Package 'tidycensus'

September 28, 2020

Type Package

Title Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames

Version 0.10.2

Date 2020-09-28

URL https://github.com/walkerke/tidycensus

BugReports https://github.com/walkerke/tidycensus/issues

Description
An integrated R interface to the decennial US Census and American Community Survey APIs and the US Census Bureau's geographic boundary files. Allows R users to return Census and ACS data as tidyverse-ready data frames, and optionally returns a list-column with feature geometry for all Census geographies.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Depends R (>= 3.3.0)

Imports httr, sf, dplyr (>= 0.8.0), tigris, stringr, jsonlite (>= 1.5.0), purrr, rvest, tidyr (>= 1.0.0), rappdirs, readr, xml2, units, utils, rlang

Suggests ggplot2, survey, srvyr

RoxygenNote 7.1.1

NeedsCompilation no

Author Kyle Walker [aut, cre], Matt Herman [aut], Kris Eberwein [ctb]

Maintainer Kyle Walker <kyle@walker-data.com>

Repository CRAN

Date/Publication 2020-09-28 12:10:02 UTC

R topics documented:

`census_api_key` *Install a CENSUS API Key in Your* `.Renviron` *File for Repeated Use*

Description

This function will add your CENSUS API key to your `.Renviron` file so it can be called securely without being stored in your code. After you have installed your key, it can be called any time by typing `Sys.getenv("CENSUS_API_KEY")` and can be used in package functions by simply typing CENSUS_API_KEY If you do not have an `.Renviron` file, the function will create on for you. If you already have an `.Renviron` file, the function will append the key to your existing file, while making a backup of your original file for disaster recovery purposes.

Usage

```
census_api_key(key, overwrite = FALSE, install = FALSE)
```

Arguments

`key`	The API key provided to you from the Census formated in quotes. A key can be acquired at `http://api.census.gov/data/key_signup.html`
`overwrite`	If this is set to TRUE, it will overwrite an existing CENSUS_API_KEY that you already have in your `.Renviron` file.
`install`	if TRUE, will install the key in your `.Renviron` file for use in future sessions. Defaults to FALSE.

Examples

```
## Not run:
census_api_key("111111abc", install = TRUE)
# First time, reload your environment so you can use the key without restarting R.
readRenviron("~/.Renviron")
# You can check it with:
Sys.getenv("CENSUS_API_KEY")

## End(Not run)

## Not run:
# If you need to overwrite an existing key:
census_api_key("111111abc", overwrite = TRUE, install = TRUE)
# First time, relead your environment so you can use the key without restarting R.
readRenviron("~/.Renviron")
# You can check it with:
Sys.getenv("CENSUS_API_KEY")

## End(Not run)
```

---

county_laea *County geometry with Alaska and Hawaii shifted and re-scaled*

Description

Built-in dataset for use with `shift_geo = TRUE`

Dataset of US counties with Alaska and Hawaii shifted and re-scaled

Usage

```
data(county_laea)

data(county_laea)
```

Format

An object of class `sf` (inherits from `data.frame`) with 3143 rows and 2 columns.

Details

Dataset with county geometry for use when shifting Alaska and Hawaii

Built-in dataset for use with the `shift_geo` parameter, with the continental United States in a Lambert azimuthal equal area projection and Alaska and Hawaii counties and Census areas shifted and re-scaled. The data were originally obtained from the albersusa R package (https://github.com/hrbrmstr/albersusa).

fips_codes *Dataset with FIPS codes for US states and counties*

Description

Built-in dataset for smart state and county lookup. To access the data directly, issue the command `data(fips_codes)`.

- `county`: County name, title-case
- `county_code`: County code. (3-digit, 0-padded, character)
- `state`: Upper-case abbreviation of state
- `state_code`: State FIPS code (2-digit, 0-padded, character)
- `state_name`: Title-case name of state

Usage

```
data(fips_codes)
```

Format

An object of class `data.frame` with 3237 rows and 5 columns.

