Package ‘outbreaks’
January 21, 2019

Title A Collection of Disease Outbreak Data
Version 1.5.0
Description Empirical or simulated disease outbreak data, provided either as
          RData or as text files.
Depends R (>= 3.0.0)
License GPL (>= 2)
Encoding UTF-8
LazyData true
RoxygenNote 6.1.1
Suggests testthat, covr, ape, incidence
URL https://github.com/reconhub/outbreaks
BugReports https://github.com/reconhub/outbreaks/issues
NeedsCompilation no
Author Thibaut Jombart [aut],
          Simon Frost [aut],
          Pierre Nouvellet [aut],
          Finlay Campbell [aut, cre],
          Bertrand Sudre [aut],
          Sang Woo Park [ctb],
          Juliet R.C. Pulliam [ctb]
Maintainer Finlay Campbell <f.campbell15@imperial.ac.uk>
Repository CRAN
Date/Publication 2019-01-21 10:00:03 UTC

R topics documented:

dengue_fais_2011 .................................................. 2
dengue_yap_2011 .................................................. 3
ebola_kikwit_1995 .................................................. 4
ebola_sim .......................................................... 5
Description

These data describe incidence of clinical cases of Dengue on the island of Fais, Micronesia.

Usage

dengue_fais_2011

Format

A data frame with 57 rows and 3 columns

- **onset_date** Date
- **nr** Days after starting date
- **value** Number of cases

The data on Dengue incidence reported by Funk et al. (2016) cover the period from 2011-09-15 to 2012-02-14, over which time a total of 157 clinical cases were reported among 294 residents. The first reported case is thought to be the index case. The population of Fais is concentrated in a single population centre, and is thought to have been immunologically naive at the time of infection.

Author(s)

Data from Funk et al. (2016), provided by Sebastian Funk (github.com/sbnfunk). Transfer to R and documentation by Finlay Campbell (<f.campbell115@imperial.ac.uk>).

Source

Funk et al. (2016)
References


Examples

```r
## show first few weeks of Dengue incidence
head(dengue_fais_RP11)
```

---

**Description**

These data describe incidence of clinical cases of Dengue on the Yap Main Islands, Micronesia.

**Usage**

dengue_yap_2011

**Format**

A data frame with 185 rows and 3 columns

- **onset_date**: Date
- **nr**: Days after starting date
- **value**: Number of cases

The data on Dengue incidence reported by Funk et al. (2016) cover the period from 2011-07-07 to 2012-04-12, over which time a total of 978 cases were reported among 7391 residents. Suspected Dengue cases were identified by the Yap Department of Health, using the WHO (2009) case definition. A small proportion of cases (9 time series for the Yap Main Islands. Dengue virus serotype 2 was confirmed by reverse transcriptase polymerase chain reaction by the CDC Dengue Branch, Puerto Rico.

**Author(s)**

Data from Funk et al. (2016), provided by Sebastian Funk (github.com/sbnfunk). Transfer to R and documentation by Finlay Campbell (<f.campbell115@imperial.ac.uk>).

**Source**

Funk et al. (2016)

**References**

Examples

```r
## show first few weeks of Dengue incidence
head(dengue_yap_2011)
```

---

**Description**

These data comprise of new cases of Ebola haemorrhagic fever in Kikwit, Democratic Republic of the Congo.

**Usage**

`ebola_kikwit_1995`

**Format**

A data frame with 192 rows and 4 columns

- **date**: Date
- **onset**: Number of new cases
- **deaths**: Number of deaths per day
- **reporting**: Whether data were reported on a daily basis

The data on daily cases reported by Khan et al. (1999) cover the period 1995-03-01 to 1995-07-16, over which time there were 291 cases and 236 deaths. The first case became ill on 1995-01-06, which is taken as the first timepoint in this version of the data. Over the entire period, there were 316 cases i.e. the onset times are not reported for 24 individuals, and the recovery times for the individuals who did not die are not reported.

