

# Package ‘psidR’

May 15, 2015

**Type** Package

**Title** Build Panel Data Sets from PSID Raw Data

**Version** 1.3

**Date** 2015-05-09

**Author** Florian Oswald

**Maintainer** Florian Oswald <florian.oswald@gmail.com>

**Description** Makes it easy to build panel data in wide format from PSID delivered raw data. Deals with data downloaded and pre-processed by Stata or SAS, or can optionally download directly from the PSID server using the SAScii package. `psidR` takes care of merging data from each wave onto a cross-period index file, so that individuals can be followed over time. The user must specify which years they are interested in, and the PSID variable names (e.g. ER21003) for each year (they differ in each year). There are different panel data designs and sample subsetting criteria implemented (`SRC`, `SEO`, `immigrant` and `latino` samples).

**URL** <https://github.com/floswald/psidR>

**Depends** data.table, RCurl, foreign, SAScii

**License** GPL-3

**Collate** 'build.panel.r' 'makeids.r' 'psidR-package.r'

**Suggests** testthat

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2015-05-15 09:20:10

## R topics documented:

build.panel	2
get.psid	5
make.char	6
makeids	6
psidR	6
testPSID	7

## Index

8

---

build.panel	<i>build.panel: Build PSID panel data set</i>
-------------	---

---

## Description

Builds a panel data set in wide format with id variables pid (unique person identifier) and year from individual PSID family files.

## Usage

```
build.panel(datadir = NULL, fam.vars, ind.vars = NULL, SAScii = FALSE,
  heads.only = FALSE, sample = NULL, design = "balanced",
  verbose = FALSE)
```

## Arguments

datadir	either NULL, in which case saves to tmpdir or path to directory containing family files ("FAMyyyy.xyz") and individual file ("IND2009ER.xyz") in admissible formats .xyz. Admissible are .dta, .csv, .RData, .rda. Please follow naming convention. Only .dta version <= 12 supported.
fam.vars	data.frame of variable to retrieve from family files. Can contain see example for required format.
ind.vars	data.frame of variables to get from individual file. In almost all cases this will be the type of survey weights you want to use. don't include id variables ER30001 and ER30002.
SAScii	logical TRUE if you want to directly download data into Rda format (no dependency on STATA/SAS/SPSS). may take a long time.
heads.only	logical TRUE if user wants current household heads only.
sample	string indicating which sample to select: "SRC" (survey research center), "SEO" (survey for economic opportunity), "immigrant" (immigrant sample), "latino" (Latino family sample). Defaults to NULL, so no subsetting takes place.
design	either character <i>balanced</i> or <i>all</i> or integer. <i>balanced</i> means only individuals who appear in each wave are considered. <i>All</i> means all are taken. An integer value stands for minimum consecutive years of participation, i.e. design=3 means present in at least 3 consecutive waves.
verbose	logical TRUE if you want verbose output.

## Details

takes desired variables from family files for specified years in folder datadir and merges using the id information in IND2011ER.xyz, which must be in the same directory. Note that only one IND file may be present in the directory (each PSID shipping comes with a new IND file). The raw data can be supplied in stata .dta format or it can be directly downloaded from the PSID server to folders datadir or tmpdir. Notice that currently only stata format <= 12 is supported (so do saveold

in stata). The user can change subsetting criteria as well as sample designs. Merge: the variables interview number in each family file map to the interview number variable of a given year in the individual file. Run `example(build.panel)` for a demonstration. Accepted input data are stata format .dta, .csv files or R data formats .rda and RData. Similar in usage to stata module `psiduse`.

## Value

<code>data</code>	resulting <code>data.table</code> . the variable <code>pid</code> is the unique person identifier, constructed from <code>ID1968</code> and <code>pnum</code> .
<code>dict</code>	data dictionary if stata data was supplied, <code>NULL</code> else

