Analysis of smart card data to understand the mobility patterns of concessionary bus users

Ffion Carney[*1], Paul Longley[†1] and Jens Kandt[‡2]

[1]Department of Geography, University College London
[2]The Bartlett Centre for Advanced Spatial Analysis, University College London

January 24, 2019

Summary

Smart card data, produced by automated fare collection (AFC) systems on public transport networks, can provide insights into the daily mobility of cardholders. Increasingly transport authorities are utilising these data to gain insights into travel behaviour, travel demand, and the efficiency of transport services. This research focuses on grouping concessionary bus users based on their travel behaviours. Subsequent analyses investigate differences in each group's demographic characteristics, and spatial and temporal mobility patterns. These analyses were conducted with the aim of detecting a relationship with bus activity levels and evidencing that the mobility patterns of concessionary bus users are not homogenous.

KEYWORDS: Big Data, Smart Card Data, Mobility Patterns, Travel Behaviour

1. Introduction

Different age groups have particular challenges when accessing transport services and maintaining high levels of mobility, and future transport policies should be sensitive to these differences. The older population are likely to be particularly vulnerable to the effects of a lack of access to transport services, as this group are often less mobile than the working-age population due to declining physical abilities and health (Spinney et al. 2009). Understanding the travel behaviours and mobility needs of the older population is therefore a significant challenge for transport planners and policy makers.

Big Data have the potential to offer insights into many complex issues with accuracy and timeliness. One of the sources of Big Data utilised in transport studies is smart card data, produced by automated fare collection (AFC) systems on public transport networks. These systems collect payment from smart travel cards, recording all journeys that are made on the transport network. Although the main purpose of these AFC systems is to manage revenue collection, they also produce large quantities of data on a daily basis, providing a valuable resource for studying the travel habits of transport users (Mohamed et al. 2014). Increasingly, transport authorities are making use of these smart technologies and the resulting data to inform decision making and policy planning. This study exploits these novel data to investigate differences in the travel behaviours of concessionary bus users in the West Midlands, and their subsequent spatial, temporal and demographic profiles.

[*] Ffion.Carney.16@ucl.ac.uk
[†] P.Longley@ucl.ac.uk
[‡] J.Kandt@ucl.ac.uk

2. Study Area and Data

2.1 Case Study

The West Midlands takes part in the English National Concessionary Travel Scheme (ENCTS), which was introduced by the Transport Act (2000) and allows for free bus travel outside of peak hours for those that have reached pensionable age or are disabled (Mackett 2014). This study utilises data collected from bus passengers using 'Over 60 Concession' smart card types. Currently, over 90% of those eligible for concessionary travel are registered cardholders, however bus usage can vary significantly between users.

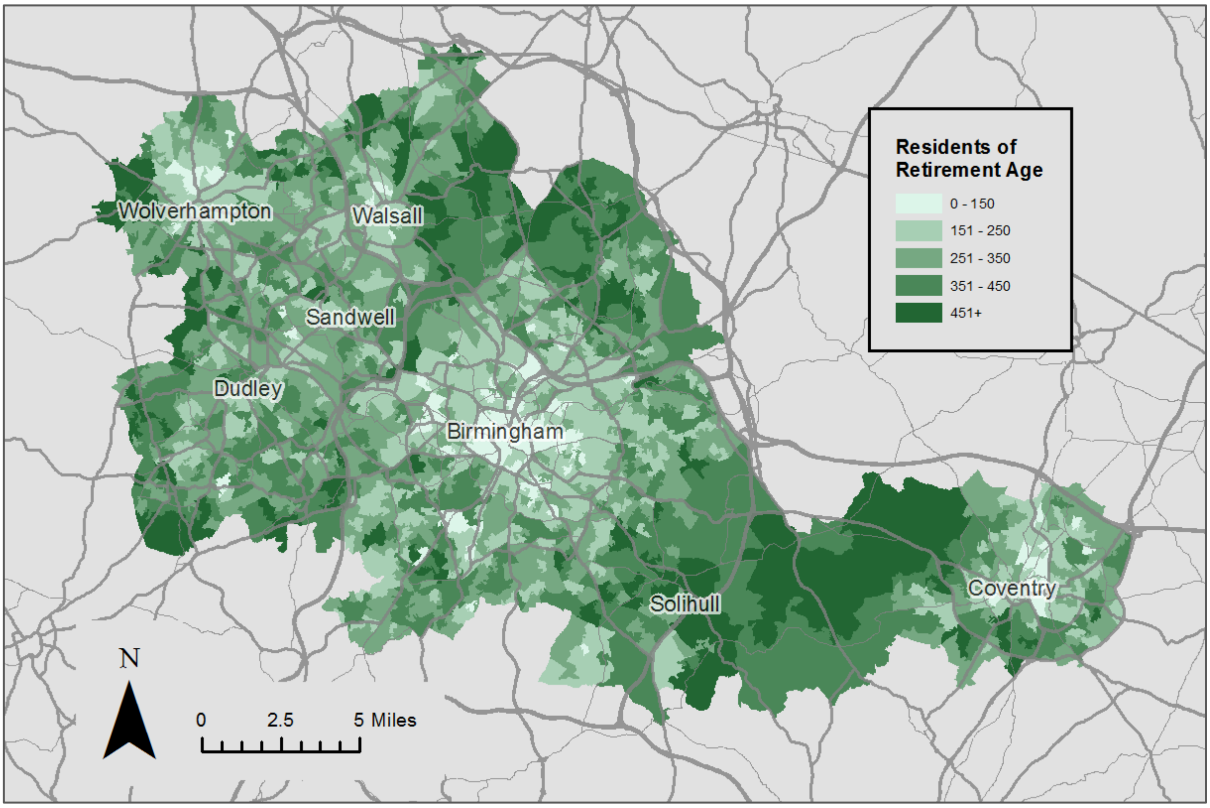

Figure 1 Study area map showing the population eligible for 60+ concessionary bus travel in the West Midlands (ONS 2015)

