Package 'diverse'

March 23, 2017

Title Diversity Measures for Complex Systems

Version 0.1.5

Author Miguel R. Guevara <miguel.guevara@upla.cl>, Dominik Hartmann
<d.hartmann@uni-hohenheim.de>, Marcelo Mendoza
<marcelo.mendoza@usm.cl>

Maintainer Miguel R. Guevara <miguel.guevara@upla.cl>

Description Computes the most common diversity measures used in social and other sciences, and includes new measures from interdisciplinary research.

Depends proxy, reshape2, foreign, stats

Suggests pheatmap, igraph, entropart

URL https://github.com/mguevara/diverse

Date 2017-03-23

BugReports https://github.com/mguevara/diverse/issues

License CC BY-SA 4.0

Note Supported by internal research project ING01-1516 from
'Universidad de Playa Ancha'

Repository CRAN

LazyData true

NeedsCompilation no

Encoding UTF-8

Date/Publication 2017-03-23 17:49:56 UTC

R topics documented:

1

---

balance	*Balance or proportions*

Description

Computes the proportions or probabilities of raw values.

Usage

```
balance(data, category_row = FALSE)
```

Arguments

data	A numeric matrix with entities i in the rows and categories j in the columns. Cells show the respective value (value of abundance) of entity i in the category j. It can also be a transpose of the previous matrix, that is, a matrix with categories in the rows and entities in the columns. Yet in that case, the argument "category_row" has to be set to TRUE. The matrix must include names for the rows and the columns. The argument "data", also accepts a dataframe with three columns in the following order: entity, category and value.
category_row	A flag to indicate that categories are in the rows. The analysis assumes that the categories are in the columns of the matrix. If the categories are in the rows and the entities in the columns, then the argument "category_row" has to be set to TRUE. The default value is FALSE.

Value

A matrix of entities-categories with proportions.

Examples

```
balance(data=geese, category_row = TRUE)
```

---

disparity   *A procedure to compute the sum and average of disparities*

Description

Computes the sum and the average of distances or disparities between the categories.

Usage

```
disparity(data, method = "euclidean", category_row = FALSE)
```

Arguments

data
: A numeric matrix with entities i in the rows and categories j in the columns. Cells show the respective value (value of abundance) of entity i in the category j. It can also be a transpose of the previous matrix, that is, a matrix with categories in the rows and entities in the columns. Yet in that case, the argument "category_row" has to be set to TRUE. The matrix must include names for the rows and the columns. The argument "data", also accepts a dataframe with three columns in the following order: entity, category and value.

method
: A distance or dissimilarity method available in "proxy" package as for example "Euclidean", "Kullback" or "Canberra". This argument also accepts a similarity method available in the "proxy" package, as for example: "cosine", "correlation" or "Jaccard" among others. In the latter case, a correspondent transformation to a dissimilarity measure will be retrieved. A list of available methods can be queried by using the function pr_DB. e.g. summary(pr_DB). The default value is Euclidean distance.

category_row
: A flag to indicate that categories are in the rows. The analysis assumes that the categories are in the columns of the matrix. If the categories are in the rows and the entities in the columns, then the argument "category_row" has to be set to TRUE. The default value is FALSE.

Value

A data frame with disparity measures for each entity in the dataset. Both the sum of disparities and the average of disparities are computed.

Examples

```
data(pantheon)
disparity(data= pantheon)
disparity(data = pantheon, method='Canberra')
#For scientific publications
#Same disparities, since all countries authored all entities
disparity(scidat)
disparity(data= scidat, method='cosine')
#Creating differences by measuring Revealed Compartive Advantages
```

```
disparity(values(scidat, norm='rca', filter=1))
#Activity Index for scientometrics
disparity(values(scidat, norm='ai', filter=0), method='cosine')
#Using binarization of values and a binary metric for dissimilarities.
disparity(values(scidat, norm='ai', filter=0, binary=TRUE), method='jaccard')
```

---

dis_categories	*A procedure to create a disparity matrix between categories.*

Description

Takes a data frame or a matrix to create a disparity matrix

Usage

```
dis_categories(data, method = "euclidean", category_row = FALSE)
```

Arguments

data
: A numeric matrix with entities i in the rows and categories j in the columns. Cells show the respective value (value of abundance) of entity i in the category j. It can also be a transpose of the previous matrix, that is, a matrix with categories in the rows and entities in the columns. Yet in that case, the argument "category_row" has to be set to TRUE. The matrix must include names for the rows and the columns. The argument "data", also accepts a dataframe with three columns in the following order: entity, category and value.

