Package 'tabula'

June 23, 2022

Title Analysis and Visualization of Archaeological Count Data

Version 1.7.0

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

Description An easy way to examine archaeological count data. This
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 https://packages.tesselle.org/tabula/,
https://github.com/tesselle/tabula

BugReports https://github.com/tesselle/tabula/issues

Depends R (>= 3.3)

Imports arkhe (>= 0.5.0), boot, kairos (>= 1.1.0), ggplot2, methods,
rlang, stats, utils

Suggests covr, folio, khroma, knitr, magrittr, rmarkdown, testthat (>=
3.0.0), vdiffr (>= 1.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

Encoding UTF-8

RoxygenNote 7.2.0

Collate 'AllClasses.R' 'AllGenerics.R' 'coerce.R' 'deprecate.R'
'ggplot2.R' 'index_diversity.R' 'index_heterogeneity.R'
'index_rarefaction.R' 'index_richness.R' 'index_similarity.R'
'index_test.R' 'index_turnover.R' 'mutators.R' 'plot_bertin.R'
'plot_diversity.R' 'plot_ford.R' 'plot_heatmap.R' 'plot_rank.R'
'plot_spot.R' 'reexport.R' 'show.R' 'statistics.R' 'subset.R'
'tabula-package.R' 'utilities.R' 'validate.R' 'zzz.R'

Author Nicolas Frerebeau [aut, cre] (<<https://orcid.org/0000-0001-5759-4944>>,
     Université Bordeaux Montaigne),
     Brice Lebrun [ctb] (<<https://orcid.org/0000-0001-7503-8685>>, Université
     Bordeaux Montaigne),
     Matthew Peeples [ctb] (<<https://orcid.org/0000-0003-4496-623X>>, Arizona
     State University),
     Ben Marwick [ctb] (<<https://orcid.org/0000-0001-7879-4531>>, University
     of Washington),
     Anne Philippe [ctb] (<<https://orcid.org/0000-0002-5331-5087>>,
     Université de Nantes),
     Jean-Baptiste Fourvel [ctb] (<<https://orcid.org/0000-0002-1061-4642>>,
     CNRS)

R topics documented:

DiversityIndex   *Diversity Index*

Description

An S4 class to represent a diversity measure.

Slots

names A character vector giving the sample names.

size A integer vector giving the sample sizes.

simulation A four columns numeric matrix giving the diversity measures for the simulated assemblage (sample size, mean estimate, lower and upper boundaries of the confidence interval).

method A character string specifying the method used.

Subset

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

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

Coerce

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

as.data.frame(x) Coerces to a data.frame.

Note

This class inherits from base numeric.

Author(s)

N. Frerebeau

---

heterogeneity   *Heterogeneity and Evenness*

Description

- heterogeneity() returns an heterogeneity or dominance index.
- evenness() returns an evenness measure.

Usage

```
heterogeneity(object, ...)

evenness(object, ...)

index_berger(x, ...)

index_brillouin(x, ...)
```

```
index_mcintosh(x, ...)

index_shannon(x, ...)

index_simpson(x, ...)

## S4 method for signature 'matrix'
heterogeneity(
  object,
  method = c("berger", "brillouin", "mcintosh", "shannon", "simpson")
)

## S4 method for signature 'data.frame'
heterogeneity(
  object,
  method = c("berger", "brillouin", "mcintosh", "shannon", "simpson")
)

## S4 method for signature 'matrix'
evenness(object, method = c("shannon", "brillouin", "mcintosh", "simpson"))

## S4 method for signature 'data.frame'
evenness(object, method = c("shannon", "brillouin", "mcintosh", "simpson"))

## S4 method for signature 'numeric'
index_berger(x, na.rm = FALSE, ...)

## S4 method for signature 'numeric'
index_brillouin(x, evenness = FALSE, na.rm = FALSE, ...)

## S4 method for signature 'numeric'
index_mcintosh(x, evenness = FALSE, na.rm = FALSE, ...)

## S4 method for signature 'numeric'
index_shannon(x, evenness = FALSE, base = exp(1), na.rm = FALSE, ...)

## S4 method for signature 'numeric'
index_simpson(x, evenness = FALSE, na.rm = FALSE, ...)
```

