Package 'nhanesA'

January 10, 2025

Version 1.3

Date 2025-01-08

Title NHANES Data Retrieval

BugReports https://github.com/cjendres1/nhanes/issues

Depends R (>= 4.1.0), methods

Imports stringr, foreign, rvest, xml2, dplyr, plyr

Description Utility to retrieve data from the National Health and Nutrition Examination Survey (NHANES) website <https://www.cdc.gov/nchs/nhanes/>.

License GPL (>= 2)

Encoding UTF-8

URL https://cran.r-project.org/package=nhanesA

Suggests knitr, DBI, odbc, RMariaDB, RPostgres, rmarkdown, survey, ggplot2

VignetteBuilder knitr

RoxygenNote 7.3.2

NeedsCompilation no

Author Christopher Endres [aut, cre], Laha Ale [aut] (<https://orcid.org/0000-0002-4070-5289>), Robert Gentleman [aut], Deepayan Sarkar [aut]

Maintainer Christopher Endres <cjendres1@gmail.com>

Repository CRAN

Date/Publication 2025-01-10 15:10:02 UTC

Contents

`browseNHANES` *Open a browser to NHANES.*

Description

The browser may be directed to a specific year, survey, or table.

Usage

```
browseNHANES(
  year = NULL,
  data_group = NULL,
  nh_table = NULL,
  local = TRUE,
  browse = TRUE
)
```

Arguments

`year`	The year in yyyy format where 1999 <= yyyy.
`data_group`	The type of survey (DEMOGRAPHICS, DIETARY, EXAMINATION, LABORATORY, QUESTIONNAIRE). Abbreviated terms may also be used: (DEMO, DIET, EXAM, LAB, Q).
`nh_table`	The name of an NHANES table.
`local`	logical flag. If `TRUE`, and a local or alternative source was specificed using the environment variable `NHANES_TABLE_BASE`, this will be used in preference to the CDC website at `https://wwwn.cdc.gov` for named tables.
`browse`	logical flag, indicating whether the specific NHANES site should be opened using a browser (which is the default behaviour).

Details

By default, browseNHANES will open a web browser to the specified NHANES site.

Value

A character string giving the URL, invisibly if the URL is also opened using `browseURL`.

Examples

```
browseNHANES(browse = FALSE)       # Defaults to the main data sets page
browseNHANES(2005)                 # The main page for the specified survey year
browseNHANES(2009, 'EXAM')         # Page for the specified year and survey group
browseNHANES(nh_table = 'VIX_D')   # Page for a specific table
browseNHANES(nh_table = 'DXA')     # DXA main page
```

---

nhanes *Download an NHANES table and return as a data frame.*

Description

Use to download NHANES data tables that are in SAS format.

Usage

```
nhanes(
  nh_table,
  includelabels = FALSE,
  translated = TRUE,
  cleanse_numeric = FALSE,
  nchar = 128,
  adjust_timeout = TRUE
)
```

Arguments

`nh_table`	The name of the specific table to retrieve.
`includelabels`	If TRUE, then include SAS labels as variable attribute (default = FALSE).
`translated`	translated whether the variables are translated.
`cleanse_numeric`	Logical flag. If `TRUE`, some special codes in numeric variables, such as 'Refused' and 'Don't know' will be converted to `NA`.
`nchar`	Maximum length of translated string (default = 128). Ignored if translated=FALSE.
`adjust_timeout`	Typically a logical flag indicating whether the default `download.file` timeout option should be adjusted by taking into account the size of the file to be downloaded, as reported by the server. The value can also be a positive numeric value, in which case it is used as a further multiplicative factor for the default calculation.

Details

Downloads a table from the NHANES website as is, i.e. in its entirety with no modification or cleansing. If the environment variable NHANES_TABLE_BASE was set during startup, the value of this variable is used as the base URL instead of https://wwwn.cdc.gov (this allows the use of a local or alternative mirror of the CDC data). NHANES tables are stored in SAS '.XPT' format but are imported as a data frame. The nhanes function cannot be used to import limited access data.

Value

The table is returned as a data frame.

