

Package ‘GraphPCA’

February 19, 2015

Type Package

Title GraphPCA, Graphical tools of histogram PCA

Version 1.0

Depends R (>= 2.15.1)

Imports ggplot2, FactoMineR, scatterplot3d, scales,

Date 2014-03-21

Author Brahim Brahim <brahim.brahim@bigdatavisualizations.com> and Sun Makosso-Kallyth <sun.makosso-kallyth@crchuq.ulaval.ca>

Maintainer Brahim Brahim <brahim.brahim@bigdatavisualizations.com>

Description This package implements a histogram principal components analysis. It provides numerical and graphical tools to deal with big data.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2014-03-28 13:53:47

R topics documented:

GraphPCA-package	2
Hist1	3
Hist2	4
Hist3	5
Hist4	6
Hist5	7
Hist6	8
HistPCA	9
Hypercube	12
Visu	12
Index	14

GraphPCA-package

Graphical tools for histogram PCA.

Description

This package implements a histogram principal components analysis. It provides numerical and graphical tools to deal with big data.

Details

Package: GraphPCA
Type: Package
Version: 1.0
Date: 2014-03-21
License: GPL (>= 2)

HistPCA, Visu.

Author(s)

Brahim Brahim <brahim.brahim@bigdatavisualizations.com> and Sun Makosso-Kallyth <sun.makosso-kallyth@crchuq.ulaval.ca>

References

Billard, L. and E. Diday (2006). Symbolic Data Analysis: conceptual statistics and data Mining. Berlin: Wiley series in computational statistics.

Diday, E., Rodriguez O. and Winberg S. (2000). Generalization of the Principal Components Analysis to Histogram Data, 4th European Conference on Principles and Practice of Knowledge Discovery in Data Bases, September 12-16, 2000, Lyon, France.

Donoho, D., & Ramos, E. (1982). Primdata: Data Sets for Use With PRIM-H. Version for second (15-18, Aug, 1983) Exposition of Statistical Graphics Technology, by American Statistical Association.

Le-Rademacher J., Billard L. (2013). Principal component histograms from interval-valued observations, Computational Statistics, v.28 n.5, p.2117-2138.

Makosso-Kallyth S. and Diday E. (2012). Adaptation of interval PCA to symbolic histogram variables, Advances in Data Analysis and Classification July, Volume 6, Issue 2, pp 147-159.

Hist1

Miles per gallon

Description

This histogram datasets Hist1 comes from the aggregation of the first variable (Miles per gallon) of the dataset collected by Ernesto Ramos and David Donoho (1982) which dealt with automobiles. The original data set contains 406 observations and the following 8 variables. We create 13 new groups of observations (Year70,...,Year82) or clusters. After that, for every cluster and all 6 other variables, we build histograms with 5 bins Every histogram has 5 bins.

Usage

```
data(Hist1)
```

Format

Data frame with the following histogram bins mpg.1, mpg.2, mpg.3, mpg.4, mpg.5.

mpg.1 a numeric vector

mpg.2 a numeric vector

mpg.3 a numeric vector

mpg.4 a numeric vector

mpg.5 a numeric vector

Source

<http://lib.stat.cmu.edu/datasets/cars.data>

References

Donoho, D., & Ramos, E. (1982). Primdata: Data Sets for Use With PRIM-H, (1982). Version for second (15-18, Aug, 1983) Exposition of Statistical Graphics Technology, by American Statistical Association.

Examples

```
data(Hist1)
```

Hist2

Number cylinders

Description

This histogram datasets Hist2 comes from the aggregation of the second variable (number cylinders) of the dataset collected by Ernesto Ramos and David Donoho (1982) which dealt with automobiles. The original data set contains 406 observations and the following 8 variables. We create 13 new groups of observations (Year70,...,Year82) or clusters. After that, for every cluster and all 6 other variables, we build histograms with 4 bins.

Usage

```
data(Hist2)
```

Format

Data frame with the following histogram bins cylinders.1, cylinders.2, cylinders.3, cylinders.4.

cylinders.1 a numeric vector

cylinders.2 a numeric vector

cylinders.3 a numeric vector

cylinders.4 a numeric vector

Source

<http://lib.stat.cmu.edu/datasets/cars.data>

References

Donoho, D., & Ramos, E. (1982). Primdata: Data Sets for Use With PRIM-H, (1982). Version for second (15-18, Aug, 1983) Exposition of Statistical Graphics Technology, by American Statistical Association.

