

Package ‘censobr’

July 7, 2025

Type Package

Title Download Data from Brazil's Population Census

Version 0.5.0

Description Easy access to data from Brazil's population censuses. The package provides a simple and efficient way to download and read the data sets and the documentation of all the population censuses taken in and after 1960 in the country. The package is built on top of the 'Arrow' platform <<https://arrow.apache.org/docs/r/>>, which allows users to work with larger-than-memory census data using 'dplyr' familiar functions. <<https://arrow.apache.org/docs/r/articles/arrow.html#analyzing-arrow-data-with-dplyr>>.

License MIT + file LICENSE

URL <https://github.com/ipeaGIT/censobr>,
<https://ipeagit.github.io/censobr/>

BugReports <https://github.com/ipeaGIT/censobr/issues>

Depends R (>= 4.1.0)

Imports arrow (>= 15.0.1), checkmate, cli, curl (>= 5.0.0), dplyr,
duckdb, fs, glue, rlang, tools

Suggests covr, DBI, dbplyr, geobr, ggplot2 (>= 3.3.1), rmarkdown,
kableExtra, knitr, scales, testthat

VignetteBuilder knitr

Encoding UTF-8

RoxygenNote 7.3.2

NeedsCompilation no

Author Rafael H. M. Pereira [aut, cre] (ORCID:
<<https://orcid.org/0000-0003-2125-7465>>),
Rogério J. Barbosa [aut] (ORCID:
<<https://orcid.org/0000-0002-6796-4547>>),
Diego Rabatone Oliveira [ctb],
Neal Richardson [ctb],
Ipea - Institute for Applied Economic Research [cph, fnd]

Maintainer Rafael H. M. Pereira <rafa.pereira.br@gmail.com>
Repository CRAN
Date/Publication 2025-07-07 02:30:02 UTC

Contents

censobr_cache	2
data_dictionary	3
get_censobr_cache_dir	4
interview_manual	5
questionnaire	6
read_emigration	7
read_families	8
read_households	10
read_mortality	11
read_population	13
read_tracts	15
set_censobr_cache_dir	16
Index	18

censobr_cache	<i>Manage cached files from the censobr package</i>
---------------	---

Description

Manage cached files from the censobr package

Usage

```
censobr_cache(  
  list_files = TRUE,  
  print_tree = FALSE,  
  delete_file = NULL,  
  verbose = TRUE  
)
```

Arguments

- | | |
|-------------|--|
| list_files | Logical. Whether to print a message with the address of all censobr data sets cached locally. Defaults to TRUE. |
| print_tree | Logical. Whether the cache files should be printed in a tree-like format. This parameter only works if list_files = TRUE. Defaults to FALSE. |
| delete_file | String. The file name or a string pattern that matches the file path of a file cached locally and which should be deleted. Defaults to NULL, so that no file is deleted. If delete_file = "all", then all of the cached files are deleted. |

verbose A logical. Whether the function should print informative messages. Defaults to TRUE.

Value

A message indicating which file exist and/or which ones have been deleted from the local cache directory.

See Also

Other Cache data: [get_censobr_cache_dir\(\)](#), [set_censobr_cache_dir\(\)](#)

Examples

```
# list all files cached
censobr_cache(list_files = TRUE)

# delete particular file
censobr_cache(delete_file = '2010_deaths')
```

data_dictionary	<i>Data dictionary of Brazil's census data</i>
-----------------	--

Description

Open on a browser the data dictionary of Brazil's census data.

Usage

```
data_dictionary(
  year,
  dataset,
  showProgress = TRUE,
  cache = TRUE,
  verbose = TRUE
)
```

Arguments

year	Numeric. Year of reference in the format yyyy.
dataset	Character. The dataset of data dictionary to be opened. Options include c("population", "households", "families", "mortality", "emigration", "tracts").
showProgress	Logical. Defaults to TRUE display download progress bar. The progress bar only reflects only the downloading time, not the time to load the data to memory.

cache	Logical. Whether the function should read the data cached locally, which is much faster. Defaults to TRUE. The first time the user runs the function, censobr will download the file and store it locally so that the file only needs to be download once. If FALSE, the function will download the data again and overwrite the local file.
verbose	A logical. Whether the function should print informative messages. Defaults to TRUE.

