Package: dgeecr (via r-universe)

November 28, 2026

Title Access to Portuguese Education Statistics by Region (DGEEC)

Version 0.1.4

Description Client for the education statistics portal of the Directorate-General for Education and Science Statistics (DGEEC) of Portugal, <https://estatisticas-educacao.dgeec.medu.pt/regioesemnumeros/>. Each page of the portal is a form whose cartesian product of options is traversed with parallel 'httr2' requests, and the returned HTML is converted into a long table (one row per observation), covering students enrolled by school year, region, school level, nature (public or private), gender and age group, together with establishment, indicator and human-resource tables. Results can be enriched with Portuguese administrative and health divisions (ULS, NUTS and the 2024 health regions) using a linkage file bundled with the package, and requested by name as well as by internal code. Downloaded pages are cached on disk, so that interrupted extractions can be resumed.

License Apache License (>= 2)

URL https://jdrdionisio.github.io/dgeecr/, https://github.com/jdrdionisio/dgeecr

BugReports https://github.com/jdrdionisio/dgeecr/issues

Copyright The bundled geographic linkage file (inst/extdata) is compiled from Portuguese official legislation and health-authority publications, and is included with permission; see inst/COPYRIGHTS.

Encoding UTF-8

Language pt

Depends R (>= 4.1)

Imports dplyr, httr2 (>= 1.0.0), nanoparquet, parallel, purrr, R6, readr, rlang, rvest, stringr, tibble, tidyr, writexl, xml2

Suggests testthat (>= 3.0.0), knitr, rmarkdown

Config/testthat/edition 3

Config/roxygen2/version 8.1.1

VignetteBuilder knitr

NeedsCompilation no

Author João Dionísio [aut, cre]

Maintainer João Dionísio <joaodrdionisio@gmail.com>

Repository https://cran.r-universe.dev

Date/Publication 2026-11-28 21:05:17 UTC

RemoteUrl https://github.com/cran/dgeecr

RemoteRef HEAD

RemoteSha bc2fcb8e15fb163e291af4a01273bade6e528452

Contents

agregar_regiao_saude 2
agregar_uls 3
dgeec_cache_dir 3
dgeec_dados 4
dgeec_dimensoes 4
dgeec_exportar 5
dgeec_extrair_tudo 6
dgeec_geo 6
dgeec_limpar_cache 7
dgeec_paginas 7
dgeec_regioes 8
dgeec_regioes_saude 9
dgeec_tamanho_cache 9
dgeec_uls 10
DgeecClient 10
enriquecer_geo 14

Index **15**

`agregar_regiao_saude` *Agrega dados ao nivel da regiao de saude 2024*

Description

Agrega dados ao nivel da regiao de saude 2024

Usage

```
agregar_regiao_saude(dados, geo = NULL)
```

Arguments

dados	tibble de DgeecClient$dados().
geo	Tabela de ligacao (por omissao dgeec_geo()).

Value

Um tibble ao nivel da regiao de saude.

---

agregar_uls *Agrega dados ao nivel da ULS*

Description

Soma os valores dos municipios pertencentes a cada ULS.

Usage

```
agregar_uls(dados, geo = NULL, multiplas = c("excluir", "duplicar", "maior"))
```

Arguments

dados	tibble de DgeecClient$dados().
geo	Tabela de ligacao (por omissao dgeec_geo()).
multiplas	Como tratar os municipios repartidos por varias ULS (Lisboa, Porto, Gondomar, Loures, Funchal), cujos dados o portal so publica ao nivel do municipio: "excluir" (por omissao) remove-os da soma e lista-os no atributo municipios_excluidos; "duplicar" atribui o valor a cada ULS (dupla contagem); "maior" atribui a ULS com mais freguesias.

Value

Um tibble ao nivel da ULS (+ atributo municipios_excluidos).

---

dgeec_cache_dir *Pasta de cache por omissao*

Description

Pasta de cache por omissao

Usage

```
dgeec_cache_dir()
```

Value

O caminho da pasta de cache do pacote.

dgeec_dados *Extrair dados de uma pagina (atalho)*

Description

Cria um DgeecClient e extrai os dados de uma pagina.

Usage

```
dgeec_dados(pagina, ..., cliente = NULL)
```

Arguments

pagina	Id de pagina (ver dgeec_paginas()).
...	Argumentos passados a DgeecClient$dados() (p.ex. regioes, natureza, incluir_totais).
cliente	Um DgeecClient opcional.

Value

Um tibble longo.

