Package 'pseudohouseholds'

July 23, 2025

Type Package

Title Generate Pseudohouseholds on Road Networks in Regions

Version 0.1.1

Description Given an arbitrary set of spatial regions and road networks, generate a set of representative points, or pseudohouseholds, that can be used for travel burden analysis. Parallel processing is supported.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Imports dplyr, furrr, sf

Depends R (>= 2.10)

Suggests covr, future, ggplot2, ggspatial, knitr, purrr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

URL https://github.com/chris31415926535/pseudohouseholds

BugReports https://github.com/chris31415926535/pseudohouseholds/issues

VignetteBuilder knitr

NeedsCompilation no

Author Christopher Belanger [aut, cre, cph] (ORCID: <https://orcid.org/0000-0003-2070-5721>)

Maintainer Christopher Belanger <christopher.a.belanger@gmail.com>

Repository CRAN

Date/Publication 2023-07-19 09:30:02 UTC

Contents

`get_phhs_parallel` *Get Pseudo-Households (PHH) for many regions, with optional parallel processing*

Description

Calculate PHHs for a set of regions using a given road network.

Usage

```
get_phhs_parallel(
  regions,
  region_idcol,
  roads,
  region_popcol = NA,
  roads_idcol = NA,
  phh_density = 0.005,
  min_phh_pop = 5,
  min_phhs_per_region = 1,
  min_phh_distance = 25,
  road_buffer_m = 5,
  delta_distance_m = 5,
  skip_unpopulated_regions = TRUE
)
```

Arguments

`regions`	simple feature object, sf tibble where each row is a region
`region_idcol`	character, name of column with unique region id
`roads`	simple feature object, lines or polylines with road network
`region_popcol`	character, name of column with region population
`roads_idcol`	character, name of column containing road unique identifiers
`phh_density`	numeric, parameter given to sf::st_line_sample()
`min_phh_pop`	numeric, minimum population per phh

`min_phhs_per_region`
: numeric, minimum phhs per region (it will try its best)

`min_phh_distance`
: numeric, minimum distance between phhs in meters

`road_buffer_m`
: numeric, buffer in meters for intersections

`delta_distance_m`
: numeric, buffer in meters for intersections

`skip_unpopulated_regions`
: boolean, should we skip regions with no population?

Details

Regions will be processed sequentially by default, but parallel processing is supported if users call future::plan() before calling this function.

This function is a wrapper around get_phhs_single(), and parameters are passed on to it.

Value

a simple feature object with one row per phh in the region

Examples

```
# Create PHHs for the first 2 dissemination blocks in Ottawa, Ontario, without
# using any parallel processing
library(sf)
library(pseudohouseholds)
phhs <- get_phhs_parallel(region = ottawa_db_shp[1:2,], region_idcol = "DBUID",
region_popcol = "dbpop2021", roads = ottawa_roads_shp, roads_idcol = "NGD_UID")


# Create PHHs for the first 5 dissemination blocks in Ottawa, Ontario, using
# parallel processing (consult documentation for the package future for details
# about parallel processing).

library(future)
future::plan(future::multisession)
phhs <- get_phhs_parallel(region = ottawa_db_shp[1:5,], region_idcol = "DBUID",
 region_popcol = "dbpop2021", roads = ottawa_roads_shp, roads_idcol = "NGD_UID")

# Shut down parallel workers
future::plan(future::sequential)
```

`get_phhs_single` *Get Pseudo-Households (PHH) for a single region*

Description

Get Pseudo-Households (PHH) for a single region

Usage

```
get_phhs_single(
  region,
  region_idcol,
  roads,
  region_popcol = NA,
  roads_idcol = NA,
  phh_density = 0.005,
  min_phh_pop = 5,
  min_phhs_per_region = 1,
  min_phh_distance = 25,
  road_buffer_m = 5,
  delta_distance_m = 5,
  skip_unpopulated_regions = TRUE,
  track_warnings = FALSE
)
```

Arguments

`region`	simple feature object, one-row sf tibble
`region_idcol`	character, name of column with unique region id
`roads`	simple feature object, lines or polylines with road network
`region_popcol`	character, name of column with region population
`roads_idcol`	character, name of column containing road unique identifiers
`phh_density`	numeric, parameter given to sf::st_line_sample()
`min_phh_pop`	numeric, minimum population per phh
`min_phhs_per_region`	numeric, minimum phhs per region (it will try its best)
`min_phh_distance`	numeric, minimum distance between phhs in meters
`road_buffer_m`	numeric, buffer in meters for intersections
`delta_distance_m`	numeric, buffer in meters for intersections
`skip_unpopulated_regions`	boolean, should we skip regions with no population?
`track_warnings`	boolean, internal parameter used when this function is called by get_phhs_parallel() to ensure warnings are only shown once.