Arguments

object	A $m \times p$ numeric [matrix] or [data.frame] of count data (absolute frequencies giving the number of individuals for each class).
...	Currently not used.
x	A [numeric] vector of count data (absolute frequencies).
method	A [character] string specifying the index to be computed (see details). Any unambiguous substring can be given.
na.rm	A [numeric] scalar: should missing values (including NaN) be removed?

| evenness | A logical scalar: should an evenness measure be computed instead of an heterogeneity/dominance index? |
| base | A positive numeric value specifying the base with respect to which logarithms are computed. |

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.

Value

- heterogeneity() returns an HeterogeneityIndex object.
- evenness() returns an EvennessIndex object.
- index_*() return a numeric vector.

Heterogeneity and Evenness Measures

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.

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. (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:10.1016/00225193(78)903028.

Magurran, A. E. (1988). *Ecological Diversity and its Measurement*. Princeton, NJ: Princeton University Press. doi:10.1007/9789401573580.

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.15387305.1948.tb01338.x.

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

See Also

Other diversity measures: occurrence(), plot_diversity, rarefaction(), richness(), similarity(), simulate(), turnover()

Examples

```
data("chevelon", package = "folio")

## Shannon diversity index
(h <- heterogeneity(chevelon, method = "shannon"))
(e <- evenness(chevelon, method = "shannon"))
```

```
## Bootstrap resampling (summary statistics)
bootstrap(h, f = NULL)

bootstrap(h, f = summary)

quant <- function(x) quantile(x, probs = c(0.25, 0.50))
bootstrap(h, f = quant)

## Jackknife resampling
jackknife(h)
```

---

independance   *Independance*

Description

Independance

Usage

```
eppm(object, ...)

pvi(object, ...)

## S4 method for signature 'matrix'
eppm(object)

## S4 method for signature 'data.frame'
eppm(object)

## S4 method for signature 'matrix'
pvi(object)

## S4 method for signature 'data.frame'
pvi(object)
```

Arguments

object   A [CountMatrix](#) object.

...   Currently not used.

Details

Computes for each cell of a numeric matrix one of the following statistic.

Value

A [numeric matrix](#).

EPPM

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

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 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.

See Also

plot_ford(), plot_heatmap(), seriate_rank()

Other statistics: test_diversity()

Examples

```
## Abundance data
## Coerce dataset to a count matrix (data from Desachy 2004)
data("compiegne", package = "folio")

## Compute EPPM
counts_eppm <- eppm(compiegne)
plot_heatmap(counts_eppm) +
  khroma::scale_fill_YlOrBr(name = "EPPM")

## Compute PVI
counts_pvi <- pvi(compiegne)
plot_heatmap(counts_pvi) +
  khroma::scale_fill_BuRd(name = "PVI", midpoint = 1)
```

---

mutator	*Get or Set Parts of an Object*

Description

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

Usage

```
get_method(x)

## S4 method for signature 'DiversityIndex'
get_method(x)
```

Arguments

x   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()

---

occurrence	*Co-Occurrence*

Description

Co-Occurrence

Usage

```
occurrence(object, ...)

## S4 method for signature 'matrix'
occurrence(object)

## S4 method for signature 'data.frame'
occurrence(object)
```

Arguments

`object`	A $m \times p$ matrix of count data.
`...`	Currently not used.

Details

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

Value

A stats::dist object.

Author(s)

N. Frerebeau

See Also

Other diversity measures: heterogeneity(), plot_diversity, rarefaction(), richness(), similarity(), simulate(), turnover()

Examples

```
## Plot spot diagram of a co-occurrence matrix
data("mississippi", package = "folio")

occ <- occurrence(mississippi)
plot_spot(occ)
```

---

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 'matrix'
plot_bertin(object, threshold = NULL, scale = NULL)

## S4 method for signature 'data.frame'
plot_bertin(object, threshold = NULL, scale = NULL)
```

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

## S4 method for signature 'data.frame'
plot_ford(object, EPPM = FALSE)
```

Arguments

object	A $m \times p$ numeric [matrix] or [data.frame] of count data (absolute frequencies giving the number of individuals for each class).
...	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 below)?

