

Package: ipumseasyr (via r-universe)

November 21, 2024

Type Package

Title Easy Access to IPUMS Data

Version 0.1.0

Maintainer Eli Pousson <eli.pousson@gmail.com>

Description A package with helper functions extending the ipumsr package for accessing NHGIS and other IPUMS data sources.

License MIT + file LICENSE

URL <https://github.com/elipousson/ipumseasyr>,
<https://elipousson.github.io/ipumseasyr/>

BugReports <https://github.com/elipousson/ipumseasyr/issues>

Depends R (>= 2.10)

Imports cli, dplyr, ipumsr, rappdirs, rlang, stringr, vctrs

Suggests ggplot2, labelled, readr, sf, tibble, tidyr, tigris (>= 2.1.1), withr

Remotes elipousson/tigris

Encoding UTF-8

Language en-US

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Config/pak/sysreqs make libicu-dev libxml2-dev libssl-dev libx11-dev
zlib1g-dev

Repository <https://elipousson.r-universe.dev>

RemoteUrl <https://github.com/elipousson/ipumseasyr>

RemoteRef HEAD

RemoteSha eaceb0ac650e425c5d671184dcfef6d29cb51098

Contents

define_nhgis_ts_extract	2
download_ipumsr_extract	4
get_ipumsr_extract_paths	5
get_nhgis_ts_data	6
join_nhgis_percent_change	8
labs_nhgis	8
list_nhgis_ts_tables	9
nhgis_ts_tables	10
read_ipums_geometry	10
read_nhgis_files	11
usa_states	12

Index	13
--------------	-----------

```
define_nhgis_ts_extract
      Define a NHGIS time series extract using
      ipumsr::define_extract_nhgis
```

Description

`define_nhgis_ts_extract()` is a wrapper for `ipumsr::define_extract_nhgis()` with defaults that support the creation of tidy data using `read_nhgis_data()` or `pivot_nhgis_data()`.

Usage

```
define_nhgis_ts_extract(
  year = NULL,
  tables = NULL,
  geography = c("county", "state"),
  extent = "us",
  output = c("tidy", "wide", "file"),
  shape_year = NULL,
  basis = 2008,
  geometry = FALSE,
  ...,
  time_series_tables = NULL,
  description = NULL,
  shapefiles = NULL,
  data_format = "csv_no_header",
  validate = TRUE,
  api_key = Sys.getenv("IPUMS_API_KEY")
)
```

Arguments

output	Used to set <code>tst_layout</code> value. <code>c("tidy", "wide", "file")</code> corresponding to "time_by_row_layout", "time_by_column_layout", or "time_by_file_layout".
geometry	If TRUE, include shapefiles in the defined extract. If shapefiles is NULL, the function uses <code>list_nhgis_shapefiles()</code> with <code>shape_year</code> as the year parameter.
...	Arguments passed on to <code>ipumsr::define_extract_nhgis</code>
datasets	List of dataset specifications for any datasets to include in the extract request. Use <code>ds_spec()</code> to create a <code>ds_spec</code> object containing a dataset specification. See examples.
geographic_extents	Vector of geographic extents to use for all of the datasets in the extract definition (for instance, to obtain data within a particular state). Use "*" to select all available extents. Required when any of the datasets included in the extract definition include <code>geog_levels</code> that require extent selection. See <code>get_metadata_nhgis()</code> to determine if a geographic level requires extent selection. At the time of writing, NHGIS supports extent selection only for blocks and block groups.
breakdown_and_data_type_layout	The desired layout of any datasets that have multiple data types or breakdown values. <ul style="list-style-type: none"> "single_file" (default) keeps all data types and breakdown values in one file "separate_files" splits each data type or breakdown value into its own file Required if any datasets included in the extract definition consist of multiple data types (for instance, estimates and margins of error) or have multiple breakdown values specified. See <code>get_metadata_nhgis()</code> to determine whether a requested dataset has multiple data types.
time_series_tables	List of time series table specifications for any time series tables to include in the extract request. Use <code>tst_spec()</code> to create a <code>tst_spec</code> object containing a time series table specification. See examples.
description	Description of the extract.
shapefiles	Names of any shapefiles to include in the extract request.
data_format	The desired format of the extract data file. <ul style="list-style-type: none"> "csv_no_header" (default) includes only a minimal header in the first row "csv_header" includes a second, more descriptive header row. "fixed_width" provides data in a fixed width format Note that by default, <code>read_nhgis()</code> removes the additional header row in "csv_header" files. Required when an extract definition includes any <code>datasets</code> or <code>time_series_tables</code> .
api_key	API key associated with your user account. Defaults to the value of the <code>IPUMS_API_KEY</code> environment variable. See <code>set_ipums_api_key()</code> .

```
download_ipumsr_extract
```