Value

Returns NULL and opens an .html, .pdf or excel file

See Also

Other Census documentation: [interview_manual\(\)](#)

Examples

```
# Open data dictionary
data_dictionary(year = 2010,
               dataset = 'population',
               showProgress = FALSE)

data_dictionary(year = 2022,
               dataset = 'tracts',
               showProgress = FALSE)

data_dictionary(year = 1980,
               dataset = 'households',
               showProgress = FALSE)
```

get_censobr_cache_dir *Get path to cache directory for censobr files*

Description

Get the path to the cache directory currently being used for for the censobr files

Usage

```
get_censobr_cache_dir()
```

Value

Path to cache dir

See Also

Other Cache data: [censobr_cache\(\)](#), [set_censobr_cache_dir\(\)](#)

Examples

```
# get path to cache directory
get_censobr_cache_dir()
```

interview_manual	<i>Interview manual of the data collection of Brazil's censuses</i>
------------------	---

Description

Open on a browser the interview manual of the data collection of Brazil's censuses

Usage

```
interview_manual(
  year = NULL,
  showProgress = TRUE,
  cache = TRUE,
  verbose = TRUE
)
```

Arguments

year	Numeric. Year of reference in the format yyyy.
showProgress	Logical. Defaults to TRUE display download progress bar. The progress bar only reflects only the downloading time, not the time to load the data to memory.
cache	Logical. Whether the function should read the data cached locally, which is much faster. Defaults to TRUE. The first time the user runs the function, censobr will download the file and store it locally so that the file only needs to be download once. If FALSE, the function will download the data again and overwrite the local file.
verbose	A logical. Whether the function should print informative messages. Defaults to TRUE.

Value

Opens a .pdf file on the browser

See Also

Other Census documentation: [data_dictionary\(\)](#)

Examples

```
# Open interview manual on the browser
interview_manual(
  year = 2010,
  showProgress = FALSE
)
```

questionnaire

Questionnaires used in the data collection of Brazil's censuses

Description

Open on a browser the questionnaire used in the data collection of Brazil's censuses

Usage

```
questionnaire(
  year = 2010,
  type = NULL,
  showProgress = TRUE,
  cache = TRUE,
  verbose = TRUE
)
```

Arguments

year	Numeric. Year of reference in the format yyyy.
type	Character. The type of questionnaire used in the survey, whether the "long" one used in the sample component of the census, or the "short" one, which is answered by more households. Options include <code>c("long", "short")</code> .
showProgress	Logical. Defaults to TRUE display download progress bar. The progress bar only reflects only the downloading time, not the time to load the data to memory.
cache	Logical. Whether the function should read the data cached locally, which is much faster. Defaults to TRUE. The first time the user runs the function, censobr will download the file and store it locally so that the file only needs to be download once. If FALSE, the function will download the data again and overwrite the local file.
verbose	A logical. Whether the function should print informative messages. Defaults to TRUE.

Value

Opens a .pdf file on the browser

Examples

```
library(censobr)

# Open questionnaire on browser
questionnaire(year = 2010, type = 'long', showProgress = FALSE)
```

read_emigration	<i>Download microdata of emigration records from Brazil's census</i>
-----------------	--

Description

Download microdata of emigration records from Brazil's census. Data collected in the sample component of the questionnaire.