Details

If EPPM is TRUE and if a relative abundance is greater than the mean percentage of the type, the exceeding part is highlighted.

Value

A [ggplot2::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."

EPPM

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 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

eppm()

Other plot methods: plot_diversity, plot_heatmap(), plot_line, plot_spot()

Examples

```
## Abundance data
data("mississippi", package = "folio")

## Plot a Bertin diagram...
## ...without threshold
plot_bertin(mississippi)
## ...with variables scaled to 0-1 and the variable mean as threshold
scale_01 <- function(x) (x - min(x)) / (max(x) - min(x))
plot_bertin(mississippi, threshold = mean, scale = scale_01)

## Abundance data
## Coerce dataset to a count matrix (data from Desachy 2004)
data("compiegne", package = "folio")

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

---

plot_diversity   *Diversity Plot*

Description

Diversity Plot

Usage

```
## S4 method for signature 'DiversityIndex'
autoplot(object, ...)

## S4 method for signature 'DiversityIndex,missing'
plot(x, y, ...)

## S4 method for signature 'RarefactionIndex'
autoplot(object, ...)

## S4 method for signature 'RarefactionIndex,missing'
plot(x, y, ...)
```

Arguments

`object, x`	A DiversityIndex object to be plotted.
`...`	Currently not used.
`y`	Currently not used.

Value

- `autoplot()` returns a ggplot object.
- `plot()` is called it for its side-effects: it results in a graphic being displayed (invisibly returns x).

Author(s)

N. Frerebeau

See Also

Other diversity measures: heterogeneity(), occurrence(), rarefaction(), richness(), similarity(), simulate(), turnover()

Other plot methods: plot_bar, plot_heatmap(), plot_line, plot_spot()

Examples

```
## Coerce data to a count matrix
data("chevelon", package = "folio")

## Assemblage diversity size comparison
## Warning: this may take a few seconds!
h <- heterogeneity(chevelon, method = "shannon")
h_sim <- simulate(h)
plot(h_sim)

r <- richness(chevelon, method = "count")
r_sim <- simulate(r)
```

```
plot(r_sim)
```

---

plot_heatmap   *Heatmap*

Description

Plots a heatmap.

Usage

```
plot_heatmap(object, ...)

## S4 method for signature 'matrix'
plot_heatmap(object, diag = TRUE, upper = TRUE, lower = TRUE, freq = FALSE)

## S4 method for signature 'data.frame'
plot_heatmap(object, diag = TRUE, upper = TRUE, lower = TRUE, freq = FALSE)

## S4 method for signature 'dist'
plot_heatmap(object, diag = FALSE, upper = FALSE, lower = !upper)
```

Arguments

object	A $m \times p$ numeric [matrix](#) or [data.frame](#) of count data (absolute frequencies giving the number of individuals for each class).
...	Currently not used.
diag	A [logical](#) scalar indicating whether the diagonal of the matrix should be plotted. Only used if object is a symmetric matrix.
upper	A [logical](#) scalar indicating whether the upper triangle of the matrix should be plotted. Only used if object is a symmetric matrix.
lower	A [logical](#) scalar indicating whether the lower triangle of the matrix should be plotted. Only used if object is a symmetric matrix.
freq	A [logical](#) scalar indicating whether relative frequency should be used instead of counts (absolute frequency).

Value

A [ggplot2::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

pvi()

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

Examples

```
## Abundance data (data from Desachy 2004)
data("compiegne", package = "folio")

## Plot matrix diagram...
plot_heatmap(compiegne)
plot_heatmap(compiegne, freq = TRUE)

## Presence/absence data
inc <- sample(0:1, size = 100, replace = TRUE)
bin <- matrix(data = as.logical(inc), nrow = 10, ncol = 10)

plot_heatmap(bin) +
  ggplot2::scale_fill_manual(values = c("TRUE" = "black", "FALSE" = "white"))
```

---

plot_line   *Line Plot*

Description

Plots a rank *vs* relative abundance diagram.

Usage

```
plot_rank(object, ...)

