Package ‘infectiousR’

June 16, 2025

Type Package

Title Access Infectious and Epidemiological Data via 'disease.sh API'

Version 0.1.0

Maintainer Renzo Caceres Rossi <arenzocaceresrossi@gmail.com>

Description Provides functions to access real-time infectious disease data from the 'disease.sh API', including COVID-19 global, US states, continent, and country statistics, vaccination coverage, influenza-like illness data from Centers for Disease Control and Prevention (CDC), and more. Also includes curated datasets on a variety of infectious diseases such as influenza, measles, dengue, Ebola, tuberculosis, meningitis, AIDS, and others. The package supports epidemiological research and data analysis by combining API access with high-quality historical and survey datasets on infectious diseases. For more details on the 'disease.sh API', see <https://disease.sh/>.

License GPL-3

URL https://github.com/lightbluetitan/infectiousr, https://lightbluetitan.github.io/infectiousr/

BugReports https://github.com/lightbluetitan/infectiousr/issues

Encoding UTF-8

LazyData true

Depends R (>= 4.1.0)

Suggests ggplot2, testthat (>= 3.0.0), knitr, rmarkdown

Imports utils, httr, jsonlite, lubridate, dplyr

RoxygenNote 7.3.2

Config/testthat/edition 3

VignetteBuilder knitr

NeedsCompilation no

Author Renzo Caceres Rossi [aut, cre]

Repository CRAN

Date/Publication 2025-06-16 11:00:06 UTC

Contents

`active_hepatitis_df` *Chronic Active Hepatitis Clinical Trial*

Description

This dataset, active_hepatitis_df, is a data frame containing information from a clinical trial of 44 patients with chronic active hepatitis. Patients were randomized to receive either the drug prednisolone or no treatment (control group).

Usage

```
data(active_hepatitis_df)
```

Format

A data frame with 44 observations and 3 variables:

treatment Integer vector indicating treatment group: 1 for prednisolone, 0 for control

time Integer vector representing the time to event or censoring (in days)

status Integer vector indicating status: 1 for death, 0 for censored

Details

The dataset name has been kept as 'active_hepatitis_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the collett package version 0.1.0

`aids_azt_df` *AIDS Symptoms and AZT Use Data*

Description

This dataset, aids_azt_df, is a data frame containing cross-classified counts of AIDS symptoms and AZT use by race of the patients, as reported in a 1991 New York Times article.

Usage

```
data(aids_azt_df)
```

Format

A data frame with 4 observations and 4 variables:

yes Numeric vector indicating the number of patients showing AIDS symptoms

no Numeric vector indicating the number of patients not showing AIDS symptoms

azt Factor with 2 levels indicating AZT use (`yes`, `no`)

race Factor with 2 levels indicating patient race (`white`, `black`)

Details

The dataset name has been kept as 'aids_azt_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the cond package version 1.2-4

`bcg_vaccine_df` *BCG Vaccine Effectiveness Against Tuberculosis*

Description

This dataset, bcg_vaccine_df, is a data frame containing results from 13 studies examining the effectiveness of the Bacillus Calmette-Guerin (BCG) vaccine against tuberculosis.

Usage

```
data(bcg_vaccine_df)
```

Format

A data frame with 13 observations and 9 variables:

trial Integer identifier for each study

author Character vector indicating the lead author of each study

year Integer year in which the study was published

tpos Integer count of tuberculosis cases in the treatment group

tneg Integer count of non-cases in the treatment group

cpos Integer count of tuberculosis cases in the control group

cneg Integer count of non-cases in the control group

ablat Integer representing absolute latitude of study location

alloc Character string describing the method of allocation

Details

The dataset name has been kept as 'bcg_vaccine_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the metadat package version 1.4-0

`campy_infections_ts` *Campylobacter Infections Time Series*

Description

This dataset, campy_infections_ts, is a time series object containing the number of cases of campylobacter infections in the north of the province Quebec (Canada) in four week intervals from January 1990 to the end of October 2000. It contains 13 observations per year and 140 observations in total.

Usage

```
data(campy_infections_ts)
```

Format

A time series object of class `ts` with 140 observations, frequency 13, starting from 1990 to 2000 (end of October).

Details

The dataset name has been kept as 'campy_infections_ts' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'ts' indicates that the dataset is a time series object. The original content has not been modified in any way.

Source

Data taken from the tscount package version 1.4.3. Original study: Ferland, R., Latour, A. and Oraichi, D. (2006) Integer-valued GARCH process. *Journal of Time Series Analysis* 27(6), 923–942.

`china_dengue_tbl_df` *Dengue Cases in Mainland China (2005–2020)*

Description

This dataset, china_dengue_tbl_df, is a tibble containing annual records of indigenous and imported dengue cases in mainland China from 2005 to 2020.

Usage

```
data(china_dengue_tbl_df)
```

Format

A tibble with 16 observations and 5 variables:

year Integer year of observation (2005–2020)

dengue.cases.indigenous Numeric vector of indigenous dengue cases

dengue.cases.imported Numeric vector of imported dengue cases

counties.with.dengue.fever.indigenous Numeric vector of counties with reported indigenous dengue fever

counties.with.dengue.fever.imported Numeric vector of counties with reported imported dengue fever

Details

The dataset name has been kept as 'china_dengue_tbl_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'tbl_df' indicates that the dataset is a tibble. The original content has not been modified in any way.

