

Package ‘HVT’

November 19, 2023

Type Package

Date 2023-11-17

Title Constructing Hierarchical Voronoi Tessellations and Overlay
Heatmaps for Data Analysis

Version 23.11.1

Description Facilitates building topology preserving maps for rich multivariate data. See <https://en.wikipedia.org/wiki/Voronoi_diagram> for more information. Credits to Mu Sigma for their continuous support throughout the development of the package.

License Apache License 2.0

Encoding UTF-8

Imports MASS, deldir, grDevices, splancs, sp, conf.design, Hmisc,
stats, dplyr, purrr, magrittr, polyclip, ggplot2, tidyr,
scales, cluster, reshape2, plyr, data.table

Depends R (>= 3.6.0)

BugReports <https://github.com/Mu-Sigma/HVT/issues>

URL <https://github.com/Mu-Sigma/HVT>

RoxygenNote 7.2.3

Suggests knitr, rmarkdown, testthat, geozoo, kableExtra, plotly, DT,
gridExtra, gtable, htmlwidgets, installr, skimr, tibble,
devtools, tidyverse, DataExplorer, htmltools, corrplot,
gganimate

VignetteBuilder knitr

NeedsCompilation no

Author Zubin Dowlaty [aut],
Mu Sigma, Inc. [cre]

Maintainer ``Mu Sigma, Inc." <ird.experience@mu-sigma.com>

Repository CRAN

Date/Publication 2023-11-19 15:20:12 UTC

R topics documented:

exploded_hmap	2
get_cell_id	3
HVT	4
hvtHmap	6
mLayerHVT	9
multiNormalDist	11
plotCells	12
plotDiag	13
plotHVT	14
predictHVT	15
predictLayerHVT	17
qeHistPlot	19
removeNovelty	20
removeOutliers	21

Index	23
--------------	-----------

exploded_hmap	<i>Function to construct an interactive 3D heatmap overlay for each HVT Level</i>
---------------	---

Description

Function to construct an interactive 3D heatmap overlay for each HVT Level

Usage

```
exploded_hmap(
  hvt.results,
  child.level = NULL,
  hmap.cols = NULL,
  n_cells.hmap = NULL,
  separation_width = 7,
  layer_opacity = c(0.5, 0.75, 0.99),
  dim_size = 1000,
  ...
)
```

Arguments

hvt.results	List. A list of hvt.results obtained from the HVT function.
child.level	Numeric. Indicating the level for which the heat map is to be plotted.
hmap.cols	Numeric or Character. The column number of column name from the dataset indicating the variables for which the heat map is to be plotted.
n_cells.hmap	Numeric. An integer indicating the number of clusters per hierarchy (level)

separation_width Numeric. An integer indicating the width between two Levels

layer_opacity Numeric. A vector indicating the opacity of each layer/ level

dim_size Numeric. An integer indicating the dimension size used to create the matrix for the plot

... color.vec and line.width can be passed from here

Value

Plot object containing Hierarchical Voronoi Tessellation 3D Surface plot with Heatmap Overlay

Author(s)

Shubhra Prakash <shubhra.prakash@mu-sigma.com>

get_cell_id *Function to generate cell ID based on 1D sammons projection*

Description

To generate cell id for the multivariate data, the data is being projected from n-dimensions to 1-dimension and the cell id is being assigned by ordering these values and finding the corresponding indexes. The output CellID gets appended to the HVT model.

Usage

```
get_cell_id(hvt.results, seed = 123)
```

Arguments

hvt.results List. A list of hvt.results obtained from the HVT function.

seed Numeric. Random Seed

Details

Cell ID

Value

Object containing Cell.ID mappings for the given hvt.results list.

Author(s)

Shubhra Prakash <shubhra.prakash@mu-sigma.com>

Examples

```

data(USArrests)
hvt.results <- list()
hvt.results <- HVT(USArrests, n_cells = 15, depth = 1, quant.err = 0.2,
                  distance_metric = "L1_Norm", error_metric = "mean",
                  projection.scale = 10, normalize = TRUE, seed = 123,
                  quant_method="kmeans",diagnose=TRUE)
plotHVT(hvt.results, line.width = c(0.8), color.vec = c('#141B41'),
        maxDepth = 1)
get_cell_id (hvt.results)

```

HVT

Constructing Hierarchical Voronoi Tessellations

Description

Main function to construct hierarchical voronoi tessellations.

