Converting Cellular GPS Data into Trips Using R

Gillian Riches[a,*], Gregory Macfarlane[a,**]

[a] *Civil and Environmental Engineering Department, 232 Engineering Building, Provo, Utah 84602*

Abstract

This is where the abstract should go.

Keywords: GPS Data, Trips, Clusters

1. Question

Global Positioning System (GPS) surveys have become a more accurate and reputable alternative to previous travel survey methods that collect activity-travel patterns. Despite GPS devices ability to record time and positional characteristics, they still require processing in order to convert the positional characteristics into trip purposes and activities.

The first step of this conversion process is cleaning the GPS data to produce trips. Currently, most researchers use subjective time and speed rule-based algorithms to perform this task (Shen and Stopher, 2014). Due to their ambiguity, using these rules is not ideal. For example, some people walk slower than others, so the speed threshold would require constant manual changing. Another issue with these rules is that every researcher must have their own definition of a trip. One researcher who considers picking somebody up to be its own activity will have a significantly smaller time threshold. Due to GPS data imputation being applied in these different contexts, accuracy ranges from 43% to 61% (Shen and Stopher, 2014).

The newest and second most common method is cluster-based: the density of GPS points within a predefined radius determines an activity. The radius and point density would not vary from person to person thus providing increased efficiency and precision. In fact, one experiment using a DBSCAN cluster-based algorithms proved to be 92% precise (Luo et al., 2017). Despite this impressive precision, three main gaps still remain: survey collection typically doesn't exceed two weeks (Feng and Timmermans, 2016), not all activities are accounted for in analysis, and this algorithm has not been published in R. Usually, researchers group all of the *Other* trip purposes into one category and analyze it as a whole (Elevelt et al., 2021).

Therefore, the question I am answering is: *How does one write a cluster-based algorithm in R that accurately transforms 6+ months worth of GPS survey data into trips and analyze the "Other" trip purposes as separate activities?* The respondents' GPS data used in my code is associated with their responses to mental health surveys, so they are not publicly available. However, their contents will be generally described in the Methods section when I perform the cluster algorithm.

2. Methods

In this chapter, you describe the approach you have taken on the problem. This usually involves a discussion about both the data you used and the models you applied.

[*] Corresponding Author

[**] Present affiliation: Committee Chair

Email addresses: `martingillian4@gmail.com` (Gillian Riches), `gregmacfarlane@byu.edu` (Gregory Macfarlane)

Table 1: Descriptive Statistics of Dataset

		regcar (N=10930)		sportuv (N=1048)		sportcar (N=880)		stwagon (N=4446)		truck (N=5628	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. De
price		4.2	1.9	4.7	1.9	4.8	2.2	4.1	1.9	4.2	2.
range		237.2	94.5	241.6	94.7	233.6	96.7	238.7	94.3	238.2	93.
size		2.4	0.8	2.1	1.0	1.4	1.0	2.3	0.8	2.4	0.
		N	Pct.	N	Pct.	N	Pct.	N	Pct.	N	Pc
fuel	gasoline	2704	24.7	280	26.7	218	24.8	1096	24.7	1413	25.
	methanol	2729	25.0	246	23.5	225	25.6	1091	24.5	1445	25.
	cng	2767	25.3	260	24.8	238	27.0	1109	24.9	1360	24.
	electric	2730	25.0	262	25.0	199	22.6	1150	25.9	1410	25.

2.1. Data

Discuss where you got your data, how you cleaned it, any assumptions you made.

Often there will be a table describing summary statistics of your dataset. Table 1 shows a nice table using the `datasummary` functions in the `modelsummary` package.

2.2. Models

If your work is mostly a new model, you probably will have introduced some details in the literature review. But this is where you describe the mathematical construction of your model, the variables it uses, and other things. Some methods are so common (linear regression) that it is unnecessary to explore them in detail. But others will need to be described, often with mathematics. For example, the probability of a multinomial logit model is

$$P_i(X_{in}) = \frac{e^{X_{in}\beta_i}}{\sum_{j \in J} e^{X_{jn}\beta_j}} \tag{1}$$

Use LaTeX mathematics. You'll want to number display equations so that you can refer to them later in the manuscript. Other simpler math can be described inline, like saying that $i, j \in J$. Details on using equations in bookdown are available here.

3. Findings

This section might be called "Results" instead of "Applications," depending on what it is that you are working on. But you'll probably say something like "The initial model estimation results are given in Table **??**." That table is created with the `modelsummary()` package and function.

With those results presented, you can go into a discussion of what they mean. first, discuss the actual results that are shown in the table, and then any interesting or unintuitive observations.

3.1. Additional Analysis

Usually, it is good to use your model for something.

- Hypothetical policy analysis
- Statistical validation effort
- Equity or impact analysis

If the analysis is substantial, it might become its own top-level section.

Acknowledgements

This is where you will put your acknowledgments

References

Elevelt, A., Bernasco, W., Lugtig, P., Ruiter, S., and Toepoel, V. (2021). Where you at? using gps locations in an electronic time use diary study to derive functional locations. *Social Science Computer Review*, 39(4):509–526.

Feng, T. and Timmermans, H. J. (2016). Comparison of advanced imputation algorithms for detection of transportation mode and activity episode using gps data. *Transportation Planning and Technology*, 39(2):180–194.

Luo, T., Zheng, X., Xu, G., Fu, K., and Ren, W. (2017). An improved dbscan algorithm to detect stops in individual trajectories. *ISPRS International Journal of Geo-Information*, 6(3).

Shen, L. and Stopher, P. R. (2014). Review of gps travel survey and gps data-processing methods. *Transport Reviews*, 34(3):316–334.