Package 'RMaCzek'

March 19, 2022

Type Package

Title Czekanowski's Diagrams

Version 1.4

Date 2022-03-18

Maintainer Krzysztof Bartoszek <krzbar@protonmail.ch>

Description Allows for production of Czekanowski's Diagrams. See K. Bartoszek, A. Vasterlund (2020) <doi:10.2478/bile-2020-0008>.

Depends R(>= 3.6)

Imports GA(>= 3.2), graphics, methods, seriation(>= 1.3.4), stats, utils

License GPL-3

Collate czek_matrix.R dot_functions.R plot.czek_matrix.R register_seriate_ga.R RMaCzek.R seriate_ga.R Um_factor.R utils_czek_matrix.R read_maczek.R

LazyLoad yes

LazyData true

NeedsCompilation no

Repository CRAN

Author Albin Vasterlund [aut], Krzysztof Bartoszek [cre, aut, ths], Piotr Jaskulski [ctb]

Date/Publication 2022-03-19 00:00:09 UTC

R topics documented:

`czek_matrix` *Preprocess data to produce a Czekanowski's Diagram.*

Description

This is a function that divided the values inside a distance matrix into classes. The output can be used in the plot function to produce a Czekanowski's Diagram.

Usage

```
czek_matrix(x, order = "OLO", n_classes = 5, interval_breaks = NULL,
  monitor = FALSE, distfun = dist, scale_data = TRUE, focal_obj=NULL,
  as_dist=FALSE, original_diagram=FALSE, column_order_stat_grouping=NULL,
  dist_args=list(), ...)
```

Arguments

`x`	a numeric matrix, data frame or a 'dist' object.
`order`	specifies which seriation method should be applied. The standard setting is the seriation method OLO. If NA or NULL, then no seriation is done and the original ordering is saved. The user may provide their own ordering, through a number vector of indices. Also in this case no rearrangement will be done.
`n_classes`	specifies how many classes the distances should be divided into. The standard setting is 5 classes.
`interval_breaks`	specifies the partition boundaries for the distances. As a standard setting, each class represents an equal amount of distances. If the interval, breaks are positive and sum up to 1, then it is assumed that they specify percentages of the distances in each interval. Otherwise if provided as a numeric vector not summing up to 1, they specify the exact boundaries for the symbols representing distance groups.
`monitor`	specifies if the distribution of the distances should be visualized. The standard setting is that the distribution will not be visualized. TRUE and "cumulativ_plot" is available.
`distfun`	specifies which distance function should be used. Standard setting is the dist function which uses the Euclidean distance. The first argument of the function has to be the matrix or data frame containing the data.
`scale_data`	specifies if the data set should be scaled. The standard setting is that the data will be scaled.

`focal_obj` numbers or names of objects (rows if x is a dataset and not 'dist' object) that are not to take part in the reordering procedure. These observations will be placed as last rows and columns of the output matrix. See Details.

`as_dist` if TRUE, then the distance matrix of x is returned, with object ordering, instead of the matrix with the levels assigned in place of the original distances. The output will, be of class czek_matrix_dist, if FALSE, then of class czek_matrix.

`original_diagram` if TRUE, then the returned matrix corresponds as close as possible to the original method proposed by Czekanowski (1909). The levels are column specific and not matrix specific. See Details.

`column_order_stat_grouping` if original_diagram is TRUE, then here one can pass the partition boundaries for the ranking in each column.

`dist_args` specifies further parameters that can be passed on to the distance function.

`...` specifies further parameters that can be passed on to the seriate function in the seriation package.

Details

In his original paper Czekanowski (1909) did not have as the output a symmetric matrix where each distance was assigned a level (symbol) depending in which numeric interval it was in. Instead having the desired ordering, the following procedure was applied to each column. The three smallest distances (in each column) obtain level (symbol) 1, the fourth smallest level (symbol) 2, fifth smallest level (symbol) 3, sixth smallest level (symbol) 4 and all the bigger distances the fifth symbol which was originally just a blank cell in the output matrix. Here, we give the user more flexibility. In column_order_stat_grouping one may specify how the order statistics should be grouped in each column. See Example. The user may also choose some observations not to influence the ordering procedure. This could be useful if e.g. a single observation is meant to be assigned to a cluster and for some reason the clusters (that are to be read of from the ordering) should not be influenced by this observation. One can pass such observations using the focal_obs parameter.

