

Package ‘tabula’

March 20, 2019

Title Analysis, Seriation and Visualization of Archaeological Count Data

Version 1.2.0

Date 2019-03-19

Maintainer Nicolas Frerebeau <nicolas.frerebeau@u-bordeaux-montaigne.fr>

Description An easy way to examine archaeological count data (artifacts, faunal remains, etc.). This package includes several measures of diversity: e.g. richness and rarefaction (Chao1, Chao2, ACE, ICE, etc.), diversity/dominance and evenness (Brillouin, Shannon, Simpson, etc.), turnover and similarity (Brainerd-Robinson, ...). Most of these methods are described and discussed in Maguran (1988) <doi:10.1007/978-94-015-7358-0>. It also provides matrix seriation methods (reciprocal ranking, CA-based seriation) for chronological modeling and dating. The package make it easy to visualize count data and statistical thresholds: rank/abundance plots, Ford (1972) <isbn:0913134082> and Bertin (1977) <isbn:2082111121> diagrams, etc.

URL <http://github.com/nfrerebeau/tabula>

BugReports <http://github.com/nfrerebeau/tabula/issues>

Depends R (>= 3.4)

License GPL-3

Encoding UTF-8

LazyData true

Collate 'tabula.R' 'AllClasses.R' 'coerce.R' 'AllGenerics.R' 'data.R' 'date.R' 'diversity.R' 'extract.R' 'statistics.R' 'plot.R' 'rarefaction.R' 'refine.R' 'richness.R' 'seriation.R' 'seriate.R' 'show.R' 'similarity.R' 'turnover.R' 'utilities.R' 'validate.R' 'zzz.R'

RoxygenNote 6.1.1

Suggests covr, khroma, knitr, pbapply, rmarkdown, testthat

Imports dplyr (>= 0.7), FactoMineR, ggplot2 (>= 3.0.0), grDevices, magrittr, methods, plyr, rlang, stats, tidyr, utils

VignetteBuilder knitr

NeedsCompilation no

Author Nicolas Frerebeau [aut, cre] (<<https://orcid.org/0000-0001-5759-4944>>),
 Brice Lebrun [ctb] (<<https://orcid.org/0000-0001-7503-8685>>),
 Matthew Peeples [ctb] (<<https://orcid.org/0000-0003-4496-623X>>)

Repository CRAN

Date/Publication 2019-03-20 13:10:03 UTC

R topics documented:

accessors	2
BootCA-class	3
boves	4
compiegne	5
CountMatrix-class	6
dateEvent	7
DateModel-class	10
diversity	12
FrequencyMatrix-class	15
IncidenceMatrix-class	16
mississippi	18
OccurrenceMatrix-class	18
PermutationOrder-class	19
plotBar	20
plotDate	22
plotMatrix	23
plotRank	24
plotSpot	25
richness	27
seriation	30
similarity	33
SimilarityMatrix-class	36
turnover	36
zuni	39
Index	40

accessors

Accessors

Description

Accessors

Usage

totals(x)

Arguments

x An object from which to extract element(s).

Author(s)

N. Frerebeau

BootCA-class

Partial bootstrap CA

Description

An S4 class to store partial bootstrap correspondance analysis results.

Usage

```
## S4 method for signature 'BootCA'
x[[i]]

## S4 method for signature 'BootCA'
x[i, j]
```

Arguments

x A BootCA object from which to extract element(s).

i, j Indices specifying elements to extract. *i* is a [character](#) string matching to the name of a slot. *j* can be [missing](#) or `NULL`, a [numeric](#) or [character](#) vector. Numeric values are coerced to [integer](#) as by `as.integer` (and hence truncated towards zero). Character vectors will be matched to the names of the object.

Methods (by generic)

- `[]`: Extracts informations from a slot selected by subscript *i*.
- `[[`: Allows to select a slot thru *j*.

Slots

rows A three columns [data.frame](#) giving the vertices coordinates of the samples convex hull and a identifier to link each row to a sample.

columns A three columns [data.frame](#) giving the vertices coordinates of the variables convex hull and a identifier to link each row to a variable.

lengths A two columns [data.frame](#) giving the convex hull maximum dimension length of each sample and a identifier to link each row to a sample.

cutoff A length-one [numeric](#) vector giving the cutoff value for sample selection.

keep A named [numeric](#) vector giving the subscript of the samples to be kept.

Author(s)

N. Frerebeau

boves

Boves ceramics

Description

A dataset containing the ceramic counts from the castle site of Boves (Somme, France). The data are grouped into eight periods ranging from the 10th to the 18th century and thirteen ceramic types.

Usage

boves

Format

A numeric matrix with 8 rows and 13 variables (ceramic types):

I**IIa****IIb****IIIa****IIIb****IIIc****IVa****IVb****Va****Vb****VI****VII****Source**

Racinet P. (2002). Le site castral et prioral de Boves du Xe au XVIIe siècle. Bilan des recherches 1996-2000. *Revue archéologique de Picardie*. Numéro spécial 20, 123 p.

compiègne

Compiègne ceramics

Description

A dataset containing the ceramic counts from the Place des Hallettes in Compiègne (Oise, France). The data are grouped into five periods of about a century, ranging from the 9th to the 14th century, and sixteen ceramic types.

Usage

compiègne

Format

A numeric matrix with 5 rows (chronological periods, numbered from the oldest to the most recent from 1 to 5) and 16 variables (ceramic types):

- A** Red to white ceramics with fine sized inclusions.
- B** Red to white ceramics with medium sized inclusions.
- C** Dark ceramics with fine sized inclusions.
- D** Dark ceramics with medium sized inclusions.
- E** Ceramics close to those of groups B or D, with similarities to group F.
- F** Black, red or beige ceramics with coarse inclusions.
- G** Red polished ceramics with fine to medium sized inclusions.
- H** Black polished ceramics with fine sized inclusions.
- I** Black polished ceramics with medium sized inclusions.
- J** Polished and painted ceramics with fine to medium sized inclusions.
- K** Painted ceramics, similar to those of group A.
- L** Painted ceramics, similar to those of group B.
- M** Painted ceramics with coarse inclusions.
- N** Glazed ceramics.
- O** Stamped ceramics.
- P** Coated ceramics.

Source

Lacroix, M. C. (1997). La céramique médiévale du site des Hallettes à Compiègne (Oise). *Revue archéologique de Picardie*. Numéro spécial, 13(1), 135-168. DOI: [10.3406/pica.1997.1945](https://doi.org/10.3406/pica.1997.1945)

CountMatrix-class *Count matrix*

Description

An S4 class to represent a count matrix.

Usage

```
CountMatrix(data = NA, nrow = 1, ncol = 1, byrow = FALSE,  
            dimnames = NULL)
```

Arguments

data	an optional data vector (including a list or expression vector). Non-atomic classed R objects are coerced by as.vector and all attributes discarded.
nrow	the desired number of rows.
ncol	the desired number of columns.
byrow	logical. If FALSE (the default) the matrix is filled by columns, otherwise the matrix is filled by rows.
dimnames	A dimnames attribute for the matrix: NULL or a list of length 2 giving the row and column names respectively. An empty list is treated as NULL, and a list of length one as row names. The list can be named, and the list names will be used as names for the dimensions.

Details

Numeric values are coerced to [integer](#) as by [as.integer](#) (and hence truncated towards zero).

Note

This class extends the base [matrix](#).

Author(s)

N. Frerebeau

See Also

[matrix](#)

Other abundance matrix: [FrequencyMatrix-class](#), [OccurrenceMatrix-class](#)

Examples

```
# Create a count data matrix
A1 <- CountMatrix(data = sample(0:10, 100, TRUE),
                  nrow = 10, ncol = 10, byrow = TRUE)

# Coerce counts to frequencies
B <- as(A1, "FrequencyMatrix")

# Row sums are internally stored before coercing to a frequency matrix
totals(B) # Get row sums

# This allows to restore the source data
A2 <- as(B, "CountMatrix")
all(A1 == A2)

# Coerce to a co-occurrence matrix
C <- as(B, "OccurrenceMatrix")
```

dateEvent

Date archaeological assemblages

Description

Experimental (see note).

Usage

```
dateEvent(object, ...)
```

```
## S4 method for signature 'CountMatrix'
dateEvent(object, dates, level = 0.95,
          cutoff = 90, jackknife = FALSE, bootstrap = FALSE, n = 1000, ...)
```

Arguments

object	A $m \times p$ matrix of count data.
...	Further arguments to be passed to CA .
dates	A list of numeric dates. Dates will be matched with assemblage by names.
level	A length-one numeric vector giving the confidence level.
cutoff	An integer giving the cumulative percentage of variance used to select CA factorial components for linear model fitting (see details). All compounds with a cumulative percentage of variance of less than the cutoff value will be retained.
jackknife	A logical scalar: should the model be checked by jackknife estimation (removing each fabric/type one at a time)?
bootstrap	A logical scalar: should the model be checked by bootstrap resampling?
n	A non-negative integer giving the number of partial bootstrap replications (see details).