## S4 method for signature 'matrix'
plot_rank(object, log = NULL, facet = FALSE)

## S4 method for signature 'data.frame'
plot_rank(object, log = NULL, facet = FALSE)
```

Arguments

object	A $m \times p$ numeric matrix or data.frame of count data (absolute frequencies giving the number of individuals for each class).
...	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?

Value

A ggplot2::ggplot object.

Author(s)

N. Frerebeau

References

Magurran, A. E. (1988). *Ecological Diversity and its Measurement.* Princeton, NJ: Princeton
University Press. [doi:10.1007/9789401573580](doi:10.1007/9789401573580).

See Also

Other plot methods: plot_bar, plot_diversity, plot_heatmap(), plot_spot()

Examples

```
## Abundance matrix
## Coerce datasets to a count matrix (data from Desachy 2004)
data("compiegne", package = "folio")

## Plot rank vs abundance
plot_rank(compiegne)
plot_rank(compiegne, facet = TRUE)
```

---

plot_spot   *Spot Plot*

Description

Plots a spot matrix.

Usage

```
plot_spot(object, ...)

## S4 method for signature 'matrix'
plot_spot(
  object,
  type = c("ring", "plain"),
  threshold = NULL,
```

```
    diag = TRUE,
    upper = TRUE,
    lower = TRUE,
    freq = FALSE,
    ...
)

## S4 method for signature 'data.frame'
plot_spot(
    object,
    type = c("ring", "plain"),
    threshold = NULL,
    diag = TRUE,
    upper = TRUE,
    lower = TRUE,
    freq = FALSE,
    ...
)

## S4 method for signature 'dist'
plot_spot(
    object,
    type = c("ring", "plain"),
    diag = FALSE,
    upper = FALSE,
    lower = !upper,
    ...
)
```

Arguments

object	A $m \times p$ numeric [matrix](#) or [data.frame](#) of count data (absolute frequencies giving the number of individuals for each class).
...	Extra parameters to be passed to threshold.
type	A [character](#) string specifying the graph to be plotted. It must be one of "ring" (the default) or "plain". Any unambiguous substring can be given.
threshold	A [function](#) that takes a numeric vector as argument and returns a numeric threshold value. If NULL (the default), no threshold is computed.
diag	A [logical](#) scalar indicating whether the diagonal of the matrix should be plotted. Only used if object is a symmetric matrix.
upper	A [logical](#) scalar indicating whether the upper triangle of the matrix should be plotted. Only used if object is a symmetric matrix.
lower	A [logical](#) scalar indicating whether the lower triangle of the matrix should be plotted. Only used if object is a symmetric matrix.
freq	A [logical](#) scalar indicating whether relative frequency should be used instead of counts (absolute frequency).

Details

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

Value

A [ggplot2::ggplot](#) object.

Note

Adapted from Dan Gopstein's original [idea](#).

Author(s)

N. Frerebeau

See Also

Other plot methods: `plot_bar`, `plot_diversity`, `plot_heatmap`(), `plot_line`

Examples

```
## Plot spot diagram of count data...
data("mississippi", package = "folio")

### ...without threshod
plot_spot(mississippi)
### ...with the  column means as threshold
plot_spot(mississippi, threshold = mean)
### ...with the column medians as threshold
plot_spot(mississippi, threshold = median)
```

---

rarefaction   *Rarefaction*

Description

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

Usage

```
rarefaction(object, ...)

index_hurlbert(x, ...)