A hopefully useful property is that the ordering inside the czek_matrix (and hence of the diagram when one calls plot) can be manually changed. One merely manipulates the order attribute as desired. However in such a case one should remember that the Path_length, criterion_value and Um attributes will have incorrect values and should be corrected (see Examples).

Value

The function returns a matrix with class czek_matrix. The return from the function is expected to be passed to the plot function. If as_dist is passed as TRUE, then a czek_matrix_dist object is returned and this is not suitable for the plotting. As an attribute of the output the optimized criterion value is returned. However, this is a guess based on seriation::seriate()'s and seriation::criterion()'s manuals. If something else was optimized, e.g. due to user's parameters, then this will be wrong. If unable to guess, then NA saved in the attribute.

Author(s)

Albin Vasterlund

Maintainer: Krzysztof Bartoszek <krzbar@protonmail.ch>

References

K. Bartoszek and A. Vasterlund (2020). "Old Techniques for New Times": the RMaCzek package for producing Czekanowski's diagrams. Biometrical Letters 57(2):89-118.

J. Czekanowski (1909). Zur Differentialdiagnose der Neandertalgruppe. Korespondentblatt der Deutschen Gesellschaft fur Anthropologie, Ethnologie und Urgeschichte, XL(6/7):44-47,

A. Soltysiak and P. Jaskulski (1999). Czekanowski's diagram. a method of multidimensional clustering. New Techniques for Old Times. CAA 98. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 26th Conference, Barcelona, March 1998, number 757 in BAR International Series, pages 175-184, Oxford.

Vasterlund, A. (2019). Czekanowski's Diagram: Implementing and exploring Czekanowski's Diagram with different seriation methods Master thesis, Linkoping University

Examples

```
# Set data ####
x<-mtcars


# Different type of input that give same result ############
czek_matrix(x)
czek_matrix(stats::dist(scale(x)))

## below a number of other options are shown
## but they take too long to run

# Change seriation method ############
#seriation::list_seriation_method()$dist
czek_matrix(x,order = "GW")
czek_matrix(x,order = "ga")
czek_matrix(x,order = sample(1:nrow(x)))


# Change number of classes ############
czek_matrix(x,n_classes = 3)


# Change the partition boundaries ############
czek_matrix(x,interval_breaks = c(0.1,0.4,0.5)) #10%, 40% and 50%
czek_matrix(x,interval_breaks = c(0,1,4,6,8.48)) #[0,1] (1,4] (4,6] (6,8.48]
czek_matrix(x,interval_breaks = "equal_width_between_classes")
#[0,1.7] (1.7,3.39]  (3.39,5.09] (5.09,6.78] (6.78,8.48]


# Change number of classes ############
czek_matrix(x,monitor = TRUE)
czek_matrix(x,monitor = "cumulativ_plot")


# Change distance function ############
czek_matrix(x,distfun = function(x) stats::dist(x,method = "manhattan"))
```

```
# Change dont scale the data ############
czek_matrix(x,scale_data = FALSE)
czek_matrix(stats::dist(x))


# Change additional settings to the seriation method ############
czek_matrix(x,order="ga",control=list(popSize=200,
                                      suggestions=c("SPIN_STS","QAP_2SUM")))


# Create matrix as originally described by Czekanowski (1909), with each column
# assigned levels according to how the order statistics of the  distances in it
# are grouped. The grouping below is the one used by Czekanowski (1909).
czek_matrix(x,original_diagram=TRUE,column_order_stat_grouping=c(3,4,5,6))

# Create matrix with two focal object that will not influence seriation
czek_matrix(x,focal_obj=c("Merc 280","Merc 450SL"))
# Same results but with object indices
czek_res<-czek_matrix(x,focal_obj=c(10,13))

## we now place the two objects in a new place
czek_res_neworder<-manual_reorder(czek_res,c(1:10,31,11:20,32,21:30),
    orig_data=x)

## the same can be alternatively done by hand
attr(czek_res,"order")<-attr(czek_res,"order")[c(1:10,31,11:20,32,21:30)]
## and then correct the values of the different criteria so that they
## are consistent with the new ordering
attr(czek_res,"Path_length")<-seriation::criterion(stats::dist(scale(x)),
    order=seriation::ser_permutation(attr(czek_res, "order")),method="Path_length")
## Here we need to know what criterion was used for the seriation procedure
## If the seriation package was used, then see the manual for seriation::seriate()
## seriation::criterion().
## If the genetic algorithm shipped with RMaCzek was used, then it was the Um factor.
attr(czek_res,"criterion_value")<-seriation::criterion(stats::dist(scale(x)),
    order=seriation::ser_permutation(attr(czek_res, "order")),method="Path_length")
attr(czek_res,"Um")<-RMaCzek::Um_factor(stats::dist(scale(x)),
    order= attr(czek_res, "order"),inverse_um=FALSE)
```