Examples

```
data(Hist2)
```

Hist3

Engine displacement

Description

This histogram datasets Hist3 comes from the aggregation of the third variable (engine displacement (cu. inches)) of the dataset collected by Ernesto Ramos and David Donoho (1982) which dealt with automobiles. The original data set contains 406 observations and the following 8 variables. We create 13 new groups of observations (Year70,...,Year82) or clusters. After that, for every cluster and all 6 other variables, we build histograms with 5 bins.

Usage

```
data(Hist3)
```

Format

A data frame with the following histogram bins displacement.1, displacement.2, displacement.3, displacement.4 and displacement.5.

displacement.1 a numeric vector

displacement.2 a numeric vector

displacement.3 a numeric vector

displacement.4 a numeric vector

displacement.5 a numeric vector

Source

<http://lib.stat.cmu.edu/datasets/cars.data>

References

Donoho, D., & Ramos, E. (1982). Primdata: Data Sets for Use With PRIM-H, (1982). Version for second (15-18, Aug, 1983) Exposition of Statistical Graphics Technology, by American Statistical Association.

Examples

```
data(Hist3)
```

Hist4

Horsepower variable.

Description

This histogram datasets Hist4 come from the aggregation of the fourth variable (horsepower) of the dataset collected by Ernesto Ramos and David Donoho (1982) which dealt with automobiles. The original data set contains 406 observations and the following 8 variables. We create 13 new groups of observations (Year70,...,Year82) or clusters. After that, for every cluster and all 6 other variables, we build histograms with 5 bins.

Usage

```
data(Hist4)
```

Format

A data frame with the following histogram bins horsepower.1, horsepower.2, horsepower.3, horsepower.4, horsepower.5.

horsepower.1 a numeric vector

horsepower.2 a numeric vector

horsepower.3 a numeric vector

horsepower.4 a numeric vector

horsepower.5 a numeric vector

Source

<http://lib.stat.cmu.edu/datasets/cars.data>

References

Donoho, D., & Ramos, E. (1982). Primdata: Data Sets for Use With PRIM-H, (1982). Version for second (15-18, Aug, 1983) Exposition of Statistical Graphics Technology, by American Statistical Association.

Examples

```
data(Hist4)
```

Hist5

Vehicle weight in lbs.

Description

This histogram datasets Hist5 come from the aggregation of the fifth variable (vehicle weight in lbs) of the dataset collected by Ernesto Ramos and David Donoho (1982) which dealt with automobiles. The original data set contains 406 observations and the following 8 variables. We create 13 new groups of observations (Year70,...,Year82) or clusters. After that, for every cluster and all 6 other variables, we build histograms with 5 bins.

Usage

```
data(Hist5)
```

Format

A data frame with the following histogram bins weight.1, weight.2, weight.3, weight.4, weight.5.

weight.1 a numeric vector

weight.2 a numeric vector

weight.3 a numeric vector

weight.4 a numeric vector

weight.5 a numeric vector

Source

<http://lib.stat.cmu.edu/datasets/cars.data>

References

Donoho, D., & Ramos, E. (1982). Primdata: Data Sets for Use With PRIM-H, (1982). Version for second (15-18, Aug, 1983) Exposition of Statistical Graphics Technology, by American Statistical Association.

Examples

```
data(Hist5)
```

Hist6

Vehicle weight in lbs

Description

This histogram datasets Hist6 come from the aggregation of the second variable (vehicle weight in lbs) of the dataset collected by Ernesto Ramos and David Donoho (1982) which dealt with automobiles. The original data set contains 406 observations and the following 8 variables. We create 13 new groups of observations (Year70,...,Year82) or clusters. After that, for every cluster and all 6 other variables, we build histograms with 5 bins.

