Package 'essurvey'

December 11, 2019

Title Download Data from the European Social Survey on the Fly

Version 1.0.5

Date 2019-12-11

BugReports https://github.com/ropensci/essurvey/issues

Description Download data from the European Social Survey directly from their website <http://www.europeansocialsurvey.org/>. There are two families of functions that allow you to download and interactively check all countries and rounds available.

Depends R (>= 3.4.0)

License MIT + file LICENSE

URL https://docs.ropensci.org/essurvey/,

https://github.com/ropensci/essurvey

Encoding UTF-8

LazyData true

RoxygenNote 7.0.1

Imports xml2, httr, haven (>= 2.1.0), rvest, tibble, utils

Suggests foreign, testthat (>= 2.1.0), knitr, rmarkdown, covr

VignetteBuilder knitr

NeedsCompilation no

Author Jorge Cimentada [aut, cre],
Thomas Leeper [rev] (Thomas reviewed the package for rOpensci,see
https://github.com/ropensci/onboarding/issues/201),
Nujcharee Haswell [rev] (Nujcharee reviewed the package for rOpensci,
see https://github.com/ropensci/onboarding/issues/201),
Jorge Lopez [ctb],
François Briatte [ctb]

Maintainer Jorge Cimentada <cimentadaj@gmail.com>

Repository CRAN

Date/Publication 2019-12-11 13:30:03 UTC

R **topics documented:**

---

import_country	*Download integrated rounds separately for countries from the European Social Survey*

Description

Download integrated rounds separately for countries from the European Social Survey

Usage

```
import_country(country, rounds, ess_email = NULL, format = NULL)

import_all_cntrounds(country, ess_email = NULL, format = NULL)

download_country(
  country,
  rounds,
  ess_email = NULL,
  output_dir = getwd(),
  format = "stata"
)
```

Arguments

country	a character of length 1 with the full name of the country. Use show_countries for a list of available countries.
rounds	a numeric vector with the rounds to download. See show_rounds for all available rounds.
ess_email	a character vector with your email, such as "your_email@email.com". If you haven't registered in the ESS website, create an account at http://www.europeansocialsurvey.org/user/new. A preferred method is to login through set_email.

format	the format from which to download the data. By default it is NULL for `import_*` functions and tries to read 'stata', 'spss' and 'sas' in the specific order. This can be useful if some countries don't have a particular format available. Alternatively, the user can specify the format which can either be 'stata', 'spss' or 'sas'. For the `download_*` functions it is set to 'stata' because the format should be specified before the downloading. When using `import_country` the data will be downloaded and read in the `format` specified. For `download_country`, the data is downloaded from the specified `format` (only 'spss' and 'stata' supported, see details).
output_dir	a character vector with the output directory in case you want to only download the files using `download_country`. Defaults to your working directory. This will be interpreted as a **directory** and not a path with a file name.

Details

Use `import_country` to download specified rounds for a given country and import them to R. `import_all_cntrounds` will download all rounds for a given country by default and `download_country` will download rounds and save them in a specified `format` in the supplied directory.

The `format` argument from `import_country` should not matter to the user because the data is read into R either way. However, different formats might have different handling of the encoding of some questions. This option was preserved so that the user can switch between formats if any encoding errors are found in the data. For more details see the discussion here. For this particular argument, 'sas' is not supported because the data formats have changed between ESS waves and separate formats require different functions to be read. To preserve parsimony and format errors between waves, the user should use 'spss' or 'stata'.