**Author(s)**

Data from Khan et al. (1999), provided by T.J. McKinley. Transfer to R and documentation by Simon Frost (<sdwfrost@gmail.com>).

**Source**

Khan et al. (1999)

**References**

**ebola_sim**

### Examples

```r
## show first few cases
dput(head(ebola_kikwit_1995))
```

### Description

This simulated outbreak of Ebola Virus Disease matches some key properties of the West African Ebola outbreak of 2014-2015. Specifically, care was taken to use realistic delays (incubation period, serial interval, time to hospitalisation, etc.) and reproduction number (see references).

### Usage

- `ebola_sim`
- `ebola_sim_clean`

### Format

An object of class `list` of length 2.

### Details

This dataset is used for teaching purposes during Imperial College’s short course on infectious disease modelling. The exercise aims to simulate the response to an Ebola outbreak taking place in a single large city, and evaluate the impact of an intervention (increased bed capacity).

Note that to ensure realism, some errors have been introduced in this dataset. These can be identified as negative incubation periods (delay from infection to onset of symptoms). See example for a simple way to identify these cases. The dataset `ebola_sim_clean` is the same dataset, only dates of infection and onset have been set to ‘NA’.

### Author(s)

Data simulated by Pierre Nouvellet (<p.nouvellet@imperial.ac.uk>). Transfer to R and documentation by Thibaut Jombart (<thibautjombart@gmail.com>).

### References


```sh
```

Examples

```r
## identify mistakes in data entry (negative incubation period)
mistakes <- which(ebola_sim$linelist$date_of_onset <= ebola_sim$linelist$date_of_infection)

## check that ebola_sim_clean is identical after removing mistakes
identical(ebola_sim_clean$linelist, ebola_sim$linelist[-mistakes, ])
```

fluh7n9_china_2013  Influenza A H7N9 in China, 2013

Description

These data comprise of 136 cases of influenza A H7N9 in China, analysed by Kucharski et al. (2014).

Usage

fluh7n9_china_2013

Format

A data frame with 136 rows and 8 columns

Author(s)

Data collated by Adam Kucharski et al. from ProMed, WHO, FluTrackers, news reports and research articles. Transfer to R and documentation by Simon Frost (<sdwfrost@gmail.com>).

Source

http://datadryad.org/resource/doi:10.5061/dryad.2g43n/1

References


Examples

```r
## show first few cases
head(fluH7N9_china_2013)
```

**Description**

These data comprise of a time series of influenza cases in a boarding school in England. The original data were available only in a figure with some additional data in the main text; hence, the exact numbers vary depending on the source. These data are from Chapter 9 of De Vries et al. (1996).

**Usage**

`influenza_england_1978_school`

**Format**

A data frame with 14 rows and 3 columns

- **date**: Date
- **in_bed**: Number in bed
- **convalescent**: Number convalescing

**Details**

The index case was infected by 1978-01-10, and had febrile illness from 1978-01-15 to 1978-01-18. 512 boys out of 763 became ill.

**Author(s)**

Data from De Vries et al. (2006), from the original Anonymous (1978) figure. Transfer to R and documentation by Simon Frost (<sdwfrost@gmail.com>).

**Source**


**References**


Examples

```r
## show first few cases
head(influenza_england_1978_school)
```

---

measles_hagelloch_1861

*Measles in Hagelloch, Germany, 1861*

Description

These data comprise of 188 cases of measles among children in the German city of Hagelloch, 1861. The data were originally collected by Dr. Albert Pfeilsticker (1863) and augmented and re-analysed by Dr. Heike Oesterle (1992).