Details

This is an implementation of the chronological modeling method developed by Bellanger and Husi (2012, 2013).

This method allows the estimation of two probability densities. The first one (*event date*) represents the *terminus post-quem* of an archaeological assemblage: an event dated in calendar time. The second represents the "chronological profile" of the assemblage: the accumulation rate (Bellanger and Husi 2012).

This method - somewhat similar to that described by Poblome and Groenen 2003 - is based on the adjustment of a Gaussian multiple linear regression model on the factors resulting from a correspondence analysis. This model results from the known dates of a selection of reliable contexts and allows to predict the *event* dates of the remaining assemblage with a 95 confidence interval.

Since correspondence analysis allows the rows and columns of a contingency table to be projected in the same space (through the transition formula), it is possible to estimate the date of each fabric using the previous model. Finally, the *accumulation* date of each context is defined as the mean of the fabric dates, weighted by their relative proportions in that context (akin to the *Mean Ceramic Date* proposed by South 1977).

This method relies on strong archaeological and statistical assumptions. Use it if you know what you are doing (see references below).

Value

An object of class [DateModel](#).

Model checking

Resampling methods can be used to check the stability of the resulting model. If `jackknife` is TRUE, one type/fabric is removed at a time and all statistics are recalculated. In this way, one can assess whether certain type/fabric has a substantial influence on the date estimate. If `bootstrap` is TRUE, a large number of new bootstrap assemblages is created, with the same sample size, by resampling each of the original assemblage with replacement. Then, examination of the bootstrap statistics makes it possible to pinpoint assemblages that require further investigation.

The use of resampling options (`jackknife` and `bootstrap`) can lead to much longer execution times and larger output objects. To monitor the execution of these re-sampling procedures, a progress bar will automatically be displayed if the [pbapply](#) package is installed on your machine.

Note

The original authors of the method did not publish the data supporting their demonstration and some elements are unclear. As such, no replication of their results is possible and this implementation should be considered **EXPERIMENTAL**. It may be subject to major changes in a future release.

Author(s)

N. Frerebeau

References

- Bellanger, L. & Husi, P. (2013). Mesurer et modéliser le temps inscrit dans la matière à partir d'une source matérielle : la céramique médiévale. In *Mesure et Histoire Médiévale*. Histoire ancienne et médiévale. Paris: Publication de la Sorbonne, p. 119-134.
- Bellanger, L. & Husi, P. (2012). Statistical Tool for Dating and Interpreting Archaeological Contexts Using Pottery. *Journal of Archaeological Science*, 39(4), 777-790. DOI: [10.1016/j.jas.2011.06.031](https://doi.org/10.1016/j.jas.2011.06.031).
- Bellanger, L., Tomassone, R. & Husi, P. (2008). A Statistical Approach for Dating Archaeological Contexts. *Journal of Data Science*, 6, 135-154.
- Bellanger, L., Husi, P. & Tomassone, R. (2006). Une approche statistique pour la datation de contextes archéologiques. *Revue de Statistique Appliquée*, 54(2), 65-81.
- Bellanger, L., Husi, P. & Tomassone, R. (2006). Statistical Aspects of Pottery Quantification for the Dating of Some Archaeological Contexts. *Archaeometry*, 48(1), 169-183. DOI: [10.1111/j.1475-4754.2006.00249.x](https://doi.org/10.1111/j.1475-4754.2006.00249.x).
- Poblome, J. & Groenen, P. J. F. (2003). Constrained Correspondence Analysis for Seriation of Sagalassos Tablewares. In Doerr, M. & Apostolis, S. (eds.), *The Digital Heritage of Archaeology*. Athens: Hellenic Ministry of Culture.
- South, S. A. (1977). *Method and Theory in Historical Archaeology*. New York: Academic Press.

See Also

[refine](#)

Examples

```
# Coerce dataset to abundance (count) matrix
zuni <- as(zuni, "CountMatrix")

# Assume that some assemblages are reliably dated (this is NOT a real example).
# The names of the vector entries must match the names of the assemblages.
dates <- list(
  LZ0569 = 1097, LZ0279 = 1119, CS16 = 1328, LZ0066 = 1111,
  LZ0852 = 1216, LZ1209 = 1251, CS144 = 1262, LZ0563 = 1206,
  LZ0329 = 1076, LZ0005Q = 859, LZ0322 = 1109, LZ0067 = 863,
  LZ0578 = 1180, LZ0227 = 1104, LZ0610 = 1074
)

# Model the event and accumulation date for each assemblage.
(model <- dateEvent(zuni, dates, cutoff = 90))

# Plot results for the first five assemblages
# Plot event date distribution
plotDate(model, type = "event", select = 1:5) +
  ggplot2::theme_bw()
# Plot accumulation date distribution
plotDate(model, type = "acc", select = 1:5) +
  ggplot2::theme_bw()
# Plot both distributions
plotDate(model, select = "LZ1105") +
```

```

ggplot2::theme_bw()

# Check model variability
checked <- dateEvent(zuni, dates, cutoff = 90,
                    jackknife = TRUE, bootstrap = TRUE, n = 1000)

# Extract results for the first 5 assemblages
## Modeled event dates
checked["rows", 1:5]
## Jackknife fabrics
checked["jackknife", 1:5]
## Bootstrap of assemblages
checked["bootstrap", 1:5]

```

DateModel-class

Date model

Description

An S4 class to store the event and accumulation times of archaeological assemblages as well as the results of resampling methods for date model checking.

Usage

```

## S4 method for signature 'DateModel'
x[[i]]

## S4 method for signature 'DateModel'
x[i, j, drop = TRUE]

```

Arguments

<code>x</code>	A <code>DateModel</code> object from which to extract element(s).
<code>i, j</code>	Indices specifying elements to extract. <code>i</code> is a <code>character</code> string matching to the name of a slot. <code>j</code> can be <code>missing</code> or <code>NULL</code> , a <code>numeric</code> or <code>character</code> vector. Numeric values are coerced to <code>integer</code> as by <code>as.integer</code> (and hence truncated towards zero). Character vectors will be matched to the names of the object.
<code>drop</code>	A <code>logical</code> scalar: should the result be coerced to the lowest possible dimension?

Methods (by generic)

- `[[`: Extracts informations from a slot selected by subscript `i`.
- `[`: Allows to select a slot thru `j`.

Slots

- counts** A numeric matrix of count data.
- dates** A two columns `data.frame` giving the known dates used for model fitting and an identifier to link each row to an assemblage.
- level** A length-one `numeric` vector giving the confidence level.
- model** A `multiple linear model`: the Gaussian multiple linear regression model fitted for event date estimation and prediction.
- residual** A length-one `numeric` vector giving the residual standard deviation.
- rows** A five columns `data.frame` giving the predicted event dates for each archaeological assemblage, with the following columns:
id An identifier to link each row to an assemblage.
estimation The event date estimation.
earliest The lower boundary of the confidence interval.
latest The upper boundary of the confidence interval.
error The standard error of predicted dates.
- columns** A five columns `data.frame` giving the predicted event dates for each archaeological type or fabric, with the following columns:
id An identifier to link each row to an assemblage.
estimation The event date estimation.
earliest The lower boundary of the confidence interval.
latest The upper boundary of the confidence interval.
error The standard error of predicted dates.
- accumulation** A two columns `data.frame` giving the point estimate of accumulation dates of archaeological assemblages and an identifier to link each row to an assemblage.
- jackknife** A six columns `data.frame` giving the results of the resampling procedure (jackknifing fabrics) for each assemblage (in rows) with the following columns:
id An identifier to link each row to an assemblage.
estimation The jackknife event date estimate.
earliest The lower boundary of the associated prediction interval.
latest The upper boundary of the associated prediction interval.
error The standard error of predicted means.
bias The jackknife estimate of bias.
- bootstrap** A six columns `data.frame` giving the bootstrap distribution statistics for each replicated assemblage (in rows) with the following columns:
id An identifier to link each row to an assemblage.
min Minimum value.
Q05 Sample quantile to 0.05 probability.
mean Mean value (event date).
Q95 Sample quantile to 0.95 probability.
max Maximum value.

Author(s)

N. Frerebeau

diversity

*Heterogeneity and evenness***Description**

diversity returns a diversity or dominance index. evenness returns an evenness measure.

