Package 'census2020download'

November 23, 2022

Title 2020 Census Blocks and Tools to Download from FTP

Version 2.1.1

Date 2022-11-11

Description This package has basic functions for downloading from FTP,
unzipping, reading the 2020 Census data for some or all US States
into a single data.frame. It provides a dataset of all US blocks
with a few key variables like population count, lat/lon of internal point,
FIPS. It downloads and also could retain some other columns like
area, and some basic race ethnicity counts,
with small changes to some source code.
EPA may at some point provide the same data in a package such as EJAMblockdata.

Depends R (>= 3.5.0)

Imports data.table,
SearchTrees,
usethis,
readr

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.1

NeedsCompilation no

Author Mark Corrales [aut, cre]

Maintainer Mark Corrales <corrales.mark@epa.gov>

R topics documented:

`blockid2fips` *block id for each block fips code for US Census blocks*

Description

BUT SEE EJAMblockdata package For populated Census blocks only, from Census block dataset.

`blockquadtree` *quad tree format data on locations of US Census 2020 blocks*

Description

BUT SEE EJAMblockdata package Used to quickly find which of these block points are nearby a specified facility or site with circular buffer.

```
Formal class 'QuadTree' [package "SearchTrees"] with 7 slots
..@ ref      :<externalptr>
..@ numNodes : int 5173
..@ dataNodes: int 3294
..@ maxDepth : int 7
..@ maxBucket: int 48705
..@ totalData: int 5806512
..@ dataType : chr "point"
..$ data  :<externalptr>
..$ points: int 5806512
..$ year  : num 2020
#  code in this package created quaddata and blockquadtree:

 quaddata <- blockdata[ , .(BLOCK_X, BLOCK_Z, BLOCK_Y, blockid)]
blockquadtree <- SearchTrees::createTree(quaddata, treeType = "quad", dataType = "point")
```

blockwts2020 *2020 Decennial Census block group weights*

Description

BUT SEE EJAMblockdata package A data.table with blockwt, and blockid or blockfips, bgid or bgfips used by doaggregate() to summarize the block group score of the average person in the buffer, as the population weighted mean of blockgroup scores of all the blocks in the buffer.

Details

This is drawn from blocks2020 dataset

```
blockfips is the 15 character Census Bureau FIPS code for each Census block
blockid will be a unique integer ID 1 through total number of blocks,
  to be used as a more efficient indexing than a FIPS code is.

bgfips from blocks2020 is the 12 character Census Bureau FIPS code
  for the parent blockgroup (i.e., the one containing the given block).
  and bgfips is used to join to a blockgroup dataset to get indicator scores.
bgid will be a unique integer ID 1 through total number of unique blockgroups,
  to be used as a more efficient indexing than a FIPS code is.

blockwt is the block's population weight,
calculated as the block population as a fraction of the parent blockgroup pop.
Based on the latest decennial Census table of population count for each block.

This table of weights can be used for calculating the
weighted mean or sum of each blockgroup score for a buffer
where only some blocks of any given blockgroup are in the buffer.
The sum of weights from some blocks tells you
what fraction of its whole parent blockgroup's population count
is in those blocks (the ones found inside a buffer, for example).
```

See https://www.epa.gov/ejscreen

census2020download *2020 Census Blocks and Tools to Download from FTP*

Description

Basic functions for downloading from FTP, unzipping, reading the 2020 Census data for some or all US States into a single data.table.

Details

This package provides a dataset of all US blocks with a few key variables like population count, lat/lon of internal point, FIPS. This package downloads and also could retain some other columns like area, and some basic race ethnicity counts, with small changes to some source code. The proxistat package has a similar data.table but also with area and population count. EPA may at some point provide the same data in a package such as EJAMblockdata.

To use quaddata and blockquadtree for fast search of blocks: localtree <- SearchTrees::createTree(EJAMblockdata::quaddata, treeType = "quad", dataType = "point")

Key functions and data.tables include

- `census2020_get_data()` Download/ Unzip/ Read/ Clean basic data on all US Census blocks.
- `census2020_save_datasets()` Use the data to create the separate data.tables, and save for use in the package.
- `census2020_download()` Download all the state files from FTP site
- `census2020_unzip()` Unzip the downloads
- `census2020_read()` Read the unzipped files
- `census2020_clean()` Clean up what was read
- `blockpoints` - data.table with latitude and longitude
- `blockid2fips` - data.table with FIPS code to blockid lookup
- `bgid2fips` - data.table with FIPS codes to block group id lookup
- `blockwts` - data.table with Census 2020 population-based weight as fraction of parent block group population
- `quaddata` - data.table with xyz format locations of blocks
- `blockquadtree` A quadtree index, for fast search to find nearby blocks

---

census2020_clean *start to clean up download block data*

Description

Renames variables (columns) based on census_col_names_map Drops columns not needed. Returns it in data.table format.

