

Comparative Analysis of App-Based Travel Diary and Self-Reported Behaviour: A Case Study from Glasgow, UK

Varun Raturi

Urban Big Data Centre, University of Glasgow, Glasgow, UK

1. Introduction

Traditional methods for travel data collection bear certain limitations. **App-based data collection** emerges as a promising alternative.

Our study investigates the **potential and challenges of app-based data collection** methods in contrast to traditional surveys for transportation policy-making.

2. Data and Methods

Data from an automatic trip detection app, **MyWays**, from TravelAI was utilised.

383 individuals from Glasgow using the **MyWays app** for over a week in 2022 was analysed, who also completed a one-day travel diary within the same app.

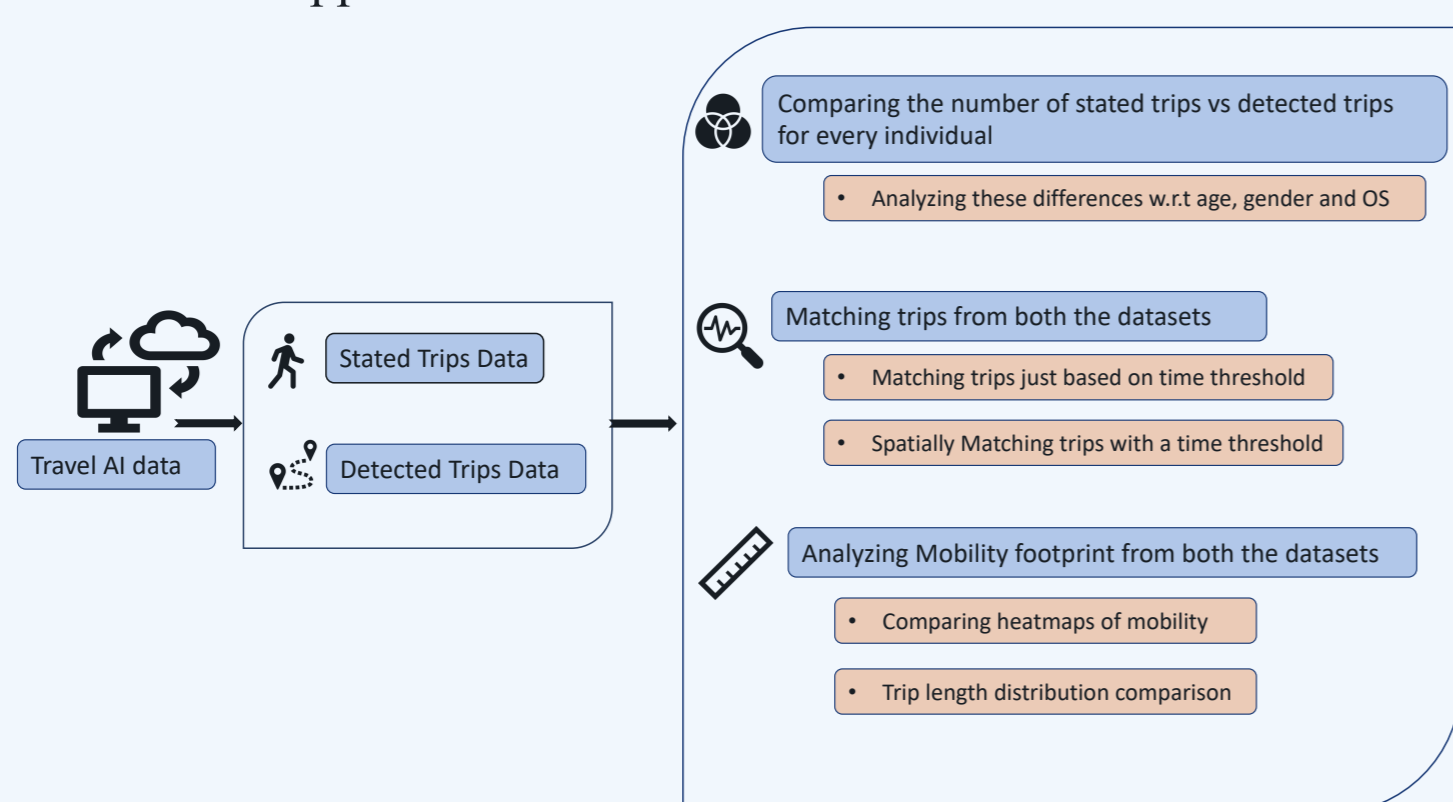