Usage

```
diversity(object, ...)
```

```
evenness(object, ...)
```

```
## S4 method for signature 'CountMatrix'
diversity(object, method = c("berger",
  "brillouin", "mcintosh", "shannon", "simpson"), simplify = FALSE, ...)
```

```
## S4 method for signature 'CountMatrix'
evenness(object, method = c("brillouin",
  "mcintosh", "shannon", "simpson"), simplify = FALSE, ...)
```

Arguments

object	A $m \times p$ matrix of count data.
...	Further arguments passed to other methods.
method	A character string or vector of strings specifying the index to be computed (see details). Any unambiguous substring can be given.
simplify	A logical scalar: should the result be simplified to a matrix? The default value, FALSE, returns a list.

Details

Diversity measurement assumes that all individuals in a specific taxa are equivalent and that all types are equally different from each other (Peet 1974). A measure of diversity can be achieved by using indices built on the relative abundance of taxa. These indices (sometimes referred to as non-parametric indices) benefit from not making assumptions about the underlying distribution of taxa abundance: they only take relative abundances of the species that are present and species richness into account. Peet (1974) refers to them as indices of *heterogeneity*.

Diversity indices focus on one aspect of the taxa abundance and emphasize either *richness* (weighting towards uncommon taxa) or dominance (weighting towards abundant taxa; Magurran 1988).

Evenness is a measure of how evenly individuals are distributed across the sample.

The following heterogeneity index and corresponding evenness measures are available (see Magurran 1988 for details):

berger Berger-Parker dominance index. The Berger-Parker index expresses the proportional importance of the most abundant type. This metric is highly biased by sample size and richness, moreover it does not make use of all the information available from sample.

brillouin Brillouin diversity index. The Brillouin index describes a known collection: it does not assume random sampling in an infinite population. Pielou (1975) and Laxton (1978) argues for the use of the Brillouin index in all circumstances, especially in preference to the Shannon index.

mcintosh McIntosh dominance index. The McIntosh index expresses the heterogeneity of a sample in geometric terms. It describes the sample as a point of a S-dimensional hypervolume and uses the Euclidean distance of this point from the origin.

shannon Shannon-Wiener diversity index. The Shannon index assumes that individuals are randomly sampled from an infinite population and that all taxa are represented in the sample (it does not reflect the sample size). The main source of error arises from the failure to include all taxa in the sample: this error increases as the proportion of species discovered in the sample declines (Peet 1974, Magurran 1988). The maximum likelihood estimator (MLE) is used for the relative abundance, this is known to be negatively biased by sample size.

simpson Simpson dominance index for finite sample. The Simpson index expresses the probability that two individuals randomly picked from a finite sample belong to two different types. It can be interpreted as the weighted mean of the proportional abundances. This metric is a true probability value, it ranges from 0 (perfectly uneven) to 1 (perfectly even).

The berger, mcintosh and simpson methods return a *dominance* index, not the reciprocal or inverse form usually adopted, so that an increase in the value of the index accompanies a decrease in diversity.

Value

If `simplify` is FALSE, then `diversity` and `evenness` return a list (default), else return a matrix.

Note

Ramanujan approximation is used for $x!$ computation if $x > 170$.

Author(s)

N. Frerebeau

References

- Berger, W. H. & Parker, F. L. (1970). Diversity of Planktonic Foraminifera in Deep-Sea Sediments. *Science*, 168(3937), 1345-1347. DOI: [10.1126/science.168.3937.1345](https://doi.org/10.1126/science.168.3937.1345).
- Brillouin, L. (1956). *Science and information theory*. New York: Academic Press.
- Laxton, R. R. (1978). The measure of diversity. *Journal of Theoretical Biology*, 70(1), 51-67. DOI: [10.1016/0022-5193\(78\)90302-8](https://doi.org/10.1016/0022-5193(78)90302-8).
- Magurran, A. E. (1988). *Ecological Diversity and its Measurement*. Princeton, NJ: Princeton University Press. DOI: [10.1007/978-94-015-7358-0](https://doi.org/10.1007/978-94-015-7358-0).
- McIntosh, R. P. (1967). An Index of Diversity and the Relation of Certain Concepts to Diversity. *Ecology*, 48(3), 392-404. DOI: [10.2307/1932674](https://doi.org/10.2307/1932674).

- Peet, R. K. (1974). The Measurement of Species Diversity. *Annual Review of Ecology and Systematics*, 5(1), 285-307. DOI: [10.1146/annurev.es.05.110174.001441](https://doi.org/10.1146/annurev.es.05.110174.001441).
- Pielou, E. C. (1975). *Ecological Diversity*. New York: Wiley. DOI: [10.4319/lo.1977.22.1.0174b](https://doi.org/10.4319/lo.1977.22.1.0174b)
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379-423. DOI: [10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x).
- Simpson, E. H. (1949). Measurement of Diversity. *Nature*, 163(4148), 688-688. DOI: [10.1038/163688a0](https://doi.org/10.1038/163688a0).

See Also

[turnover similarity](#)

Other alpha-diversity: [richness](#)

Examples

```
# Shannon diversity index
# Data from Magurran 1988, p. 145-149
birds <- CountMatrix(
  data = c(35, 26, 25, 21, 16, 11, 6, 5, 3, 3,
           3, 3, 3, 2, 2, 2, 1, 1, 1, 1, 0, 0,
           30, 30, 3, 65, 20, 11, 0, 4, 2, 14,
           0, 3, 9, 0, 0, 5, 0, 0, 0, 0, 1, 1),
  nrow = 2, byrow = TRUE, dimnames = list(c("oakwood", "spruce"), NULL))

diversity(birds, "shannon") # 2.40 2.06
evenness(birds, "shannon") # 0.80 0.78

# Brillouin diversity index
# Data from Magurran 1988, p. 150-151
moths <- CountMatrix(data = c(17, 15, 11, 4, 4, 3, 3, 3, 2, 2, 1, 1, 1),
  nrow = 1, byrow = TRUE)

diversity(moths, "brillouin") # 1.88
evenness(moths, "brillouin") # 0.83

# Simpson dominance index
# Data from Magurran 1988, p. 152-153
trees <- CountMatrix(
  data = c(752, 276, 194, 126, 121, 97, 95, 83, 72, 44, 39,
           16, 15, 13, 9, 9, 9, 8, 7, 4, 2, 2, 1, 1, 1),
  nrow = 1, byrow = TRUE
)

diversity(trees, "simpson") # 1.19
evenness(trees, "simpson") # 0.21

# McIntosh dominance index
# Data from Magurran 1988, p. 154-155
invertebrates <- CountMatrix(
  data = c(254, 153, 90, 69, 68, 58, 51, 45, 40, 39, 25, 23, 19, 18, 16, 14, 14,
           11, 11, 11, 11, 10, 6, 6, 6, 6, 5, 3, 3, 3, 3, 3, 1, 1, 1, 1, 1, 1),
  nrow = 1, byrow = TRUE
```

```

)

diversity(invertebrates, "mcintosh") # 0.71
evenness(invertebrates, "mcintosh") # 0.82

# Berger-Parker dominance index
# Data from Magurran 1988, p. 156-157
fishes <- CountMatrix(
  data = c(394, 3487, 275, 683, 22, 1, 0, 1, 6, 8, 1, 1, 2,
           1642, 5681, 196, 1348, 12, 0, 1, 48, 21, 1, 5, 7, 3,
           90, 320, 180, 46, 2, 0, 0, 1, 0, 0, 2, 1, 5,
           126, 17, 115, 436, 27, 0, 0, 3, 1, 0, 0, 1, 0,
           32, 0, 0, 5, 0, 0, 0, 0, 13, 9, 0, 0, 4),
  nrow = 5, byrow = TRUE,
  dimnames = list(c("station 1", "station 2", "station 3",
                    "station 4", "station 5"), NULL)
)

diversity(fishes, "berger") # 0.71 0.63 0.50 0.60 0.51

```

FrequencyMatrix-class *Frequency matrix*

Description

An S4 class to represent a relative frequency matrix.

Usage

```
## S4 method for signature 'FrequencyMatrix'
totals(x)
```

Arguments

x A FrequencyMatrix object from which to extract element.

Details

To ensure data integrity, a FrequencyMatrix can only be created by coercion from a [CountMatrix](#) (see examples).

Methods (by generic)

- totals: Returns the row sums (counts).

Slots

total A [numeric](#) vector.

Note

This class extends the base [matrix](#).

Author(s)

N. Frerebeau

See Also

[matrix](#)

Other abundance matrix: [CountMatrix-class](#), [OccurrenceMatrix-class](#)

Examples

```
# Create a count data matrix
A1 <- CountMatrix(data = sample(0:10, 100, TRUE),
                  nrow = 10, ncol = 10, byrow = TRUE)

# Coerce counts to frequencies
B <- as(A1, "FrequencyMatrix")

# Row sums are internally stored before coercing to a frequency matrix
totals(B) # Get row sums

# This allows to restore the source data
A2 <- as(B, "CountMatrix")
all(A1 == A2)

# Coerce to a co-occurrence matrix
C <- as(B, "OccurrenceMatrix")
```

IncidenceMatrix-class *Incidence matrix*

Description

An S4 class to represent an incidence (presence/absence) matrix.

