Package 'tabula'

March 19, 2020

Title Analysis, Seriation and Visualization of Archaeological Count
Data

Version 1.5.1

Maintainer Nicolas Frerebeau

<nicolas.frerebeau@u-bordeaux-montaigne.fr>

Description An easy way to examine archaeological count data. This package
provides a convenient and reproducible toolkit for relative and absolute
dating and analysis of (chronological) patterns. It includes functions for
matrix seriation (reciprocal ranking, CA-based seriation), chronological
modeling and dating of archaeological assemblages and/or objects.
Beyond these, the package provides several tests and measures of diversity:
heterogeneity and evenness (Brillouin, Shannon, Simpson, etc.),
richness and rarefaction (Chao1, Chao2, ACE, ICE, etc.),
turnover and similarity (Brainerd-Robinson, etc.).
The package make it easy to visualize count data and statistical thresholds:
rank vs. abundance plots, heatmaps, Ford (1962) and Bertin (1977) diagrams.

License GPL-3

URL <http://tabula.archaeo.science>,

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

<https://cran.r-project.org/package=tabula>

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

Depends R (>= 3.2), arkhe

Imports ca (>= 0.71), ggplot2 (>= 3.2.1), grDevices, methods, rlang
(>= 0.4.0), stats, utils

Suggests covr (>= 3.3.1), khroma (>= 1.3.0), knitr, magrittr (>= 1.5),
pbapply (>= 1.4-2), rmarkdown, testthat (>= 2.2.1), vdiffr (>=
0.3.1)

VignetteBuilder knitr

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

Collate 'utilities.R' 'AllClasses.R' 'AllGenerics.R' 'coerce.R'
    'data.R' 'date.R' 'deprecate.R' 'extract.R' 'index-diversity.R'
    'index-heterogeneity.R' 'index-rarefaction.R'
    'index-richness.R' 'index-similarity.R' 'index-turnover.R'
    'plot-prepare.R' 'plot-bar.R' 'plot-date.R' 'plot-diversity.R'
    'plot-matrix.R' 'plot-rank.R' 'plot-spot.R' 'plot-time.R'
    'refine.R' 'seriation.R' 'seriate.R' 'show.R' 'statistics.R'
    'tabula-package.R' 'test.R' 'validate.R' 'zzz.R'

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>),
    Ben Marwick [ctb] (<https://orcid.org/0000-0001-7879-4531>),
    Anne Philippe [ctb],
    Jean-Baptiste Fourvel [ctb] (<https://orcid.org/0000-0002-1061-4642>)

Repository CRAN

Date/Publication 2020-03-19 11:40:03 UTC

R **topics documented:**

---

access   *Get or Set Parts of an Object*

Description

Getters and setters to extract or replace parts of an object.

Usage

```
get_id(object)
```

Arguments

object   An object from which to get or set element(s).

Value

An object of the same sort as object with the new values assigned.

Author(s)

N. Frerebeau

See Also

Other mutator: subset()

BootCA-class   *Partial bootstrap CA*

Description

An S4 class to store partial bootstrap correspondence analysis results.

Slots

id A character string specifying the unique identifier of the corresponding matrix (UUID v4).

rows A list of length three giving the vertices coordinates (x, y) of the samples convex hull and a identifier (id) to link each row to a sample.

columns A list of length three giving the vertices coordinates (x, y) of the variables convex hull and a identifier (id) to link each row to a variable.

lengths A list of two named numeric vectors giving the convex hull maximum dimension length of samples and variables, respectively.

cutoff A length-two numeric vector giving the cutoff value for samples and variables selection, respectively.

keep A list of two integer vectors giving the subscript of the samples and variables to be kept, respectively.

Subset

In the code snippets below, x is a `BootCA` object.

`x[i, j, drop]` Extracts informations from a slot selected by subscript i thru j (see examples). i should be one of `"rows"` or `"columns"`. Any unambiguous substring can be given. j is a [numeric](#), [integer](#) or [character](#) vector or empty (missing) or `NULL`. Numeric values are coerced to [integer](#) as by [as.integer](#) (and hence truncated towards zero). Character vectors will be matched to the name of the elements. An empty index (a comma separated blank) indicates that all entries in that dimension are selected.

`x[[i]]` Extracts informations from a slot selected by subscript i. i should be one of `"id"`, `"rows"`, `"columns"`, `"lengths"`, `"cutoff"` or `"keep"`. Any unambiguous substring can be given.

Author(s)

N. Frerebeau

See Also

Other class: [DateModel-class](#), [DiversityIndex](#), [PermutationOrder-class](#)

---

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.

See Also

Other datasets: compiegne, merzbach, mississippi, zuni

---

compiegne	*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

```
compiegne
```

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

See Also

Other datasets: boves, merzbach, mississippi, zuni

---

date   *Date Archaeological Assemblages*

Description

date_mcd estimates the Mean Ceramic Date of an assemblage.