This is part of how the output of census2020_read() can be cleaned up and split into smaller data files, to be used in EJAM.

Usage

```
census2020_clean(x, cols_to_keep = c("blockfips", "lat", "lon", "pop", "area"))
```

Arguments

x	data from census2020_read()
cols_to_keep	optional, which (renamed) columns to retain and return

Value

data.table with these columns by default: blockfips lat lon pop area

census2020_download *Download Census 2020 data files from FTP*

Description

Warning: Code is not tested. Attempts to download data files for specified states from the US Census Bureau's FTP site for Decennial Census file data. see https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/summary-file/2020Census_PL94_171Redistricting_StatesTechDoc_English.pdf

Usage

```
census2020_download(folder = getwd(), mystates)
```

Arguments

folder	Default is getwd()
mystates	Character vector of 2 letter abbreviations, now optional - Default is 50 states + DC + PR here

Value

Effect is to download and save locally a number of data files.

See Also

census2020_read census2020_unzip

Examples

```
## Not run:
 library(census2020download)
 census2020_download('./census2020zip', mystates = c('MD', 'DC'))
 census2020_unzip('./census2020zip','./census2020out')
 c2 <- census2020_read(folder = './census2020out', mystates = c('MD', 'DC'))
 save(c2,file = 'census2020blocks.rdata')
 dim(c2)
 str(c2)
 head(c2)
 sum(c2$POP100)
 plot(c2$INTPTLON[substr(c2$GEOCODE,1,2)=='24'], c2$INTPTLAT[substr(c2$GEOCODE,1,2)=='24'], pch='.')

## End(Not run)
```

census2020_get_data *Download and clean up block data census 2020*

Description

Download and clean up block data census 2020

Usage

```
census2020_get_data(
  folder = "~/../Downloads/census2020zip",
  folderout = "~/../Downloads/census2020out",
  mystates = c(state.abb, "DC"),
  ...
)
```

Arguments

folder	path for downloaded files
folderout	path for assembled results files
mystates	default is DC and the 50 states, not PR not VI/GU/MU/AS
...	passed to census2020_read()

Value

a data.table of US Census blocks with columns like blockid lat lon pop area (area in square meters I think)

census2020_read *Compile Census 2020 block data for all US states once downloaded and unzipped*

Description

Not tested. Works for file 1, but maybe not files 2 and 3. Not used for EJAM - that Census 2020 data was from EJScreen team. ttempts to read files already downloaded and unzipped, data files for specified states from the US Census Bureau's FTP site for Decennial Census file data. see https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/summary-file/2020Census_PL94_171Redistricting_StatesTechDoc_English.pdf

Usage

```
census2020_read(
  folder = ".",
  filenumbers = 1,
  mystates,
  sumlev = 750,
 best_header_cols = c("LOGRECNO", "GEOCODE", "AREALAND", "AREAWATR", "POP100", "HU100",
    "INTPTLAT", "INTPTLON"),
  best_data_cols = paste0("P00", (20001:20011))
)
```

Arguments

folder	path
filenumbers	a vector with any or all of 1,2,3 – default is file 1. File01 has Tables P1 and P2. File02 has Tables P3, P4, and H1. File03 has Table P5.
mystates	can be vector of 2-letter abbreviations of states
sumlev	default is 750, for blocks
best_header_cols	default is a few key columns like POP100, GEOCODE (fips), etc.
best_data_cols	default is key race ethnicity fields

Details

Also look at the package totalcensus https://github.com/GL-Li/totalcensus see Census website for list of possible fields etc.

```
   for example:
#  AREALAND      Area (Land)
#  AREAWATR      Area (Water)

#  BASENAME      Area Base Name
# NAME       Area Name-Legal/Statistical Area Description (LSAD) Term-Part Indicator
#  FUNCSTAT      Functional Status Code
#  GCUNI         Geographic Change User Note Indicator

#  POP100        Population Count (100
#  HU100         Housing Unit Count (100
#  INTPTLAT      Internal Point (Latitude)
#  INTPTLON      Internal Point (Longitude)

File 1 has Table P1 and
 Table P2. HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE
Universe: Total population
P0020001 P0020002 P0020003 P0020004 P0020005 P0020006 P0020007 P0020008 P0020009 P0020010 P0020011
Total: P0020001
 Hispanic or Latino P0020002
Not Hispanic or Latino: P0020003
 Population of one race: P0020004
 White alone P0020005
Black or African American alone P0020006
American Indian and Alaska Native alone P0020007
Asian alone P0020008
Native Hawaiian and Other Pacific Islander alone P0020009
Some Other Race alone P0020010
Population of two or more races: Population of two races: P0020011
```