Source

Data taken from the denguedatahub package version 2.1.1

`contagious_diseases_df`

Contagious Disease Data for US States

Description

This dataset, `contagious_diseases_df`, is a data frame containing yearly counts for Hepatitis A, Measles, Mumps, Pertussis, Polio, Rubella, and Smallpox for US states. The original data is courtesy of the Tycho Project.

Usage

```
data(contagious_diseases_df)
```

Format

A data frame with 16,065 observations and 6 variables:

disease Factor with 7 levels indicating the disease type

state Factor with 51 levels indicating the US state

year Numeric vector indicating the year of observation

weeks_reporting Numeric vector indicating the number of weeks reported

count Numeric vector indicating the number of cases reported

population Numeric vector indicating the population of the state in that year

Details

The dataset name has been kept as `contagious_diseases_df` to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix `_df` indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the dslabs package version 0.8.0. Original data courtesy of the Tycho Project (`http://www.tycho.pitt.edu/`).

`covid_mortality_df` *COVID-19 Cardiovascular Mortality*

Description

This dataset, `covid_mortality_df`, is a data frame containing several effect estimates (β) and their standard errors for the impact of cardiovascular disease on the mortality of COVID-19 reported in the literature.

Usage

```
data(covid_mortality_df)
```

Format

A data frame with 6 observations and 3 variables:

study Character vector with the name or reference of each study

beta Numeric vector representing the estimated effect size (β)

se Numeric vector representing the standard error associated with each estimate

Details

The dataset name has been kept as `covid_mortality_df` to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix `_df` indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the PRP package version 0.1.1

`covid_new_york_df` *New York City COVID-19 Data*

Description

This dataset, covid_new_york_df, is a data frame containing daily proportions of COVID-19 cases, hospitalizations, and deaths by borough in New York City through 2020-06-30.

Usage

```
data(covid_new_york_df)
```

Format

A data frame with 615 observations and 5 variables:

date Date of observation

borough Character vector indicating the borough (e.g., Manhattan, Bronx, etc.)

case Integer vector representing the number of reported COVID-19 cases

hospitalization Integer vector representing the number of hospitalizations

death Integer vector representing the number of deaths

Details

The dataset name has been kept as 'covid_new_york_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the incidental package version 0.1

`covid_severity_df` *COVID-19 Cardiovascular Severity*

Description

This dataset, `covid_severity_df`, is a data frame containing several effect estimates (β) and their standard errors for the impact of cardiovascular disease on the severe case rate of COVID-19 as reported in the literature.

Usage

```
data(covid_severity_df)
```

Format

A data frame with 6 observations and 3 variables:

study Character vector with the name or reference of each study

beta Numeric vector representing the estimated effect size (β)

se Numeric vector representing the standard error associated with each estimate

Details

The dataset name has been kept as `covid_severity_df` to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix `_df` indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the PRP package version 0.1.1

`diphtheria_philly_df` *Weekly Diphtheria Incidence in Philadelphia*

Description

This dataset, diphtheria_philly_df, is a data frame containing the weekly incidence of diphtheria in Philadelphia between 1914 and 1947.

Usage

```
data(diphtheria_philly_df)
```

Format

A data frame with 1774 observations and 4 variables:

YEAR Integer vector representing the year of observation (1914–1947)

WEEK Integer vector representing the epidemiological week (1–52)

PHILADELPHIA Integer vector representing the weekly incidence of diphtheria in Philadelphia

TIME Numeric vector representing the continuous time index

Details

The dataset name has been kept as 'diphtheria_philly_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the epimdr package version 0.6-5

`ebola_cases_df` *Time Series Counts of Ebola Cases*

Description

This dataset, ebola_cases_df, is a data frame containing daily time series counts of new individuals exhibiting clinical signs of Ebola virus disease, as well as the number of daily removals (e.g., deaths or recoveries), during the 1995 Ebola epidemic in the Democratic Republic of Congo (DRC).

Usage

```
data(ebola_cases_df)
```

Format

A data frame with 192 observations and 3 variables:

time Integer indicating the number of days since the beginning of observation

clin_signs Integer indicating the number of new individuals with clinical signs of Ebola

removals Integer indicating the number of new removals (e.g., deaths or recoveries)

Details

The dataset name has been kept as 'ebola_cases_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the SimBIID package version 0.2.2

`ebola_sleone_df` *Ebola Cases in Sierra Leone, Africa*

Description

This dataset, ebola_sleone_df, is a data frame containing the cumulative number of Ebola virus disease cases in Sierra Leone, Africa, recorded from May 1, 2014 to December 16, 2015.

Usage

```
data(ebola_sleone_df)
```

Format

A data frame with 110 observations and 2 variables:

Day Integer indicating the number of days since May 1, 2014

Cases Integer representing the cumulative number of Ebola cases reported

Details

The dataset name has been kept as 'ebola_sleone_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the MMAC package version 0.1.2

`ebola_survey_tbl_df` *Survey on Ebola Quarantine*

Description

This dataset, ebola_survey_tbl_df, is a tibble containing responses from a poll conducted in New York City between October 26th and 28th, 2014. The poll was conducted shortly after a doctor who had treated Ebola patients in Guinea was diagnosed with Ebola in New York City. Participants were asked whether they favored a "mandatory 21-day quarantine for anyone who has come in contact with an Ebola patient". The survey included responses from 1,042 adults residing in New York.