Usage

```
date_mcd(object, ...)

date_event(object, ...)

refine_dates(object, ...)

## S4 method for signature 'CountMatrix'
date_mcd(object, dates, errors = NULL, level = 0.95, n = 1000, ...)

## S4 method for signature 'CountMatrix'
date_event(object, level = 0.95, cutoff = 90, ...)

## S4 method for signature 'DateModel'
refine_dates(object, method = c("jackknife", "bootstrap"), n = 1000, ...)
```

Arguments

object	A $m \times p$ matrix of count data (typically of class CountMatrix).
...	Further arguments to be passed to internal methods.
dates	A length-p numeric vector giving the mid-date of each type (year AD).
errors	A length-p numeric vector giving the absolute error of each date.
level	A length-one numeric vector giving the confidence level.
n	A non-negative integer giving the number of bootstrap replications (see below).
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.
method	A character string specifying the resampling method to be used. This must be one of "jackknife", "bootstrap" (see details). Any unambiguous substring can be given.

Details

`date_event` estimates the event and accumulation dates of an assemblage.

`refine_dates` checks the stability of a date model with resampling methods.

Value

`date_mcd` returns a [`data.frame`](#) with the following columns:

id An identifier to link each row to an assemblage.

date The Mean Ceramic Date.

error The error on the MCD.

lower The lower boundary of the confidence interval.

upper The upper boundary of the confidence interval.

`date_event` returns an object of class [DateModel](#).

`refine_dates` returns a [`data.frame`](#).

Mean Ceramic Date

The Mean Ceramic Date (MCD) is a point estimate of the occupation of an archaeological site (South 1977). The MCD is estimated as the weighted mean of the date midpoints of the ceramic types (based on absolute dates or the known production interval) found in a given assemblage. The weights are the relative frequencies of the respective types in the assemblage.

A bootstrapping procedure is used to estimate the confidence interval of a given MCD. For each assemblage, a large number of new bootstrap replicates is created, with the same sample size, by resampling the original assemblage with replacement. MCDs are calculated for each replicates and upper and lower boundaries of the confidence interval associated with each MCD are then returned. Confidence interval are not estimated for assemblages with only a single type (NAs are returned).

Event and Accumulation Dates

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

Event and accumulation dates are density estimates of the occupation and duration of an archaeological site (Bellanger and Husi 2012, 2013). The event date is an estimation of the *terminus post-quem* of an archaeological assemblage. The accumulation date represents the "chronological profile" of the assemblage. According to Bellanger and Husi (2012), accumulation date can be interpreted "at best [...] as a formation process reflecting the duration or succession of events on the scale of archaeological time, and at worst, as imprecise dating due to contamination of the context by residual or intrusive material." In other words, accumulation dates estimate occurrence of archaeological events and rhythms of the long term.

This method relies on strong archaeological and statistical assumptions. Use it only if you know what you are doing (see references below and the vignette: `utils::vignette("dating",package = "tabula")`).

Date Model Checking

refine_date can be used to check the stability of the resulting DateModel with resampling methods.

If jackknife is used, 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. A six columns data.frame is returned, 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.

date The jackknife event date estimate.

lower The lower boundary of the associated prediction interval.

upper The upper boundary of the associated prediction interval.

error The standard error of predicted means.

bias The jackknife estimate of bias.

If bootstrap is used, 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. A six columns data.frame is returned, 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.

Note

Bellanger *et al.* did not publish the data supporting their demonstration: no replication of their results is possible and this implementation must be considered **experimental**. date_event may be subject to major changes in a future release.

Refining methods 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.

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.

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.

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

set_dates

Examples

```
## Mean Ceramic Date
## Coerce the zuni dataset to an abundance (count) matrix
zuni_counts <- as_count(zuni)

## Set the start and end dates for each ceramic type
zuni_dates <- list(
  LINO = c(600, 875), KIAT = c(850, 950), RED = c(900, 1050),
  GALL = c(1025, 1125), ESC = c(1050, 1150), PUBW = c(1050, 1150),
  RES = c(1000, 1200), TULA = c(1175, 1300), PINE = c(1275, 1350),
  PUBR = c(1000, 1200), WING = c(1100, 1200), WIPO = c(1125, 1225),
  SJ = c(1200, 1300), LSJ = c(1250, 1300), SPR = c(1250, 1300),
  PINER = c(1275, 1325), HESH = c(1275, 1450), KWAK = c(1275, 1450)
)

## Calculate date midpoints and errors
zuni_mid <- vapply(X = zuni_dates, FUN = mean, FUN.VALUE = numeric(1))
zuni_error <- vapply(X = zuni_dates, FUN = diff, FUN.VALUE = numeric(1)) / 2

## Calculate MCD
## (we use a bootstrapping procedure to estimate the confidence interval)
zuni_mcd <- date_mcd(zuni_counts, dates = zuni_mid, errors = zuni_error)
head(zuni_mcd)

## Plot dates
keep_sites <- c("CS11", "CS12", "CS144", "CS195", "CS40", "LZ0219", "LZ0280",
```

```
                "LZ0367", "LZ0508", "LZ0560", "LZ1076", "LZ1087")
set_dates(zuni_counts) <- list(value = zuni_mcd$date, error = zuni_mcd$error)
plot_date(zuni_counts, select = keep_sites, sort = "asc") +
  ggplot2::theme_bw()

## Event and accumulation dates (Bellanger et al.)
## See the vignette:

utils::vignette("dating", package = "tabula")
```