Download IPUMS extract using `ipumsr::wait_for_extract` and `ipumsr::download_extract`

Description

`download_ipumsr_extract()` is a wrapper for `ipumsr::wait_for_extract()` and `ipumsr::download_extract()` to wait until an extract is ready for download before attempting to download it.

Usage

```
download_ipumsr_extract(
  extract = NULL,
  download_dir = getwd(),
  overwrite = FALSE,
  progress = TRUE,
  ...,
  api_key = Sys.getenv("IPUMS_API_KEY")
)
```

Arguments

<code>extract</code>	<p>One of:</p> <ul style="list-style-type: none"> An <code>ipums_extract</code> object The data collection and extract number formatted as a string of the form "collection:number" or as a vector of the form <code>c("collection", number)</code> An extract number to be associated with your default IPUMS collection. See <code>set_ipums_default_collection()</code>. <p>For a list of codes used to refer to each collection, see <code>ipums_data_collections()</code>.</p>
<code>download_dir</code>	Path to the directory where the files should be written. Defaults to current working directory.
<code>overwrite</code>	If TRUE, overwrite any conflicting files that already exist in <code>download_dir</code> . Defaults to FALSE.
<code>progress</code>	If TRUE, output progress bar showing the status of the download request. Defaults to TRUE.
<code>...</code>	<p>Arguments passed on to <code>ipumsr::wait_for_extract</code></p> <p><code>initial_delay_seconds</code> Seconds to wait before first status check. The wait time will automatically increase by 10 seconds between each successive check.</p> <p><code>max_delay_seconds</code> Maximum interval to wait between status checks. When the wait interval reaches this value, checks will continue to occur at <code>max_delay_seconds</code> intervals until the extract is complete or <code>timeout_seconds</code> is reached. Defaults to 300 seconds (5 minutes).</p>

timeout_seconds	Maximum total number of seconds to continue waiting for the extract before throwing an error. Defaults to 10,800 seconds (3 hours).
verbose	If TRUE, print status updates to the R console at the beginning of each wait interval and upon extract completion. Defaults to TRUE.
api_key	API key associated with your user account. Defaults to the value of the IPUMS_API_KEY environment variable. See set_ipums_api_key() .

get_ipumsr_extract_paths

Get extract paths for extract with optional support for cached extract files

Description

Download extract with [download_ipumsr_extract\(\)](#) and return a list of file paths for the data and shape files.

Usage

```
get_ipumsr_extract_paths(
  extract = NULL,
  data_file = NULL,
  shape_file = NULL,
  submit_extract = TRUE,
  download_extract = TRUE,
  download_dir = getwd(),
  overwrite = FALSE,
  progress = TRUE,
  refresh = FALSE,
  api_key = Sys.getenv("IPUMS_API_KEY")
)
```

Arguments

extract	An ipums_extract object.
submit_extract	If extract is not NULL and submit_extract = TRUE, use ipumsr::submit_extract to submit the extract.
download_dir	Path to the directory where the files should be written. Defaults to current working directory.
overwrite	If TRUE, overwrite any conflicting files that already exist in download_dir. Defaults to FALSE.
progress	If TRUE, output progress bar showing the status of the download request. Defaults to TRUE.
api_key	API key associated with your user account. Defaults to the value of the IPUMS_API_KEY environment variable. See set_ipums_api_key() .

