

Package ‘datazoom.saude’

October 6, 2026

Title Download Brazilian Health Data from DATASUS

Version 0.1.1

Description Provides a comprehensive set of functions to easily download, clean, and standardize various public health datasets from DATASUS <<https://datasus.saude.gov.br/>>, the Department of Informatics of the Brazilian Unified Health System (SUS). This package streamlines access to crucial health information, including mortality (SIM), hospital admissions (SIH), live births (SINASC), hospital beds (CNES-LT), and outpatient procedures (SIA-SUS), making the data ready for epidemiological and public health analyses.

License MIT + file LICENSE

URL https://datazoom.com.br/en/dz_saude/

Depends R (>= 4.0)

Imports data.table, dplyr, Hmisc, janitor, magrittr, purrr, readr, readxl, sf, stringr, tibble, utils, XML, lubridate, tidyselect, jsonlite, tidyr, stats

Suggests foreign, googledrive, knitr, RCurl, rmarkdown, terra, testthat (>= 3.0.0), units

LinkingTo Rcpp

VignetteBuilder knitr

Encoding UTF-8

Config/testthat/edition 3

Config/roxygen2/version 8.0.0

BugReports <https://github.com/datazoompuc/datazoom.saude/issues>

NeedsCompilation yes

Author Laura Tavares Regadas [aut, cre],
Luana Laport Santos [aut],
João Paulo dos Santos P. Barbosa [aut],
DataZoom (PUC-Rio) [fnd],
Vanndher Amós B. A. de Souza [aut],
Daniela Petruzalek [ctb],
Mark Adler [ctb]

Maintainer Laura Tavares Regadas <lauratregadas@gmail.com>
Repository CRAN
Date/Publication 2026-10-06 13:40:02 UTC

Contents

dbc2dbf_wrapper	2
load_births	3
load_hospital_admissions	4
load_hospital_beds	6
load_mortality	7
load_oncology_case	8
load_outpatient_procedures	9
load_vaccines	11
Index	13

dbc2dbf_wrapper	<i>Decompress a DBC (compressed DBF) file</i>
-----------------	---

Description

This function allows you decompress a DBC file into its DBF counterpart. Please note that this is the file format used by the Brazilian Ministry of Health (DATASUS), and it is not related to the FoxPro or CANDb DBC file formats.

Usage

```
dbc2dbf_wrapper(input.file, output.file)
```

Arguments

- input.file The name of the DBC file (including extension)
- output.file The output file name (including extension)

Details

DBC is the extension for compressed DBF files (from the 'XBASE' family of databases). This is a proprietary file format used by the brazilian government to make available public healthcare datasets (by it's agency called DATASUS).

It uses internally the PKWare's Data Compression Library (DCL) "implode" compression algorithm. When decompressed, it becomes a regular DBF file.

Value

Return TRUE if succeeded, FALSE otherwise.

Author(s)

Daniela Petruzalek, <daniela.petruszalek@gmail.com>

Source

The internal C code for dbc2dbf is based on blast decompressor and blast-dbf (see *References*).

References

The PKWare ZIP file format documentation (contains the "implode" algorithm specification) available at <https://pkware.my.site.com/s/>, current version <https://pkware.cachefly.net/webdocs/casestudies/APPNOTE.TXT>.

blast source code in C: <https://github.com/madler/zlib/tree/master/contrib/blast>

blast-dbf, DBC to DBF command-line decompression tool: <https://github.com/eaglebh/blast-dbf>

See Also

read.dbc

Examples

```
in.f <- system.file("files/sids.dbc", package = "datazoom.saude")
out.f <- tempfile(fileext = ".dbf")
if (file.exists(in.f)) {
  dbc2dbf_wrapper(input.file = in.f, output.file = out.f)
}
```

load_births

Loads birth data of newborns from Brazil provided by DATASUS

Description

This function downloads and organizes data from SINASC (Live Birth Information System), part of DATASUS, used in public health analyses.

Usage

```
load_births(time_period, states = "all", raw_data = FALSE, language = "eng")
```

Arguments

time_period	A numeric value or vector indicating the year(s) of the data to be downloaded. For example, 2020 or 2015:2020.
states	A string or vector of strings indicating the Brazilian state(s) for which the data should be downloaded. Use "all" (by default) to download data for the entire country. For specific states, use abbreviations like "SP", "RJ", or c("SP", "RJ").
raw_data	Logical. If TRUE, returns the raw data exactly as provided by DATASUS. If FALSE (default), returns a cleaned and standardized version of the dataset.
language	A string indicating the desired language of variable names and labels. Accepts "eng" (default) for English or "pt" for Portuguese.

