

Package ‘socialmixr’

August 6, 2020

Version 0.1.7

Date 2020-08-06

Title Social Mixing Matrices for Infectious Disease Modelling

Depends R (>= 3.5.0)

Imports data.table, curl, httr, jsonlite, oai, wpp2015, countrycode,
stringr, XML, lubridate

Description Provides methods for sampling contact matrices from diary data for use in infectious disease modelling, as discussed in Mossong et al. (2008) <doi:10.1371/journal.pmed.0050074>.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

Suggests testthat, knitr, rmarkdown, ggplot2, reshape2, formatR

VignetteBuilder knitr

NeedsCompilation no

Author Sebastian Funk [aut, cre],
Maria Bekker-Nielsen Dunbar [ctb],
Carl A. B. Pearson [ctb],
Sam Clifford [ctb],
Christopher Jarvis [ctb],
Alexis Robert [ctb]

Maintainer Sebastian Funk <sebastian.funk@lshtm.ac.uk>

Repository CRAN

Date/Publication 2020-08-06 11:30:03 UTC

R topics documented:

check	2
cite	3
clean	4

contact_matrix	4
get_survey	6
limits_to_agegroups	7
list_surveys	8
polymod	8
pop_age	9
reduce_agegroups	10
survey	10
survey_countries	11
wpp_age	12
wpp_countries	12

Index 13

check	<i>Check contact survey data</i>
-------	----------------------------------

Description

Checks that a survey fulfills all the requirements to work with the 'contact_matrix' function

Usage

```
## S3 method for class 'survey'
check(
  x,
  columns = FALSE,
  quiet = FALSE,
  error = FALSE,
  id.column = "part_id",
  participant.age.column = "part_age",
  country.column = "country",
  year.column = "year",
  contact.age.column = "cnt_age",
  ...
)
```

Arguments

x	A survey object
columns	if given, a named character vector containing the name of the "id", "participant.age" and "contact.age" columns
quiet	if TRUE, will not exit quietly if the test is passed.
error	if TRUE, will stop if an error is found in the structure of the participants and contacts data frame
id.column	the column in both the participants and contacts data frames that links contacts to participants

<code>participant.age.column</code>	the column in the participants data frame containing participants' age
<code>country.column</code>	the column in the participants data frame containing the country in which the participant was queried
<code>year.column</code>	the column in the participants data frame containing the year in which the participant was queried
<code>contact.age.column</code>	the column in the contacts data frame containing the age of contacts; if this does not exist, at least columns "..._exact", "..._est_min" and "..._est_max" must (see the <code>estimated.contact.age</code> option in contact_matrix)
<code>...</code>	ignored

Value

invisibly returns a character vector of the relevant columns

Examples

```
data(polymod)
check(polymod)
```

cite

Citation for a survey

Description

Gets a full citation for a [survey](#). If `quiet` is `FALSE` (default)

Usage

```
## S3 method for class 'survey'
cite(x, quiet = FALSE, ...)
```

Arguments

<code>x</code>	a character vector of surveys to cite
<code>quiet</code>	if set to <code>TRUE</code> , do not print entry, just return bibentry object
<code>...</code>	ignored

Value

citation as bibentry

Examples

```
data(polymod)
cite(polymod)
```

clean	<i>Clean contact survey data</i>
-------	----------------------------------

Description

Cleans survey data to work with the 'contact_matrix' function

Usage

```
## S3 method for class 'survey'
clean(x, country.column = "country", participant.age.column = "part_age", ...)
```

Arguments

x	A survey object
country.column	the name of the country in which the survey participant was interviewed
participant.age.column	the column in x\$participants containing participants' age
...	ignored

Value

a cleaned survey in the correct format

Examples

```
data(polymod)
cleaned <- clean(polymod) # not really necessary as the 'polymod' data set has already been cleaned
```

contact_matrix	<i>Generate a contact matrix from diary survey data</i>
----------------	---

Description

Samples a contact survey using a bootstrap

Usage

```
contact_matrix(
  survey,
  countries = c(),
  survey.pop,
  age.limits,
  filter,
  n = 1,
```

```

bootstrap,
counts = FALSE,
symmetric = FALSE,
split = FALSE,
estimated.participant.age = c("mean", "sample", "missing"),
estimated.contact.age = c("mean", "sample", "missing"),
missing.participant.age = c("remove", "keep"),
missing.contact.age = c("remove", "sample", "keep"),
weights = c(),
weigh.dayofweek = FALSE,
sample.all.age.groups = FALSE,
quiet = FALSE,
...
)