Usage

```
IncidenceMatrix(data = NA, nrow = 1, ncol = 1, byrow = FALSE,
                dimnames = NULL)
```


Arguments

data	an optional data vector (including a list or expression vector). Non-atomic classed R objects are coerced by as.vector and all attributes discarded.
nrow	the desired number of rows.
ncol	the desired number of columns.
byrow	logical. If FALSE (the default) the matrix is filled by columns, otherwise the matrix is filled by rows.
dimnames	A dimnames attribute for the matrix: NULL or a list of length 2 giving the row and column names respectively. An empty list is treated as NULL, and a list of length one as row names. The list can be named, and the list names will be used as names for the dimensions.

Details

Numeric values are coerced to [logical](#) as by [as.logical](#).

Note

This class extends the base [matrix](#).

Author(s)

N. Frerebeau

See Also

[matrix](#)

Examples

```
# Create an incidence (presence/absence) matrix
# Data will be coerced with as.logical()
A <- IncidenceMatrix(data = sample(0:1, 100, TRUE, c(1, 0.3)),
                    nrow = 10, ncol = 10)

# Coerce to a co-occurrence matrix
B <- as(A, "OccurrenceMatrix")

# Create a count data matrix
C <- CountMatrix(data = sample(0:10, 100, TRUE),
                nrow = 10, ncol = 10, byrow = TRUE)

# Coerce to presence/absence
D <- as(C, "IncidenceMatrix")

# Coerce to a co-occurrence matrix
E <- as(C, "OccurrenceMatrix")
```

mississippi

Mississippi ceramics

Description

A dataset containing ceramic counts from the Mississippi region.

Usage

mississippi

Format

A numeric matrix with 20 rows and 10 variables (ceramic types):

ParkinPunctate

BartonKentMPI

Painted

FortuneNoded

RanchIncised

WallsEngraved

WallaceIncised

RhodesIncised

VernonPaulApplique

HullEngraved

Source

Lipo, C. P., Madsen, M. E. & Dunnell, R. C. (2015). A Theoretically-Sufficient and Computationally-Practical Technique for Deterministic Frequency Seriation. *PLOS ONE*, 10(4), e0124942. DOI: [10.1371/journal.pone.0124942](https://doi.org/10.1371/journal.pone.0124942).

OccurrenceMatrix-class

Co-occurrence matrix

Description

An S4 class to represent a co-occurrence matrix.

Details

A co-occurrence matrix is a symmetric matrix with zeros on its main diagonal, which works out how many times (expressed in percent) each pairs of taxa occur together in at least one sample.

Note

This class extends the base [matrix](#).

Author(s)

N. Frerebeau

See Also

[matrix](#)

Other abundance matrix: [CountMatrix-class](#), [FrequencyMatrix-class](#)

Examples

```
# Create a count data matrix
A1 <- CountMatrix(data = sample(0:10, 100, TRUE),
                  nrow = 10, ncol = 10, byrow = TRUE)

# Coerce counts to frequencies
B <- as(A1, "FrequencyMatrix")

# Row sums are internally stored before coercing to a frequency matrix
totals(B) # Get row sums

# This allows to restore the source data
A2 <- as(B, "CountMatrix")
all(A1 == A2)

# Coerce to a co-occurrence matrix
C <- as(B, "OccurrenceMatrix")
```

PermutationOrder-class

Permutation order

Description

An S4 class to represent a permutation order.

Usage

```
## S4 method for signature 'PermutationOrder'
x[[i]]
```

Arguments

x A PermutationOrder object from which to extract element(s).

i A [character](#) string specifying the element to extract. Character vectors will be matched to the name of the slots.

Methods (by generic)

- `[[`: Extracts informations from a slot selected by subscript `i`.

Slots

`rows` An [integer](#) vector giving the rows permutation.

`columns` An [integer](#) vector giving the columns permutation.

`seriation` A [character](#) string indicating the seriation method used.

Author(s)

N. Frerebeau

plotBar

Bar plot

Description

Plots a Bertin or a Ford (battleship curve) diagram.

Usage

```
plotBar(object, ...)
```

```
## S4 method for signature 'CountMatrix'
plotBar(object, level = FALSE, EPPM = FALSE,
        center = TRUE, horizontal = FALSE)
```

```
## S4 method for signature 'FrequencyMatrix'
plotBar(object, level = FALSE,
        EPPM = FALSE, center = TRUE, horizontal = FALSE)
```

Arguments

<code>object</code>	An object to be plotted.
<code>...</code>	Further arguments passed to other methods.
<code>level</code>	A length-one numeric vector giving the confidence level to be drawn.
<code>EPPM</code>	A logical scalar: should the EPPM be drawn (see details)?
<code>center</code>	A logical scalar: should the bar plot be centered? The default, TRUE, produces a Ford diagram, otherwise it produces a Bertin diagram.
<code>horizontal</code>	A logical scalar: should the bar plot be horizontal? The default, FALSE, means variables in rows and cases in columns (i.e. Bertin diagram). Only used if center is FALSE.

Details

If EPPM is TRUE and if a relative abundance is greater than the mean percentage of the type, the exceeding part is highlighted. This positive difference from the column mean percentage (in french "écart positif au pourcentage moyen", EPPM) represents a deviation from the situation of statistical independence. As independence can be interpreted as the absence of relationships between types and the chronological order of the assemblages, EPPM is a usefull graphical tool to explore significance of relationship between rows and columns related to [seriation](#).

Author(s)

N. Frerebeau

References

Bertin, J. (1977). *La graphique et le traitement graphique de l'information*. Paris: Flammarion. Nouvelle Bibliothèque Scientifique.

Desachy, B. (2004). Le sériographe EPPM: un outil informatisé de sériation graphique pour tableaux de comptages. *Revue archéologique de Picardie*, 3(1), 39-56. DOI: [10.3406/pica.2004.2396](https://doi.org/10.3406/pica.2004.2396).

Ford, J. A. (1962). *A quantitative method for deriving cultural chronology*. Washington, DC: Pan American Union. Technical manual 1.

See Also

Other plot: [plotDate](#), [plotMatrix](#), [plotRank](#), [plotSpot](#)

Examples

```
# Plot Ford diagram...
count <- as(compiegne, "CountMatrix") # Data from Desachy 2004
## ...without threshold
plotBar(count)
## ...with EPPM (i.e. Bruno Desachy's sériographe)
plotBar(count, EPPM = TRUE)

# Plot Bertin diagram...
freq <- as(boves, "FrequencyMatrix") # Data from Desachy 2004
## ...without EPPM...
plotBar(freq, center = FALSE, horizontal = TRUE)
## ...and confidence intervals
plotBar(freq, level = 0.05, center = FALSE, horizontal = FALSE)
## ...with EPPM (i.e. Bruno Desachy's sériographe)
plotBar(freq, EPPM = TRUE, center = FALSE, horizontal = TRUE)
```

plotDate *Date plot*

Description

Plots date estimates.

Usage

```
plotDate(object, ...)

## S4 method for signature 'DateModel'
plotDate(object, type = c("event", "accumulation"),
         select = 1, n = 500)
```

Arguments

object	An object of class <code>DateModel</code> to be plotted.
...	Further arguments passed to other methods.
type	A <code>character</code> string or vector of character strings indicating the modelled dates to be plotted. It must be one or both (default) of event and accumulation. Any unambiguous substring can be given.
select	A <code>numeric</code> or <code>character</code> vector giving the selection of the assemblage that are drawn.
n	A length-one non-negative <code>numeric</code> vector giving the desired length of the vector of quantiles for density computation.

Details

Plots the two probability estimate density curves of archaeological assemblage dates (*event* and *accumulation* dates; Bellanger and Husi 2012).

The estimated probability density of an event date is approached by a normal distribution. The distribution of the accumulation time of each context is approached by a Gaussian mixture.

The *event date* is plotted as a line, while the *accumulation time* is shown as a grey filled area.

Author(s)

N. Frerebeau

References

Bellanger, L. & Husi, P. (2012). Statistical Tool for Dating and Interpreting Archaeological Contexts Using Pottery. *Journal of Archaeological Science*, 39(4), 777-790. DOI: [10.1016/j.jas.2011.06.031](https://doi.org/10.1016/j.jas.2011.06.031).

See Also[dateEvent](#)Other plot: [plotBar](#), [plotMatrix](#), [plotRank](#), [plotSpot](#)

`plotMatrix`*Matrix plot*

Description

Plots a heatmap.

