Package: yowie (via r-universe)

July 5, 2026

Title Longitudinal Wages Data from the National Longitudinal Survey of Youth 1979

Version 0.1.1

Description Longitudinal wages datasets and several demographic variables from the National Longitudinal Survey of Youth from 1979 to 2018. There are three data sets in this package: The wages data from the NLSY79 cohort; The wages data of the high school dropouts and; The demographic data of NLSY79 cohort.

License GPL (>= 3)

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

URL https://github.com/numbats/yowie

BugReports https://github.com/numbats/yowie/issues

Depends R (>= 3.5.0), tsibble (>= 1.0.1)

Suggests ggplot2, knitr, markdown, rmarkdown, tidyr, forcats, janitor, kableExtra

VignetteBuilder rmarkdown

Imports magrittr, dplyr

Repository https://emitanaka.r-universe.dev

Date/Publication 2022-06-07 04:14:23 UTC

RemoteUrl https://github.com/numbats/yowie

RemoteRef HEAD

RemoteSha ae7c4316ae9f2b2f337e67aafde4b2108aee8642

Contents

`adj_inflation` *A Function to get inflation-adjusted wages.*

Description

This function aims to adjust the `wage` column in `wages` and `wages_hs_do` with the inflation rate of the base year of your choice. The inflation rate is calculated from the CPI data published by the US Bureau of Labor Statistics (BLS). This function needs internet connection to download the data from the BLS website. The code is adapted from Kris Eberwein in R Bloggers website. https://www.r-bloggers.com/2015/12/calculate-inflation-with-r/

Usage

```
adj_inflation(base_year)
```

Arguments

`base_year`	The base year of your choice (4 digits-year without quotation mark).

Value

wages and wages_hs_do data with a new column of inflation-adjusted wages value.

Examples

```
## Not run:
   inflation_adjust(1990)

## End(Not run)
```

demog_nlsy79 — *Demographic Data from the National Longitudinal Survey of Youth (NLSY79)*

Description

A data set contains demographic information of the NLSY79 cohort. The cohort included 12,686 respondents ages 14-22 when first interviewed in 1979.

Format

A data frame contains 12,686 rows and 8 variables:

id A unique individual's ID number.

age_1979 The age of the subject in 1979.

sex Sex of the subject: f = Female and m = Male.

race Race of the subject: NBH = Non-Black,Non-Hispanic; H = Hispanic; B = Black.

hgc Highest grade completed.

hgc_i Integer of highest grade completed.

hgc_1979 The highest grade completed in 1979 (integer value).

ged Wether the respondent had a high school diploma or Graduate Equivalency Degree (GED). 1: High school diploma; 2: GED; 3: Both

Source

The U.S. Bureau of Labor Statistics. (2021, January 6). *National Longitudinal Survey of Youth 1979*. https://www.nlsinfo.org/content/cohorts/nlsy79/get-data

wages — *Wages Data from the National Longitudinal Survey of Youth (NLSY79)*

Description

A data set contains longitudinal data of mean hourly wages along with several demographic variables of Americans from the National Longitudinal Survey of Youth (NLSY79) held by the U.S. Bureau of Labor Statistics from Round 1 (1979 survey year) to Round 28 (2018 survey year). The cohort provided in this data set of the NLSY79 cohort who participated in at least 3 survey rounds.

Format

A tsibble with 206,795 rows and 18 variables:

id A unique individual's ID number. This is the `key` of the data.

year The year the observation was taken. This could be the `index` of the data.

wage The mean of the hourly wages the individual gets at each of their different jobs. The value could be a weighted or an arithmetic mean. The weighted mean is used when the information of hours of work as the weight is available. The mean hourly wage could also be a predicted value if the original value is considered influential by the robust linear regression as part of data cleaning.

age_1979 The age of the subject in 1979.

sex Sex of the subject: f = Female and m = Male.

race Race of the subject: NBH = Non-Black,Non-Hispanic; H = Hispanic; B = Black.

grade Integer value of the highest grade completed corresponding to `year`.

hgc Highest grade completed.

hgc_i Integer of highest grade completed.

hgc_1979 The highest grade completed in 1979 (integer value).

ged Wether the respondent had a high school diploma or Graduate Equivalency Degree (GED). 1: High school diploma; 2: GED; 3: Both

njobs Number of jobs that an individual has.

hours The total number of hours the individual usually works per week.

stwork The year when the individual starting to work.

yr_workforce The length of time in the workforce in years (year - stwork).

exp Work experience, i.e., the number of years worked

is_wm Whether the mean hourly wage is weighted mean, using the hour work as the weight, or regular/arithmetic mean. TRUE = is weighted mean. FALSE = is regular mean.

is_pred Whether the mean hourly wage is a predicted value or not.

Source

The U.S. Bureau of Labor Statistics. (2021, January 6). *National Longitudinal Survey of Youth 1979*. https://www.nlsinfo.org/content/cohorts/nlsy79/get-data

Examples

```
# data summary
wages

library(ggplot2)
library(dplyr)
library(tsibble)
wages_ids <- key_data(wages) %>% select(id)
wages %>%
  dplyr::filter(id %in% sample_n(wages_ids, 10)$id) %>%
```

```
ggplot() +
  geom_line(aes(x = year,
                y = wage,
                group = id), alpha = 0.8)
```

---

wages_hs_do *Wages Data of High School Dropout from the National Longitudinal Survey of Youth (NLSY79)*

Description

A data set contains longitudinal data of mean hourly wages along with several demographic variables from the National Longitudinal Survey of Youth (NLSY79) held by the U.S. Bureau of Labor Statistics from Round 1 (1979 survey year) to Round 28 (2018 survey year). The cohort provided in this data set is high school dropouts in NLSY79 cohort and participated in at least 3 survey rounds.

Format

A tsibble contains 14,174 rows and 18 variables:

#'

id A unique individual's ID number. This is the `key` of the data.

year The year the observation was taken. This could be the `index` of the data.

wage The mean of the hourly wages the individual gets at each of their different jobs. The value could be a weighted or an arithmetic mean. The weighted mean is used when the information of hours of work as the weight is available. The mean hourly wage could also be a predicted value if the original value is considered influential by the robust linear regression as part of data cleaning.

age_1979 The age of the subject in 1979.

sex Sex of the subject: f = Female and m = Male.

race Race of the subject: NBH = Non-Black,Non-Hispanic; H = Hispanic; B = Black.

grade Integer value of the highest grade completed corresponding to `year`.

hgc Highest grade completed.

hgc_i Integer of highest grade completed.

hgc_1979 The highest grade completed in 1979 (integer value).

ged Wether the respondent had a high school diploma or Graduate Equivalency Degree (GED). 1: High school diploma; 2: GED; 3: Both

njobs Number of jobs that an individual has.

hours The total number of hours the individual usually works per week.

stwork The year when the individual starting to work.

yr_workforce The length of time in the workforce in years (year - stwork).

exp Work experience, i.e., the number of years worked

is_wm Whether the mean hourly wage is weighted mean, using the hour work as the weight, or regular/arithmetic mean. TRUE = is weighted mean. FALSE = is regular mean.

is_pred Whether the mean hourly wage is a predicted value or not.

Source

The U.S. Bureau of Labor Statistics. (2021, January 6). *National Longitudinal Survey of Youth 1979*. https://www.nlsinfo.org/content/cohorts/nlsy79/get-data

Examples

```
# saving the data into a new object
library(yowie)
library(tsibble)
wages_hs_do %>% key_data()
```

Index