Usage

```
read_emigration(
  year,
  columns = NULL,
  add_labels = NULL,
  merge_households = FALSE,
  as_data_frame = FALSE,
  showProgress = TRUE,
  cache = TRUE,
  verbose = TRUE
)
```

Arguments

year	Numeric. Year of reference in the format yyyy.
columns	String. A vector of column names to keep. The rest of the columns are not read. Defaults to NULL and read all columns.
add_labels	Character. Whether the function should add labels to the responses of categorical variables. When add_labels = "pt", the function adds labels in Portuguese. Defaults to NULL.
merge_households	Logical. Indicate whether the function should merge household variables to the output data. Defaults to FALSE.
as_data_frame	Logical. When FALSE (Default), the function returns an Arrow Dataset, which allows users to work with larger-than-memory data. If TRUE, the function returns data.frame.
showProgress	Logical. Defaults to TRUE display download progress bar. The progress bar only reflects only the downloading time, not the time to load the data to memory.

cache	Logical. Whether the function should read the data cached locally, which is much faster. Defaults to TRUE. The first time the user runs the function, censobr will download the file and store it locally so that the file only needs to be download once. If FALSE, the function will download the data again and overwrite the local file.
verbose	A logical. Whether the function should print informative messages. Defaults to TRUE.

Value

An arrow Dataset or a "data.frame" object.

See Also

Other Microdata: [read_families\(\)](#), [read_households\(\)](#), [read_mortality\(\)](#), [read_population\(\)](#)

Examples

```
# return data as arrow Dataset
df <- read_emigration(
  year = 2010,
  showProgress = FALSE
)

# return data as data.frame
df <- read_emigration(
  year = 2010,
  as_data_frame = TRUE,
  showProgress = FALSE
)
```

read_families

Download microdata of family records from Brazil's census

Description

Download microdata of family records from Brazil's census. Data collected in the sample component of the questionnaire.

Usage

```
read_families(
  year,
  columns = NULL,
  add_labels = NULL,
  as_data_frame = FALSE,
```



```

    showProgress = TRUE,
    cache = TRUE,
    verbose = TRUE
  )

```

Arguments

year	Numeric. Year of reference in the format yyyy.
columns	String. A vector of column names to keep. The rest of the columns are not read. Defaults to NULL and read all columns.
add_labels	Character. Whether the function should add labels to the responses of categorical variables. When add_labels = "pt", the function adds labels in Portuguese. Defaults to NULL.
as_data_frame	Logical. When FALSE (Default), the function returns an Arrow Dataset, which allows users to work with larger-than-memory data. If TRUE, the function returns data.frame.
showProgress	Logical. Defaults to TRUE display download progress bar. The progress bar only reflects only the downloading time, not the time to load the data to memory.
cache	Logical. Whether the function should read the data cached locally, which is much faster. Defaults to TRUE. The first time the user runs the function, censobr will download the file and store it locally so that the file only needs to be download once. If FALSE, the function will download the data again and overwrite the local file.
verbose	A logical. Whether the function should print informative messages. Defaults to TRUE.

Value

An arrow Dataset or a "data.frame" object.

See Also

Other Microdata: [read_emigration\(\)](#), [read_households\(\)](#), [read_mortality\(\)](#), [read_population\(\)](#)

Examples

```

# return data as arrow Dataset
df <- read_families(
  year = 2000,
  showProgress = FALSE
)

```

read_households	<i>Download microdata of household records from Brazil's census</i>
-----------------	---

Description

Download microdata of household records from Brazil's census. Data collected in the sample component of the questionnaire.

Usage

```
read_households(
  year,
  columns = NULL,
  add_labels = NULL,
  as_data_frame = FALSE,
  showProgress = TRUE,
  cache = TRUE,
  verbose = TRUE
)
```

Arguments

year	Numeric. Year of reference in the format yyyy.
columns	String. A vector of column names to keep. The rest of the columns are not read. Defaults to NULL and read all columns.
add_labels	Character. Whether the function should add labels to the responses of categorical variables. When <code>add_labels = "pt"</code> , the function adds labels in Portuguese. Defaults to NULL.
as_data_frame	Logical. When FALSE (Default), the function returns an Arrow Dataset, which allows users to work with larger-than-memory data. If TRUE, the function returns <code>data.frame</code> .
showProgress	Logical. Defaults to TRUE display download progress bar. The progress bar only reflects only the downloading time, not the time to load the data to memory.
cache	Logical. Whether the function should read the data cached locally, which is much faster. Defaults to TRUE. The first time the user runs the function, <code>censobr</code> will download the file and store it locally so that the file only needs to be download once. If FALSE, the function will download the data again and overwrite the local file.
verbose	A logical. Whether the function should print informative messages. Defaults to TRUE.