Usage

```

HVT(
  dataset,
  min_compression_perc = NA,
  n_cells = NA,
  depth = 1,
  quant.err = 0.2,
  projection.scale = 10,
  normalize = FALSE,
  seed = 279,
  distance_metric = c("L1_Norm", "L2_Norm"),
  error_metric = c("mean", "max"),
  quant_method = c("kmeans", "kmedoids"),
  scale_summary = NA,
  diagnose = FALSE,
  hvt_validation = FALSE,
  train_validation_split_ratio = 0.8
)

```

Arguments

dataset	Data frame. A data frame with different columns is given as input.
min_compression_perc	Numeric. An integer indicating the minimum percent compression rate to be achieved for the dataset
n_cells	Numeric. An integer indicating the number of cells per hierarchy (level)
depth	Numeric. An integer indicating the number of levels. (1 = No hierarchy, 2 = 2 levels, etc ...)

quant.err	Numeric. A number indicating the quantization error threshold.
projection.scale	Numeric. A number indicating the scale factor for the tessellations so as to visualize the sub-tessellations well enough.
normalize	Logical. A logical value indicating if the columns in your dataset should be normalized. Default value is TRUE.
seed	Numeric. Random Seed.
distance_metric	character. The distance metric can be "L1_Norm" or "Manhattan". L1_Norm is selected by default.
error_metric	character. The error metric can be "mean" or "max". mean is selected by default
quant_method	character. The quant_method can be "kmeans" or "kmedoids". kmeans is selected by default
scale_summary	List. A list with mean and standard deviation values for all the features in the dataset. Pass the scale summary when the input dataset is already scaled or normalize is set to False.
diagnose	Logical. A logical value indicating if the diagnose is required. Default value is TRUE.
hvt_validation	Logical. A logical value indicating if the MAD values are to tested for validation set. Default value is FALSE.
train_validation_split_ratio	Numeric. A numeric value indicating the train and validation split ratio.

Details

This is the main function to construct hierarchical voronoi tessellations. The hvq function is called from this function. The output of the hvq function is hierarchical clustered data which will be the input for constructing tessellations. The data is then represented in 2d coordinates and the tessellations are plotted using these coordinates as centroids. For subsequent levels, transformation is performed on the 2d coordinates to get all the points within its parent tile. Tessellations are plotted using these transformed points as centroids. The lines in the tessellations are chopped in places so that they do not protrude outside the parent polygon. This is done for all the subsequent levels.

Value

A list that contains the hierarchical tessellation information. This list has to be given as input argument to plot the tessellations.

[[1]]	List. Information about the tessellation co-ordinates - level wise
[[2]]	List. Information about the polygon co-ordinates - level wise
[[3]]	List. Information about the hierarchical vector quantized data - level wise
[[4]]	List. Information about the model diagnosis- selected level
[[5]]	List. Information about the MAD values and percentage anomalies for validation dataset

Author(s)

Shubhra Prakash <shubhra.prakash@mu-sigma.com>, Sangeet Moy Das <sangeet.das@mu-sigma.com>, Shantanu Vaidya <shantanu.vaidya@mu-sigma.com>

See Also

[plotHVT](#)
[hvtHmap](#)

Examples

```
data(USArrests)
hvt.results <- list()
hvt.results <- HVT(USArrests, min_compression_perc = 70, quant.err = 0.2,
                  distance_metric = "L1_Norm", error_metric = "mean",
                  projection.scale = 10, normalize = TRUE, seed = 279,
                  quant_method="kmeans")
plotHVT(hvt.results, line.width = c(0.8), color.vec = c('#141B41'),
        maxDepth = 1)

hvt.results <- list()
hvt.results <- HVT(USArrests, n_cells = 15, depth = 3, quant.err = 0.2,
                  distance_metric = "L1_Norm", error_metric = "mean",
                  projection.scale = 10, normalize = TRUE, seed = 279,
                  quant_method="kmeans")
plotHVT(hvt.results, line.width = c(1.2,0.8,0.4), color.vec = c('#141B41', '#0582CA', '#8BA0B4'),
        maxDepth = 3)
```

hvtHmap

Heat Map over Hierarchical Voronoi Tessellations

Description

Main function to construct heatmap overlay for hierarchical voronoi tessellations.

