

# R Notebook

## Open access data for transport research: tools, modelling and simulation (OpenInfra)

### Introductory sentence

OpenStreetMap data has the potential to transform how transport infrastructure planning is done, yet currently it does not provide enough data for (accessible) active travel infrastructure planning due to the lack of more refined information about highway attributes.

### Project overview

Dependence on motor modes of travel has been associated with adverse physical, mental, and environmental health (reference 1)(image 1). A shift towards a more sustainable - active - transport, such as walking, wheeling, and cycling, might help to tackle the aforementioned problems (reference 2). However, to encourage the uptake of active modes of travel, it is important to meet the infrastructural needs of citizens. For this, citizen participation in the planning process might be essential. A potential way to foster engagement is to encourage citizens to generate data about their local communities which, consequently, could be used make decisions informed by both data and local contexts. Given this, the project aims to explore how open crowd-sourced data, specifically OpenStreetMap (OSM), can be used to *understand, prioritise, and design* active travel infrastructure<sup>1</sup>.

### Data and methods

OSM data is open, thus free to use regardless of the purpose as long as the source is credited. To query OSM database, an `osmextract` package (reference 3) in R has been used. Three metropolitan counties - West Yorkshire, Greater Manchester, and Merseyside - and an administrative area - Greater London - have been chosen as case studies for the first 6 months of the project with a goal to scale it up to the entire England in months 7-12.

It is important to understand the data structure of OSM in order to utilize it. Essentially, OSM data is composed of three key elements: nodes, ways, and relations. To describe geographic features, *tags*, or key=value pairs, are utilized. For more details on OSM data structure, read an introductory article written on the subject as part of the project.

To explore OSM data, exploratory data analysis (EDA) was applied. EDA was used to understand the potential and limitations of the OSM data. Bar charts have been chosen as a visual tool to demonstrate the presence/absence of data. In particular, bar charts in this report are used to demonstrate the proportions of (some of) the tags relevant to active transport research to all the mapped highways in a given case study<sup>2</sup>. The size of data per area can be seen in Table 1. It is important to note that OSM data is constantly being updated, hence the size of data might fluctuate depending on when it was downloaded. Finally, values in Table 2 have been recategorized to minimize the number of bins and, thus, improve the readability of the bar plots.

---

<sup>1</sup>For a detailed outline of project goals, visit GitHub repository or a website

<sup>2</sup>It should be mentioned that motorways and motorway links have been excluded as neither pedestrians nor cyclists are legally allowed to use them.

Table 1: Number of highways per area (as of 1 April 2022)

Area	Number of highways
Greater London	381590
Greater Manchester	179976
Merseyside	75090
West Yorkshire	183112

The project has a GitHub repository with an aim to promote the reproducibility of project outputs. In order to encourage engagement with OSM data, a website has been developed which currently contains two articles: one introducing OSM data from a theoretical point of view and another one demonstrates how to download and start working with OSM data in R.

## Key findings

The analysis of the chosen areas has shown that `highway=footway` and `highway=cycleway` tags constitute about 1/3 of all the mapped highways in OSM. Greater London stands out as the proportion reaches 1/2 mark (Table 2). This is not surprising given London’s investment in active travel (reference 4). Surprisingly, however, Greater Manchester has a very similar highway proportion to other two metropolitan counties even though it has, in 2018, proposed an ambition to deliver the most comprehensive active travel network in the England (reference 5). Also, it is important to note that the proportions shown in Figure 1 might be underestimated because of the existing alternative method to mapping footpaths and cycleways that relies on a different highway value - path.

Arguably the most important characteristic of OSM is not its ability to represent various types of highways but rather attributes about them. For example, it is possible to refine `highway=cycleway` tag by using `cycleway` to specify if it is a separate or shared cycleway. Similarly, a `bicycle=*` tag can be used to indicate if a highway, such as a footway, can also be cycled on. This variety of tags and their values allow to represent different elements of active travel, such as shared spaces. `Footway=*` tag can be used to indicate the presence of a sidewalk rather than a crossing (coded as “other” in Figure 2). Although these tags provide valuable information about the active travel infrastructure, not each highway is refined. However, when the tags are used, proportionally they tend to be used for cycling rather than pedestrian infrastructure. It could be argued that sometimes the existence of pedestrian infrastructure, such as presence of a sidewalk, can be implied. Hence leading to more sparse use of tags. The lack of sidewalk information has also been reported in Germany (Mobasheri et al., 2017), thus indicating that absence of data is not limited to England only but rather is a limitation in OSM mapping practice.