Value

A data frame containing birth records from SINASC for the specified period and states.

Examples

```
load_births(time_period = 2023,
            states = "RJ",
            raw_data = FALSE)
```

```
load_hospital_admissions
```

Loads hospital admission data from Brazil provided by DATASUS (SIHSUS)

Description

This function downloads and organizes data from SIHSUS (Hospital Information System), part of DATASUS, used in public health and hospital care analyses.

Usage

```
load_hospital_admissions(
  dataset,
  time_period,
  states = "all",
  raw_data = FALSE,
  language = "eng"
)
```

Arguments

dataset	A string indicating the type of SIHSUS dataset to download. Accepted values are: "reduced_aih", "professional_services", "rejected_aih", or "rejected_aih_error". See the 'Details' section for explanations.
time_period	A numeric value or vector indicating the year(s) of the data to be downloaded. For example, 2020 or 2015:2020.
states	A string or vector of strings indicating the Brazilian state(s) for which the data should be downloaded. Use "all" to download data for the entire country. For specific states, use abbreviations like "SP", "RJ", or c("SP", "RJ").
raw_data	Logical. If TRUE, returns the raw data exactly as provided by DATASUS. If FALSE (default), returns a cleaned and standardized version of the dataset.
language	A string indicating the desired language of variable names and labels. Accepts "eng" (default) for English or "pt" for Portuguese.

Details

SIHSUS provides several datasets related to hospital admissions in Brazil:

reduced_aih (RD) - Reduced AIH (Hospital Admission Authorization) A simplified database with the main information from approved and processed AIHs. It is the most commonly used dataset for statistical and epidemiological analyses, including data on the main procedure, diagnoses, and total values of each admission.

professional_services (SP) - Professional Services A stratified dataset containing details about services provided during the hospital stay, such as medical procedures, professional identification (CBO/CNS), and values related to professional and hospital services.

rejected_aih (RJ) - Rejected AIHs Contains rejected AIHs and summarizes the reasons for rejection. It is useful for analyzing the volume and impact of rejected records but does not include detailed information on each rejection.

rejected_aih_error (ER) - Rejected AIHs with Error Code Includes AIHs rejected due to inconsistencies identified during processing. These records contain specific error codes indicating why the rejection occurred (e.g., patient data inconsistency, procedure incompatibility).

Value

A data frame containing hospital admission records from SIHSUS for the specified period and states.

Examples

```
load_hospital_admissions(dataset = "professional_services",
  time_period = 2020,
  states = "AC")
```

load_hospital_beds	<i>Load monthly data on hospital beds from Brazil's National Health Facilities Registry</i>
--------------------	---

Description

Retrieves and processes structured microdata on hospital beds from the 'CNES' (Cadastro Nacional de Estabelecimentos de Saude), which is part of the 'DATASUS' system, maintained by the Brazilian Ministry of Health. This dataset includes information on the availability and distribution of hospital beds in Brazil.

Usage

```
load_hospital_beds(
  time_period,
  states = "all",
  raw_data = FALSE,
  language = "eng"
)
```

Arguments

time_period	A character vector of years (e.g., c("2019", "2020")) for which data should be loaded.
states	A character vector of state abbreviations (e.g., c("RJ", "SP")) to filter the data. Use "all" to include all states. Defaults to "all".
raw_data	Logical. If TRUE, returns a list of raw data frames (one per file). If FALSE, returns a cleaned and labeled data frame. Default is FALSE.
language	Character string indicating the variable label language. Options are "pt" for Portuguese or "eng" for English. Default is "eng".

Details

The data is sourced from the 'CNES-LT' subsystem (Leitos - Beds), which provides monthly information on beds by establishment and type, and can be filtered by state and time period.

File downloads are performed via the official DATASUS FTP server. Users can choose to retrieve raw or pre-processed data, and label variables in English or Portuguese.

This function downloads and processes data from the CNES-LT subsystem, which is part of the Brazilian National Health Facilities Registry. It uses the DATASUS FTP server and handles compressed data in DBF format, using packages such as `foreign` and `Rcurl`.

The function labels variables using a built-in dictionary and offers harmonized column names depending on the selected language. A filter by state and year is applied to reduce file size and processing time.

