Package 'maptree'

October 13, 2022

Version 1.4-8

Date 2022-04-03

Title Mapping, Pruning, and Graphing Tree Models

Author Denis White, Robert B. Gramacy <rbg@vt.edu>

Maintainer Robert B. Gramacy <rbg@vt.edu>

Depends R (>= 2.14), cluster, rpart

Description Functions with example data for graphing, pruning, and
mapping models from hierarchical clustering, and classification
and regression trees.

License Unlimited

Repository CRAN

Date/Publication 2022-04-06 11:52:39 UTC

NeedsCompilation no

R topics documented:

---

clip.clust	*Prunes a Hierarchical Cluster Tree*

Description

Reduces a hierarchical cluster tree to a smaller tree either by pruning until a given number of observation groups remain, or by pruning tree splits below a given height.

Usage

```
clip.clust (cluster, data=NULL, k=NULL, h=NULL)
```

Arguments

cluster   object of class hclust or twins.

data   clustered dataset for hclust application.

k   desired number of groups.

h   height at which to prune for grouping.

                             At least one of k or h must be specified; k takes precedence if both are given.

Details

Used with draw.clust. See example.

Value

Pruned cluster object of class hclust.

Author(s)

Denis White

See Also

hclust, twins.object, cutree, draw.clust

Examples

```
library (cluster)
data (oregon.bird.dist)

draw.clust (clip.clust (agnes (oregon.bird.dist), k=6))
```

---

clip.rpart	*Prunes an Rpart Classification or Regression Tree*

Description

Reduces a prediction tree produced by rpart to a smaller tree by specifying either a cost-complexity parameter, or a number of nodes to which to prune.

Usage

```
clip.rpart (tree, cp=NULL, best=NULL)
```

Arguments

tree	object of class rpart.
cp	cost-complexity parameter.
best	number of nodes to which to prune.
	If both cp and best are not NULL, then cp is used.

Details

A minor enhancement of the existing prune.rpart to incorporate the parameter best as it is used in the (now defunct) prune.tree function in the old **tree** package. See example.

Value

Pruned tree object of class rpart.

Author(s)

Denis White

See Also

rpart, prune.rpart

Examples

```
library (rpart)
data (oregon.env.vars, oregon.border, oregon.grid)

draw.tree (clip.rpart (rpart (oregon.env.vars), best=7),
  nodeinfo=TRUE, units="species", cases="cells", digits=0)

group <- group.tree (clip.rpart (rpart (oregon.env.vars), best=7))
names(group) <- row.names(oregon.env.vars)
map.groups (oregon.grid, group)
lines (oregon.border)
```

```
map.key (0.05, 0.65, labels=as.character(seq(6)),
    size=1, new=FALSE, sep=0.5, pch=19, head="node")
```

---

draw.clust   *Graph a Hierarchical Cluster Tree*

Description

Graph a hierarchical cluster tree of class twins or hclust using colored symbols at observations.

Usage

```
draw.clust (cluster, data=NULL, cex=par("cex"), pch=par("pch"), size=2.5*cex,
    col=NULL, nodeinfo=FALSE, cases="obs", new=TRUE)
```

Arguments

cluster	object of class hclust or twins.
data	clustered dataset for hclust application.
cex	size of text, par parameter.
pch	shape of symbol at leaves, par parameter.
size	size in cex units of symbol at leaves.
col	vector of colors from hsv, rgb, etc, or if NULL, then use rainbow.
nodeinfo	if TRUE, add a line at each node with number of observations included in each leaf.
cases	label for type of observations.
new	if TRUE, call plot.new.

Details

An alternative to pltree and plot.hclust.

Value

The vector of colors supplied or generated.

Author(s)

Denis White

See Also

agnes, diana, hclust, draw.tree, map.groups

Examples

```
library (cluster)
data (oregon.bird.dist)

draw.clust (clip.clust (agnes (oregon.bird.dist), k=6))
```

---

draw.tree	*Graph a Classification or Regression Tree*

Description

Graph a classification or regression tree with a hierarchical tree diagram, optionally including colored symbols at leaves and additional info at intermediate nodes.