Value

A named list with "data" and "shape" elements containing extract file paths.

get_nhgis_ts_data *Get NHGIS time series data*

Description

Use `define_nhgis_ts_extract()`, `ipumsr::submit_extract()`, `ipumsr::download_extract()`, and `read_nhgis_files()` to define, submit, download, and read a NHGIS time series extract. This function is *only* recommended for interactive use and is *not* recommended if you are requesting a large number of tables or geographies.

Usage

```
get_nhgis_ts_data(  
  year = NULL,  
  tables = NULL,  
  geography = c("county", "state"),  
  extent = "us",  
  output = c("tidy", "wide", "file"),  
  basis = 2008,  
  shape_year = NULL,  
  geometry = FALSE,  
  extract = NULL,  
  data_file = NULL,  
  shape_file = NULL,  
  state = NULL,  
  ...,  
  time_series_tables = NULL,  
  description = NULL,  
  shapefiles = NULL,  
  data_format = "csv_no_header",  
  validate = TRUE,  
  submit_extract = TRUE,  
  download_extract = TRUE,  
  read_files = TRUE,  
  download_dir = getwd(),  
  overwrite = FALSE,  
  progress = TRUE,  
  verbose = progress,  
  api_key = Sys.getenv("IPUMS_API_KEY")  
)
```

Arguments

output	Used to set <code>tst_layout</code> value. <code>c("tidy", "wide", "file")</code> corresponding to "time_by_row_layout", "time_by_column_layout", or "time_by_file_layout".
geometry	If TRUE, include shapefiles in the defined extract. If shapefiles is NULL, the function uses <code>list_nhgis_shapefiles()</code> with <code>shape_year</code> as the year parameter.
extract	An <code>ipums_extract</code> object.
data_file	Path to a .zip archive containing an NHGIS extract or a single file from an NHGIS extract.
shape_file	Path to a single .shp file or a .zip archive containing at least one .shp file. See Details section.
time_series_tables	List of time series table specifications for any time series tables to include in the extract request. Use <code>tst_spec()</code> to create a <code>tst_spec</code> object containing a time series table specification. See examples.
description	Description of the extract.
shapefiles	Names of any shapefiles to include in the extract request.
data_format	The desired format of the extract data file. <ul style="list-style-type: none"> • "csv_no_header" (default) includes only a minimal header in the first row • "csv_header" includes a second, more descriptive header row. • "fixed_width" provides data in a fixed width format <p>Note that by default, <code>read_nhgis()</code> removes the additional header row in "csv_header" files.</p> <p>Required when an extract definition includes any datasets or <code>time_series_tables</code>.</p>
download_dir	Path to the directory where the files should be written. Defaults to current working directory.
overwrite	If TRUE, overwrite any conflicting files that already exist in <code>download_dir</code> . Defaults to FALSE.
progress	If TRUE, output progress bar showing the status of the download request. Defaults to TRUE.
verbose	Logical controlling whether to display output when loading data. If TRUE, displays IPUMS conditions, a progress bar, and column types. Otherwise, all are suppressed.
	Will be overridden by <code>readr.show_progress</code> and <code>readr.show_col_types</code> options, if they are set.
api_key	API key associated with your user account. Defaults to the value of the <code>IPUMS_API_KEY</code> environment variable. See <code>set_ipums_api_key()</code> .

`join_nhgis_percent_change`*Join a percent change in variable relative to a reference year*

Description

`join_nhgis_percent_change()` joins a percent change column relative to a reference year. Optionally join a rank from the reference year using `dplyr::ntile()`.