Examples

```
bpx_e = nhanes('BPX_E')
dim(bpx_e)
folate_f = nhanes('FOLATE_F', includelabels = TRUE)
dim(folate_f)
```

---

nhanesAttr *Returns the attributes of an NHANES data table.*

Description

Returns attributes such as number of rows, columns, and memory size, but does not return the table itself. This function is deprecated. Use nhanesTableSummary instead.

Usage

```
nhanesAttr(nh_table)
```

Arguments

nh_table	The name of the specific table to retrieve

Details

nhanesAttr allows one to check the size and other charactersistics of a data table before importing into R. To retrieve these characteristics, the specified table is downloaded, characteristics are determined, then the table is deleted. Downloads a table from the NHANES website as is, i.e. in its entirety with no modification or cleansing.

If the environment variable NHANES_TABLE_BASE was set during startup, the value of this variable is used as the base URL instead of https://wwwn.cdc.gov (this allows the use of a local or alternative mirror of the CDC data).

Value

The following attributes are returned as a list
nrow = number of rows
ncol = number of columns
names = name of each column
unique = true if all SEQN values are unique
na = number of 'NA' cells in the table
size = total size of table in bytes
types = data types of each column

Examples

```
## Not run: bpx_e = nhanesAttr('BPX_E')
## Not run: length(bpx_e)
## Not run: folate_f = nhanesAttr('FOLATE_F')
## Not run: length(folate_f)
```

---

nhanesCodebook *Display codebook for selected variable.*

Description

Returns full NHANES codebook including Variable Name, SAS Label, English Text, Target, and Value distribution.

Usage

```
nhanesCodebook(nh_table, colname = NULL, dxa = FALSE)
```

Arguments

nh_table	The name of the NHANES table that contains the desired variable.
colname	The name of the table column (variable).
dxa	If TRUE then the 2005-2006 DXA codebook will be used (default=FALSE).

Details

Each NHANES variable has a codebook that provides a basic description as well as the distribution or range of values. This function returns the full codebook information for the selected variable. If the environment variable NHANES_TABLE_BASE was set during startup, the value of this variable is used as the base URL instead of https://wwwn.cdc.gov (this allows the use of a local or alternative mirror of the CDC documentation).

Value

The codebook is returned as a list object. Returns NULL upon error.

Examples

```
nhanesCodebook('AUX_D', 'AUQ020D')
nhanesCodebook('BPX_J', 'BPACSZ')
bpx_code = nhanesCodebook('BPX_J')
length(bpx_code)
```

---

nhanesCodebookFromURL *Parse NHANES doc URL*

Description

Download and parse an NHANES doc file from a URL

Usage

```
nhanesCodebookFromURL(url)
```

Arguments

url	URL to be downloaded

Details

Downloads and parses an NHANES doc file from a URL and returns it as a list

Value

list with one element for each variable

---

nhanesDXA *Import Dual Energy X-ray Absorptiometry (DXA) data.*

Description

DXA data were acquired from 1999-2006.

Usage

```
nhanesDXA(year, suppl = FALSE, destfile = NULL, adjust_timeout = TRUE)
```

Arguments

year	The year of the data to import, where 1999<=year<=2006.
suppl	If TRUE then retrieve the supplemental data (default=FALSE).
destfile	The name of a destination file. If NULL then the data are imported into the R environment but no file is created.
adjust_timeout	Typically a logical flag indicating whether the default download.file timeout option should be adjusted by taking into account the size of the file to be downloaded, as reported by the server. The value can also be a positive numeric value, in which case it is used as a further multiplicative factor for the default calculation.

Details

Provide destfile in order to write the data to file. If destfile is not provided then the data will be imported into the R environment.

Value

By default the table is returned as a data frame. When downloading to file, the return argument is the integer code from download.file where 0 means success and non-zero indicates failure to download.

Examples

```
dxa_b <- nhanesDXA(2001)
dxa_c_s <- nhanesDXA(2003, suppl=TRUE)
## Not run: dxa = nhanesDXA(1999, destfile="dxx.xpt")
```

---

nhanesFromURL *Download NHANES table from URL*

Description

Downloads an NHANES table from a URL and returns it as a data frame.

