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 fails to meet the needs of (accessible) active travel infrastructure planning due to the lack of more refined information on 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 health 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 is essential. A potential way to foster engagement is to encourage citizens to generate data of their local communities which, consequently, could be used make local context and data informed decisions. In this case, open crowd-sourced data might become invaluable. Given this, the project aims to explore how open crowd-sourced data, specifically OpenStreetMap(OSM) data, can be used to *understand*, *prioritise*, and *design* active travel infrastructure.

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 in the first 6 months of the project with a goal to scale it up to the entire England in months 7-12.

Moreover, 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 each of the aspects, read an introductory article written on the subject matter 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 present the proportions of (some of) the tags relevant to active transport research to all the mapped highways in a given case study. 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.

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=cycleways` 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 (still) has a very similar network 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 it 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.

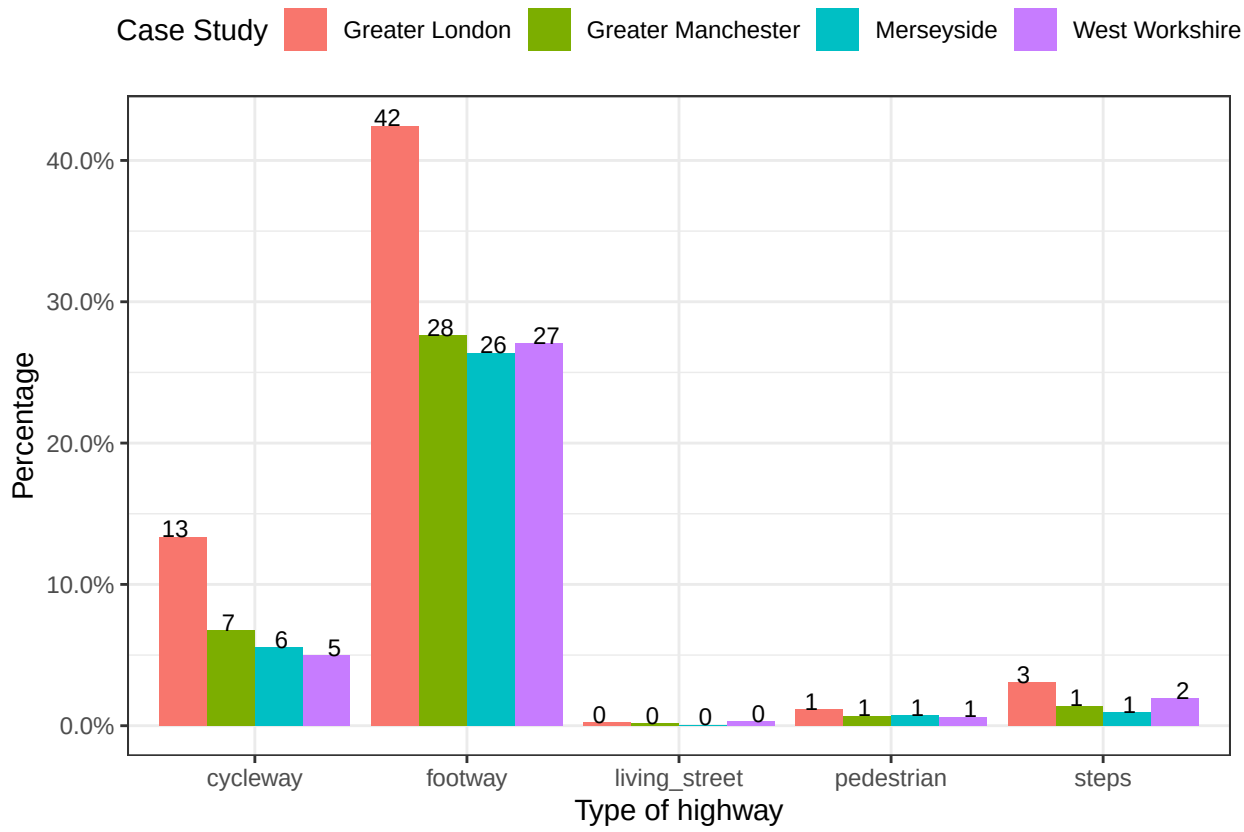

Figure 1: Proportions of different highway values in four case studies

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=footway` tag using additional `footway` tag to indicate that it is a sidewalk rather than a crossing (coded as "other" in Figure 2). Nevertheless, given the high proportion of mapped footways, it is surprising that not many of them are refined to permit a better understanding of a highway network. This echoes Mobasheri et al's (2018 reference 6) observation on the limited availability of sidewalk data in OSM, thus allowing to argue for the persistence of the issue.