DateModel-class   *Date Model*

Description

An S4 class to store the event and accumulation times of archaeological assemblages.

Slots

id A [character](#) string specifying the unique identifier of the corresponding matrix (UUID v4).

counts A numeric matrix of count data.

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.

rows A four columns [numeric](#) matrix giving the predicted event dates for each archaeological assemblage, with the following columns:

   date The event date estimation.

   lower The lower boundary of the confidence interval.

   upper The upper boundary of the confidence interval.

   error The standard error of predicted dates.

columns A four columns [numeric](#) matrix giving the predicted event dates for each archaeological type or fabric, with the following columns:

   date The event date estimation.

   lower The lower boundary of the confidence interval.

   upper The upper boundary of the confidence interval.

   error The standard error of predicted dates.

accumulation A two columns [numeric](#) matrix giving the point estimate of the accumulation dates and the corresponding error.

Subset

x[[i]] Extracts informations from a slot selected by subscript i. i is a length-one [character](#) vector. Returns the corresponding slot values.

Author(s)

N. Frerebeau

See Also

Other class: [BootCA-class](), [DiversityIndex](), [PermutationOrder-class]()

DiversityIndex *Diversity Index*

Description

An S4 class to represent a diversity measure.

Slots

id A [character]() string specifying the unique identifier of the corresponding matrix (UUID v4).

index A [numeric]() vector giving the diversity index values.

size A [integer]() vector giving the sample sizes.

jackknife A numeric [matrix]() vector giving the jackknifed estimates.

boostrap A numeric [matrix]() vector giving the boostraped estimates.

simulated A numeric [matrix]() vector giving the diversity measures for the simulated assemblage.

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

Subset

In the code snippets below, x is a DiversityIndex object.

x[[i]] Extracts informations from a slot selected by subscript i. i is a length-one [character]() vector. Returns the corresponding slot values.

Author(s)

N. Frerebeau

See Also

Other class: [BootCA-class](), [DateModel-class](), [PermutationOrder-class]()

---

heterogeneity-index   *Heterogeneity and Evenness*

Description

`index_heterogeneity` returns an heterogeneity or dominance index.

`index_evenness` returns an evenness measure.

Usage

```
index_heterogeneity(object, ...)

index_evenness(object, ...)

## S4 method for signature 'CountMatrix'
index_heterogeneity(
  object,
  method = c("berger", "brillouin", "mcintosh", "shannon", "simpson"),
  jackknife = TRUE,
  bootstrap = TRUE,
  simulate = FALSE,
  level = 0.8,
  n = 1000,
  ...
)

## S4 method for signature 'CountMatrix'
index_evenness(
  object,
  method = c("shannon", "brillouin", "mcintosh", "simpson"),
  jackknife = TRUE,
  bootstrap = TRUE,
  simulate = FALSE,
  level = 0.8,
  n = 1000,
  ...
)
```

Arguments

`object`	A $m \times p$ matrix of count data (typically a CountMatrix object).
`...`	Further arguments to be passed to internal methods.
`method`	A character string specifying the index to be computed (see details). Any unambiguous substring can be given.
`jackknife`	A logical scalar: should jackknifed estimates be computed?

bootstrap	A [logical](#) scalar: should boostraped estimates be computed?
simulate	A [logical](#) scalar: simulated assemblages be computed?
level	A length-one [numeric](#) vector giving the confidence level.
n	A non-negative [integer](#) giving the number of bootstrap replications.

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

index_heterogeneity and index_evenness return a [DiversityIndex](#) object.

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.

Brillouin, L. (1956). *Science and information theory*. New York: Academic Press.

Kintigh, K. W. (1984). Measuring Archaeological Diversity by Comparison with Simulated Assemblages. *American Antiquity*, 49(1), 44-54. DOI: 10.2307/280511.