Figure 1: Methodology

3. Results section 1

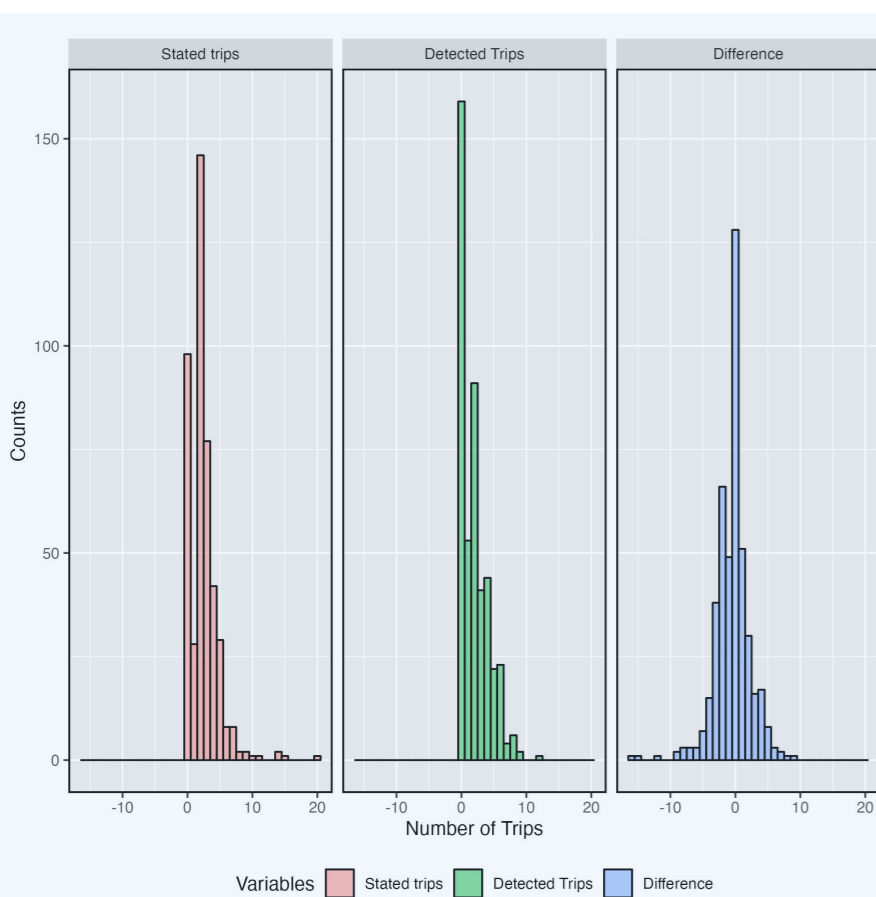


Figure 2: Differences between Detected trips and Stated trips

- Differences (Detected trips – Stated trips) vary from **-16 to 9**.
- This indicates **stated trips and possible dates** of travel stated by individual might be unreliable.
- Detected trips indicate several individuals making **no trips** in a day.

- These differences were more pronounced among **iOS users** compared to **Android users**, attributable to **iOS's stricter privacy controls for apps**.
- For iOS users to grant full access to their location data, **additional steps** in the settings were required.
- This discrepancy was even more evident among **older age groups**, compared to younger ones.