`internet_availability` *Distances between Internet availability in some counties of the Silesia Voivodeship, Poland*

Description

This is a dataset that describes Internet availability in 36 counties of the Silesia Voivodeship, Poland to school children. It is derived from counts the number of students per computer with Internet access at various levels of school education (Warzecha 2015).

Usage

```
skulls_distances
```

Format

The format is a list with three fields. The first field, `internet_availability_distances`, is a 36 x 36 matrix. The second field, `full_county_names`, is a vector of length 36. The third field, `MaCzek_order_Warzecha`, is a vector of length 36.

Details

A number of row/column names in the distances matrix are abbreviated in order to fit into the diagram. The full names of the counties is in the field `full_county_names`. The field `MaCzek_order_Warzecha` contains the ordering found by Warzecha (2015) using the MaCzek program, for comparison purposes.

Source

The data is obtained personally from Katarzyna Warzecha and is included in the package with her permission.

References

Warzecha, K. (2015); The use of quantitative methods in research on selected behavioral addictions of young people. In: Studia Ekonomiczne, 247, 121–139.

Examples

```
RNGversion(min(as.character(getRversion()),"3.6.1"))
set.seed(12345, kind = "Mersenne-Twister", normal.kind = "Inversion")

internet_availability
internet_availability_distances<-as.dist(internet_availability$internet_availability_distances)

##  Warzecha (2015)'s original ordering by MaCzek
czek_matrix_internet_availability_sym<-czek_matrix(internet_availability_distances,
    original_diagram=FALSE,order=internet_availability$MaCzek_order_Warzecha)
plot(czek_matrix_internet_availability_sym,plot_title="",label.cex=0.5)

## Try to find a better ordering using the OLO method.
## We find the same clusters of counties as MaCzek did.
czek_matrix_internet_availability_OLO_sym<-czek_matrix(internet_availability_distances,
    original_diagram=FALSE,order="OLO")
plot(czek_matrix_internet_availability_OLO_sym,plot_title="",label.cex=0.5)
## Try to find a better ordering using the QAP_2SUM method.
## Even though Um is lower, the ordering does not seem to capture the
## clusters, suggesting that Hamiltonian path length minimization
## is a better strategy than Um minimization.
czek_matrix_internet_availability_qap2sum_sym<-czek_matrix(internet_availability_distances,
    original_diagram=FALSE,order="QAP_2SUM")
plot(czek_matrix_internet_availability_qap2sum_sym,plot_title="",label.cex=0.5)
```

```
## We try to see if anything more can be found from Czekanowski's original asymmetric
## diagram, but it seems that the plots are much less informative.
## We first try to make the grouping of similar objects in each column more
## related to the found clusters through the symmetric diagram.
column_order_stat_grouping<-c(8,10,12,16)
czek_matrix_internet_availability<-czek_matrix(internet_availability_distances,
    original_diagram=TRUE,order=internet_availability$MaCzek_order_Warzecha,
    column_order_stat_grouping=column_order_stat_grouping)
plot(czek_matrix_internet_availability,plot_title="",label.cex=0.5)
czek_matrix_internet_availability_OLO<-czek_matrix(internet_availability_distances,
    original_diagram=TRUE,order=attr(czek_matrix_internet_availability_OLO_sym,"order"),
    column_order_stat_grouping=column_order_stat_grouping)
plot(czek_matrix_internet_availability_OLO,plot_title="",label.cex=0.5)
czek_matrix_internet_availability_qap2sum<-czek_matrix(internet_availability_distances,
   original_diagram=TRUE,order=attr(czek_matrix_internet_availability_qap2sum_sym,"order"),
    column_order_stat_grouping=column_order_stat_grouping)
plot(czek_matrix_internet_availability_qap2sum,plot_title="",label.cex=0.5)

RNGversion(as.character(getRversion()))
```