Limited information on sidewalks and their attributes demonstrates OSM's currently restricted potential to be utilized in planning accessible active travel infrastructure. Scarcity of data relevant to accessible planning can be exemplified via other tags too. Table 3 shows how information on, for instance, kerbs and width is barely present, yet these street elements are essential for accessible infrastructure (see Inclusive Mobility 2021 reference 7). On the other hand, the flexibility of OSM data holds potential to be used to support the examination of inclusivity as outlined in the Inclusive Mobility guide (2021). In other words, OSM data

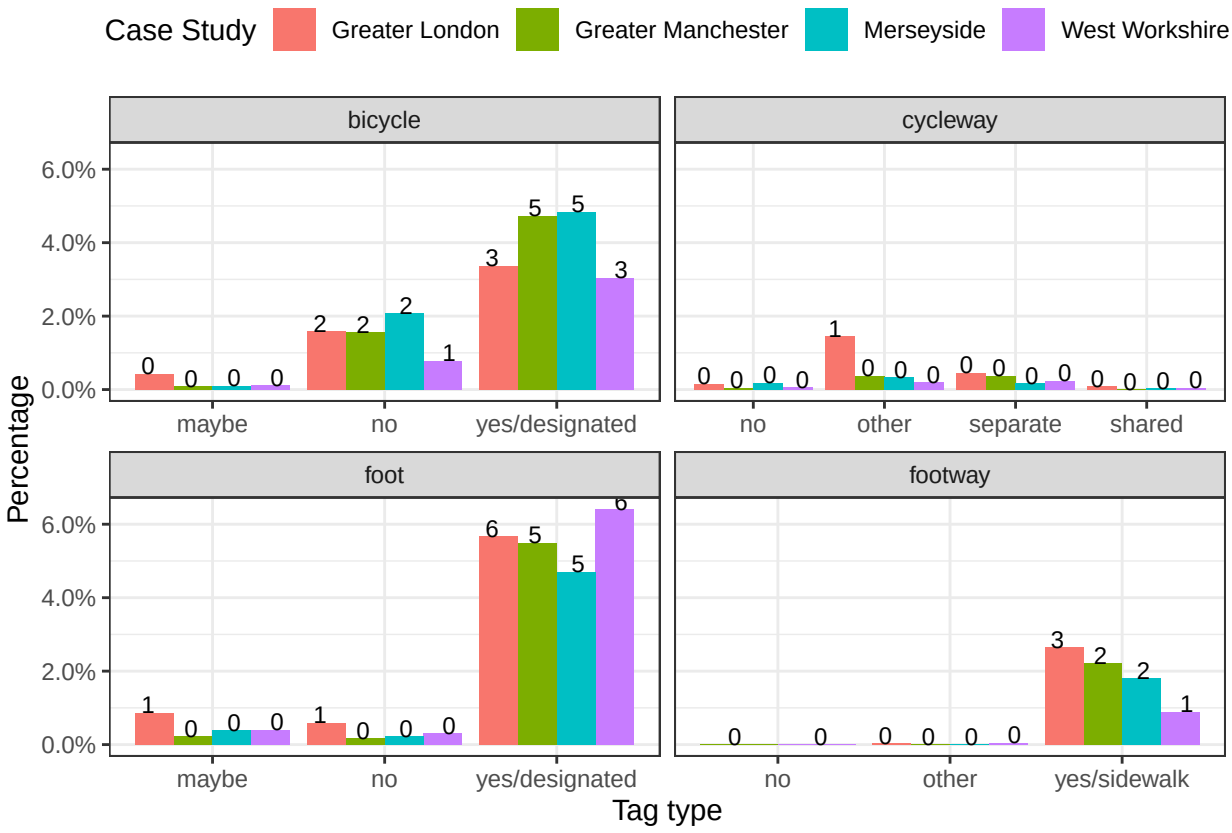

Figure 2: Proportions of (some) tags relevant to active travel infrastructure planning

could be recategorized. This does not mean that all the intricacies of the requirements would be captured but that it could, perhaps 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 started being done, namely writing a function in R (hyperlink to the code), that returns recategorised OSM data.