Details

Dataset with FIPS codes for US states and counties

Built-in dataset for use with the `lookup_code` function. To access the data directly, issue the command `data(fips_codes)`.

Note: this dataset includes FIPS codes for all counties that have appeared in the decennial Census or American Community Survey from 2010 to the present. This means that counties that have been renamed or absorbed into other geographic entities since 2010 remain in this dataset along with newly added or renamed counties.

If you need the FIPS codes and names for counties for a particular Census year, you can use the counties function from the tigris package and set the year parameter as required.

get_acs *Obtain data and feature geometry for the five-year American Community Survey*

Description

Obtain data and feature geometry for the five-year American Community Survey

Usage

```
get_acs(
  geography,
  variables = NULL,
  table = NULL,
  cache_table = FALSE,
  year = 2018,
  endyear = NULL,
  output = "tidy",
  state = NULL,
  county = NULL,
  geometry = FALSE,
  keep_geo_vars = FALSE,
  shift_geo = FALSE,
  summary_var = NULL,
  key = NULL,
  moe_level = 90,
  survey = "acs5",
  show_call = FALSE,
  ...
)
```

Arguments

geography	The geography of your data.
variables	Character string or vector of character strings of variable IDs. tidycensus automatically returns the estimate and the margin of error associated with the variable.
table	The ACS table for which you would like to request all variables. Uses lookup tables to identify the variables; performs faster when variable table already exists through `load_variables(cache = TRUE)`. Only one table may be requested per call.
cache_table	Whether or not to cache table names for faster future access. Defaults to FALSE; if TRUE, only needs to be called once per dataset. If variables dataset is already cached via the `load_variables` function, this can be bypassed.
year	The year, or endyear, of the ACS sample. 2009 through 2018 are available. Defaults to 2018.

endyear	Deprecated and will be removed in a future release.
output	One of "tidy" (the default) in which each row represents an enumeration unit-variable combination, or "wide" in which each row represents an enumeration unit and the variables are in the columns.
state	An optional vector of states for which you are requesting data. State names, postal codes, and FIPS codes are accepted. Defaults to NULL.
county	The county for which you are requesting data. County names and FIPS codes are accepted. Must be combined with a value supplied to 'state'. Defaults to NULL.
geometry	if FALSE (the default), return a regular tibble of ACS data. if TRUE, uses the tigris package to return an sf tibble with simple feature geometry in the 'geometry' column. state, county, tract, block group, block, and ZCTA geometry are supported.
keep_geo_vars	if TRUE, keeps all the variables from the Census shapefile obtained by tigris. Defaults to FALSE.
shift_geo	if TRUE, returns geometry with Alaska and Hawaii shifted for thematic mapping of the entire US. Geometry was originally obtained from the albersusa R package.
summary_var	Character string of a "summary variable" from the ACS to be included in your output. Usually a variable (e.g. total population) that you'll want to use as a denominator or comparison.
key	Your Census API key. Obtain one at http://api.census.gov/data/key_signup.html
moe_level	The confidence level of the returned margin of error. One of 90 (the default), 95, or 99.
survey	The ACS contains one-year, three-year, and five-year surveys expressed as "acs1", "acs3", and "acs5". The default selection is "acs5."
show_call	if TRUE, display call made to Census API. This can be very useful in debugging and determining if error messages returned are due to tidycensus or the Census API. Copy to the API call into a browser and see what is returned by the API directly. Defaults to FALSE.
...	Other keyword arguments

