Package 'lehdr'

November 2, 2026

Type Package

Title Grab Longitudinal Employer-Household Dynamics (LEHD) Flat Files

Version 1.2.1

Maintainer Jamaal Green <jamaal.green@gmail.com>

Description Designed to query Longitudinal Employer-Household Dynamics (LEHD) workplace/residential association and origin-destination flat files and optionally aggregate Census block-level data to block group, tract, county, or state. Also provides analytical functions for commute flow statistics, longitudinal change, and earnings tier distributions. Data comes from the LODES FTP server <https://lehd.ces.census.gov/data/lodes/LODES8/>.

License MIT + file LICENSE

URL https://jamgreen.github.io/lehdr/, https://github.com/jamgreen/lehdr/

BugReports https://github.com/jamgreen/lehdr/issues/

Depends R (>= 4.1.0)

Imports dplyr, glue, httr2, magrittr, readr, rlang, stringr, tidyr

Suggests ggplot2, knitr, pkgdown, rmarkdown, sf, testthat (>= 3.0.0), tigris, withr

VignetteBuilder knitr

Config/testthat/edition 3

Encoding UTF-8

Language en

Config/roxygen2/version 8.1.1

NeedsCompilation no

Author Jamaal Green [cre, aut] (ORCID: <https://orcid.org/0000-0002-8322-8654>), Liming Wang [aut], Dillon Mahmoudi [aut] (ORCID: <https://orcid.org/0000-0002-8816-1639>), Matthew Rogers [ctb],

Kyle Walker [ctb],
Eli Pousson [ctb] (ORCID: <https://orcid.org/0000-0001-8280-1706>)

Repository CRAN

Date/Publication 2026-11-01 23:20:02 UTC

Contents

compute_commute_stats 2
compute_earnings_share 3
compute_lodes_change 5
grab_crosswalk 6
grab_lodes 7

Index **10**

`compute_commute_stats` *Compute commute flow statistics from LODES OD data*

Description

Derives three key commute flow metrics from an OD tibble: inflow, outflow, net flow, and the self-containment ratio. These metrics are widely used in transportation planning and economic geography to characterize labor market catchment areas and job/housing balance.

Self-containment is defined as the proportion of workers who both live and work within the same geographic unit, relative to all employed residents. A value close to 1 indicates a highly self-contained labor market; values near 0 indicate heavy out-commuting.

Net flow is defined as inbound workers minus outbound workers (including internal flows as both in and out). This is an unsigned flow balance indicator: positive values signal net job importers (more workers arrive than leave); negative values signal net exporters.

Cross-state commuters: When `state_part = "main"` is used in `grab_lodes()`, only workers who live and work in the same state are included. Workers who cross state lines (e.g., Maryland residents working in DC) appear only in `state_part = "aux"` files for the *workplace* state. To capture full commute flows for border counties, retrieve both `"main"` and `"aux"` files and bind the rows before calling `compute_commute_stats()`.

Usage

```
compute_commute_stats(od_df, agg_geo = "tract")
```

Arguments

`od_df`	A data frame (tibble) of LODES origin-destination data returned by `grab_lodes()` with `lodes_type = "od"`. Must include columns `h_{agg_geo}` (home geography), `w_{agg_geo}` (work geography), and `S000` (total job count). The data frame may be at any aggregation level supported by `agg_geo`.

Note on row structure: LODES OD files are a flow matrix. A call to `grab_lodes()` with `lodes_type = "od"` returns one row per *origin-destination pair*, not one row per geography – even after aggregation via `agg_geo`. For example, a county-level OD pull for West Virginia returns ~2,800 rows (one per observed county-county flow pair), not 55 (the number of counties). Pass the result directly to `compute_commute_stats()` to reduce the pair table to one row per geography with inflow, outflow, net flow, and self-containment.

`agg_geo` The geographic level of the OD data. Must match the level at which `od_df` was retrieved or aggregated. One of `"block"`, `"bg"`, `"tract"`, `"county"`, or `"state"`. Defaults to `"tract"`.