Kintigh, K. W. (1989). Sample Size, Significance, and Measures of Diversity. In Leonard, R. D. and Jones, G. T., *Quantifying Diversity in Archaeology*. New Directions in Archaeology. Cambridge: Cambridge University Press, p. 25-36.

Laxton, R. R. (1978). The measure of diversity. *Journal of Theoretical Biology*, 70(1), 51-67. DOI: 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.

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.

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.

Pielou, E. C. (1975). *Ecological Diversity*. New York: Wiley. DOI: 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.

Simpson, E. H. (1949). Measurement of Diversity. *Nature*, 163(4148), 688-688. DOI: 10.1038/163688a0.

See Also

turnover similarity

Other diversity: richness-index, similarity(), turnover-index

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))
```

```
index_heterogeneity(birds, "shannon") # 2.40 2.06
index_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)

index_heterogeneity(moths, "brillouin") # 1.88
index_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
)

index_heterogeneity(trees, "simpson") # 1.19
index_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
)

index_heterogeneity(invertebrates, "mcintosh") # 0.71
index_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)
)

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

---

merzbach   *Merzbach ceramics*

Description

A dataset containing the ceramic counts from the Merzbach assemblage (Germany). The data are grouped into eight phases.

Usage

```
merzbach
```

Format

A numeric matrix with 8 rows (phases, numbered from VII to XIV) and 36 variables (pottery motifs).

BT1, BT2, ... Counts of a motif.

Source

Crema, E. R., Kandler, A. & Shennan, S. (2016). Revealing Patterns of Cultural Transmission from Frequency Data: Equilibrium and Non-Equilibrium Assumptions. *Scientific Reports*, 6(1). DOI: 10.1038/srep39122.

Crema, E. R. (2016). Sample codes and data for "Revealing patterns of cultural transmission from frequency data: equilibrium and non-equilibrium assumptions". *Zenodo*, v1.0. DOI: 10.5281/zenodo.187558.

See Also

Other datasets: boves, compiegne, mississippi, zuni

---

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.

See Also

Other datasets: boves, compiegne, merzbach, zuni

PermutationOrder-class

Permutation order

Description

An S4 class to represent a permutation order.

Slots

id A character string specifying the unique identifier of the corresponding matrix (UUID v4).

rows An integer vector giving the rows permutation.

columns An integer vector giving the columns permutation.

method A character string indicating the seriation method used.

Subset

x[[i]] Extracts informations from a slot selected by subscript i. i should be one of "id", "rows", "columns" or "method". Any unambiguous substring can be given.

Note

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

Author(s)

N. Frerebeau

See Also

Other class: BootCA-class, DateModel-class, DiversityIndex

---

plot_bar   *Bar Plot*

Description

Plots a Bertin, Ford (battleship curve) or Dice-Leraas diagram.

Usage

```
plot_bertin(object, ...)

plot_ford(object, ...)

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

## S4 method for signature 'CountMatrix'
plot_ford(object, EPPM = FALSE)
```

Arguments

object	An object to be plotted.
...	Currently not used.
threshold	A function that takes a numeric vector as argument and returns a numeric threshold value (see below). If NULL (the default), no threshold is computed.
scale	A function used to scale each variable, that takes a numeric vector as argument and returns a numeric vector. If NULL (the default), no scaling is performed.
EPPM	A logical scalar: should the EPPM be drawn (see details)?

Value

A ggplot object.

Bertin Matrix

As de Falguerolles *et al.* (1997) points out: "In abstract terms, a Bertin matrix is a matrix of displays. [...] To fix ideas, think of a data matrix, variable by case, with real valued variables. For each variable, draw a bar chart of variable value by case. High-light all bars representing a value above some sample threshold for that variable."

Ford Diagram

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 useful graphical tool to explore significance of relationship between rows and columns related to seriation (Desachy 2004).

Author(s)

N. Frerebeau

References

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

de Falguerolles, A., Friedrich, F. & Sawitzki, G. (1997). A Tribute to J. Bertin's Graphical Data Analysis. In W. Badilla & F. Faulbaum (eds.), *SoftStat '97: Advances in Statistical Software 6.* Stuttgart: Lucius & Lucius, p. 11-20.

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.

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

See Also

Other plot: plot_date(), plot_diversity(), plot_line, plot_matrix, plot_spot()

Examples

```
## Abundance data
## Coerce dataset to a count matrix
mississippi_count <- as_count(mississippi)

## Plot a Bertin diagram...
## ...without threshold
plot_bertin(mississippi_count, threshold = NULL)
## ...with the variable mean as threshold
plot_bertin(mississippi_count, threshold = mean)

## ...with variables scaled to 0-1
scale_01 <- function(x) (x - min(x)) / (max(x) - min(x))
plot_bertin(mississippi_count, threshold = mean, scale = scale_01)

## Abundance data
## Coerce dataset to a count matrix (data from Desachy 2004)
compiegne_count <- as_count(compiegne)

## Plot a Ford diagram...
## ...without threshold
plot_ford(compiegne_count)
## ...with EPPM
plot_ford(compiegne_count, EPPM = TRUE)
```

plot_date *Date and Time Plot*

Description

plot_date produces an activity or tempo plot.