Value

for `import_country` if `length(rounds)` is 1, it returns a tibble with the latest version of that round. Otherwise it returns a list of `length(rounds)` containing the latest version of each round. For `download_country`, if `output_dir` is a valid directory, it returns the saved directories invisibly and saves all the rounds in the chosen `format` in `output_dir`

Examples

```
## Not run:

set_email("your_email@email.com")

# Get first three rounds for Denmark
dk_three <- import_country("Denmark", 1:3)

# Only download the files, this will return nothing

temp_dir <- tempdir()

download_country(
 "Turkey",
 rounds = c(2, 4),
 output_dir = temp_dir
```

```
)

# By default, download_country downloads 'stata' files but
# you can also download 'spss' or 'sas' files.

download_country(
 "Turkey",
 rounds = c(2, 4),
 output_dir = temp_dir,
 format = 'spss'
)

# If email is not registered at ESS website, error will arise
uk_one <- import_country("United Kingdom", 5, "wrong_email@email.com")
# Error in authenticate(ess_email) :
# The email address you provided is not associated with any registered user.
# Create an account at http://www.europeansocialsurvey.org/user/new

# If selected rounds don't exist, error will arise

czech_two <- import_country("Czech Republic", c(1, 22))

# Error in country_url(country, rounds) :
# Only rounds ESS1, ESS2, ESS4, ESS5, ESS6, ESS7, ESS8 available
# for Czech Republic

## End(Not run)
```

---

import_rounds   *Download integrated rounds from the European Social Survey*

Description

Download integrated rounds from the European Social Survey

Usage

```
import_rounds(rounds, ess_email = NULL, format = NULL)

import_all_rounds(ess_email = NULL, format = NULL)

download_rounds(
  rounds,
  ess_email = NULL,
  output_dir = getwd(),
  format = "stata"
)
```

Arguments

rounds	a numeric vector with the rounds to download. See [show_rounds](#) for all available rounds.
ess_email	a character vector with your email, such as "your_email@email.com". If you haven't registered in the ESS website, create an account at http://www.europeansocialsurvey.org/user/new. A preferred method is to login through [set_email](#).
format	the format from which to download the data. By default it is NULL for import_* functions and tries to read 'stata', 'spss' and 'sas' in the specific order. This can be useful if some countries don't have a particular format available. Alternatively, the user can specify the format which can either be 'stata', 'spss' or 'sas'. For the download_* functions it is set to 'stata' because the format should be specified before downloading. When using import_country the data will be downloaded and read in the format specified. For download_country, the data is downloaded from the specified format (only 'spss' and 'stata' supported, see details).
output_dir	a character vector with the output directory in case you want to only download the files using the download_rounds. Defaults to your working directory. This will be interpreted as a **directory** and not a path with a file name.

Details

Use import_rounds to download specified rounds and import them to R. import_all_rounds will download all rounds by default and download_rounds will download rounds and save them in a specified format in the supplied directory.

The format argument from import_rounds should not matter to the user because the data is read into R either way. However, different formats might have different handling of the encoding of some questions. This option was preserved so that the user can switch between formats if any encoding errors are found in the data. For more details see the discussion here. For this particular argument in, 'sas' is not supported because the data formats have changed between ESS waves and separate formats require different functions to be read. To preserve parsimony and format errors between waves, the user should use 'spss' or 'stata'.

Value

for import_rounds if length(rounds) is 1, it returns a tibble with the latest version of that round. Otherwise it returns a list of length(rounds) containing the latest version of each round. For download_rounds, if output_dir is a valid directory, it returns the saved directories invisibly and saves all the rounds in the chosen format in output_dir

Examples

```
## Not run:

set_email("your_email@email.com")

# Get first three rounds
three_rounds <- import_rounds(1:3)
```

```
temp_dir <- tempdir()

# Only download the files to output_dir, this will return nothing.
download_rounds(
 rounds = 1:3,
 output_dir = temp_dir,
)

# By default, download_rounds saves a 'stata' file. You can
# also download 'spss' and 'sas' files.

download_rounds(
 rounds = 1:3,
 output_dir = temp_dir,
 format = 'spss'
)

# If rounds are repeated, will download only unique ones
two_rounds <- import_rounds(c(1, 1))

# If email is not registered at ESS website, error will arise
two_rounds <- import_rounds(c(1, 2), "wrong_email@email.com")

# Error in authenticate(ess_email) :
# The email address you provided is not associated with any registered user.
# Create an account at https://www.europeansocialsurvey.org/user/new

# If selected rounds don't exist, error will arise

two_rounds <- import_rounds(c(1, 22))
# Error in round_url(rounds) :
# ESS round 22 is not a available. Check show_rounds()

## End(Not run)
```

