

Package ‘MicroDatosEs’

January 20, 2025

Type Package

Title Utilities for Official Spanish Microdata

Version 0.8.15

Date 2023-11-24

Encoding UTF-8

Description

Provides utilities for reading and processing microdata from Spanish official statistics with R.

URL <https://www.datanalytics.com/2012/08/06/un-paseo-por-el-paquete-microdatoses-y-la-epa-de-nuevo/>

Depends readr

License GPL-3

LazyLoad yes

RoxygenNote 7.2.1

NeedsCompilation no

Author Carlos J. Gil Bellosta [aut, cre],
Carlos Neira [ctb],
Diego Paniagua Sánchez [ctb],
Fiorella Mori Peláez [ctb],
Jorge López Pérez [ctb],
Estitxu Villamor [ctb],
José Luis Cañadas Reche [ctb]

Maintainer Carlos J. Gil Bellosta <cgb@datanalytics.com>

Repository CRAN

Date/Publication 2023-11-24 16:10:02 UTC

Contents

censo2010	2
defun2011	3
ees2010	4

epa2005	5
epf.2011.gastos	6
epf.2011.hogares	7
epf.2011.miembros	8
padron2016	10
test.metadata	11

Index	13
--------------	-----------

censo2010	<i>Download data from 2010 Spanish Census from INE</i>
-----------	--

Description

This function downloads the data from 2010 (or later) Spanish Census using the information provided by the Spanish Statistical Office (INE).

Usage

```
censo2010(file)
```

Arguments

<code>file</code>	Character string with the name of the microdata file provided by the INE on the Census section. It allows either a path to a file, or literal data (single string or raw vector). It also allows compressed files in <code>.gz</code> , <code>.bz2</code> , <code>.xz</code> , or <code>.zip</code> format.
-------------------	--

Details

This function reads microdata from the 2010 census in Spain. It was originally built using the metadata for 2010 file but it may be used for later years.

Note that the loaded object will be of considerable size and may require a computer from 8 GB of RAM. In order to avoid this kind of computational issues, you can also download the splitted version provided by the INE.

Value

`tibble` with all available requested data where each row corresponds to an anonymised citizen.

Note

The `file` parameter allows any flat file with fixed width.

Author(s)

Carlos Neira and Carlos J. Gil Bellosta

References

<http://www.ine.es/prodyser/microdatos.htm>

See Also

[read_fwf](#) to read fixed width files.

Examples

```
# Downloaded \href{ftp://www.ine.es/temas/censopv/cen11/Microdatos_personas_nacional.zip}{data}
## Not run:
raw <- censo2010("MicrodatosCP_NV_per_nacional_3VAR.txt")
summary(raw)

## End(Not run)
```

defun2011

Download data for deaths from INE

Description

This function downloads deaths statistics from the Spanish Statistical Office (INE).

Usage

```
defun2011(file)
```

Arguments

file Character string with the name of the microdata file provided by the INE on the **Deaths** section. It allows either a path to a file or literal data (single string or raw vector). It also allows compressed files in .gz, .bz2, .xz, or .zip format.

Details

This function reads deaths microdata in Spain.

It was originally built using the metadata for the 2011 file and may work with microdata files for other yearly submissions. Previous files have a different, non compatible, format.

Value

[tibble](#) with all available requested data where each row corresponds to an anonymised citizen.

Note

The file parameter allows any flat file with fixed width.

Author(s)

Carlos J. Gil Bellosta

References

<http://www.ine.es/prodyser/microdatos.htm>

See Also

[read_fwf](#) to read fixed width files.

Examples

```
## Not run:
raw <- defun2011("datos_2016.zip") # It will be automatically uncompressed.
summary(raw)

## End(Not run)
```

ees2010

Download data from the Encuesta de Estructura Salarial (EES)

Description

This function downloads data from the Spanish Survey of Salary Structure using the information provided by the Spanish Statistical Office (INE).

Usage

```
ees2010(file)
```

Arguments

file Character string with the name of the microdata file provided by the INE on the **EES** section. It allows either a path to a file, or literal data (single string or raw vector). It also allows compressed files in .gz, .bz2, .xz, or .zip format.

Details

This function reads microdata from the Survey of Salary Structure in Spain. It was originally built using the metadata for 2010 file but it may be used for later years. Previous files have a different, non compatible, format.

Value

[tibble](#) with all available requested data where each row corresponds to an anonymised citizen.

Note

The file parameter allows any flat file with fixed width.

Author(s)

Carlos J. Gil Bellosta, with major contributions from José Luis Cañadas Reche

References

<http://www.ine.es/prodyser/microdatos.htm>

See Also

[read_fwf](#) to read fixed width files.