method
: A distance or dissimilarity method available in "proxy" package as for example "Euclidean", "Kullback" or "Canberra". This argument also accepts a similarity method available in the "proxy" package, as for example: "cosine", "correlation" or "Jaccard" among others. In the latter case, a correspondent transformation to a dissimilarity measure will be retrieved. A list of available methods can be queried by using the function [pr_DB](). e.g. summary(pr_DB). The default value is Euclidean distance.

category_row
: A flag to indicate that categories are in the rows. The analysis assumes that the categories are in the columns of the matrix. If the categories are in the rows and the entities in the columns, then the argument "category_row" has to be set to TRUE. The default value is FALSE.

Value

A distance or dissimilarity square matrix

Examples

```
Xdis <- dis_categories(pantheon)
```

---

dis_entities   *A procedure to create a disparity matrix between entities*

Description

Takes a data frame or a matrix to create a disparity matrix

Usage

```
dis_entities(data, method = "euclidean", category_row = FALSE)
```

Arguments

data

A numeric matrix with entities i in the rows and categories j in the columns. Cells show the respective value (value of abundance) of entity i in the category j. It can also be a transpose of the previous matrix, that is, a matrix with categories in the rows and entities in the columns. Yet in that case, the argument "category_row" has to be set to TRUE. The matrix must include names for the rows and the columns. The argument "data", also accepts a dataframe with three columns in the following order: entity, category and value.

method

A distance or dissimilarity method available in "proxy" package as for example "Euclidean", "Kullback" or "Canberra". This argument also accepts a similarity method available in the "proxy" package, as for example: "cosine", "correlation" or "Jaccard" among others. In the latter case, a correspondent transformation to a dissimilarity measure will be retrieved. A list of available methods can be queried by using the function [pr_DB](). e.g. summary(pr_DB). The default value is Euclidean distance.

category_row

A flag to indicate that categories are in the rows. The analysis assumes that the categories are in the columns of the matrix. If the categories are in the rows and the entities in the columns, then the argument "category_row" has to be set to TRUE. The default value is FALSE.

Value

A distance or dissimilarity square matrix

Examples

```
Xdis <- dis_entities(pantheon)
#for science dataset
dis_entities(scidat, method='cosine')
```

---

diverse   *diverse: A package to compute diversity measures.*

Description

The package **diverse** allows the user to compute the most common measures of diversity.

Details

The main function of the **diverse** package is diversity. Other functions are useful according to user needs, as variety to compute simple variety-richness or balance to compute measures related to balance. Also values and normalized values are accesible by using the function values.

---

diversity   **Main** *function to compute diversity measures*

Description

Main function of the package. The diversity function computes diversity measures for a dataset with entities, categories and values.

Usage

```
diversity(data, type = "all", category_row = FALSE, dis = NULL,
  method = "euclidean", q = 0, alpha = 1, beta = 1, base = exp(1))
```

Arguments

data
: A numeric matrix with entities i in the rows and categories j in the columns. Cells show the respective value (value of abundance) of entity i in the category j. It can also be a transpose of the previous matrix, that is, a matrix with categories in the rows and entities in the columns. Yet in that case, the argument "category_row" has to be set to TRUE. The matrix must include names for the rows and the columns. The argument "data", also accepts a dataframe with three columns in the following order: entity, category and value.

type
: A string or a vector of strings of nemonic strings referencing to the available diversity measures. The available measures are: "variety", (Shannon) "entropy", "blau","gini-simpson", "simpson", "hill-numbers", "herfindahl-hirschman", "berger-parker", "renyi", (Pielou) "evenness", "rao", "rao-stirling". A list of short mnemonics for each measure: "v", "e", "gs", "s", "td", "hh", "bp", "re", "ev", "r", and "rs". The default for type is "all" which computes all available formulas.

category_row
: A flag to indicate that categories are in the rows. The analysis assumes that the categories are in the columns of the matrix. If the categories are in the rows and the entities in the columns, then the argument "category_row" has to be set to TRUE. The default value is FALSE.