Usage

measles_hagelloch_1861

Format

A data frame with 188 rows and 12 columns

- **case_ID** Case ID number
- **infector** Number of patient who is the putative source of infection
- **date_of_prodrome** Date of onset of prodromal symptoms
- **date_of_rash** Date of onset of rash
- **date_of_death** Date of death (NA implies recovered)
- **age** Age in years (fractions ignored)
- **gender** Gender of the individual (factor: f, m)
- **family_ID** Family ID number
- **class** School class (factor: 0, preschool; 1, 1st class; 2, 2nd class )
- **complications** Complications (factor: no, yes)
- **x_loc** x coordinate of house (in metres). Scaling in metres is obtained by multiplying the original coordinates by 2.5 (see details in Neal and Roberts (2004))
- **y_loc** y coordinate of house (in metres). See x_loc above.

Author(s)

This version of the data was formatted from hagelloch.df in the surveillance package, which in turn was provided by Niels Becker via Peter Neal. Formatting to fit in with the other datasets in the outbreaks package by Simon Frost (<sdwfrost@gmail.com>).
Source

Pfeilsticker (1863) and Oesterle (1992).

References

Pfeilsticker, A. 1863. Beiträge zur Pathologie der Masern mit besonderer Berücksichtigung der
www.archive.org/details/beitrgezurpatho@pfeigoog.

Eberhard-Karls-Universitäat Tübingen.


Höhle M. 2007. surveillance: An R package for the monitoring of infectious diseases. Computa-
tional Statistics, 22:571-582.


Examples

```r
## show first few cases
head(mesles_hagelloch_1861)
```

Description

These datasets correspond to the initial information collected by the Epidemic Intelligence group at
European Centre for Disease Prevention and Control (ECDC) during the first weeks of the outbreak
of Middle East respiratory syndrome (MERS-CoV) outbreak (South Korea) in 2015. The data were
used to follow the daily evolution of this outbreak using public information available.

Usage

```r
mers_korea_2015
```

Format

A list of two dataframes:

- $linelist: A dataframe of MERS-CoV cases and their attributes
  - id: Unique identifier
  - age: Age
  - age_class: Age using 10-year groups
• sex: Sex
• place_infect: Probable region of infection
• reporting_ctry: Country reporting the case
• loc_hosp: Local hospital name where the case was hospitalized
• dt_onset: Date of onset of symptoms
• dt_report: Date of reporting
• week_report: Week number of date of reporting
• dt_start_exp: Date of first probable exposure to another MERS Co-V case
• dt_end_exp: Date of last probable exposure to another MERS Co-V case
• dt_diag: Date of MERS Co-V diagnosis
• outcome: Outcome (alive or dead)
• dt_death: Date of death

$contacts$: A dataframe describing the relationship between MERS Co-V cases

• from: Unique identifier of the probably source patient
• to: Unique identifier of the secondary case
• exposure: Probable place of exposure
• diff_dt_onset: Time in days between two successive cases

Details
This dataset is meant for teaching purposes; it represents neither the final outbreak investigation results nor a consolidated and complete description of the transmission chain.

Author(s)
Data collected by the European Centre for Disease Prevention and Control (Epidemic Intelligence and Response section, contact: Bertrand Sudre (<bertrand.sudre@ecdc.europa.eu>) and Kaja Kaasik-Aaslav(<kaja.KaasikAaslav@ecdc.europa.eu>). Transfer to R and documentation by Bertrand Sudre (<bertrand.sudre@ecdc.europa.eu>).

References

Examples
```r
# show the line list describing MERS Co-V cases and their attributes
head(mers_korea_2015$linelist)

# show the relationships between MERS Co-V cases
head(mers_korea_2015$contacts)
```
These data describe incidence of human cases of Nipah virus encephalitis in Malaysia and Singapore from January 1997 through April 1999.