Examples

```
## Not run:  
raw <- ees2010("datos_2014.zip") # It will be automatically uncompressed.  
summary(raw)  
  
## End(Not run)
```

epa2005

Download data from the Encuesta de Población Activa (EPA)

Description

This function downloads the data from the Spanish Active Population Survey using the information provided by the Spanish Statistical Office (INE).

Usage

```
epa2005(file)
```

Arguments

file	Character string with the name of the microdata file provided by the INE on the EPA section. It allows either a path to a file, or literal data (single string or raw vector). It also allows compressed files in .gz, .bz2, .xz, or .zip format.
------	--

Details

This function reads microdata from the quarterly Active Population Survey in Spain. It was originally built using the metadata for 2005 file but it may be used for later years. Previous files have a different, non compatible, format.

Value

`tibble` with all available requested data where each row corresponds to an anonymised citizen.

Note

The file parameter allows any flat file with fixed width.

Author(s)

Carlos J. Gil Bellosta

References

<http://www.ine.es/prodyser/microdatos.htm>

See Also

`read_fwf` to read fixed width files.

Examples

```
## Not run:  
raw <- epa2005("EPA4T017")  
summary(raw)  
  
## End(Not run)
```

epf.2011.gastos

Download data from the Encuesta de Presupuestos Familiares (EPF)

Description

This function downloads data from the Spanish Survey of Family Budgets (**Expenditure approach**) using the information provided by the Spanish Statistical Office (INE).

Usage

```
epf.2011.gastos(file)
```

Arguments

`file` Character string with the name of the microdata file provided by the INE on the **EPF** section. It allows either a path to a file, or literal data (single string or raw vector). It also allows compressed files in `.gz`, `.bz2`, `.xz`, or `.zip` format.

Details

This function reads microdata from the expenditure approach of the Survey of Family Budgets in Spain. It was originally built using the metadata for 2011 file but it may be used for later years. Previous files have a different, non compatible, format.

Value

`tibble` with all available requested data where each row corresponds to an anonymised citizen.

Note

The `file` parameter allows any flat file with fixed width.

Author(s)

Diego Paniagua Sánchez and Carlos J. Gil Bellosta

References

<http://www.ine.es/prodyser/microdatos.htm>

See Also

`read_fwf` to read fixed width files.

Examples

```
## Not run:  
raw <- epf.2011.gastos("sample_gastos_a2011.txt")  
summary(raw)  
  
## End(Not run)
```

epf.2011.hogares

Download data from the Encuesta de Presupuestos Familiares (EPF)

Description

This function downloads data from the Spanish Survey of Family Budgets (**Homes approach**) using the information provided by the Spanish Statistical Office (INE).

Usage

```
epf.2011.hogares(file)
```

Arguments

`file` Character string with the name of the microdata file provided by the INE on the **EPF** section. It allows either a path to a file, or literal data (single string or raw vector). It also allows compressed files in `.gz`, `.bz2`, `.xz`, or `.zip` format.

Details

This function reads microdata from the homes approach of the Survey of Family Budgets in Spain. It was originally built using the metadata for 2011 file but it may be used for later years. Previous files have a different, non compatible, format.

Value

`tibble` with all available requested data where each row corresponds to an anonymised citizen.

Note

The `file` parameter allows any flat file with fixed width.

Author(s)

Diego Paniagua Sánchez and Carlos J. Gil Bellosta

References

<http://www.ine.es/prodyser/microdatos.htm>

See Also

`read_fwf` to read fixed width files.

Examples

```
## Not run:  
raw <- epf.2011.gastos("sample_hogares_a2011.txt")  
summary(raw)  
  
## End(Not run)
```

epf.2011.miembros

Download data from the Encuesta de Presupuestos Familiares (EPF)

Description

This function downloads data from the Spanish Survey of Family Budgets (**Members approach**) using the information provided by the Spanish Statistical Office (INE).

Usage

```
epf.2011.miembros(file)
```

Arguments

file	Character string with the name of the microdata file provided by the INE on the EPF section. It allows either a path to a file, or literal data (single string or raw vector). It also allows compressed files in .gz, .bz2, .xz, or .zip format.
------	--

Details

This function reads microdata from the members approach of the Survey of Family Budgets in Spain. It was originally built using the metadata for 2011 file but it may be used for later years. Previous files have a different, non compatible, format.

Value

[tibble](#) with all available requested data where each row corresponds to an anonymised citizen.

Note

The file parameter allows any flat file with fixed width.