Usage

```
plot_date(object, ...)

plot_time(object, ...)

## S4 method for signature 'Matrix'
plot_date(object, select = NULL, sort = "dsc")

## S4 method for signature 'DateModel'
plot_date(
  object,
  type = c("activity", "tempo"),
  event = FALSE,
  select = 1,
  n = 500
)

## S4 method for signature 'CountMatrix'
plot_time(
  object,
  highlight = NULL,
  level = 0.95,
  roll = FALSE,
  window = 5,
  facet = TRUE,
  ...
)
```

Arguments

object	An object of class [DateModel](#) to be plotted.
...	Further arguments to be passed to internal methods.
select	A [numeric](#) or [character](#) vector giving the selection of the assemblage that are drawn.
sort	A [character](#) string indicating whether the dates should be sorted. It can be one of "asc" or "dsc" (default). Any unambiguous substring can be given. If NULL no sorting is performed.

type	A [character](#) string indicating the type of plot. It must be one of "activity" (default) or "tempo". Any unambiguous substring can be given.
event	A [logical](#) scalar: should the distribution of the event date be displayed? Only used if type is "activity".
n	A length-one non-negative [numeric](#) vector giving the desired length of the vector of quantiles for density computation.
highlight	A [character](#) string indicating the type of plot. It must be one of "FIT" or NULL (default). Any unambiguous substring can be given.
level	A length-one [numeric](#) vector giving the confidence level.
roll	A [logical](#) scalar: should each time series be subsetted to look for episodes of selection? Only used if highlight is "FIT" (see details).
window	An odd [integer](#) giving the size of the rolling window. Only used if roll is TRUE.
facet	A [logical](#) scalar: should a matrix of panels defined by type/taxon be drawn? Only used if highlight is NULL.

Details

plot_time produces an abundance *vs.* time diagram.

Value

A [ggplot](#) object.

Event and Acccumulation Dates

plot_date plots the probability estimate density curves of archaeological assemblage dates (*event* and *accumulation* dates; Bellanger and Husi 2012). The *event date* is plotted as a line, while the *accumulation date* is shown as a grey filled area.

The accumulation date can be displayed as a tempo plot (Dye 2016) or an activity plot (Philippe and Vibet 2017):

Tempo plot A tempo plot estimates the cumulative occurrence of archaeological events, such as the slope of the plot directly reflects the pace of change.

Activity plot An activity plot displays the first derivative of the tempo plot.

Detection of Selective Processes

Results of the frequency increment test can be displayed on an abundance *vs.* time diagram aid in the detection and quantification of selective processes in the archaeological record. If roll is TRUE, each time series is subsetted according to window to see if episodes of selection can be identified among decoration types that might not show overall selection. If so, shading highlights the data points where [test_fit](#) identifies selection.

Note

Displaying FIT results on an abundance *vs.* time diagram is adapted from Ben Marwick's original [idea](#).

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.

Dye, T. S. (2016). Long-Term Rhythms in the Development of Hawaiian Social Stratification. *Journal of Archaeological Science*, 71, 1-9. DOI: 10.1016/j.jas.2016.05.006.

Philippe, A. & Vibet, M.-A. (2017). Analysis of Archaeological Phases using the CRAN Package ArchaeoPhases. HAL, hal-01347895, v3.

See Also

date_event, test

Other plot: plot_bar, plot_diversity(), plot_line, plot_matrix, plot_spot()

Examples

```
## Coerce the merzbach dataset to a count matrix
## Keep only decoration types that have a maximum frequency of at least 50
keep <- apply(X = merzbach, MARGIN = 2, FUN = function(x) max(x) >= 50)
merzbach_count <- as_count(merzbach[, keep])

## The data are grouped by phase
## We use the row names as time coordinates (roman numerals)
set_dates(merzbach_count) <- rownames(merzbach)
## Plot time vs abundance
plot_time(merzbach_count)
## Plot time vs abundance and highlight selection
plot_time(merzbach_count, highlight = "FIT", roll = TRUE)
```

---

plot_diversity   *Diversity Plot*

Description

Diversity Plot

Usage

```
plot_diversity(object, ...)

## S4 method for signature 'DiversityIndex'
plot_diversity(object)
```

Arguments

object	A DiversityIndex object to be plotted.
...	Currently not used.