Usage

```
hvtHmap(  
  hvt.results,  
  dataset,  
  child.level,  
  hmap.cols,  
  color.vec = NULL,  
  line.width = NULL,  
  centroid.size = 3,  
  pch = 21,  
  palette.color = 6,  
  previous_level_heatmap = TRUE,
```

```

    show.points = FALSE,
    asp = 1,
    ask = TRUE,
    tess.label = NULL,
    quant.error.hmap = NULL,
    n_cells.hmap = NULL,
    label.size = 0.5,
    ...
)

```

Arguments

<code>hvt.results</code>	List. A list of <code>hvt.results</code> obtained from the HVT function.
<code>dataset</code>	Data frame. The input data set.
<code>child.level</code>	Numeric. Indicating the level for which the heat map is to be plotted.
<code>hmap.cols</code>	Numeric or Character. The column number of column name from the dataset indicating the variables for which the heat map is to be plotted.
<code>color.vec</code>	Vector. A color vector such that <code>length(color.vec) = (child.level - 1)</code> . (default = NULL)
<code>line.width</code>	Vector. A line width vector such that <code>length(line.width) = (child.level - 1)</code> . (default = NULL)
<code>centroid.size</code>	Numeric. Indicating the centroid size of the first level. (default = 3)
<code>pch</code>	Numeric. Indicating the centroid's symbol type. (default = 21)
<code>palette.color</code>	Numeric. Indicating the heat map color palette. 1 - rainbow, 2 - heat.colors, 3 - terrain.colors, 4 - topo.colors, 5 - cm.colors, 6 - seas color. (default = 6)
<code>previous_level_heatmap</code>	Logical. If TRUE, the heatmap of previous level will be overlayed on the heatmap of selected level. If # FALSE, the heatmap of only selected level will be plotted
<code>show.points</code>	Logical. Indicating if the centroids should be plotted on the tessellations. (default = FALSE)
<code>asp</code>	Numeric. Indicating the aspect ratio type. For flexible aspect ratio set, <code>asp = NA</code> . (default = 1)
<code>ask</code>	Logical. If TRUE (and the R session is interactive) the user is asked for input, before a new figure is drawn. (default = TRUE)
<code>tess.label</code>	Vector. A vector for labelling the tessellations. (default = NULL)
<code>quant.error.hmap</code>	Numeric. A number indicating the quantization error threshold.
<code>n_cells.hmap</code>	Numeric. An integer indicating the number of cells/clusters per hierarchy (level)
<code>label.size</code>	Numeric. The size by which the tessellation labels should be scaled. (default = 0.5)
<code>...</code>	The ellipsis is passed to it as additional argument. (Used internally)

Details

The output of the HVT function has all the required information about the HVT. Now a heat map is overlaid over this HVT. The user defines the level and also those variables of the data for which the heat map is to be plotted. Now for each variable a separate heat map is plotted. The plot area is divided into 2 screens where the first screen is relatively large and will have the heat map. The second screen is small and contains the gradient scale. To plot the heat map, the data is first normalized. The gradient scale is divided into 'n' regions(500 is the set default). Using the normalized data, the different regions into which the data items fall are found. Each data item is now having a region on the gradient scale. This color is filled in the tile corresponding to the data item. This procedure is done for all the tiles for that level to get the complete heat map. Once the heat map is ready, the higher level tessellations are plotted to represent the hierarchies. The size of the centroids, the thickness of the lines and the color of the tessellation lines can be given as input by the user. Appropriate values for these parameters should be given to identify the hierarchies properly. In the second screen the gradient scale is plotted. The heat maps and hierarchical tessellations are obtained for all the desired variables.

