Package 'educabR'

March 5, 2026

Title Download and Process Brazilian Education Data from INEP

Version 0.1.3

Description Download and process public education data from INEP (Instituto
Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira). Provides
functions to access microdata from the School Census (Censo Escolar),
ENEM (Exame Nacional do Ensino Médio), IDEB (Índice de Desenvolvimento
da Educação Básica), and other educational datasets. Returns data in
tidy format ready for analysis. Data source: INEP Open Data Portal
<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos>.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.3

URL https://github.com/SidneyBissoli/educabR,

https://sidneybissoli.github.io/educabR/

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

Depends R (>= 4.1.0)

Imports cli, dplyr, httr2, purrr, readr, rlang, stringr, tidyr, tools

Suggests ggplot2, knitr, readxl, rmarkdown, testthat (>= 3.0.0),
tibble, withr

Config/testthat/edition 3

VignetteBuilder knitr

NeedsCompilation no

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

Maintainer Sidney da Silva Pereira Bissoli <sbissoli76@gmail.com>

Repository CRAN

Date/Publication 2026-03-05 12:40:02 UTC

Contents

---

available_years	*Check available years for a dataset*

Description

Returns the years available for a given INEP dataset.

Usage

```
available_years(dataset)
```

Arguments

dataset	The dataset name.

Value

An integer vector of available years.

Examples

```
available_years("censo_escolar")
available_years("enem")
```

---

clear_cache *Clear the educabR cache*

Description

Removes all cached files from the educabR cache directory.

Usage

```
clear_cache(dataset = NULL)
```

Arguments

dataset   Optional. A character string specifying which dataset cache to clear. If NULL,
               clears all caches.

Value

Invisibly returns TRUE if successful.

Examples

```
## Not run:
# clear all cached data
clear_cache()

# clear only ENEM cache
clear_cache("enem")

## End(Not run)
```

---

enem_summary *Summary statistics for ENEM scores*

Description

Calculates summary statistics for ENEM scores, optionally grouped by demographic variables.

Usage

```
enem_summary(data, by = NULL)
```

Arguments

data   A tibble with ENEM data (from get_enem()).

by   Optional grouping variable(s) as character vector.

Value

A tibble with summary statistics for each score area.

Examples

```
## Not run:
enem <- get_enem(2023, n_max = 10000)

# overall summary
enem_summary(enem)

# summary by sex
enem_summary(enem, by = "tp_sexo")

## End(Not run)
```

---

get_cache_dir   *Get the current cache directory*

Description

Returns the current cache directory used by educabR.

Usage

```
get_cache_dir()
```

Value

A character string with the path to the cache directory.

Examples

```
get_cache_dir()
```

---

get_censo_escolar   *Get School Census (Censo Escolar) data*

Description

Downloads and processes microdata from the Brazilian School Census (Censo Escolar), conducted annually by INEP. Returns school-level data with information about infrastructure, location, and administrative details.

Usage

```
get_censo_escolar(year, uf = NULL, n_max = Inf, keep_zip = TRUE, quiet = FALSE)
```

Arguments

year	The year of the census (2007-2024).
uf	Optional. Filter by state (UF code or abbreviation).
n_max	Maximum number of rows to read. Default is Inf (all rows).
keep_zip	Logical. If TRUE, keeps the downloaded ZIP file in cache.
quiet	Logical. If TRUE, suppresses progress messages.

Details

The School Census is the main statistical survey on basic education in Brazil. It collects data from all public and private schools offering basic education (early childhood, elementary, and high school).

Important notes:

- The microdata contains one row per school (~217,000 schools in 2023).
- Column names are standardized to lowercase with underscores.
- Use the uf parameter to filter by state for faster processing.

Value

A tibble with school data in tidy format.

Data dictionary

For detailed information about variables, see INEP's documentation: https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/censo-escolar

Examples

```
## Not run:
# get schools data for 2023
escolas <- get_censo_escolar(2023)

# get schools from Sao Paulo state only
escolas_sp <- get_censo_escolar(2023, uf = "SP")

# read only first 1000 rows for exploration
escolas_sample <- get_censo_escolar(2023, n_max = 1000)

## End(Not run)
```

get_enem *Get ENEM (Exame Nacional do Ensino Médio) data*

Description

Downloads and processes microdata from ENEM, the Brazilian National High School Exam. ENEM is used for university admissions and as a high school equivalency exam.

Usage

```
get_enem(year, n_max = Inf, keep_zip = TRUE, quiet = FALSE)
```

Arguments

year	The year of the exam (2009-2023).
n_max	Maximum number of rows to read. Default is Inf (all rows). Consider using a smaller value for exploration, as ENEM files contain millions of rows.
keep_zip	Logical. If TRUE, keeps the downloaded ZIP file in cache.
quiet	Logical. If TRUE, suppresses progress messages.

Details

ENEM is conducted annually by INEP and is the largest exam in Brazil, with millions of participants. The microdata includes:

- Participant demographics (age, sex, race, etc.)
- Socioeconomic questionnaire responses
- Scores for each test area
- Essay scores
- School information (when applicable)

Important notes:

- ENEM files are very large (several GB when extracted).
- Use n_max to read a sample first for exploration.
- Column names are standardized to lowercase with underscores.
- Score variables start with nu_nota_ prefix.