Usage

```
nhanesFromURL(
  url,
  translated = TRUE,
  cleanse_numeric = TRUE,
  nchar = 128,
  adjust_timeout = TRUE
)
```

Arguments

url	URL of XPT file to be downloaded.
translated	logical, whether variable codes should be translated
cleanse_numeric	Logical flag. If TRUE, some special codes in numeric variables, such as 'Refused' and 'Don't know' will be converted to NA.
nchar	integer, labels are truncated after this
adjust_timeout	Typically a logical flag indicating whether the default download.file timeout option should be adjusted by taking into account the size of the file to be downloaded, as reported by the server. The value can also be a positive numeric value, in which case it is used as a further multiplicative factor for the default calculation.

Details

This function downloads a table from the NHANES website using its URL. It is similar to nhanes, except that it requires the URL to be explicitly specified, and does not try to infer it from the table name. It also performs some limited cleansing of the data by default.

The URL may be completely specified, but the initial standard prefix "https://wwwn.cdc.gov" may be optionally dropped. More precisely, if the URL starts with "/nchs/", then the prefix is automatically added. It is possible to override the default prefix by setting the environment variable NHANES_TABLE_BASE (this allows the use of a local or alternative mirror of the CDC data).

Value

A data frame containing the data in the XPT file avalable at the URL.

Examples

```
vix_e = nhanesFromURL("https://wwwn.cdc.gov/Nchs/Data/Nhanes/Public/2007/DataFiles/VIX_E.xpt")
bpx_e = nhanesFromURL("/Nchs/Data/Nhanes/Public/2007/DataFiles/BPX_E.xpt", translated = FALSE)
dim(bpx_e)
```

---

nhanesManifest *Download and parse NHANES manifests*

Description

Downloads and parses NHANES manifests for data tables and variables, and returns them as data frames.

Usage

```
nhanesManifest(
  which = c("public", "limitedaccess", "variables"),
  sizes = FALSE,
  dxa = FALSE,
  component = NULL,
  verbose = getOption("verbose"),
  use_cache = TRUE,
  max_age = 24 * 60 * 60
)
```

Arguments

which	Either "public" or "limitedaccess" to get a manifest of available tables, or "variables" to get a manifest of available variables.
sizes	Logical, whether to compute data file sizes (as reported by the server) and include them in the result.
dxa	Logical, whether to include information on DXA tables. These tables contain imputed imputed Dual Energy X-ray Absorptiometry measurements, and are listed separately, not in the main listing.
component	An optional character string specifying the component for which the public data manifest is to be downloaded. Valid values are "demographics", "dietary", "examination", "laboratory", and "questionnaire". Partial matching is allowed, and case is ignored. Specifying a component for the public manifest will return a subset of the tables, but has the advantage that the result will include a description of each table.
verbose	Logical flag indicating whether information on progress should be reported.
use_cache	Logical flag indicating whether a cached version (from a previous download in the same session) should be used.
max_age	Maximum allowed age of the cache in seconds (defaults to 24 hours). Cached versions that are older are ignored, even if available.

Details

The NHANES website maintains several listings (manifests) of tables and associated variables, which can be downloaded using these functions.

The list of tables for which data is available publicly can be found at https://wwwn.cdc.gov/Nchs/Nhanes/search/DataPage.aspx, with further restriction to specific components possible by specifying an additional query parameter as below. This is the *public* manifest.

Limited access tables (also referred to as RDC only tables) are listed at https://wwwn.cdc.gov/Nchs/Nhanes/search/DataPage.aspx?Component=LimitedAccess. This is the *limited access* manifest.

Available variables are listed according to component at https://wwwn.cdc.gov/nchs/nhanes/search/variablelist.aspx?Component=Demographics, etc. These are the *variable* manifests.

Value

A data frame, with columns that depend on `which`.

For a manifest of tables, columns are "Table", "DocURL", "DataURL", "Years", "Date.Published". If `component` is specified, an additional column "Description" giving a description of the table will be included. If `sizes = TRUE`, an additional column "DataSize" giving the data file sizes in bytes (as reported by the server) is included.

For limited access tables, the "DataURL" and "DataSize" columns are omitted.

For a manifest of variables, columns are "VarName", "VarDesc", "Table", "TableDesc", "BeginYear", "EndYear", "Component", and "UseConstraints".