Value

A tibble with one row per geography, containing:

`{agg_geo}` The geographic identifier.

`year` Year of the data (if present in `od_df`).

`state` State FIPS abbreviation (if present in `od_df`).

`workers_in` Total workers arriving (working in this geography).

`workers_out` Total workers departing (living here, working elsewhere).

`workers_internal` Workers whose home and work are both in this geography (internal flows).

`net_flow` Net worker flow: `workers_in - workers_out`. Positive values indicate net job importers.

`self_containment` Share of resident workers who also work in this geography: `workers_internal / workers_out_total`, where `workers_out_total` includes internal flows.

Examples

```
## Not run:
  od <- grab_lodes(
    state = "md", year = 2019,
    lodes_type = "od", job_type = "JT00",
    segment = "S000", state_part = "main",
    agg_geo = "county"
  )
  compute_commute_stats(od, agg_geo = "county")

## End(Not run)
```

`compute_earnings_share` *Compute earnings tier shares from LODES RAC or WAC data*

Description

Computes the share of jobs (or workers) in each of the three LODES monthly earnings tiers:

Low (`CE01`) Earnings up to $1,250/month.

Mid (`CE02`) Earnings $1,251-$3,333/month.

High (`CE03`) Earnings above $3,333/month.

Earnings shares are useful for tracking wage polarization, identifying low-wage job concentration, and examining how the earnings structure of a labor market has shifted over time, especially when combined with `compute_lodes_change()`.

The total denominator is the sum of the three tiers, ensuring shares sum to 1 within rounding error.

Usage

```
compute_earnings_share(
  lodes_df,
  type = c("wac", "rac"),
  geo_col = NULL,
  output = c("wide", "long")
)
```

Arguments

`lodes_df`	A data frame (tibble) of LODES RAC or WAC data returned by `grab_lodes()` with `segment = "S000"` (the total count segment, which includes all three earnings columns). Must contain columns `CE01`, `CE02`, and `CE03`. Both WAC and RAC files include these earnings tier columns; the `type` argument controls only which geography prefix is used for auto-detection.
`type`	One of `"wac"` (workplace area characteristics, default) or `"rac"` (residential area characteristics). Controls the expected geography column prefix (`w_` for WAC, `h_` for RAC) when `geo_col = NULL`. Does not affect which earnings columns are used, as both file types share the `CE01`/`CE02`/`CE03` schema.
`geo_col`	The name of the geography column to group by, e.g. `"w_tract"` or `"h_county"`. Defaults to `NULL`, which auto-detects the first geography column in `lodes_df`.
`output`	One of `"wide"` (default) or `"long"`. Wide format appends three share columns. Long format returns one row per geography-tier combination.

Value

A tibble with earnings tier counts and shares. In `"wide"` format, columns are added for `share_low`, `share_mid`, and `share_high`. In `"long"` format, columns are `tier`, `label`, `count`, and `share`.

Examples

```
## Not run:
  wac <- grab_lodes(
    state = "md", year = 2019,
    lodes_type = "wac", job_type = "JT00",
```

```
  segment = "S000", agg_geo = "county"
)
compute_earnings_share(wac, type = "wac", geo_col = "w_county")

# Long format, suitable for ggplot2
compute_earnings_share(
  wac, type = "wac", geo_col = "w_county", output = "long"
)

## End(Not run)
```

`compute_lodes_change` *Compute longitudinal change in LODES data across years*

Description

Computes absolute and percentage change in LODES variables between two years for each geographic unit. This is particularly useful for tracking shifts in employment, industrial composition, or earnings structure over time.

Percentage change is computed as $(\text{compare} - \text{base})/\text{base} \times 100$. Returns `NA` where `base_value` is zero or missing.