Usage

```
join_nhgis_percent_change(  
  data,  
  reference_year = NULL,  
  value_col = "value",  
  reference_prefix = "reference_",  
  variable_col = "variable",  
  year_col = "YEAR",  
  rank_col = "rank",  
  rank = NULL,  
  rank_n = NULL,  
  rank_by = NULL,  
  ...,  
  perc_prefix = "perc_change_",  
  digits = 2  
)
```

Arguments

`reference_year` Reference year to use when calculating a percent change column.

`rank`, `rank_n` Passed to `x` and `n` arguments of `dplyr::ntile()` to join a reference rank value.

`rank_by` Used as `.by` argument of `dplyr::mutate()` if `rank_n` is not `NULL`.

`labs_nhgis`*Label ggplot2 plots with the appropriate credit caption for NHGIS data*

Description

`labs_nhgis()` adds a standard credit caption for NHGIS data to make consistent attribution easier.

Usage

```
labs_nhgis(
  ...,
  caption = NULL,
  credit = "IPUMS NHGIS, University of Minnesota, www.nhgis.org.",
  prefix = "Source: ",
  collapse = " ",
  width = 80
)
```

Arguments

...	Arguments passed on to <code>ggplot2::labs</code>
title	The text for the title.
subtitle	The text for the subtitle for the plot which will be displayed below the title.
caption	The text for the caption which will be displayed in the bottom-right of the plot by default.
tag	The text for the tag label which will be displayed at the top-left of the plot by default.
alt, alt_insight	Text used for the generation of alt-text for the plot. See get_alt_text for examples.
credit	Credit line for IPUMS.
collapse	String to collapse caption and credit. Defaults to " ". Set to "\n" to place the credit line on a separate line following the caption. Ignored if caption is NULL.
width	Maximum width of caption line passed to <code>stringr::str_wrap()</code> .

list_nhgis_ts_tables *List NHGIS time series tables using ipumsr::get_metadata_nhgis*

Description

Use `ipumsr::get_metadata_nhgis()` with `type = "time_series_tables"` to return a data frame of time series tables. Optionally filter by geographical integration type "nominal" or "standardized" ("2010" or "standardized to 2010" also work).

Usage

```
list_nhgis_ts_tables(
  ...,
  cache = TRUE,
  cache_file = "nhgis_time_series_tables.rds",
  refresh = FALSE,
  integration = NULL
)
```

Arguments

...	Additional parameters passed to <code>ipumsr::get_metadata_nhgis()</code>
refresh	If FALSE, do not read a file from cache. If TRUE, read a file from cache if it exists at the supplied path.
integration	Optional filter for geographical integration.

nhgis_ts_tables	<i>NHGIS Time Series Table names</i>
-----------------	--------------------------------------

Description

A vector of NHGIS time series table names named with table descriptions.

Usage

```
nhgis_ts_tables
```

Format

A character vector with 389 time series table names.

read_ipums_geometry	<i>Read IPUMS geometry using ipumsr::read_ipums_sf</i>
---------------------	--

Description

Read IPUMS geometry using `ipumsr::read_ipums_sf`

Usage

```
read_ipums_geometry(
  shape_file = NULL,
  path = NULL,
  file_select = NULL,
  vars = "GISJOIN",
  encoding = NULL,
  bind_multiple = TRUE,
  add_layer_var = NULL,
  verbose = FALSE
)
```

Arguments

shape_file	Path to a single .shp file or a .zip archive containing at least one .shp file. See Details section.
file_select	If shape_file is a .zip archive that contains multiple files, an expression identifying the files to load. Accepts a character string specifying the file name, a tidyselect selection , or index position. If multiple files are selected, bind_multiple must be equal to TRUE.
vars	Names of variables to include in the output. Accepts a character vector of names or a tidyselect selection . If NULL, includes all variables in the file.
encoding	Encoding to use when reading the shape file. If NULL, defaults to "latin1" unless the file includes a .cpg metadata file with encoding information. The default value should generally be appropriate.
bind_multiple	If TRUE and shape_file contains multiple .shp files, row-bind the files into a single sf object. Useful when shape_file contains multiple files that represent the same geographic units for different extents (e.g. block-level data for multiple states).
add_layer_var	If TRUE, add a variable to the output data indicating the file that each row originates from. Defaults to FALSE unless bind_multiple = TRUE and multiple files exist in shape_file. The column name will always be prefixed with "layer", but will be adjusted to avoid name conflicts if another column named "layer" already exists in the data.
verbose	If TRUE report additional progress information on load.

read_nhgis_files *Read NHGIS data and geometry*

Description

Read NHGIS data and geometry to return a named list or a combined sf object.

