

Package ‘subincomeR’

December 19, 2024

Title Access to Global Sub-National Income Data

Version 0.2.2

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/13773040>>). 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>

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

Suggests extrafont, fixest, ggplot2, ggtext, knitr, rmarkdown,
spelling

VignetteBuilder knitr

Config/testthat/edition 3

Encoding UTF-8

Language en-US

RoxygenNote 7.3.1

NeedsCompilation no

Author Pablo García Guzmán [aut, cre, cph]

Maintainer Pablo García Guzmán <garciagp@ebrd.com>

Repository CRAN

Date/Publication 2024-12-19 20:40:02 UTC

Contents

getDOSE	2
getDOSE_geom	3
matchDOSE	4

`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  
)
```

Arguments

<code>years</code>	Optional vector of years for which to filter the DOSE dataset. If NULL (the default), data for all years are returned.
<code>countries</code>	Optional vector of countries for which to filter the DOSE dataset. Countries can be specified in ISO2C, ISO3C, or country name format. Use the <code>format_countries</code> parameter to specify the format of the countries vector.
<code>format_countries</code>	The format of the countries provided in the <code>countries</code> parameter. Acceptable values are 'iso2c', 'iso3c', or 'country.name'. Default is 'country.name'. This parameter is used only if the <code>countries</code> parameter is not NULL.
<code>path</code>	Optional character string specifying where to store the downloaded data. If NULL (default), uses <code>tempdir()</code> .

Value

A dataframe containing the filtered DOSE dataset based on the input parameters.

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')
```

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)
```

Arguments

path	Optional character string specifying where to store the files. If NULL (default), uses tempdir().
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.

Value

An sf object containing the GADM-DOSE geometries

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
)
```

 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
)
```

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.

Index

getDOSE, [2](#)
getDOSE_geom, [3](#)
matchDOSE, [4](#)