Usage

```
compute_lodes_change(
  lodes_df,
  geo_col = NULL,
  base_year = NULL,
  compare_year = NULL,
  variables = NULL,
  output = c("wide", "long")
)
```

Arguments

`lodes_df`	A data frame (tibble) of LODES data covering at least two years, as returned by `grab_lodes()` with a vector of years. Must contain a `year` column and one or more numeric columns to difference.
`geo_col`	The name of the geography column to group by, e.g. `"w_tract"`, `"h_county"`, or `"w_geocode"`. Defaults to `NULL`, which auto-detects the first column ending in `_geocode`, `_tract`, `_county`, `_bg`, or `_state`.
`base_year`	The reference year for computing change. Defaults to the earliest year present in `lodes_df`.
`compare_year`	The target year for computing change. Defaults to the latest year present in `lodes_df`.
`variables`	Optional character vector of numeric column names to include in the output. Defaults to all numeric columns (excluding `year`).

output — One of "wide" (default) or "long". In wide format, absolute and percentage change columns are appended for each variable. In long format, each variable is a row with columns variable, base_value, compare_value, change, and pct_change.

Value

A tibble of change statistics. In "wide" format, columns follow the pattern {variable}_base, {variable}_compare, {variable}_change, and {variable}_pct_change. In "long" format, columns are variable, base_value, compare_value, change, and pct_change.

Examples

```
## Not run:
  wac_multi <- grab_lodes(
    state = "md", year = c(2015, 2019),
    lodes_type = "wac", job_type = "JT00",
    segment = "S000", agg_geo = "county"
  )
  compute_lodes_change(wac_multi, geo_col = "w_county")

  # Long format
  compute_lodes_change(
    wac_multi,
    geo_col  = "w_county",
    output   = "long",
    variables = c("C000", "CE01", "CE02", "CE03")
  )

## End(Not run)
```

---

grab_crosswalk — *Download and load LODES geographic crosswalk into a data frame (tibble)*

Description

Download the LODES geographic crosswalk for one or more states. The crosswalk maps Census block GEOIDs to higher-level geographies and is useful for custom aggregations outside the built-in agg_geo argument in grab_lodes().

Usage

```
grab_crosswalk(
  state,
  version = c("LODES8", "LODES7", "LODES5"),
 download_dir = normalizePath(file.path(tools::R_user_dir("lehdr", which = "cache")),
    mustWork = FALSE)
)
```

Arguments

state	US state abbreviation in lower case. Can be a vector of state abbreviations, e.g. c("wy", "nd", "sd").
version	The LODES version whose crosswalk to download. Must be one of "LODES8" (default), "LODES7", or "LODES5". The crosswalk maps Census blocks to higher-level geographies for that vintage.
download_dir	Directory where the crosswalk file will be downloaded. Defaults to the user-level cache directory for lehdr.

Value

A tibble containing the geographic crosswalk at the Census block level, with columns linking blocks to block groups, tracts, counties, and states.

Examples

```
## Not run:
  # Download crosswalk for Vermont
  vt_xwalk <- grab_crosswalk("vt")

  # Download crosswalk for several small states
  small_xwalk <- grab_crosswalk(c("wy", "nd", "sd"))

  # Download a LODES7 crosswalk (2010 Census block vintage)
  vt_xwalk_7 <- grab_crosswalk("vt", version = "LODES7")

## End(Not run)
```

---

grab_lodes *Download and load LODES data into a data frame (tibble)*

Description

Download LODES OD, RAC, and WAC tables from the LEHD FTP server and return a tidy data frame.

Usage

```
grab_lodes(
  state,
  year,
  version = c("LODES8", "LODES7", "LODES5"),
  lodes_type = c("od", "rac", "wac"),
  job_type = c("JT00", "JT01", "JT02", "JT03", "JT04", "JT05"),
 segment = c("S000", "SA01", "SA02", "SA03", "SE01", "SE02", "SE03", "SI01", "SI02",
    "SI03"),
  agg_geo = c("block", "bg", "tract", "county", "state"),
```

```
  state_part = c("", "main", "aux"),
 download_dir = normalizePath(file.path(tools::R_user_dir("lehdr", which = "cache")),
   mustWork = FALSE),
  geometry = FALSE,
  use_cache = getOption("lehdr_use_cache", FALSE),
  ...
)
```