---

import_sddf_country   *Download SDDF data by round for countries from the European So-*
                        cial Survey

Description

Download SDDF data by round for countries from the European Social Survey

Usage

```
import_sddf_country(country, rounds, ess_email = NULL, format = NULL)

import_all_sddf_cntrounds(country, ess_email = NULL, format = NULL)
```

```
download_sddf_country(
  country,
  rounds,
  ess_email = NULL,
  output_dir = getwd(),
  format = "stata"
)
```

Arguments

country
a character of length 1 with the full name of the country. Use show_countries for a list of available countries.

rounds
a numeric vector with the rounds to download. See show_sddf_cntrounds for all available rounds for any given country.

ess_email
a character vector with your email, such as "your_email@email.com". If you haven't registered in the ESS website, create an account at http://www.europeansocialsurvey.org/user/new. A preferred method is to login through set_email.

format
the format from which to download the data. By default it is NULL for import_* functions and tries to read 'stata', 'spss' and 'sas' in the specific order. This can be useful if some countries don't have a particular format available. Alternatively, the user can specify the format which can either be 'stata', 'spss' or 'sas'. For the download_* functions it is set to 'stata' because the format should be specified before downloading. Setting it to NULL will iterate over 'stata', 'spss' and 'sas' and download the first that is available. When using import_country the data will be downloaded and read in the format specified. For download_country, the data is downloaded from the specified format (only 'spss' and 'stata' supported, see details).

output_dir
a character vector with the output directory in case you want to only download the files using download_sddf_country. Defaults to your working directory. This will be interpreted as a **directory** and not a path with a file name.

Details

SDDF data (Sample Design Data Files) are data sets that contain additional columns with the sample design and weights for a given country in a given round. These additional columns are required to perform any complex weighted analysis of the ESS data. Users interested in using this data should read the description of SDDF files here and should read here for the sampling design of the country of analysis for that specific round.

Use import_sddf_country to download the SDDF data by country into R. import_all_sddf_cntrounds will download all available SDDF data for a given country by default and download_sddf_country will download SDDF data and save them in a specified format in the supplied directory.

The format argument from import_country should not matter to the user because the data is read into R either way. However, different formats might have different handling of the encoding of some questions. This option was preserved so that the user can switch between formats if any encoding errors are found in the data. For more details see the discussion here.

Additionally, given that the SDDF data is not very complete, some countries do not have SDDF data in Stata or SPSS formats. For that reason, the `format` argument is not used in `import_sddf_country`. Internally, `Stata` is chosen over SPSS and SPSS over SAS in that order of preference.

For this particular argument, 'sas' is not supported because the data formats have changed between ESS waves and separate formats require different functions to be read. To preserve parsimony and format errors between waves, the user should use 'stata' or 'spss'.

Starting from round 7 (including), the ESS switched the layout of SDDF data. Before the rounds, SDDF data was published separately by wave-country combination. From round 7 onwards, all SDDF data is released as a single integrated file with all countries combined for that given round. `import_sddf_country` takes care of this nuance by reading the data and filtering the chosen country automatically. `download_sddf_country` downloads the raw file but also reads the data into memory to subset the specific country requested. This process should be transparent to the user but beware that reading/writing the data might delete some of it's properties such as dropping the labels or label attribute.