Author(s)

N. Frerebeau

See Also

index_heterogeneity, index_evenness, index_richness

Other plot: `plot_bar`, `plot_date()`, `plot_line`, `plot_matrix`, `plot_spot()`

Examples

```
library(magrittr)

merzbach %>%
  as_count() %>%
  index_evenness(method = "shannon", simulate = FALSE) %>%
  plot_diversity()


## Assemblage diversity size comparison
## Warning: this may take a few seconds!
merzbach %>%
  as_count() %>%
  index_evenness(method = "shannon", simulate = TRUE) %>%
  plot_diversity()

merzbach %>%
  as_count() %>%
  index_richness(method = "none", simulate = TRUE) %>%
  plot_diversity()
```

---

plot_line	*Line Plot*

Description

`plot_rank` plots a rank *vs* relative abundance diagram.

Usage

```
plot_rank(object, ...)

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

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

Arguments

object	An object to be plotted.
...	Further arguments to be passed to internal 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

TODO

Value

A ggplot object.

Author(s)

N. Frerebeau

References

Magurran, A. E. (1988). *Ecological Diversity and its Measurement*. Princeton, NJ: Princeton University Press. DOI: 10.1007/978-94-015-7358-0.

See Also

Other plot: `plot_bar`, `plot_date()`, `plot_diversity()`, `plot_matrix`, `plot_spot()`

Examples

```
## Abundance matrix
## Coerce datasets to a count matrix (data from Desachy 2004)
compiegne_count <- as_count(compiegne)

## Plot rank vs abundance
plot_rank(compiegne_count)
plot_rank(compiegne_count, facet = FALSE)
```

---

plot_matrix *Heatmap*

Description

Plots a heatmap.

Usage

```
plot_heatmap(object, ...)

## S4 method for signature 'CountMatrix'
plot_heatmap(object, PVI = FALSE, frequency = TRUE)

## S4 method for signature 'AbundanceMatrix'
plot_heatmap(object)

## S4 method for signature 'IncidenceMatrix'
plot_heatmap(object)
```

Arguments

object	An object to be plotted.
...	Further arguments to be passed to internal methods.
PVI	A [logical](#) scalar: should the PVI be drawn instead of frequencies (see details)?
frequency	A [logical](#) scalar: should relative frequencies be drawn? If FALSE, raw data are plotted.

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).

Value

A [ggplot](#) object.

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.

See Also

Other plot: plot_bar, plot_date(), plot_diversity(), plot_line, plot_spot()

Examples

```
## Abundance data
## Coerce dataset to a count matrix (data from Desachy 2004)
compiegne_count <- as_count(compiegne)

## Plot matrix diagram...
## ...without threshod (i.e. heatmap)
plot_heatmap(compiegne_count)
## ...with PVI as threshold (i.e. Bruno Desachy's matrigraphe)
plot_heatmap(compiegne_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_heatmap(bin) +
  ggplot2::scale_fill_manual(values = c("TRUE" = "black", "FALSE" = "white"))
```

---

plot_spot   *Spot Plot*

Description

Plots a spot matrix.

Usage

```
plot_spot(object, ...)

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

## S4 method for signature 'AbundanceMatrix'
plot_spot(object, threshold = NULL)

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

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

Arguments

`object`	An object to be plotted.
`...`	Currently not used.
`threshold`	A [function](#) that takes a numeric vector as argument and returns a numeric threshold value. If `NULL` (the default), no threshold is computed.

Details

The spot matrix can be considered as a variant of the [Bertin diagram](#) where the data are first transformed to relative frequencies.

Value

A [ggplot](#) object.

Note

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

Author(s)

N. Frerebeau

See Also

Other plot: [plot_bar](#), [plot_date](#)(), [plot_diversity](#)(), [plot_line](#), [plot_matrix](#)

Examples

```
## Plot spot diagram...

## ...of count data...
mississippi_count <- as_count(mississippi)
### ...without threshod
plot_spot(mississippi_count)
### ...with the  column means as threshold
plot_spot(mississippi_count, threshold = mean)
### ...with the column medians as threshold
plot_spot(mississippi_count, threshold = median)

## ...of a similarity matrix
sim <- similarity(mississippi_count, method = "brainerd")
plot_spot(sim)

## ...of a co-occurrence matrix
occ <- as_occurrence(mississippi)
plot_spot(occ)
```

richness-index   *Richness and Rarefaction*

Description

index_richness returns sample richness. index_composition returns asymptotic species richness.