`manual_reorder` *Manually reorder Czekanowski's Diagram*

Description

The function returns a Czekanowski's Diagram with the new order and recalculated factors

Usage

```
manual_reorder(x, v_neworder, ...)
```

Arguments

`x`	a matrix with class czek_matrix, czek_matrix_dist or data matrix/data.frame or dist object.
`v_neworder`	a numeric vector with the new ordering.
`...`	specifies further parameters that can be passed on to the czek_matrix function or will be present in derived functions. See details and example code.

Details

This is an S3 generic function. The RMaCzek package defines five methods for it:

- manual_reorder.czek_matrixIn this case the input is of czek_matrix class and the user has to specify a further parameter orig_data. This is as the czek_matrix object does not hold information on the distances and hence they will need to be recalculated. The user should pass all the parameters that were passed to the original call to czek_matrix(), except order and as_dist.

- manual_reorder.czek_matrix_distIn this case the input is of czek_matrix_dist class and nothing additionally needs to be provided.
- manual_reorder.data.frameIn this case the input is a data.frame. The user should pass all the parameters that they would have passed to the original call to czek_matrix(), except order.
- manual_reorder.matrixIn this case the input is a matrix of measurements (not distance matrix). The user should pass all the parameters that they would have passed to the original call to czek_matrix(), except order.
- manual_reorder.distIn this case the input is a dist object (e.g. distance matrix after calling as.dist()). The user should pass all the parameters that they would have passed to the original call to czek_matrix(), except order.

Author(s)

Krzysztof Bartoszek <krzbar@protonmail.ch>

Examples

```
#'# Set data ####
x<-mtcars
# Calculate Czekanowski's diagram
czkm<-czek_matrix(x)
czkm_dist<-czek_matrix(x,as_dist=TRUE)
# new ordering
neworder<-attr(czkm,"order")
neworder[1:2]<-neworder[2:1]
# reorder the diagram
#if the output was Czekanowski's diagram without the distances
#remembered, then the original data has to be passed so that
#factors can be recalculated.
new_czkm<-manual_reorder(czkm,v_neworder=neworder,orig_data=x)
new_czkm_dist<-manual_reorder(czkm_dist,v_neworder=neworder)
#we can also pass the original data directly
new_czkm<-manual_reorder(x,v_neworder=neworder)
#and this is equivalent to calling
czkm<-czek_matrix(x,order=neworder)
#up to the value of the "criterion_value" attribute
#which in the second case can be lost, as no information is passed
#on which one was originally used, while in the first case it might
#be impossible to recalculate-only criteria values from seriate are supported
#if a user has a custom seriation function, then they need to recalculate this
#value themselves
```

---

plot.czek_matrix *Produce a Czekanowski's Diagram*

Description

This is a function that can produce Czekanowski's Diagram.