dis	Optional square matrix of distances or dissimilarities between categories. It allows the user to provide her own matrix of dissimilarities between categories. The category names have to be both in the rows and in the columns, and these must be the exact same names used by the categories in the argument "data". Only the upper triangle will be used. If the argument "dis" is not defined, and the user requires a measure that uses disparities (e.g. Rao), then a matrix of disparities is computed internally using the method defined by the argument 'method'. The default value is NULL.
method	The "rao-stirling" and "rao"-diversity indices use a disparity function to measure the distance between objects. If the user does not provide a matrix with disparities by using the argument 'dis', then a matrix of disparities is computed using the method specified in this argument (method). Possible values for this argument are distance or dissimilarity methods available in "proxy" package as for example "Euclidean", "Kullback" or "Canberra". This argument also accepts a similarity method available in the "proxy" package, as for example: "cosine", "correlation" or "Jaccard" among others. In the latter case, a correspondent transformation to a dissimilarity measure will be retrieved. A list of available methods can be queried by using the function `pr_DB`. e.g. summary(pr_DB). The default value is Euclidean distance.
q	The parameter used for the hill numbers. This argument is also used for the Renyi entropy and HCDT entropy. The default value is 0.
alpha	Parameter for Rao-Stirling diversity. The default value is 1.
beta	Parameter for Rao-Stirling diversity. The default value is 1.
base	Base of the logarithm. Used in Entropy calculations. The default value is exp(1).

Details

Notation used in the following formulas: N, category count; p_i, proportion of entity comprises category i; d_{ij}, disparity between i and j; q, α and β, arguments.

The available diversity measures included in the package are listed above. The titles of the formulas are the possible mnemonic values that the argument "type" might take to compute that formula (i.e. diversity(data, type='variety') or diversity(data, type='v'):

variety, v: Category counts per entity [MacArthur 1965]

$$\sum_i (p_i^0)$$

.

entropy, e: Shannon entropy per entity [Shannon 1948]

$$-\sum_i (p_i \log p_i)$$

Herfindahl-Hirschman, hh, hhi: The Herfindahl-Hirschman Index used in economy to measure the concentration of markets.

$$\sum_i (p_i^2)$$

gini-simpson, gs: Gini-Simpson index per object [Gini 1912]. This measure is also known as the Gibbs-Martin index or the Blau index in sociology, psychology and management studies.

$$1 - \sum_i (p_i^2)$$

simpson, s: Simpson index per entity [Simpson 1949].

$$D = \sum_i n_i(n_i - 1)/N(N - 1)$$

When this measure is required, then also associated variations Simpson's Index of Diversity $1 - D$ and the Reciprocal Simpson $1/D$ will be computed.

hill-numbers, td,hn: Hill Numbers [Hill 1973]. This measure is q parameterized. When $q = 1$, it results in the exponential of Shannon Entropy. Default for q is 0, this is the variety or richness.

$$\left(\sum_i p_i^q\right)^{1/(1-q)}$$

berger-parker, bp: Berger-Parker index is equals to the maximum p_i value in the entity, i.e. the proportional abundance of the most abundant type. When this measure is required, the reciprocal measure is also computed.

renyi, re: Renyi entropy per object. This measure is a generalization of the Shannon entropy parameterized by q. It corresponds to the logarithm of the hill numbers. The default value for q is 0.

$$(1 - q)^{-1} \log\left(\sum_i p_i^q\right)$$

evenness, ev: Pielou evenness per object across categories [Pielou, 1969]. It is based in Shannon Entropy

$$-\sum_i (p_i \log p_i)/\log v$$

rao: Rao diversity.

$$\sum_{ij} d_{ij} p_i p_j$$

rao-stirling, rs: Rao-Stirling diversity per object across categories [Stirling, 2007]. Default values are $\alpha = 1$ and $\beta = 1$. For the pairwise disparities the measure allows to consider the Jaccard Index, Euclidean distances, Cosine Similarity among others.

$$\sum_{ij} d_{ij}{}^\alpha (p_i p_j)^\beta$$

Value

A data frame with diversity measures as columns for each entity.

References

Gini, C. (1912). "Italian: Variabilita e mutabilita" 'Variability and Mutability', Memorie di metodologica statistica.

Hill, M. (1973). "Diversity and evenness: a unifying notation and its consequences". Ecology 54: 427-432.

MacArthur, R. (1965). "Patterns of Species Diversity". Biology Reviews 40: 510-533.

Pielou, E. (1969). "An Introduction to Mathematical Ecology". Wiley.

Shannon, C. (1948). "A Mathematical Theory of Communication". Bell entity Technical Journal 27 (3): 379-423.