Limited information on sidewalks and their attributes reveals currently restricted OSM’s potential to be utilized in planning accessible active travel infrastructure. Scarcity of data relevant to accessible planning can be exemplified via other tags, such as `kerb` or `surface`, too (see Figure 3). Nonetheless, in the future it might be possible to take full advantage of the OSM data to support the examination of accessibility as outlined in the Inclusive Mobility guide (2021). In other words, OSM data could be recategorized. This does not mean that all the intricacies of the requirements would be captured but that it could, perhaps when joined with other (open) datasets too, be used to help to determine which streets are inclusive and which need to be prioritized for redesign. Some of this work has been started, namely writing a function in R, that returns recategorised OSM data. For instance, based on the footway (pavements adjacent to the road) and footpath (any other right of way that is not adjacent to road) definitions in the Inclusive Mobility guide, about 1/4 of all the mapped highways could be classified as footpaths; yet only a few percents (with an exception of Greater London) are footways<sup>3</sup>. Moreover, given there is enough data, OSM data could be used to find out not only if surface is paved or not but, more importantly for accessibility, if it is even or not and if the kerb is flush for a wheelchair-friendly crossing.

<sup>3</sup>It should be noted that this is work-in-progress, so results following the recategorization are initial.

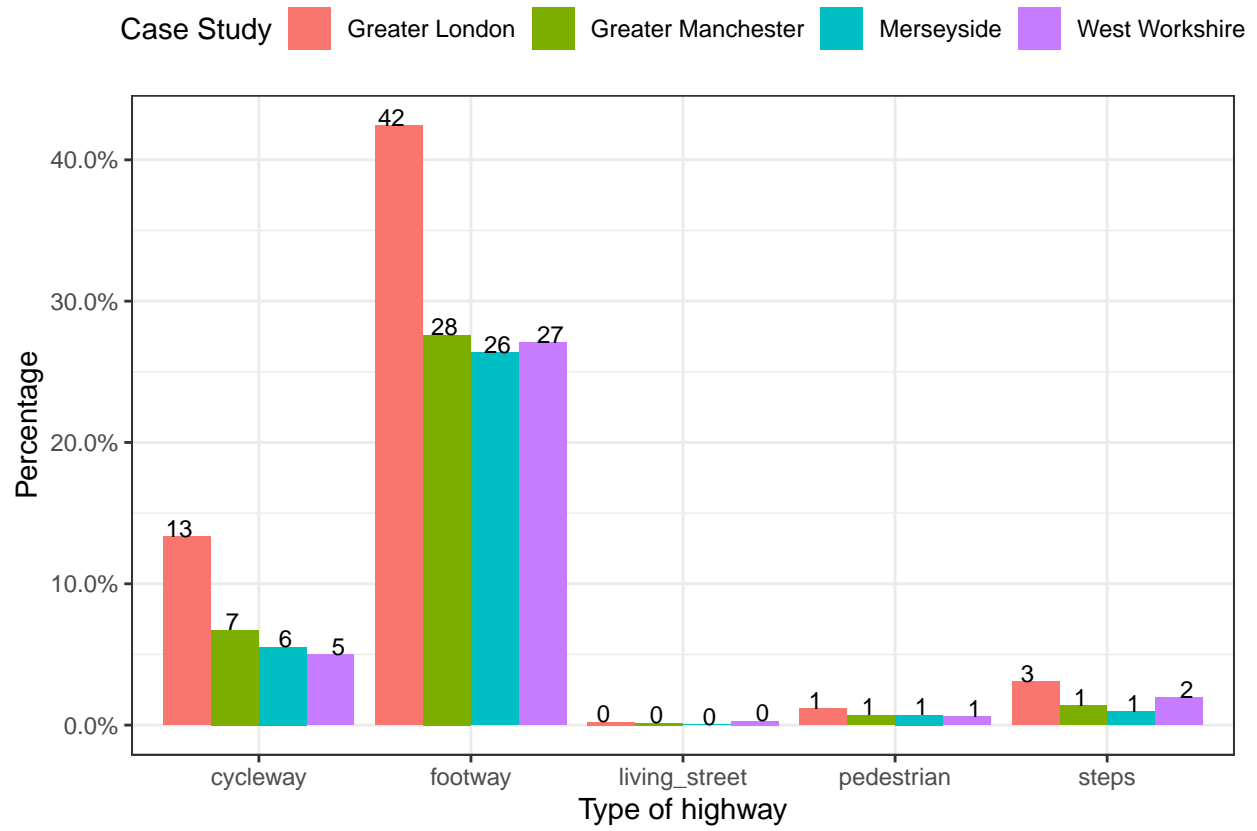


Figure 1: Proportions of different highway values in four case studies

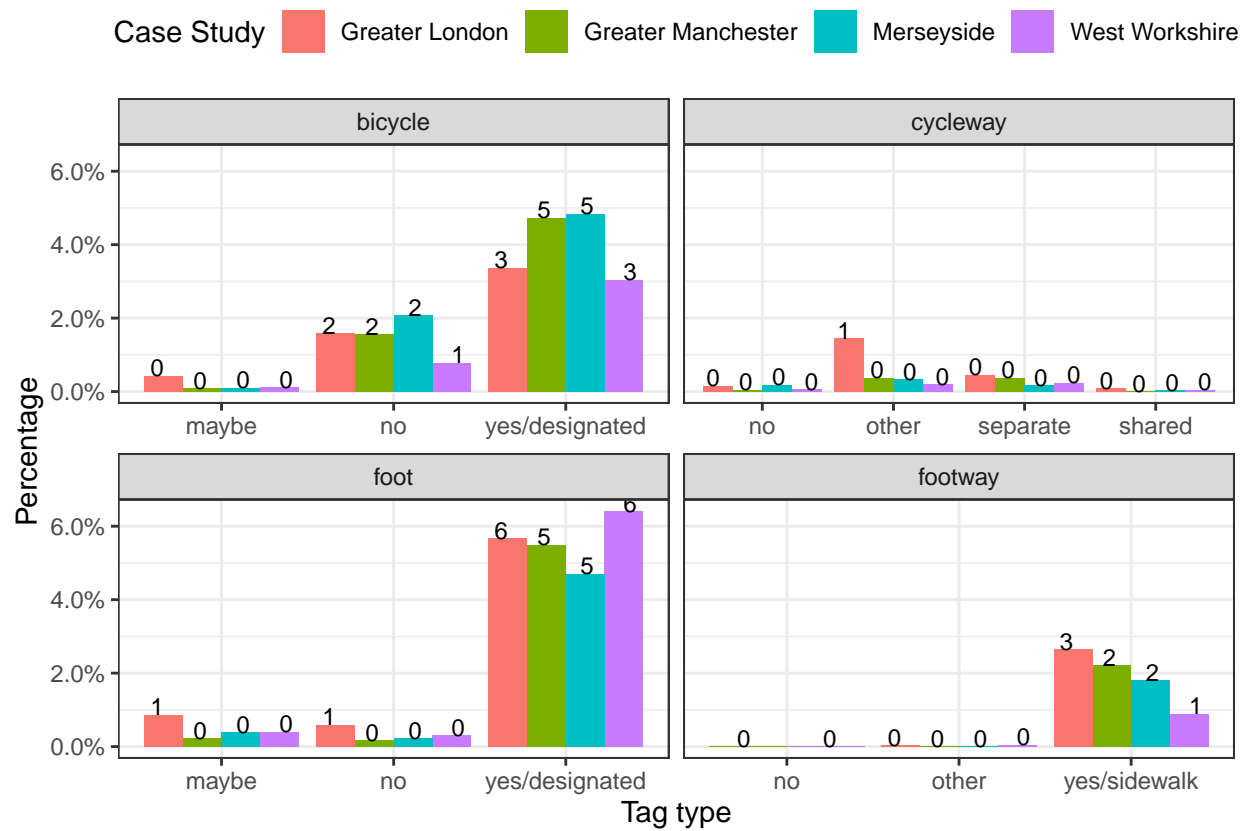


Figure 2: Proportions of (some) tags relevant to active travel infrastructure planning