Usage

```
## S3 method for class 'czek_matrix'
plot(x, values = NULL, type = "symbols",
  plot_pch = NULL, plot_cex = 1.5, label.cex = 0.6,
  plot_title = "Czekanowski's diagram", legend = FALSE, axis = TRUE,
  ...)
```

Arguments

x	a matrix with class czek_matrix.
values	specifies the color or the size of the symbols in the graph. The standard setting is a grey scale for a color graph and a vector with the values 2,1,0.5,0.25 and 0 for a graph with symbols.
type	specifies if the graph should use color or symbols. The standard setting is symbols.
plot_pch	specifies which symbols the graph should use. The standard setting is 19, which is a black circle.
plot_cex	specifies the size of the cells in a color graph. The standard setting is 1.5.
label.cex	specifies the size of the labels for the objects. The standard setting is 0.6.
plot_title	specifies the main title in the graph.
legend	specifies if a legend should be included or not. The standard setting is that the legend will not be included.
axis	specifies if the labels for the objects should be included. The standard setting is that the labels are included.
...	specifies further parameters that can be passed on to the plot function.

Author(s)

Albin Vasterlund

Maintainer: Krzysztof Bartoszek <krzbar@protonmail.ch>

References

K. Bartoszek and A. Vasterlund (2020). "Old Techniques for New Times": the RMaCzek package for producing Czekanowski's diagrams. Biometrical Letters 57(2):89-118.

J. Czekanowski (1909). Zur Differentialdiagnose der Neandertalgruppe. Korespondentblatt der Deutschen Gesellschaft fur Anthropologie, Ethnologie und Urgeschichte, XL(6/7):44-47,

A. Soltysiak and P. Jaskulski (1999). Czekanowski's diagram. a method of multidimensional clustering. New Techniques for Old Times. CAA 98. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 26th Conference, Barcelona, March 1998, number 757 in BAR International Series, pages 175-184, Oxford.

A. Vasterlund (2019). Czekanowski's Diagram: Implementing and exploring Czekanowski's Diagram with different seriation methods Master thesis, Linkoping University

Examples

```
# Set data ####
x<-czek_matrix(mtcars)

# Standard plot ############
plot(x)
plot.czek_matrix(x)

# Specify values ############
plot(x,values=c(1.5,1,0.75,0.25,0 ))
plot(x,values=grDevices::colorRampPalette(c("black","red","white"))(5))

# Specify type ############
plot(x,type = "symbols")
plot(x,type = "col")

# Specify plot_pch ############
plot(x,plot_pch = 15)

# Specify plot_cex ############
plot(x,type="col",plot_cex = 1)

# Specify plot_cex ############
plot(x,label.cex = 0.45)

# Specify the main title ############
plot(x,plot_title = "Czekanowski's Diagram of mtcars")

# Add legend ############
plot(x,legend = TRUE)

# Remove axis name ############
plot(x,axis = FALSE)

# Change additional settings to the plot function ############
plot(x,col.main="blue",font.main=9,cex.main=2)
```

---

print.czek_matrix *Print a Czekanowski's Diagram*

Description

This is a function that prints out information on a Czekanowski's Diagram.

Usage

```
## S3 method for class 'czek_matrix'
print(x, print_raw = FALSE, ...)
```

Arguments

x	a matrix with class czek_matrix.
print_raw	logical, if TRUE print out raw, as if the object was a matrix, in particular this prints out the matrix itself, if FALSE (default) print out a summary. Furthermore, with print_raw=TRUE the attributes "levels", "partition_boundaries" and "n_classes" defining the diagram will be printed out.
...	specifies further parameters that can be passed on to the print function.

Author(s)

Krzysztof Bartoszek <krzbar@protonmail.ch>

Examples

```
czkm<-czek_matrix(mtcars)
#'# Standard print ############
print(czkm)
print.czek_matrix(czkm)
# Print out the raw object ############
print(czkm,print_raw=TRUE)
print.czek_matrix(czkm,print_raw=TRUE)
```

---

print.czek_matrix_dist

Print a Czekanowski's Diagram

Description

This is a function that prints out information on a Czekanowski's Diagram.

Usage

```
## S3 method for class 'czek_matrix_dist'
print(x, print_raw = FALSE, ...)
```

Arguments

x	a matrix with class czek_matrix_dist.
print_raw	logical, if TRUE print out raw, as if the object was a matrix, in particular this prints out the matrix itself, if FALSE (default) print out a summary. Furthermore, with print_raw=TRUE the attributes "levels", "partition_boundaries" and "n_classes" defining the diagram will be printed out.
...	specifies further parameters that can be passed on to the print function.

