# Package 'RNHANES'

October 12, 2022

**Type** Package

**Title** Facilitates Analysis of CDC NHANES Data

**Version** 1.1.0

**Date** 2016-11-28

**URL** http://github.com/silentspringinstitute/RNHANES

**BugReports** https://github.com/silentspringinstitute/RNHANES/issues

**Description** Tools for downloading and analyzing CDC NHANES data, with a focus on analytical laboratory data.

**License** Apache License 2.0 | file LICENSE

**LazyData** TRUE

**Depends** R (>= 2.10)

**Imports** foreign, survey, rvest, xml2, methods, dplyr

**Suggests** testthat, knitr, rmarkdown

**VignetteBuilder** knitr

**RoxygenNote** 5.0.1

**NeedsCompilation** no

**Author** Herb Susmann [cre, aut], Silent Spring Institute [cph]

**Maintainer** Herb Susmann <susmann@silentspring.org>

**Repository** CRAN

**Date/Publication** 2016-11-29 02:45:46

## R topics documented:

---

`RNHANES-package` *RNHANES simplifies downloading and analyzing NHANES data.*

---

## Description

RNHANES simplifies downloading and analyzing NHANES data.

---

`demography_filename` *Translates cycle years into the correct demography filename suffix, e.g. '2001-2002' returns 'B'*

---

## Description

Translates cycle years into the correct demography filename suffix, e.g. '2001-2002' returns 'B'

## Usage

```
demography_filename(year)
```

## Arguments

| | |
|---|---|
| `year` | NHANES cycle, e.g. "2001-2002" |

## Value

suffix character e.g. "B"

## download_nhanes_file *Download an NHANES data file from a given cycle*

### Description

Download an NHANES data file from a given cycle

### Usage

```
download_nhanes_file(file_name, year, destination = tempdir(), cache = TRUE)
```

### Arguments

| | |
|---|---|
| file_name | file name |
| year | NHANES cycle |
| destination | directory to download the file into |
| cache | whether to cache the file |

### Value

path to the downloaded file

## file_suffix *Returns the NHANES file suffix for the given year*

### Description

Returns the NHANES file suffix for the given year

### Usage

```
file_suffix(year)
```

### Arguments

| | |
|---|---|
| year | NHANES cycle year (e.g. "2001-2002") |

### Value

suffix character (e.g. "B" or "C")

load_nhanes_description *Download an NHANES description file*

## Description

Download an NHANES description file

## Usage

```
load_nhanes_description(file_name, year, destination = tempdir(),
  cache = FALSE)
```

## Arguments

| | |
|---|---|
| file_name | file name |
| year | NHANES cycle |
| destination | directory to download the file into |
| cache | whether to cache the file |

## Value

data frame containing the file description

nhanes_analyze *Compute quantiles from NHANES weighted survey data*

## Description

Compute quantiles from NHANES weighted survey data

## Usage

```
nhanes_analyze(analysis_fun, nhanes_data, column, comment_column = "",
  weights_column = "", filter = NULL)
```

## Arguments

| | |
|---|---|
| analysis_fun | function to use to analyze each variable |
| nhanes_data | data frame containing NHANES data |
| column | column name of the variable to compute quantiles for |
| comment_column | comment column name of the variable |
| weights_column | name of the weights column |
| filter | logical expression used to subset the data |

### Value

a data frame

---

`nhanes_cycle_years` *List the valid NHANES cycle years*

---

### Description

List the valid NHANES cycle years

### Usage

```
nhanes_cycle_years()
```

### Value

vector of NHANES cycle years

---

`nhanes_data_files` *List the NHANES data files*

---

### Description

List the NHANES data files

### Usage

```
nhanes_data_files(components = "all", destination = tempfile(),
  cache = TRUE)
```

### Arguments

| | |
|---|---|
| `components` | one of "all", "demographics", "dietary", "examination", "laboratory", "questionnaire" |
| `destination` | destinatino to save the file lists |
| `cache` | whether to cache the downloaded file lists so they don't have to be re-downloaded every time |

