Package: ppmf (via r-universe)

October 31, 2026

Title Read Census Privacy Protected Microdata Files

Version 0.2.1

Date 2025-09-03

Description Implements data processing described in <doi:10.1126/sciadv.abk3283> to align modern differentially private data with formatting of older US Census data releases. The primary goal is to read in Census Privacy Protected Microdata Files data in a reproducible way. This includes tools for aggregating to relevant levels of geography by creating geographic identifiers which match the US Census Bureau's numbering. Additionally, there are tools for grouping race numeric identifiers into categories, consistent with OMB (Office of Management and Budget) classifications. Functions exist for downloading and linking to existing sources of privacy protected microdata.

License MIT + file LICENSE

Config/Needs/website christopherkenny/ctktemplate

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

BugReports https://github.com/christopherkenny/ppmf/issues

URL https://github.com/christopherkenny/ppmf/, https://christophertkenny.com/ppmf/

RoxygenNote 7.3.2

Imports censable, dplyr, readr, rlang (>= 0.4.11), stringr, tibble, tidyr, zip

Suggests roxygen2, spelling

Depends R (>= 4.1.0)

Language en-US

Config/pak/sysreqs libabsl-dev cmake libgdal-dev gdal-bin libgeos-dev libicu-dev libssl-dev libproj-dev libsqlite3-dev libudunits2-dev libx11-dev

Repository https://christopherkenny.r-universe.dev

Date/Publication 2026-05-01 16:15:25 UTC

RemoteUrl https://github.com/christopherkenny/ppmf

RemoteRef HEAD

RemoteSha 94f561fdb1a80dadf10adafde62e20144714226a

Contents

add_geoid . . . 2
add_ppmf12_path . . . 3
add_ppmf19_path . . . 4
add_ppmf19r_path . . . 5
add_ppmf4_path . . . 5
agg . . . 6
breakdown_geoid . . . 7
download_ppmf . . . 8
get_ppmf_links . . . 8
overwrite_hisp_race . . . 9
ppmf_ex . . . 10
races . . . 10
read_merge_ppmf . . . 11
read_ppmf . . . 12
replace_race . . . 12
states . . . 13

Index **14**

`add_geoid` *Add Standard GEOID to PPMF Data*

Description

Adds the GEOID identifier common to spatial census data sets, such as those loaded by tigris. This allows for easier merging or aggregation by a single variable.

Usage

```
add_geoid(
  ppmf,
  state = TABBLKST,
  county = TABBLKCOU,
  tract = TABTRACT,
  block_group = TABBLKGRP,
  block = TABBLK,
  level = "block"
)
```

Arguments

ppmf	tibble of ppmf data
state	Column in ppmf with state (fips) ID. Default is TABBLKST.
county	Column in ppmf with county (fips) ID. Default is TABBLKCOU.
tract	Column in ppmf with tract ID. Default is TABBLKTRACT.
block_group	Column in ppmf with block group ID. Default is TABBLKGRP
block	Column in ppmf with block ID. Default is TABBLK.
level	Geographic level to write the GEOID for. Options are block (default), block_group, tract, and county.

Value

input data ppmf with added column GEOID

Examples

```
data(ppmf_ex)
ppmf_ex <- ppmf_ex |> add_geoid()
```

---

add_ppmf12_path *Add ppmf12 path to Renviron*

Description

Add ppmf12 path to Renviron

Usage

```
add_ppmf12_path(path, overwrite = FALSE, install = FALSE)
```

Arguments

path	path where ppmf12 data is stored
overwrite	Defaults to FALSE. Should existing ppmf12 in Renviron be overwritten?
install	Defaults to FALSE. Should ppmf12 be added to '~/.Renviron' file?

Value

path, invisibly

Examples

```
## Not run:
tp <- tempfile(fileext = '.csv')
add_ppmf12_path(tp)
path12 <- Sys.getenv('path12')

## End(Not run)
```

`add_ppmf19_path` *Add ppmf19 path to Renviron*

Description

Add ppmf19 path to Renviron

Usage

```
add_ppmf19_path(path, overwrite = FALSE, install = FALSE)
```

Arguments

`path`	path where ppmf19 data is stored
`overwrite`	Defaults to FALSE. Should existing ppmf19 in Renviron be overwritten?
`install`	Defaults to FALSE. Should ppmf19 be added to '~/.Renviron' file?

Value

path, invisibly

Examples

```
## Not run:
tp <- tempfile(fileext = '.csv')
add_ppmf19_path(tp)
path19 <- Sys.getenv('path19')

## End(Not run)
```

add_ppmf19r_path *Add ppmf19r path to Renviron*

Description

Path for the 19.61 *r*eplication in 2023.

Usage

```
add_ppmf19r_path(path, overwrite = FALSE, install = FALSE)
```

Arguments

path	path where ppmf19r data is stored
overwrite	Defaults to FALSE. Should existing ppmf19 in Renviron be overwritten?
install	Defaults to FALSE. Should ppmf19r be added to '~/.Renviron' file?

