

Package ‘easyPSID’

February 16, 2020

Title Reading, Formatting, and Organizing the Panel Study of Income Dynamics (PSID)

Version 0.1.1

Description Provides various functions for reading and preparing the Panel Study of Income Dynamics (PSID) for longitudinal analysis, including functions that read the PSID’s fixed width format files directly into R, rename all of the PSID’s longitudinal variables so that recurring variables have consistent names across years, simplify assembling longitudinal datasets from cross sections of the PSID Family Files, and export the resulting PSID files into file formats common among other statistical programming languages ('SAS', 'STATA', and 'SPSS').

Depends R (>= 3.0.1), stringr (>= 1.0.0), LaF (>= 0.6.0), foreign (>= 0.8-67)

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.0.2

NeedsCompilation no

Author Brian Aronson [aut, cre]

Maintainer Brian Aronson <bdaronson@gmail.com>

Repository CRAN

Date/Publication 2020-02-16 21:40:02 UTC

R topics documented:

easyPSID-package	2
convert_from_rds	2
convert_to_rds	3
create_custom_panel	3
create_extract	4
find_description	5
find_name	5
find_years	6

rename_fam_vars	6
rename_ind_vars	7
unzip_all_files	7

Index	9
--------------	----------

easyPSID-package	<i>easyPSID: Reading, Formatting, and Organizing the Panel Study of Income Dynamics (PSID)</i>
------------------	--

Description

Provides various functions for reading and preparing the Panel Study of Income Dynamics (PSID) for longitudinal analysis, including functions that read the PSID's fixed width format files directly into R, rename all of the PSID's longitudinal variables so that recurring variables have consistent names across years, simplify assembling longitudinal datasets from cross sections of the PSID Family Files, and export the resulting PSID files into file formats common among other statistical programming languages ('SAS', 'STATA', and 'SPSS').

Details

This package is designed for use with the PSID's packaged data Family Files, available at <https://simba.isr.umich.edu/>. See easyPSID's readme at <https://github.com/BrianAronson/easyPSID/blob/master/README.md> for a more detailed overview and guide for using this package.

Author(s)

Maintainer: Brian Aronson <bdaronson@gmail.com>

convert_from_rds	<i>Export PSID files to another statistical language</i>
------------------	--

Description

Exports all .rds files in the chosen directory into a common file format used by one of three other statistical programming languages (SPSS, SAS, and STATA). Unlike most alternatives, this function retains all variable labels provided by the PSID.

Usage

```
convert_from_rds(language, in_dir, out_dir)
```

Arguments

language	Language to export PSID .rds files into (options include SPSS, SAS, and STATA)
in_dir	Directory of PSID .rds files to export. Note that large files can take a long time to export.
out_dir	Directory for exported files to be placed

Examples

```

convert_from_rds(
  language="STATA",
  in_direc=system.file("extdata","rds_dir", package = "easyPSID"),
  out_direc=tempdir()
)

```

convert_to_rds	<i>Convert all PSID files from .txt format to .rds format</i>
----------------	---

Description

Converts all PSID fixed width format .txt files in a selected directory into .rds format. Importantly, this function assumes that all files contained in the original PSID .zip files (especially those ending in .do) are present in the same directory as the PSID .txt files, and that all files within that directory have the same names as when first unzipped.

Usage

```
convert_to_rds(in_direc, out_direc)
```

Arguments

in_direc	Directory containing unzipped PSID .txt and .do files
out_direc	Directory to place PSID .rds files into

Examples

```

convert_to_rds(
  in_direc=system.file("extdata","unzip_dir", package = "easyPSID"),
  out_direc=tempdir()
)

```

create_custom_panel	<i>Create custom longitudinal (panel) dataset with PSID Family Files</i>
---------------------	--

Description

Uses the longitudinal PSID Family Files to create a custom longitudinal dataset in long format based on all PSID .rds Family files in a selected directory.

This function can work with data that has been renamed via the `rename_fam_vars` function or data just converted to .rds format via the `convert_to_rds` function. It will create NAs for years when a given variable was not available, and creates a new variable ("Year") to specify the panel of data included in the custom dataset. If a provided variable exists in other waves of the family files under a different name, all waves of that variable will be included in the resulting dataset.

To create a longitudinal family file of the PSID with all variables in the PSID Family Files, it is recommended that one uses the `create_extract` function instead. However, such a file can be very large when using many waves of the PSID. Users with more than five waves of the PSID Family Files are highly recommended to avoid creating a longitudinal dataset with all unique Family File variables.

Usage

```
create_custom_panel(var_names, in_dir, out_dir)
```

Arguments

<code>var_names</code>	Variable names to include in custom longitudinal dataset (as vector of strings)
<code>in_dir</code>	Directory of PSID .rds to use for custom longitudinal dataset
<code>out_dir</code>	Directory to place resulting longitudinal dataset into

Examples

```
create_custom_panel(
  var_names = c("V534", "V442", "V398"),
  in_dir=system.file("extdata", "rds_dir", package = "easyPSID"),
  out_dir=tempdir()
)
```

<code>create_extract</code>	<i>Create subset of common family file variables in long format</i>
-----------------------------	---

Description

Creates an extract dataset in long format consisting of the 500 most frequently reoccurring PSID Family Variables across all supplied waves of the PSID.