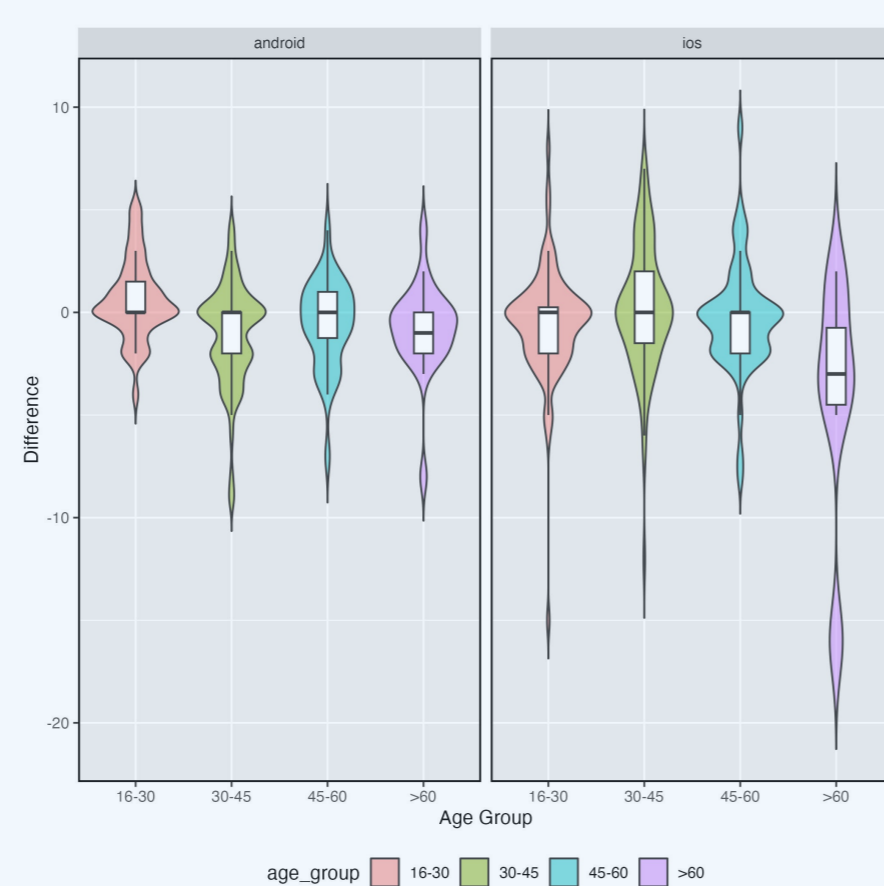
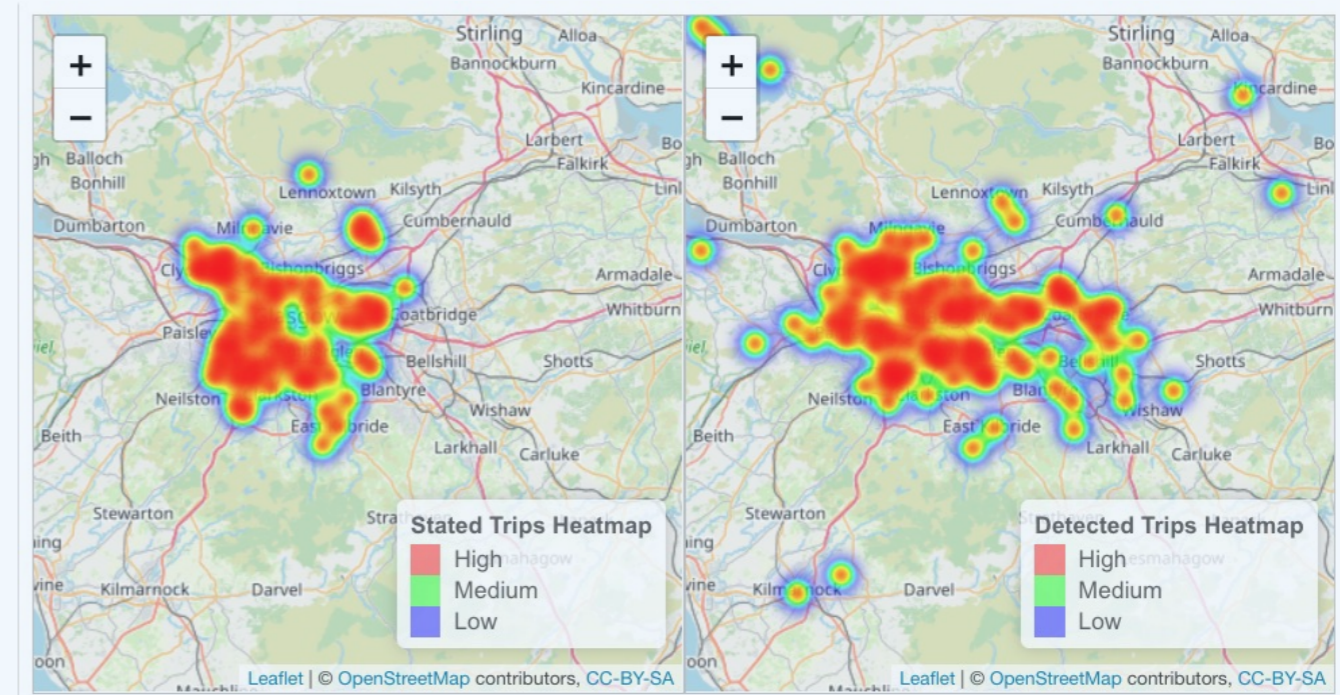