Simpson, A. (1949). "Measurement of Diversity". Nature 163: 41-48.

Stirling, A. (2007). "A General Framework for Analysing Diversity in Science, Technology and Society". Journal of the Royal Society Interface 4: 707-719.

Rafols, I., & Meyer, M. (2009). Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. Scientometrics, 82(2), 263-287.

Rafols, I. (2014). Knowledge Integration and Diffusion: Measures and Mapping of Diversity and Coherence. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), Measuring Scholarly Impact (pp. 169-190). Springer International Publishing.

Chavarro, D., Tang, P., & Rafols, I. (2014). Interdisciplinarity and research on local issues: evidence from a developing country. Research Evaluation, 23(3), 195-209.

Examples

```
data(pantheon)
diversity(pantheon)
diversity(pantheon, type='variety')
diversity(geese, type='berger-parker', category_row=TRUE)
#reading csv data matrix
path_to_file <- system.file("extdata", "PantheonMatrix.csv", package = "diverse")
X <- read_data(path = path_to_file)
diversity(data=X, type="gini")
diversity(data=X, type="rao-stirling", method="cosine")
diversity(data=X, type="all", method="jaccard")

#reading csv dataframe
path_to_file <- system.file("extdata", "PantheonEdges.csv", package = "diverse")
X <- read_data(path = path_to_file)
#hill numbers
diversity(data=X, type="td", q=1)
#rao stirling with differente arguments
diversity(data=X, type="rao-stirling", method="euclidean", alpha=0, beta=1)
#more than one diversity measure
diversity(data=X, type=c('e','ev','bp','s'))
```

---

geese   *Geese dataset*

Description

A matrix of species of geese. The dataset includes the quantity of 4 species of geese observed by year in the Netherlands. The data comes from the Dutch bird protection organisation Sovon.

Usage

```
geese
```

Format

A matrix with the variables:

Columns Year of observation

Rows Species

Source

<https://www.sovon.nl/en>

<http://www.compass-project.eu/applets/3/index_EN.html>

Examples

```
str(geese)
summary(geese)
geese[,"2000"]
geese["Mute Swan",]
```

---

pantheon   *Pantheon dataset*

Description

Dataframe of globally famous people according to MIT's Pantheon 1.0. Dataset includes the number of globally famous people for a **sample** of 10 countries and 53 different occupations. The **complete** dataset is described in [Yu et al., 2015].

Usage

```
pantheon
```

Format

A dataframe with the variables:

Country Name of the country

Occupation Occupation according to the taxonomy of Pantheon

Value Quantity of globally famous people that were born in that country

Source

<http://pantheon.media.mit.edu/>

References

Yu, A. Z., Ronen, S., Hu, K., Lu, T., & Hidalgo, C. A. (2016). Pantheon 1.0, a manually verified dataset of globally famous biographies. Scientific Data, 3.

Examples

```
data(pantheon)
str(pantheon)
summary(pantheon)
pantheon[pantheon$Country=="Chile",]
```

---

read_data   *A procedure to read data of a data file in formats csv, dta or spss*

Description

This function reads a file with data shaped as a matrix or as edges list. Several types of formats are allowed.

Usage

```
read_data(path, type = "csv", sep = ",", category_row = FALSE)
```

Arguments

path   A string representing the path to data file. If the data contained in the file is shaped as a matrix, the first column must include the names of the categories. If the data is shaped as edges list, it must contain three columns: Entity, category and value.

type   It indicates the type of data to be read. This argument facilitates the input of diverse types of data files, such as spss or stata. Possible options are the names of the mentioned software. The default value is csv.

sep   Separator character used in the file to separate columns. Only for CSV file. Default value is comma.

category_row   A flag to indicate that categories are in the rows. The analysis assumes that the categories are in the columns of the matrix. If the categories are in the rows and the entities in the columns, then the argument "category_row" has to be set to TRUE. The default value is FALSE.

Value

A data frame with three columns (entity, category, value).

Examples

```
#reading an edges list or panel shape, source data must include three columns
path <-   system.file("extdata", "PantheonEdges.csv", package = "diverse")
sep <- ","
data <- read_data(path)
#reading a table
path  <- system.file("extdata", "PantheonMatrix.csv", package = "diverse")
sep <- ","
data <- read_data(path)
#reading a table which includes the entities in the columns
path <- system.file("extdata", "Geese.csv", package = "diverse")
data <- read_data(path, category_row=TRUE)
```

---

scidat   *Scidat dataset*

Description

A matrix of the number of papers authored by 10 countries in 27 areas of science in the year 2013. Data was retrieved and aggregated from SCimago.