## S4 method for signature 'matrix'
rarefaction(object, sample = NULL, method = c("hurlbert"), step = 1)
```

```
## S4 method for signature 'data.frame'
rarefaction(object, sample = NULL, method = c("hurlbert"), step = 1)

## S4 method for signature 'numeric'
index_hurlbert(x, sample, ...)
```

Arguments

object	A $m \times p$ numeric [matrix](#) or [data.frame](#) of count data (absolute frequencies).
...	Currently not used.
x	A [numeric](#) vector of count data (absolute frequencies).
sample	A length-one [numeric](#) vector giving the sub-sample size. The size of sample should be smaller than total community size.
method	A [character](#) string or vector of strings specifying the index to be computed (see details). Any unambiguous substring can be given.
step	An [integer](#) giving the increment of the sample size.

Value

- rarefaction() returns a [RarefactionIndex](#) object.
- index_*() return a [numeric](#) vector.

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.

Author(s)

N. Frerebeau

References

Hurlbert, S. H. (1971). The Nonconcept of Species Diversity: A Critique and Alternative Parameters. *Ecology*, 52(4), 577-586. doi:10.2307/1934145.

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

See Also

Other diversity measures: heterogeneity(), occurrence(), plot_diversity, richness(), similarity(), simulate(), turnover()

Examples

```
## Richness
## Margalef and Menhinick index
## Data from Magurran 1988, p. 128-129
trap <- matrix(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 = "margalef") # 2.55 1.88
richness(trap, method = "menhinick") # 1.95 1.66

## Asymptotic species richness
## Chao1-type estimators
## Data from Chao & Chiu 2016
brazil <- matrix(
  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
)

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

## Rarefaction
rarefaction(trap, sample = 13) # 6.56 6.00
```

---

resample   *Resampling Methods*

Description

- resample() simulate observations from a multinomial distribution.
- bootstrap() generate bootstrap estimations of a statistic.
- jackknife() generate jackknife estimations of a statistic.

Usage

```
resample(object, ...)

bootstrap(object, ...)
```

```
jackknife(object, ...)

## S4 method for signature 'DiversityIndex'
bootstrap(object, n = 1000, f = NULL)

## S4 method for signature 'DiversityIndex'
jackknife(object)

## S4 method for signature 'numeric'
resample(object, do, n, size = sum(object), ..., f = NULL)
```

Arguments

object	A [numeric](#) vector of count data (absolute frequencies).
...	Extra arguments passed to do.
n	A non-negative [integer](#) specifying the number of bootstrap replications.
f	A [function](#) that takes a single numeric vector (the result of do) as argument.
do	A [function](#) that takes `object` as an argument and returns a single numeric value.
size	A non-negative [integer](#) specifying the sample size.

Value

If f is NULL, resample() returns the n values of do. Else, returns the result of f applied to the n values of do.

If f is NULL, bootstrap() and jackknife() return a [data.frame](#) with the following elements (else, returns the result of f applied to the n values of do) :

original The observed value of do applied to `object`.

mean The bootstrap/jackknife estimate of mean of do.

bias The bootstrap/jackknife estimate of bias of do.

error The boostrap/jackknife estimate of standard error of do.

Author(s)

N. Frerebeau

See Also

[stats::rmultinom()](#)

Examples

```
## Sample observations from a multinomial distribution
x <- sample(1:100, 50, TRUE)
resample(x, do = median, n = 100)

## Estimate the 25th, 50th and 95th percentiles
```

```
quant <- function(x) { quantile(x, probs = c(0.25, 0.50, 0.75)) }
resample(x, n = 100, do = median, f = quant)
```

---

richness   *Richness*

Description

- `richness()` returns sample richness.
- `composition()` returns asymptotic species richness.

Usage

```
richness(object, ...)

composition(object, ...)

index_ace(x, ...)

index_ice(x, ...)

index_chao1(x, ...)

index_chao2(x, ...)

index_margalef(x, ...)

index_menhinick(x, ...)

## S4 method for signature 'matrix'
richness(object, method = c("count", "margalef", "menhinick"))

## S4 method for signature 'data.frame'
richness(object, method = c("count", "margalef", "menhinick"))

## S4 method for signature 'matrix'
composition(
  object,
  method = c("chao1", "ace", "chao2", "ice"),
  unbiased = FALSE,
  improved = FALSE,
  k = 10
)

## S4 method for signature 'data.frame'
composition(
  object,
```

```
  method = c("chao1", "ace", "chao2", "ice"),
  unbiased = FALSE,
  improved = FALSE,
  k = 10
)