Usage

```
plotMatrix(object, ...)  
  
## S4 method for signature 'CountMatrix'  
plotMatrix(object, PVI = FALSE)  
  
## S4 method for signature 'FrequencyMatrix'  
plotMatrix(object, PVI = FALSE)  
  
## S4 method for signature 'IncidenceMatrix'  
plotMatrix(object)
```

Arguments

<code>object</code>	An object to be plotted.
<code>...</code>	Further arguments passed to other methods.
<code>PVI</code>	A logical scalar: should the PVI be drawn instead of frequencies (see details)?

Details

If PVI is FALSE, it plots a heatmap of relative abundances (frequency), otherwise percentages of the independence value are drawn (in french, "pourcentages de valeur d'indépendance", PVI).

PVI is calculated for each cell as the percentage to the column theoretical independence value: PVI greater than 1 represent positive deviations from the independence, whereas PVI smaller than 1 represent negative deviations (Desachy 2004).

The PVI matrix allows to explore deviations from independence (an intuitive graphical approach to χ^2), in such a way that a high-contrast matrix has quite significant deviations, with a low risk of being due to randomness (Desachy 2004).

Author(s)

N. Frerebeau

References

Desachy, B. (2004). Le sériographe EPPM: un outil informatisé de sériation graphique pour tableaux de comptages. *Revue archéologique de Picardie*, 3(1), 39-56. DOI: [10.3406/pica.2004.2396](https://doi.org/10.3406/pica.2004.2396).

See Also

Other plot: [plotBar](#), [plotDate](#), [plotRank](#), [plotSpot](#)

Examples

```
# Abundance data
## Coerce dataset to abundance matrix
## Data from Desachy 2004
count <- as(compiegne, "CountMatrix")

# Plot matrix diagram...
## ...without threshod (i.e. heatmap)
plotMatrix(count)

## ...with PVI as threshold (i.e. Bruno Desachy's matrigraphe)
plotMatrix(count, PVI = TRUE) +
  ggplot2::scale_fill_gradient2(midpoint = 1)

# Presence/absence data
bin <- IncidenceMatrix(data = sample(0:1, size = 100, replace = TRUE),
  nrow = 10, ncol = 10)

# Plot matrix diagram
plotMatrix(bin) +
  ggplot2::scale_fill_manual(values = c("TRUE" = "black", "FALSE" = "white"))
```

plotRank

Rank vs abundance plot

Description

Plots a rank vs relative abundance diagram.

Usage

```
plotRank(object, ...)

## S4 method for signature 'CountMatrix'
plotRank(object, log = NULL, facet = TRUE)

## S4 method for signature 'FrequencyMatrix'
plotRank(object, log = NULL, facet = TRUE)
```


Arguments

object	An object to be plotted.
...	Further arguments passed to other methods.
log	A character string which contains "x" if the x axis is to be logarithmic, "y" if the y axis is to be logarithmic and "xy" or "yx" if both axes are to be logarithmic (base 10).
facet	A logical scalar: should a matrix of panels defined by case/sample be drawn?

Details

Note that rows are scaled to 0-1 (frequencies).

Author(s)

N. Frerebeau

See Also

Other plot: [plotBar](#), [plotDate](#), [plotMatrix](#), [plotSpot](#)

Examples

```
# Coerce dataset to abundance matrix
# Data from Desachy 2004
count <- as(compiegne, "CountMatrix")
freq <- as(boves, "FrequencyMatrix")

# Plot rank vs abundance
plotRank(count)
plotRank(count, log = "xy")

# Plot rank vs abundance
plotRank(freq, facet = FALSE)
plotRank(freq, facet = FALSE, log = "xy")
```

plotSpot

Spot plot

Description

Plots a spot matrix.

Usage

```
plotSpot(object, ...)

## S4 method for signature 'CountMatrix'
plotSpot(object, threshold = NULL)

## S4 method for signature 'FrequencyMatrix'
plotSpot(object, threshold = NULL)

## S4 method for signature 'SimilarityMatrix'
plotSpot(object)

## S4 method for signature 'OccurrenceMatrix'
plotSpot(object)
```

Arguments

object	An object to be plotted.
...	Further arguments passed to other methods.
threshold	A function that takes a numeric vector as argument and returns a single numeric value (see details). If NULL, no threshold is computed.

Details

Note that rows are scaled to 0-1 (frequencies).

Note

Adapted from Dan Gopstein's original [spot matrix](#). Credit should be given to him.

Author(s)

N. Frerebeau

See Also

Other plot: [plotBar](#), [plotDate](#), [plotMatrix](#), [plotRank](#)

Examples

```
# Plot spot diagram...

## ...of count data...
count <- as(mississippi, "CountMatrix")
### ...without threshold
plotSpot(count)
### ...with the column means as threshold
plotSpot(count, threshold = mean)
### ...with the column medians as threshold
```

```

plotSpot(count, threshold = median)

## ...of a similarity matrix
sim <- similarity(count, "brainerd")
plotSpot(sim)

## ...of a co-occurrence matrix
occ <- as(mississippi, "OccurrenceMatrix")
plotSpot(occ)

```

richness	<i>Richness and rarefaction</i>
----------	---------------------------------

Description

richness returns sample richness. rarefaction returns Hurlbert's unbiased estimate of Sander's rarefaction.

Usage

```

richness(object, ...)

rarefaction(object, ...)

## S4 method for signature 'CountMatrix'
rarefaction(object, sample)

## S4 method for signature 'CountMatrix'
richness(object, method = c("ace", "chao1",
  "chao1i", "margalef", "menhinick", "none"), unbiased = FALSE, k = 10,
  simplify = FALSE)

## S4 method for signature 'IncidenceMatrix'
richness(object, method = c("chao2",
  "chao2i", "ice"), unbiased = FALSE, k = 10, simplify = FALSE)

```

Arguments

object	A $m \times p$ matrix of count data.
...	Further arguments passed to other methods.
sample	A length-one numeric vector giving the sub-sample size.
method	A character string or vector of strings specifying the index to be computed (see details). Any unambiguous substring can be given.
unbiased	A logical scalar. Should the bias-corrected estimator be used? Only used with "chao1" or "chao2" (improved) estimator.
k	A length-one numeric vector giving the threshold between rare/infrequent and abundant/frequent species. Only used if method is "ace" or "ice".

`simplify` A `logical` scalar: should the result be simplified to a matrix? The default value, `FALSE`, returns a list.

Details

The number of different taxa, provides an instantly comprehensible expression of diversity. While the number of taxa within a sample is easy to ascertain, as a term, it makes little sense: some taxa may not have been seen, or there may not be a fixed number of taxa (e.g. in an open system; Peet 1974). As an alternative, *richness* (S) can be used for the concept of taxa number (McIntosh 1967).

It is not always possible to ensure that all sample sizes are equal and the number of different taxa increases with sample size and sampling effort (Magurran 1988). Then, *rarefaction* ($E(S)$) is the number of taxa expected if all samples were of a standard size (i.e. taxa per fixed number of individuals). Rarefaction assumes that imbalances between taxa are due to sampling and not to differences in actual abundances.

The following richness measures are available for count data:

ace Abundance-based Coverage Estimator.

chao1 Chao1 estimator.

chao1i Improved Chao1 estimator.

margalef Margalef richness index.

menhinick Menhinick richness index.

none Returns the number of observed taxa/types.

The following richness measures are available for replicated incidence data:

ice Incidence-based Coverage Estimator.

chao2 Chao2 estimator.

chao2i Improved Chao2 estimator.

Value

`rarefaction` returns a numeric vector.

If `simplify` is `FALSE`, then `richness` method returns a list (default), else returns a matrix (for `CountMatrix`) or a numeric vector (for `IncidenceMatrix`).

Author(s)

N. Frerebeau

References

Chao, A. (1984). Nonparametric Estimation of the Number of Classes in a Population. *Scandinavian Journal of Statistics*, 11(4), 265-270.

Chao, A. (1987). Estimating the Population Size for Capture-Recapture Data with Unequal Catchability. *Biometrics* 43(4), 783-791. DOI: [10.2307/2531532](https://doi.org/10.2307/2531532).