Value

Returns the ggplot object containing scatter plot for the selected depth level containing the centroids of each cell

Author(s)

Shubhra Prakash <shubhra.prakash@mu-sigma.com>, Sangeet Moy Das <sangeet.das@mu-sigma.com>

See Also

[plotHVT](#)

Examples

```
data(USArrests)
hvt.results <- list()
hvt.results <- HVT(USArrests,
  n_cells = 15, depth = 1, quant.err = 0.2,
  distance_metric = "L1_Norm", error_metric = "mean",
  projection.scale = 10, normalize = TRUE,
  quant_method = "kmeans", diagnose = TRUE
)
hvtHmap(hvt.results, USArrests,
  child.level = 1, hmap.cols = "Murder",
  line.width = c(0.2), color.vec = c("#141B41"), palette.color = 6,
  quant.error.hmap = 0.2, n_cells.hmap = 6
)
```

mlayerHVT	<i>Predict which cell and what level each point in the test dataset belongs to</i>
-----------	--

Description

Predict which cell and what level each point in the test dataset belongs to

Usage

```
mlayerHVT(
  data,
  hvt_mapA,
  hvt_mapB,
  hvt_mapC,
  mad.threshold = 0.2,
  normalize = TRUE,
  seed = 300,
  distance_metric = "L1_Norm",
  error_metric = "max",
  child.level = 1,
  line.width = c(0.6, 0.4, 0.2),
  color.vec = c("#141B41", "#6369D1", "#D8D2E1"),
  yVar = NULL,
  ...
)
```

Arguments

data	Data Frame. A dataframe containing test dataset. The dataframe should have atleast one variable used while training. The variables from this dataset can also be used to overlay as heatmap
hvt_mapA	A list of hvt.results.model obtained from HVT function while performing hierarchical vector quantization on train data
hvt_mapB	A list of removed outlier rows using removedOutliers function
hvt_mapC	A list of hvt.results.model obtained from HVT function while performing hierarchical vector quantization on train data without outlier(s)
mad.threshold	A numeric values indicating the permissible Mean Absolute Deviation
normalize	Logical. A logical value indicating if the columns in your dataset should be normalized. Default value is TRUE.
seed	Numeric. Random Seed.
distance_metric	character. The distance metric can be 'Euclidean' or 'Manhattan'. Euclidean is selected by default.
error_metric	character. The error metric can be "mean" or "max". mean is selected by default

child.level	A number indicating the level for which the heat map is to be plotted.(Only used if hmap.cols is not NULL)
line.width	Vector. A line width vector
color.vec	Vector. A color vector
yVar	character. Name of the dependent variable(s)
...	color.vec and line.width can be passed from here

Value

Dataframe containing Cell.ID predictions for the given test/validation dataset

Author(s)

Shubhra Prakash <shubhra.prakash@mu-sigma.com>, Sangeet Moy Das <sangeet.das@mu-sigma.com>, Shantanu Vaidya <shantanu.vaidya@mu-sigma.com>

See Also

[HVT](#)
[hvtHmap](#)

Examples

```
data(USArrests)

# Split in train and test
train <- USArrests[1:40, ]
test <- USArrests[41:50, ]

hvt_mapA <- list()
hvt_mapA <- HVT(train,
  min_compression_perc = 70, quant.err = 0.2,
  distance_metric = "L1_Norm", error_metric = "mean",
  projection.scale = 10, normalize = TRUE,
  quant_method = "kmeans"
)

identified_outlier_cells <- c(2, 10)
output_list <- removeOutliers(identified_outlier_cells, hvt_mapA)
hvt_mapB <- output_list[[1]]
dataset_without_outliers <- output_list[[2]]

mapA_scale_summary <- hvt_mapA[[3]]$scale_summary
hvt_mapC <- list()
hvt_mapC <- HVT(dataset_without_outliers,
  n_cells = 15,
  depth = 2, quant.err = 0.2, distance_metric = "L1_Norm",
  error_metric = "max", quant_method = "kmeans",
```

```
    projection.scale = 10, normalize = FALSE, scale_summary = mapA_scale_summary
  )

predictions <- list()
predictions <- mlayerHVT(test, hvt_mapA, hvt_mapB, hvt_mapC)
```

multiNormalDist	<i>Multivariate normal distribution</i>
-----------------	---

Description

Function to generate multivariate normal distribution where each variable has a standard normal distribution $N(0,1)$

Usage

```
multiNormalDist(sample.size = 2500, ncol = 5)
```

Arguments

sample.size	Numeric. Indicating the sample size for distribution
ncol	Numeric. Indicating the number of columns

Details

The multivariate normal distribution, multivariate Gaussian distribution, or joint normal distribution is a generalization of the one-dimensional (univariate) normal distribution to higher dimensions. It is a vector in multiple normally distributed variables, such that any linear combination of the variables is also normally distributed.

