

# Package ‘healthbR’

February 4, 2026

**Title** Access Brazilian Public Health Data

**Version** 0.1.1

**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,  
purrr

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

**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** 2026-02-04 08:20:36 UTC

Contents

list_sources . . . . .	2
utils . . . . .	2
vigitel_cache_status . . . . .	3
vigitel_clear_cache . . . . .	3
vigitel_data . . . . .	4
vigitel_dictionary . . . . .	5
vigitel_download . . . . .	6
vigitel_info . . . . .	7
vigitel_variables . . . . .	7
vigitel_years . . . . .	8
<b>Index</b>	<b>9</b>

---

list_sources	<i>List Available Data Sources</i>
--------------	------------------------------------

---

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	<i>Utility Functions for healthbR</i>
-------	---------------------------------------

---

Description

Utility Functions for healthbR

---

vigitel_cache_status	<i>Get VIGITEL cache status</i>
----------------------	---------------------------------

---

**Description**

Shows which years are cached and file sizes.

**Usage**

```
vigitel_cache_status(cache_dir = NULL)
```

**Arguments**

cache_dir	Character. Optional custom cache directory. If NULL (default), uses the standard user cache directory.
-----------	--

**Value**

A tibble with cache information

**Examples**

```
# check cache status
vigitel_cache_status()
```

---

vigitel_clear_cache	<i>Clear VIGITEL cache</i>
---------------------	----------------------------

---

**Description**

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

**Usage**

```
vigitel_clear_cache(keep_parquet = FALSE, cache_dir = NULL)
```

**Arguments**

keep_parquet	Logical. If TRUE, keep Parquet files and only remove Excel files. Default is FALSE (remove all).
cache_dir	Character. Optional custom cache directory. If NULL (default), uses the standard user cache directory.

**Value**

NULL (invisibly)

## Examples

```
# remove all cached files from default cache
vigitel_clear_cache()
```

---

vigitel_data	<i>Load VIGITEL microdata</i>
--------------	-------------------------------

---

## 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,
  cache_dir = NULL
)
```

## Arguments

year	Year(s) of the survey. Can be: <ul style="list-style-type: none"> <li>• Single year: 2023</li> <li>• Range: 2021:2023</li> <li>• Vector: c(2021, 2023)</li> <li>• Character: c("2021", "2023")</li> <li>• All years: "all"</li> </ul>
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.
cache_dir	Character. Optional custom cache directory. If NULL (default), uses the standard user cache directory. Use tempdir() for temporary storage that won't persist.

## 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.

The arrow package is required for Parquet file support. If not installed, an informative error message will be shown with installation instructions.

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 (uses tempdir to avoid leaving files on system)
df <- vigitel_data(2023, cache_dir = tempdir())

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

---

<code>vigitel_dictionary</code>	<i>Get VIGITEL variable dictionary</i>
---------------------------------	--

---

## Description

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

## Usage

```
vigitel_dictionary(force_download = FALSE, cache_dir = NULL)
```

**Arguments**

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

`cache_dir` Character. Optional custom cache directory. If NULL (default), uses the standard user cache directory. Use `tempdir()` for temporary storage that won't persist.

**Value**

A tibble with variable metadata

**Examples**

```
# get the dictionary (uses tempdir to avoid leaving files)
dict <- vigitel_dictionary(cache_dir = tempdir())

# view column names
names(dict)
```

---

<code>vigitel_download</code>	<i>Download VIGITEL microdata for a specific year</i>
-------------------------------	---

---

**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, cache_dir = NULL)
```

**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.

`cache_dir` Character. Optional custom cache directory. If NULL (default), uses the standard user cache directory. Use `tempdir()` for temporary storage that won't persist.

**Value**

Path to the downloaded file (invisibly)

**Examples**

```
# download 2023 data (uses tempdir to avoid leaving files)
vigitel_download(2023, cache_dir = tempdir())
```

---

vigitel_info	<i>Get VIGITEL survey information</i>
--------------	---------------------------------------

---

**Description**

Returns metadata about the VIGITEL survey.

**Usage**

```
vigitel_info()
```

**Value**

A list with survey information

**Examples**

```
vigitel_info()
```

---

vigitel_variables	<i>List VIGITEL variables</i>
-------------------	-------------------------------

---

**Description**

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

**Usage**

```
vigitel_variables(year, cache_dir = NULL)
```

**Arguments**

year	Integer. Year of the survey.
cache_dir	Character. Optional custom cache directory. If NULL (default), uses the standard user cache directory. Use <code>tempdir()</code> for temporary storage that won't persist.

**Value**

A character vector of variable names

**Examples**

```
# list variables for 2023 (uses tempdir to avoid leaving files)
vigitel_variables(2023, cache_dir = tempdir())
```

---

vigitel_years	<i>List available VIGITEL survey years</i>
---------------	--

---

**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

`list_sources`, [2](#)

`utils`, [2](#)

`vigitel_cache_status`, [3](#)

`vigitel_clear_cache`, [3](#)

`vigitel_data`, [4](#)

`vigitel_dictionary`, [5](#)

`vigitel_download`, [6](#)

`vigitel_info`, [7](#)

`vigitel_variables`, [7](#)

`vigitel_years`, [8](#)