Usage

```
scidat
```

Format

A matrix with the variables:

Columns Areas of Science according SCimago

Rows Name of the country

Source

Raw data before the aggregation was queried from <http://www.scimagojr.com/> in 2014.

References

SCImago. (2007). SJR-SCImago Journal & Country Rank.

Examples

```
str(scidat)
summary(scidat)
scidat["United States",]
scidat[,"Chemistry"]
```

---

sim_dataset   *A procedure to simulate datasets*

Description

Simulates a dataset with values of variety for each entity and possible values of abundance.

Usage

```
sim_dataset(n_categ, category_prefix = "", entity_prefix = "",
  values = "log-normal", size = -1, mean = 0, sd = 1,
  category_random = FALSE)
```

Arguments

n_categ	a vector with number of categories for each entity. The number of entities to create is defined by the length of this vector.
category_prefix	
	a prefix to be used as part of the category label
entity_prefix	a prefix to be used as part of the entity label
values	values of abundance. This argument can be both, a distribution name or a vector of integers. The distribution is used to simulate individuals that are aggregated in frequencies or values of abundance. Use 'log-normal' for log normal distribution or 'normal' for normal distribution. In the second case, an integer or a vector of integers of possible values of abundance to be used randomly. Default value is 'log-normal'
size	number of individuals. A number or a vector of numbers for each entity. Default value is 7 times variety.
mean	parameter for normal or log-normal distribution. Default value is 0.
sd	parameter for normal or log-normal distribution. Default value is 1.
category_random	
	boolean argument to determine if categories should be taken randomly (TRUE) or sequentially (FALSE). Default is FALSE

Value

A data frame with three columns: entity, category and value of abundance.

Examples

```
sim_dataset(n_categ=50,  category_prefix='ctg', values=1) #equal value, just one entity
#Several entities with random values
n_entities <- 50
v_n_c <- sample(1:100, size = n_entities, replace=TRUE)
v_v <- sample(10:5000, size= n_entities, replace=TRUE)
d <- sim_dataset(n_categ = v_n_c, values= v_v, category_random = TRUE)
```

---

sim_entity   *A procedure to simulate entities*

Description

Simulates an entity with values of abundance for some categories.

Usage

```
sim_entity(n_categ, category_prefix = "", values = "log-normal",
  size = -1, mean = 0, sd = 1)
```

Arguments

n_categ	number of categories
category_prefix	
	a prefix to be used as part of the category label
values	values of abundance. This argument can be both, a distribution name or a vector of integers. The distribution is used to simulate individuals that are aggregated in frequencies or values of abundance. In the second case, an integer or a vector of integers of possible values of abundance to be used randomly. Default value is 'log-normal'
size	number of individuals. Default value is 7 times n_categ.
mean	parameter for normal or log-normal distribution. Default value is 0.
sd	parameter for normal or log-normal distribution. Default value is 1.

Value

A data frame with two columns: category and value of abundance.

Examples

```
sim_entity(n_categ=50,  category_prefix='ctg', values=1) #equal value
#random numbers for values of abundance
sim_entity(n_categ=50,  category_prefix='ctg', values=sample(1:100, replace=TRUE))
sim_entity(n_categ=50,  category_prefix='ctg', values='log-normal') #equal value
```

---

sim_individuals *A procedure to simulate labeled individuals for one category*

Description

Simulates a number of individuals tagged in N different categories, given a distribution such as log normal or normal.

Usage

```
sim_individuals(n_categ, size, category_prefix = "", type = "log-normal",
  mean = 0, sd = 1)
```

Arguments

n_categ	number of categories
size	number of individuals.
category_prefix	
	a prefix to be used as part of the category label
type	distribution name. The distribution is used to simulate how individuals are created. Use 'log-normal' for log normal distribution or 'normal' for normal distribution. Default value is 'log-normal'
mean	parameter for normal or log-normal distribution. Default value is 0.
sd	parameter for normal or log-normal distribution. Default value is 1.

Value

A vector of category labels.

Examples

```
sim_individuals(n_categ=50, size=10000, type='log-normal', mean=0.507, sd=1.183)
```

---

to_entropart *Transforms data to be used in Entropart package*

Description

Transform a dataframe used in diverse to values of abundance to be used in Entropart.