Note

Duplicate rows are removed from the result. Most of these duplicates arise from duplications in the source tables for multi-cycle tables (which are repeated once for each cycle). One special case is the WHQ table which has two variables, WHD120 and WHQ030, duplicated with differing variable descriptions. These are removed explicitly, keeping only the first occurrence.

Examples

```
manifest <- nhanesManifest(sizes = FALSE)
dim(manifest)
varmf <- nhanesManifest("variables", component = "lab")
head(varmf)
```

`nhanesOptions` *Options for the nhanesA package*

Description

Set and retrieve global options controlling the behaviour of certain functions in the package.

Usage

```
nhanesOptions(...)
```

Arguments

`...`	either one or more named arguments giving options to be set (in the form `key = value`), or a single unnamed character string to retrieve a setting.

Details

The 'nhanesOptions()' function can be used in two forms, to set or get options. Options can be set using 'nhanesOptions(key1 = value1, key2 = value2)'. Options can be retrieved (one at a time) using 'nhanesOptions("key")'. When called with no arguments, all currently set options are returned as a list.

Options currently used in the package are 'use.db' (logical flag controlling whether a database should be used if available), and 'log.access', a logical flag that logs any attempted URL access by printing the URL).

Value

When retrieving an option, the value of the option, or `NULL` if the option has not been set. When setting one or more options, a list (invisibly) containing the previous values (possibly `NULL`) of the options being set.

Author(s)

Deepayan Sarkar <deepayan.sarkar@gmail.com>

Examples

```
nhanesOptions(foo = "bar")
nhanesOptions()
print(nhanesOptions(foo = NULL))
```

---

nhanesSearch *Perform a search over the comprehensive NHANES variable list.*

Description

The descriptions in the master variable list will be filtered by the provided search terms to retrieve a list of relevant variables. The search can be restricted to specific survey years by specifying ystart and/or ystop.

Usage

```
nhanesSearch(
  search_terms = NULL,
  exclude_terms = NULL,
  data_group = NULL,
  ignore.case = FALSE,
  ystart = NULL,
  ystop = NULL,
  includerdc = FALSE,
  nchar = 128,
  namesonly = FALSE
)
```

Arguments

search_terms	List of terms or keywords.
exclude_terms	List of exclusive terms or keywords.
data_group	Which data groups (e.g. DIET, EXAM, LAB) to search. Default is to search all groups.
ignore.case	Ignore case if TRUE. (Default=FALSE).
ystart	Four digit year of first survey included in search, where ystart >= 1999.
ystop	Four digit year of final survey included in search, where ystop >= ystart.
includerdc	If TRUE then RDC only tables are included in list (default=FALSE).
nchar	Truncates the variable description to a max length of nchar.
namesonly	If TRUE then only the table names are returned (default=FALSE).

Details

nhanesSearch is useful to obtain a comprehensive list of relevant tables. Search terms will be matched against the variable descriptions in the NHANES Comprehensive Variable Lists. Matching variables must have at least one of the search_terms and not have any exclude_terms. The search may be restricted to specific surveys using ystart and ystop. If no arguments are given, then nhanesSearch returns the complete variable list.

Value

Returns a data frame that describes variables that matched the search terms. If namesonly=TRUE, then a character vector of table names that contain matched variables is returned.

Examples

```
bladder = nhanesSearch("bladder", ystart=2001, ystop=2008, nchar=50)
dim(bladder)
urin = nhanesSearch("urin", exclude_terms="During", ystart=2009)
dim(urin)
urine = nhanesSearch(c("urine", "urinary"), ignore.case=TRUE, ystop=2006, namesonly=TRUE)
length(urine)
```

---

nhanesSearchTableNames

Search for matching table names

Description

Returns a list of table names that match a specified pattern.

Usage

```
nhanesSearchTableNames(
  pattern = NULL,
  ystart = NULL,
  ystop = NULL,
  includerdc = FALSE,
  includewithdrawn = FALSE,
  nchar = 128,
  details = FALSE
)
```

Arguments

pattern	Pattern of table names to match
ystart	Four digit year of first survey included in search, where ystart >= 1999.
ystop	Four digit year of final survey included in search, where ystop >= ystart.
includerdc	If TRUE then RDC only tables are included (default=FALSE).
includewithdrawn	IF TRUE then withdrawn tables are included (default=FALSE).
nchar	Truncates the variable description to a max length of nchar.
details	If TRUE then complete table information from the comprehensive data list is returned (default=FALSE).

