Package 'Cluster.OBeu'

January 25, 2018

Type Package

Title Cluster Analysis OpenBudgets.eu

Version 1.2.1

Date 2017-10-30

Description Estimate and return the needed parameters for visualizations designed for OpenBudgets.eu <http://openbudgets.eu/> datasets. Calculate cluster analysis measures in Budget data of municipalities across Europe, according to the OpenBudgets.eu data model. It involves a set of techniques and algorithms used to find and divide the data into groups of similar observations. Also, can be used generally to extract visualization parameters convert them to JSON format and use them as input in a different graphical interface.

Maintainer Kleanthis Koupidis <koupidis.okfgr@gmail.com>

URL https://github.com/okgreece/Cluster.OBeu

BugReports https://github.com/okgreece/Cluster.OBeu/issues

License GPL-2 | file LICENSE

Encoding UTF-8

LazyData TRUE

Imports car, cluster, clValid, data.tree, dendextend, graphics, jsonlite, mclust, RCurl, reshape, reshape2, stringr, tibble, utils, methods

RoxygenNote 6.0.1

Author Kleanthis Koupidis [aut, cre], Charalampos Bratsas [aut], Jaroslav Kuchar [ctb]

R topics documented:

`cl.analysis` *Cluster analusis*

Description

Clustering Analysis for OBEU datasets.

Usage

```
cl.analysis(cl.data, cl_feature = NULL, amount = NULL, cl.aggregate = "sum",
cl.meth = NULL, clust.numb = NULL, dist = "euclidean", tojson = FALSE)
```

Arguments

`cl.data`	The input data
`cl_feature`	The feature to be clustered (nominal variables)
`amount`	The numeric variables
`cl.aggregate`	Select a different aggregation in case of filtering the input data
`cl.meth`	The clustering method algorithm
`clust.numb`	The number of clusters
`dist`	The distance metric
`tojson`	If TRUE the results are returned in json format, default returns a list

Details

There are different clustering models to be selected through an evaluation process. The user should define the cl_feature, cl.aggregate and amount parameters to form the structure of cluster data. The clustering algorithm, the number of clusters and the distance metric of the clustering model are set to the best selection using internal and stability measures. The end user can also interact with the cluster analysis and these parameters by specifying the cl.method, cl.num and cl.dist parameters respectively.

Value

The final returns are the parameters needed for visualizing the cluster data depending on the selected algorithm and the specification parameters, as long as some comparison measure matrices.

- cluster.method - Label of the clustering algorithm
- raw.data - Input data
- data.pca - The principal components to visualize the input data
- modelparam - The results of this parameter depend of the selected clustering model
- compare - Clustering measures

Author(s)

Kleanthis Koupidis, Jaroslav Kuchar

See Also

`cl.features`, `clValid`, `diana`, `agnes`, `pam`, `clara`, `fanny`, `Mclust`

cl.features *Clustering features*

Description

Select clustering characteristic to form the clustering data

Usage

```
cl.features(data, features = NULL, amounts = NULL, aggregate = "sum", tojson = FALSE )
```

Arguments

data	The input data
features	The clustering features
amounts	The amount measures of the dataset
aggregate	The function to aggregate
tojson	If TRUE the results are returned in json format, default returns a list

Details

This function adapts the dataset according to the selected dimension of the dataset and the aggregation function.

Value

This function returns the dataset for cluster analysis adapted to the desired features.

Author(s)

Kleanthis Koupidis

See Also

cl.analysis

cl.plot *Clustering model plotting*

Description

cl.plot function plots the clustering model constructed by the cl.analysis function.

Usage

```
cl.plot(clustering.model, parameters = list())
```

Arguments

clustering.model	Object returned by the cl.analysis function.
parameters	List of parameters to indicate plotting of ellipses or convex hulls. Default values: list(ellipses=FALSE, convex.hulls=FALSE).

Author(s)

Jaroslav Kuchar <https://github.com/jaroslav-kuchar>

See Also

cl.analysis

Examples

```
library(Cluster.OBeu)
inputs.clustering <- cl.analysis(sample_city_data, cl.meth="kmeans", clust.numb=3)
cl.plot(inputs.clustering, parameters = list(ellipses=TRUE))
```

---

cl.summary — *Extract the proposed clustering method and the number of clusters from clvalid method*

Description

Extract the most frequent

Usage

```
cl.summary(clv)
```

Arguments

clv	A clValid object

Details

This function returns the proposed method or number of clusters or both according to the majority clustering indices of a clValid process

Value

A value that indicates the proposed method and number of clusters.

Author(s)

Kleanthis Koupidis

nums — *Select the numeric columns of a given dataset*

Description

Extract and return a data frame with the columns that include only numeric values

Usage

```
nums(data)
```

Arguments

data	The input data frame, matrix

Value

This function returns a data frame with the numeric columns of the input dataset.

Author(s)

Kleanthis Koupidis

open_spending.cl — *Read and Calculate the Basic Information for Cluster Analysis Tasks from Open Spending API*

Description

Extract and analyze the input data provided from Open Spending API, using the `cl.analysis` function.

Usage

```
open_spending.cl(json_data, dimensions=NULL, amounts=NULL, measured.dimensions=NULL,
cl.aggregate="sum", cl.method=NULL, cl.num=NULL, cl.dist="euclidean")
```

Arguments

json_data	The json string, URL or file from Open Spending API
dimensions	The dimensions/feature of the input data
amounts	The measures of the input data
measured.dimensions	The dimensions to which correspond amount/numeric variables
cl.aggregate	Aggregate function of the input data
cl.method	The clustering algorithm
cl.num	The number of clusters
cl.dist	The distance metric

Details

This function is used to read data in json format from Open Spending API, in order to implement cluster analysis through `cl.analysis` function.

Value

A json string with the resulted parameters of the `cl.analysis` function.

Author(s)

Kleanthis Koupidis

See Also

`cl.analysis`

`sample_city_data` *Sample city data*

Description

This dataset is an example data frame of the budget phase data

- Administrative_Unit
- Approved
- Draft
- Executed
- Revised

Format

A data frame with the previous characteristics as columns

Index