Usage

```
read_nhgis_files(
  path = NULL,
  data_file = NULL,
  data_file_select = NULL,
  shape_file = NULL,
  shape_file_select = NULL,
  verbose = FALSE,
  geometry = FALSE,
  ...
)
```

Arguments

path	Optional if data_file is supplied. A named list with a "data" and "shape" element containing the paths to the data_file and shape_file arguments of used by <code>ipumsr::read_nhgis()</code> and <code>ipumsr::read_ipums_sf()</code> .
data_file	Path to a .zip archive containing an NHGIS extract or a single file from an NHGIS extract.
data_file_select, shape_file_select	Passed to file_select parameter of <code>read_nhgis_data()</code> or <code>read_ipums_geometry()</code> .
shape_file	Path to a single .shp file or a .zip archive containing at least one .shp file. See Details section.
verbose	Logical controlling whether to display output when loading data. If TRUE, displays IPUMS conditions, a progress bar, and column types. Otherwise, all are suppressed. Will be overridden by <code>readr.show_progress</code> and <code>readr.show_col_types</code> options, if they are set.

Value

A named list with "data" and "shape" elements or a combined sf data frame.

usa_states	<i>U.S. States Reference data</i>
------------	-----------------------------------

Description

Reference data with U.S. state names, USPS abbreviations, and Census divisions, and regions. Includes 50 U.S. States and the District of Columbia.

Usage

```
usa_states
```

Format

A data frame with 51 rows and 4 variables:

STATE State name

STUSPS State USPS abbreviation

division U.S. Census Division name

region U.S. Census Region name

Index

- * **datasets**
 - nhgis_ts_tables, 10
 - usa_states, 12
- * **ggplot2**
 - labs_nhgis, 8
- define_nhgis_ts_extract, 2
- define_nhgis_ts_extract(), 2
- download_ipumsr_extract, 4
- download_ipumsr_extract(), 4, 5
- dplyr::mutate(), 8
- dplyr::ntile(), 8
- ds_spec(), 3

- get_alt_text, 9
- get_ipumsr_extract_paths, 5
- get_metadata_nhgis(), 3
- get_nhgis_ts_data, 6
- ggplot2::labs, 9

- ipums_data_collections(), 4
- ipums_extract, 4, 5, 7
- ipumsr::define_extract_nhgis, 3
- ipumsr::define_extract_nhgis(), 2
- ipumsr::download_extract(), 4
- ipumsr::read_ipums_sf(), 12
- ipumsr::read_nhgis(), 12
- ipumsr::submit_extract, 5
- ipumsr::wait_for_extract, 4
- ipumsr::wait_for_extract(), 4

- join_nhgis_percent_change, 8
- join_nhgis_percent_change(), 8

- labs_nhgis, 8
- labs_nhgis(), 8
- list_nhgis_shapefiles(), 3, 7
- list_nhgis_ts_tables, 9

- nhgis_ts_tables, 10

- pivot_nhgis_data(), 2

- read_ipums_geometry, 10
- read_ipums_geometry(), 12
- read_nhgis(), 3, 7
- read_nhgis_data(), 2, 12
- read_nhgis_files, 11

- set_ipums_api_key(), 3, 5, 7
- set_ipums_default_collection(), 4
- stringr::str_wrap(), 9

- tidyselect selection, 11
- tst_spec(), 3, 7

- usa_states, 12