- Chao, A. & Chiu, C.-H. (2016). Species Richness: Estimation and Comparison. In Balakrishnan, N., Colton, T., Everitt, B., Piegorisch, B., Ruggeri, F. & Teugels, J. L. (Eds.), *Wiley StatsRef: Statistics Reference Online*. Chichester, UK: John Wiley & Sons, Ltd., 1-26. DOI: [10.1002/9781118445112.stat03432.pub2](https://doi.org/10.1002/9781118445112.stat03432.pub2)
- Chao, A. & Lee, S.-M. (1992). Estimating the Number of Classes via Sample Coverage. *Journal of the American Statistical Association*, 87(417), 210-217. DOI: [10.1080/01621459.1992.10475194](https://doi.org/10.1080/01621459.1992.10475194).
- Chiu, C.-H., Wang, Y.-T., Walther, B. A. & Chao, A. (2014). An improved nonparametric lower bound of species richness via a modified good-turing frequency formula. *Biometrics*, 70(3), 671-682. DOI: [10.1111/biom.12200](https://doi.org/10.1111/biom.12200).
- Hurlbert, S. H. (1971). The Nonconcept of Species Diversity: A Critique and Alternative Parameters. *Ecology*, 52(4), 577-586. DOI: [10.2307/1934145](https://doi.org/10.2307/1934145).
- Magurran, A. E. (1988). *Ecological Diversity and its Measurement*. Princeton, NJ: Princeton University Press. DOI:[10.1007/978-94-015-7358-0](https://doi.org/10.1007/978-94-015-7358-0).
- Margalef, R. (1958). Information Theory in Ecology. *General Systems*, 3, 36-71.
- Menhinick, E. F. (1964). A Comparison of Some Species-Individuals Diversity Indices Applied to Samples of Field Insects. *Ecology*, 45(4), 859-861. DOI: [10.2307/1934933](https://doi.org/10.2307/1934933).
- McIntosh, R. P. (1967). An Index of Diversity and the Relation of Certain Concepts to Diversity. *Ecology*, 48(3), 392-404. DOI: [10.2307/1932674](https://doi.org/10.2307/1932674).
- Sander, H. L. (1968). Marine Benthic Diversity: A Comparative Study. *The American Naturalist*, 102(925), 243-282.

See Also

[turnover similarity](#)

Other alpha-diversity: [diversity](#)

Examples

```
# Richness
## Margalef and Menhinick index
## Data from Magurran 1988, p. 128-129
trap <- CountMatrix(data = c(9, 3, 0, 4, 2, 1, 1, 0, 1, 0, 1, 1,
                             1, 0, 1, 0, 0, 0, 1, 2, 0, 5, 3, 0),
                    nrow = 2, byrow = TRUE, dimnames = list(c("A", "B"), NULL))
richness(trap, method = c("margalef", "menhinick"), simplify = TRUE)
## A 2.55 1.88
## B 1.95 1.66

## Chao1-type estimators
## Data from Chao & Chiu 2016
brazil <- CountMatrix(
  data = rep(x = c(1:21, 23, 25, 27, 28, 30, 32, 34:37, 41,
                  45, 46, 49, 52, 89, 110, 123, 140),
            times = c(113, 50, 39, 29, 15, 11, 13, 5, 6, 6, 3, 4,
                    3, 5, 2, 5, 2, 2, 2, 2, 1, 2, 1, 1, 1, 1, 1,
                    0, 0, 2, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0)),
  nrow = 1, byrow = TRUE
)
```

```
richness(brazil, method = c("chao1", "chao1i", "ace"),
         unbiased = FALSE, simplify = TRUE)
## 461.625 488.284 445.822

# Rarefaction
rarefaction(trap, sample = 13) # 6.56
```

seriation

Matrix seriation

Description

seriate computes a permutation order for rows and/or columns.

permute rearranges a data matrix according to a permutation order.

refine performs a partial bootstrap correspondance analysis seriation refinement.

Usage

```
seriate(object, subset, ...)
```

```
permute(object, order, ...)
```

```
refine(object, ...)
```

```
## S4 method for signature 'CountMatrix'
refine(object, cutoff, n = 1000, axes = c(1,
  2), ...)
```

```
## S4 method for signature 'CountMatrix,missing'
seriate(object,
  method = c("correspondance", "reciprocal"), EPPM = FALSE,
  margin = c(1, 2), stop = 100, ...)
```

```
## S4 method for signature 'IncidenceMatrix,missing'
seriate(object,
  method = c("correspondance", "reciprocal"), margin = c(1, 2),
  stop = 100, ...)
```

```
## S4 method for signature 'CountMatrix,BootCA'
seriate(object, subset, margin = c(1, 2),
  ...)
```

```
## S4 method for signature 'CountMatrix,PermutationOrder'
permute(object, order)
```

```
## S4 method for signature 'IncidenceMatrix,PermutationOrder'
permute(object, order)
```

Arguments

object	An $m \times p$ data matrix.
subset	A BootCA object giving the subset of object to be used.
...	Further arguments passed to other methods.
order	A PermutationOrder object giving the permutation order for rows and columns.
cutoff	A function that takes a numeric vector as argument and returns a single numeric value (see details).
n	A non-negative integer giving the number of partial bootstrap replications (see details).
axes	A numeric vector giving the subscripts of the CA axes to use (see details).
method	A character string specifying the method to be used. This must be one of "reciprocal", "correspondance" (see details). Any unambiguous substring can be given.
EPPM	A logical scalar: should the seriation be computed on EPPM instead of raw data?
margin	A numeric vector giving the subscripts which the rearrangement will be applied over: 1 indicates rows, 2 indicates columns, c(1, 2) indicates rows then columns, c(2, 1) indicates columns then rows.
stop	A length-one numeric vector giving the stopping rule (i.e. maximum number of iterations) to avoid infinite loop.

Value

seriate returns a [PermutationOrder](#) object.

permute returns either a [CountMatrix](#), [FrequencyMatrix](#) or [IncidenceMatrix](#) (the same as object).

refine returns a [BootCA](#) object containing the subscript of samples to be kept (i.e. samples with maximum dimension length of the convex hull smaller than the cutoff value).

Seriation

The matrix seriation problem in archaeology is based on three conditions and two assumptions, which Dunell (1970) summarizes as follows.

The homogeneity conditions state that all the groups included in a seriation must:

1. Be of comparable duration.
2. Belong to the same cultural tradition.
3. Come from the same local area.

The mathematical assumptions state that the distribution of any historical or temporal class:

1. Is continuous through time.
2. Exhibits the form of a unimodal curve.

These assumptions create a distributional model and ordering is accomplished by arranging the matrix so that the class distributions approximate the required pattern. The resulting order is inferred to be chronological.

The following seriation methods are available:

correspondance Correspondance analysis-based seriation. Correspondance analysis (CA) is an effective method for the seriation of archaeological assemblages. The order of the rows and columns is given by the coordinates along one dimension of the CA space, assumed to account for temporal variation. The direction of temporal change within the correspondance analysis space is arbitrary: additional information is needed to determine the actual order in time.

reciprocal Reciprocal ranking (incidence data) or averaging (frequency data) seriation. These procedures iteratively rearrange rows and/or columns according to their weighted rank in the data matrix until convergence. Note that this procedure could enter into an infinite loop. If no convergence is reached before the maximum number of iterations, it stops with a warning.

CA seriation refining

`refine` allows to identify samples that are subject to sampling error or samples that have underlying structural relationships and might be influencing the ordering along the CA space.

This relies on a partial bootstrap approach to CA-based seriation where each sample is replicated n times. The maximum dimension length of the convex hull around the sample point cloud allows to remove samples for a given cutoff value.

According to Peebles and Schachner (2012), "[this] point removal procedure [results in] a reduced dataset where the position of individuals within the CA are highly stable and which produces an ordering consistent with the assumptions of frequency seriation."

If the results of `refine` is used as an input argument in `seriate`, a correspondance analysis is performed on the subset of object which matches the samples to be kept. Then excluded samples are projected onto the dimensions of the CA coordinate space using the row transition formulae. Finally, row coordinates onto the first dimension give the seriation order.

To monitor the execution of these re-sampling procedure, a progress bar will automatically be displayed if the `pbapply` package is installed on your machine.

Author(s)

N. Frerebeau

References

- Desachy, B. (2004). Le sériographe EPPM: un outil informatisé de sériation graphique pour tableaux de comptages. *Revue archéologique de Picardie*, 3(1), 39-56. DOI: [10.3406/pica.2004.2396](https://doi.org/10.3406/pica.2004.2396).
- Dunnell, R. C. (1970). Seriation Method and Its Evaluation. *American Antiquity*, 35(03), 305-319. DOI: [10.2307/278341](https://doi.org/10.2307/278341).
- Ihm, P. (2005). A Contribution to the History of Seriation in Archaeology. In C. Weihs & W. Gaul (Eds.), *Classification: The Ubiquitous Challenge* (p. 307-316). Berlin Heidelberg: Springer. DOI: [10.1007/3-540-28084-7_34](https://doi.org/10.1007/3-540-28084-7_34).
- Peebles, M. A., & Schachner, G. (2012). Refining correspondance analysis-based ceramic seriation of regional data sets. *Journal of Archaeological Science*, 39(8), 2818-2827. DOI: [10.1016/j.jas.2012.04.040](https://doi.org/10.1016/j.jas.2012.04.040).