Usage

```
data(ebola_survey_tbl_df)
```

Format

A tibble with 1,042 observations and 1 variable:

quarantine Factor with two levels indicating whether the respondent supports a mandatory 21-day quarantine for individuals who have come in contact with an Ebola patient

Details

The dataset name has been kept as 'ebola_survey_tbl_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'tbl_df' indicates that the dataset is a tibble. The original content has not been modified in any way.

Source

Data taken from the openintro package version 2.5.0

`ecoli_infections_df` *E. coli Infections Time Series*

Description

This dataset, ecoli_infections_df, is a data frame containing the weekly number of reported disease cases caused by *Escherichia coli* in the state of North Rhine-Westphalia (Germany) from January 2001 to May 2013. The data excludes cases of EHEC (enterohemorrhagic *E. coli*) and HUS (hemolytic uremic syndrome).

Usage

```
data(ecoli_infections_df)
```

Format

A data frame with 646 observations and 3 variables:

year Numeric variable indicating the calendar year of observation

week Numeric variable indicating the calendar week (1 to 52 or 53)

cases Numeric variable representing the number of reported *E. coli* cases

Details

The dataset name has been kept as 'ecoli_infections_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the tscount package version 1.4.3

`ehec_infections_df` *EHEC Infections Time Series*

Description

This dataset, ehec_infections_df, is a data frame containing the weekly number of reported EHEC/HUS infections in the state of North Rhine-Westphalia (Germany) from January 2001 to May 2013.

Usage

```
data(ehec_infections_df)
```

Format

A data frame with 646 observations and 3 variables:

year Numeric variable indicating the calendar year of observation

week Numeric variable indicating the calendar week (1 to 52 or 53)

cases Numeric variable representing the number of reported EHEC/HUS cases

Details

The dataset name has been kept as 'ehec_infections_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the tscount package version 1.4.3

`flu_enrich_df` *Flu Enrichment Gene Data*

Description

This dataset, flu_enrich_df, is a data frame containing gene-set enrichment information for genes that have been identified as having an effect on influenza-virus replication.

Usage

```
data(flu_enrich_df)
```

Format

A data frame with 5719 observations and 3 variables:

nflugenes Numeric vector representing gene identifiers with an effect on influenza-virus replication

setsize Integer vector representing the size of each gene set

GO_terms Factor vector representing Gene Ontology terms associated with each gene set

Details

The dataset name has been kept as 'flu_enrich_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the rvalues package version 0.7.1

`fungal_infections_df` *Fungal Infections Treatment Data*

Description

This dataset, fungal_infections_df, is a data frame containing results from a clinical trial on the success of a particular treatment for fungal infections across five research units. Interest in the study focuses on the treatment effect.

Usage

```
data(fungal_infections_df)
```

Format

A data frame with 10 observations and 4 variables:

success Numeric vector indicating the number of treatment successes

failure Numeric vector indicating the number of treatment failures

group Factor with 2 levels indicating treatment group (`control`, `treated`)

center Factor with 5 levels indicating the research center where the trial was conducted

Details

The dataset name has been kept as 'fungal_infections_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the cond package version 1.2-4

`get_covid_stats_by_continent` *Get COVID-19 Statistics for All Continents*

Description

Retrieves real-time COVID-19 totals for all continents from the 'disease.sh' API.

Usage

```
get_covid_stats_by_continent(
  yesterday = FALSE,
  twoDaysAgo = FALSE,
  sort = NULL,
  allowNull = FALSE
)
```

Arguments

`yesterday`	Logical. If `TRUE`, retrieves data reported from the previous day. Default is `FALSE`.
`twoDaysAgo`	Logical. If `TRUE`, retrieves data reported two days ago. Default is `FALSE`.
`sort`	Character. Field to sort results by. Options include: `"cases"`, `"todayCases"`, `"deaths"`, `"recovered"`, `"active"`, etc.
`allowNull`	Logical. If `TRUE`, missing values are returned as `NA` instead of `0`. Default is `FALSE`.

Details

This function retrieves COVID-19 summary data for each continent. You may specify whether to get data from today, yesterday, or two days ago.

Value

A data frame containing:

- `continent`: Continent name.
- `updated`: Last updated timestamp (as POSIXct in UTC).
- `cases`: Total confirmed cases.
- `todayCases`: New confirmed cases today.
- `deaths`: Total deaths.
- `todayDeaths`: New deaths today.
- `population`: Continent population estimate.

Note

Requires internet access.

References

API Docs: https://disease.sh/docs/#/COVID-19

Examples

```
# Get current COVID-19 stats for all continents
get_covid_stats_by_continent()

# Get yesterday's data sorted by number of cases
get_covid_stats_by_continent(yesterday = TRUE, sort = "cases")
```

`get_covid_stats_by_country`

Get COVID-19 Statistics for All Countries

Description

Retrieves real-time COVID-19 totals for all countries from the ’disease.sh’ API.