Value

a simple feature object with one row per phh in the region

Examples

```
phhs <- get_phhs_single(region = region_shp, region_idcol = "region_id",
region_popcol = "population", roads = road_shp, roads_idcol = "road_id")
```

`ottawa_db_shp` *2021 Statistics Canada Dissemination Block Boundaries and Populations for Ottawa, Ontario*

Description

Spatial dataset for dissemination blocks (DBs) in Ottawa, Ontario, provided by Statistics Canada, https://www12.statcan.gc.ca/census-recensement/2021/geo/sip-pis/boundary-limites/index2021-eng.cfm?year=21.

Usage

```
ottawa_db_shp
```

Format

A data frame with class sf in CRS NAD/MTM zone 9 (32189) and 8,559 rows and 3 variables:

DBUID Unique dissemination block identifier

dbpop2021 Dissemination block population as given in the 2021 Statistics Canada geographic attribute file, https://www12.statcan.gc.ca/census-recensement/2021/geo/aip-pia/attribute-attribs/index2021-eng.cfm

geometry MULTIPOLYGON defining DB geometry

Details

Dissemination blocks are the smallest unit of geography at which Statistics Canada publishes population data. DBs are generally bounded by road segments or natural features like waterways. In urban areas DBs are generally the size of a city block, but in rural areas they can be much larger.

ottawa_roads_shp *2021 Statistics Canada Road Network for Ottawa, Ontario*

Description

Spatial dataset for road networks in Ottawa, Ontario, provided by Statistics Canada, https://www12.statcan.gc.ca/census-recensement/2021/geo/sip-pis/rnf-frr/index2021-eng.cfm?year=21.

Usage

```
ottawa_roads_shp
```

Format

A data frame with class sf in CRS NAD/MTM zone 9 (32189) and 33,983 rows and 5 variables:

NGD_UID Unique road segment identifier

NAME Road segment name

RANK Road rank, lower numbers generally mean bigger/faster roads, https://www12.statcan.gc.ca/census-recensement/2021/geo/ref/domain-domaine/index2021-eng.cfm?lang=e&id=RANK

CLASS Road class, lower numbers generally but do not always mean bigger/faster roads, https://www12.statcan.gc.ca/census-recensement/2021/geo/ref/domain-domaine/index2021-eng.cfm?lang=e&id=CLASS

geometry LINESTRING defining road segment geometry

This data is licensed under the Statistics Canada Open Data License (https://www.statcan.gc.ca/en/reference/licence). Adapted from Statistics Canada, 2021 Census Road Network File, 2022-09-28. This does not constitute an endorsement by Statistics Canada of this product.

region_shp *Synthetic region shapefile for testing*

Description

A small shapefile for testing.

Usage

```
region_shp
```

Format

An object of class `sf` (inherits from `data.frame`) with 1 rows and 3 columns.

road_shp *Synthetic road shapefile for testing*

Description

A small shapefile for testing.

Usage

```
road_shp
```

Format

An object of class `sf` (inherits from `data.frame`) with 1 rows and 2 columns.

validate_phhs *Validate Pseudohouseholds (PHHs)*

Description

This function runs two tests to ensure that PHHs meet minimal criteria for validity: it checks to see whether PHH populations sum accurately to region populations, and whether each populated region has at least one PHH. Results are returned in a data frame, and any failing regions are returned in a list- column that can be used for filtering and further analysis. Note that these tests may fail if PHHs were generated without using population data.

Usage

```
validate_phhs(phhs, regions, region_idcol, region_popcol)
```

Arguments

`phhs`	A data frame containing a set of PHHs.
`regions`	A simple feature object, sf tibble where each row is a region, used to generate the PHHs.
`region_idcol`	Character, the name of the column in both ‘phhs‘ and ‘regions‘ containing regional identifiers.
`region_popcol`	Character, the name of the column in both ‘phhs‘ and ‘regions‘ containing population data.

Value

A data frame containing test outputs.

Examples

```
phhs <- get_phhs_single(region = region_shp, region_idcol = "region_id",
region_popcol = "population", roads = road_shp, roads_idcol = "road_id")
validate_phhs(phhs = phhs, regions = region_shp, region_idcol = "region_id",
region_popcol = "population")
```

Index