Value

A tibble or sf tibble of ACS data

Examples

```
## Not run:
library(tidycensus)
library(tidyverse)
library(viridis)
census_api_key("YOUR KEY GOES HERE")

tarr <- get_acs(geography = "tract", variables = "B19013_001",
```

```
                state = "TX", county = "Tarrant", geometry = TRUE)

ggplot(tarr, aes(fill = estimate, color = estimate)) +
  geom_sf() +
  coord_sf(crs = 26914) +
  scale_fill_viridis(option = "magma") +
  scale_color_viridis(option = "magma")


vt <- get_acs(geography = "county", variables = "B19013_001", state = "VT")

vt %>%
mutate(NAME = gsub(" County, Vermont", "", NAME)) %>%
 ggplot(aes(x = estimate, y = reorder(NAME, estimate))) +
  geom_errorbarh(aes(xmin = estimate - moe, xmax = estimate + moe)) +
  geom_point(color = "red", size = 3) +
  labs(title = "Household income by county in Vermont",
       subtitle = "2012-2016 American Community Survey",
       y = "",
       x = "ACS estimate (bars represent margin of error)")


## End(Not run)
```

`get_decennial` *Obtain data and feature geometry for the decennial Census*

Description

Obtain data and feature geometry for the decennial Census

Usage

```
get_decennial(
  geography,
  variables = NULL,
  table = NULL,
  cache_table = FALSE,
  year = 2010,
  sumfile = "sf1",
  state = NULL,
  county = NULL,
  geometry = FALSE,
  output = "tidy",
  keep_geo_vars = FALSE,
  shift_geo = FALSE,
  summary_var = NULL,
  key = NULL,
  show_call = FALSE,
```

```
  ...
)
```

Arguments

geography	The geography of your data.
variables	Character string or vector of character strings of variable IDs.
table	The Census table for which you would like to request all variables. Uses lookup tables to identify the variables; performs faster when variable table already exists through `load_variables(cache = TRUE)`. Only one table may be requested per call.
cache_table	Whether or not to cache table names for faster future access. Defaults to FALSE; if TRUE, only needs to be called once per dataset. If variables dataset is already cached via the `load_variables` function, this can be bypassed.
year	The year for which you are requesting data. 1990, 2000, and 2010 are available.
sumfile	The Census summary file. Defaults to sf1; the function will look in sf3 if it cannot find a variable in sf1.
state	The state for which you are requesting data. State names, postal codes, and FIPS codes are accepted. Defaults to NULL.
county	The county for which you are requesting data. County names and FIPS codes are accepted. Must be combined with a value supplied to 'state'. Defaults to NULL.
geometry	if FALSE (the default), return a regular tibble of ACS data. if TRUE, uses the tigris package to return an sf tibble with simple feature geometry in the 'geometry' column. state, county, tract, and block group are supported for 1990 through 2010; block and ZCTA geometry are supported for 2000 and 2010.
output	One of "tidy" (the default) in which each row represents an enumeration unit-variable combination, or "wide" in which each row represents an enumeration unit and the variables are in the columns.
keep_geo_vars	if TRUE, keeps all the variables from the Census shapefile obtained by tigris. Defaults to FALSE.
shift_geo	if TRUE, returns geometry with Alaska and Hawaii shifted for thematic mapping of the entire US. Geometry was originally obtained from the albersusa R package.
summary_var	Character string of a "summary variable" from the decennial Census to be included in your output. Usually a variable (e.g. total population) that you'll want to use as a denominator or comparison.
key	Your Census API key. Obtain one at http://api.census.gov/data/key_signup.html
show_call	if TRUE, display call made to Census API. This can be very useful in debugging and determining if error messages returned are due to tidycensus or the Census API. Copy to the API call into a browser and see what is returned by the API directly. Defaults to FALSE.
...	Other keyword arguments

Value

a tibble or sf tibble of decennial Census data

Examples

```
## Not run:
# Plot of race/ethnicity by county in Illinois for 2010
library(tidycensus)
library(tidyverse)
library(viridis)
census_api_key("YOUR KEY GOES HERE")
vars10 <- c("P005003", "P005004", "P005006", "P004003")

il <- get_decennial(geography = "county", variables = vars10, year = 2010,
                    summary_var = "P001001", state = "IL", geometry = TRUE) %>%
  mutate(pct = 100 * (value / summary_value))

ggplot(il, aes(fill = pct, color = pct)) +
  geom_sf() +
  facet_wrap(~variable)



## End(Not run)
```