Examples

```
## Not run:
# requer acesso a Internet
dgeec_dados("pre_escolar", regioes = "Portugal")

## End(Not run)
```

dgeec_dimensoes *Dimensoes (filtros) de uma pagina*

Description

Le a pagina do portal e devolve os valores possiveis de cada <select>, com o respetivo nivel hierarquico (no caso das regioes).

Usage

```
dgeec_dimensoes(pagina, timeout = 45L)
```

Arguments

pagina	Id de pagina (ver dgeec_paginas()) ou o respetivo path.
timeout	Timeout por pedido, em segundos.

Value

Um tibble com select, codigo, rotulo e nivel.

Examples

```
## Not run:
# requer acesso a Internet
dgeec_dimensoes("pre_escolar")

## End(Not run)
```

---

dgeec_exportar *Exportar dados*

Description

Escreve um data.frame em disco. O CSV e o Parquet suportam o conjunto completo; o XLSX so e gerado se o numero de linhas couber no limite de uma aba do Excel (limite_xlsx), caso contrario e ignorado com aviso.

Usage

```
dgeec_exportar(
  dados,
  ficheiro,
  formatos = c("csv", "parquet", "xlsx"),
  limite_xlsx = 1048575L
)
```

Arguments

dados	data.frame/tibble a exportar.
ficheiro	Caminho base, sem extensao.
formatos	Formatos a escrever: qualquer subconjunto de "csv", "parquet", "xlsx".
limite_xlsx	Maximo de linhas para gerar XLSX (limite do Excel: 1 048 576; por seguranca usa-se 1 048 575).

Value

(invisivelmente) o vetor de ficheiros escritos.

dgeec_extrair_tudo *Extrair e exportar todas as paginas*

Description

Percorre o catalogo, extrai cada pagina e escreve os formatos pedidos.

Usage

```
dgeec_extrair_tudo(
  dir,
  paginas = NULL,
  formatos = c("csv", "parquet"),
  cliente = NULL,
  ...
)
```

Arguments

dir	Pasta de saida.
paginas	Vetor de ids (NULL = todas).
formatos	Formatos a escrever (ver dgeec_exportar()).
cliente	Um DgeecClient. Se NULL, cria um.
...	Argumentos passados a cliente$dados().

Value

(invisivelmente) um tibble com pagina e ficheiro.

dgeec_geo *Tabela de ligacao municipio -> divisoes administrativas e de saude*

Description

Le o ficheiro de ligacao geografica incluido no pacote (inst/extdata/geo_linkage_2024_v8.csv) e agrega-o ao nivel do municipio, resolvendo as ULS que dividem um municipio ao nivel da freguesia em uls_1 e uls_2.

Usage

```
dgeec_geo(path = NULL)
```

Arguments

path	Caminho alternativo para o CSV de ligacao.

Value

Um `tibble` por municipio, com `uls_1`, `uls_2`, `n_uls`, `uls_todas`, `uls_principal` (ULS com mais freguesias), `uls_hierarquia_csp` (nomenclatura do BI-CSP), `regiao_2024`, `nuts1`, `nuts2`, `nuts3`, `aces_2022`, `ars_2022` e `freguesias`.

---

dgeec_limpar_cache *Limpar o cache em disco*

Description

Remove o HTML cru descarregado e/ou o resultado ja interpretado (RDS).

Usage

```
dgeec_limpar_cache(
  dir = NULL,
  html = TRUE,
  dados = TRUE,
  confirmar = interactive()
)
```

Arguments

`dir`	Pasta de cache (NULL = `dgeec_cache_dir()`).
`html`	Remover o HTML cru?
`dados`	Remover os resultados interpretados (RDS)?
`confirmar`	Pedir confirmacao (por omissao, apenas em sessao interativa).

Value

(invisivelmente) o numero de ficheiros removidos.

---

dgeec_paginas *Catalogo de paginas disponiveis*

Description

Devolve o conjunto de paginas (tabelas) publicadas no portal.

Usage

```
dgeec_paginas()
```

Value

Um `tibble` com as colunas `id`, `grupo`, `path` e `descricao`.

Examples

```
dgeec_paginas()
```

`dgeec_regioes` *Catalogo de regioes do portal*

Description

Devolve a lista de regioes aceites em `regioes=` por DgeecClient`$dados()`, com o codigo e o nome, para nao ser necessario conhecer os codigos.

