

# Package ‘socialmixr’

September 24, 2019

**Version** 0.1.5

**Date** 2019-09-24

**Title** Social Mixing Matrices for Infectious Disease Modelling

**Depends** R (>= 3.5.0)

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

**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** 6.1.1

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

**VignetteBuilder** knitr

**NeedsCompilation** no

**Author** Sebastian Funk [aut, cre]

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

**Repository** CRAN

**Date/Publication** 2019-09-24 12:40:02 UTC

## R topics documented:

check . . . . .	2
cite . . . . .	3
clean . . . . .	4
contact_matrix . . . . .	4
get_survey . . . . .	6
limits_to_agegroups . . . . .	7
list_surveys . . . . .	7
polymod . . . . .	8

pop_age . . . . .	8
reduce_agegroups . . . . .	9
survey . . . . .	9
survey_countries . . . . .	10
wpp_age . . . . .	11
wpp_countries . . . . .	11

<b>Index</b>	<b>12</b>
--------------	-----------

---

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 <a href="#">survey</a> 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
participant.age.column	the column in the participants data frame containing participants' age
country.column	the column in the participants data frame containing the country in which the participant was queried
year.column	the column in the participants data frame containing the year in which the participant was queried
contact.age.column	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 estimated.contact.age option in <a href="#">contact_matrix</a> )
...	ignored

**Value**

invisibly returns a character vector of the relevant columns

**Examples**

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

---

cite	<i>Citation for a survey</i>
------	------------------------------

---

**Description**

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

**Usage**

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

**Arguments**

x	a character vector of surveys to cite
quiet	if set to TRUE, do not print entry, just return bibentry object
...	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", ...)
```

### Arguments

x	A <a href="#">survey</a> object
country.column	the name of the country denoting the country in which the survey participant was interviewed
...	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.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 <a href="#">survey</a> 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
split	whether to split the number of contacts and assortativity
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 and 2 for weekends)
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 <a href="#">get_survey</a> , <a href="#">check</a> and <a href="#">pop_age</a> (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 <a href="#">list_surveys</a> ), or a character vector of file names, or a <a href="#">survey</a> object (in which case only cleaning is done).
quiet	if TRUE, suppress messages
...	options for <a href="#">clean</a> , 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     *List all surveys available for download*

---

**Description**

List all surveys available for download

**Usage**

```
list_surveys()
```

**Value**

character vector of surveys

**Examples**

```
## Not run:  
list_surveys()  
  
## End(Not run)
```

---

 polymod

*Social contact data from 8 European countries*


---

### 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 ...

### Source

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

---

pop\_age

*Change age groups in population data*


---

### 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 <a href="#">list_surveys</a> ), or a character vector of file names, or a <a href="#">survey</a> object (in which case only cleaning is done).
country.column	column in the survey indicating the country
...	further arguments for <a href="#">get_survey</a>

**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

## \*Topic **datasets**

polymod, [8](#)

check, [2](#), [5](#)

cite, [3](#)

clean, [4](#), [6](#)

contact\_matrix, [2](#), [4](#)

get\_survey, [5](#), [6](#), [10](#)

limits\_to\_agegroups, [7](#)

list\_surveys, [6](#), [7](#), [10](#)

polymod, [8](#)

pop\_age, [5](#), [8](#)

reduce\_agegroups, [9](#)

survey, [2-6](#), [9](#), [10](#)

survey\_countries, [10](#)

wpp\_age, [11](#)

wpp\_countries, [11](#)