Package 'subincomeR'

August 21, 2026

Title Access to Global Sub-National Income Data

Version 0.6.0

Description Provides access to granular sub-national income data from the MCC-PIK Database Of Sub-national Economic Output (DOSE). The package downloads and processes the data from its open repository on 'Zenodo' (<https://zenodo.org/records/20035157>). Functions are provided to fetch data at multiple geographic levels, match coordinates to administrative regions, and access associated geometries.

License MIT + file LICENSE

URL https://github.com/pablogguz/subincomeR, https://pablogguz.github.io/subincomeR/

BugReports https://github.com/pablogguz/subincomeR/issues

Depends R (>= 4.1.0)

Imports countrycode, curl, dplyr, rlang, sf, tidygeocoder, tools, utils, zip

Suggests fixest, ggplot2, ggtext, knitr, rmarkdown, spelling, testthat (>= 3.1.7), withr

VignetteBuilder knitr

Config/testthat/edition 3

Encoding UTF-8

Language en-US

RoxygenNote 7.3.3

NeedsCompilation no

Author Pablo García Guzmán [aut, cre, cph]

Maintainer Pablo García Guzmán <garciagp@ebrd.com>

Repository CRAN

Date/Publication 2026-08-21 21:00:02 UTC

Contents

getDOSE 2
getDOSE_geom 3
matchDOSE 4
subincomeR_cache_clear 6
subincomeR_cache_dir 7

Index **8**

`getDOSE` *Download and load the DOSE dataset into memory*

Description

This function downloads the DOSE dataset from Zenodo and loads it into memory as a dataframe. It allows for optional filtering of the dataset based on specific years and/or countries. The country format can be specified to ensure correct filtering. The function automatically handles different download methods based on system capabilities.

Usage

```
getDOSE(
  years = NULL,
  countries = NULL,
  format_countries = "country.name",
  path = NULL,
  cache = FALSE
)
```

Arguments

`years`	Optional vector of years for which to filter the DOSE dataset. If NULL (the default), data for all years are returned.
`countries`	Optional vector of countries for which to filter the DOSE dataset. Countries can be specified in ISO2C, ISO3C, or country name format. Use the format_countries parameter to specify the format of the countries vector.
`format_countries`	The format of the countries provided in the countries parameter. Acceptable values are 'iso2c', 'iso3c', or 'country.name'. Default is 'country.name'. This parameter is used only if the countries parameter is not NULL.
`path`	Optional character string specifying where to store the downloaded data. If NULL (default), uses tempdir(). Cannot be combined with `cache = TRUE`.
`cache`	Logical. If TRUE, stores the downloaded data in a persistent per-user cache directory (see `subincomeR_cache_dir()`) so it is reused across R sessions. Default is FALSE, which stores the data in tempdir() for the current session only.

Value

A dataframe containing the filtered DOSE dataset based on the input parameters, or NULL (invisibly, with an informative message) if the dataset could not be downloaded or read, for example when there is no internet connection or the Zenodo repository is temporarily unavailable.

Examples

```
# Load the entire dataset
data_all <- getDOSE()

# Load dataset filtered by specific years
data_2018_2019 <- getDOSE(years = c(2018, 2019))

# Load dataset filtered by specific countries (using ISO3C codes)
data_usa_can <- getDOSE(countries = c('USA', 'CAN'), format_countries = 'iso3c')

# Load dataset filtered by year and countries (using country names)
data_mex_2019 <- getDOSE(years = 2019, countries = c('Mexico'),
                         format_countries = 'country.name')


## Not run:
# Store the data in a persistent cache so it is reused across sessions
data_cached <- getDOSE(cache = TRUE)

## End(Not run)
```

---

getDOSE_geom *Download and load GADM-DOSE geometries*

Description

This function downloads and loads GADM-DOSE geometries from a remote source. The geometries are stored in a temporary directory by default, or in a user-specified location if provided. The uncompressed geometries file is approximately 769 MB.

Usage

```
getDOSE_geom(path = NULL, countries = NULL, download = FALSE, cache = FALSE)
```

Arguments

path	Optional character string specifying where to store the files. If NULL (default), uses tempdir(). Cannot be combined with cache = TRUE.
countries	Optional vector of ISO3C country codes to filter geometries. If NULL (default), all available geometries are returned.

download — Logical indicating whether to download without confirmation. Default is FALSE, which will prompt for confirmation in interactive sessions. Set to TRUE to skip confirmation.

cache — Logical. If TRUE, stores the downloaded geometries in a persistent per-user cache directory (see subincomeR_cache_dir()) so they are reused across R sessions. Default is FALSE, which stores the files in tempdir() for the current session only.

Value

An sf object containing the GADM-DOSE geometries, or NULL (invisibly, with an informative message) if the geometries could not be downloaded or read, for example when there is no internet connection, the remote resource is temporarily unavailable, or the download was canceled by the user.

