Reproducibility review of: Mobility Vitality: Measuring urban vitality through active and micro-mobility modes

Nina Wiedemann

June 9, 2025

This report is part of the reproducibility review at the AGILE conference. For more information see https://reproducible-agile.github.io/. This document is published on OSF at https://doi.org/10.17605/OSF.IO/gv2z4. To cite the report use

Wiedemann, N. (2025). Reproducibility review of: Mobility Vitality: Measuring urban vitality through active and micro-mobility modes https://doi.org/10.17605/OSF.IO/gv2z4

1. Reviewed paper

Qiang, D. and McKenzie, G.: Mobility Vitality in Active and Micro-Mobility Modes: Measuring Urban Vitality Through Spatiotemporal Similarity, AGILE GIScience Ser., 6, 9, https://doi.org/10.5194/agile-giss-6-9-2025, 2025.

2. Summary

The paper comes with a GitHub repository that includes all code to reproduce the figures in the paper from the raw data. However, part of the data can not be shared. While they are available via an API, the time to collect the data is not feasible within the reproducibility review. Using only the available data, the corresponding figures could be successfully reproduced after minor code fixes done upon request. Overall, the paper is **partly reproducible**.

3. Reproducibility reviewer notes

3.1. Repository structure and installation

The GitHub repository is structured in three subfolders: 1) Data collection, 2) Data cleaning and STA, 3) Other figures except STA. Each folder contains mainly Jupyter notebooks. The authors recommended to execute these notebooks on Colab, and thus no conda or venv requirements file is provided.
There are three datasets:

- Capital bikesharing: This data is publicly available and easy to access (only 12 csv files).
- Lime scooter data: These data were collected from the Lime API in real time. A script is available to crawl the data, but, to obtain *similar* results as in the paper, the script would have to run for a whole year.
- Strava: The historical data is available via an API, and the corresponding observational IDs (10k samples) are provided within the repo. However, due to the rate limit of the API, acquiring the data would take a few days.

Thus, this reproducibility review only comprises the analysis based on the Capital bikesharing dataset.

3.2. Reproducing results for the Capital bikesharing data

The preprocessing of the data is all done in one Jupyter notebook and takes a while. Upon recommendation, the authors added runtime notes to the notebook, implying that one cell can take up to 50h to process on Colab. The time estimates generally aligned with the ones experienced in the review. Some minor typos in the list of files were fixed through the review process.

Further notebooks provide the code to reproduce the figures. All notebooks are commented and structured well.

Figure 1 of the paper shows one histogram per dataset. The histogram for the Capital bike sharing data could be reproduced successfully. Only the "WD" (weekday) and "WE" (weekend) labels were added manually by the authors and are not shown in the plot. However, this does not matter for the interpretability of this figure. Figure 2 shows aggregated volumes over all four datasets, and can thus not be reproduced. Figure 3 simply shows the neighborhoods of the analysis and can be reproduced by downloading the corresponding shape file. Instructions are provided in the Readme. Figure 4 shows the trip volume per dataset. In the process of the reproducibility review, a structured notebook was provided that allows to reproduce all parts of the figure separately. Thus, Figure 4c can be easily reproduced, showing the aggregated volume for the Capital bike sharing data. Figure 5 is the STA analysis by dataset, which is reproducible for the Capital bike sharing data. Only colored boxes were manually added to the figure to highlight specific parts. Figure 6 and 7 again combine all modes and are therefore not reproducible with the available data.

Since the figures for the Capital bike sharing data could be reproduced successfully, and complete code is provided for the other datasets, it is more than likely that all results are reproducible after crawling the necessary data. For the Lime dataset, there would of course be some differences to the figures in the published paper since no historic data is available.