```

Arguments

survey	a survey object
countries	limit to one or more countries; if not given, will use all countries in the survey; these can be given as country names or 2-letter (ISO Alpha-2) country codes
survey.pop	survey population – either a data frame with columns 'lower.age.limit' and 'population', or a character vector giving the name(s) of a country or countries from the list that can be obtained via <code>wpp_countries</code> ; if not given, will use the country populations from the chosen countries, or all countries in the survey if countries is not given
age.limits	lower limits of the age groups over which to construct the matrix
filter	any filters to apply to the data, given as list of the form (column=filter_value) - only contacts that have 'filter_value' in 'column' will be considered
n	number of matrices to sample
bootstrap	whether to sample participants and contacts randomly using a bootstrap; by default, will use bootstrap if $n > 1$
counts	whether to return counts (instead of means)
symmetric	whether to make matrix symmetric, such that $c_{ijN_i} = c_{jiN_j}$.
split	whether to split the number of contacts and assortativity
estimated.participant.age	if set to "mean" (default), people whose ages are given as a range (in columns named "..._est_min" and "..._est_max") but not exactly (in a column named "..._exact") will have their age set to the mid-point of the range; if set to "sample", the age will be sampled from the range; if set to "missing", age ranges will be treated as missing
estimated.contact.age	if set to "mean" (default), people whose ages are given as a range (in columns named "..._est_min" and "..._est_max") but not exactly (in a column named "..._exact") will have their age set to the mid-point of the range; if set to "sample", the age will be sampled from the range; if set to "missing", age ranges will be treated as missing

missing.participant.age	if set to "remove" (default), participants without age information are removed; if set to "keep", participants with missing age are kept and treated as a separate age group
missing.contact.age	if set to "remove" (default), participants that that have contacts without age information are removed; if set to "sample", contacts without age information are sampled from all the contacts of participants of the same age group; if set to "keep", contacts with missing age are kept and treated as a separate age group
weights	columns that contain weights
weigh.dayofweek	whether to weigh the day of the week (weight 5 for weekdays ans 2 for week-ends)
sample.all.age.groups	what to do if bootstrapping fails to sample participants from one or more age groups; if FALSE (default), corresponding rows will be set to NA, if TRUE the sample will be discarded and a new one taken instead
quiet	if set to TRUE, output is reduced
...	further arguments to pass to get_survey , check and pop_age (especially column names)

Value

a list of sampled contact matrices, and the underlying demography of the surveyed population

Author(s)

Sebastian Funk

Examples

```
data(polymod)
contact_matrix(polymod, countries = "United Kingdom", age.limits = c(0, 1, 5, 15))
```

get_survey	<i>Get a survey, either from its Zenodo repository, a set of files, or a survey variable</i>
------------	--

Description

Downloads survey data, or extracts them from files, and returns a clean data set.

Usage

```
get_survey(survey, quiet = FALSE, ...)
```

Arguments

survey a DOI (see [list_surveys](#)), or a character vector of file names, or a [survey](#) object (in which case only cleaning is done).

quiet if TRUE, suppress messages

... options for [clean](#), which is called at the end of this

Value

a survey in the correct format

Examples

```
## Not run:  
list_surveys()  
peru_survey <- get_survey("https://doi.org/10.5281/zenodo.1095664")  
  
## End(Not run)
```

limits_to_agegroups *Convert lower age limits to age groups.*

Description

Mostly used for plot labelling

Usage

```
limits_to_agegroups(x, limits)
```

Arguments

x age limits to transform

limits lower age limits; if not given, will use all limits in x

Value

Age groups (limits separated by dashes)

Examples

```
limits_to_agegroups(c(0, 5, 10))
```

list_surveys	<i>List all surveys available for download</i>
--------------	--

Description

List all surveys available for download

Usage

```
list_surveys()
```

Value

character vector of surveys

Examples

```
## Not run:
  list_surveys()

## End(Not run)
```

polymod	<i>Social contact data from 8 European countries</i>
---------	--

Description

A dataset containing social mixing diary data from 8 European countries: Belgium, Germany, Finland, Great Britain, Italy, Luxembourg, The Netherlands and Poland. The Data are fully described in Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. (2008) Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. PLoS Med 5(3): e74.