Usage

```
get_covid_stats_by_country(
  yesterday = FALSE,
  twoDaysAgo = FALSE,
  sort = NULL,
  allowNull = FALSE
)
```

Arguments

`yesterday`	Logical. If `TRUE`, retrieves data reported from the previous day. Default is `FALSE`.
`twoDaysAgo`	Logical. If `TRUE`, retrieves data reported two days ago. Default is `FALSE`.
`sort`	Character. Field to sort results by. Options include: `"cases"`, `"todayCases"`, `"deaths"`, `"recovered"`, `"active"`, etc.
`allowNull`	Logical. If `TRUE`, missing values are returned as `NA` instead of `0`. Default is `FALSE`.

Details

This function fetches COVID-19 summary statistics for each country. Useful for global surveillance or international comparisons.

Value

A data frame containing:

- `country`: Country name.
- `updated`: Last updated timestamp (as POSIXct in UTC).

- `cases`: Total confirmed cases.
- `todayCases`: New confirmed cases today.
- `deaths`: Total deaths.
- `todayDeaths`: New deaths today.
- `population`: Population estimate for each country.

Note

Requires internet access.

References

API Docs: https://disease.sh/docs/#/COVID-19

Examples

```
# Get real-time COVID-19 data for all countries
get_covid_stats_by_country()

# Get sorted data by number of deaths reported yesterday
get_covid_stats_by_country(yesterday = TRUE, sort = "deaths")
```

`get_covid_stats_by_country_name`
Get COVID-19 Statistics for a Specific Country

Description

Retrieves COVID-19 totals for a given country using the ’disease.sh’ API.

Usage

```
get_covid_stats_by_country_name(
  country,
  yesterday = FALSE,
  twoDaysAgo = FALSE,
  strict = TRUE,
  allowNull = FALSE
)
```

Arguments

`country`	Character. A country name, ISO2, ISO3 code, or country ID.
`yesterday`	Logical. If `TRUE`, gets data reported from the previous day. Default is `FALSE`.
`twoDaysAgo`	Logical. If `TRUE`, gets data reported two days ago. Default is `FALSE`.
`strict`	Logical. If `TRUE` (default), disables fuzzy matching (e.g., avoids confusion between "Oman" and "Romania").
`allowNull`	Logical. If `TRUE`, allows null values (returned as `NA`). Default is `FALSE`.

Details

This function accesses COVID-19 data for a specific country based on its name or ISO code.

Value

A data frame with the following columns:

- `country`: Country name.
- `updated`: Timestamp of last update (POSIXct in UTC).
- `cases`: Total confirmed cases.
- `todayCases`: New confirmed cases today.
- `deaths`: Total deaths.
- `recovered`: Total recoveries.
- `population`: Estimated population.

Note

Requires internet connection.

References

API Docs: https://disease.sh/docs/#/COVID-19

Examples

```
# Get data for Brazil
get_covid_stats_by_country_name("Brazil")

# Get data for the USA using ISO2 code
get_covid_stats_by_country_name("US", yesterday = TRUE)
```

`get_covid_stats_for_state`

Get COVID-19 Statistics for Specific US State(s)

Description

Retrieves real-time COVID-19 totals for one or more U.S. states from the 'disease.sh' API.

Usage

```
get_covid_stats_for_state(states, yesterday = FALSE, allowNull = FALSE)
```

Arguments

`states`	A character string with the name of a U.S. state or a comma-separated list of state names. Names must be spelled correctly.
`yesterday`	Logical. If `TRUE`, returns data from the previous day. Default is `FALSE`.
`allowNull`	Logical. If `TRUE`, missing values are returned as `NA` instead of `0`. Default is `FALSE`.

Details

This function sends a GET request to the 'disease.sh' API for COVID-19 statistics in one or more U.S. states. If multiple states are passed, they must be comma-separated and correctly spelled. The 'updated' field is returned in milliseconds and is converted to a POSIXct datetime.

Value

A data frame containing the following columns:

- `state`: State name.
- `updated`: Last updated timestamp (converted to human-readable datetime in UTC).
- `cases`: Total confirmed cases.
- `todayCases`: New confirmed cases today.
- `deaths`: Total deaths.
- `todayDeaths`: New deaths today.
- `population`: State population estimate.

Note

Requires an internet connection.

References

API Docs: https://disease.sh/docs/#/COVID-19

Examples

```
# Retrieve COVID-19 data for California
ca <- get_covid_stats_for_state("California")

# Retrieve yesterday's data for New York and Texas
ny_tx <- get_covid_stats_for_state("New York,Texas", yesterday = TRUE)
```

`get_global_covid_stats`

Get Global COVID-19 Statistics

Description

Retrieves real-time global statistics on COVID-19 from the ’disease.sh’ API.

Usage

```
get_global_covid_stats()
```

Details

This function sends a GET request to the ’disease.sh’ API and parses the returned JSON into a structured and user-friendly data frame. The timestamp is converted to a readable date-time format (in UTC).