Usage

nipah_malaysia

Format

A data frame with 49 rows and 5 columns

date  Onset date (weekly)
perak  Number of cases (Perak State, Malaysia)
negeri_sembilan  Number of cases (Negeri Sembilan State, Malaysia)
selangor  Number of cases (Selangor State, Malaysia)
singapore  Number of cases (Singapore)

Nipah virus is a paramyxovirus that occurs in flying fox (fruit bat) populations throughout Asia. The data provided are from the first known emergence of Nipah virus into humans. During this outbreak, the virus was transmitted from bats to pigs, where it circulated in commercial pig farms, infecting mostly farm and abattoir workers. The outbreak started in Perak State, later spreading to Negeri Sembilan and Selangor through sale of infected pigs. There were also 11 cases reported among abattoir workers in Singapore. The data, as published in Pulliam _et al._ (2011), include all 257 clinical cases recorded in humans from 1997-01-11 to 1999-04-14, when the outbreak ended following large-scale depopulation of pig farms. Human cases represent zoonotic infections, with little or no human-to-human transmission. Thus, the epidemic curve reflects transmission and spatial spread within pigs.

Author(s)

Data from Funk et al. (2016), provided by Juliet Pulliam (github.com/jrcpulliam).

Source

Pulliam et al. (2011)

References

Examples

```r
## show first few weeks of Dengue incidence
head(nipah_malaysia)

## convert data to incidence object and plot epicurve using the incidence package
library(incidence)
cases <- subset(nipah_malaysia, select = c("perak", "negeri_sembilan", "selangor", "singapore"))
i <- as.incidence(cases, dates = nipah_malaysia$date, interval = 7L)
plot(i)
```

---

norovirus_derbyshire_2001_school

*Norovirus in a primary school in Derbyshire, England, 2001*

Description

These data describe an outbreak of norovirus in the summer of 2001 in a primary school and nursery in Derbyshire, England.

Usage

norovirus_derbyshire_2001_school

Format

A data frame with 492 rows and 5 columns

- **class** School class of the child
- **day_absent** First day of absence from school
- **start_illness** First day of illness
- **end_illness** Last day of illness
- **day_vomiting** Day of vomiting in school

The data on norovirus cases were analysed by O’Neill and Marks (2005). As described in the paper, out of a total of 492 children in the school, 186 were absent from school with gastrointestinal symptoms. The school was cleaned on days 13 and 14, and on days 20 and 21, both of which were weekends, and the school was shut on days 18 and 19. Following the second cleaning, there were no further absences, although three children reported symptoms on day 22, the last day of the outbreak.

Author(s)

Data from O’Neill and Marks (2005), provided by Philip O’Neill. Transfer to R and documentation by Simon Frost (<sdwfrost@gmail.com>) and Finlay Campbell (<f.campbell15@imperial.ac.uk>).
rabies_car_2003

Source

O’Neill and Marks (2005)

References


Examples

```r
## show first few cases
head(norovirus_derbyshire_2001_school)
```

---

rabies_car_2003  
*Dog Rabies in Central African Republic, 2003-2012*

Description

These data document a dog rabies epidemic from 2003 to 2012 in Bangui, Central African Republic, and its surroundings. Data comprise dates and locations of the cases, as well as viral sequences of the pathogen for most cases.

Usage

```r
rabies_car_2003
```

Format

A list comprising a data.frame (`linelist`) and a DNAbin matrix (`dna`). `linelist` contains the following variables:

- `$index`: numeric identifier of the case
- `$date`: date of case reporting
- `$latitude`: the latitude of the collection point
- `$longitude`: the longitude of the collection point
- `$has_dna`: a logical indicating of the case has a matching pathogen sequence in `dna`

`dna` is a DNAbin matrix whose labels are to be matched against `linelist$index`.

Author(s)

Data from Transfer to R and documentation by Thibaut Jombart.
Source

The data were provided by the Institut Pasteur de Bangui, Bangui, République Centrafricaine, and the Institut Pasteur, Unit Lyssavirus Dynamics and Host Adaptation, WHO Collaborating Centre for Reference and Research on Rabies, Paris, France.