Value

A tibble (if `raw_data = FALSE`) containing harmonized monthly hospital bed data from Brazil, or a list of raw data frames (if `raw_data = TRUE`).

Note

The `foreign` and `Rcurl` packages must be installed to use this function. They are listed in `Suggests` to avoid installing them by default.

Examples

```
# Load beds data for R0 in 2020
beds <- load_hospital_beds(time_period = "2020",
                           states = "R0",
                           raw_data = FALSE,
                           language = "eng")
```

load_mortality

Load Brazilian mortality data from the SIM system via DATASUS

Description

Retrieves mortality records from Brazil's official Mortality Information System (SIM), made available through DATASUS (Department of Informatics of the Brazilian Unified Health System). This dataset includes detailed information on deaths by cause, location, sex, age, and other demographic characteristics. Data is useful for public health research, epidemiology, and demographic analysis.

Usage

```
load_mortality(
  dataset,
  time_period,
  states = "all",
  raw_data = FALSE,
  keep_all = FALSE,
  language = "eng"
)
```

Arguments

`dataset` A string identifying the specific mortality dataset to download. Accepted values are:

- "general": General death records
- "fetal": Fetal deaths
- "external_causes": Deaths by external causes

	• "infant": Infant deaths • "maternal": Maternal deaths
time_period	A numeric value or vector indicating the year(s) of the data to be downloaded. For example, 2020 or 2015:2020.
states	A string or vector of strings indicating the Brazilian state(s) for which the data should be downloaded. Use "all" to download data for the entire country. For specific states, use abbreviations like "SP", "RJ", or c("SP", "RJ").
raw_data	Logical. If TRUE, returns the raw data exactly as provided by DATASUS. If FALSE (default), returns a cleaned and standardized version of the dataset.
keep_all	A boolean choosing whether to aggregate the data by municipality, losing individual-level variables (FALSE) or to keep all original variables (TRUE). Only applies when raw_data is FALSE.
language	A string indicating the desired language of variable names and labels. Accepts "eng" (default) for English or "pt" for Portuguese.

Value

A data frame containing the mortality records.

Examples

```
load_mortality(dataset = "general",
               time_period = 2022,
               states = "RJ")

load_mortality(dataset = "maternal",
               time_period = 2020,
               raw_data = FALSE,
               language = "pt")
```

load_oncology_case	<i>Loads oncology case data from Brazil provided by DATASUS</i>
--------------------	---

Description

This function downloads and organizes data from the Oncology Panel (Painel de Oncologia), part of DATASUS, used in public health and epidemiological analyses.

Usage

```
load_oncology_case(time_period, raw_data = FALSE, language = "eng")
```

Arguments

time_period	A numeric value or vector indicating the year(s) of the data to be downloaded. For example, 2020 or 2015:2020.
raw_data	Logical. If TRUE, returns the raw data exactly as provided by DATASUS. If FALSE (default), returns a cleaned and standardized version of the dataset.
language	A string indicating the desired language of variable names and labels. Accepts "eng" (default) for English or "pt" for Portuguese.

Value

A data table containing oncology case records from the Oncology Panel of all Brazilian states for the specified period.

Examples

```
# Download processed oncology data for the year 2023.
# This will return data from the Oncology Panel for all Brazilian states.
oncology_cases_treated <- load_oncology_case(time_period = 2023,
                                             raw_data = FALSE,
                                             language = "eng"
                                             )
```

load_outpatient_procedures

Loads outpatient procedure data from Brazil provided by 'SIASUS' (Ambulatory Information System)

Description

This function downloads and organizes public health data from 'SIASUS' – the Ambulatory Information System from Brazil's national health database, 'DATASUS'. It contains records of outpatient medical procedures performed throughout the country.

Usage

```
load_outpatient_procedures(
  dataset,
  time_period,
  states = "all",
  raw_data = FALSE,
  language = "eng"
)
```

Arguments

dataset	A string indicating the type of 'SIASUS' dataset to download. Accepted values include: "bariatric_surgery", "diverse_reports", "medicines", "nephrology", "ambulatory_production", "psychosocial", "bariatric_surgery_follow_up", "fistula_confection", "dialytic_treatment", "home_care". See the 'Details' section for descriptions.
time_period	A numeric value or vector indicating the year(s) of the data to be downloaded. For example, 2020 or 2015:2020.
states	A string or vector of strings indicating the Brazilian state(s) for which the data should be downloaded. Use "all" to download data for the entire country. For specific states, use abbreviations like "SP", "RJ", or c("SP", "RJ").
raw_data	Logical. If TRUE, returns the raw data exactly as provided by DATASUS. If FALSE (default), returns a cleaned and standardized version of the dataset.
language	A string indicating the desired language of variable names and labels. Accepts "eng" (default) for English or "pt" for Portuguese.