Usage

```
dgeec_regioes(pagina = "pre_escolar", timeout = 45L)
```

Arguments

`pagina`	Pagina de onde ler a lista (o catalogo e igual em todas as paginas; ver `dgeec_paginas()`). Por omissao `"pre_escolar"`.
`timeout`	Timeout por pedido, em segundos.

Value

Um `tibble` com `regiao_codigo`, `regiao_nome`, `regiao_tipo` e `regiao_nivel` (0 = Portugal e NUTS I, 1 = NUTS II, 2 = NUTS III, 3 = municipio).

Examples

```
## Not run:
# requer acesso a Internet
dgeec_regioes()
dgeec_regioes() |> subset(regiao_nivel == 3)   # so municipios

## End(Not run)
```

dgeec_regioes_saude *Catalogo de regioes de saude 2024*

Description

Catalogo de regioes de saude 2024

Usage

```
dgeec_regioes_saude(geo = NULL)
```

Arguments

geo	Tabela de ligacao (por omissao dgeec_geo()).

Value

Um tibble com regiao_2024, n_municipios e municipios.

dgeec_tamanho_cache *Tamanho do cache em disco*

Description

Tamanho do cache em disco

Usage

```
dgeec_tamanho_cache(dir = NULL)
```

Arguments

dir	Pasta de cache (NULL = dgeec_cache_dir()).

Value

Um tibble por pagina (ou "(outros)"), com n_ficheiros, bytes e mb, mais o total em atributo total.

dgeec_uls *Catalogo de ULS (2024)*

Description

Catalogo de ULS (2024)

Usage

```
dgeec_uls(geo = NULL)
```

Arguments

geo	Tabela de ligacao (por omissao dgeec_geo()).

Value

Um tibble com uls, regiao_2024, n_municipios e municipios.

DgeecClient *Cliente DGEEC*

Description

Cliente R6 para o portal "Regioes em Numeros - Educacao". Gere a paralelizacao (httr2 na rede, parallel no parsing), o cache em disco (com resume) e a conversao em tabela longa.

Public fields

cache_dir Pasta de cache (NULL = tools::R_user_dir("dgeecr","cache")).

concorrencia No de pedidos simultaneos (rede).

nucleos No de processos para o parsing (NULL = todos menos 2, ou 2 durante a verificacao da CRAN).

timeout Timeout por pedido (s).

tentativas Tentativas por pedido.

usar_cache Reutilizar HTML ja descarregado?

cache_dados Guardar/ler tambem o resultado ja interpretado (RDS)?

lote No de pedidos por lote.

Methods

Public methods:

- DgeecClient$new()
- DgeecClient$paginas()
- DgeecClient$dimensoes()
- DgeecClient$regioes()
- DgeecClient$dados()
- DgeecClient$exportar()
- DgeecClient$tamanho_cache()
- DgeecClient$limpar_cache()
- DgeecClient$.parsear()
- DgeecClient$.descarregar()
- DgeecClient$clone()

`DgeecClient$new()`: Cria um cliente.

Usage:

```
DgeecClient$new(
  cache_dir = NULL,
  concorrencia = 8L,
  nucleos = NULL,
  timeout = 45L,
  tentativas = 3L,
  usar_cache = TRUE,
  cache_dados = TRUE,
  lote = 250L
)
```

Arguments:

`cache_dir` Pasta de cache.

`concorrencia` No de pedidos simultaneos.

`nucleos` No de processos para o parsing.

`timeout` Timeout por pedido (s).

`tentativas` Tentativas por pedido.

`usar_cache` Reutilizar cache em disco.

`cache_dados` Guardar/ler o resultado interpretado.

`lote` No de pedidos por lote.

`DgeecClient$paginas()`: Catalogo de paginas.

Usage:

```
DgeecClient$paginas()
```

`DgeecClient$dimensoes()`: Dimensoes (filtros) de uma pagina.

Usage:

```
DgeecClient$dimensoes(pagina)
```

Arguments:

`pagina` Id de pagina.

`DgeecClient$regioes()`: Catalogo de regioes (codigo e nome).

Usage:

```
DgeecClient$regioes(pagina = "pre_escolar")
```

Arguments:

`pagina` Pagina de onde ler a lista.

`DgeecClient$dados()`: Extrai dados de uma pagina.

Usage:

```
DgeecClient$dados(
  pagina,
  regioes = NULL,
  uls = NULL,
  regiao_saude = NULL,
  natureza = NULL,
  genero = NULL,
  orientacao = NULL,
  incluir_totais = FALSE,
  progress = TRUE
)
```

Arguments:

`pagina` Id de pagina (ver `dgeec_paginas()`).