References

Cori et al. (submitted) A graph-based evidence synthesis approach to detecting outbreak clusters: an application to dog rabies.

Examples

```r
if (require(incidence) && require(ape)) {
  i <- incidence(rabies_car_2003$linelist$date, 28L)
  plot(i)

  tre <- nj(dist.dna(rabies_car_2003$dna))
  plot(tre, main = "Neighbour-Joining tree")
}
```

---

**sars_canada_2003**

Severe Acute Respiratory Syndrome in Canada, 2003

---

Description

These data comprise new cases of SARS in Canada in 2003. These data are from Chapter 9 of De Vries et al. (1996).

Usage

`sars_canada_2003`

Format

A data frame with 110 rows and 5 columns

- **date** Date
- **cases_travel** New cases attributed to travel
- **cases_household** New cases attributed to household transmission
- **cases_healthcare** New cases in a healthcare setting
- **cases_other** Other new cases

Author(s)

Data from De Vries et al. (2006), based on original data from Health Canada. Transfer to R and documentation by Simon Frost (<sdwfrost@gmail.com>).
Source

De Vries et al. (2006)

References


Examples

```r
## show first few cases
head(sars_canada_2003)
```

---

**smallpox_abakaliki_1967**

*Smallpox in Abakaliki, Nigeria, 1967*

**Description**

These data comprise of 32 cases of smallpox in Abakaliki, Nigeria in 1967, first described by Thompson and Foege (1968), which occurred predominantly in a religious group that refused medical interventions.

**Usage**

`smallpox_abakaliki_1967`

**Format**

A data frame with 32 rows and 8 columns

- `case_ID` Case identification number
- `date_of_onset` Date of onset of symptoms
- `age` Age in years
- `gender` Gender: female (f) or male (m) (factor)
- `vaccinated` Previously vaccinated: no (n) or yes (y) (factor)
- `vaccscar` Vaccination scar present: no (n) or yes (y) (factor)
- `ftc` Member of the Faith Tabernacle: no (n) or yes (y) (factor)
- `compound` Compound number (factor)

**Author(s)**

Data from Thompson and Foege (1968). Transfer to R and documentation by Simon Frost (<sdwfrost@gmail.com>).
Source

http://apps.who.int/iris/bitstream/10665/67462/1/WHO_SE_68.3.pdf

References


Examples

```r
## show first few cases
head(smallpox_abakaliki_1967)
```

---

**s_enteritidis_pt59**  *Salmonella Enteritidis PT59 outbreak*

**Description**

This dataset covers a food-borne outbreak of Salmonella Enteritidis PT59 investigated by Public Health England. The data includes a distribution network, and genetic clusters of bacteria isolated in contaminated patients. The data is anonymised: identifiers of the food distribution network have been hashed. The object `s_enteritidis_pt59` is a list containing:

**Usage**

```r
s_enteritidis_pt59
```

**Format**

An object of class `list` of length 2.

**Details**

- `Dgraph`: a `data.frame` containing with two columns `from` and `to` specifying (directed) edges of the food distribution network, alongside reported cases (terminal branches, or 'tips').
- `Dcluster`: a factor indicating the genetic cluster of named tips.

**Author(s)**

Thomas Inns <Thomas.Inns@phe.gov.uk>, Philip Ashton, Tim Dallman, Roberto Vivancos. Hashing and port to R by Thibaut Jombart.

**Source**

Multi-agency Salmonella Enteritidis PT59 Outbreak Control Team, chaired by Public Health England
Description
These data describe the daily incidence of Zika virus disease in Girardot, Colombia.

Usage
zika_girardot_2015

Format
A data frame with 93 rows and 2 columns

<table>
<thead>
<tr>
<th>Date</th>
<th>Daily incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>cases</td>
<td></td>
</tr>
</tbody>
</table>

The data on Zika virus disease incidence reported by Rojas et al. (2016) cover the period from October 2015 to January 2016, over which time a total of 1936 cases were reported to health authorities of Girardot (population of 102,225). Suspected cases were confirmed by reverse transcription-polymerase chain reaction (RT-PCR) in the serum of acute cases within five days of symptom onset.