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

Usage

```
index_richness(object, ...)

index_composition(object, ...)

rarefaction(object, ...)

## S4 method for signature 'CountMatrix'
rarefaction(object, sample, method = c("hurlbert"), simplify = FALSE, ...)

## S4 method for signature 'CountMatrix'
index_richness(
  object,
  method = c("none", "margalef", "menhinick"),
  jackknife = TRUE,
  bootstrap = TRUE,
  simulate = FALSE,
  level = 0.8,
  n = 1000,
  ...
)

## S4 method for signature 'CountMatrix'
index_composition(
  object,
  method = c("chao1", "ace"),
  unbiased = FALSE,
  improved = FALSE,
  k = 10
)

## S4 method for signature 'IncidenceMatrix'
index_composition(
  object,
  method = c("chao2", "ice"),
  unbiased = FALSE,
```

```
  improved = FALSE,
  k = 10
)
```

Arguments

object	A $m \times p$ matrix of count data.
...	Further arguments to be passed to internal 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.
simplify	A [logical](#) scalar: should the result be simplified to a matrix? The default value, FALSE, returns a list.
jackknife	A [logical](#) scalar: should jackknifed estimates be computed?
bootstrap	A [logical](#) scalar: should boostraped estimates be computed?
simulate	A [logical](#) scalar: simulated assemblages be computed?
level	A length-one [numeric](#) vector giving the confidence level. Only used if simulate is TRUE.
n	A non-negative [integer](#) giving the number of bootstrap replications. Only used if simulate is TRUE.
unbiased	A [logical](#) scalar. Should the bias-corrected estimator be used? Only used with "chao1" or "chao2" (improved) estimator.
improved	A [logical](#) scalar. Should the improved estimator be used? Only used with "chao1" or "chao2".
k	A length-one [numeric](#) vector giving the threshold between rare/infrequent and abundant/frequent species. Only used if method is "ace" or "ice".

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:

margalef Margalef richness index.

menhinick Menhinick richness index.

none Returns the number of observed taxa/types.

Value

`index_richness` and `index_composition` return a [DiversityIndex](#) object.

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

Asymptotic Species Richness

The following measures are available for count data:

ace Abundance-based Coverage Estimator.

chao1 (improved/unbiased) Chao1 estimator.

The following measures are available for replicated incidence data:

ice Incidence-based Coverage Estimator.

chao2 (improved/unbiased) Chao2 estimator.

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](#).

Chao, A. & Chiu, C.-H. (2016). Species Richness: Estimation and Comparison. *In* Balakrishnan, N., Colton, T., Everitt, B., Piegorsch, 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](#)

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](#).

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](#).

Hurlbert, S. H. (1971). The Nonconcept of Species Diversity: A Critique and Alternative Parameters. *Ecology*, 52(4), 577-586. DOI: [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](#).

Magurran, A E. & Brian J. McGill (2011). *Biological Diversity: Frontiers in Measurement and Assessment*. Oxford: Oxford University Press.

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](#).

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](#).

Sander, H. L. (1968). Marine Benthic Diversity: A Comparative Study. *The American Naturalist*, 102(925), 243-282.

See Also

Other diversity: [heterogeneity-index](), [similarity]()(), [turnover-index]()

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))
index_richness(trap, method = "margalef")
## A 2.55 1.88
index_richness(trap, method = "menhinick")
## B 1.95 1.66

## Asymptotic species richness
## 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
)

index_composition(brazil, method = c("chao1"), unbiased = FALSE)
## 461.625
index_composition(brazil, method = c("ace"), k = 10)
## 445.822

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

richness(trap, method = c("margalef", "menhinick", "chao1"))
```

---

seriation	*Matrix Seriation*

Description

`seriate_*` computes a permutation order for rows and/or columns.

`permute` rearranges a data matrix according to a permutation order.

`get_order` returns the seriation order for rows and columns.

`refine_seriation` performs a partial bootstrap correspondence analysis seriation.

Usage

```
seriate_reciprocal(object, ...)

seriate_correspondence(object, subset, ...)

permute(object, order, ...)

refine_seriation(object, ...)

get_order(object)

## S4 method for signature 'PermutationOrder'
get_order(object)

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

## S4 method for signature 'CountMatrix'
seriate_reciprocal(object, EPPM = FALSE, margin = c(1, 2), stop = 100)

## S4 method for signature 'IncidenceMatrix'
seriate_reciprocal(object, margin = c(1, 2), stop = 100)

## S4 method for signature 'CountMatrix,missing'
seriate_correspondence(object, margin = c(1, 2), ...)

## S4 method for signature 'IncidenceMatrix,missing'
seriate_correspondence(object, margin = c(1, 2), ...)

## S4 method for signature 'CountMatrix,BootCA'
seriate_correspondence(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 (typically an object of class [CountMatrix](#) or [Incidence-Matrix](#).
...	Further arguments to be passed to internal methods.
subset	A [BootCA](#) object giving the subset of `object` to be used.
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 below).

n	A non-negative [integer](#) giving the number of partial bootstrap replications (see below).
axes	A [numeric](#) vector giving the subscripts of the CA axes to be used (see below).
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	An [integer](#) 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](#) or an [IncidenceMatrix](#) (the same as object).

refine_seriation returns a [BootCA](#) object.