Usage

```
to_entropart(data)
```

Arguments

data	a dataframe used in diverse with three columns, entity, category and value of abundance.

Value

An object of type matrix of abundance to be used in entropart to create metacommunities.

Examples

```
ab <- to_entropart(sim_dataset(c(1,2)))
```

---

ubiquity	*Ubiquity of categories across entities*

Description

Computes the ubiquity or the rareness of the categories

Usage

```
ubiquity(data, category_row = FALSE)
```

Arguments

data	A numeric matrix with entities i in the rows and categories j in the columns. Cells show the respective value (value of abundance) of entity i in the category j. It can also be a transpose of the previous matrix, that is, a matrix with categories in the rows and entities in the columns. Yet in that case, the argument "category_row" has to be set to TRUE. The matrix must include names for the rows and the columns. The argument "data", also accepts a dataframe with three columns in the following order: entity, category and value.
category_row	A flag to indicate that categories are in the rows. The analysis assumes that the categories are in the columns of the matrix. If the categories are in the rows and the entities in the columns, then the argument "category_row" has to be set to TRUE. The default value is FALSE.

Value

A dataframe with values of number of entities where the category is present. Ordered in decreasing order.

Examples

```
ub <- ubiquity(data=pantheon)
```

---

values   *Pre-process the raw data*

Description

Allows to filter, binarize and/or normalize raw data. Also filter and binarization is available.

Usage

```
values(data, category_row = FALSE, norm = NULL, filter = NULL,
  binary = FALSE)
```

Arguments

data

A numeric matrix with entities i in the rows and categories j in the columns. Cells show the respective value (value of abundance) of entity i in the category j. It can also be a transpose of the previous matrix, that is, a matrix with categories in the rows and entities in the columns. Yet in that case, the argument "category_row" has to be set to TRUE. The matrix must include names for the rows and the columns. The argument "data", also accepts a dataframe with three columns in the following order: entity, category and value.

category_row

A flag to indicate that categories are in the rows. The analysis assumes that the categories are in the columns of the matrix. If the categories are in the rows and the entities in the columns, then the argument "category_row" has to be set to TRUE. The default value is FALSE.

norm

Methods to compute normalized values. Possible values are 'p', 'proportions', 'rca', 'rca_norm' and 'ai'. RCA refers to Revealed Comparative Advantages [Balassa 1986], rca_norm normalizes the RCAs between -1 and with 1, ai refers to the Activity Index.

filter

A threshold below which values are replaced with NA.

binary

A boolean value to indicate if values distinct from NA are replaced with 1.

Details

If the three arguments 'norm', 'filter' and 'binary' are used, then the same sequential order is applied in the calculations.

Value

A matrix with the raw, normalized, filtered andVor binarized data.

References

Balassa, B. (1986). Comparative advantage in manufactured goods: a reappraisal. The Review of Economics and Statistics, 315-319.

Examples

```
#raw values
values(data=pantheon)
values(data = scidat)
#proportions
values(data = scidat, norm='p')
#revealed comparative advantages
values(data = scidat, norm='rca')
values(data = scidat, norm='rca', filter=1)
values(data = scidat, norm='rca', filter=1, binary=TRUE)
```

---

variety   *Variety or Richness*

Description

Computes the variety (number of distinct types) or simple diversity of an entity. It is also known as richness.

Usage

```
variety(data, sort = TRUE, decreasing = TRUE, category_row = FALSE)
```

Arguments

data	A numeric matrix with entities i in the rows and categories j in the columns. Cells show the respective value (value of abundance) of entity i in the category j. It can also be a transpose of the previous matrix, that is, a matrix with categories in the rows and entities in the columns. Yet in that case, the argument "category_row" has to be set to TRUE. The matrix must include names for the rows and the columns. The argument "data", also accepts a dataframe with three columns in the following order: entity, category and value.
sort	Indicates whether results should be ordered or not. Define it to FALSE to avoid ordering.
decreasing	If argument "sort" is set to TRUE, this argument indicates descending order. The default value is TRUE.
category_row	A flag to indicate that categories are in the rows. The analysis assumes that the categories are in the columns of the matrix. If the categories are in the rows and the entities in the columns, then the argument "category_row" has to be set to TRUE. The default value is FALSE.

Value

A dataframe with values of variety for each entity.

Examples

```
variety(data=pantheon)
variety(data=pantheon, sort=FALSE)
```

Index