Author(s)

Krzysztof Bartoszek <krzbar@protonmail.ch>

Examples

```
czkm<-czek_matrix(mtcars,as_dist=TRUE)
#'# Standard print ############
print(czkm)
print.czek_matrix(czkm)
# Print out the raw object ############
print(czkm,print_raw=TRUE)
print.czek_matrix_dist(czkm,print_raw=TRUE)
```

`read_maczek_file` *Read in MaCzek data.*

Description

The function reads in an mdt file that contains the data prepared for the MaCzek program (MaCzek 3.3 - http://www.antropologia.uw.edu.pl/MaCzek/maczek.html). The example data file (slabosz.mdt), concerning skull measurements from Slaboszewo, comes from Piontek (1981).

Usage

```
read_maczek_file( filepath )
```

Arguments

`filepath`	the mdt MaCzek file to be read in.

Value

The function returns a data frame with the data.

Author(s)

Piotr Jaskulski

Maintainer: Krzysztof Bartoszek <krzbar@protonmail.ch>

References

J. Piontek J (1981). Biologiczna charakterystyka sredniowiecznej populacji ze Slaboszewa, woj. bydgoskie. Zrodla do badan biologii i historii populacji slowianskich, pages 39-83, Poznan UAM. (in Polish, Biological characteristics of the medieval population of Slaboszewo, Bydgoszcz Voivodeship. Materials for studying the biology and history of Slavic populations.)

Examples

```
filepath <- system.file("extdata", "slabosz.mdt", package="RMaCzek")
slabosz <- read_maczek_file(filepath)
res <- czek_matrix(slabosz)
```

RMaCzek *RMaCzek: A package that can produce Czekanowski's diagram*

Description

This package produces Czekanowski's diagram.

This software comes AS IS in the hope that it will be useful WITHOUT ANY WARRANTY, NOT even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. Please understand that there may still be bugs and errors. Use it at your own risk. We take no responsibility for any errors or omissions in this package or for any misfortune that may befall you or others as a result of its use. Please send comments and report bugs to Krzysztof Bartoszek at krzbar@protonmail.ch .

Details

Package:	RMaCzek
Type:	Package
Version:	1.4
Date:	2022-03-18
License:	GPL-3
LazyLoad:	yes

This package produces Czekanowski's diagram.

The packages functions

czek_matrix A function that returns a distance matrix where the distances are divided into classes. The return from the function is expected to be passed into the plot function.

plot.czek_matrix A function that returns Czekanowski's Diagram.

Author(s)

Albin Vasterlund

Piotr Jaskulski

Maintainer: Krzysztof Bartoszek <krzbar@protonmail.ch>

References

K. Bartoszek and A. Vasterlund (2020). "Old Techniques for New Times": the RMaCzek package for producing Czekanowski's diagrams. Biometrical Letters 57(2):89-118.

J. Czekanowski (1909). Zur Differentialdiagnose der Neandertalgruppe. Korespondentblatt der Deutschen Gesellschaft fur Anthropologie, Ethnologie und Urgeschichte, XL(6/7):44-47.

A. Soltysiak and P. Jaskulski (1999). Czekanowski's diagram. a method of multidimensional clustering. New Techniques for Old Times. CAA 98. Computer Applications and Quantitative Methods

in Archaeology. Proceedings of the 26th Conference, Barcelona, March 1998, number 757 in BAR International Series, pages 175-184, Oxford.

A. Vasterlund (2019). Czekanowski's Diagram: Implementing and exploring Czekanowski's Diagram with different seriation methods Master thesis, Linkoping University.

Examples

```
# Set data ####
x<-czek_matrix(mtcars)


# Standard plot #############
plot(x)
plot.czek_matrix(x)
```

`seals_similarities` *Similarities between Akkadian cylinder seals depicting the Serpent God*

Description

A data matrix containing the similarities between 37 seals depicting the Serpent God (Table 10, Soltysiak 2000).