### Value

data frame of NHANES data files available to download

## Examples

```
## Not run:

# Download a data frame of all the NHANES data files
files <- nhanes_data_files()

# Download a data frame of just the laboratory files
lab_files <- nhanes_data_files(component = "laboratory")


## End(Not run)
```

---

`nhanes_detection_frequency` *Compute detection frequencies of NHANES data*

---

## Description

Compute detection frequencies of NHANES data

## Usage

```
nhanes_detection_frequency(nhanes_data, column, comment_column,
  weights_column = "", filter = NULL)
```

## Arguments

| | |
|---|---|
| `nhanes_data` | data frame containing NHANES data |
| `column` | column names of the variables to compute detection frequencies for |
| `comment_column` | comment column names of the variables to compute detection frequencies for |
| `weights_column` | sample weight column |
| `filter` | logical expression used to subset the data |

## Value

named vector of detection frequencies

## Examples

```
## Not run:
dat <- nhanes_load_data("UHG_G", "2011-2012", demographics = TRUE)

# Compute detection frequency
nhanes_detection_frequency(dat, c("URXUHG"), c("URDUHGLC"))

## End(Not run)
```

nhanes_hist *Plot a weighted histogram of an NHANES variable*

## Description

Plot a weighted histogram of an NHANES variable

## Usage

```
nhanes_hist(nhanes_data, column, comment_column, weights_column = "",
  filter = "", transform = "", ...)
```

## Arguments

| | |
|---|---|
| nhanes_data | data frame containing NHANES data |
| column | column name of the variable to plot |
| comment_column | comment column of the variable to plot |
| weights_column | name of the weights column |
| filter | logical expression used to subset the data |
| transform | transformation to apply to the column. Accepts any function name, for example: "log" |
| ... | parameters passed through to svyhist function |

## Value

a data frame

## Examples

```
## Not run:
dat <- nhanes_load_data("PFC_G", "2011-2012", demographics = TRUE)

nhanes_hist(dat, "LBXPFOA")

## End(Not run)
```

nhanes_load_data *Download NHANES data files.*

## Description

Download NHANES data files.

## Usage

```
nhanes_load_data(file_name, year, destination = tempdir(),
  demographics = FALSE, cache = TRUE, recode = FALSE,
  recode_data = FALSE, recode_demographics = FALSE,
  allow_duplicate_files = FALSE)
```

## Arguments

| | |
|---|---|
| file_name | NHANES file name (e.g. "EPH") or a vector of filenames (e.g c("EPH", "GHB")) |
| year | NHANES cycle year (e.g. "2007-2008") or a vector of cycle years |
| destination | directory to download the files to |
| demographics | include demographics data into the dataset |
| cache | whether to cache the file to disk |
| recode | whether to recode the data and demographics (overrides other parameters) |
| recode_data | whether to recode just the data |
| recode_demographics | whether to recode just the demographics |
| allow_duplicate_files | how to handle a request that has duplicate file names/cycle years. By default duplicates will be removed. |

## Details

If you supply vectors for both file_name and year, then the vectors are paired and each file_name/year pair is downloaded. For example, file_name = c("EPH, GHB"), year = c("2009-2010", "2011-2012") will download "EPH_F.XPT" and "EPH_G.XPT". In other words, the function does not download every possible combination of file_name and year.

You can specify file names in several formats. In order of specificity: You can supply the complete filename: "EPH_F.XPT" You can supply the filename without an extension: "EPH_F" You can supply the filename without a suffix: "EPH", year = "2009-2010"

If you are loading the same file across multiple years, you must supply the filename without a suffix so that the correct suffix for each year can be used.

This function returns either a list or a data frame. If you load multiple files, the return value will always be a list. This is because the columns may not match in between files. If you load one file, the result will be a data frame.

## Value

if file_name or year is a vector, returns a list containing a data frame for each file_name. If file_name and year are both singletons, then a data frame is returned.