Examples

```
# Load all geometries with download confirmation
geom_all <- getDOSE_geom()

# Load geometries with automatic download
geom_auto <- getDOSE_geom(download = TRUE)

# Load geometries for specific countries
geom_subset <- getDOSE_geom(
  countries = c("USA", "CAN", "MEX"),
  download = TRUE
)


## Not run:
# Store the geometries in a persistent cache to avoid re-downloading
# the ~769 MB file in every R session
geom_cached <- getDOSE_geom(download = TRUE, cache = TRUE)

## End(Not run)
```

---

matchDOSE *Match coordinates to DOSE dataset*

Description

This function matches input coordinates (latitude and longitude) to the DOSE dataset. It accepts either vectors of latitudes and longitudes or a dataframe containing these coordinates. Before matching, it ensures that only unique coordinates are processed to avoid duplicating operations on identical coordinates. It uses GADM-1 geometries to match coordinates to regions and returns a dataframe with unique input coordinates and matched DOSE data.

Usage

```
matchDOSE(
  lat = NULL,
  long = NULL,
  df = NULL,
  lat_col = "lat",
  long_col = "long",
  years = NULL,
  countries = NULL,
  format_countries = "iso3c",
  path = NULL,
  download = FALSE,
  cache = FALSE
)
```

Arguments

lat	Optional vector of latitudes of the points to match. Required if no dataframe is provided.
long	Optional vector of longitudes of the points to match. Required if no dataframe is provided.
df	Optional dataframe containing coordinates and possibly additional columns. If provided, 'lat' and 'long' vectors should not be provided. The dataframe must include columns specified by 'lat_col' and 'long_col' parameters.
lat_col	Optional name of the latitude column in 'df'. Only used if 'df' is provided. Defaults to "lat".
long_col	Optional name of the longitude column in 'df'. Only used if 'df' is provided. Defaults to "long".
years	Optional vector of years for which to filter the DOSE dataset. If NULL (the default), a 1:m matching is performed and data for all years are returned.
countries	Optional vector or dataframe column name of country identifiers. If provided, the function skips the country matching step. Can significantly reduce processing time.
format_countries	Specifies the format of the country identifiers in 'countries'. Options are "country.name" (default), "iso3c", and "iso2c". This parameter is ignored if 'countries' is NULL.
path	Optional character string specifying where to store downloaded files. If NULL (default), uses tempdir(). Cannot be combined with cache = TRUE.
download	Logical indicating whether to download without confirmation. Default is FALSE, which will prompt for confirmation in interactive sessions. Set to TRUE to skip confirmation.
cache	Logical. If TRUE, stores the downloaded files in a persistent per-user cache directory (see subincomeR_cache_dir()) so they are reused across R sessions. Default is FALSE, which stores the files in tempdir() for the current session only.

Value

A dataframe with input coordinates (and any additional input dataframe columns) and matched DOSE data, or NULL (invisibly, with an informative message) if any required internet resource — the geocoding service, the DOSE dataset, or the geometries — is unavailable.

Examples

```
# Match coordinates using vectors
matched_data <- matchDOSE(lat = c(19.4326, 51.5074),
                          long = c(-99.1332, -0.1276))

# Match coordinates using a dataframe
df <- data.frame(ID = 1:2,
                 latitude = c(19.4326, 51.5074),
                 longitude = c(-99.1332, -0.1276))
matched_data_df <- matchDOSE(df = df,
                             lat_col = "latitude",
                             long_col = "longitude")

# Match coordinates for a specific year
matched_data_2019 <- matchDOSE(lat = c(19.4326),
                               long = c(-99.1332),
                               years = 2019)

# Match coordinates with known countries
matched_data_countries <- matchDOSE(lat = c(19.4326, 51.5074),
                                    long = c(-99.1332, -0.1276),
                                    countries = c("MEX", "GBR"),
                                    format_countries = "iso3c")
```

`subincomeR_cache_clear`

Clear the subincomeR persistent cache

Description

Deletes the persistent cache directory used when functions are called with `cache = TRUE` (see `subincomeR_cache_dir()`), including the downloaded DOSE dataset and geometries stored there.

Usage

```
subincomeR_cache_clear(ask = interactive())
```

Arguments

`ask`	Logical. If TRUE (the default in interactive sessions), asks for confirmation before deleting. Set to FALSE to delete without confirmation.

Value

NULL, invisibly. Called for its side effect.

Examples

```
## Not run:
subincomeR_cache_clear()

## End(Not run)
```

`subincomeR_cache_dir` *Path to the subincomeR persistent cache directory*

Description

Returns the directory used to store downloaded data when `getDOSE()`, `getDOSE_geom()` or `matchDOSE()` are called with `cache = TRUE`. The location follows `tools::R_user_dir()`, so it respects the `R_USER_CACHE_DIR` environment variable. The directory is not created by this function.

Usage

```
subincomeR_cache_dir()
```

Value

A character string with the path to the cache directory.

Examples

```
subincomeR_cache_dir()
```

Index

getDOSE, 2
getDOSE(), *7*
getDOSE_geom, 3
getDOSE_geom(), *7*

matchDOSE, 4
matchDOSE(), *7*

subincomeR_cache_clear, 6
subincomeR_cache_dir, 7
subincomeR_cache_dir(), *2*, *4–6*

tools::R_user_dir(), *7*