

Package: tidylodes (via r-universe)

September 9, 2024

Type Package

Title tidylodes: A package for extracting, processing and 'spatialising' data from the US Census Bureau's LODES database

Version 0.1.0

Author Patrick Ballantyne <P.J.Ballantyne@liverpool.ac.uk> [aut, cre]

Maintainer Pukar Bhandari <pukar.bhandari@outlook.com>

Description This package enables users to extract datasets directly from the US Census Bureau's database of Longitudinal Origin-Destination Employment Statistics (LODES). The database contains three different datasets - the first two Workplace Area Characteristics (WAC) and Residential Area Characteristics (RAC) contain estimates of total jobs and jobs in specific sectors (e.g. Retail Trade, Manufacturing). The third dataset, contains OD flows of jobs from residential areas to workplace areas. This package is novel in that it not only cleans these datasets and improves the recording of census geographies within them, but also converts these datasets to spatial formats that enable spatial manipulation/analysis to be carried out.

Depends R (>= 3.5.0)

Imports data.table, dplyr, R.utils, sf, stringr, tigris,

VignetteBuilder knitr

License CC0

Encoding UTF-8

RoxygenNote 7.1.0

Repository <https://ar-puuk.r-universe.dev>

RemoteUrl <https://github.com/ar-puuk/tidylodes>

RemoteRef HEAD

RemoteSha b16f8dedabd9830129792e6e6a15c6269be08f62

Contents

get_jobsector_rac	2
get_jobsector_wac	3
get_od_data	4
get_od_spatial	5
get_od_subset	6
get_rac_data	7
get_rac_spatial	8
get_wac_data	9
get_wac_spatial	10

Index	11
--------------	-----------

get_jobsector_rac	<i>get_jobsector_rac</i>
-------------------	--------------------------

Description

This function allows users to subset the output dataframe of `get_wac_data` to focus on one specific job sector. This will be useful for those using RAC data to perform analysis on the geography of specific job types.

Usage

```
get_jobsector_rac(df, job_code, job_proportion = T)
```

Arguments

<code>df</code>	The input for this function is the output dataframe from <code>get_rac_data</code> .
<code>job_code</code>	Here the user can select a specific job code from the output dataframe of <code>get_rac_data</code> , which enables the function to drop all other job sectors, but keep total jobs and the chosen job sector. Users should use <code>colnames(df)</code> to obtain a list of all available job codes.
<code>job_proportion</code>	This argument enables users to calculate the proportion of the chosen job sector in relation to the total number of jobs in each census block. By default the argument is 'T', so will calculate a <code>job_proportion</code> column unless set to 'F'.

Value

A dataframe of cleaned RAC data, focusing on a specific job sector. If `job_proportion = T` then the dataframe will also contain an additional column where the proportion of total jobs that the chosen job sector occupies is calculated.

Author(s)

Ballantyne, Patrick

References

LODES data available to download manually from: <https://lehd.ces.census.gov/data/lodes/LODES7/>

Examples

```
## Users need to use the get_rac_data function first:
df <- get_rac_data("nj", "2008")
## Get a list of all available job sectors
colnames(df)

## Use the get_jobsector_rac function to extract retail trade jobs, and calculate the proportion of retail trade jobs
df2 <- get_jobsector_rac(df, job_code = "Retail_Trade", job_proportion = T)
```

get_jobsector_wac *get_jobsector_rac*

Description

This function allows users to subset the output dataframe of `get_wac_data` to focus on one specific job sector. This will be useful for those using WAC data to perform analysis on the geography of specific job types.

Usage

```
get_jobsector_wac(df, job_code, job_proportion = T)
```

Arguments

<code>df</code>	The input for this function is the output dataframe from <code>get_wac_data</code> .
<code>job_code</code>	Here the user can select a specific job code from the output dataframe of <code>get_wac_data</code> , which enables the function to drop all other job sectors, but keep total jobs and the chosen job sector. Users should use <code>colnames(df)</code> to obtain a list of all available job codes.
<code>job_proportion</code>	This argument enables users to calculate the proportion of the chosen job sector in relation to the total number of jobs in each census block. By default the argument is 'T', so will calculate a <code>job_proportion</code> column unless set to 'F'.

Value

A dataframe of cleaned WAC data, focusing on a specific job sector. If `job_proportion = T` then the dataframe will also contain an additional column where the proportion of total jobs that the chosen job sector occupies is calculated.