Usage

```
skulls_distances
```

Format

The format is a 37 x 37 matrix.

Details

In the original similarities table, there are no values on the diagonal (i.e. self-similarity). Hence for further analyses, something has to be filled in, e.g. 100.

Source

The similarity matrix is taken from Soltysiak (2000).

References

A. Soltysiak (2000). "Przedstawienie Boga-Weza na pieczeciach cylindrycznych z okresu akadyjskiego: analiza ikonograficzna." Studia i Materialy Archeologiczne, 10, 189-214. (Depiction of the Serpent God on cylinder seals from the Akkadian era: and iconographic analysis, in Polish).

Examples

```
seals_similarities
## there are only NAs on the diagonal (self-similarity)
diag(seals_similarities)<-100
## need to change it to distance matrix for analysis by czek_matrix
seals_distances<-as.dist(100-seals_similarities)
int_breaks<-c(0,40,60,80,100)

## Soltysiak (2000)'s original ordering
czkm_seals<-czek_matrix(seals_distances,original_diagram=FALSE,order=NA,
    n_classes=4,interval_breaks=int_breaks)
plot(czkm_seals,plot_title="",label.cex=0.7)

## try to find a better ordering
czkm_seals_OLO<-czek_matrix(seals_distances,original_diagram=FALSE,order="OLO",
    n_classes=4,interval_breaks=int_breaks)
plot(czkm_seals_OLO,plot_title="",label.cex=0.7)

## Construct Czekanowski's original non-symmetric diagram
czkm_seals_OLO_2<-czek_matrix(seals_distances,original_diagram=TRUE,order="OLO")
plot(czkm_seals_OLO_2,plot_title="",label.cex=0.7)
```

---

skulls_distances *Distances between archaic human skulls*

Description

A data matrix containing the distances between 13 skulls of archaic humans (Table II, Czekanowski, 1909).

Usage

```
skulls_distances
```

Format

The format is a 13 x 13 matrix.

Details

In the original paper (Table II, Czekanowski, 1909) there are minor typographic error. Firstly, distance(Neandertal,Galey Hill)=10.54 while distance(Galey Hill,Neandertal)=10.504 . However, if one analysis the source of the data (Stolyhwa 1908), one will see that it should be 10.504 . Furthermore, the third row is called "Krapina G", while it should be "Krapina C".

Source

The distance matrix is taken from Czekanowski (1909), while the original skull measurements can be found in Stolyhwa (1908).

References

J. Czekanowski (1909). Zur Differentialdiagnose der Neandertalgruppe. In: Korespondentblatt der Deutschen Gesellschaft fur Anthropologie, Ethnologie und Urgeschichte, XL(6/7), 44–47.

K. Stolyhwa (1908). Czaszka z Nowosiolki jako dowod istnienia w okresie historycznym ksztaltow pokrewnych z Homo primigenius. In: Rozprawy Wydzialu matematyczno–przyrodnicznego Akademii Umiejetnosci, XLVIII(B), 1–27. (The skull from Nowosiolka as proof of existence during the era of history shapes common with Homo primigenius, in Polish).

Examples

```
## correcting typo rowname
rownames(skulls_distances)[3]<-"Krapina C"
## d(Neandertal,Galey Hill)!=d(Galey Hill,Neandertal)
## in the original paper so we need to correct,
## e.g. as based on Stolyhwa (1908)'s data
skulls_distances
sym_skulls_distances<-skulls_distances
sym_skulls_distances[5,9]<-10.504
## To obtain Czekanowski (1909)'s original diagram (with different symbols)
czek_matrix_skulls<-czek_matrix(as.dist(sym_skulls_distances),order=NA,original_diagram=TRUE)
plot(czek_matrix_skulls,plot_title="",label.cex=0.5)
## Obtain a symmetric version of the diagram
czek_matrix_skulls_sym<-czek_matrix(as.dist(sym_skulls_distances),order=NA,original_diagram=FALSE)
plot(czek_matrix_skulls_sym,plot_title="",label.cex=0.5)
## Try to find a better ordering
czek_matrix_skulls_OLO<-czek_matrix(as.dist(sym_skulls_distances),order="OLO",original_diagram=TRUE)
plot(czek_matrix_skulls_OLO,plot_title="",label.cex=0.5)
## Better ordering with original symmetric diagram
czek_matrix_skulls_OLO_sym<-czek_matrix(as.dist(sym_skulls_distances),order="OLO",
    original_diagram=FALSE)
plot(czek_matrix_skulls_OLO_sym,plot_title="",label.cex=0.5)
```