Usage

```
draw.tree (tree, cex=par("cex"), pch=par("pch"), size=2.5*cex,
    col=NULL, nodeinfo=FALSE, units="", cases="obs",
    digits=getOption("digits"), print.levels=TRUE,
    new=TRUE)
```

Arguments

tree	object of class rpart or tree.
cex	size of text, par parameter.
pch	shape of symbol at leaves, par parameter.
size	if size=0, draw terminal symbol at leaves else a symbol of size in cex units.
col	vector of colors from hsv, rgb, etc, or if NULL, then use rainbow.
nodeinfo	if TRUE, add a line at each node with mean value of response, number of observations, and percent deviance explained (or classified correct).
units	label for units of mean value of response, if regression tree.
cases	label for type of observations.
digits	number of digits to round mean value of response, if regression tree.
print.levels	if TRUE, print levels of factors at splits, otherwise only the factor name.
new	if TRUE, call plot.new.

Details

As in plot.rpart(,uniform=TRUE), each level has constant depth. Specifying nodeinfo=TRUE, shows the deviance explained or the classification rate at each node.

A split is shown, for numerical variables, as variable <> value when the cases with lower values go left, or as variable >< value when the cases with lower values go right. When the splitting variable is a factor, and print.levels=TRUE, the split is shown as levels = factor = levels with the cases on the left having factor levels equal to those on the left of the factor name, and correspondingly for the right.

Value

The vector of colors supplied or generated.

Author(s)

Denis White

See Also

[rpart](), [draw.clust](), [map.groups]()

Examples

```
library (rpart)
data (oregon.env.vars)

draw.tree (clip.rpart (rpart (oregon.env.vars), best=7),
    nodeinfo=TRUE, units="species", cases="cells", digits=0)
```

---

group.clust   *Observation Groups for a Hierarchical Cluster Tree*

Description

Alternative to [cutree]() that orders pruned groups from left to right in draw order.

Usage

```
group.clust (cluster, k=NULL, h=NULL)
```

Arguments

cluster	object of class hclust or twins.
k	desired number of groups.
h	height at which to prune for grouping.
	At least one of k or h must be specified; k takes precedence if both are given.

Details

Normally used with [map.groups](). See example.

Value

Vector of pruned cluster membership

Author(s)

Denis White

See Also

[hclust](), [twins.object](), [cutree](), [map.groups]()

Examples

```
data (oregon.bird.dist, oregon.grid)

group <- group.clust (hclust (dist (oregon.bird.dist)), k=6)
names(group) <- row.names(oregon.bird.dist)
map.groups (oregon.grid, group)
```

---

group.tree	*Observation Groups for Classification or Regression Tree*

Description

Alternative to tree[["where"]] that orders groups from left to right in draw order.

Usage

```
group.tree (tree)
```

Arguments

tree   object of class rpart or tree.

Details

Normally used with [map.groups.]() See example.

Value

Vector of rearranged tree[["where"]]

Author(s)

Denis White

See Also

[rpart](), [map.groups]()

Examples

```
library (rpart)
data (oregon.env.vars, oregon.grid)

group <- group.tree (clip.rpart (rpart (oregon.env.vars), best=7))
names(group) <- row.names(oregon.env.vars)
map.groups (oregon.grid, group=group)
```

---

kgs   *KGS Measure for Pruning Hierarchical Clusters*

Description

Computes the Kelley-Gardner-Sutcliffe penalty function for a hierarchical cluster tree.

Usage

```
kgs (cluster, diss, alpha=1, maxclust=NULL)
```

Arguments

cluster	object of class hclust or twins.
diss	object of class dissimilarity or dist.
alpha	weight for number of clusters.
maxclust	maximum number of clusters for which to compute measure.

Details

Kelley et al. (see reference) proposed a method that can help decide where to prune a hierarchical cluster tree. At any level of the tree the mean across all clusters of the mean within clusters of the dissimilarity measure is calculated. After normalizing, the number of clusters times alpha is added. The minimum of this function corresponds to the suggested pruning size.

The current implementation has complexity $O(n*n*maxclust)$, thus very slow with large n. For improvements, at least it should only calculate the spread for clusters that are split at each level, rather than over again for all.