Author(s)

Ballantyne, Patrick

References

LODES data available to download manually from: <https://lehd.ces.census.gov/data/lodes/LODES7/>

Examples

```
## Users need to use the get_wac_data function first:
df <- get_wac_data("de", "2015")
## Get a list of all available job sectors
colnames(df)

## Use the get_jobsector_wac function to extract retail trade jobs, and calculate the proportion of retail trade jobs
df2 <- get_jobsector_wac(df, job_code = "Retail_Trade", job_proportion = T)
```

get_od_data

get_od_data

Description

This function enables users to extract job flows from Residence Areas (Origin Areas) to Workplace areas (Destination Areas) for a chosen state and a chosen year. The function grabs the OD data, merges with a lookup of census block geographies and processes the dataset. The output dataframe contains a "Total_Jobs" column which indicates the total estimated job flows from a residence area to a workplace area. These dataframes are bigger than those obtained by `get_wac_data` or `get_rac_data` as they contain multiple rows of data for each residence area, corresponding to the various workplace areas that job flows occur between.

Usage

```
get_od_data(state_name, year, main = T)
```

Arguments

<code>state_name</code>	Users need to give the lowercase abbreviated state name of any US state, to enable the function to grab OD data for that state.
<code>year</code>	Users need to give a year between 2002-2017, to enable the function to grab the OD data for that year, and for the state identified with the <code>state_name</code> argument.
<code>main</code>	As default, <code>main</code> is set to 'T', which means the function is grabbing OD flows where residence and workplace areas are in the same state. Setting <code>main</code> to 'F' will give OD data where the workplace areas are in the chosen state, but the residence areas are outside the state.

Value

Function returns a cleaned OD dataset for the chosen state and chosen year.

Author(s)

Ballantyne, Patrick

References

LODES data available to download manually from: <https://lehd.ces.census.gov/data/lodes/LODES7/>

Examples

```
## Get OD data for Illinois, from 2014
df <- get_od_data("il", "2014", main = T)

## Get OD data for Illinois, from 2014, but where residence areas are outside Illinois state
df2 <- get_od_data("il", "2014", main = F)
```

get_od_spatial *get_od_spatial*

Description

Function that converts output of `get_od_data` or `get_od_subset` into a format that enables plotting of job flows from residence to workplace areas. The function creates centroids for each census block and joins the coordinates of these onto the workplace and residence census blocks.

Usage

```
get_od_spatial(df)
```

Arguments

`df` Requires as input an output dataframe of `get_od_data` or `get_od_subset`.

Value

Returns a dataframe with two individual simple feature collection columns containing the point geometries for the workplace area census blocks and the residence area census blocks.

Author(s)

Ballantyne, Patrick

References

LODES data available to download manually from: <https://lehd.ces.census.gov/data/lodes/LODES7/>

Examples

```
## Get OD data for Illinois, for 2014
df <- get_od_data("il", "2014")

## Convert to format that enables linestring plotting
df2 <- get_od_spatial(df)
```

get_od_subset	<i>get_od_subset</i>
---------------	----------------------

Description

Function that takes the output of `get_od_data`, and subsets it to include only rows of data where flows are greater than a specified threshold.

Usage

```
get_od_subset(df, flow_threshold)
```

Arguments

<code>df</code>	Function requires as input the output dataframe of <code>get_od_data</code>
<code>flow_threshold</code>	Function requires specification of a threshold of job flows. The function will subset the dataframe based on this value, to keep only rows of data with flows greater than the specified value.

Value

Returns a dataframe with identical column structure of outputs of `get_od_data`, but with fewer rows as determined by value.

Author(s)

Ballantyne, Patrick

References

LODES data available to download manually from: <https://lehd.ces.census.gov/data/lodes/LODES7/>

Examples

```
## Get OD data for Illinois, for 2014
df <- get_od_data("il", "2014")

## Subset to include only flows of over 45
df2 <- get_od_subset(df, flow_threshold = 45)
```

get_rac_data	<i>get_rac_data</i>
--------------	---------------------

Description

This function enables users to extract datasets on Residence Area Characteristics (RAC) at a census block level for a chosen state, and for a chosen year between 2002-17. The RAC dataset(s) give employment estimates for a variety of specific job sectors (e.g. manufacturing), and estimates of total jobs, for each census block that is classified as a 'residence area'. This function grabs RAC data for the state and year chosen, and cleans it returning the output as a dataframe.