Value

path, invisibly

Examples

```
## Not run:
tp <- tempfile(fileext = '.csv')
add_ppmf19r_path(tp)
path19 <- Sys.getenv('path19')

## End(Not run)
```

add_ppmf4_path *Add ppmf4 path to Renviron*

Description

Add ppmf4 path to Renviron

Usage

```
add_ppmf4_path(path, overwrite = FALSE, install = FALSE)
```

Arguments

path	path where ppmf4 data is stored
overwrite	Defaults to FALSE. Should existing ppmf4 in Renviron be overwritten?
install	Defaults to FALSE. Should ppmf4 be added to '~/.Renviron' file?

Value

path, invisibly

Examples

```
## Not run:
tp <- tempfile(fileext = '.csv')
add_ppmf4_path(tp)
path4 <- Sys.getenv('path4')

## End(Not run)
```

---

agg *Aggregate PPMF Data*

Description

Aggregate PPMF Data

Usage

```
agg(ppmf, group = GEOID, age = VOTING_AGE, race = CENRACE, hisp = CENHISP)
```

Arguments

ppmf	tibble of ppmf data
group	Column in ppmf to group by, typically GEOID
age	Column in ppmf containing 1 for not voting age and 2 for voting age
race	Column in ppmf containing race codes
hisp	Column in ppmf containing 1 for Not Hispanic and 2 for Hispanic

Value

tibble of ppmf data aggregated by group with race classified with columns:

- group: named by entry group
- pop: total population
- pop_hisp: total population - Hispanic or Latino (of any race)
- pop_white: total population - White alone, not Hispanic or Latino
- pop_black: total population - Black or African American alone, not Hispanic or Latino
- pop_aian: total population - American Indian and Alaska Native alone, not Hispanic or Latino
- pop_asian: total population - Asian alone, not Hispanic or Latino
- pop_nhpi: total population - Native Hawaiian and Other Pacific Islander alone, not Hispanic or Latino

- pop_other: total population - Some Other Race alone, not Hispanic or Latino
- pop_two: total population - Population of two or more races, not Hispanic or Latino
- vap: voting age population
- vap_hisp: voting age population - Hispanic or Latino (of any race)
- vap_white: voting age population - White alone, not Hispanic or Latino
- vap_black: voting age population - Black or African American alone, not Hispanic or Latino
- vap_aian: voting age population - American Indian and Alaska Native alone, not Hispanic or Latino
- vap_asian: voting age population - Asian alone, not Hispanic or Latino
- vap_nhpi: voting age population - Native Hawaiian and Other Pacific Islander alone, not Hispanic or Latino
- vap_other: voting age population - Some Other Race alone, not Hispanic or Latino
- vap_two: voting age population - Population of two or more races, not Hispanic or Latino

Examples

```
data(ppmf_ex)
ppmf_ex <- ppmf_ex |> add_geoid()
blocks <- agg(ppmf_ex)
```

---

breakdown_geoid *Breakdown GEOID into Components*

Description

Breakdown GEOID into Components

Usage

```
breakdown_geoid(ppmf, GEOID = GEOID)
```

Arguments

ppmf	tibble of ppmf data
GEOID	Column in ppmf with GEOID. Default is GEOID.

Value

tibble. ppmf with columns added for state, county, tract, block group, and/or block

Examples

```
data(ppmf_ex)
ppmf_ex <- ppmf_ex |> add_geoid()
ppmf_ex <- ppmf_ex |> censable::breakdown_geoid()
```

download_ppmf *Download PPMF Files*

Description

Downloads zipped ppmf files from GitHub.

Usage

```
download_ppmf(dsn, dir = "", version = "19r", overwrite = FALSE)
```

Arguments

dsn	(data save name) string to unzip the data to
dir	the folder or directory to save the file in
version	string in '19r', '19', '12' or '4' signifying the revised 19.61, original 19.61, 12.2 or 4.5 versions respectively
overwrite	If a file is found at path/dsn, should it be overwritten? Defaults to FALSE.

Value

a string path to where the file was downloaded to

Examples

```
## Not run:
# Takes a few minutes and requires read access to files
temp <- tempdir()
path <- download_ppmf(dsn = 'ppmf_12', dir = temp)

## End(Not run)
```

get_ppmf_links *Get PPMF File Links*

Description

Returns the urls for the data. This will be expanded to link to prior or any new releases.

Usage

```
get_ppmf_links(version = "19r", compressed = TRUE)
```

Arguments

version	string in '19r',, '19', '12' or '4' signifying the 19.61, 12.2, or 4.5 versions respectively
compressed	boolean. Return a compressed version (TRUE). FALSE gives the Census Bureau link to the uncompressed data.