Value

A data frame with the following columns:

- `updated`: Last updated time (as a human-readable date-time).
- `cases`: Total confirmed cases worldwide.
- `todayCases`: Number of new confirmed cases today.
- `deaths`: Total confirmed deaths worldwide.
- `recovered`: Total number of recovered patients.
- `todayRecovered`: Number of recovered patients today.
- `active`: Current active cases.
- `critical`: Current number of critical cases.
- `tests`: Total number of tests performed.
- `population`: Estimated global population.
- `affectedCountries`: Number of countries affected.

Note

An internet connection is required to use this function.

References

API Docs: https://disease.sh/docs/#/COVID-19

Examples

```
global_stats <- get_global_covid_stats()
print(global_stats)
```

`get_influenza_cdc_ili` *Get CDC Influenza-like Illness (ILI) Data*

Description

Retrieves ILI data for the 2019 and 2020 influenza outbreaks from the US CDC.

Usage

```
get_influenza_cdc_ili()
```

Details

This endpoint provides historical data for flu-like symptoms reported in the United States, sourced from the CDC ILINet.

Value

A list containing:

- `updated`: Last update timestamp (POSIXct).
- `source`: Source of the data.
- `data`: A data frame with the following columns:
  - `week`: Week of reporting.
  - `age 5-24`, `age 25-49`, `age 50-64`, `age 64+`: ILI counts per age group.
  - `totalILI`: Total ILI cases.
  - `totalPatients`: Total patients.

Note

Requires internet connection.

References

API Docs: https://disease.sh/docs/#/Influenza/get_v3_influenza_cdc_ILINet

Examples

```
get_influenza_cdc_ili()
```

`get_us_states_covid_stats`

Get COVID-19 Statistics for U.S. States and Territories

Description

Retrieves real-time COVID-19 totals from the 'disease.sh' API for all 50 U.S. states, as well as U.S. territories (e.g., Puerto Rico, Guam), special jurisdictions (e.g., Veteran Affairs, U.S. Military), and others (e.g., cruise ships, repatriated individuals).

Usage

```
get_us_states_covid_stats()
```

Details

This function sends a GET request to the 'disease.sh' API endpoint for US state-level COVID-19 statistics and parses the response into a structured data frame. The timestamp is converted to a readable date-time format (in UTC).

Value

A data frame with the following columns:

- `state`: Name of the U.S. state.
- `cases`: Total confirmed cases in the state.
- `todayCases`: New confirmed cases today.
- `deaths`: Total deaths in the state.
- `todayDeaths`: New deaths today.
- `active`: Current active cases.
- `population`: Estimated state population.

Note

An internet connection is required to use this function.

References

API Docs: https://disease.sh/docs/#/COVID-19:

Examples

```
us_states_stats <- get_us_states_covid_stats()
head(us_states_stats)
```

`gonorrhea_ma_df` *Weekly Gonorrhea Cases in Massachusetts*

Description

This dataset, gonorrhea_ma_df, is a data frame containing weekly cases of gonorrhea in Massachusetts between 2006 and 2015.

Usage

```
data(gonorrhea_ma_df)
```

Format

A data frame with 422 observations and 4 variables:

number Integer vector representing the number of weekly gonorrhea cases

year Numeric vector representing the year of observation (2006–2015)

week Numeric vector representing the epidemiological week (1–52)

time Numeric vector representing the continuous time index

Details

The dataset name has been kept as 'gonorrhea_ma_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the epimdr package version 0.6-5

`hepatitisA_df` *Hepatitis A Prevalence in Bulgaria*

Description

This dataset, hepatitisA_df, is a data frame containing information from a cross-sectional survey conducted in 1964 on the prevalence of hepatitis A in individuals from Bulgaria. The surveyed population includes individuals aged between 1 and 86 years.

Usage

```
data(hepatitisA_df)
```

Format

A data frame with 83 observations and 3 variables:

t Integer vector indicating the age of the individuals

freq1 Integer vector representing the frequency of individuals tested

freq2 Integer vector representing the frequency of individuals with antibodies to hepatitis A

Details

The dataset name has been kept as 'hepatitisA_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the curstatCI package version 0.1.1

`india_dengue_tbl_df` *Dengue/DHF Situation in India Since 2017*

Description

This dataset, india_dengue_tbl_df, is a tibble containing state and union territory-wise annual dengue/DHF (Dengue Hemorrhagic Fever) cases and deaths in India since 2017.

Usage

```
data(india_dengue_tbl_df)
```

Format

A tibble with 432 observations and 5 variables:

area Character vector indicating the State or Union Territory

type Character vector indicating whether the entry refers to 'cases' or 'deaths'

year Character vector indicating the year of observation

additional_information Character vector providing supplemental information

value Numeric vector indicating the number of cases or deaths

Details

The dataset name has been kept as 'india_dengue_tbl_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'tbl_df' indicates that the dataset is a tibble (enhanced data frame). The original content has not been modified in any way.

Source

Data taken from the denguedatahub package version 2.1.1

`infectiousR` *infectiousR: Access Infectious and Epidemiological Data via 'disease.sh API'*

Description

This package provides functions to access real-time infectious disease data from the 'disease.sh API', including COVID-19 global, US states, continent, and country statistics, vaccination coverage,influenza-like illness data from Centers for Disease Control and Prevention (CDC), also includes curated datasets on a variety of infectious diseases such as influenza, measles, dengue, Ebola, tuberculosis, meningitis, AIDS, and others.