Value

for `import_sddf_country` if `length(rounds)` is 1, it returns a tibble with the latest version of that round. Otherwise it returns a list of `length(rounds)` containing the latest version of each round. For `download_sddf_country`, if `output_dir` is a valid directory, it returns the saved directories invisibly and saves all the rounds in the chosen `format` in `output_dir`

Examples

```
## Not run:

set_email("your_email@email.com")

sp_three <- import_sddf_country("Spain", 5:6)

show_sddf_cntrounds("Spain")

# Only download the files, this will return nothing

temp_dir <- tempdir()

download_sddf_country(
 "Spain",
 rounds = 5:6,
 output_dir = temp_dir
)

# By default, download_sddf_country downloads 'stata' files but
# you can also download 'spss' or 'sas' files.

download_sddf_country(
 "Spain",
 rounds = 1:8,
 output_dir = temp_dir,
 format = 'spss'
)
```

```
## End(Not run)
```

---

recode_missings   *Recode pre-defined missing values as NA*

Description

This function is not needed any more, please see the details section.

Usage

```
recode_missings(ess_data, missing_codes)

recode_numeric_missing(x, missing_codes)

recode_strings_missing(y, missing_codes)
```

Arguments

ess_data	data frame or [tibble](#) with data from the European Social Survey. This data frame should come either from [import_rounds](#), [import_country](#) or read with [read_dta](#) or [read_spss](#). This is the case because it identifies missing values using [labelled](#) classes.
missing_codes	a character vector with values 'Not applicable', 'Refusal', 'Don't Know', 'No answer' or 'Not available'. By default all values are chosen. Note that the wording is case sensitive.
x	a [labelled](#) numeric
y	a character vector

Details

Data from the European Social Survey is always accompanied by a script that recodes the categories 'Not applicable', 'Refusal', 'Don't Know', 'No answer' and 'Not available' to missing. This function recodes these categories to NA

The European Social Survey now provides these values recoded automatically in Stata data files. These missing categories are now read as missing values by [read_dta](#), reading the missing categories correctly from Stata.For an example on how these values are coded, see [here](#).

Old details:

When downloading data directly from the European Social Survey's website, the downloaded .zip file contains a script that recodes some categories as missings in Stata and SPSS formats.

For recoding numeric variables recode_numeric_missings uses the labels provided by the [labelled](#) class to delete the labels matched in missing_codes. For the character variables matching is done

with the underlying number assigned to each category, namely 6, 7, 8, 9 and 9 for 'Not applicable', Refusal', 'Don't Know', No answer' and 'Not available'.

The functions are a direct translation of the Stata script that comes along when downloading one of the rounds. The Stata script is the same for all rounds and all countries, meaning that these functions work for all rounds.

Value

The same data frame or [tibble](#) but with values 'Not applicable', 'Refusal', 'Don't Know', 'No answer' and 'Not available' recoded as NA.

Examples

```
## Not run:
seven <- import_rounds(7, your_email)

attr(seven$tvtot, "labels")
mean(seven$tvtot, na.rm = TRUE)

names(table(seven$lnghom1))
# First three are actually missing values

seven_recoded <- recode_missings(seven)

attr(seven_recoded$tvtot, "labels")
# All missings have been removed
mean(seven_recoded$tvtot, na.rm = TRUE)

names(table(seven_recoded$lnghom1))
# All missings have been removed

# If you want to operate on specific variables
# you can use other recode_*_missing

seven$tvtot <- recode_numeric_missing(seven$tvtot)

# Recode only 'Don't know' and 'No answer' to missing
seven$tvpol <- recode_numeric_missing(seven$tvpol, c("Don't know", "No answer"))


# The same can be done with recode_strings_missing

## End(Not run)
```

---

set_email   *Save your ESS email as an environment variable*

Description

Save your ESS email as an environment variable

Usage

```
set_email(ess_email)
```

Arguments

ess_email   a character string with your registered email.

Details

You should only run `set_email()` once and every `import_` and `download_` function should work fine. Make sure your email is registered at <http://www.europeansocialsurvey.org/> before setting the email.