Value

Vector of the penalty function for trees of size 2:maxclust. The names of vector elements are the respective numbers of clusters.

Author(s)

Denis White

References

Kelley, L.A., Gardner, S.P., Sutcliffe, M.J. (1996) An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally-related subfamilies, *Protein Engineering*, **9**, 1063-1065.

See Also

twins.object, dissimilarity.object, hclust, dist, clip.clust,

Examples

```
library (cluster)
data (votes.repub)

a <- agnes (votes.repub, method="ward")
b <- kgs (a, a$diss, maxclust=20)
plot (names (b), b, xlab="# clusters", ylab="penalty")
```

---

map.groups   *Map Groups of Observations*

Description

Draws maps of groups of observations created by clustering, classification or regression trees, or some other type of classification.

Usage

```
map.groups (pts, group, pch=par("pch"), size=2, col=NULL,
    border=NULL, new=TRUE)
```

Arguments

pts	matrix or data frame with components "x", and "y" for each observation (see details).
group	vector of integer class numbers corresponding to pts (see details), and indexing colors in col.
pch	symbol number from par("pch") if < 100, otherwise parameter n for ngon.
size	size in cex units of point symbol.
col	vector of fill colors from hsv, rgb, etc, or if NULL, then use rainbow.
border	vector of border colors from hsv, rgb, etc, or if NULL, then use rainbow.
new	if TRUE, call plot.new.

Details

If the number of rows of pts is not equal to the length of group, then (1) pts are assumed to represent polygons and polygon is used, (2) the identifiers in group are matched to the polygons in pts through names(group) and pts$x[is.na(pts$y)], and (3) these identifiers are mapped to dense integers to reference colours. Otherwise, group is assumed to parallel pts, and, if pch < 100, then points is used, otherwise ngon, to draw shaded polygon symbols for each observation in pts.

Value

The vector of fill colors supplied or generated.

Author(s)

Denis White

See Also

[ngon](), [polygon](), [group.clust](), [group.tree](), [map.key]()

Examples

```
data (oregon.bird.names, oregon.env.vars, oregon.bird.dist)
data (oregon.border, oregon.grid)

# range map for American Avocet
spp <- match ("American avocet", oregon.bird.names[["common.name"]])
group <- oregon.bird.dist[,spp] + 1
names(group) <- row.names(oregon.bird.dist)
kol <- gray (seq(0.8,0.2,length.out=length (table (group))))
map.groups (oregon.grid, group=group, col=kol)
lines (oregon.border)

# distribution of January temperatures
cuts <- quantile (oregon.env.vars[["jan.temp"]], probs=seq(0,1,1/5))
group <- cut (oregon.env.vars[["jan.temp"]], cuts, labels=FALSE,
  include.lowest=TRUE)
names(group) <- row.names(oregon.env.vars)
kol <- gray (seq(0.8,0.2,length.out=length (table (group))))
map.groups (oregon.grid, group=group, col=kol)
lines (oregon.border)

# January temperatures using point symbols rather than polygons
map.groups (oregon.env.vars, group, col=kol, pch=19)
lines (oregon.border)
```

---

map.key   *Draw Key to accompany Map of Groups*

Description

Draws legends for maps of groups of observations.

Usage

```
map.key (x, y, labels=NULL, cex=par("cex"), pch=par("pch"),
    size=2.5*cex, col=NULL, head="", sep=0.25*cex, new=FALSE)
```

Arguments

x, y	coordinates of lower left position of key in proportional units (0-1) of plot.
labels	vector of labels for classes, or if NULL, then integers 1:length(col), or 1.
size	size in cex units of shaded key symbol.
pch	symbol number for par if < 100, otherwise parameter n for ngon.
cex	pointsize of text, par parameter.
head	text heading for key.
sep	separation in cex units between adjacent symbols in key. If sep=0, assume a continuous scale, use square symbols, and put labels at breaks between squares.
col	vector of colors from hsv, rgb, etc, or if NULL, then use rainbow.
new	if TRUE, call plot.

Details

Uses points or ngon, depending on value of pch, to draw shaded polygon symbols for key.