Value

Dataframe containing multinomial distribution with the given sample size and number of columns

Author(s)

Shubhra Prakash <shubhra.prakash@mu-sigma.com>

Examples

```
multiNormalDist(2500, 2)
x <- multiNormalDist(2500, 2)
hist(x[, 1])
```

plotCells	<i>Plot the identified outlier cell(s) in the voronoi tessellations map. Plotting function to construct hierarchical voronoi tessellations and highlight the cells using the compressed HVT map.</i>
-----------	--

Description

Plot the identified outlier cell(s) in the voronoi tessellations map.

Plotting function to construct hierarchical voronoi tessellations and highlight the cells using the compressed HVT map.

Usage

```
plotCells(
  plot.cells,
  hvt.map,
  line.width = c(0.6),
  color.vec = c("#141B41"),
  pch1 = 21,
  centroid.size = 0.5,
  title = NULL,
  maxDepth = 1
)
```

Arguments

plot.cells	Vector. A vector indicating the cells to be highlighted in the map
hvt.map	List. A list containing the output of HVT function which has the details of the tessellations to be plotted.
line.width	Numeric Vector. A vector indicating the line widths of the tessellation boundaries for each level.
color.vec	Vector. A vector indicating the colors of the boundaries of the tessellations at each level.
pch1	Numeric. Symbol type of the centroids of the tessellations (parent levels). Refer points . (default = 21)
centroid.size	Numeric. Size of centroids of first level tessellations. (default = 3)
title	String. Set a title for the plot. (default = NULL)
maxDepth	Numeric. An integer indicating the number of levels. (default = NULL)

Value

ggplot object containing hierarchical voronoi tessellations plot highlighting the outlier cells in the map

Author(s)

Shantanu Vaidya <shantanu.vaidya@mu-sigma.com>

See Also

[HVT](#)
[hvtHmap](#)

plotDiag	<i>Make the diagnostic plots for hierarchical voronoi tessellations model.</i>
----------	--

Description

Make the diagnostic plots for hierarchical voronoi tessellations model.

Usage

```
plotDiag(model_obj)
```

Arguments

model_obj List. A list of model_obj obtained from the HVT function or prediction object

Value

plot object containing Diagnostics plots for the HVT model or HVT predictions. for HVT model, Minimum Intra-DataPoint Distance Plot, Minimum Intra-Centroid Distance Plot Mean Absolute Deviation Plot, Distribution of Number of Observations in Cells, for Training Data and Mean Absolute Deviation Plot for Validation Data are plotted. for HVT Predictions, "Mean Absolute Deviation Plot for Training Data and Validation Data are plotted

Author(s)

Shubhra Prakash <shubhra.prakash@mu-sigma.com>

See Also

[plotHVT](#)

plotHVT	<i>Plot the hierarchical tessellations. Main plotting function to construct hierarchical voronoi tessellations.</i>
---------	---

Description

Plot the hierarchical tessellations.

Main plotting function to construct hierarchical voronoi tessellations.