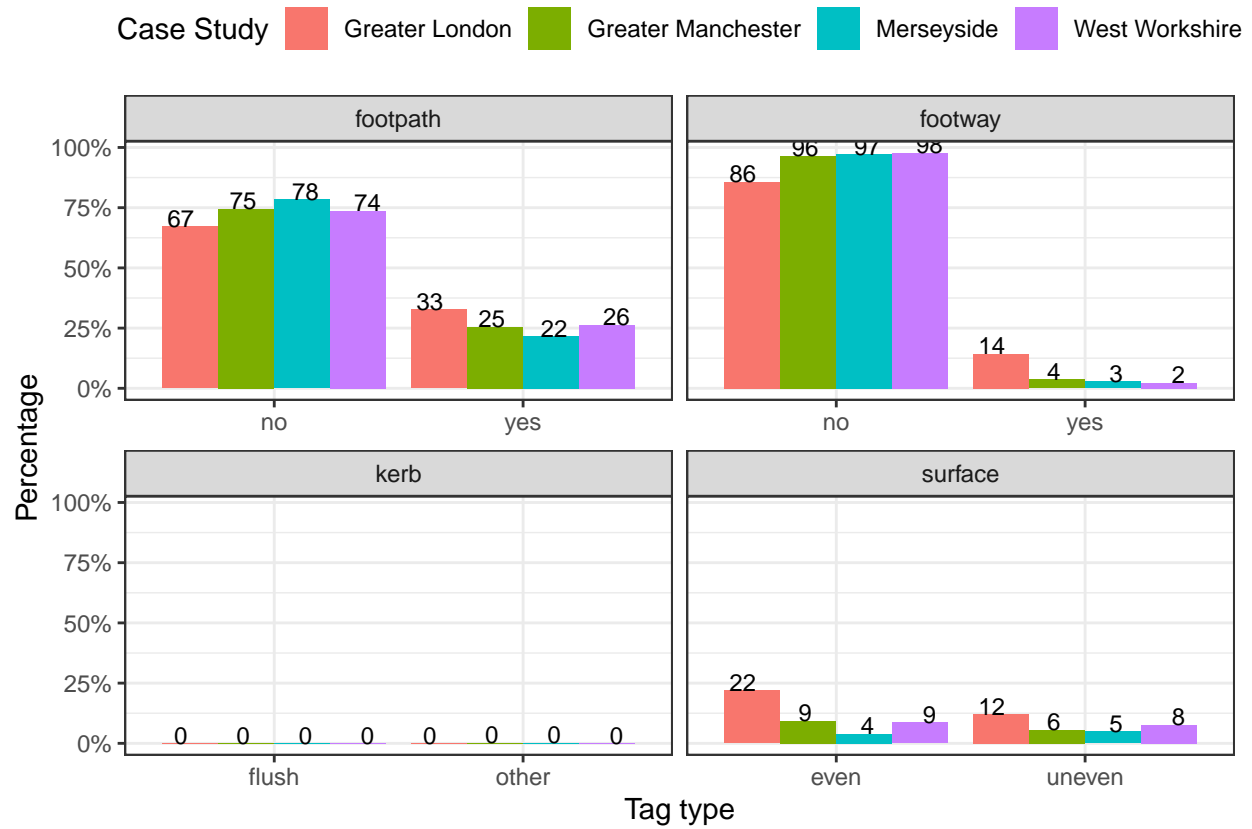


Figure 3: Recategorised OSM tags based on Inclusive Mobility guide

## Value of the research

Value of research: - focus on accessibility - explore the limitations and benefits of OSM in transport research, specifically accessible infrastructure planning - improve the the accessibility of OSM data -> vignettes, functions, etc

## Quote from project partner

### Insights

Advantages of using OSM data in transport research:

- high-level transport network overview;
- provides a (comprehensive?) active travel network;
- tags are flexible enough to represent cycling and walking infrastructures (e.g., cycling lanes and tracks, shared spaces);
- up-to-dateness;
- encourages citizen science and bottom-up approach to planning;

Limitations:

- the more specific highway attribute is, the less data there is;
- tag values lack homogeneity (e.g., width = -1 (!) in WY but also 20cm, 2m, 1-1.5m);
- not straightforward to recategorise;

## Research theme

Urban analytics

## People

Robin Lovelace, University of Leeds

Victoria Houlden, University of Leeds

Ed Manley, University of Leeds

Dustin Carlino, A/B Street

## Funders

This work was supported by the ESRC funded Consumer Data Research Centre (CDRC) under grant references ES/S007164/1 and ES/L011891/1.

## Images + o

Please send a couple of high res images that can be used for the case study (e.g. graph, visualisation).

links:

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/523460/Working\\_Together\\_to\\_Promote\\_Active\\_Travel\\_A\\_briefing\\_for\\_local\\_authorities.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/523460/Working_Together_to_Promote_Active_Travel_A_briefing_for_local_authorities.pdf)

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1033990/net-zero-strategy-beis.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1033990/net-zero-strategy-beis.pdf)

Notes: - for inclusive mobility filter it out so only pedestrian network is left - add a table showing that cycleways tend to be refined more than cycleway (goes with table 2)