




Osmenrich: An R package to enrich geocoded data

Erik-Jan van Kesteren^{1, 2}, Leonardo Jaya Vida^{1, 2}, and Jonathan de Bruin^{1, 2}

1 Utrecht University 2 ODISSEI

DOI: [DOIunavailable](#)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Pending Editor](#) 

Reviewers:

- [@Pending Reviewers](#)

Submitted: N/A

Published: N/A

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

Our package provides a user-friendly way to enrich geographic datasets in R, for example observations on a map, with geographic features around those observations. This package builds on existing infrastructure to interface with OpenStreetMaps (osmdata, Padgham et al. (2017)), and works well with the existing ecosystem of packages for (geospatial) data analysis, namely sf (Pebesma, 2018) and the tidyverse (Hadley Wickham et al., 2019). The osmenrich package streamlines and standardizes the process going from a raw dataset with observations and locations to a tidy (Wickham, 2014), rich dataset with multiple relevant geographical features about those locations. Figure [Figure 1](#) shows graphically the basic workflow of osmenrich.

The R package osmenrich is available on [GitHub](#).

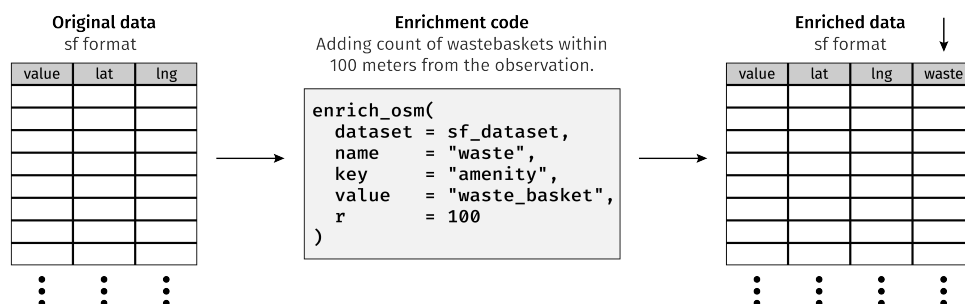


Figure 1: Basic workflow of osmenrich.

Statement of need

Geographic data is valuable in research where the environment influences the process under investigation. For example, the osmenrich package is useful in the analysis of data retrieved from citizen science projects such as [plastic spotter](#) or the [great backyard bird count](#). At the same time, multiple software solutions exist for extracting data from geographic information systems (Padgham et al., 2017) (MORE CIT). However, in order to include these data in further analysis (e.g., universal kriging in gstat, citation), these data often need further processing and, crucially, *aggregation*. Within this problem space, the contributions of osmenrich are as follows:

- Creates a user-friendly interface to OpenStreetMaps, abstracting away the necessary API calls.
- Defines standardized ways to aggregate geographic information based on kernels (see [section](#)).

- Allows distance measures based on routing, such as duration by foot or distance by car (see [section](#)).

In this way, researchers can focus on the research questions themselves, rather than spending time figuring out how to aggregate geographic data: which features at which distance are relevant for the process under investigation?

To enrich data, the `osmenrich` package uses its main function `enrich_osm()`. This function takes `XXX` and `XXX` to output `XXX`.

Kernels

Kernels are a tool to weight and aggregate the features retrieved from OpenStreetMap and play a center role in the data enrichment process used in `osmenrich`. Specifying the kernel variable in `enrich_osm()` allows to choose the kernel function applied on the features retrieved from OpenStreetMap.

In detail, kernels first weight the points retrieved by their distances (or durations) from the reference points and then convert these vectors into single numbers. In this package, this conversion can be done by just specifying a kernel, or by specifying a kernel and the aggregation function used to reduce these weighted vectors of points into single numbers.

The `osmenrich` package provides three different kernels to be used out-of-the-box (uniform, gaussian and parabola). For each of this kernels, and also in case the user created a custom kernel, the aggregation function can be specified by setting the variable `reduce_fun` to any standard R function, such as, but not limited to, `sum`, `mean` or `median`.

Routing

To retrieve feature points and the distances from these points, the `osmenrich` package makes use of a public test instance of OpenStreetMap and OSRM respectively. The OpenStreetMap server is used to retrieve features points and can also be .

The OSRM server server to query the driving distances (distance = "distance_by_car"). Follow the instructions in section `osmenrich` Docker repository to set it up. Otherwise, out-of-the-box this package will support querying only driving distances. If you are interested in querying distances or durations for other means of transportation, you will need to set up local OSRM instances. # Usage

`osmenrich` is available on [GitHub](#) and can be installed and loaded into the R session using:

```
# install.packages("remotes") # To install osmenrich we need 'remotes'
# remotes::install_github("sodascience/osmenrich") # Main version
# remotes::install_github("sodascience/osmenrich@develop") # Development version
library(osmenrich)
```

As a brief example, suppose we have a dataset of common swift's nests provided openly by the city of Utrecht, the Netherlands. We want to enrich this dataset retrieving a proxy of natural material availability. For the sake of this example, we retrieve only "trees," as they represent a close enough proxy to natural materials. However, we could easily retrieve other data types, using the available list on OpenStreetMap at [Map Features](#)

```
# Retrieve points
data_url <- "https://ckan.dataplatform.nl/dataset/8ceaae10-fb90-4ec0-961c-ef02691b"
# Conver points into sf
bird_sf <-
  read_csv(data_url) %>%
  drop_na(latitude, longitude, `aantal nesten`) %>%
```

```
st_as_sf(coords = c("longitude", "latitude"), crs = 4326) %>%
  select(nestcount = "aantal nesten", geometry)

bird_sf <-
  bird_sf %>%
  enrich_osm(
    name = "tree_1km",
    key = "natural",
    value = "tree",
    kernel = "gaussian",
    r = 1000
  )

bird_sf

Change kernel example

Change routing example
```

Acknowledgements

References

- Hadley Wickham et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Padgham, M., Rudis, B., Lovelace, R., & Salmon, M. (2017). Osmdata. *The Journal of Open Source Software*, 2(14). <https://doi.org/10.21105/joss.00305>
- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1), 439–446. <https://doi.org/10.32614/RJ-2018-009>
- Wickham, H. (2014). Tidy data. *Journal of Statistical Software, Articles*, 59(10), 1–23. <https://doi.org/10.18637/jss.v059.i10>