---

get_estimates *Get data from the US Census Bureau Population Estimates APIs*

Description

Get data from the US Census Bureau Population Estimates APIs

Usage

```
get_estimates(
  geography,
  product = NULL,
  variables = NULL,
  breakdown = NULL,
  breakdown_labels = FALSE,
  year = 2018,
  state = NULL,
  county = NULL,
  time_series = FALSE,
  output = "tidy",
  geometry = FALSE,
  keep_geo_vars = FALSE,
  shift_geo = FALSE,
```

```
  key = NULL,
  show_call = FALSE,
  ...
)
```

Arguments

geography	The geography of your data.
product	The data product (optional). "population", "components" "housing", and "characteristics" are supported.
variables	A character string of requested variables to get specific variables from the population, components, and housing APIs.
breakdown	The population breakdown used when product = "characteristics". Acceptable values are "AGEGROUP", "RACE", "SEX", and "HISP", for Hispanic/Not Hispanic. These values can be combined in a vector, returning population estimates in the value column for all combinations of these breakdowns.
breakdown_labels	Whether or not to label breakdown elements returned when product = "characteristics". Defaults to FALSE.
year	The data year (defaults to 2018)
state	The state for which you are requesting data. State names, postal codes, and FIPS codes are accepted. Defaults to NULL.
county	The county for which you are requesting data. County names and FIPS codes are accepted. Must be combined with a value supplied to 'state'. Defaults to NULL.
time_series	If TRUE, the function will return a time series of observations back to the decennial Census of 2010. The returned column is either "DATE", representing a particular estimate date, or "PERIOD", representing a time period (e.g. births between 2016 and 2017), and contains integers representing those values. Integer to date or period mapping is available at https://www.census.gov/data/developers/data-sets/popest-popproj/popest/popest-vars/2018.html.
output	One of "tidy" (the default) in which each row represents an enumeration unit-variable combination, or "wide" in which each row represents an enumeration unit and the variables are in the columns.
geometry	if FALSE (the default), return a regular tibble of ACS data. if TRUE, uses the tigris package to return an sf tibble with simple feature geometry in the 'geometry' column.
keep_geo_vars	if TRUE, keeps all the variables from the Census shapefile obtained by tigris. Defaults to FALSE.
shift_geo	if TRUE, returns geometry with Alaska and Hawaii shifted for thematic mapping of the entire US.
key	Your Census API key. Obtain one at http://api.census.gov/data/key_signup.html. Can be stored in your .Renviron with census_api_key("YOUR KEY",install = TRUE)

show_call	if TRUE, display call made to Census API. This can be very useful in debugging and determining if error messages returned are due to tidycensus or the Census API. Copy to the API call into a browser and see what is returned by the API directly. Defaults to FALSE.
...	other keyword arguments

Value

A tibble, or sf tibble, of population estimates data

`get_pums` *Load data from the American Community Survey Public Use Microdata Series API*

Description

Load data from the American Community Survey Public Use Microdata Series API

Usage

```
get_pums(
  variables,
  state = NULL,
  puma = NULL,
  year = 2018,
  survey = "acs5",
  rep_weights = NULL,
  recode = FALSE,
  show_call = FALSE,
  key = NULL
)
```

Arguments

variables	A vector of variables from the PUMS API.
state	A state, or vector of states, for which you would like to request data. The entire US can be requested with `state = "all"` - though be patient with the data download!
puma	A vector of PUMAs from a single state, for which you would like to request data. To get data from PUMAs in more than one state, specify a named vector of state/PUMA pairs and set `state = "multiple"`.
year	The data year of the 1-year ACS sample or the endyear of the 5-year sample. Defaults to 2018.
survey	The ACS survey; one of either `"acs1"` or `"acs5"` (the default).