Value

An arrow Dataset or a "data.frame" object.

1960 Census

The 1960 microdata version available in **censobr** is a combination of two versions of the Demographic Census sample. The 25% sample data from the 1960 Census was never fully processed by IBGE - several states did not have their questionnaires digitized. Currently, this dataset only has data from 16 states of the Federation (and from a contested border region between Minas Gerais and Espírito Santo called Serra dos Aimores). Information is missing for the states of the former Northern Region, Maranhão, Piauí, Guanabara, Santa Catarina, and Espírito Santo. In 1965, IBGE decided to draw a probabilistic sub-sample of approximately 1.27% of the population, including all units of the federation. With this data, IBGE produced several official reports at the time. The data from **censobr** is the combination of these two datasets.

We pre-processed the 1.27% sample data to ensure data consistency, given the original data was partially corrupted. We also created a sample weight variable to correct for unbalanced data and to expand the sample to the total population. For the data from the 25% sample, the weights expand to the municipal totals. Meanwhile, for the data from the 1.27% sample, the weights expand to the state totals. Additionally, we constructed a few variables that allow for the approximate incorporation of the complex sample design, enabling the proper calculation of standard errors and confidence intervals.

You can read more about the 1960 Census and find a thorough documentation of how this dataset was processed on this link <https://github.com/antrologos/ConsistenciaCenso1960Br>.

See Also

Other Microdata: [read_emigration\(\)](#), [read_families\(\)](#), [read_mortality\(\)](#), [read_population\(\)](#)

Examples

```
# return data as arrow Dataset
df <- read_households(
  year = 2010,
  showProgress = FALSE
)
```

read_mortality

Download microdata of death records from Brazil's census

Description

Download microdata of death records from Brazil's census. Data collected in the sample component of the questionnaire.

Usage

```
read_mortality(
  year,
  columns = NULL,
  add_labels = NULL,
  merge_households = FALSE,
  as_data_frame = FALSE,
  showProgress = TRUE,
  cache = TRUE,
  verbose = TRUE
)
```

Arguments

year	Numeric. Year of reference in the format yyyy.
columns	String. A vector of column names to keep. The rest of the columns are not read. Defaults to NULL and read all columns.
add_labels	Character. Whether the function should add labels to the responses of categorical variables. When add_labels = "pt", the function adds labels in Portuguese. Defaults to NULL.
merge_households	Logical. Indicate whether the function should merge household variables to the output data. Defaults to FALSE.
as_data_frame	Logical. When FALSE (Default), the function returns an Arrow Dataset, which allows users to work with larger-than-memory data. If TRUE, the function returns data.frame.
showProgress	Logical. Defaults to TRUE display download progress bar. The progress bar only reflects only the downloading time, not the time to load the data to memory.
cache	Logical. Whether the function should read the data cached locally, which is much faster. Defaults to TRUE. The first time the user runs the function, censobr will download the file and store it locally so that the file only needs to be download once. If FALSE, the function will download the data again and overwrite the local file.
verbose	A logical. Whether the function should print informative messages. Defaults to TRUE.

Value

An arrow Dataset or a "data.frame" object.