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.

Theses 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:

correspondence Correspondence analysis-based seriation. Correspondence 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 correspondence analysis space is arbitrary: additional information is needed to determine the actual order in time.

reciprocal Reciprocal ranking 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_seriation 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 correspondence 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.

Note

Refining method 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.

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](#).

Dunnell, R. C. (1970). Seriation Method and Its Evaluation. *American Antiquity*, 35(03), 305-319. DOI: [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*. Berlin Heidelberg: Springer, p. 307-316. DOI: [10.1007/3-540-28084-7_34](#).

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](#).

See Also

[ca](#)

Examples

```
## Matrix seriation
## Replicates Desachy 2004 results
## Coerce dataset to abundance matrix
compiegne_count <- as_count(compiegne)

## 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_reciprocal(compiegne_count, EPPM = TRUE,
                                         margin = 2))

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

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

## Refined seriation
## See the vignette:

utils::vignette("seriation", package = "tabula")
```

---

similarity   *Similarity*

Description

Similarity

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 to be passed to internal 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 symmetric matrix of class SimilarityMatrix.

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.

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.

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.

Robinson, W. S. (1951). A Method for Chronologically Ordering Archaeological Deposits. *American Antiquity*, 16(04), 293-301. DOI: 10.2307/276978.

See Also

Other diversity: heterogeneity-index, richness-index, turnover-index

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")
plot_spot(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
```

---

subset   *Extract or Replace Parts of an Object*

Description

Operators acting on objects to extract or replace parts.

Usage

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

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

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

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

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

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

Arguments

x	An object from which to extract element(s) or in which to replace element(s).
i, j	Indices specifying elements to extract or replace. Indices are [numeric](), [integer]() or [character]() vectors or empty (missing) or NULL. Numeric values are coerced to [integer]() as by [as.integer]() (and hence truncated towards zero). Character vectors will be matched to the name of the elements. An empty index (a comma separated blank) indicates that all entries in that dimension are selected.
drop	A [logical]() scalar: should the result be coerced to the lowest possible dimension? This only works for extracting elements, not for the replacement.

Value

A subsetted object.

Author(s)

N. Frerebeau

See Also

Other mutator: [access]()

test   *Tests on Abundance Data*

Description

Tests on Abundance Data

Usage

```
test_diversity(object, ...)

test_fit(object, ...)

## S4 method for signature 'CountMatrix'
test_fit(object, simplify = FALSE, ...)

## S4 method for signature 'CountMatrix'
test_diversity(object, adjust = "holm", ...)
```

Arguments

object	A $m \times p$ matrix of count data.
...	Further arguments to be passed to internal methods.
simplify	A logical scalar: should the result be simplified to a matrix?
adjust	A character string specifying the method for adjusting p values (see p.adjust).

Details

The following methods are available:

test_diversity Compare Shannon diversity between samples. This test produces two sided pairwise comparisons: it returns a matrix of adjusted p values.

test_fit The Frequency Increment Test (Feder et al. 2014). This test rejects neutrality if the distribution of normalized variant frequency increments exhibits a mean that deviates significantly from zero.

Value

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

Author(s)

N. Frerebeau

References

Feder, A. F., Kryazhimskiy, S. & Plotkin, J. B. (2014). Identifying Signatures of Selection in Genetic Time Series. *Genetics*, 196(2), 509-522. DOI: 10.1534/genetics.113.158220.

Magurran, A. E. (1988). *Ecological Diversity and its Measurement*. Princeton, NJ: Princeton University Press. DOI: 10.1007/978-94-015-7358-0.

Examples

```
## Shannon diversity test
merzbach_count <- as_count(merzbach)
div <- test_diversity(merzbach_count)

## Frequency Increment Test
## Coerce the merzbach dataset to a count matrix
## Keep only decoration types that have a maximum frequency of at least 50
keep <- apply(X = merzbach, MARGIN = 2, FUN = function(x) max(x) >= 50)
merzbach_count <- as_count(merzbach[, keep])

## The data are grouped by phase
## We use the row names as time coordinates (roman numerals)
set_dates(merzbach_count) <- rownames(merzbach)

fit <- test_fit(merzbach_count, simplify = TRUE)
```

---

turnover-index *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

`object`	A $m \times p$ matrix of count data.
`...`	Further arguments to be passed to internal methods.
`method`	A `character` string specifying the method to be used (see details). Any unambiguous substring can be given.
`simplify`	A `logical` scalar: should the result be simplified to a matrix?

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.

Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, 30(3), 279-338. DOI: 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.

See Also

Other diversity: `heterogeneity-index`, `richness-index`, `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.

See Also

Other datasets: boves, compiegne, merzbach, mississippi

Index