

Package ‘psidread’

October 29, 2025

Title Streamline Building Panel Data from Panel Study of Income Dynamics ('PSID') Raw Files

Version 1.0.6

Description

Streamline the management, creation, and formatting of panel data from the Panel Study of Income Dynamics ('PSID') <<https://psidonline.isr.umich.edu>> using this user-friendly tool. Simply define variable names and input code book details directly from the 'PSID' official website, and this toolbox will efficiently facilitate the data preparation process, transforming raw 'PSID' files into a well-organized format ready for further analysis.

License GPL (>= 3)

Depends R (>= 4.1.0)

Encoding UTF-8

RoxygenNote 7.3.3

URL <https://github.com/Qcrates/psidread>

BugReports <https://github.com/Qcrates/psidread/issues>

Imports asciiSetupReader, dplyr, stringr, tidyr, utils

Suggests knitr, rmarkdown, testthat (>= 3.0.0), DiagrammeR, spelling

Config/testthat/edition 3

VignetteBuilder knitr

Language en-US

NeedsCompilation no

Author Shuyi Qiu [aut, cre] (ORCID: <<https://orcid.org/0009-0000-9730-5262>>)

Maintainer Shuyi Qiu <shuyi.qiu@duke.edu>

Repository CRAN

Date/Publication 2025-10-29 16:00:02 UTC

Contents

psid_read	2
psid_reshape	3
psid_str	4
psid_unzip	6

Index	8
--------------	----------

psid_read	<i>Read PSID data from the packaged data file or customized data file</i>
-----------	---

Description

psid_read() is the core function which enables the user to read variables from multiple packaged PSID data files using just one line of code.

Usage

```
psid_read(indir, str_df, idvars = NA, type, filename = NA)
```

Arguments

indir	A character value of the directory path where the user store the .rda data files. This value should be the same as the exdir in the psid_unzip()
str_df	A data frame of the data structure, generated from the psid_str() function.
idvars	A vector of character values, including the variables that do not change across years. Labelled as "ALL YEAR" in PSID website.
type	The type of data that the user downloaded from PSID. Set to "package" if the user downloaded packaged dataset, "single" if the user downloaded selected data set.
filename	A character value of the name of the file. You can use the filename value you used when you run psid_unzip(), or a filename without any file extension.

Details

This function also offers the option to read a customized single data files with selected variables. It is important to note that psid_read() does not change the original variable names as they are in the source data. To execute it effectively, please make sure that:

- psid_str() has been executed beforehand and the table of data structure has been in the environment.
- psid_unzip() has been executed to prepare the data in .rda format.

Value

A data frame with all the selected variables inside but name unchanged.

Examples

```
# Example 1: Read from multiple package data files (Whole procedure)
psid_varlist = c(" hh_age || [13]ER53017 [17]ER66017",
                 " p_age || [13]ER34204")
str_df <- psid_str(varlist = psid_varlist,
                  type = "separated")
# Below is the file path for the package test data, set this to your own directory
indir <- system.file(package = "psidread","extdata")
psid_read(indir = indir,
          str_df = str_df,
          idvars = c("ER30000"),
          type = "package",
          filename = NA)

# Example 2: Read from your customized data file (Whole procedure)
filename = "J327825.zip"
psid_read(indir = indir,
          str_df = str_df,
          idvars = c("ER30000"),
          type = "single",
          filename = filename)
```

psid_reshape

Rename and reshape your PSID dataset

Description

psid_reshape() serves as the final step in processing your PSID data which helps the user rename and reshape the data frame to produce output in your desired format.

Usage

```
psid_reshape(psid_df, str_df, shape = "wide", level = "individual")
```

Arguments

psid_df	The data frame generated from psid_read() function. The user should not change anything on the data frame.
str_df	The structure data frame generated from psid_str() function.
shape	The shape you would like the data frame to be. "long" if you want to have each line represent the data for each person at each year; "wide" if you want to have each line represent the data for all waves for each person;
level	The level of output. Default value is set to 'individual'. The user can also set this value to 'household' if needed. Deduplication will be performed and leave only the record of household head for each household.

Details

This function offers options in data output at both the household and individual levels. When the user specifies for household-level output, the output will only retain the household head's record for each household. This option will be useful if the user aims to conduct family-level analysis. In contrast, individual-level output includes details for all family members.

Additionally, `psid_reshape()` allows for choosing between wide and long formats. In the wide format, variables are named as `VARNAME_YYYY`.

Value

A data frame with self-defined variable name

Examples

```
# Import dataset use `psid_str()`, `psid_unzip()` and `psid_read()`
psid_varlist = c(" hh_age || [13]ER53017 [17]ER66017", " p_age || [13]ER34204")
str_df <- psid_str(varlist = psid_varlist, type = "separated")
# Below is the file path for the package test data, set this to your own directory
indir <- system.file(package = "psidread", "extdata")
df <- psid_read(indir = indir,
                str_df = str_df,
                idvars = c("ER30000"),
                type = "package",
                filename = NA)
# Example 1: Individual-level output in long format
ind_long_df <- psid_reshape(psid_df = df,
                           str_df = str_df,
                           shape = "long",
                           level = "individual")
# Example 2: Household-level output in wide format
fam_wide_df <- psid_reshape(psid_df = df,
                           str_df = str_df,
                           shape = "wide",
                           level = "household")
```

psid_str

Construct the table of PSID data structure

Description

The `psid_str()` function provides a simplified solution to build a table of data structure for your PSID dataset with your customized variable names.

