Package 'censobr'

March 23, 2024

Type Package

Title Download Data from Brazil's Population Census

Version 0.3.2

Description Download data from Brazil's population census. 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 (>= 4.0)

Imports arrow, checkmate, dplyr, httr (>= 1.4.1), tools

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

VignetteBuilder knitr

Encoding UTF-8

RoxygenNote 7.3.1

NeedsCompilation no

Author Rafael H. M. Pereira [aut, cre]
(<https://orcid.org/0000-0003-2125-7465>),
Rogério J. Barbosa [aut] (<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 2024-03-23 16:10:05 UTC

R topics documented:

censobr_cache *Manage cached files from the censobr package*

Description

Manage cached files from the censobr package

Usage

```
censobr_cache(list_files = TRUE, delete_file = NULL)
```

Arguments

`list_files`	Logical. Whether to print a message with the address of all censobr data sets cached locally. Defaults to `TRUE`.
`delete_file`	String. The file name (basename) of a censobr data set cached locally that should be deleted. Defaults to `NULL`, so that no file is deleted. If `delete_file = "all"`, then all cached censobr files are deleted.

Value

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

See Also

Other Cache data: `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` *Data dictionary of Brazil's census data*

Description

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

Usage

```
data_dictionary(year = 2010, dataset = NULL, showProgress = TRUE, cache = TRUE)
```

Arguments

`year`	Numeric. Year of reference in the format `yyyy`. Defaults to `2010`.
`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.

Value

Returns `NULL` and opens .html or .pdf file on the browser

See Also

Other Census documentation: `interview_manual()`

Examples

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

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

data_dictionary(year = 2010, dataset = 'tracts', showProgress = FALSE)
```

interview_manual *Interview manual of the data collection of Brazil's censuses*

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)
```

Arguments

year	Numeric. Year of reference in the format yyyy. Defaults to 2010.
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.

Value

Opens a .pdf file on the browser

See Also

Other Census documentation: data_dictionary()

Examples

```
# Open interview manual on 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)
```

Arguments

`year`	Numeric. Year of reference in the format `yyyy`. Defaults to `2010`.
`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 `c("long", "short")`.
`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.

Value

Opens a `.pdf` file on the browser

Examples

```
library(censobr)

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

`read_emigration` *Download microdata of emigration records from Brazil's census*

Description

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

Usage

```
read_emigration(
  year = 2010,
  columns = NULL,
  add_labels = NULL,
  as_data_frame = FALSE,
  showProgress = TRUE,
  cache = TRUE
)
```

Arguments

`year`	Numeric. Year of reference in the format `yyyy`. Defaults to `2010`.
`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.

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 = 2000,
  columns = NULL,
  add_labels = NULL,
  as_data_frame = FALSE,
  showProgress = TRUE,
  cache = TRUE
)
```

Arguments

`year`	Numeric. Year of reference in the format `yyyy`. Defaults to `2000`.
`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.

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 *Download microdata of household records from Brazil's census*

Description

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

Usage

```
read_households(
  year = 2010,
  columns = NULL,
  add_labels = NULL,
  as_data_frame = FALSE,
  showProgress = TRUE,
  cache = TRUE
)
```

Arguments

year Numeric. Year of reference in the format yyyy. Defaults to 2010.

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.

Value

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

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 = 2010,
  columns = NULL,
  add_labels = NULL,
  as_data_frame = FALSE,
  showProgress = TRUE,
  cache = TRUE
)
```

Arguments

year — Numeric. Year of reference in the format yyyy. Defaults to 2010.

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.

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 = 2010,
  columns = NULL,
  add_labels = NULL,
  as_data_frame = FALSE,
  showProgress = TRUE,
  cache = TRUE
)
```

Arguments

year	Numeric. Year of reference in the format yyyy. Defaults to 2010.
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.

Value

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

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 = 2010,
  dataset = NULL,
  as_data_frame = FALSE,
  showProgress = TRUE,
  cache = TRUE
)
```

Arguments

`year`	Numeric. Year of reference in the format `yyyy`. Defaults to `2010`.
`dataset`	Character. The dataset to be opened. Options currently include `c("Basico", "Domicilio", "DomicilioRenda", "Responsavel", "ResponsavelRenda", "Pessoa", "PessoaRenda", "Entorno")`. Preliminary results of the 2022 census are available with `"Preliminares"`.
`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.

Value

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

Examples

```
# return data as arrow Dataset
df <- read_tracts(year = 2010,
                  dataset = 'PessoaRenda',
                  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. If users want to set a custom cache directory, the function needs to be run again in each new R session.

Usage

```
set_censobr_cache_dir(path = NULL)
```

Arguments

`path`	String. The path to an existing directory. It defaults to `path = NULL`, to use the default directory

Value

A message indicating the directory where censobr files are cached.

See Also

Other Cache data: `censobr_cache()`

Examples

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

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

Index