Value

a string with url

Examples

```
# 04.28.2021 version 12.2
get_ppmf_links()
# 04.28.2021 version 4.5
get_ppmf_links(version = '4')
```

---

overwrite_hisp_race *Overwrite Races with Hispanic*

Description

Overwrite Races with Hispanic

Usage

```
overwrite_hisp_race(ppmf, race = CENRACE, hisp = CENHISP)
```

Arguments

ppmf	tibble of ppmf data
race	Column in ppmf containing race codes
hisp	Column in ppmf containing 1 for Not Hispanic and 2 for Hispanic

Value

tibble with race column entries replaced if the individual is Hispanic

Examples

```
data(ppmf_ex)
ppmf_ex |> replace_race() |> overwrite_hisp_race()
```

ppmf_ex *Example PPMF Data*

Description

Includes Perry County, Alabama PPMF data from the April 28, 2021 PPMF data release. This is a subset taken from the 12-2 P data.

As each observation is a person, this does not cover every block in the county and due to DAS, not every block with population appears in this data.

Usage

```
data('ppmf_ex')
```

Value

tibble with sample ppmf data

Examples

```
data('ppmf_ex')
```

races *Race Classifications*

Description

This data includes the basic race classifications used for redistricting to get to an easier to work with set of values. This does not include `hisp` grouping which is controlled separately by race within the census

Usage

```
data('races')
```

Value

tibble with three columns

- code: the two digit code used to code races
- desc: the description of the races
- group: the summary group used

Examples

```
data('races')
```

read_merge_ppmf *Read PPMF data and Merge with Census 2010 Data*

Description

Read PPMF data and Merge with Census 2010 Data

Usage

```
read_merge_ppmf(
  state,
  level,
  versions = c("19"),
  prefixes = paste0("v", versions, "_"),
  paths = Sys.getenv(paste0("ppmf", versions))
)
```

Arguments

state	state abbreviation
level	geography level. One of 'block', 'block group', 'tract', 'county'
versions	character vector of ppmf versions. Currently '19', '12', and/or '4'
prefixes	prefixes to give pop and vap columns in output. Default is `paste0('v', versions, '_')`
paths	paths to PPMF data. Default is `Sys.getenv(paste0('ppmf', versions))`

Value

sf tibble of PPMF merged with Census 2010 data

Examples

```
## Not run:
# Requires Census Bureau API
de_bg <- read_merge_ppmf('DE', 'block group')

## End(Not run)
```

read_ppmf *Read in PPMF Data*

Description

This reads in PPMF data from a file. Use `download_ppmf()` if you do not have a local copy of the ppmf data.

Usage

```
read_ppmf(state, path, ...)
```

Arguments

`state`	two letter state (+ DC + PR) abbreviation or two digit state fips code
`path`	where the data is saved to
`...`	additional arguments passed on to `readr::read_csv()`

Value

tibble of ppmf data

Examples

```
## Not run:
# Takes a few minutes and requires read access to files
temp <- tempdir()
path <- download_ppmf('ppmf_12.csv', dir = temp)
# If you already have it downloaded, point to it with path:
ppmf <- read_ppmf('AL', path)

## End(Not run)
```

replace_race *Replace Race Categories*

Description

Replaces the Census's numeric categories for race with less specific racial classifications, typically useful for redistricting purposes.

Usage

```
replace_race(ppmf, race = CENRACE)
```

Arguments

ppmf	tibble of ppmf data
race	Column in ppmf containing race codes

Value

tibble with race column replaced by simpler racial classifications

Examples

```
data(ppmf_ex)
ppmf_ex |> replace_race()
```

---

states *State Rows*

Description

This data includes the 52 geographies (50 states plus D.C. and P.R.). Within the 2010 PPMF, skip and n_max indicate the relevant rows for a geography.

Usage

```
data('states')
```

Value

tibble with sample ppmf data

Examples

```
data('states')
```

Index

∗ **basic**
  agg, 6
  overwrite_hisp_race, 9
  replace_race, 12
∗ **data**
  ppmf_ex, 10
  races, 10
  states, 13
∗ **geoid**
  add_geoid, 2
  breakdown_geoid, 7
∗ **getdata**
  download_ppmf, 8
  get_ppmf_links, 8
  read_merge_ppmf, 11
  read_ppmf, 12
∗ **save**
  add_ppmf12_path, 3
  add_ppmf19_path, 4
  add_ppmf19r_path, 5
  add_ppmf4_path, 5

add_geoid, 2
add_ppmf12_path, 3
add_ppmf19_path, 4
add_ppmf19r_path, 5
add_ppmf4_path, 5
agg, 6

breakdown_geoid, 7

download_ppmf, 8

get_ppmf_links, 8

overwrite_hisp_race, 9

ppmf_ex, 10

races, 10
read_merge_ppmf, 11
read_ppmf, 12
readr::read_csv(), *12*
replace_race, 12

states, 13