## S4 method for signature 'numeric'
index_margalef(x, na.rm = FALSE, ...)

## S4 method for signature 'numeric'
index_menhinick(x, na.rm = FALSE, ...)

## S4 method for signature 'numeric'
index_ace(x, k = 10, ...)

## S4 method for signature 'numeric'
index_chao1(x, unbiased = FALSE, improved = FALSE, ...)

## S4 method for signature 'matrix'
index_ice(x, k = 10, ...)

## S4 method for signature 'matrix'
index_chao2(x, unbiased = FALSE, improved = FALSE, ...)
```

Arguments

object	A $m \times p$ numeric [matrix](#) or [data.frame](#) of count data (absolute frequencies).
...	Further arguments to be passed to internal methods.
x	A [numeric](#) vector or matrix of count data (absolute frequencies).
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.
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".
na.rm	A [numeric](#) scalar: should missing values (including NaN) be removed?

Value

- richness() returns a [RichnessIndex](#) object.
- composition() returns a [CompositionIndex](#) object.
- index_*() return a [numeric](#) vector.

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.

Richness Measures

The following richness measures are available for count data:

count Returns the number of observed taxa/types.

margalef Margalef richness index.

menhinick Menhinick richness index.

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.

Magurran, A. E. (1988). *Ecological Diversity and its Measurement*. Princeton, NJ: Princeton University Press. doi:10.1007/9789401573580.

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.

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.

See Also

plot_diversity()

Other diversity measures: heterogeneity(), occurrence(), plot_diversity, rarefaction(), similarity(), simulate(), turnover()

Examples

```
## Richness
## Margalef and Menhinick index
## Data from Magurran 1988, p. 128-129
trap <- matrix(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 = "margalef") # 2.55 1.88
richness(trap, method = "menhinick") # 1.95 1.66

## Asymptotic species richness
## Chao1-type estimators
## Data from Chao & Chiu 2016
brazil <- matrix(
  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
)

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

## Rarefaction
```

```
rarefaction(trap, sample = 13) # 6.56 6.00
```

---

similarity   *Similarity*

Description

Similarity

Usage

```
similarity(object, ...)

index_jaccard(x, y, ...)

index_sorenson(x, y, ...)

index_bray(x, y, ...)

index_morisita(x, y, ...)

index_brainerd(x, y, ...)

index_binomial(x, y, ...)

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

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

## S4 method for signature 'character,character'
index_jaccard(x, y)

## S4 method for signature 'logical,logical'
index_jaccard(x, y)

## S4 method for signature 'numeric,numeric'
index_jaccard(x, y)

## S4 method for signature 'logical,logical'
```

```
index_sorenson(x, y)

## S4 method for signature 'numeric,numeric'
index_sorenson(x, y)

## S4 method for signature 'numeric,numeric'
index_bray(x, y)

## S4 method for signature 'numeric,numeric'
index_morisita(x, y)

## S4 method for signature 'numeric,numeric'
index_brainerd(x, y)

## S4 method for signature 'numeric,numeric'
index_binomial(x, y)
```

Arguments

`object`	A $m \times p$ matrix of count data.
`...`	Currently not used.
`x, y`	A length-p [numeric](#) vector of count data.
`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 [stats::dist](#) object.
- `index_*()` return a [numeric](#) vector.

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:10.1007/9789401573580.

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 measures: heterogeneity(), occurrence(), plot_diversity, rarefaction(), richness(), simulate(), turnover()

Examples

```
## Data from Huntley 2008
ceramics <- matrix(
  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
)
rownames(ceramics) <- c("Atsinna", "Cienega", "Mirabal", "PdMuertos",
                        "Hesh", "LowPesc", "BoxS", "Ojo Bon", "S170")
colnames(ceramics) <- c("DLH-1", "DLH-2a", "DLH-2b", "DLH-2c", "DLH-4")

## Brainerd-Robinson measure
(C <- similarity(ceramics, "brainerd"))
plot_spot(C)

## Data from Magurran 1988, p. 166
birds <- matrix(
  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),
  nrow = 2, byrow = TRUE
)
rownames(birds) <- c("unmanaged", "managed")