Author(s)

Diego Paniagua Sánchez and Carlos J. Gil Bellosta

References

<http://www.ine.es/prodyser/microdatos.htm>

See Also

[read_fwf](#) to read fixed width files.

Examples

```
## Not run:  
raw <- epf.2011.gastos("sample_miembros_a2011.txt")  
summary(raw)  
  
## End(Not run)
```

`padron2016`*Download data from the "Padrón"*

Description

This function downloads data from the Spanish "padrón" using the information provided by the Spanish Statistical Office (INE).

Usage

```
padron2016(file)
```

Arguments

`file` Character string with the name of the microdata file provided by the INE on the **Padrón** section. It allows either a path to a file, or literal data (single string or raw vector). It also allows compressed files in `.gz`, `.bz2`, `.xz`, or `.zip` format.

Details

This function reads microdata from the members approach of the "Padrón" in Spain. It was originally built using the metadata for 2016 file but it may be used for later years. Previous files have a different, non compatible, format.

Value

`tibble` with all available requested data where each row corresponds to an anonymised citizen.

Note

The `file` parameter allows any flat file with fixed width.

Author(s)

Carlos J. Gil Bellosta; metadata provided by Fiorella Mori.

References

<http://www.ine.es/prodyser/microdatos.htm>

See Also

`read_fwf` to read fixed width files.

Examples

```
## Not run:
raw <- epf.2011.gastos("sample_padron_2016.txt")
summary(raw)

## End(Not run)
```

test.metadata	<i>Reads microdata using user provided metadata files</i>
---------------	---

Description

This function reads a microdata file based on metadata provided by the user. It is particularly intended for the user to test the metadata before submitting it for inclusion in the package.

Usage

```
test.metadata(file, md.1, md.2, encoding = "UTF-8" )
```

Arguments

file	The name of the microdata file
md.1	First metadata file
md.2	Second metadata file
encoding	Either 'latin1' or 'UTF-8', the encoding used in the metadata files

Details

The function reads the microdata file using two standardized metadata files. For some microdata files, the package provides the required metadata files; however, users may be willing to load other microdata files using the package infrastructure or test metadata files before providing them to the maintainer for inclusion in it.

The first metadata file deals with columns. It is a tab-separated file with five columns. The first one contains the column name. Columns 2-4, the position where the column data starts and ends and its length. The last field is the column label.

The second metadata file deals with the column contents. It is also a tab separated file with five columns. The first one is the column name.

The second one indicates the column type. There are two main types: "D" for dictionary and "N" for numeric. Dictionary columns require a map from keys to values (columns 4 and 5): values read in the key column will be replaced by the value column content. These columns can be left empty for numeric columns. The third column deals with options in case NULL/NA values are found. There are other specialized column types not documented here. Adding new specialized column types requires changing the inner code of other functions.

The third column indicates which code does the column use for NA values (in case, say, NULL values are marked with "."). Occurrences of such code will be replaced with NA in the final table.

Note that the metadata files need to be stored in UTF-8 encoding.

The example section of this document shows the contents of a set of metadata files.

Value

A `data.frame`.

Author(s)

Carlos J. Gil Bellosta

Examples

```
# This command reads a few lines sampled from
# the EPA for 2011Q1
x <- test.metadata(system.file("extdata", "sampleEPA0111.txt", package = "MicroDatosEs"),
                   system.file("metadata", "epa_mdat1.txt", package = "MicroDatosEs"),
                   system.file("metadata", "epa_mdat2.txt", package = "MicroDatosEs"))
# Example of first metadata file
mdat1_example <- read.table(system.file("metadata", "epa_mdat1.txt",
                                       package = "MicroDatosEs"),
                            header = TRUE, sep = "\t",
                            fileEncoding = "UTF-8",
                            stringsAsFactors = FALSE)
head(mdat1_example)

# Example of second metadata file
mdat2_example <- read.table(system.file("metadata", "epa_mdat2.txt",
                                       package = "MicroDatosEs"),
                            header = TRUE, sep = "\t",
                            fileEncoding = "UTF-8",
                            stringsAsFactors = FALSE)

head(mdat2_example)
```

Index

* **manip**

test.metadata, [11](#)

censo2010, [2](#)

data.frame, [12](#)

defun2011, [3](#)

ees2010, [4](#)

epa2005, [5](#)

epf.2011.gastos, [6](#)

epf.2011.hogares, [7](#)

epf.2011.miembros, [8](#)

padron2016, [10](#)

read_fwf, [3–10](#)

test.metadata, [11](#)

tibble, [2–4, 6–10](#)