Value

The vector of colors supplied or generated.

Author(s)

Denis White

See Also

ngon, map.groups

Examples

```
data (oregon.env.vars)

# key for examples in help(map.groups)
# range map for American Avocet
kol <- gray (seq(0.8,0.2,length.out=2))
map.key (0.2, 0.2, labels=c("absent","present"), pch=106,
  col=kol, head="key", new=TRUE)
# distribution of January temperatures
cuts <- quantile (oregon.env.vars[["jan.temp"]], probs=seq(0,1,1/5))
kol <- gray (seq(0.8,0.2,length.out=5))
map.key (0.2, 0.2, labels=as.character(round(cuts,0)),
  col=kol, sep=0, head="key", new=TRUE)

# key for example in help file for group.tree
map.key (0.2, 0.2, labels=as.character(seq(6)),
  pch=19, head="node", new=TRUE)
```

---

ngon   *Outline or Fill a Regular Polygon*

Description

Draws a regular polygon at specified coordinates as an outline or shaded.

Usage

```
ngon (xydc, n=4, angle=0, type=1)
```

Arguments

xydc	four element vector with x and y coordinates of center, d diameter in mm, and c color.
n	number of sides for polygon (>8 => circle).
angle	rotation angle of figure, in degrees.
type	type=1 => interior filled, type=2 => edge, type=3 => both.

Details

Uses [polygon](#) to draw shaded polygons and [lines](#) for outline. If n is odd, there is a vertex at (0, d/2), otherwise the midpoint of a side is at (0, d/2).

Value

Invisible.

Author(s)

Denis White

See Also

[polygon](#), [lines](#), [map.key](#), [map.groups](#)

Examples

```
plot (c(0,1), c(0,1), type="n")
ngon (c(.5, .5, 10, "blue"), angle=30, n=3)
apply (cbind (runif(8), runif(8), 6, 2), 1, ngon)
```

---

oregon.bird.dist *Presence/Absence of Bird Species in Oregon, USA*

Description

Binary matrix (1 = present) for distributions of 248 native breeding bird species for 389 grid cells in Oregon, USA.

Usage

```
data (oregon.bird.dist)
```

Format

A data frame with 389 rows and 248 columns.

Details

Row names are hexagon identifiers from White et al. (1992). Column names are species element codes developed by The Nature Conservancy (TNC), the Oregon Natural Heritage Program (ONHP), and NatureServe.

Source

Denis White

References

Master, L. (1996) Predicting distributions for vertebrate species: some observations, *Gap Analysis: A Landscape Approach to Biodiversity Planning*, Scott, J.M., Tear, T.H., and Davis, F.W., editors, American Society for Photogrammetry and Remote Sensing, Bethesda, MD, pp. 171-176.

White, D., Preston, E.M., Freemark, K.E., Kiester, A.R. (1999) A hierarchical framework for conserving biodiversity, *Landscape ecological analysis: issues and applications*, Klopatek, J.M., Gardner, R.H., editors, Springer-Verlag, pp. 127-153.

White, D., Kimerling, A.J., Overton, W.S. (1992) Cartographic and geometric components of a global sampling design for environmental monitoring, *Cartography and Geographic Information Systems*, **19**(1), 5-22.

TNC, https://www.nature.org/en-us/

ONHP, https://inr.oregonstate.edu/orbic/

NatureServe, https://www.natureserve.org/

See Also

oregon.env.vars, oregon.bird.names, oregon.grid, oregon.border

---

oregon.bird.names   *Names of Bird Species in Oregon, USA*

Description

Scientific and common names for 248 native breeding bird species in Oregon, USA.

Usage

```
data (oregon.bird.names)
```

Format

A data frame with 248 rows and 2 columns.

Details

Row names are species element codes. Columns are "scientific.name" and "common.name". Data are provided by The Nature Conservancy (TNC), the Oregon Natural Heritage Program (ONHP), and NatureServe.

Source

Denis White

References

Master, L. (1996) Predicting distributions for vertebrate species: some observations, *Gap Analysis: A Landscape Approach to Biodiversity Planning*, Scott, J.M., Tear, T.H., and Davis, F.W., editors, American Society for Photogrammetry and Remote Sensing, Bethesda, MD, pp. 171-176.