Details

infectiousR: Access Infectious and Epidemiological Data via 'disease.sh API'

Access Infectious and Epidemiological Data via 'disease.sh API'.

Author(s)

Maintainer: Renzo Caceres Rossi <`arenzocaceresrossi@gmail.com`>

See Also

Useful links:

- `https://github.com/lightbluetitan/infectiousr`

`influenza_ice_df` *Monthly Influenza Incidence in Iceland*

Description

This dataset, influenza_ice_df, is a data frame containing monthly incidence data of influenza-like illness (ILI) in Iceland between 1980 and 2009.

Usage

```
data(influenza_ice_df)
```

Format

A data frame with 360 observations and 3 variables:

month Integer vector representing the month of observation (1–12)

year Integer vector representing the year of observation (1980–2009)

ili Integer vector representing the monthly incidence of influenza-like illness

Details

The dataset name has been kept as 'influenza_ice_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the epimdr package version 0.6-5

`influenza_infections_df` *Influenza Infections Time Series*

Description

This dataset, influenza_infections_df, is a data frame containing the weekly number of reported influenza cases in the state of North Rhine-Westphalia (Germany) from January 2001 to May 2013.

Usage

```
data(influenza_infections_df)
```

Format

A data frame with 646 observations and 3 variables:

year Numeric variable indicating the calendar year of observation

week Numeric variable indicating the calendar week (1 to 52 or 53)

cases Numeric variable representing the number of reported influenza cases

Details

The dataset name has been kept as 'influenza_infections_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the tscount package version 1.4.3

`influenza_pneumonia_ts`

US Pneumonia and Influenza Death Rates

Description

This dataset, `influenza_pneumonia_ts`, is a time series containing monthly pneumonia and influenza deaths per 10,000 people in the United States over a period of 11 years, from 1968 to 1978.

Usage

```
data(influenza_pneumonia_ts)
```

Format

A time series object with 132 monthly observations:

value Monthly pneumonia and influenza deaths per 10,000 people in the United States from 1968 to 1978.

Details

The dataset name has been kept as `influenza_pneumonia_ts` to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the **infectiousR** package and assists users in identifying its specific characteristics. The suffix `_ts` indicates that the dataset is a time series object. The original content has not been modified in any way.

Source

Data taken from the **astsa** package version 2.2.

`influenza_vax_survey_df`

Influenza Vaccination Survey

Description

This dataset, influenza_vax_survey_df, is a data frame containing aggregated responses from three RAND American Life Panel (ALP) surveys regarding individuals' probability of vaccinating for influenza. The responses were discretized to "Never" (0%), "Always" (100%), or "Sometimes" (any other value). After merging, missing responses were coded as "Missing", and respondents were grouped and counted by all three coded responses.

Usage

```
data(influenza_vax_survey_df)
```

Format

A data frame with 117 observations and 6 variables:

survey Factor indicating which of the three ALP surveys the response came from

freq Integer indicating frequency count of grouped respondents

subject Integer identifier for each subject

response Factor with 4 levels: "Never", "Sometimes", "Always", and "Missing"

start_date Date indicating the start of the survey

end_date Date indicating the end of the survey

Details

The dataset name has been kept as 'influenza_vax_survey_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the ggalluvial package version 0.12.5

`korea_dengue_tbl_df` *Imported Dengue Cases in Korea*

Description

This dataset, korea_dengue_tbl_df, is a tibble containing information on imported dengue cases in Korea from the years 2011 to 2015. The data were collected by the Korea Centers for Disease Control and Prevention (KCDC).

Usage

```
data(korea_dengue_tbl_df)
```

Format

A tibble with 33 observations and 7 variables:

Country Character vector indicating the country of origin of the dengue cases

Region Character vector indicating the region within the country

2011 Character vector indicating the number of imported cases in 2011

2012 Character vector indicating the number of imported cases in 2012

2013 Character vector indicating the number of imported cases in 2013

2014 Character vector indicating the number of imported cases in 2014

2015 Character vector indicating the number of imported cases in 2015

Details

The dataset name has been kept as 'korea_dengue_tbl_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'tbl_df' indicates that the dataset is a tibble. The original content has not been modified in any way.

Source

Data taken from the denguedatahub package version 2.1.1

`malaria_mice_df` *Daily Measures of Malaria-Infected Mice*

Description

This dataset, malaria_mice_df, is a data frame containing daily data on laboratory mice infected with various strains of *Plasmodium chaubaudi*.

Usage

```
data(malaria_mice_df)
```

Format

A data frame with 1300 observations and 11 variables:

Line Integer vector indicating the parasite line

Day Integer vector representing the day of observation

Box Integer vector identifying the box where the mouse was housed

Mouse Integer vector identifying the individual mouse

Treatment Factor indicating the treatment group (6 levels)

Ind2 Integer vector used to identify individual measurements

Weight Numeric vector indicating the weight of the mouse

Glucose Integer vector indicating glucose levels

RBC Numeric vector representing red blood cell counts

Sample Integer vector identifying sample number

Para Numeric vector indicating parasitemia levels

Details

The dataset name has been kept as 'malaria_mice_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the epimdr package version 0.6-5

`measles_infections_df` *Measles Infections Time Series*

Description

This dataset, measles_infections_df, is a data frame containing the weekly number of reported measles infections in the state of North Rhine-Westphalia (Germany) from January 2001 to May 2013.