Arguments

<code>hvt.results</code>	List. A list containing the output of HVT function which has the details of the tessellations to be plotted.
<code>line.width</code>	Numeric Vector. A vector indicating the line widths of the tessellation boundaries for each level.
<code>color.vec</code>	Vector. A vector indicating the colors of the boundaries of the tessellations at each level.
<code>pch1</code>	Numeric. Symbol type of the centroids of the tessellations (parent levels). Refer points . (default = 21)
<code>centroid.size</code>	Numeric. Size of centroids of first level tessellations. (default = 3)
<code>title</code>	String. Set a title for the plot. (default = NULL)
<code>maxDepth</code>	Numeric. An integer indicating the number of levels. (default = NULL)

Value

ggplot object containing the main HVT plot for the given HVT results

Author(s)

Shubhra Prakash <shubhra.prakash@mu-sigma.com>, Sangeet Moy Das <sangeet.das@mu-sigma.com>

See Also

[HVT](#)
[hvtHmap](#)

Examples

```
data("USArrests", package = "datasets")

hvt.results <- list()
hvt.results <- HVT(USArrests,
  n_cells = 15, depth = 1, quant.err = 0.2,
  distance_metric = "L1_Norm", error_metric = "mean",
  projection.scale = 10, normalize = TRUE,
```

```

    quant_method = "kmeans", diagnose = TRUE
  )
  plotHVT(hvt.results,
    line.width = c(0.8), color.vec = c("#141B41"),
    maxDepth = 1
  )

```

predictHVT	<i>Predict which cell and what level each point in the test dataset belongs to</i>
------------	--

Description

Predict which cell and what level each point in the test dataset belongs to

Usage

```

predictHVT(
  data,
  hvt.results.model,
  child.level = 1,
  mad.threshold = 0.2,
  line.width = c(0.6, 0.4, 0.2),
  color.vec = c("#141B41", "#6369D1", "#D8D2E1"),
  normalize = TRUE,
  seed = 300,
  distance_metric = "L1_Norm",
  error_metric = "max",
  yVar = NULL,
  ...
)

```

Arguments

data	List. A dataframe containing test dataset. The dataframe should have atleast one variable used while training. The variables from this dataset can also be used to overlay as heatmap
hvt.results.model	A list of hvt.results.model obtained from HVT function while performing hierarchical vector quantization on train data
child.level	A number indicating the level for which the heat map is to be plotted.(Only used if hmap.cols is not NULL)
mad.threshold	A numeric values indicating the permissible Mean Absolute Deviation
line.width	Vector. A line width vector
color.vec	Vector. A color vector

normalize	Logical. A logical value indicating if the columns in your dataset should be normalized. Default value is TRUE.
seed	Numeric. Random Seed.
distance_metric	character. The distance metric can be "Euclidean" or "Manhattan". Euclidean is selected by default.
error_metric	character. The error metric can be "mean" or "max". mean is selected by default
yVar	character. Name of the dependent variable(s)
...	color.vec and line.width can be passed from here

Value

Dataframe containing scored predicted data, prediction plots and mean absolute deviation plots

Author(s)

Shubhra Prakash <shubhra.prakash@mu-sigma.com>, Sangeet Moy Das <sangeet.das@mu-sigma.com>

See Also

[HVT](#)
[hvtHmap](#)

Examples

```
data(USArrests)
# Split in train and test

train <- USArrests[1:40, ]
test <- USArrests[41:50, ]

hvt.results <- list()
hvt.results <- HVT(train,
  n_cells = 15, depth = 1, quant.err = 0.2,
  distance_metric = "L1_Norm", error_metric = "mean",
  projection.scale = 10, normalize = TRUE,
  quant_method = "kmeans", diagnose = TRUE
)

predictions <- predictHVT(test, hvt.results, child.level = 2, mad.threshold = 0.2)
print(predictions$scoredPredictedData)
```

predictLayerHVT	<i>Predict which cell and what level each point in the test dataset belongs to</i>
-----------------	--

Description

Predict which cell and what level each point in the test dataset belongs to

Usage

```
predictLayerHVT(
  data,
  hvt_mapA,
  hvt_mapB,
  hvt_mapC,
  mad.threshold = 0.2,
  normalize = TRUE,
  seed = 300,
  distance_metric = "L1_Norm",
  error_metric = "max",
  child.level = 1,
  line.width = c(0.6, 0.4, 0.2),
  color.vec = c("#141B41", "#6369D1", "#D8D2E1"),
  yVar = NULL,
  ...
)
```