TNC, https://www.nature.org/en-us/

ONHP, https://inr.oregonstate.edu/orbic/

NatureServe, https://www.natureserve.org/

See Also

oregon.bird.dist

---

oregon.border   *Boundary of Oregon, USA*

Description

The boundary of the state of Oregon, USA, in lines format.

Usage

```
data (oregon.border)
```

Format

A data frame with 485 rows and 2 columns (the components "x" and "y").

Details

The map projection for this boundary, as well as the point coordinates in oregon.env.vars, is the Lambert Conformal Conic with standard parallels at 33 and 45 degrees North latitude, with the longitude of the central meridian at 120 degrees, 30 minutes West longitude, and with the projection origin latitude at 41 degrees, 45 minutes North latitude.

Source

Denis White

---

oregon.env.vars   *Environmental Variables for Oregon, USA*

Description

Distributions of 10 environmental variables for 389 grid cells in Oregon, USA.

Usage

```
data (oregon.env.vars)
```

Format

A data frame with 389 rows and 10 columns.

Details

Row names are hexagon identifiers from White et al. (1992). Variables (columns) are

bird.spp	number of native breeding bird species
x	x coordinate of center of grid cell
y	y coordinate of center of grid cell
jan.temp	mean minimum January temperature (C)
jul.temp	mean maximum July temperature (C)
rng.temp	mean difference between July and January temperatures (C)
ann.ppt	mean annual precipitation (mm)
min.elev	minimum elevation (m)
rng.elev	range of elevation (m)
max.slope	maximum slope (percent)

Source

Denis White

References

White, D., Preston, E.M., Freemark, K.E., Kiester, A.R. (1999) A hierarchical framework for conserving biodiversity, *Landscape ecological analysis: issues and applications*, Klopatek, J.M., Gardner, R.H., editors, Springer-Verlag, pp. 127-153.

White, D., Kimerling, A.J., Overton, W.S. (1992) Cartographic and geometric components of a global sampling design for environmental monitoring, *Cartography and Geographic Information Systems*, **19**(1), 5-22.

See Also

oregon.bird.dist, oregon.grid, oregon.border

---

oregon.grid	*Hexagonal Grid Cell Polygons covering Oregon, USA*

Description

Polygon borders for 389 hexagonal grid cells covering Oregon, USA, in polygon format.

Usage

```
data (oregon.grid)
```

Format

A data frame with 3112 rows and 2 columns (the components "x" and "y").

Details

The polygon format used for these grid cell boundaries is a slight variation from the standard R/S format. Each cell polygon is described by seven coordinate pairs, the last repeating the first. Prior to the first coordinate pair of each cell is a row containing NA in the ″y″ column and, in the ″x″ column, an identifier for the cell. The identifiers are the same as the row names in `oregon.bird.dist` and `oregon.env.vars`. See `map.groups` for how the linkage is made in mapping.

These grid cells are extracted from a larger set covering the conterminous United States and adjacent parts of Canada and Mexico, as described in White et al. (1992). Only cells with at least 50 percent of their area contained within the state of Oregon are included.

The map projection for the coordinates, as well as the point coordinates in `oregon.env.vars`, is the Lambert Conformal Conic with standard parallels at 33 and 45 degrees North latitude, with the longitude of the central meridian at 120 degrees, 30 minutes West longitude, and with the projection origin latitude at 41 degrees, 45 minutes North latitude.

Source

Denis White

References

White, D., Kimerling, A.J., Overton, W.S. (1992) Cartographic and geometric components of a global sampling design for environmental monitoring, *Cartography and Geographic Information Systems*, **19**(1), 5-22.

---

twins.to.hclust *Converts agnes or diana object to hclust object*

Description

Alternative to `as.hclust` that retains cluster data.

Usage

```
twins.to.hclust (cluster)
```

Arguments

cluster   object of class `twins`.

Details

Used internally in with `clip.clust` and `draw.clust`.

Value

hclust object

Author(s)

Denis White

See Also

hclust, twins.object

Index