Usage

```
get_rac_data(state_name, year)
```

Arguments

state_name	Users need to give the lowercase abbreviated state name of any US state, to enable the function to grab RAC data for that state.
year	Users need to give a year between 2002-2017, to enable the function to grab the RAC data for that year, and for the state identified with the state_name argument.

Value

A dataframe of cleaned RAC data for the chosen state and chosen year.

Author(s)

Ballantyne, Patrick

References

LODES data available to download manually from: <https://lehd.ces.census.gov/data/lodes/LODES7/>

Examples

```
## e.g. Get RAC data for New Jersey from 2008  
df <- get_rac_data("nj", "2008")
```

<code>get_rac_spatial</code>	<code>get_rac_spatial</code>
------------------------------	------------------------------

Description

This function allows users to convert output dataframe(s) of `get_rac_data` or `get_jobsector_wac` into simple features, by joining the dataframe onto a census block simple feature for the state of interest, using the TIGRIS package to obtain the census blocks sf.

Usage

```
get_rac_spatial(df)
```

Arguments

`df` This function requires as input a cleaned RAC dataframe either as a direct output of `get_rac_data`, or of `get_jobsector_rac`.

Value

Returns WAC data in a spatial (simple features) format, that can be easily manipulated, mapped or used in spatial operations.

Author(s)

Ballantyne, Patrick

References

LODES data available to download manually from: <https://lehd.ces.census.gov/data/lodes/LODES7/>
For more information on the TIGRIS package visit: <https://CRAN.R-project.org/package=tigris>

Examples

```
## Obtain a RAC dataset
df <- get_rac_data("ak", "2013")

## Convert to simple features
sf <- get_rac_spatial(df)
```

get_wac_data *get_wac_data*

Description

This function enables users to extract datasets on Workplace Area Characteristics (WAC) at a census block level for a chosen state, and for a chosen year between 2002-17. The WAC dataset(s) give employment estimates for a variety of specific job sectors (e.g. manufacturing), and estimates of total jobs, for each census block that is classified as a 'workplace area'. This function grabs WAC data for the state and year chosen, and cleans it returning the output as a dataframe.

Usage

```
get_wac_data(state_name, year)
```

Arguments

state_name	Users need to give the lowercase abbreviated state name of any US state, to enable the function to grab WAC data for that state.
year	Users need to give a year between 2002-2017, to enable the function to grab the WAC data for that year, and for the state identified with the state_name argument.

Value

A dataframe of cleaned WAC data for the chosen state and chosen year.

Author(s)

Ballantyne, Patrick

References

LODES data available to download manually from: <https://lehd.ces.census.gov/data/lodes/LODES7/>

Examples

```
## e.g. Get WAC data for delaware from 2015  
df <- get_wac_data("de", "2015")
```

<code>get_wac_spatial</code>	<i>get_wac_spatial</i>
------------------------------	------------------------

Description

This function allows users to convert output dataframe(s) of `get_wac_data` or `get_jobsector_wac` into simple features, by joining the dataframe onto a census block simple feature for the state of interest, using the `TIGRIS` package to obtain the census blocks `sf`.

Usage

```
get_wac_spatial(df)
```

Arguments

`df` This function requires as input a cleaned WAC dataframe either as a direct output of `get_wac_data`, or of `get_jobsector_wac`.

Value

Returns WAC data in a spatial (simple features) format, that can be easily manipulated, mapped or used in spatial operations.

Author(s)

Ballantyne, Patrick

References

LODES data available to download manually from: <https://lehd.ces.census.gov/data/lodes/LODES7/>
For more information on the `TIGRIS` package visit: <https://CRAN.R-project.org/package=tigris>

Examples

```
## Obtain a WAC dataset
df <- get_wac_data("ak", "2013")

## Convert to simple features
sf <- get_wac_spatial(df)
```

Index

`get_jobsector_rac`, 2
`get_jobsector_wac`, 3
`get_od_data`, 4
`get_od_spatial`, 5
`get_od_subset`, 6
`get_rac_data`, 7
`get_rac_spatial`, 8
`get_wac_data`, 9
`get_wac_spatial`, 10