Usage

```
create_extract(in_dir, out_dir, num_vars = 500, all_years = F)
```

Arguments

<code>in_dir</code>	Directory containing waves of the Family Files in .rds format
<code>out_dir</code>	Directory to place export file into
<code>num_vars</code>	Number of variables to include in export dataset (default = 500). High variable counts with many waves of data require a significant amount of RAM, and may cause this function to throw errors if a computer's RAM is insufficient
<code>all_years</code>	Select most common variables based on all years of the PSID rather than based in the data actually supplied

Examples

```
create_extract(
  in_direc=system.file("extdata","rds_dir", package = "easyPSID"),
  out_direc=tempdir(),
  num_vars=25,
)
```

find_description	<i>Find description of PSID variable</i>
------------------	--

Description

Finds the descriptions of selected PSID variables.

Usage

```
find_description(variables)
```

Arguments

variables Variable names to look up (as individual string or vector of strings)

Examples

```
find_description(variables=c("V2","V30"))
```

find_name	<i>Lookup new longitudinal variables names</i>
-----------	--

Description

Finds the new name of any longitudinal variable in the PSID Family Files or Individual files following implementation of the rename_fam_vars and rename_ind_vars functions.

Usage

```
find_name(variable, var_year = TRUE)
```

Arguments

variable Variable to look up
var_year Report year when renamed variable first entered the PSID dataset (default=TRUE)

Examples

```
find_name(variable="V1244",var_year=FALSE)
```

find_years	<i>Find years of PSID variable</i>
------------	------------------------------------

Description

Finds the years and corresponding variable names for any longitudinal PSID variable in the PSID Family Files and Individual file.

Usage

```
find_years(variable, var_names = TRUE)
```

Arguments

variable	Variable name to look up
var_names	Report names for longitudinal PSID variable across years (default=TRUE)

Examples

```
find_years(variable="V3", var_names=FALSE)
```

rename_fam_vars	<i>Rename longitudinal Family File variables</i>
-----------------	--

Description

Renames all longitudinal variables in every PSID Family File of a given directory, such that variables are labeled with the variable name used when the variable was first made available in the PSID. For example, the "Release Number" variable was first recorded in the PSID dataset in 1968 as variable "V1" but its name in the 1969 family file is "V441". This program changes the "Release Number" variable name to "V1" in 1968 and all subsequent waves.

Usage

```
rename_fam_vars(in_dir, out_dir)
```

Arguments

in_dir	Directory of PSID .rds files to rename
out_dir	Directory for renamed PSID .rds files to be saved to. Warning: If no directory specified, this function will overwrite the Family Files in the current directory.

Examples

```
rename_fam_vars(
  in_dir=system.file("extdata", "rds_dir", package = "easyPSID"),
  out_dir=tempdir()
)
```

rename_ind_vars	<i>Renames longitudinal Cross-year Individual variables and saves in long format</i>
-----------------	--

Description

Renames all repeated variables in the Cross-year Individual file so that matching variables across waves have the same name, and transforms the resulting dataset into long format. The longitudinal file does not include rows for respondents who were missing in a given wave, and cross-sectional variables are marked as NA during waves when they were not asked. In addition, the resulting file adds two variables for ease of use: "Year" and "fam_id_68".

This function may require up to 8gb of RAM, and will likely throw "cannot allocate memory" errors to users with less RAM on their computer. Users with memory issues should implement the "only_long_vars" or "cust_vars" options.

Usage

```
rename_ind_vars(in_dir, out_dir, only_long_vars = F, cust_vars = NULL)
```

Arguments

in_dir	Directory of PSID Cross-year Individual file .rds file
out_dir	Directory for renamed and transformed PSID Cross-year Individual file to be saved to
only_long_vars	Keep only longitudinal variables in dataset
cust_vars	Custom variables to keep in dataset (as character vector). Output will always include "ER30001", "fam_id_68", and "Year"

Examples

```
rename_ind_vars(
  only_long_vars=TRUE,
  in_dir=system.file("extdata","rds_dir", package = "easyPSID"),
  out_dir=tempdir()
)
```

unzip_all_files	<i>Unzip all PSID files</i>
-----------------	-----------------------------

Description

Unzips all .zip_files files in the specified directory.

Usage

```
unzip_all_files(in_direc, out_direc)
```

Arguments

<code>in_direc</code>	Directory of .zip files to be unzipped
<code>out_direc</code>	Directory for unzipped PSID files to be placed

Examples

```
unzip_all_files(  
  in_direc=system.file("extdata", "zip_dir", package = "easyPSID"),  
  out_direc=tempdir()  
)
```


Index

*Topic **PSID**

- convert_from_rds, [2](#)
- convert_to_rds, [3](#)
- create_custom_panel, [3](#)
- create_extract, [4](#)
- find_description, [5](#)
- find_name, [5](#)
- find_years, [6](#)
- rename_fam_vars, [6](#)
- rename_ind_vars, [7](#)
- unzip_all_files, [7](#)

- convert_from_rds, [2](#)
- convert_to_rds, [3](#)
- create_custom_panel, [3](#)
- create_extract, [4](#)

- easyPSID (easyPSID-package), [2](#)
- easyPSID-package, [2](#)

- find_description, [5](#)
- find_name, [5](#)
- find_years, [6](#)

- rename_fam_vars, [6](#)
- rename_ind_vars, [7](#)

- unzip_all_files, [7](#)