Um_factor *Calculate the Um factor*

Description

The function calculates the Um factor associated with an ordering of the rows and columns of a distance matrix. Lower values indicate a better grouping of similar objects. This was the original objective function proposed in the MaCzek program for producing Czekanowski's Diagram.

Usage

```
Um_factor(
  distMatrix,
  order = NULL,
  matrix_conversion_coefficient = 1,
  inverse_um = TRUE
)
```

Arguments

distMatrix	a 'dist' object, matrix of distances between observations.
order	a vector, if NULL, then the value of the factor is calculate for the distance matrix as is, otherwise the rows and columns are reordered according to the vector order.
matrix_conversion_coefficient	numeric, value to be added to the distances, so that a division by 0 error is not thrown.
inverse_um	logical, if TRUE, then the negative is returned. Default TRUE as the function is called in the genetic algorithm maximization procedures.

Value

The function returns a numeric value equalling the Um_factor.

References

K. Bartoszek and A. Vasterlund (2020). "Old Techniques for New Times": the RMaCzek package for producing Czekanowski's diagrams. Biometrical Letters 57(2):89-118.

A. Soltysiak and P. Jaskulski (1999). Czekanowski's diagram. a method of multidimensional clustering. New Techniques for Old Times. CAA 98. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 26th Conference, Barcelona, March 1998, number 757 in BAR International Series, pages 175-184, Oxford.

Examples

```
# Set data ####
x<-mtcars

mD<-stats::dist(scale(x))
mCz<-czek_matrix(x)
Um_factor(mD)
Um_factor(mD,order=attr(mCz,"order"))
```

---

urns *Measurements of urns from cremation graves*

Description

Measurements of urns from cremation graves excavated at Paprotki Kolonia 12 in Poland (Table p. 181, Soltysiak and Jaskulski 1999).

Usage

```
urns
```

Format

The format is a 15 x 9 matrix.

Details

Urn "gr.52.1" (row 9) has too many missing values and probably should be removed for further analyses.

The column names correspond to height (WYS), rim diameter (SW), maximal diameter (MWB), bottom diameter (SD), average wall thickness (GS), average bottom thickness (GD) and three indices describing proportions of the vessel (W-A, W-B, W-D).

Source

The matrix is taken from Soltysiak and Jaskulski (1999).

References

A. Soltysiak and P. Jaskulski (1999). "Czekanowski's Diagram. A Method of Multidimensional Clustering." New Techniques for Old Times. CAA 98. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 26th Conference, Barcelona, March 1998, number 757 in BAR International Series, pp. 175-184. Oxford.

Examples

```
urns
urns_use<-urns[-9,] ## removed as too many missing values, observation "gr.52.1"

## proposed order by Soltysiak and Jaskulski (1999) from the MaCzek program
urns_use_proposedorder<-c(1,3,5,7,2,4,8,11,10,12,6,9,14,13)
czkm_urns<-czek_matrix(urns_use_proposedorder,original_diagram=FALSE,
    order=urns_use_proposedorder,scale_data=TRUE)
plot(czkm_urns,plot_title="",label.cex=0.9)

## try to find a better ordering
czkm_urns_OLO<-czek_matrix(urns_use,original_diagram=FALSE,order="OLO",scale_data=TRUE)
plot(czkm_urns_OLO,plot_title="",label.cex=0.9)

## Construct Czekanowski's original non-symmetric diagram
czkm_urns_OLO_2<-czek_matrix(urns_use,original_diagram=TRUE,order="OLO",scale_data=TRUE)
plot(czkm_urns_OLO_2,plot_title="",label.cex=0.9)
```

Index