See Also

Other Microdata: [read_emigration\(\)](#), [read_families\(\)](#), [read_households\(\)](#), [read_population\(\)](#)

Examples

```
library(censobr)

# return data as arrow Dataset
df <- read_mortality(
  year = 2010,
  showProgress = FALSE
)

# dplyr::glimpse(df)

# return data as data.frame
df <- read_mortality(
  year = 2010,
  as_data_frame = TRUE,
  showProgress = FALSE
)

# dplyr::glimpse(df)
```

read_population

*Download microdata of population records from Brazil's census***Description**

Download microdata of population records from Brazil's census. Data collected in the sample component of the questionnaire.

Usage

```
read_population(
  year,
  columns = NULL,
  add_labels = NULL,
  as_data_frame = FALSE,
  showProgress = TRUE,
  cache = TRUE,
  verbose = TRUE
)
```

Arguments

year	Numeric. Year of reference in the format yyyy.
columns	String. A vector of column names to keep. The rest of the columns are not read. Defaults to NULL and read all columns.

add_labels	Character. Whether the function should add labels to the responses of categorical variables. When <code>add_labels = "pt"</code> , the function adds labels in Portuguese. Defaults to <code>NULL</code> .
as_data_frame	Logical. When <code>FALSE</code> (Default), the function returns an Arrow Dataset, which allows users to work with larger-than-memory data. If <code>TRUE</code> , the function returns <code>data.frame</code> .
showProgress	Logical. Defaults to <code>TRUE</code> display download progress bar. The progress bar only reflects only the downloading time, not the time to load the data to memory.
cache	Logical. Whether the function should read the data cached locally, which is much faster. Defaults to <code>TRUE</code> . The first time the user runs the function, <code>censobr</code> will download the file and store it locally so that the file only needs to be download once. If <code>FALSE</code> , the function will download the data again and overwrite the local file.
verbose	A logical. Whether the function should print informative messages. Defaults to <code>TRUE</code> .

Value

An arrow Dataset or a "data.frame" object.

1960 Census

The 1960 microdata version available in **censobr** is a combination of two versions of the Demographic Census sample. The 25% sample data from the 1960 Census was never fully processed by IBGE - several states did not have their questionnaires digitized. Currently, this dataset only has data from 16 states of the Federation (and from a contested border region between Minas Gerais and Espírito Santo called Serra dos Aimores). Information is missing for the states of the former Northern Region, Maranhão, Piauí, Guanabara, Santa Catarina, and Espírito Santo. In 1965, IBGE decided to draw a probabilistic sub-sample of approximately 1.27% of the population, including all units of the federation. With this data, IBGE produced several official reports at the time. The data from **censobr** is the combination of these two datasets.

We pre-processed the 1.27% sample data to ensured data consistency, given the original data was partially corrupted. We also created a sample weight variable to correct for unbalanced data and to expand te sample to the total population. For the data from the 25% sample, the weights expand to the municipal totals. Meanwhile, for the data from the 1.27% sample, the weights expand to the state totals. Additionally, we constructed a few variables that allow for the approximate incorporation of the complex sample design, enabling the proper calculation of standard errors and confidence intervals.

You can read more about the 1960 Census and find a thorough documentation of how this dataset was processed on this link <https://github.com/antropologos/ConsistenciaCenso1960Br>.

See Also

Other Microdata: `read_emigration()`, `read_families()`, `read_households()`, `read_mortality()`

Examples

```
# return data as arrow Dataset
df <- read_population(
  year = 2010,
  showProgress = FALSE
)
```

read_tracts

Download census tract-level data from Brazil's censuses

Description

Download census tract-level aggregate data from Brazil's censuses.

Usage

```
read_tracts(
  year,
  dataset,
  as_data_frame = FALSE,
  showProgress = TRUE,
  cache = TRUE,
  verbose = TRUE
)
```

Arguments

year Numeric. Year of reference in the format yyyy.

dataset Character. The dataset to be opened. The following options are available for each edition of the census:

2000 Census

- c("Basico", "Domicilio", "Responsavel", "Pessoa", "Instrucao", "Morador").

2010 Census

- c("Basico", "Domicilio", "DomicilioRenda", "Responsavel", "ResponsavelRenda", "Pessoa", "PessoaRenda", "Entorno").