Author(s)
Data from Rojas et al. (2016), provided by Diana P. Rojas (<dprojas@epi.ufl.edu>). Transfer to R and documentation by Finlay Campbell (<f.campbell15@imperial.ac.uk>).

Source
Rojas et al. (2016)

References

These data were provided under a Creative Commons Attribution (CC BY 4.0) Licence (http://creativecommons.org/licenses/by/4.0/).

Examples
```R
# show first few days of Zika incidence
table(zika_girardot_2015)
```
Description
These data describe the daily incidence of Zika virus disease in San Andres, Colombia.

Usage
zika_sanandres_2015

Format
A data frame with 101 rows and 2 columns

date  Date
cases  Daily incidence

The data on Zika virus disease incidence reported by Rojas et al. (2016) cover the period from September 2015 to January 2016, over which time a total of 928 cases were reported to health authorities of San Andres (population of 54,513). Suspected cases were confirmed by reverse transcription-polymerase chain reaction (RT-PCR) in the serum of acute cases within five days of symptom onset.

Author(s)
Data from Rojas et al. (2016), provided by Diana P. Rojas (<dprojas@epi.ufl.edu>). Transfer to R and documentation by Finlay Campbell (<f.campbel115@imperial.ac.uk>).

Source
Rojas et al. (2016)

References

These data were provided under a Creative Commons Attribution (CC BY 4.0) Licence (http://creativecommons.org/licenses/by/4.0/).

Examples
## show first few days of Zika incidence
head(zika_sanandres_2015)
**Description**

These data describe weekly incidence of probable and confirmed cases of Zika virus on the Yap Main Islands, Micronesia.

**Usage**

**Format**

A data frame with 29 rows and 3 columns

- **onset_date** Date
- **nr** Days after starting date
- **value** Number of cases per week

The data on weekly cases reported by Funk et al. (2016) cover the period from 2007-02-18 to 2007-09-02, over which time there were a total of 108 cases classified as probable (59) or confirmed (49) in a population of 7391. Cases were identified by a combination of prospective and retrospective surveillance at all health centres on Yap.

**Author(s)**

Data from Funk et al. (2016), provided by Sebastian Funk (github.com/sbnfunk). Transfer to R and documentation by Finlay Campbell (<f.campbell115@imperial.ac.uk>).

**Source**

Funk et al. (2016)

**References**


**Examples**

```r
# show first few weeks of Zika incidence
table(zika_yap_2007)
```
Index

*Topic datasets
  dengue_fais_2011, 2
  dengue_yap_2011, 3
  ebola_kikwit_1995, 4
  ebola_sim, 5
  fluH7N9_china_2013, 6
  influenza_england_1978_school, 7
  measles_hagelloch_1861, 8
  mers_korea_2015, 9
  nipah_malaysia, 11
  norovirus_derbyshire_2001_school, 12
  rabies_car_2003, 13
  s_enteritidis_pt59, 16
  sars_canada_2003, 14
  smallpox_abakaliki_1967, 15
  zika_girardot_2015, 17
  zika_sanandres_2015, 18
  zika_yap_2007, 19

  dengue_fais_2011, 2
  dengue_yap_2011, 3

  ebola_kikwit_1995, 4
  ebola_sim, 5
  ebola_sim_clean (ebola_sim), 5

  fluH7N9_china_2013, 6

  influenza_england_1978_school, 7

  measles_hagelloch_1861, 8
  mers_korea_2015, 9

  nipah_malaysia, 11
  norovirus_derbyshire_2001_school, 12

  rabies_car_2003, 13

  s_enteritidis_pt59, 16
  sars_canada_2003, 14