Usage

```
psid_str(varlist, type = "separated")
```

Arguments

<code>varlist</code>	A vector of string values or a single string value, including user's self-defined variable name and the year and variable code from PSID website.
<code>type</code>	A string value of either "separated" or "integrated", indicating the type of varlist

Details

This function is influenced by the methodology implemented in the `psidtools` package developed by Professor Ulrich Kohler. To utilize it, users only need to provide either a string vector (which is recommended) or a single value (if copying and pasting from their .do file). Minimal additional formatting is needed, and users can easily copy and paste year-variable names directly from the PSID website's codebook.

- This data frame only contains cross-year variables selected by the user to include in their panel dataset. Do not specify ALL-YEAR variables here (e.g., individual's sex, individual's birth order).
- The recommended approach for providing the variable list is to separate them into distinct character values and then wrap them within a vector. For example: `c(" hh_age || [13]ER53017 [17]ER66017", " p_age || [13]ER34204")` `hh_age` and `p_age` are self-defined variable name. It is up to you! The final data output will show this variable name instead of the variable code like `ER34204`. The `[YY]VARCODE` sequence can be found from the code book of PSID. You do not need to make any changes on them. Please ensure proper separation between the user-defined variable name and the `[YY]VARCODE` using `"||"`. Each variable sequence should be placed in a separate string value. If this method is used, users should set the `type` argument to `"separated"`.
- This function also offers an option for users who wish to copy and paste their code directly from Stata. Simply copy and paste your code into a single string value without making any alterations. For instance: `"|| hh_age /// [13]ER53017 [17]ER66017 /// || p_age /// [13]ER34204///"` Note the different position of `"||"` between the two input methods. If you utilize this input method, please specify the `type` argument as `"integrated"`.

The output of this function will be utilized by `psid_read()` for reading the dataset into the environment and by `psid_reshape()` for renaming and reshaping the panel dataset. Therefore, it is recommended to execute this function prior to running other functions within this package.

Value

A data frame of the data structure of your PSID dataset, with each row represents the year and each column represents the variable

Examples

```
# Example 1: Separated string input
psid_varlist = c(" hh_age || [13]ER53017 [17]ER66017",
                " p_age || [13]ER34204")
psid_str(varlist = psid_varlist, type = "separated")

# Example 2: Integrated string input
psid_varlist <- "|| hh_age ///
```

```

[13]ER53017 [17]ER66017 ///
|| p_age ///
[13]ER34204///" # DO NOT CHANGE ANYTHING
psid_str(varlist = psid_varlist, type = "integrated")

```

psid_unzip

*Unzip and transfer the downloaded PSID data files***Description**

This `psid_unzip()` function streamlines the process of transforming ASCII data downloaded from the PSID website to R data files (.rda).

Usage

```
psid_unzip(indir, exdir, zipped = TRUE, type = "package", filename = NA)
```

Arguments

<code>indir</code>	A string value of the directory path where the user store the downloaded data files.
<code>exdir</code>	A string value of the directory path where the user wish to put the generated .rda files.
<code>zipped</code>	A logic value indicating whether the data files are zipped or not.
<code>type</code>	A string value of either "package" or "single", indicating whether the data files are packaged data file or a single customized dataset with only selected variables.
<code>filename</code>	A string value of the name of the single file. Default to be NA, but requires to be specified if the type is "single"

Details

This function executes two primary operations:

- Unzip the zipped data files
- Converts the ASCII file into a .rda format for the reading steps

For optimal functionality, please ensure that you have satisfied the following prerequisites:

- If you are using packaged data files, please do not make any changes to the name of the data files
- If you download the dataset with only selected variables, please choose ASCII Data With SAS Statements as the data output type

The user will only need to execute this function once. If you have already executed this function before and have all the .rda files settled down, you do not have to run this again. This function may take several minutes if you have multiple packaged file to unzip and convert.

Value

.rda data files stored in the specified directory file folder

Examples

```
## Not run:
#' # Example 1: Unzip and convert packaged files
exdir <- tempdir()
indir <- system.file(package = "psidread","extdata") # Define the input directory
psid_unzip(indir = indir, exdir = exdir, zipped = TRUE, type = "package", filename = NA)
# Example 2: Unzip and convert customized single data files
exdir <- tempdir()
indir <- system.file(package = "psidread","extdata") # Define the output directory
filename = "J327825.zip"
psid_unzip(indir = indir, exdir = exdir, zipped = TRUE, type = "single", filename = filename)

## End(Not run)
```

Index

psid_read, [2](#)
psid_reshape, [3](#)
psid_str, [4](#)
psid_unzip, [6](#)