rep_weights	Whether or not to return housing unit, person, or both housing and person-level replicate weights for calculation of standard errors; one of "person", "housing", or "both".
recode	If TRUE, recodes variable values using Census data dictionary and creates a new *_label column for each variable that is recoded. Only available for 2017 and 2018 data. Defaults to FALSE.
show_call	If TRUE, display call made to Census API. This can be very useful in debugging and determining if error messages returned are due to tidycensus or the Census API. Copy to the API call into a browser and see what is returned by the API directly. Defaults to FALSE.
key	Your Census API key. Obtain one at http://api.census.gov/data/key_signup.html

Value

A tibble of microdata from the ACS PUMS API.

Examples

```
## Not run:
get_pums(variables = "AGEP", state = "VT")
get_pums(variables = c("AGEP", "ANC1P"), state = "VT", recode = TRUE)
get_pums(variables = "AGEP", state = "VT", survey = "acs1", rep_weights = "person")

## End(Not run)
```

---

load_variables *Load variables from a decennial Census or American Community Survey dataset to search in R*

Description

Load variables from a decennial Census or American Community Survey dataset to search in R

Usage

```
load_variables(year, dataset, cache = FALSE)
```

Arguments

year	The year for which you are requesting variables. Either the year of the decennial Census, or the endyear for a 5-year ACS sample.
dataset	One of "sf1", "sf3", "acs1", "acs3", "acs5", "acs1/profile", "acs3/profile, "acs5/profile", "acs1/subject", "acs3/subject", or "acs5/subject".
cache	Whether you would like to cache the dataset for future access, or load the dataset from an existing cache. Defaults to FALSE.

Value

A tibble of variables from the requested dataset.

Examples

```
## Not run:
v15 <- load_variables(2015, "acs5", cache = TRUE)
View(v15)

## End(Not run)
```

`moe_product` *Calculate the margin of error for a derived product*

Description

Calculate the margin of error for a derived product

Usage

```
moe_product(est1, est2, moe1, moe2)
```

Arguments

`est1`	The first factor in the multiplication equation (an estimate)
`est2`	The second factor in the multiplication equation (an estimate)
`moe1`	The margin of error of the first factor
`moe2`	The margin of error of the second factor

Value

A margin of error for a derived product

`moe_prop` *Calculate the margin of error for a derived proportion*

Description

Calculate the margin of error for a derived proportion

Usage

```
moe_prop(num, denom, moe_num, moe_denom)
```

Arguments

num	The numerator involved in the proportion calculation (an estimate)
denom	The denominator involved in the proportion calculation (an estimate)
moe_num	The margin of error of the numerator
moe_denom	The margin of error of the denominator

Value

A margin of error for a derived proportion

---

moe_ratio *Calculate the margin of error for a derived ratio*

Description

Calculate the margin of error for a derived ratio

Usage

```
moe_ratio(num, denom, moe_num, moe_denom)
```

Arguments

num	The numerator involved in the ratio calculation (an estimate)
denom	The denominator involved in the ratio calculation (an estimate)
moe_num	The margin of error of the numerator
moe_denom	The margin of error of the denominator

Value

A margin of error for a derived ratio

moe_sum *Calculate the margin of error for a derived sum*

Description

Generates a margin of error for a derived sum. The function requires a vector of margins of error involved in a sum calculation, and optionally a vector of estimates associated with the margins of error. If the associated estimates are not specified, the user risks inflating the derived margin of error in the event of multiple zero estimates. It is recommended to inspect your data for multiple zero estimates before using this function and setting the inputs accordingly.

Usage

```
moe_sum(moe, estimate = NULL, na.rm = FALSE)
```

Arguments

moe	A vector of margins of error involved in the sum calculation
estimate	A vector of estimates, the same length as moe, associated with the margins of error
na.rm	A logical value indicating whether missing values (including NaN) should be removed

Value

A margin of error for a derived sum

See Also

https://www2.census.gov/programs-surveys/acs/tech_docs/accuracy/MultiyearACSAccuracyofData2015.pdf

pums_variables *Dataset with PUMS variables and codes*

Description

Built-in dataset for variable name and code label lookup. To access the data directly, issue the command `data(pums_variables)`.