Examples

```
## Not run:
set_email("my_registered@email.com")

ess_country(1)

## End(Not run)
```

---

show_countries   *Return available countries in the European Social Survey*

Description

Return available countries in the European Social Survey

Usage

```
show_countries()
```

Value

character vector with available countries

Examples

```
## Not run:
show_countries()

## End(Not run)
```

---

show_country_rounds   *Return available rounds for a country in the European Social Survey*

Description

Return available rounds for a country in the European Social Survey

Usage

```
show_country_rounds(country)
```

Arguments

country   A character of length 1 with the full name of the country. Use [show_countries](#)for
                 a list of available countries.

Value

numeric vector with available rounds for country

Examples

```
## Not run:

show_country_rounds("Spain")

show_country_rounds("Turkey")


## End(Not run)
```

---

show_rounds   *Return available rounds in the European Social Survey*

Description

Return available rounds in the European Social Survey

Usage

```
show_rounds()
```

Value

numeric vector with available rounds

Examples

```
## Not run:
show_rounds()

## End(Not run)
```

---

show_rounds_country   *Return countries that participated in* **all** *of the specified rounds.*

Description

Return countries that participated in **all** of the specified rounds.

Usage

```
show_rounds_country(rounds, participate = TRUE)
```

Arguments

rounds   A numeric vector specifying the rounds from which to return the countries. Use
                [show_rounds](#)for a list of available rounds.

participate   A logical that controls whether to show participating countries in that/those
                rounds or countries that didn't participate. Set to TRUE by default.

Details

show_rounds_country returns the countries that participated in **all** of the specified rounds. That is, show_rounds_country(1:2) will return countries that participated both in round 1 and round 2. Conversely, if participate = FALSE it will return the countries that did not participate in **both** round 1 and round 2.

Value

A character vector with the country names

Examples

```
## Not run:

# Return countries that participated in round 2

show_rounds_country(2)

# Return countries that participated in all rounds

show_rounds_country(1:8)

# Return countries that didn't participate in the first three rounds

show_rounds_country(1:3, participate = FALSE)


## End(Not run)
```

---

show_sddf_cntrounds	*Return available SDDF rounds for a country in the European Social Survey*

Description

Return available SDDF rounds for a country in the European Social Survey

Usage

```
show_sddf_cntrounds(country, ess_email = NULL)
```

Arguments

country   A character of length 1 with the full name of the country. Use show_countries for a list of available countries.

ess_email   a character vector with your email, such as "your_email@email.com". If you haven't registered in the ESS website, create an account at http://www.europeansocialsurvey.org/user/new. A preferred method is to login through set_email.

Details

SDDF data are the equivalent weight data used to analyze the European Social Survey properly. For more information, see the details section of [import_sddf_country](). As an exception to the show_* family of functions, show_sddf rounds needs your ESS email to check which rounds are available. Be sure to add it with [set_email]().

Value

numeric vector with available rounds for `country`

Examples

```
## Not run:
set_email("your_email@email.com")

show_sddf_cntrounds("Spain")

## End(Not run)
```

---

show_themes   *Return available themes in the European Social Survey*

Description

This function returns the available themes in the European Social Survey. However, contrary to [show_countries]() and [show_country_rounds](), themes can not be downloaded as separate datasets. This and [show_theme_rounds]() serve purely for informative purposes.

Usage

```
show_themes()
```

Value

character vector with available themes

Examples

```
## Not run:
show_themes()

## End(Not run)
```

---

show_theme_rounds *Return available rounds for a theme in the European Social Survey*

Description

This function returns the available rounds for any theme from [show_themes](#). However, contrary to [show_country_rounds](#) themes can not be downloaded as separate datasets. This and the [show_themes](#) function serve purely for informative purposes.

Usage

```
show_theme_rounds(theme)
```

Arguments

theme   A character of length 1 with the full name of the theme. Use [show_themes](#)for a
       list of available themes.

Value

numeric vector with available rounds for `country`

Examples

```
## Not run:
chosen_theme <- show_themes()[3]

# In which rounds was the topic of 'Democracy' asked?
show_theme_rounds(chosen_theme)

# And politics?
show_theme_rounds("Politics")


## End(Not run)
```

Index