## 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
```

---

simulate   *Measure Diversity by Comparing to Simulated Assemblages*

Description

Measure Diversity by Comparing to Simulated Assemblages

Usage

```
## S4 method for signature 'DiversityIndex'
simulate(
  object,
  n = 1000,
  step = 1,
  interval = c("percentiles", "student", "normal"),
  level = 0.8,
  progress = getOption("tabula.progress")
)
```

Arguments

object	A [DiversityIndex](#) object.
n	A non-negative [integer](#) giving the number of bootstrap replications.
step	An [integer](#) giving the increment of the sample size.
interval	A [character](#) string giving the type of confidence interval to be returned. It must be one "percentiles" (sample quantiles, as described in Kintigh 1984; the default), "student" or "normal". Any unambiguous substring can be given.
level	A length-one [numeric](#) vector giving the confidence level.
progress	A [logical](#) scalar: should a progress bar be displayed?

Value

Returns a DiversityIndex object.

Author(s)

N. Frerebeau

References

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

See Also

[plot_diversity()](plot_diversity), [resample()](resample)

Other diversity measures: [heterogeneity()](heterogeneity), [occurrence()](occurrence), plot_diversity, [rarefaction()](rarefaction), [richness()](richness), [similarity()](similarity), [turnover()](turnover)

Examples

```
## Coerce data to a count matrix
data("chevelon", package = "folio")

## Assemblage diversity size comparison
## Warning: this may take a few seconds!
h <- heterogeneity(chevelon, method = "shannon")
h_sim <- simulate(h)
plot(h_sim)

r <- richness(chevelon, method = "count")
r_sim <- simulate(r)
plot(r_sim)
```

---

subset   *Extract or Replace Parts of an Object*

Description

Operators acting on objects to extract or replace parts.

Arguments

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

i   A character string specifying elements to extract. Any unambiguous substring can be given (see details).

Value

A subsetted object.

Author(s)

N. Frerebeau

See Also

Other mutator: [mutator](#)

---

test_diversity   *Diversity Test*

Description

Compares Shannon diversity between samples.

Usage

```
test_diversity(object, ...)

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

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

Arguments

object	A $m \times p$ matrix of count data.
...	Further arguments to be passed to internal methods.
adjust	A [character](#) string specifying the method for adjusting p values (see [stats::p.adjust()](#)).

Details

This test produces two sided pairwise comparisons: it returns a matrix of adjusted p values.

Value

A [numeric](#) [matrix](#).

Author(s)

N. Frerebeau

References

Magurran, A. E. (1988). *Ecological Diversity and its Measurement*. Princeton, NJ: Princeton University Press. doi:10.1007/9789401573580.

See Also

Other statistics: independance

Examples

```
## Shannon diversity test
data("merzbach", package = "folio")
test_diversity(merzbach)
```

---

turnover	*Turnover*

Description

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

Usage

```
turnover(object, ...)

index_whittaker(x, ...)

index_cody(x, ...)

index_routledge1(x, ...)

index_routledge2(x, ...)

index_routledge3(x, ...)

index_wilson(x, ...)

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

## S4 method for signature 'data.frame'
turnover(
  object,
```

```
  method = c("whittaker", "cody", "routledge1", "routledge2", "routledge3", "wilson"),
  ...
)

## S4 method for signature 'matrix'
index_whittaker(x)

## S4 method for signature 'matrix'
index_cody(x)

## S4 method for signature 'matrix'
index_routledge1(x)

## S4 method for signature 'matrix'
index_routledge2(x)

## S4 method for signature 'matrix'
index_routledge3(x)

## S4 method for signature 'matrix'
index_wilson(x)
```

Arguments

object, x	A $m \times p$ matrix of count data or incidence 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

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

A numeric vector.

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 measures: heterogeneity(), occurrence(), plot_diversity, rarefaction(), richness(), similarity(), simulate()

Examples

```
## Data from Magurran 1988, p. 162
trees <- matrix(
  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
)
colnames(trees) <- 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, "routledge1") # 0.29
turnover(trees, "routledge2") # 0.56
turnover(trees, "routledge3") # 1.75

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

Index