Value

A tibble with the ENEM microdata in tidy format.

Data dictionary

For detailed information about variables, see INEP's documentation: https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem

Examples

```
## Not run:
# get a sample of 10000 rows for exploration
enem_sample <- get_enem(2023, n_max = 10000)

# get full data (warning: large file)
enem_2023 <- get_enem(2023)

## End(Not run)
```

---

get_enem_itens   *Get ENEM item response data*

Description

Downloads and processes ENEM item response (gabarito) data, which contains detailed information about each question.

Usage

```
get_enem_itens(year, n_max = Inf, quiet = FALSE)
```

Arguments

year	The year of the exam (2009-2023).
n_max	Maximum number of rows to read.
quiet	Logical. If TRUE, suppresses progress messages.

Value

A tibble with item response data.

Examples

```
## Not run:
# get item data for 2023
itens <- get_enem_itens(2023)

## End(Not run)
```

---

get_ideb   *Get IDEB (Índice de Desenvolvimento da Educação Básica) data*

Description

Downloads and processes IDEB data from INEP. IDEB is the main indicator of education quality in Brazil, combining student performance (from SAEB) with grade promotion rates.

Usage

```
get_ideb(
  year,
  level = c("escola", "municipio"),
  stage = c("anos_iniciais", "anos_finais", "ensino_medio"),
  uf = NULL,
  quiet = FALSE
)
```

Arguments

year	The year of the IDEB (available: 2017, 2019, 2021, 2023).
level	The aggregation level:

- "escola": School level
- "municipio": Municipality level

stage	The education stage:

- "anos_iniciais": Early elementary (1st-5th grade)
- "anos_finais": Late elementary (6th-9th grade)
- "ensino_medio": High school

uf	Optional. Filter by state (UF code or abbreviation).
quiet	Logical. If TRUE, suppresses progress messages.

Details

IDEB is calculated every two years since 2005 based on:

- **Learning**: Average scores in Portuguese and Mathematics from SAEB
- **Flow**: Grade promotion rate (inverse of repetition/dropout)

The index ranges from 0 to 10. Brazil's national goal is to reach 6.0 by 2022 (the level of developed countries in PISA).

Note: IDEB data is relatively small compared to other INEP datasets, so no n_max parameter is provided.

Value

A tibble with IDEB data in tidy format.

Data source

Official IDEB portal: `https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/`
`ideb`

Examples

```
## Not run:
# get school-level IDEB for early elementary in 2021
ideb_escolas <- get_ideb(2021, level = "escola", stage = "anos_iniciais")

# get municipality-level IDEB for São Paulo state
ideb_sp <- get_ideb(2021, level = "municipio", stage = "anos_iniciais", uf = "SP")

# get high school IDEB for all municipalities
ideb_em <- get_ideb(2023, level = "municipio", stage = "ensino_medio")

## End(Not run)
```

---

get_ideb_series　　　　　*Get IDEB historical series*

Description

Downloads and combines IDEB data across multiple years to create a historical series.

Usage

```
get_ideb_series(
  years = NULL,
  level = c("escola", "municipio"),
  stage = c("anos_iniciais", "anos_finais", "ensino_medio"),
  uf = NULL,
  quiet = FALSE
)
```

Arguments

years	Vector of years to include (default: all available).
level	The aggregation level.
stage	The education stage.
uf	Optional. Filter by state.
quiet	Logical. If TRUE, suppresses progress messages.

Value

A tibble with IDEB data for all requested years.

Examples

```
## Not run:
# get IDEB history for municipalities
ideb_hist <- get_ideb_series(
  years = c(2017, 2019, 2021),
  level = "municipio",
  stage = "anos_iniciais"
)

## End(Not run)
```

---

list_cache   *List cached files*

Description

Lists all files currently in the educabR cache.

Usage

```
list_cache(dataset = NULL)
```

Arguments

dataset   Optional. Filter by dataset name.

Value

A tibble with information about cached files.

Examples

```
## Not run:
list_cache()

## End(Not run)
```

---

list_censo_files   *List available Censo Escolar files*

Description

Lists the data files available in a downloaded School Census.

Usage

```
list_censo_files(year)
```

Arguments

year   The year of the census.

Value

A character vector of file names found.

Examples

```
## Not run:
list_censo_files(2023)

## End(Not run)
```

---

list_ideb_available   *List available IDEB data*

Description

Lists the IDEB data files available in the INEP portal.

Usage

```
list_ideb_available()
```

Value

A tibble with available IDEB datasets.

Examples

```
list_ideb_available()
```

---

set_cache_dir *Set the cache directory for educabR*

Description

Sets the directory where downloaded files will be cached. This avoids repeated downloads of the same data.

Usage

```
set_cache_dir(path = NULL, persistent = FALSE)
```

Arguments

path	A character string with the path to the cache directory. If NULL, uses a temporary directory (default).
persistent	Logical. If TRUE, the cache directory setting is saved to the user's R profile for future sessions.

Value

Invisibly returns the cache directory path.

Examples

```
## Not run:
# set a custom cache directory (use tempdir() in examples)
set_cache_dir(file.path(tempdir(), "educabR_cache"))

## End(Not run)
```

Index