Usage

```
polymod
```

Format

A list of two data frames:

participants the study participant, with age, country, year and day of the week (starting with 1 = Monday)

contacts reported contacts of the study participants. The variable phys_contact has two levels (1 denotes physical contact while 2 denotes non-physical contact), duration_multi has five levels (1 is less than 5 minutes while 5 is more than 4 hours, increasing in the order found in Figure 1 in Mossong et al.), and frequency_multi has five levels (1 is daily, 2 is weekly, 3 is monthly, 4 is less often, and 5 is first time) All other variables are described on the Zenodo repository of the data, available at <https://doi.org/10.5281/zenodo.1043437>

Source

<http://dx.doi.org/10.1371/journal.pmed.0050074>

pop_age	<i>Change age groups in population data</i>
---------	---

Description

This changes population data to have age groups with the given age.limits, extrapolating linearly between age groups (if more are requested than available) and summing populations (if fewer are requested than available)

Usage

```
pop_age(
  pop,
  age.limits,
  pop.age.column = "lower.age.limit",
  pop.column = "population",
  ...
)
```

Arguments

pop	a data frame with columns indicating lower age limits and population sizes (see 'age.column' and 'pop.column')
age.limits	lower age limits of age groups to extract
pop.age.column	column in the 'pop' data frame indicating the lower age group limit
pop.column	column in the 'pop' data frame indicating the population size
...	ignored

Value

data frame of age-specific population data

reduce_agegroups	<i>Reduce the number of age groups given a broader set of limits</i>
------------------	--

Description

Operates on lower limits

Usage

```
reduce_agegroups(x, limits)
```

Arguments

x	vector of limits
limits	new limits

Value

vector with the new age groups

Examples

```
reduce_agegroups(seq_len(20), c(0, 5, 10))
```

survey	<i>Contact survey</i>
--------	-----------------------

Description

A survey object contains the results of a contact survey. In particular, it contains two data frames called participants and contacts that are linked by a column specified as `id.column`

Usage

```
survey(participants, contacts, reference = NULL)
```

Arguments

participants	a data.frame containing information on participants
contacts	a data.frame containing information on contacts
reference	a list containing information needed to reference the survey, in particular it can contain a "title", "bibtype", "author", "doi", "publisher", "note", "year"

Value

a new survey object

Author(s)

Sebastian Funk

Examples

```
data(polymod)
new_survey <- survey(polymod$participants, polymod$contacts)
```

survey_countries	<i>List all countries contained in a survey</i>
------------------	---

Description

List all countries contained in a survey

Usage

```
survey_countries(survey, country.column = "country", ...)
```

Arguments

survey	a DOI (see list_surveys), or a character vector of file names, or a survey object (in which case only cleaning is done).
country.column	column in the survey indicating the country
...	further arguments for get_survey

Value

list of countries

Examples

```
data(polymod)
survey_countries(polymod)
```

wpp_age	<i>Get age-specific population data according to the World Population Prospects 2015 edition</i>
---------	--

Description

This uses data from the wpp2015 package but combines male and female, and converts age groups to lower age limits

Usage

```
wpp_age(countries, years)
```

Arguments

countries	countries, will return all if not given
years	years, will return all if not given

Value

data frame of age-specific population data

Examples

```
wpp_age("Italy", c(1990, 2000))
```

wpp_countries	<i>List all countries and regions for which socialmixr has population data</i>
---------------	--

Description

Uses the World Population Prospects data from the wpp2015 package

Usage

```
wpp_countries()
```

Value

list of countries

Examples

```
## Not run: wpp_countries()
```

Index

* datasets

polymod, 8

check, 2, 6

cite, 3

clean, 4, 7

contact_matrix, 3, 4

get_survey, 6, 6, 11

limits_to_agegroups, 7

list_surveys, 7, 8, 11

polymod, 8

pop_age, 6, 9

reduce_agegroups, 10

survey, 2–5, 7, 10, 11

survey_countries, 11

wpp_age, 12

wpp_countries, 12