`regioes` Codigos **ou nomes** de regiao (NULL = todas). Os nomes podem ser de Portugal, NUTS I/II/III ou municipio; ver `dgeec_regioes()`. Ex.: `regioes = c("Portugal", "Moncao")`.

`uls` Nome(s) de ULS (ver `dgeec_uls()`) a que restringir a extracao.

`regiao_saude` Nome(s) de regiao de saude 2024 (ver `dgeec_regioes_saude()`).

`natureza` Codigos de natureza (NULL = desagregado 1/2, ou todos se `incluir_totais`).

`genero` Codigos de genero (NULL = desagregado 1/2, ou todos se `incluir_totais`).

`orientacao` Codigos de orientacao (NULL = desagregado 10/20, ou todos se `incluir_totais`).

`incluir_totais` Incluir tambem os niveis totais (0/...)?

`progress` Mostrar progresso.

Returns: Um `tibble` longo (1 linha = 1 observacao).

`DgeecClient$exportar()`: Exporta um data frame em CSV/Parquet/XLSX.

Usage:

```
DgeecClient$exportar(
  dados,
  ficheiro,
  formatos = c("csv", "parquet", "xlsx"),
  limite_xlsx = 1048575L
)
```

Arguments:

`dados` `tibble` a exportar.

`ficheiro` Caminho base (sem extensao).

`formatos` Formatos a escrever.

`limite_xlsx` Maximo de linhas para gerar XLSX (limite do Excel).

`DgeecClient$tamanho_cache()`: Tamanho do cache em disco.

Usage:

```
DgeecClient$tamanho_cache()
```

`DgeecClient$limpar_cache()`: Limpa o cache em disco.

Usage:

```
DgeecClient$limpar_cache(...)
```

Arguments:

`...` Passado a `dgeec_limpar_cache()` (`html`, `dados`, `confirmar`).

`DgeecClient$.parsear()`:

Usage:

```
DgeecClient$.parsear(chaves, ctx)
```

`DgeecClient$.descarregar()`:

Usage:

```
DgeecClient$.descarregar(path, combos, chaves, progress = TRUE)
```

`DgeecClient$clone()`: The objects of this class are cloneable with this method.

Usage:

```
DgeecClient$clone(deep = FALSE)
```

Arguments:

`deep` Whether to make a deep clone.

Examples

```
## Not run:
# requer acesso a Internet
cli <- DgeecClient$new()
d <- cli$dados("pre_escolar", regioes = c("Portugal", "Moncao"))

## End(Not run)
```

enriquecer_geo *Enriquece dados DGEEC com divisoes administrativas e de saude*

Description

Liga as linhas de nivel "Municipio" ao ficheiro de ligacao geografica e acrescenta `uls_1`, `uls_2`, `uls_principal`, `n_uls`, `regiao_2024`, `nuts1`, `nuts2`, `nuts3`, `aces_2022` e `ars_2022`. Municipios com varias ULS (Lisboa, Porto, Gondomar, Loures, Funchal) ficam com `uls_1`/`uls_2`.

Usage

```
enriquecer_geo(dados, geo = NULL)
```

Arguments

`dados`	Um `tibble` produzido por DgeecClient`$dados()` / `dgeec_dados()`, com as colunas `regiao_codigo`, `regiao_nome` e `regiao_tipo`.
`geo`	Tabela de ligacao (por omissao `dgeec_geo()`).

Value

`dados` com as colunas geograficas acrescentadas.

Index

agregar_regiao_saude, 2
agregar_uls, 3

dgeec_cache_dir, 3
dgeec_cache_dir(), *7*, *9*
dgeec_dados, 4
dgeec_dados(), *14*
dgeec_dimensoes, 4
dgeec_exportar, 5
dgeec_exportar(), *6*
dgeec_extrair_tudo, 6
dgeec_geo, 6
dgeec_geo(), *3*, *9, 10*, *14*
dgeec_limpar_cache, 7
dgeec_limpar_cache(), *13*
dgeec_paginas, 7
dgeec_paginas(), *4*, *8*, *12*
dgeec_regioes, 8
dgeec_regioes(), *12*
dgeec_regioes_saude, 9
dgeec_regioes_saude(), *12*
dgeec_tamanho_cache, 9
dgeec_uls, 10
dgeec_uls(), *12*
DgeecClient, *3, 4*, *6*, *8*, 10, *14*

enriquecer_geo, 14