Figure 3: Differences between Detected trips and Stated trips w.r.t age and operating system

4. Results section 2



- The heatmap comparison indicates **locations information outside Glasgow** is missing from the stated trips data.
- This is due to individuals **not knowing the postcodes** outside Glasgow city.

Figure 4: Mobility Heatmaps from stated and detected trips data

- Mean of stated trips length is **smaller** than the mean of detected trip length.
- This is due to the **missing postcodes** data in the stated trips especially for longer trips.
- KS test indicates a **significant difference** between the two distributions.

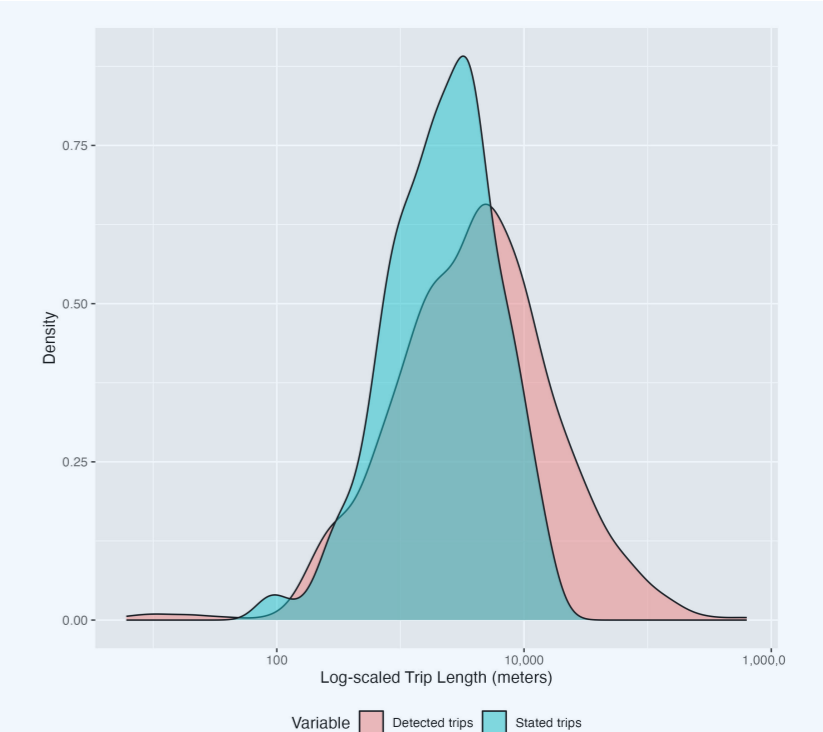


Figure 5: Trip length distribution of stated and detected trips

Table 1: Matching trips based on time threshold

Threshold (in hours)	Mode detection Accuracy	No. of matched trips
0.5	78.98	352
1	77.79	635
1.5	76.93	841
2	76.48	969

Table 2: Matching trips spatially and based on time threshold

Threshold (in hours)	Mode detection Accuracy	No. of matched trips
0.5	81.88	287
1.0	83.13	332
1.5	82.96	358
2.0	82.33	368

- When matching the trips based on a **time threshold**, 969 trips out of a total 1485 stated trips were matched.
- When matching the trips **spatially**, 368 trips out of a total 497 stated trips (where postcodes were given) for a two-hour threshold were matched.
- Mode detection accuracy** of the app increases when we accurately match the trips.

5. Conclusions

The app-based data **provided more detailed insights**, recording separate legs of each trip, unlike the traditional self-reported data.

A potential limitation of app-based methods is the **access to location data**, emphasizing the importance of **careful installation** with all necessary permissions granted by the users.

6. References

Hesjevoll et. al (2021); App-based automatic collection of travel behaviour: A field study comparison with self-reported behaviour, Transportation Research Interdisciplinary Perspectives.