Usage

```
data(Hist6)
```

Format

A data frame with the following histogram bins weight.1, weight.2, weight.3, weight.4, weight.5.

weight.1 a numeric vector

weight.2 a numeric vector

weight.3 a numeric vector

weight.4 a numeric vector

weight.5 a numeric vector

Source

<http://lib.stat.cmu.edu/datasets/cars.data>

References

Donoho, D., & Ramos, E. (1982). Primdata: Data Sets for Use With PRIM-H, (1982). Version for second (15-18, Aug, 1983) Exposition of Statistical Graphics Technology, by American Statistical Association.

Examples

```
data(Hist6)
```

 HistPCA

HistPCA

Description

Performs a PCA of multiple tables of histogram variables.

Usage

```
HistPCA(Variable = list, score = NULL, t = 1.1,
  axes = c(1, 2), Row.names = NULL, xlim = NULL,
  ylim = NULL, xlegend = NULL, ylegend = NULL,
  Col.names = NULL, transformation = 1,
  method = "hypercube", proc = 0,
  plot3d.table = NULL, axes2 = c(1, 2, 3))
```

Arguments

Variable	List of all data frames containing initial histogram variable. Every histogram is a data frames and every columns of data frame contains histogram bins.
score	List of bins score of every histogram variable. By default these scores are the ranks of histogram bins.
t	t is a real number used for transforming histogram to interval via Tchebychev's inequality. By default, t=1.1.
axes	a length 2 vector specifying the components to plot
Row.names	Retrieve or set the row names of a matrix-like object.
xlim	range for the plotted "x" values, defaulting to the range of the finite values of "x".
ylim	range for the plotted "y" values, defaulting to the range of the finite values of "y".
xlegend	This function could be used to add legends to plots.
ylegend	This function could be used to add legends to plots.
Col.names	Retrieve or set the row names of a matrix-like object.
transformation	type of tranformation for data. If transformation=2, angular is used.
method	method used (method='hypercube',method='longueur')
proc	option valid when method='longueur'. If proc=1, the procuste analysis is used.
plot3d.table	specification for the scatterplot3d. if plot3d.table=1, the scatterplot3d will appear.
axes2	a length 2 vector specifying the components to plot

Value

Correlation	Correlations between means of histogram and their principal components
Tablemean	Table containing the average of histogram mean
eigenvector	eigen vector of PCA of histogram mean
eigen_value_summary	a matrix containing all the eigenvalues, the percentage of variance and the cumulative percentage of variance
PCinterval	Data frame containing the coordinates of the individuals on the principal axes

Author(s)

Brahim Brahim <brahim.brahim@bigdatavisualizations.com> and Sun Makosso-Kallyth <sun.makosso-kallyth@crchuq.ulaval.ca>

References

Billard, L. and E. Diday (2006). Symbolic Data Analysis: conceptual statistics and data Mining. Berlin: Wiley series in computational statistics.

Diday, E., Rodriguez O. and Winberg S. (2000). Generalization of the Principal Components Analysis to Histogram Data, 4th European Conference on Principles and Practice of Knowledge Discovery in Data Bases, September 12-16, 2000, Lyon, France.

Donoho, D., & Ramos, E. (1982). Primdata: Data Sets for Use With PRIM-H. Version for second (15-18, Aug, 1983) Exposition of Statistical Graphics Technology, by American Statistical Association.

Le-Rademacher J., Billard L. (2013). Principal component histograms from interval-valued observations, Computational Statistics, v.28 n.5, p.2117-2138.

Makosso-Kallyth S. and Diday E. (2012). Adaptation of interval PCA to symbolic histogram variables, Advances in Data Analysis and Classification July, Volume 6, Issue 2, pp 147-159.

