Package 'healthbR'

February 3, 2026

Title Access Brazilian Public Health Data

Version 0.1.0

Description Provides easy access to Brazilian public health data from multiple sources including VIGITEL (Surveillance of Risk Factors for Chronic Diseases by Telephone Survey), PNS (National Health Survey), SIM (Mortality Information System), SINASC (Live Birth Information System), and other health information systems. Data is downloaded from the Brazilian Ministry of Health VIGITEL repository <https://svs.aids.gov.br/download/Vigitel/>. Data is returned in tidy format following tidyverse conventions.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.3

Depends R (>= 4.2.0)

Imports tibble, dplyr, readxl, curl, cli, rlang, stringr, janitor, arrow, purrr

Suggests testthat (>= 3.0.0), knitr, rmarkdown, furrr, future

Config/testthat/edition 3

VignetteBuilder knitr

URL https://github.com/SidneyBissoli/healthbR

BugReports https://github.com/SidneyBissoli/healthbR/issues

NeedsCompilation no

Author Sidney Bissoli [aut, cre] (ORCID: <https://orcid.org/0009-0001-0442-3700>)

Maintainer Sidney Bissoli <sbissoli76@gmail.com>

Repository CRAN

Date/Publication 2026-02-03 10:40:07 UTC

Contents

`list_sources` *List Available Data Sources*

Description

Returns information about all data sources available in healthbR.

Usage

```
list_sources()
```

Value

A tibble with columns:

- `source`: Source code (e.g., "vigitel", "sim")
- `name`: Full name of the data source
- `description`: Brief description
- `years`: Range of available years
- `status`: Implementation status ("available", "planned")

Examples

```
list_sources()
```

`utils` *Utility Functions for healthbR*

Description

Utility Functions for healthbR

vigitel_cache_status *Get VIGITEL cache status*

Description

Shows which years are cached and file sizes.

Usage

```
vigitel_cache_status()
```

Value

A tibble with cache information

Examples

```
vigitel_cache_status()
```

vigitel_clear_cache *Clear VIGITEL cache*

Description

Removes all cached VIGITEL data files (Excel and Parquet).

Usage

```
vigitel_clear_cache(keep_parquet = FALSE)
```

Arguments

keep_parquet	Logical. If TRUE, keep Parquet files and only remove Excel files. Default is FALSE (remove all).

Value

NULL (invisibly)

Examples

```
# remove all cached files
vigitel_clear_cache()

# remove only Excel files, keep Parquet
vigitel_clear_cache(keep_parquet = TRUE)
```

`vigitel_data` *Load VIGITEL microdata*

Description

Downloads (if necessary) and loads VIGITEL survey microdata into R. Data is automatically converted to Parquet format for faster subsequent loading. The data includes survey weights for proper statistical analysis.

Usage

```
vigitel_data(
  year,
  vars = NULL,
  force_download = FALSE,
  parallel = TRUE,
  lazy = FALSE
)
```

Arguments

`year`	Year(s) of the survey. Can be: • Single year: `2023` • Range: `2021:2023` • Vector: `c(2021, 2023)` • Character: `c("2021", "2023")` • All years: `"all"`
`vars`	Character vector. Variable names to select, or NULL for all variables. Default is NULL.
`force_download`	Logical. If TRUE, re-download and reconvert data. Default is FALSE.
`parallel`	Logical. If TRUE, download and process multiple years in parallel. Default is TRUE when multiple years are requested.
`lazy`	Logical. If TRUE, return an Arrow Dataset for lazy evaluation instead of loading all data into memory. Useful for filtering large datasets before collecting. Use `collect()` to retrieve results. Default is FALSE.

Details

On first access, data is downloaded from the Ministry of Health and converted to Parquet format. Subsequent loads read directly from the Parquet file, which is significantly faster.

For parallel downloads, the function uses the `furrr` and `future` packages if installed. Install them with `install.packages(c("furrr", "future"))` to enable parallel processing. The number of workers is automatically set based on available CPU cores. If these packages are not installed, processing falls back to sequential mode.

When `lazy = TRUE`, the function returns an Arrow Dataset that supports dplyr operations (filter, select, mutate, etc.) without loading data into memory. This is useful for working with large datasets or when you only need a subset of the data. Call `collect()` to retrieve the results as a tibble.

The VIGITEL survey uses complex sampling weights. For proper statistical analysis, use survey packages like `survey` or `srvyr`. The weight variable is named `pesorake`.

Value

A tibble with the VIGITEL microdata. When multiple years are requested, a `year` column is added to identify the source year. If `lazy = TRUE`, returns an Arrow Dataset that can be queried with dplyr verbs before calling `collect()`.

Examples

```
# single year
df <- vigitel_data(2023)

# multiple years
df <- vigitel_data(2021:2023)
df <- vigitel_data(c(2018, 2020, 2023))

# all available years
df <- vigitel_data("all")

# specific variables
df <- vigitel_data(2023, vars = c("cidade", "sexo", "idade", "pesorake"))

# multiple years with specific variables
df <- vigitel_data(2020:2023, vars = c("cidade", "sexo", "idade", "pesorake"))

# lazy evaluation - filter before loading into memory
vigitel_data(2023, lazy = TRUE) |>
  dplyr::filter(cidade == 1) |>
  dplyr::select(pesorake) |>
  dplyr::collect()

# lazy with multiple years
vigitel_data(2020:2023, lazy = TRUE) |>
  dplyr::filter(q6 == 1) |>
  dplyr::collect()
```

---

vigitel_dictionary *Get VIGITEL variable dictionary*

Description

Returns the data dictionary with variable descriptions, labels, and coding information for VIGITEL surveys.

Usage

```
vigitel_dictionary(force_download = FALSE)
```

Arguments

force_download	Logical. If TRUE, re-download the dictionary.

Value

A tibble with variable metadata

Examples

```
# get the dictionary
dict <- vigitel_dictionary()

# view column names
names(dict)
```

---

vigitel_download *Download VIGITEL microdata for a specific year*

Description

Downloads the VIGITEL survey microdata file from the Ministry of Health website. Files are cached locally to avoid repeated downloads.

Usage

```
vigitel_download(year, force = FALSE)
```

Arguments

year	Integer. Year of the survey (use `vigitel_years()` to see available years).
force	Logical. If TRUE, re-download even if file exists in cache. Default is FALSE.

Value

Path to the downloaded file (invisibly)

Examples

```
# download 2023 data
vigitel_download(2023)

# force re-download
vigitel_download(2023, force = TRUE)
```

vigitel_info *Get VIGITEL survey information*

Description

Returns metadata about the VIGITEL survey.

Usage

```
vigitel_info()
```

Value

A list with survey information

Examples

```
vigitel_info()
```

vigitel_variables *List VIGITEL variables*

Description

Returns a character vector of variable names available in a VIGITEL survey year.

Usage

```
vigitel_variables(year)
```

Arguments

year	Integer. Year of the survey.

Value

A character vector of variable names

Examples

```
# list variables for 2023
vigitel_variables(2023)
```

`vigitel_years` *List available VIGITEL survey years*

Description

Returns a vector of years for which VIGITEL microdata is available for download from the Ministry of Health website.

Usage

```
vigitel_years()
```

Value

An integer vector of available years

Examples

```
vigitel_years()
```

Index