## Examples

```
## Not run:

nhanes_load_data("UHG", "2011-2012")

# Load data with demographics
nhanes_load_data("UHG", "2011-2012", demographics = TRUE)

# Download to /tmp directory and overwrite the file if it already exists
nhanes_load_data("HDL_E", "2007-2008", destination = "/tmp", cache = FALSE)

## End(Not run)
```

---

`nhanes_load_demography_data` *Download NHANES demography files for a specific cycle.*

---

## Description

Download NHANES demography files for a specific cycle.

## Usage

```
nhanes_load_demography_data(year, destination = tempdir(), cache = FALSE)
```

## Arguments

| | |
|---|---|
| `year` | NHANES cycle year (e.g. "2011-2012") |
| `destination` | directory to download the file to |
| `cache` | whether load the file if it already exists on disk |

## Examples

```
## Not run:
nhanes_load_demography_data("2011-2012")

## End(Not run)
```

nhanes_quantile *Compute quantiles from NHANES weighted survey data*

## Description

Compute quantiles from NHANES weighted survey data

## Usage

```
nhanes_quantile(nhanes_data, column, comment_column = "",
  weights_column = "", quantiles = seq(0, 1, 0.25), filter = NULL)
```

## Arguments

| | |
|---|---|
| nhanes_data | data frame containing NHANES data |
| column | column name of the variable to compute quantiles for |
| comment_column | comment column name of the variable for checking if computed quantiles are below the LOD |
| weights_column | name of the weights column |
| quantiles | numeric or vector numeric of quantiles to compute |
| filter | logical expression used to subset the data |

## Value

a data frame

## Examples

```
## Not run:
dat <- nhanes_load_data("UHG_G", "2011-2012", demographics = TRUE)

# Compute 50th, 95th, and 99th quantiles
nhanes_quantile(dat, "URXUHG", "URDUHGLC", "WTSA2YR", c(0.5, 0.95, 0.99))

## End(Not run)
```

## nhanes_sample_size *Compute the sample size of NHANES data*

### Description

Compute the sample size of NHANES data

### Usage

```
nhanes_sample_size(nhanes_data, column, comment_column = "",
  weights_column = "", filter = NULL)
```

### Arguments

| | |
|---|---|
| nhanes_data | data frame containing NHANES data |
| column | column name of the variable to compute quantiles for |
| comment_column | comment column name of the variable for checking if computed quantiles are below the LOD |
| weights_column | name of the weights column |
| filter | logical expression used to subset the data |

### Value

a data frame

### Examples

```
## Not run:
dat <- nhanes_load_data("UHG_G", "2011-2012", demographics = TRUE)

nhanes_sample_size(dat, "URXUHG", "URDUHGLC")

## End(Not run)
```

## nhanes_search *Search the results from nhanes_variables or nhanes_data_files*

### Description

Search the results from nhanes_variables or nhanes_data_files

### Usage

```
nhanes_search(nhanes_data, query, ..., fuzzy = FALSE, ignore_case = TRUE,
  max_distance = 0.2)
```

### Arguments

| | |
|---|---|
| nhanes_data | nhanes variable list, from nhanes_variables function, or data file list, from nhanes_data_files |
| query | regular expression search query |
| ... | additional arguments to pass to dplyr::filter |
| fuzzy | whether to use fuzzy string matching for search (based on edit distances) |
| ignore_case | whether search query is case-sensitive |
| max_distance | parameter for tuning fuzzy string matching, 0-1 |

### Value

data frame filtered by search query

### Examples

```
## Not run:
nhanes_files <- nhanes_data_files()

# Search for data files about pesticides
nhanes_search(nhanes_files, "pesticides")

## End(Not run)
```

---

## nhanes_survey *Apply a function from the survey package to NHANES data*

---

### Description

Apply a function from the survey package to NHANES data

### Usage

```
nhanes_survey(survey_fun, nhanes_data, column, comment_column = "",
  weights_column = "", filter = NULL, analyze = "values",
  callback = NULL, ...)
```

## Arguments

| | |
|---|---|
| survey_fun | the survey package function (e.g. svyquantile or svymean) |
| nhanes_data | data frame containing NHANES data |
| column | column name of the variable to compute quantiles for |
| comment_column | comment column name of the variable |
| weights_column | name of the weights column |
| filter | logical expression used to subset the data |
| analyze | one of "values" or "comments", whether to apply the survey function to the value or comment column. |
| callback | optional function to execute on each row of the dataframe |
| ... | other arguments to pass to the survey function |

## Details

This function provides a generic way to apply any function from the survey package to NHANES data. RNHANES provides specific wrappers for computing quantiles (nhanes_quantile) and detection frequencies (nhanes_detection_frequency), and this function provides a general way to use any survey function.