Examples

```
data(Hist1)
data(Hist2)
data(Hist3)
data(Hist4)
data(Hist5)
data(Hist6)
```

```
example0=HistPCA(Variable=list
(Hist1,Hist2,Hist3,Hist4,Hist5
,Hist6),axes=c(1,2),Row.names=
paste('Year',70:82,sep='-'),
Col.names=c('mpg',
'cylinders',
'displacement',
'horsepower',
'weight',
'acceleration')
```

```
),score=list(c(1:5),c(1:4),c(1:5),c(1:5),c(1:5),c(1:5)))
example0
```

```
example1=HistPCA(Variable=list
(Hist1,Hist2,Hist3,Hist4,Hist5
,Hist6),axes=c(1,2),Row.names=
paste('Year',70:82,sep='-'),
Col.names=c('mpg',
'cylinders',
'displacement',
'horsepower',
'weight',
'acceleration'
))
example1
```

```
PC_example1=HistPCA(Variable=list
(Hist1,Hist2,Hist3,Hist4,Hist5
,Hist6),axes=c(1,2),Row.names=
paste('Year',70:82,sep='-'),
Col.names=c('mpg',
'cylinders',
'displacement',
'horsepower',
'weight',
'acceleration'
))$PCinterval
example1
```

```
example2=HistPCA(Variable=list
(Hist1,Hist2,Hist3,Hist4,Hist5
,Hist6),axes=c(1,2),Row.names=
paste('Year',70:82,sep='-'),
Col.names=c('mpg',
'cylinders',
'displacement',
'horsepower',
'weight',
'acceleration'),
transformation=2
)
example2
```

```
example3=HistPCA(Variable=list
(Hist1,Hist2,Hist3,Hist4,Hist5
,Hist6),axes=c(1,2),Row.names=
paste('Year',70:82,sep='-'),
Col.names=c('mpg',
'cylinders',
'displacement',
'horsepower',
```

```
'weight',
'acceleration'),
method='longueur')
example3
```

```
example4=HistPCA(Variable=list
(Hist1,Hist2,Hist3,Hist4,Hist5
,Hist6),axes=c(1,2),Row.names=
paste('Year',70:82,sep='-'),
Col.names=c('mpg',
'cylinders',
'displacement',
'horsepower',
'weight',
'acceleration'),
method='longueur',proc=1)
example4
```

Hypercube

Hypercube function

Description

used for building of hypercube

Visu

Visu

Description

This function plots a scatterplot of histogram variables, using the R package ggplot.

Usage

```
Visu(PC, Row.names = NULL, labs = NULL, axes = c(1, 2))
```

Arguments

PC	data frame with for columns (xmin, xmax, ymin, ymax) containing symbolic 2 symbolic interval variables.
Row.names	Retrieve or set the row names of a matrix-like object.
labs	set the names of the axes.
axes	a length 2 vector specifying the components to plot.

Author(s)

Brahim Brahim Brahim Brahim <brahim.brahim@bigdatavisualizations.com> and Sun Makosso-Kallyth <sun.makosso-kallyth@crchuq.ulaval.ca>

References

Billard, L. and E. Diday (2006). Symbolic Data Analysis: conceptual statistics and data Mining. Berlin: Wiley series in computational statistics.

Donoho, D., & Ramos, E. (1982). Primdata: Data Sets for Use With PRIM-H. Version for second (15-18, Aug, 1983) Exposition of Statistical Graphics Technology, by American Statistical Association.

Le-Rademacher J., Billard L. (2013). Principal component histograms from interval-valued observations, Computational Statistics, v.28 n.5, p.2117-2138.

Makosso-Kallyth S. and Diday E. (2012). Adaptation of interval PCA to symbolic histogram variables, Advances in Data Analysis and Classification July, Volume 6, Issue 2, pp 147-159.

Examples

```
data(Hist1)
data(Hist2)
data(Hist3)
data(Hist4)
data(Hist5)
data(Hist6)
```

```
PC_example1=HistPCA(Variable=
list(Hist1,Hist2,Hist3,Hist4,
Hist5,Hist6),axes=c(1,2),
Row.names=paste('Year',
70:82,sep='-'),Col.names=
c('mpg','cylinders',
'displacement',
'horsepower',
'weight',
'acceleration'
))$PCinterval
```

```
Visu(PC_example1, axes=c(1,2),
Row.names=rownames(PC_example1))
```

Index

*Topic **PCA, Histogram variable, Big Data, Data Visualization, Data Analysis,**
GraphPCA-package, [2](#)

GraphPCA (GraphPCA-package), [2](#)
GraphPCA-package, [2](#)

Hist1, [3](#)
Hist2, [4](#)
Hist3, [5](#)
Hist4, [6](#)
Hist5, [7](#)
Hist6, [8](#)
HistPCA, [9](#)
Hypercube, [12](#)
Visu, [12](#)