Arguments

data	Data Frame. A dataframe containing test dataset. The dataframe should have atleast one variable used while training. The variables from this dataset can also be used to overlay as heatmap
hvt_mapA	A list of hvt.results.model obtained from HVT function while performing hierarchical vector quantization on train data
hvt_mapB	A list of hvt.results.model obtained from HVT function while performing hierarchical vector quantization on train data with novelty(s)
hvt_mapC	A list of hvt.results.model obtained from HVT function while performing hierarchical vector quantization on train data without novelty(s)
mad.threshold	A numeric values indicating the permissible Mean Absolute Deviation
normalize	Logical. A logical value indicating if the columns in your dataset should be normalized. Default value is TRUE.
seed	Numeric. Random Seed.
distance_metric	character. The distance metric can be 'Euclidean' or 'Manhattan'. Euclidean is selected by default.

<code>error_metric</code>	character. The error metric can be "mean" or "max". mean is selected by default
<code>child.level</code>	A number indicating the level for which the heat map is to be plotted.(Only used if <code>hmap.cols</code> is not NULL)
<code>line.width</code>	Vector. A line width vector
<code>color.vec</code>	Vector. A color vector
<code>yVar</code>	character. Name of the dependent variable(s)
<code>...</code>	<code>color.vec</code> and <code>line.width</code> can be passed from here

Value

Dataframe containing scored predicted layer output

Author(s)

Shubhra Prakash <shubhra.prakash@mu-sigma.com>, Sangeet Moy Das <sangeet.das@mu-sigma.com>, Shantanu Vaidya <shantanu.vaidya@mu-sigma.com>, Somya Shambhawi <somya.shambhawi@mu-sigma.com>

See Also

[HVT](#)
[hvtHmap](#)

Examples

```
data(USArrests)
library("dplyr")

# Split in train and test
train <- USArrests[1:40, ]
test <- USArrests[41:50, ]

hvt_mapA <- list()
hvt_mapA <- HVT(train,
  min_compression_perc = 70, quant.err = 0.2,
  distance_metric = "L1_Norm", error_metric = "mean",
  projection.scale = 10, normalize = TRUE,
  quant_method = "kmeans"
)

identified_Novelty_cells <- c(2, 10)
output_list <- removeNovelty(identified_Novelty_cells, hvt_mapA)

data_with_novelty <- output_list[[1]] %>% dplyr::select(!c("Cell.ID", "Cell.Number"))

hvt_mapB <- HVT(data_with_novelty,
  n_cells = 3, quant.err = 0.2,
  distance_metric = "L1_Norm", error_metric = "mean",
```

```
    projection.scale = 10, normalize = TRUE,
    quant_method = "kmeans"
  )

dataset_without_novelty <- output_list[[2]]

mapA_scale_summary <- hvt_mapA[[3]]$scale_summary
hvt_mapC <- list()
hvt_mapC <- HVT(dataset_without_novelty,
  n_cells = 15,
  depth = 2, quant.err = 0.2, distance_metric = "L1_Norm",
  error_metric = "max", quant_method = "kmeans",
  projection.scale = 10, normalize = FALSE, scale_summary = mapA_scale_summary
)

predictions <- list()
predictions <- predictLayerHVT(test, hvt_mapA, hvt_mapB, hvt_mapC)
```

qeHistPlot

plotDiag Make the diagnostic plots for hierarchical voronoi tessellations model.

Description

plotDiag

Make the diagnostic plots for hierarchical voronoi tessellations model.

Usage

```
qeHistPlot(hvt.results, hvt.predictions)
```

Arguments

hvt.results List. A list of hvt.results obtained from the HVT function.
hvt.predictions List. A list of hvt.predictions obtained from the Predict function.

Value

Returns the ggplot object containing the Quantized Error distribution plots for the given HVT results and predictions

Author(s)

Shubhra Prakash <shubhra.prakash@mu-sigma.com>

See Also[plotHVT](#)

`removeNovelty`*removeNovelty*

Description

Remove identified outlier cell(s) from the dataset

Usage

```
removeNovelty(outlier_cells, hvt_results)
```

Arguments

`outlier_cells` Vector. A vector with the cell number of the identified outliers
`hvt_results` List. A list having the results of the compressed map i.e. output of HVT function

Details

This function is used to remove the identified outlier cell(s) from the dataset. It is recommended to run the HVT function before running this function. It takes input in the form of cell number of the outlier cell(s) identified using the output of the HVT function and the compressed map (`hvt_mapA`) generated using the HVT function. The output of this function is a list of two items: a new map having the data of removed outlier cell(s) and the subset of dataset without outliers.