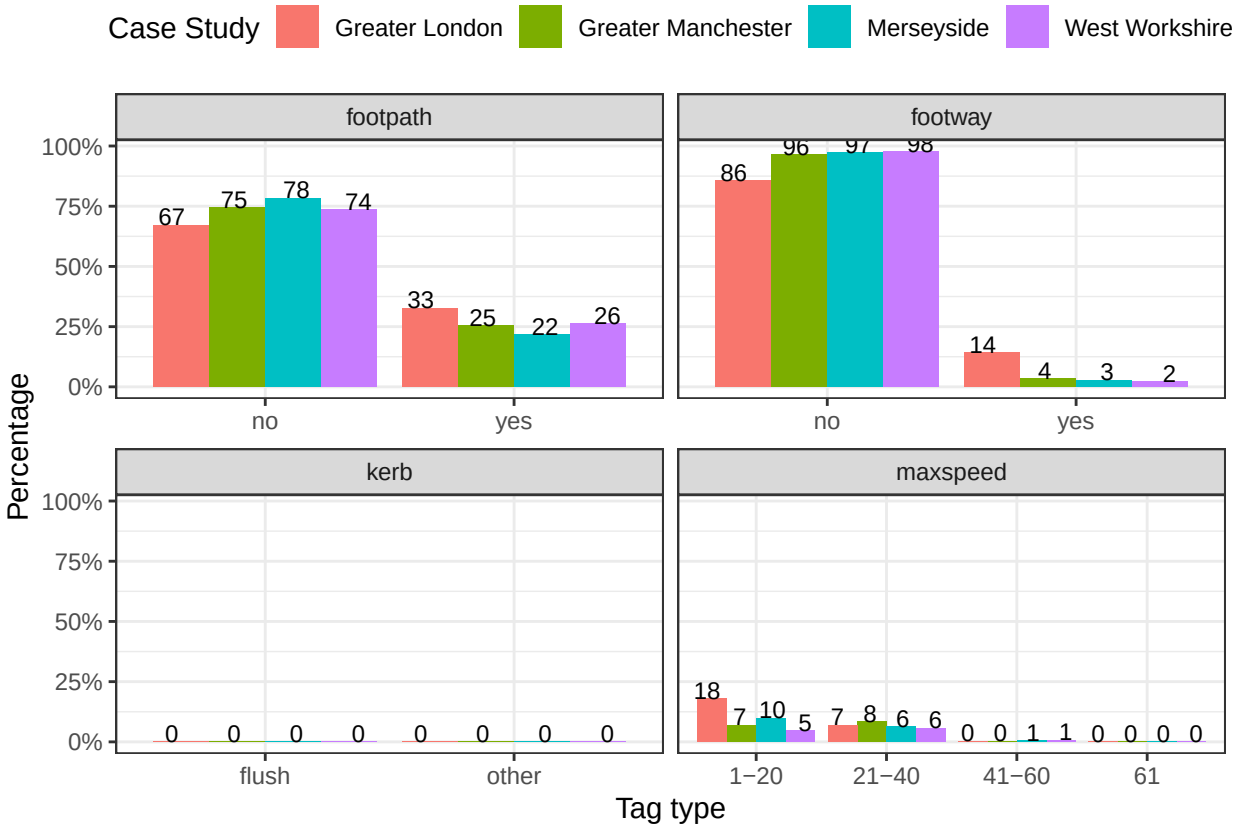

Figure 3: Recategorised OSM tags based on Inclusive Mobility guide

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/1033990/net-zero-strategy-beis.pdf

Notes: - for inclusive mobility filter it out so only pedestrian network is left - greater manchester looks suspicious compared to GISRUK paper