Usage

```
data(measles_infections_df)
```

Format

A data frame with 646 observations and 3 variables:

year Numeric variable indicating the calendar year of observation

week Numeric variable indicating the calendar week (1 to 52 or 53)

cases Numeric variable representing the number of reported measles cases

Details

The dataset name has been kept as 'measles_infections_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the tscount package version 1.4.3

`measles_survey_df` *Measles Non-Vaccination Parent Survey*

Description

This dataset, `measles_survey_df`, is a data frame containing the results of a survey conducted by Roberts et al. (1995) on parents whose children had not been immunized against measles during a recent campaign targeting all children in the first five years of secondary school.

Usage

```
data(measles_survey_df)
```

Format

A data frame with 307 observations and 11 variables:

school Factor with 10 levels indicating the school

form Factor with 2 levels indicating school form

returnf Factor with 2 levels indicating if the form was returned

consent Factor with 2 levels indicating if consent was given

hadmeas Factor with 2 levels indicating if the child had measles

previmm Factor with 2 levels indicating previous immunization

sideeff Factor with 2 levels indicating concerns about side effects

gp Factor with 2 levels indicating whether GP advised

noshot Factor with 2 levels indicating general refusal to vaccinate

notser Factor with 2 levels indicating the child was not seriously ill

gpadv Factor with 2 levels indicating GP advice against immunization

Details

The dataset name has been kept as `measles_survey_df` to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix `_df` indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the SDaA package version 0.1-5

`meningitis_df` *Meningococcal Data with Missing Response*

Description

This dataset, meningitis_df, is a data frame containing data from a brief outbreak of meningococcal disease at the University of Illinois, Urbana-Champaign campus during the years 1991 and 1992.

Usage

```
data(meningitis_df)
```

Format

A data frame with 60 observations and 6 variables:

- **Set** Integer indicating the matched set identifier
- **CaseCntrl** Integer indicator variable for case (1) or control (0)
- **Reftime** Numeric value representing the reference time (e.g., time of exposure)
- **Numnill** Integer indicating the number of ill roommates
- **Numsleep** Integer indicating the number of roommates who slept in the room
- **Smoke** Integer indicator for whether the subject smokes (1 = yes, 0 = no)

Details

The dataset name has been kept as 'meningitis_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the glmfitmiss package version 2.1.0

`rubella_austria_df` *Rubella Prevalence in Austrian Males*

Description

This dataset, rubella_austria_df, is a data frame containing prevalence data of rubella in 230 Austrian males older than three months, for whom the exact date of birth was known. Each individual was tested at the Institute of Virology, Vienna during the period 1–25 March 1988 for immunization against Rubella.

Usage

```
data(rubella_austria_df)
```

Format

A data frame with 225 observations and 3 variables:

- **t** Numeric vector representing age or time (in months or years as recorded)
- **freq1** Integer vector representing frequency count 1
- **freq2** Integer vector representing frequency count 2

Details

The dataset name has been kept as 'rubella_austria_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the curstatCI package version 0.1.1

`rubella_peru_df` *Rubella in Peru Data*

Description

This dataset, `rubella_peru_df`, is a data frame containing rubella incidence data by age as studied by Metcalf et al. (2011) in Peru.

Usage

```
data(rubella_peru_df)
```

Format

A data frame with 95 observations and 4 variables:

age Numeric vector indicating the age of individuals

incidence Integer vector indicating the number of rubella cases per age group

cumulative Integer vector indicating the cumulative number of cases by age

n Integer vector representing the sample size for each age group

Details

The dataset name has been kept as `rubella_peru_df` to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix `_df` indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the epimdr package version 0.6-5

`sars_canada_df` *Severe Acute Respiratory Syndrome in Canada, 2003*

Description

This dataset, sars_canada_df, is a data frame containing information on the daily incidence of SARS (Severe Acute Respiratory Syndrome) cases in Canada during the 2003 outbreak. The data include new cases attributed to travel, household transmission, healthcare settings, and other sources.

Usage

```
data(sars_canada_df)
```

Format

A data frame with 110 observations and 5 variables:

date Date object representing the reporting date

cases_travel Integer vector indicating new SARS cases linked to travel

cases_household Integer vector indicating new SARS cases from household transmission

cases_healthcare Integer vector indicating new SARS cases from healthcare settings

cases_other Integer vector indicating new SARS cases from other or unknown sources

Details

The dataset name has been kept as 'sars_canada_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the outbreaks package version 1.9.0

`smallpox_nigeria_df` *Smallpox in Abakaliki, Nigeria, 1967*

Description

This dataset, smallpox_nigeria_df, is a data frame containing data on 32 cases of smallpox that occurred in Abakaliki, Nigeria, in 1967. These cases were first described by Thompson and Foege (1968) and occurred predominantly in a religious group that refused medical interventions.