Arguments

state	US state abbreviation in lower case, as character. Can be a vector of states, like c("or", "md", "tx") for Oregon, Maryland, and Texas. Two-letter FIPS abbreviations only.
year	Year of the LODES data, as a numeric integer. Can be a vector of years, like c(2014, 2020) for 2014 and 2020. Must be between 2002 and 2023.
version	The LODES version to use. "LODES8" (the default) is enumerated at 2020 Census blocks and covers 2002-2023. "LODES7" is enumerated at 2010 Census blocks and covers 2002-2019. "LODES5" is enumerated at 2000 Census blocks and covers 2002-2009.
lodes_type	The LODES table type. Values can be the default origin-destination ("od"), residential area characteristics ("rac"), or workplace area characteristics ("wac"). OD files give a home and workplace census block for each worker flow. RAC files give job totals at worker home blocks; WAC files give job totals at worker job blocks.
job_type	Job type segment: "JT00" for all jobs (default), "JT01" for Primary Jobs, "JT02" for All Private Jobs, "JT03" for Private Primary jobs, "JT04" for All Federal jobs, "JT05" for Federal Primary jobs.
segment	Workforce segment. "S000" total jobs (default); "SA01" workers aged 29 or younger; "SA02" workers aged 30-54; "SA03" workers 55 and older; "SE01" earnings $1,250/month or less; "SE02" earnings $1,251-$3,333/month; "SE03" earnings above $3,333/month; "SI01" Goods Producing industries; "SI02" Trade, Transportation, & Utilities industries; "SI03" All Other Services.
agg_geo	Aggregate to a geography other than Census Block (default). Values can be "bg" (block group), "tract", "county", or "state". The string "block group" is also accepted as an alias for "bg".
state_part	Part of the state file; required when lodes_type = "od". "main" includes workers whose home and workplace are both in-state. "aux" includes workers who live out-of-state but work in the state of interest. Defaults to "main" with a warning when lodes_type = "od" and state_part is not explicitly supplied.
download_dir	Directory where the LODES file will be downloaded. Defaults to the user-level cache directory for lehdr.
geometry	If TRUE, use the tigris package to download and attach spatial geometries from the U.S. Census Bureau for the specified year at the level specified by agg_geo. Returns an sf object when lodes_type is "rac" or "wac".
use_cache	Boolean. If TRUE, reuse previously downloaded files rather than re-downloading. Defaults to FALSE. You can also set the lehdr_use_cache global option to TRUE to make caching the default for a session.

... Additional arguments passed to tigris functions when geometry = TRUE.

Value

A tibble of LODES data aggregated to block, block group, tract, county, or state level. If geometry = TRUE and lodes_type is "rac" or "wac", returns an sf object.

Examples

```
## Not run:
  # Download 2014 block-level OD data for Vermont
  blk_od <- grab_lodes(
    state = "vt", year = 2014,
    lodes_type = "od", job_type = "JT01",
    segment = "SA01", state_part = "main"
  )

  # Download 2014 OD data for Vermont aggregated to tract level
  trt_od <- grab_lodes(
    state = "vt", year = 2014,
    lodes_type = "od", job_type = "JT01",
    segment = "SA01", state_part = "main",
    agg_geo = "tract"
  )

  # Download 2020 RAC data for Vermont aggregated to tract level
  trt_rac <- grab_lodes(
    state = "vt", year = 2020,
    lodes_type = "rac", job_type = "JT01",
    segment = "SA01", agg_geo = "tract"
  )

  # Download 2020 WAC data for Vermont aggregated to tract level
  trt_wac <- grab_lodes(
    state = "vt", year = 2020,
    lodes_type = "wac", job_type = "JT01",
    segment = "SA01", agg_geo = "tract"
  )

## End(Not run)
```

Index

compute_commute_stats, 2
compute_earnings_share, 3
compute_lodes_change, 5
compute_lodes_change(), *4*

grab_crosswalk, 6
grab_lodes, 7
grab_lodes(), *2–6*