2022 Census

- c("Basico", "Domicilio", "ResponsavelRenda", "Pessoas", "Indigenas", "Quilombolas", "Entorno", "Obitos", "Preliminares").

The censobr package exposes all original IBGE census tracts datasets, regrouping them into broader themes and appending geographic identifiers so that they align seamlessly with geobr shapefiles.

For a complete description of the datasets, themes, and variables, check

	<ul style="list-style-type: none"> • <code>data_dictionary(year = 2000, dataset = "tracts")</code>, • <code>data_dictionary(year = 2010, dataset = "tracts")</code>, • <code>data_dictionary(year = 2022, dataset = "tracts")</code>.
<code>as_data_frame</code>	Logical. When FALSE (Default), the function returns an Arrow Dataset, which allows users to work with larger-than-memory data. If TRUE, the function returns <code>data.frame</code> .
<code>showProgress</code>	Logical. Defaults to TRUE display download progress bar. The progress bar only reflects only the downloading time, not the time to load the data to memory.
<code>cache</code>	Logical. Whether the function should read the data cached locally, which is much faster. Defaults to TRUE. The first time the user runs the function, censobr will download the file and store it locally so that the file only needs to be download once. If FALSE, the function will download the data again and overwrite the local file.
<code>verbose</code>	A logical. Whether the function should print informative messages. Defaults to TRUE.

Value

An arrow Dataset or a "data.frame" object.

Examples

```
library(censobr)

# return data as arrow Dataset
df <- read_tracts(
  year = 2022,
  dataset = 'Domicilio',
  showProgress = FALSE
)

# return data as data.frame
df <- read_tracts(
  year = 2010,
  dataset = 'Basico',
  as_data_frame = TRUE,
  showProgress = FALSE
)
```

`set_censobr_cache_dir` *Set custom cache directory for censobr files*

Description

Set custom directory for caching files from the censobr package. The user only needs to run this function once. This set directory is persistent across R sessions.

Usage

```
set_censobr_cache_dir(path, verbose = TRUE)
```

Arguments

path	String. The path to an existing directory. It defaults to path = NULL, to use the default directory
verbose	A logical. Whether the function should print informative messages. Defaults to TRUE.

Value

A message pointing to the directory where censobr files are cached.

See Also

Other Cache data: [censobr_cache\(\)](#), [get_censobr_cache_dir\(\)](#)

Examples

```
# Set custom cache directory
tempd <- tempdir()
set_censobr_cache_dir(path = tempd)

# back to default path
set_censobr_cache_dir(path = NULL)
```

Index

- * **Cache data**
 - censobr_cache, [2](#)
 - get_censobr_cache_dir, [4](#)
 - set_censobr_cache_dir, [16](#)
- * **Census documentation**
 - data_dictionary, [3](#)
 - interview_manual, [5](#)
- * **Census tract data**
 - read_tracts, [15](#)
- * **Microdata**
 - read_emigration, [7](#)
 - read_families, [8](#)
 - read_households, [10](#)
 - read_mortality, [11](#)
 - read_population, [13](#)
- * **Questionnaire**
 - questionnaire, [6](#)

censobr_cache, [2](#), [5](#), [17](#)

data_dictionary, [3](#), [5](#)

get_censobr_cache_dir, [3](#), [4](#), [17](#)

interview_manual, [4](#), [5](#)

questionnaire, [6](#)

read_emigration, [7](#), [9](#), [11](#), [12](#), [14](#)

read_families, [8](#), [8](#), [11](#), [12](#), [14](#)

read_households, [8](#), [9](#), [10](#), [12](#), [14](#)

read_mortality, [8](#), [9](#), [11](#), [11](#), [14](#)

read_population, [8](#), [9](#), [11](#), [12](#), [13](#)

read_tracts, [15](#)

set_censobr_cache_dir, [3](#), [5](#), [16](#)