Value

data.frame of 1 row per block, for example

See Also

census2020_download census2020_unzip

Examples

```
## Not run:
 library(census2020download)
 census2020_download('./census2020zip', mystates = c('MD', 'DC'))
 census2020_unzip('./census2020zip','./census2020out')
 c2 <- census2020_read(folder = './census2020out', mystates = c('MD', 'DC'))
 dim(c2)
 str(c2)
 head(c2)
 sum(c2$POP100)
 plot(c2$INTPTLON[substr(c2$GEOCODE,1,2)=='24'], c2$INTPTLAT[substr(c2$GEOCODE,1,2)=='24'], pch='.')
 c2$LOGRECNO <- NULL
 colnames(c2) <- census_col_names_map$Rname[match(colnames(blocks2020), census_col_names_map$ftpname)]

## End(Not run)
```

---

census2020_save_datasets

create each separate data.table, and optionally save for use in package This was just done once, to create the datasets in this package

---

Description

create each separate data.table, and optionally save for use in package This was just done once, to create the datasets in this package

Usage

```
census2020_save_datasets(x, metadata = NULL, usethis = FALSE, overwrite = TRUE)
```

Arguments

x	results of census2020_get_data()
metadata	default is Census 2020 related
usethis	default is FALSE, but if TRUE will install each dataset in package
overwrite	default is TRUE, but only relevant if usethis = TRUE

Value

a list of huge data.tables, bgid2fips, blockid2fips, blockpoints, blockwts, quaddata, blockquadtree

census2020_unzip *unzip Census 2020 zipped files already downloaded*

Description

Warning: Code is not tested. Attempts to read files already downloaded and unzipped, data files for specified states from the US Census Bureau's FTP site for Decennial Census file data. see https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/summary-file/2020Census_PL94_171Redistricting_StatesTechDoc_English.pdf

Usage

```
census2020_unzip(folder = ".", folderout = folder, filenumbers = 1, mystates)
```

Arguments

folder	path to where zip files are, default is working directory
folderout	path to where you want to put files, created if does not exist
filenumbers	a vector with any or all of 1,2,3 – default is file 1. File01 has Tables P1 and P2, which have population counts by race ethnicity. File02 has Tables P3, P4, and H1. File03 has Table P5.
mystates	optional vector of 2letter state abbreviations for which to unzip

Value

Vector of filenames of unzipped contents

See Also

census2020_download census2020_read

Examples

```
## Not run:
 library(census2020download)
 census2020_download('./census2020zip', mystates = c('MD', 'DC'))
 census2020_unzip('./census2020zip','./census2020out')
 c2 <- census2020_read(folder = './census2020out', mystates = c('MD', 'DC'))
 save(c2,file = 'census2020blocks.rdata')
 dim(c2)
 str(c2)
 head(c2)
 sum(c2$POP100)
 plot(c2$INTPTLON[substr(c2$GEOCODE,1,2)=='24'], c2$INTPTLAT[substr(c2$GEOCODE,1,2)=='24'], pch='.')

## End(Not run)
```

census_col_names_map *table that maps from a few Census variable names to short friendly variable names*

Description

table that maps from a few Census variable names to short friendly variable names

lookup_states *basic information about US States*

Description

FIPS code for each State, name, etc., and also for DC, PR, and Island Areas

quaddata *quad tree data on locations of US Census blocks*

Description

BUT SEE EJAMblockdata package This has selected columns from Census block dataset.

```
Code that created blockdata, quaddata, blockquadtree:
  (also localtree has to be created during runtime.
  cannot be saved as data file beforehand)

blockdata <- 0
quaddata <- blockdata[ , .(BLOCK_X, BLOCK_Z, BLOCK_Y, blockid)]
blockquadtree <- SearchTrees::createTree(quaddata, treeType = "quad", dataType = "point")

2020 data format:

str(quaddata)
Classes ‘data.table’ and 'data.frame': 5806512 obs. of  4 variables:
   $ BLOCK_X: num  205 205 204 204 204 ...
 $ BLOCK_Z: num  2125 2125 2125 2126 2126 ...
 $ BLOCK_Y: num  -3334 -3334 -3334 -3334 -3334 ...
 $ blockid: int  1 2 3 4 5 6 7 8 9 10 ...
 - attr(*, ".internal.selfref")=<externalptr>
   - attr(*, "year")= num 2020
```

See Also

blockquadtree

Index