Usage

```
data(smallpox_nigeria_df)
```

Format

A data frame with 32 observations and 8 variables:

case_ID Integer identifier for each smallpox case

date_of_onset Date of symptom onset

age Age of the individual (integer)

gender Factor with two levels indicating gender

vaccinated Factor with two levels indicating if the individual was vaccinated

vaccscar Factor with two levels indicating presence of vaccination scar

ftc Factor with two levels; additional epidemiological classification

compound Factor with nine levels indicating compound of residence

Details

The dataset name has been kept as 'smallpox_nigeria_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the outbreaks package version 1.9.0

`spanish_flu_df` *Daily 1918 Flu Deaths*

Description

This dataset, spanish_flu_df, is a data frame containing daily mortality data from the 1918 flu pandemic covering the period from 1918-09-01 through 1918-12-31 in Indiana, Kansas, and Philadelphia.

Usage

```
data(spanish_flu_df)
```

Format

A data frame with 122 observations and 4 variables:

Date Date of recorded mortality

Indiana Integer vector representing daily flu-related deaths in Indiana

Kansas Integer vector representing daily flu-related deaths in Kansas

Philadelphia Integer vector representing daily flu-related deaths in Philadelphia

Details

The dataset name has been kept as 'spanish_flu_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the incidental package version 0.1

`streptomycin_tbl_df` *Tuberculosis Streptomycin RCT (1948)*

Description

This dataset, streptomycin_tbl_df, is a tibble containing the results of a randomized, placebo-controlled, prospective 2-arm trial evaluating the use of streptomycin (2 grams daily) versus placebo in the treatment of tuberculosis among 107 young patients. The study was conducted by the Streptomycin in Tuberculosis Trials Committee and published in the British Medical Journal in 1948.

Usage

```
data(streptomycin_tbl_df)
```

Format

A tibble with 107 observations and 13 variables:

patient_id Character identifier for each patient

arm Factor indicating treatment arm: streptomycin (A2) or placebo (A1)

dose_strep_g Numeric dose of streptomycin in grams

dose_PAS_g Numeric dose of para-aminosalicylic acid (PAS) in grams

gender Factor with two levels indicating patient gender

baseline_condition Factor indicating the baseline clinical condition of the patient

baseline_temp Factor indicating baseline temperature category

baseline_esr Factor indicating baseline erythrocyte sedimentation rate (ESR) category

baseline_cavitation Factor indicating the presence or absence of lung cavitation at baseline

strep_resistance Factor indicating the level of resistance to streptomycin

radiologic_6m Factor describing radiological outcomes at 6 months

rad_num Numeric radiologic score at 6 months

improved Logical indicator of clinical improvement

Details

The dataset name has been kept as 'streptomycin_tbl_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'tbl_df' indicates that the dataset is a tibble (a modern form of data frame). The original content has not been modified in any way.

Source

Data taken from the medicaldata package version 0.2.0

`us_covid_cases_df` *US Lab-Confirmed COVID-19 Cases*

Description

This dataset, `us_covid_cases_df`, is a data frame containing the number of laboratory-confirmed COVID-19 cases in the United States, as reported by the Centers for Disease Control and Prevention (CDC), between January 1, 2020 and May 11, 2023, the end of the public health emergency declaration.

Usage

```
data(us_covid_cases_df)
```

Format

A data frame with 1227 observations and 2 variables:

date Date of report (class `Date`)

cases Integer vector indicating the number of confirmed cases reported on each date

Details

The dataset name has been kept as `us_covid_cases_df` to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix `_df` indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the cpr package version 0.4.0

`view_datasets_infectiousR` *View Available Datasets in infectiousR*

Description

This function lists all datasets available in the 'infectiousR' package. If the 'infectiousR' package is not loaded, it stops and shows an error message. If no datasets are available, it returns a message and an empty vector.

Usage

```
view_datasets_infectiousR()
```

Value

A character vector with the names of the available datasets. If no datasets are found, it returns an empty character vector.

Examples

```
if (requireNamespace("infectiousR", quietly = TRUE)) {
  library(infectiousR)
  view_datasets_infectiousR()
}
```

`zika_girardot_df` *Zika in Girardot, Colombia, 2015*

Description

This dataset, zika_girardot_df, is a data frame containing the daily incidence of Zika virus disease in Girardot, Colombia, during 2015.

Usage

```
data(zika_girardot_df)
```

Format

A data frame with 93 observations and 2 variables:

date Date object representing the date of reported Zika cases

cases Integer vector indicating the number of daily reported Zika cases

Details

The dataset name has been kept as 'zika_girardot_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the outbreaks package version 1.9.0

`zika_sanandres_df` *Zika in San Andres, Colombia, 2015*

Description

This dataset, zika_sanandres_df, is a data frame containing the daily incidence of Zika virus disease in San Andres, Colombia, during 2015.

Usage

```
data(zika_sanandres_df)
```

Format

A data frame with 101 observations and 2 variables:

date Date object representing the date of reported Zika cases

cases Integer vector indicating the number of daily reported Zika cases

Details

The dataset name has been kept as 'zika_sanandres_df' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the infectiousR package and assists users in identifying its specific characteristics. The suffix 'df' indicates that the dataset is a data frame. The original content has not been modified in any way.

Source

Data taken from the outbreaks package version 1.9.0

Index