## Examples

```
## Not run:
# specify variables from family files you want

the family files dataframe can contain NAs.
E.g. if there are years where a variable is missing
and you want to fix that later on somehow.
famvars <- data.frame(year=c(2001,2003),
                      house.value=c("ER17044","ER21043"),
                      total.income=c("ER20456","ER24099"),
                      education=c("ER20457",NA))

# specify variables from individual index file
# these cannot contain NAs at the moment.

indvars = data.frame(year=c(2001,2003),
                     longitud.wgt=c("ER33637","ER33740"))

# call builder
# mydir is a directory that contains FAM2001ER.dta,
# FAM2003ER.dta and IND2011ER.dta

# default
d <- build.panel(datadir=mydir,
                fam.vars=famvars,
                ind.vars=indvars)

# also non-heads
d <- build.panel(datadir=mydir,
                fam.vars=famvars,
                ind.vars=indvars,
                heads.only=FALSE)

# non-balanced panel design
d <- build.panel(datadir=mydir,
                fam.vars=famvars,
                ind.vars=indvars,
                heads.only=FALSE,
                design=2) # keep if stay 2+ periods
```

```

# subset the sample to "latino" only
d <- build.panel(datadir=mydir,
                 fam.vars=famvars,
                 ind.vars=indvars,
                 sample="latino")

## End(Not run)

# #####
# reproducible example on artificial data.
# run this with example(build.panel).
# #####

## make reproducible family data sets for 2 years
## variables are: family income (Money) and age

## Data acquisition step: you download data or
## run build.panel with sascii=TRUE

# testPSID creates artificial PSID data
td <- testPSID(N=12,N.attr=0)
fam1985 <- copy(td$famvars1985)
fam1986 <- copy(td$famvars1986)
IND2009ER <- copy(td$IND2009ER)

# create a temporary datadir
my.dir <- tempdir()
#save those in the datadir
# notice different R formats admissible
save(fam1985,file=paste0(my.dir,"/FAM1985ER.rda"))
save(fam1986,file=paste0(my.dir,"/FAM1986ER.RData"))
save(IND2009ER,file=paste0(my.dir,"/IND2009ER.RData"))

## end Data acquisition step.

# now define which famvars
famvars <- data.frame(year=c(1985,1986),
                      money=c("Money85","Money86"),
                      age=c("age85","age86"))

# create ind.vars
indvars <- data.frame(year=c(1985,1986),ind.weight=c("ER30497","ER30534"))

# call the builder
# data will contain column "relation.head" holding the relationship code.

d <- build.panel(datadir=my.dir,fam.vars=famvars,
                 ind.vars=indvars,
                 heads.only=FALSE,verbose=TRUE)

# see what happens if we drop non-heads
# only the ones who are heads in BOTH years
# are present (since design='balanced' by default)

```

```
d <- build.panel(datadir=my.dir,fam.vars=famvars,
                 ind.vars=indvars,
                 heads.only=TRUE,verbose=FALSE)
print(d$data[order(pid)],nrow=Inf)

# change sample design to "all":
# we'll keep individuals if they are head in one year,
# and drop in the other
d <- build.panel(datadir=my.dir,fam.vars=famvars,
                 ind.vars=indvars,heads.only=TRUE,
                 verbose=FALSE,design="all")
print(d$data[order(pid)],nrow=Inf)

file.remove(paste0(my.dir,"/FAM1985ER.rda"),
            paste0(my.dir,"/FAM1986ER.RData"),
            paste0(my.dir,"/IND2009ER.RData"))

# END psidR example

# #####
# Please go to https://github.com/floswald/psidR for more example usage
# #####
```

---

get.psid

*get.psid connects to PSID database and downloads into Rda*


---

## Description

see <http://www.asdfree.com/> for other usage and <http://stackoverflow.com/questions/15853204/how-to-login-and-then-download-a-file-from-asp-web-pages-with-r>

## Usage

```
get.psid(file, name, params, curl)
```

## Arguments

file	string psid file number
name	string of filename on disc
params	postFormRCurl parameters
curl	postFormRCurl curl handle

## Author(s)

Anthony Damico <ajdamico@gmail.com>

---

<code>make.char</code>	<i>Convert factor to character</i>
------------------------	------------------------------------

---

**Description**

helper function to convert factor to character in a `data.table`

**Usage**

```
make.char(x)
```

**Arguments**

<code>x</code>	a factor
----------------	----------

**Value**

a character

---

<code>makeids</code>	<i>ID list for merging PSID</i>
----------------------	---------------------------------

---

**Description**

this list is taken from <http://ideas.repec.org/c/boc/bocode/s457040.html>

**Usage**

```
makeids()
```

**Details**

this function hardcodes the PSID variable names of "interview number" from both family and individual file for each wave, as well as "sequence number", "relation to head" and numeric value `x` of that variable such that "relation to head" == `x` means the individual is the head. Varies over time.

---

<code>psidR</code>	<i>psidR</i>
--------------------	--------------

---

**Description**

`psidR` is a package that helps the task of building longitudinal datasets from the Panel Study of Income Dynamics (PSID). The user must supply the PSID variable names that correspond to the variables of interest in each desired wave. The data may be in `.dta`, `.csv` format on disk. Creation of `.dta` or `.csv` datasets requires access to Stata or SAS software. There is an option to bypass this requirement by directly downloading the data from the server into a `data.frame`.

---

testPSID	<i>Create a test PSID dataset</i>
----------	-----------------------------------

---

**Description**

makes artificial PSID data with variables age and income for two consecutive years 1985 and 1986.

**Usage**

```
testPSID(N = 100, N.attr = 0)
```

**Arguments**

N	number of people in each wave
N.attr	number of people lost to attrition

**Value**

list with (fake) individual index file IND2009ER and family files for 1985 and 1986

# Index

`build.panel`, [2](#)

`get.psid`, [5](#)

`make.char`, [6](#)

`makeids`, [6](#)

`psidR`, [6](#)

`psidR-package (psidR)`, [6](#)

`testPSID`, [7](#)