

Package ‘bluebike’

May 5, 2022

Title Blue Bike Comprehensive Data

Version 0.0.3

Description Facilitates the importation of the Boston Blue Bike trip data since 2015. Functions include the computation of trip distances of given trip data. It can also map the location of stations within a given radius and calculate the distance to nearby stations. Data is from <<https://www.bluebikes.com/system-data>>.

License MIT + file LICENSE

Depends R (>= 2.10)

Imports dplyr, janitor, leaflet, lubridate, magrittr, readr, sf, stringr, tidyselect, utils

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

NeedsCompilation no

Author Ziyue Yang [aut, cre] (<<https://orcid.org/0000-0002-9299-8327>>),
Tianshu Zhang [aut] (<<https://orcid.org/0000-0002-3004-4472>>)

Maintainer Ziyue Yang <zyang2k@gmail.com>

Repository CRAN

Date/Publication 2022-05-05 06:00:05 UTC

R topics documented:

bluebike	2
import_month_data	3
station_data	4
station_distance	4

station_radius	5
trip_distance	6
trip_history_sample	6

Index	8
--------------	----------

bluebike	<i>bluebike - A Data Package for Bluebike Users</i>
----------	---

Description

bluebike includes functions and dataset that aids bluebike users to retrieve data and perform data wrangling and visualizations

Details

This package includes data from the Boston Blue Bike trip history data acquired from the **Blue Bikes System Data**. The users can import all monthly trip history data from 2020 to 2022 into a cleaned data set that can easily be used for data analysis. The package also includes a sample data set that includes 1000 sampled trip history from Feb. 2022, and a full data set that contains information about all available stations. The package also serves as a visualization tool for user to browse for closest stations as well as trip-planning via computing trip distances.

Available functions are:

- `import_month_data`
Takes in numeric year/month values and imports data from **Blue Bikes System Data** for the specified time
- `station_distance`
Returns stations with distance in ascending order given the user's current location
- `station_radius`
Plots the position of the stations within walking distance (500 m), and present the basic information about the stations via leaflet
- `trip_distance`
Computes the geographical distance between the start and end stations

Available datasets are:

- `trip_history_sample`
A sample of 1000 trip data entries from February 2022
- `station_data`
A dataset that includes identification, position, and other basic information about bluebike stations

Examples

```
library(dplyr)
# Find most used stations:
stations <- trip_history_sample %>%
  group_by(`start_station_name`) %>%
  summarize(trips_from = n())
head(stations)
```

import_month_data	<i>Import monthly data from bluebike system data</i>
-------------------	--

Description

This function takes in numeric year/month values and imports data for the specified time

Usage

```
import_month_data(year, month)
```

Arguments

year	numeric value of year
month	numeric value of month

Value

A `spec_tbl_df` object

Examples

```
# Pull Jan., 2015 data from web
library(dplyr)
jan_2015 <- import_month_data(2015, 1)

# Pull first quarter of 2015 data from web
spring2015 <- c(1, 2, 3)
quarter_1_2015 <- lapply(spring2015, import_month_data, year = 2015)
quarter_1_2015 <- bind_rows(quarter_1_2015)
```

station_data	<i>Blue bike station data</i>
--------------	-------------------------------

Description

A dataset that includes identification, position, and other basic information about bluebike stations

Usage

```
station_data
```

Format

A data frame of 423 rows and 8 columns

number Station ID

name Station name

latitude Latitude of the station

longitude Longitude of the station

district District of the station

public Character vector showing if a station is public

total_docks The number of docks at each station

deployment_year The year that the station was put into work

Source

The original source of the data are bluebikes system data retrieved from <https://www.bluebikes.com/system-data>

station_distance	<i>Compute the distance from stations given current location</i>
------------------	--

Description

This function returns stations with distance in ascending order given the user's current location

Usage

```
station_distance(long, lat)
```

Arguments

long longitude of user location

lat latitude of user location

Value

a tbl_df object showing the distance between the user and top five closest stations with ID, name, number of docks, and position

Examples

```
# Calculate distance for user at (-71.11467361, 42.34414899) and show the closest five stations
top_5_station <- head(station_distance(-71.11467361, 42.34414899), 5)
```

station_radius	<i>Plot bike stations within a given radius</i>
----------------	---

Description

This function plots the position of the stations within walking distance

Usage

```
station_radius(long, lat, r = 1000)
```

Arguments

long	numeric value of longitude
lat	numeric value of latitude
r	numeric value of set radius in meters

Value

A leaflet map

Examples

```
# Show user at (-71.11467, 42.34415) and set the radius to 500 m
station_radius(long = -71.11467, lat = 42.34415, r = 2000)
```

trip_distance	<i>Compute trip distance for a specific dataset</i>
---------------	---

Description

This function computes the geographical distance between the start and end stations for trips in a given dataset

Usage

```
trip_distance(data)
```

Arguments

data trip data pulled from the Blue Bike System data

Value

a tbl_df object with an additional distance column

Examples

```
# Calculate distance for sample trip data
sample_distance <- trip_distance(trip_history_sample)$distance
```

trip_history_sample	<i>Random 1000 samples from the Blue Bikes System Data website</i>
---------------------	--

Description

a random sample of bluebike trip history data from February, 2022

Usage

```
trip_history_sample
```

Format

A data frame of 1,000 rows representing each sample of trip history

trip_duration Trip duration of each trip measured in seconds

start_time Start time and date of each trip

stop_time Stop time and date of each trip

start_station_id The identification variable of the start station

start_station_name The name of the end station

start_station_latitude The latitude of the start station
start_station_longitude The longitude of the start station
end_station_id The identification variable of the end station
end_station_name The name of the end station
end_station_latitude The latitude of the end station
end_station_longitude The longitude of the start station
bike_id The identification variable of the bike corresponding to each trip
user_type Type of user in each trip (Casual = Single Trip or Day Pass user; Member = Annual or Monthly Member)
postal_code Postal code of the user

Source

The original source of the data are bluebikes system data retrieved from <https://www.bluebikes.com/system-data>

Index

* datasets

station_data, 4

trip_history_sample, 6

bluebike, 2

import_month_data, 3

station_data, 4

station_distance, 4

station_radius, 5

trip_distance, 6

trip_history_sample, 6