Details

Searches the Doc File field in the NHANES Comprehensive Data List (see https://wwwn.cdc.gov/nchs/nhanes/search/DataPag for tables that match a given name pattern. Only a single pattern may be entered.

Value

Returns a character vector of table names that match the given pattern. If details=TRUE, then a data frame of table attributes is returned. NULL is returned when an HTML read error is encountered.

Examples

```
bmx = nhanesSearchTableNames('BMX')
length(bmx)
hepbd = nhanesSearchTableNames('HEPBD')
length(hepbd)
hpvs = nhanesSearchTableNames('HPVS', includerdc=TRUE, details=TRUE)
dim(hpvs)
```

nhanesSearchVarName *Search for tables that contain a specified variable.*

Description

Returns a list of table names that contain the variable

Usage

```
nhanesSearchVarName(
  varname = NULL,
  ystart = NULL,
  ystop = NULL,
  includerdc = FALSE,
  nchar = 128,
  namesonly = TRUE
)
```

Arguments

varname	Name of variable to match.
ystart	Four digit year of first survey included in search, where ystart >= 1999.
ystop	Four digit year of final survey included in search, where ystop >= ystart.
includerdc	If TRUE then RDC only tables are included in list (default=FALSE).
nchar	Truncates the variable description to a max length of nchar.
namesonly	If TRUE then only the table names are returned (default=TRUE).

Details

The NHANES Comprehensive Variable List is scanned to find all data tables that contain the given variable name. Only a single variable name may be entered, and only exact matches will be found.

Value

By default, a character vector of table names that include the specified variable is returned. If namesonly=FALSE, then a data frame of table attributes is returned.

Examples

```
bmxleg = nhanesSearchVarName('BMXLEG')
length(bmxleg)
bmxhead = nhanesSearchVarName('BMXHEAD', ystart=2003)
length(bmxhead)
```

nhanesTables *Returns a list of table names for the specified survey group.*

Description

Enables quick display of all available tables in the survey group.

Usage

```
nhanesTables(
  data_group,
  year,
  nchar = 128,
  details = FALSE,
  namesonly = FALSE,
  includerdc = FALSE
)
```

Arguments

data_group	The type of survey (DEMOGRAPHICS, DIETARY, EXAMINATION, LABORATORY, QUESTIONNAIRE). Abbreviated terms may also be used: (DEMO, DIET, EXAM, LAB, Q).
year	The year in yyyy format where 1999 <= yyyy.
nchar	Truncates the table description to a max length of nchar.
details	If TRUE then a more detailed description of the tables is returned (default=FALSE).
namesonly	If TRUE then only the table names are returned (default=FALSE).
includerdc	If TRUE then RDC only tables are included in list (default=FALSE).

Details

Function nhanesTables retrieves a list of tables and a description of their contents from the NHANES website. This provides a convenient way to browse the available tables. NULL is returned when an HTML read error is encountered.

Value

Returns a data frame that contains table attributes. If namesonly=TRUE, then a character vector of table names is returned.

Examples

```
exam = nhanesTables('EXAM', 2007)
dim(exam)
lab = nhanesTables('LAB', 2009, details=TRUE, includerdc=TRUE)
dim(lab)
```

```
q = nhanesTables('Q', 2005, namesonly=TRUE)
length(q)
diet = nhanesTables('DIET', 'P')
dim(diet)
exam = nhanesTables('EXAM', 'Y')
dim(exam)
```

---

nhanesTableSummary *Summarize an NHANES table*

Description

Computes a per-variable summary of a NHANES table either using the actual data or its corresponding codebook

Usage

```
nhanesTableSummary(nh_table, use = c("data", "codebook", "both"), ...)
```

Arguments

nh_table	the name of a valid NHANES table
use	character string, whether to create a summary from the data itself or the codebook, which respectively use either the NHANES SAS data files or the HTML documentation files. If use = "both" then both are computed and merged; additional arguments (...) are ignored in this case.
...	additional arguments, usually passed on to either nhanes or nhanesCodebook as appropriate. Alternatively, the src argument can be used to pass on an already available data frame or codebook, but this must be consistent with the use argument.

