data, *Test* 8, 1-73, 1999.

**See Also**

[pcout](#), [sign1](#)

**Examples**

```

# geochemical data from northern Europe
data(bsstop)
x=bsstop[,5:14]
# identify multivariate outliers
x.out=sign2(x,makeplot=FALSE)
# visualize multivariate outliers in the map
op <- par(mfrow=c(1,2))
data(bss.background)
pbb(asp=1)
points(bsstop$XC00,bsstop$YC00,pch=16,col=x.out$wfinal01+2)
title("Outlier detection based on signout")
legend("topleft",legend=c("potential outliers","regular observations"),pch=16,col=c(2,3))

# compare with outlier detection based on MCD:
require(robustbase)
x.mcd=covMcd(x)
pbb(asp=1)
points(bsstop$XC00,bsstop$YC00,pch=16,col=x.mcd$mcd.wt+2)
title("Outlier detection based on MCD")
legend("topleft",legend=c("potential outliers","regular observations"),pch=16,col=c(2,3))
par(op)

```

---

symbol.plot

*Symbol Plot*


---

**Description**

The function `symbol.plot` plots the (two-dimensional) data using different symbols according to the robust mahalanobis distance based on the mcd estimator with adjustment.

**Usage**

```
symbol.plot(x, quan=1/2, alpha=0.025, ...)
```

**Arguments**

<code>x</code>	two dimensional matrix or data.frame containing the data.
<code>quan</code>	amount of observations which are used for mcd estimations. has to be between 0.5 and 1, default ist 0.5
<code>alpha</code>	amount of observations used for calculating the adjusted quantile (see function <code>arw</code> ).
<code>...</code>	additional graphical parameters

**Details**

The function `symbol.plot` plots the (two-dimensional) data using different symbols. In addition a legend and four ellipsoids are drawn, on which mahalanobis distances are constant. As the legend shows, these constant values correspond to the 25%, 50%, 75% and adjusted (see function `arw`) quantiles of the chi-square distribution.

**Value**

<code>outliers</code>	boolean vector of outliers
<code>md</code>	robust mahalanobis distances of the data

**Author(s)**

Moritz Gschwandtner <<e0125439@student.tuwien.ac.at>>  
 Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

**References**

P. Filzmoser, R.G. Garrett, and C. Reimann. Multivariate outlier detection in exploration geochemistry. *Computers & Geosciences*, 31:579-587, 2005.

**See Also**

[dd.plot](#), [color.plot](#), [arw](#)

**Examples**

```
# create data:
x <- cbind(rnorm(100), rnorm(100))
y <- cbind(rnorm(10, 5, 1), rnorm(10, 5, 1))
z <- rbind(x,y)
# execute:
symbol.plot(z, quan=0.75)
```

---

`uni.plot`

*Univariate Presentation of Multivariate Outliers*

---

**Description**

The function `uni.plot` plots each variable of `x` parallel in a one-dimensional scatter plot and in addition marks multivariate outliers.

**Usage**

```
uni.plot(x, symb=FALSE, quan=1/2, alpha=0.025, ...)
```

**Arguments**

x	matrix or data.frame containing the data.
symb	logical. if FALSE, only two colors and no special symbols are used. outliers are marked red. if TRUE different symbols (cross means big value, circle means little value) according to the robust mahalanobis distance based on the mcd estimator and different colors (red means big value, blue means little value) according to the euclidean distances of the observations are used.
quan	amount of observations which are used for mcd estimations. has to be between 0.5 and 1, default ist 0.5
alpha	amount of observations used for calculating the adjusted quantile (see function arw).
...	additional graphical parameters

**Details**

The function uni.plot shows the multivariate outliers in the single variables by one-dimensional scatter plots. If symb=FALSE (default), only two colors and no special symbols are used to mark multivariate outliers (the outliers are marked red). If symb=TRUE different symbols and colors are used. The symbols (cross means big value, circle means little value) are selected according to the robust mahalanobis distance based on the adjusted mcd estimator (see function symbol.plot) Different colors (red means big value, blue means little value) according to the euclidean distances of the observations (see function color.plot) are used. For details see Filzmoser et al. (2005).

**Value**

outliers	boolean vector of outliers
md	robust multivariate mahalanobis distances of the data
euclidean	(only if symb=TRUE) multivariate euclidean distances of the observations according to the minimum of the data.

**Author(s)**

Moritz Gschwandtner <<e0125439@student.tuwien.ac.at>>

Peter Filzmoser <<P.Filzmoser@tuwien.ac.at>> <http://www.statistik.tuwien.ac.at/public/filz/>

**References**

P. Filzmoser, R.G. Garrett, and C. Reimann. Multivariate outlier detection in exploration geochemistry. *Computers & Geosciences*, 31:579-587, 2005.

**See Also**

[map.plot](#), [symbol.plot](#), [color.plot](#), [arw](#)

**Examples**

```
data(swiss)
uni.plot(swiss)
#
# Geostatistical data:
data(humus) # Load humus data
uni.plot(log(humus[, c("As", "Cd", "Co", "Cu", "Mg", "Pb", "Zn")]), symb=TRUE)
```

# Index

## \*Topic **datasets**

bhorizon, 5  
bss.background, 7  
bssbot, 8  
bsstop, 10  
chorizon, 13  
humus, 20  
kola.background, 22  
moss, 24  
pbb, 28  
pkb, 31

## \*Topic **dplot**

aq.plot, 2  
arw, 3  
chisq.plot, 12  
color.plot, 17  
cor.plot, 18  
dd.plot, 19  
map.plot, 23  
symbol.plot, 37  
uni.plot, 38

## \*Topic **multivariate**

mvoutlier.CoDa, 26  
pcout, 29  
plot.mvoutlierCoDa, 32  
sign1, 34  
sign2, 35

## \*Topic **robust**

mvoutlier.CoDa, 26  
pcout, 29  
plot.mvoutlierCoDa, 32  
sign1, 34  
sign2, 35

aq.plot, 2  
arw, 3, 18, 20, 24, 27, 33, 38, 39

bhorizon, 5  
bss.background, 7  
bssbot, 8

bsstop, 10  
chisq.plot, 12  
chorizon, 13  
color.plot, 17, 20, 24, 38, 39  
cor.plot, 18  
covMcd, 19  
covPlot, 20  
dd.plot, 18, 19, 38  
humus, 20  
kola.background, 22  
map.plot, 23, 27, 33, 39  
moss, 24  
mvoutlier.CoDa, 26, 33  
pbb, 28  
pcout, 29, 35, 36  
pkb, 28, 31  
plot.mvoutlierCoDa, 27, 32  
sign1, 30, 34, 36  
sign2, 30, 35, 35  
symbol.plot, 18, 20, 24, 37, 39  
uni.plot, 27, 33, 38