Details

SIASUS provides multiple datasets that cover different aspects of outpatient care in Brazil:

bariatric_surgery_follow_up General preventive and primary care procedures, such as check-ups and vaccinations.

diverse_reports Outpatient dental procedures including extractions, restorations, and preventive services.

medicines Higher complexity procedures including specialized consultations and diagnostic exams.

nephrology Procedures related to early-life screening tests, including metabolic and sensory testing.

chemotherapy Outpatient procedures provided through structured home care programs.

radiotherapy Procedures focused on physical, cognitive, and functional rehabilitation.

ambulatory_production High-cost procedures that require prior authorization, such as cancer treatment or dialysis.

psychosocial Aggregated or simplified records of outpatient procedures with limited detail.

post_bariatric_surgery_follow_up Records from specialized dental centers providing oral health care.

fistula_confection Data on medication dispensing and pharmaceutical consultations.

dialytic_treatment Procedures related to therapies such as chemotherapy and radiotherapy.

home_care Home-based specialized care such as oxygen therapy and related services.

Value

A data frame containing outpatient procedure records from SIASUS for the specified period and states.

Examples

```
load_outpatient_procedures(dataset = "diverse_reports",
                           time_period = 2016,
                           states = "PI",
                           language = "pt")
```

load_vaccines

Load Brazilian vaccination data from the SI-PNI system via DATASUS

Description

Retrieves vaccination records from Brazil's official National Immunization Program Information System (SI-PNI), made available through DATASUS. This function supports two data ingestion modes. For historical data (1994-2022), it can perform automated web scraping on the legacy SI-PNI Web portal to extract consolidated records of applied doses. For more recent years (2023 to present), the user must provide a locally downloaded file, in which case the function performs only data validation, cleaning, harmonization, and standardization.

Usage

```
load_vaccines(
  year,
  state,
  strategy = NULL,
  product = NULL,
  dose = NULL,
  data = NULL,
  language = "eng"
)
```

Arguments

year	A numeric value indicating the year of the data to be downloaded (Supported range: 1994-2022).
state	A string indicating the Brazilian state abbreviation (e.g., "SP", "RJ", "AC").
strategy	A string identifying the vaccination strategy (e.g., "Rotina", "Especial", "Bloqueio"). If NULL and in an interactive session, a menu will appear.
product	A string identifying the specific vaccine product (e.g., "BCG - BCG", "Hepatite B - HB"). Must be a valid product for the chosen strategy. If NULL and in an interactive session, a menu will appear.
dose	A character vector indicating the dose categories that were selected when downloading the data. This argument is only required when using manually downloaded files up to 2022.

data	An optional path to a local Excel (.xlsx) file downloaded manually from the official DATASUS vaccination dashboard. If provided, web scraping is skipped and the function performs only data cleaning and harmonization. This argument is mandatory for data from 2023 onwards.
language	A string indicating the desired language for variable names and labels. Accepts "eng" (default) or "pt".

Details

Regardless of the ingestion mode, the returned dataset is fully harmonized and consistent with the historical SI-PNI data structure.

Value

A tibble (data frame) containing state, year, municipality codes, municipality names, strategy, product, month, dose type, and the quantity of doses applied.

Examples

```
# Example: Loading Yellow Fever vaccine data for Acre in 2020
data <- load_vaccines(year = 2020,
                      state = "AC",
                      strategy = "Rotina",
                      product = "Febre amarela - FA",
                      language = "eng")
```

Index

- * **dbc**
 - dbc2dbf_wrapper, [2](#)
- * **dbf**
 - dbc2dbf_wrapper, [2](#)
- dbc2dbf_wrapper, [2](#)
- load_births, [3](#)
- load_hospital_admissions, [4](#)
- load_hospital_beds, [6](#)
- load_mortality, [7](#)
- load_oncology_case, [8](#)
- load_outpatient_procedures, [9](#)
- load_vaccines, [11](#)