See Also[refine CA](#)**Examples**

```
# Matrix seriation
# Reproduces Desachy 2004 results
## Coerce dataset to abundance matrix
compiegne_count <- as(compiegne, "CountMatrix")

## Get seriation order for columns on EPPM using the reciprocal averaging method
## Expected column order: N, A, C, K, P, L, B, E, I, M, D, G, O, J, F, H
(compiegne_indices <- seriate(compiegne_count, method = "reciprocal",
                             EPPM = TRUE, margin = 2))

## Permute columns
compiegne_new <- permute(compiegne_count, compiegne_indices)

## Plot new matrix
plotBar(compiegne_new, EPPM = FALSE)

# Refine matrix seriation (this is a long running example)
# Reproduces Peeples and Schachner 2012 results
zuni_count <- as(zuni, "CountMatrix")

## Samples with convex hull maximum dimension length greater than the cutoff
## value will be marked for removal.
## Define cutoff as one standard deviation above the mean
fun <- function(x) { mean(x) + sd(x) }

## Get indices of samples to be kept
## Warning: this may take a few seconds!
set.seed(123)
(zuni_refined <- refine(zuni_count, cutoff = fun))

## Get CA-based seriation order
(zuni_indices <- seriate(zuni_count, zuni_refined, margin = 1))
```

similarity

Similarity

Description

`similarity` returns a similarity matrix.

Usage

```

similarity(object, ...)

## S4 method for signature 'CountMatrix'
similarity(object, method = c("brainerd", "bray",
  "jaccard", "morisita", "sorenson", "binomial"), ...)

## S4 method for signature 'IncidenceMatrix'
similarity(object, method = c("jaccard",
  "sorenson"), ...)

```

Arguments

object	A $m \times p$ matrix of count data.
...	Further arguments passed to other methods.
method	A character string specifying the method to be used (see details). Any unambiguous substring can be given.

Details

β -diversity can be measured by addressing *similarity* between pairs of samples/cases (Brainerd-Robinson, Jaccard, Morisita-Horn and Sorenson indices). Similarity between pairs of taxa/types can be measured by assessing the degree of co-occurrence (binomial co-occurrence).

Jaccard, Morisita-Horn and Sorenson indices provide a scale of similarity from 0-1 where 1 is perfect similarity and 0 is no similarity. The Brainerd-Robinson index is scaled between 0 and 200. The Binomial co-occurrence assessment approximates a Z-score.

binomial Binomial co-occurrence assessment. This assesses the degree of co-occurrence between taxa/types within a dataset. The strongest associations are shown by large positive numbers, the strongest segregations by large negative numbers.

brainerd Brainerd-Robinson quantitative index. This is a city-block metric of similarity between pairs of samples/cases.

bray Sorenson quantitative index (Bray and Curtis modified version of the Sorenson index).

jaccard Jaccard qualitative index.

morisita Morisita-Horn quantitative index.

sorenson Sorenson qualitative index.

Value

`similarity` returns a $m \times m$ symmetric matrix.

Author(s)

N. Frerebeau

References

- Brainerd, G. W. (1951). The Place of Chronological Ordering in Archaeological Analysis. *American Antiquity*, 16(04), 301-313. DOI: [10.2307/276979](https://doi.org/10.2307/276979).
- Bray, J. R. & Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27(4), 325-349. DOI: [10.2307/1942268](https://doi.org/10.2307/1942268).
- Kintigh, K. (2006). Ceramic Dating and Type Associations. In J. Hantman and R. Most (eds.), *Managing Archaeological Data: Essays in Honor of Sylvia W. Gaines*. Anthropological Research Paper, 57. Tempe, AZ: Arizona State University, p. 17-26.
- Magurran, A. E. (1988). *Ecological Diversity and its Measurement*. Princeton, NJ: Princeton University Press. DOI:[10.1007/978-94-015-7358-0](https://doi.org/10.1007/978-94-015-7358-0).
- Robinson, W. S. (1951). A Method for Chronologically Ordering Archaeological Deposits. *American Antiquity*, 16(04), 293-301. DOI: [10.2307/276978](https://doi.org/10.2307/276978).

See Also

[richness](#) [rarefaction](#) [diversity](#) [evenness](#)

Other beta-diversity: [turnover](#)

Examples

```
# Data from Huntley 2008
ceramics <- CountMatrix(
  data = c(16, 9, 3, 0, 1,
           13, 3, 2, 0, 0,
           9, 5, 2, 5, 0,
           14, 12, 3, 0, 0,
           0, 26, 4, 0, 0,
           1, 26, 4, 0, 0,
           0, 11, 3, 13, 0,
           0, 0, 17, 0, 16,
           0, 0, 18, 0, 14),
  nrow = 9, byrow = TRUE,
  dimnames = list(c("Atsinna", "Cienega", "Mirabal", "PdMuertos",
                    "Hesh", "LowPesc", "BoxS", "Ojo Bon", "S170"),
                  c("DLH-1", "DLH-2a", "DLH-2b", "DLH-2c", "DLH-4")))
)

# Brainerd-Robinson measure (count data)
C <- similarity(ceramics, "brainerd")
plotSpot(C)

# Data from Magurran 1988, p. 166
birds <- CountMatrix(
  data = c(1.4, 4.3, 2.9, 8.6, 4.2, 15.7, 2.0, 50, 1, 11.4, 11.4, 4.3, 13.0,
           14.3, 8.6, 7.1, 10.0, 1.4, 2.9, 5.7, 1.4, 11.4, 2.9, 4.3, 1.4, 2.9,
           0, 0, 0, 2.9, 0, 0, 0, 10, 0, 0, 5.7, 2.5, 5.7, 8.6, 5.7, 2.9, 0, 0,
           2.9, 0, 0, 5.7, 0, 2.9, 0, 2.9) * 10,
  nrow = 2, byrow = TRUE, dimnames = list(c("unmanaged", "managed"), NULL)
)
```

```
# Jaccard measure (presence/absence data)
similarity(birds, "jaccard") # 0.46

# Sorenson measure (presence/absence data)
similarity(birds, "sorenson") # 0.63

# Jaccard measure (Bray's formula ; count data)
similarity(birds, "bray") # 0.44

# Morisita-Horn measure (count data)
similarity(birds, "morisita") # 0.81
```

SimilarityMatrix-class

Similarity matrix

Description

An S4 class to represent a (dis)similarity matrix.

Arguments

x A SimilarityMatrix object from which to extract element.

Note

This class extends the base [matrix](#).

Author(s)

N. Frerebeau

See Also

[matrix](#)

turnover

Turnover

Description

Returns the degree of turnover in taxa composition along a gradient or transect.

Usage

```
turnover(object, ...)

## S4 method for signature 'CountMatrix'
turnover(object, method = c("whittaker", "cody",
  "routledge1", "routledge2", "routledge3", "wilson"), simplify = FALSE,
  ...)

## S4 method for signature 'IncidenceMatrix'
turnover(object, method = c("whittaker",
  "cody", "routledge1", "routledge2", "routledge3", "wilson"),
  simplify = FALSE, ...)
```

Arguments

<code>object</code>	A $m \times p$ matrix of count data.
<code>...</code>	Further arguments passed to other methods.
<code>method</code>	A character string specifying the method to be used (see details). Any unambiguous substring can be given.
<code>simplify</code>	A logical scalar: should the result be simplified to a matrix? The default value, FALSE, returns a list.

Details

The following methods can be used to ascertain the degree of *turnover* in taxa composition along a gradient (β -diversity) on qualitative (presence/absence) data. This assumes that the order of the matrix rows (from 1 to n) follows the progression along the gradient/transect.

whittaker Whittaker measure.

cody Cody measure.

routledge1 Routledge first measure.

routledge2 Routledge second measure.

routledge3 Routledge third measure. This is the exponential form of the second measure.

wilson Wilson measure.

Value

If `simplify` is FALSE, returns a list (default), else returns a matrix.

Author(s)

N. Frerebeau

References

- Cody, M. L. (1975). Towards a theory of continental species diversity: Bird distributions over Mediterranean habitat gradients. In M. L. Cody & J. M. Diamond (Eds.), *Ecology and Evolution of Communities*. Cambridge, MA: Harvard University Press, p. 214–257.
- Routledge, R. D. (1977). On Whittaker's Components of Diversity. *Ecology*, 58(5), 1120-1127. DOI: [10.2307/1936932](https://doi.org/10.2307/1936932).
- Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, 30(3), 279-338. DOI: [10.2307/1943563](https://doi.org/10.2307/1943563).
- Wilson, M. V., & Shmida, A. (1984). Measuring Beta Diversity with Presence-Absence Data. *The Journal of Ecology*, 72(3), 1055-1064. DOI: [10.2307/2259551](https://doi.org/10.2307/2259551).