Details

This function computes useful summaries of each variable included in the specified NHANES table. The resulting data frame contains one row for each variable in the table, and includes summary measures that indicate the total number of observations, the number of missing observations, whether the variable is most likely numeric or categorical, whether the variable is related to skipping other variables, etc. The precise details depend on the use argument and are subject to change.

Value

A data frame with one row per variable, with columns depending on the value of the use argument.

Examples

```
nhanesTableSummary('DEMO_D', use = "data")
nhanesTableSummary('DEMO_D', use = "codebook")
```

nhanesTableVars *Displays a list of variables in the specified NHANES table.*

Description

Enables quick display of table variables and their definitions.

Usage

```
nhanesTableVars(
  data_group,
  nh_table,
  details = FALSE,
  nchar = 128,
  namesonly = FALSE
)
```

Arguments

data_group	The type of survey (DEMOGRAPHICS, DIETARY, EXAMINATION, LABORATORY, QUESTIONNAIRE). Abbreviated terms may also be used: (DEMO, DIET, EXAM, LAB, Q).
nh_table	The name of the specific table to retrieve.
details	If TRUE then all columns in the variable description are returned (default=FALSE).
nchar	The number of characters in the Variable Description to print. Default length is 128, which is set to enhance readability cause variable descriptions can be very long.
namesonly	If TRUE then only the variable names are returned (default=FALSE).

Details

NHANES tables may contain more than 100 variables. Function nhanesTableVars provides a concise display of variables for a specified table, which helps to ascertain quickly if the table is of interest. NULL is returned when an HTML read error is encountered.

Value

Returns a data frame that describes variable attributes for the specified table. If namesonly=TRUE, then a character vector of the variable names is returned.

Examples

```
lab_cbc = nhanesTableVars('LAB', 'CBC_E')
dim(lab_cbc)
exam_ohx = nhanesTableVars('EXAM', 'OHX_E', details=TRUE, nchar=50)
dim(exam_ohx)
demo = nhanesTableVars('DEMO', 'DEMO_F', namesonly = TRUE)
length(demo)
```

`nhanesTranslate` *Display code translation information.*

Description

Returns code translations for categorical variables, which appear in most NHANES tables.

Usage

```
nhanesTranslate(
  nh_table,
  colnames = NULL,
  data = NULL,
  nchar = 128,
  mincategories = 1,
  details = FALSE,
  dxa = FALSE,
  cleanse_numeric = FALSE
)
```

Arguments

`nh_table`	The name of the NHANES table to retrieve.
`colnames`	The names of the columns to translate. It will translate all the columns by default.
`data`	If a data frame is passed, then code translation will be applied directly to the data frame. In that case the return argument is the code-translated data frame.
`nchar`	Applies only when data is defined. Code translations can be very long. Truncate the length by setting nchar (default = 128).
`mincategories`	The minimum number of categories needed for code translations to be applied to the data (default=1).
`details`	If TRUE then all available table translation information is displayed (default=FALSE).
`dxa`	If TRUE then the 2005-2006 DXA translation table will be used (default=FALSE).
`cleanse_numeric`	Logical flag. If `TRUE`, some special codes in numeric variables, such as 'Refused' and 'Don't know' will be converted to `NA`.

Details

Most NHANES data tables have encoded values. E.g. 1 = 'Male', 2 = 'Female'. Thus it is often helpful to view the code translations and perhaps insert the translated values in a data frame. Only a single table may be specified, but multiple variables within that table can be selected. Code translations are retrieved for each variable. If the environment variable `NHANES_TABLE_BASE` was set during startup, the value of this variable is used as the base URL instead of `https://wwwn.cdc.gov` (this allows the use of a local or alternative mirror of the CDC documentation).

Value

The code translation table (or translated data frame when data is defined). Returns NULL upon error.

Examples

```
nhanesTranslate('DEMO_B', c('DMDBORN','DMDCITZN'))
nhanesTranslate('BPX_F', 'BPACSZ', details=TRUE)
nhanesTranslate('BPX_F', 'BPACSZ', data=nhanes('BPX_F'))
trans_demo = nhanesTranslate('DEMO_B')
length(trans_demo)
```

Index