- survey: acs1 or acs5
- year: Year of data. For 5-year data, last year in range.
- var_code: Variable name
- var_label: Variable label

- data_type: chr or num
- level: housing or person
- val_min: For numeric variables, the minimum value
- val_max: For numeric variables, the maximum value
- val_label: Value label
- recode: Use labels to recode values
- val_length: Length of value returned
- val_na: Value of NA value returned by API (if known)

Usage

```
data(pums_variables)
```

Format

An object of class tbl_df (inherits from tbl, data.frame) with 21114 rows and 12 columns.

Details

Dataset with PUMS variables and codes

Built-in dataset that is created from the Census PUMS data dictionaries. Use this dataset to lookup the names of variables to use in get_pums. This dataset also contains labels for the coded values returned by the Census API and is used when recode = TRUE in get_pums.

Because variable names and codes change from year to year, you should filter this dataset for the survey and year of interest. NOTE: only 2017 and 2018 (acs1 and acs5) variables are available.

---

significance *Evaluate whether the difference in two estimates is statistically significant.*

Description

Evaluate whether the difference in two estimates is statistically significant.

Usage

```
significance(est1, est2, moe1, moe2, clevel = 0.9)
```

Arguments

est1	The first estimate.
est2	The second estimate
moe1	The margin of error of the first estimate
moe2	The margin of error of the second estimate
clevel	The confidence level. May by 0.9, 0.95, or 0.99

Value

TRUE if the difference is statistically signifiant, FALSE otherwise.

See Also

https://www.census.gov/content/dam/Census/library/publications/2018/acs/acs_general_handbook_2018_ch07.pdf

---

state_laea *State geometry with Alaska and Hawaii shifted and re-scaled*

Description

Built-in dataset for use with `shift_geo = TRUE`

Dataset of US states with Alaska and Hawaii shifted and re-scaled

Usage

```
data(state_laea)

data(state_laea)
```

Format

An object of class `sf` (inherits from `data.frame`) with 51 rows and 2 columns.

Details

Dataset with state geometry for use when shifting Alaska and Hawaii

Built-in dataset for use with the `shift_geo` parameter, with the continental United States in a Lambert azimuthal equal area projection and Alaska and Hawaii shifted and re-scaled. The data were originally obtained from the albersusa R package (https://github.com/hrbrmstr/albersusa).

---

tidycensus *Return tidy data frames from the US Census Bureau API*

Description

This packages uses US Census Bureau data but is neither endorsed nor supported by the US Census Bureau.

Author(s)

Kyle Walker

to_survey *Convert a data frame returned by get_pums() to a survey object*

Description

This helper function takes a data frame returned by get_pums and converts it to a tbl_svy from the srvyr as_survey package or a svyrep.design object from the svrepdesign package. You can then use functions from the srvyr or survey to calculate weighted estimates with replicate weights included to provide accurate standard errors.

Usage

```
to_survey(
  df,
  type = c("person", "housing"),
  class = c("srvyr", "survey"),
  design = c("rep_weights", "cluster")
)
```

Arguments

df	A data frame with PUMS person or housing weight variables, most likely returned by get_pums.
type	Whether to use person or housing-level weights; either "housing" or "person" (the default).
class	Whether to convert to a srvyr or survey object; either "survey" or "srvyr" (the default).
design	Whether to use a cluster or replicate weight survey design; either "cluster" or "rep_weights" (the default).

Value

A tbl_svy or svyrep.design object.

Examples

```
## Not run:
pums <- get_pums(variables = "AGEP", state = "VT", rep_weights = "person")
pums_design <- to_survey(pums, type = "person", class = "srvyr")
survey::svymean(~AGEP, pums_design)

## End(Not run)
```

Index