See Also

[richness](#) [rarefaction](#) [diversity](#) [evenness](#)

Other beta-diversity: [similarity](#)

Examples

```
# Data from Magurran 1988, p. 162
trees <- IncidenceMatrix(
  data = c(1, 1, 1, 0, 0, 0,
           1, 1, 1, 1, 1, 1,
           0, 0, 1, 0, 1, 0,
           0, 0, 0, 1, 1, 1,
           0, 0, 0, 0, 1, 1,
           0, 0, 0, 1, 0, 1),
  nrow = 6, byrow = FALSE,
  dimnames = list(c("1", "2", "3", "4", "5", "6"),
                  c("Birch", "Oak", "Rowan", "Beech", "Hazel", "Holly"))
)

# Whittaker's measure
turnover(trees, "whittaker") # 1

# Cody's measure
turnover(trees, "cody") # 3

# Routledge's measures
turnover(trees, method = c("routledge1", "routledge2", "routledge3"),
         simplify = TRUE)
# 0.29 0.56 1.75

# Wilson and Shmida's measure
turnover(trees, "wilson") # 1
```

zuni

*Zuni ceramics***Description**

A dataset containing ceramic counts from the Zuni region of the American Southwest.

Usage

zuni

Format

A numeric matrix with 420 rows (assemblages) and 18 variables (ceramic types). The numbers in brackets correspond to the date range of each type (in AD years):

LINO Lino Gray (575-875).

KIAT Kiatuthlanna Black-on-white (850-910).

RED Red Mesa Black-on-white (900-1030).

GALL Gallup Black-on-white (1025-1150).

ESC Escavada Black-on-white (1050-1150).

PUBW Puerco Black-on-white (1050-1200).

RES Reserve Black-on-white (1071-1115).

TULA Tularosa Black-on-white (1175-1300).

PINE Pinedale Black-on-white (1275-1325).

PUBR Puerco Black-on-red (1050-1200).

WING Wingate Black-on-red (1070-1200).

WIPO Wingate Polychrome (1150-1250).

SJ St. Johns Black-on-red/Polychrome (1200-1300).

LSJ St. Johns glaze, Techado Polychrome (1275-1300).

SPR Springerville Polychrome (1250-1300).

PINER Pinedale Black-on-red/Polychrome (1275-1325).

HESH Heshotauthla Polychrome (1285-1400).

KWAK Kwakina Polychrome (1285-1400).

Source

Peeples, M. A., & Schachner, G. (2012). Refining correspondence analysis-based ceramic seriation of regional data sets. *Journal of Archaeological Science*, 39(8), 2818-2827. DOI: [10.1016/j.jas.2012.04.040](https://doi.org/10.1016/j.jas.2012.04.040).

Index

*Topic **datasets**

- boves, [4](#)
- compiegne, [5](#)
- mississippi, [18](#)
- zuni, [39](#)
- [,BootCA-method (BootCA-class), [3](#)
- [,DateModel-method (DateModel-class), [10](#)
- [[,BootCA-method (BootCA-class), [3](#)
- [[,DateModel-method (DateModel-class), [10](#)
- [[,PermutationOrder-method (PermutationOrder-class), [19](#)

- accessors, [2](#)
- as.integer, [3](#), [6](#), [10](#)
- as.logical, [17](#)
- as.vector, [6](#), [17](#)

- BootCA, [31](#)
- BootCA-class, [3](#)
- boves, [4](#)

- CA, [7](#), [33](#)
- character, [3](#), [10](#), [12](#), [19](#), [20](#), [22](#), [25](#), [27](#), [31](#), [34](#), [37](#)
- compiegne, [5](#)
- CountMatrix, [15](#), [31](#)
- CountMatrix (CountMatrix-class), [6](#)
- CountMatrix-class, [6](#)

- data.frame, [3](#), [11](#)
- dateEvent, [7](#), [23](#)
- dateEvent, CountMatrix-method (dateEvent), [7](#)
- dateEvent-method (dateEvent), [7](#)
- DateModel, [8](#), [22](#)
- DateModel-class, [10](#)
- dimnames, [6](#), [17](#)
- diversity, [12](#), [29](#), [35](#), [38](#)
- diversity, CountMatrix-method (diversity), [12](#)

- diversity-method (diversity), [12](#)

- evenness, [35](#), [38](#)
- evenness (diversity), [12](#)
- evenness, CountMatrix-method (diversity), [12](#)
- evenness-method (diversity), [12](#)
- expression, [6](#), [17](#)

- FrequencyMatrix, [31](#)
- FrequencyMatrix-class, [15](#)
- function, [26](#)

- IncidenceMatrix, [31](#)
- IncidenceMatrix (IncidenceMatrix-class), [16](#)
- IncidenceMatrix-class, [16](#)
- integer, [3](#), [6](#), [7](#), [10](#), [20](#), [31](#)

- logical, [7](#), [10](#), [12](#), [17](#), [20](#), [23](#), [25](#), [27](#), [28](#), [31](#), [37](#)

- matrigraphe (plotMatrix), [23](#)
- matrix, [6](#), [16](#), [17](#), [19](#), [36](#)
- missing, [3](#), [10](#)
- mississippi, [18](#)

- NULL, [3](#), [10](#)
- numeric, [3](#), [7](#), [10](#), [11](#), [15](#), [20](#), [22](#), [27](#), [31](#)

- OccurrenceMatrix-class, [18](#)

- pbapply, [8](#), [32](#)
- PermutationOrder, [31](#)
- PermutationOrder-class, [19](#)
- permute (seriation), [30](#)
- permute, CountMatrix, PermutationOrder-method (seriation), [30](#)
- permute, IncidenceMatrix, PermutationOrder-method (seriation), [30](#)
- permute-method (seriation), [30](#)

- plotBar, [20](#), [23–26](#)
- plotBar, CountMatrix-method (plotBar), [20](#)
- plotBar, FrequencyMatrix-method (plotBar), [20](#)
- plotBar-method (plotBar), [20](#)
- plotDate, [21](#), [22](#), [24–26](#)
- plotDate, DateModel-method (plotDate), [22](#)
- plotDate-method (plotDate), [22](#)
- plotMatrix, [21](#), [23](#), [23](#), [25](#), [26](#)
- plotMatrix, CountMatrix-method (plotMatrix), [23](#)
- plotMatrix, FrequencyMatrix-method (plotMatrix), [23](#)
- plotMatrix, IncidenceMatrix-method (plotMatrix), [23](#)
- plotMatrix-method (plotMatrix), [23](#)
- plotRank, [21](#), [23](#), [24](#), [24](#), [26](#)
- plotRank, CountMatrix-method (plotRank), [24](#)
- plotRank, FrequencyMatrix-method (plotRank), [24](#)
- plotRank-method (plotRank), [24](#)
- plotSpot, [21](#), [23–25](#), [25](#)
- plotSpot, CountMatrix-method (plotSpot), [25](#)
- plotSpot, FrequencyMatrix-method (plotSpot), [25](#)
- plotSpot, OccurrenceMatrix-method (plotSpot), [25](#)
- plotSpot, SimilarityMatrix-method (plotSpot), [25](#)
- plotSpot-method (plotSpot), [25](#)

- rarefaction, [35](#), [38](#)
- rarefaction (richness), [27](#)
- rarefaction, CountMatrix-method (richness), [27](#)
- rarefaction-method (richness), [27](#)
- refine, [9](#), [32](#), [33](#)
- refine (seriation), [30](#)
- refine, CountMatrix-method (seriation), [30](#)
- refine-method (seriation), [30](#)
- richness, [14](#), [27](#), [35](#), [38](#)
- richness, CountMatrix-method (richness), [27](#)
- richness, IncidenceMatrix-method (richness), [27](#)
- richness-method (richness), [27](#)

- seriate (seriation), [30](#)
- seriate, CountMatrix, BootCA-method (seriation), [30](#)
- seriate, CountMatrix, missing-method (seriation), [30](#)
- seriate, IncidenceMatrix, missing-method (seriation), [30](#)
- seriate-method (seriation), [30](#)
- seriation, [21](#), [30](#)
- seriographe (plotBar), [20](#)
- similarity, [14](#), [29](#), [33](#), [38](#)
- similarity, CountMatrix-method (similarity), [33](#)
- similarity, IncidenceMatrix-method (similarity), [33](#)
- similarity-method (similarity), [33](#)
- SimilarityMatrix-class, [36](#)

- totals (accessors), [2](#)
- totals, FrequencyMatrix-method (FrequencyMatrix-class), [15](#)
- turnover, [14](#), [29](#), [35](#), [36](#)
- turnover, CountMatrix-method (turnover), [36](#)
- turnover, IncidenceMatrix-method (turnover), [36](#)
- turnover-method (turnover), [36](#)

- zuni, [39](#)