Value

A list of two items: a map having the data of removed outlier cells and the subset of the dataset without outlier(s) which has to be passed as input argument to HVT function to generate another map

[[1]] Dataframe. Information about the removed outlier cell(s)

[[2]] Dataframe. Subset of dataset without the outlier cell(s)

Author(s)

Shantanu Vaidya <shantanu.vaidya@mu-sigma.com>

See Also

[HVT](#)
[predictLayerHVT](#)

Examples

```

data(USArrests)
hvt_mapA <- list()
hvt_mapA <- HVT(USArrests, min_compression_perc = 70, quant.err = 0.2,
               distance_metric = "L1_Norm", error_metric = "mean",
               projection.scale = 10, normalize = TRUE,
               quant_method="kmeans")
plotHVT(hvt_mapA, line.width = c(0.8), color.vec = c('#141B41'),
        maxDepth = 1)

identified_Novelty_cells <- c(2, 10)
output_list <- removeNovelty(identified_Novelty_cells, hvt_mapA)
hvt_mapB <- output_list[[1]]
dataset_without_novelty <- output_list[[2]]

```

removeOutliers	<i>removeOutliers</i>
----------------	-----------------------

Description

Remove identified outlier cell(s) from the dataset

Usage

```
removeOutliers(outlier_cells, hvt_results)
```

Arguments

outlier_cells Vector. A vector with the cell number of the identified outliers

hvt_results List. A list having the results of the compressed map i.e. output of HVT function

Details

This function is used to remove the identified outlier cell(s) from the dataset. It is recommended to run the HVT function before running this function. It takes input in the form of cell number of the outlier cell(s) identified using the output of the HVT function and the compressed map (`hvt_mapA`) generated using the HVT function. The output of this function is a list of two items: a new map having the data of removed outlier cell(s) and the subset of dataset without outliers.

Value

A list of two items: a map having the data of removed outlier cells and the subset of the dataset without outlier(s) which has to be passed as input argument to HVT function to generate another map

[[1]] Dataframe. Information about the removed outlier cell(s)

[[2]] Dataframe. Subset of dataset without the outlier cell(s)

Author(s)

Shantanu Vaidya <shantanu.vaidya@mu-sigma.com>

See Also

[HVT](#)
[mLayerHVT](#)

Examples

```
data(USArrests)
hvt_mapA <- list()
hvt_mapA <- HVT(USArrests, min_compression_perc = 70, quant.err = 0.2,
               distance_metric = "L1_Norm", error_metric = "mean",
               projection.scale = 10, normalize = TRUE,
               quant_method="kmeans")
plotHVT(hvt_mapA, line.width = c(0.8), color.vec = c('#141B41'),
        maxDepth = 1)

identified_outlier_cells <- c(2, 10)
output_list <- removeOutliers(identified_outlier_cells, hvt_mapA)
hvt_mapB <- output_list[[1]]
dataset_without_outliers <- output_list[[2]]
```

Index

- * **diagnostics**
 - plotDiag, 13
 - qeHistPlot, 19
- * **hplot**
 - HVT, 4
 - hvtHmap, 6
 - plotCells, 12
 - plotHVT, 14
- * **mlayer**
 - mlayerHVT, 9
- * **predictLayer**
 - predictLayerHVT, 17
- * **predict**
 - predictHVT, 15

exploded_hmap, 2

get_cell_id, 3

HVT, 4, 10, 13, 14, 16, 18, 20, 22

hvtHmap, 6, 6, 10, 13, 14, 16, 18

mlayerHVT, 9, 22

multiNormalDist, 11

plotCells, 12

plotDiag, 13

plotHVT, 6, 8, 13, 14, 20

points, 12, 14

predictHVT, 15

predictLayerHVT, 17, 20

qeHistPlot, 19

removeNovelty, 20

removeOutliers, 21