## Value

a data frame

## Examples

```
## Not run:
library(survey)

nhanes_data <- nhanes_load_data("EPH", "2011-2012", demographics = TRUE)

# Compute the mean of triclosan using the svymean function
nhanes_survey(svymean, nhanes_data, "URXTRS", "URDTRSLC", na.rm = TRUE)

# Compute the variance using svyvar
nhanes_survey(svyvar, nhanes_data, "URXTRS", "URDTRSLC", na.rm = TRUE)


## End(Not run)
```

## nhanes_survey_design *Build survey objects for NHANES data*

### Description

Build survey objects for NHANES data

### Usage

```
nhanes_survey_design(nhanes_data, weights_column = "")
```

### Arguments

| | |
|---|---|
| nhanes_data | data frame containing NHANES data |
| weights_column | name of the weights column |

### Value

a survey design object

### Examples

```
## Not run:
dat <- nhanes_load_data("UHG_G", "2011-2012", demographics = TRUE)

design <- nhanes_survey_design(dat, "WTSA2YR")

svymean(~RIDAGEYR, design)

svyglm(URXUHG ~ RIDAGEYR + RIAGENDR, design)

## End(Not run)
```

## nhanes_variables *Load the NHANES comprehensive variable list*

### Description

Load the NHANES comprehensive variable list

### Usage

```
nhanes_variables(components = "all", destination = tempfile(),
  cache = TRUE)
```

### Arguments

| | |
|---|---|
| components | one of "all", "demographics", "dietary", "examination", "laboratory", "questionnaire" |
| destination | where to save the variable list |
| cache | whether to cache the downloaded variable list so it doesn't have to be re-downloaded every time |
| | Helper function for nhanes_variables function |

### Value

dat

### Examples

```
## Not run:

# Download the comprehensive NHANES variable list
variables <- nhanes_variables()

# Download the variable list and cache it in a specific file
variables <- nhanes_variables(destination = "./nhanes_data")


## End(Not run)
```

---

## nhanes_vcov *Extract variance/covariance matrix from parameters of svymean*

---

### Description

Extract variance/covariance matrix from parameters of svymean

### Usage

```
nhanes_vcov(nhanes_data, columns, weights_column = "", filter = "")
```

### Arguments

| | |
|---|---|
| nhanes_data | data frame containing NHANES data |
| columns | columns to include in svymean for |
| weights_column | name of the weights column |
| filter | logical expression used to subset the data |

### Value

a data frame

## Examples

```
## Not run:
dat <- nhanes_load_data("PFC_G", "2011-2012", demographics = TRUE)

nhanes_vcov(dat, c("LBXPFOA", "LBXPFOS"))

## End(Not run)
```

---

`process_file_name` *Processes a file name to make sure it is valid and has the correct suffix and extension File names with an extension (e.g. ".XPT") are not altered*

---

## Description

Processes a file name to make sure it is valid and has the correct suffix and extension File names with an extension (e.g. ".XPT") are not altered

## Usage

```
process_file_name(file_name, year, extension = ".XPT")
```

## Arguments

| | |
|---|---|
| `file_name` | name of the file |
| `year` | NHANES cycle year |
| `extension` | file extension |

---

`validate_year` *Check that the year is in the correct format e.g. '2001-2002' is correct and returns TRUE, '2001' is not correct and returns FALSE*

---

## Description

Check that the year is in the correct format e.g. '2001-2002' is correct and returns TRUE, '2001' is not correct and returns FALSE

## Usage

```
validate_year(year, throw_error = TRUE)
```

## Arguments

| | |
|---|---|
| `year` | the year or years to validate |
| `throw_error` | whether to throw an error if the year is invalid |

# Index