Package ‘Cluster.OBeu’

October 12, 2022

Type Package

Title Cluster Analysis 'OpenBudgets.eu'

Version 1.2.3

Date 2019-12-17

Description Estimate and return the needed parameters for visualisations designed for 'OpenBudgets' <http://openbudgets.eu/> data. Calculate cluster analysis measures in Budget data of municipalities across Europe, according to the 'OpenBudgets' 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 visualisation parameters convert them to 'JSON' format and use them as input in a different graphical interface.

Maintainer Kleanthis Koupidis <koupidis@okfn.gr>

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, methods, RCurl, reshape, reshape2, stringr, utils

RoxygenNote 7.0.0

Depends R (>= 3.5.0)

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

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

Repository CRAN

Date/Publication 2019-12-17 14:20:07 UTC
R topics documented:

- city_data ................................................................. 2
- cl.analysis .............................................................. 2
- cl.features .............................................................. 4
- cl.plot ................................................................. 5
- cl.summary ............................................................. 5
- convex.hulls ......................................................... 6
- ellipses ................................................................. 7
- nums ................................................................. 7
- open_spending.cl ................................................... 8

Index

---
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

---
cl.analysis

Cluster analysis

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
c1.aggregate   Select a different aggregation in case of filtering the input data
c1.meth       The clustering method algorithm
c1ust.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

Examples

cl.analysis(city_data, cl.meth = "pam", clust.numb = 3)
cl.features

Clustering features

Description

Select clustering characteristic to form the clustering data

Usage

```r
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`

Examples

```r
cl.features(city_data, features = 'Administrative_Unit')
```

# works also for other datasets
```r
cl.features(iris, features = 'Species')
```
**cl.plot**

*Clustering model plotting*

**Description**

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

**Usage**

```r
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**

```r
inputs.clustering <- cl.analysis(city_data, cl.meth="pam", clust.numb=2)
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**

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

**Arguments**

- `clv` A `clValid` object
convex.hulls

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

convex.hulls  Convex hull points

Description

Computes points to plot a convex hull for each cluster of the clustering model.

Usage

convex.hulls(clustering.model, data.pca)

Arguments

clustering.model
Object returned by the cl.analysis function.

data.pca
data as result of the stats::prcomp(clustering.model$data, scale. = T, center = T).

Value

List of vectors with points for each convex hull.
ellipses  

**Description**

Computes points to plot an ellipse for each cluster of the clustering model

**Usage**

`ellipses(clustering.model, data.pca)`

**Arguments**

- **clustering.model**  
  Object returned by the `cl.analysis` function.

- **data.pca**  
  Data as result of the `stats::prcomp(clustering.model$data, scale. = T, center = T)`.

**Value**

List of vectors with points for each ellipse.

---

**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

**Examples**

`nums(city_data)`
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`
Index

agnes, 3

city_data, 2
cl.analysis, 2, 4–8
cl.features, 3, 4
cl.plot, 5
cl.summary, 5
clara, 3
clValid, 3
convex.hulls, 6

diana, 3

diagester, 7

fanny, 3

Mclust, 3

nums, 7

open_spending.cl, 8

pam, 3