2.2 Data Description

The data used in this study are a weekly subset of smart card data; ranging from 18/01/2016 – 24/01/2016, and consisting of 247,725 transactions made by 72,351 cardholders. The smart card dataset includes a unique record ID for each transaction, a unique account ID for each cardholder, the time and date of each boarding and alighting, and the bus stop code for the origin and modelled destination of each trip.

3. Method

This study utilised a k-means clustering approach with the aim of grouping concessionary cardholders that exhibit similar patterns of bus usage. A number of analyses were conducted to generate bus usage variables, the methods of which are shown in Table 1.

Table 1 Mobility variables methods

Variable	Method
Trip Distance	- Calculated as the road network distance between origin and destination stops - Average trip distance was then calculated
Stops Visited	- Calculated as the total number of distinct destinations visited by each cardholder
Frequency	- Calculated as the total number of trips made by each cardholder over the study period
Average Number of Days Between Transactions	- Total number of days between first and last transaction calculated - Result then divided by the frequency of transactions

The results of these analyses were then scaled using the 'scale' function in R, to standardise the data so that the variables are comparable.

4. Results

4.1 Cardholder Classification

The k-means clustering resulted in three clusters with distinct bus activity patterns. The results of which are shown in Figure 2 and a summary of each of the clusters is presented in Table 2.

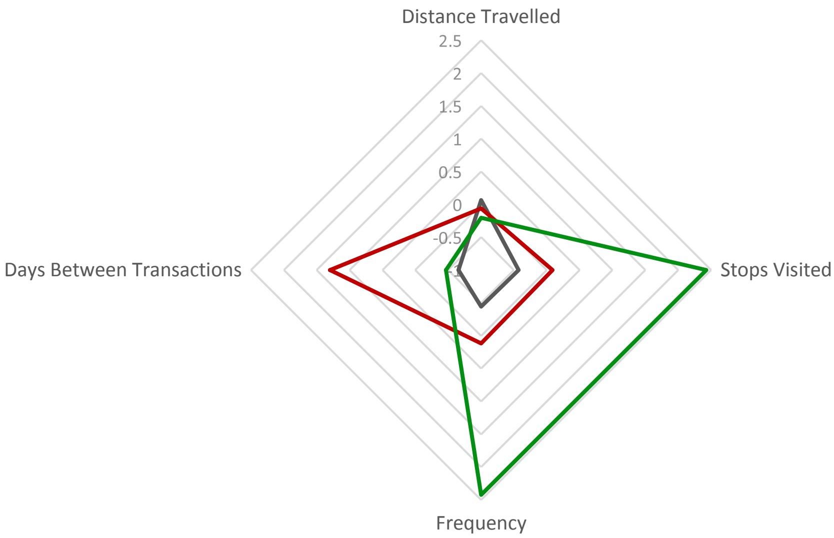

Figure 2 Radial plot showing cluster centres of Cluster 1 (grey), Cluster 2 (red) and Cluster 3 (green)

Table 2 Cluster descriptions

Cluster	Membership	Summary (averages)	Bus Activity
Cluster 1	42,099	Distance (km): 3.9 Distinct destinations: 2.4 Frequency: 2.7 Days between transactions: NA	Low
Cluster 2	23,678	Distance (km): 3.7 Distinct: 3.5 Frequency: 4.7 Days between transactions: 1.1	Medium
Cluster 3	6,574	Distance travelled (km): 3.3 Distinct destinations: 10.6 Frequency: 15.5 Days between transactions: 0.1	High

Cluster 1 exhibits the lowest average frequency of transactions and destinations visited, suggesting the lowest bus activity. Days between transactions is recorded as 'NA' as a large number of cardholders were only active on one day throughout the study week. The average distance travelled is also highest in this cluster suggesting this group undertake infrequent, long-distance bus journeys. Cluster 2 was classified as having medium activity. This cluster exhibited relatively high average frequency of use, bus stops visited and days between transactions, suggesting irregular bus use with high activity on only select days throughout the week. In contrast, Cluster 3 shows high activity throughout the study period, with high average frequency and bus stops visited. In addition, the average days between transactions and distance travelled is low, suggesting frequent, short-distance trips.

4.2 Spatial and Temporal Patterns

Comparing the spatial and temporal mobility patterns of each clusters (Figures 3 to 8) demonstrates how travel behaviours differ between active and non-active users. The spatial profiles show the most frequently visited stops by each cluster, with the size of each point representative of the frequency of alightings. The temporal profiles show the total number of hourly boardings for each cluster.

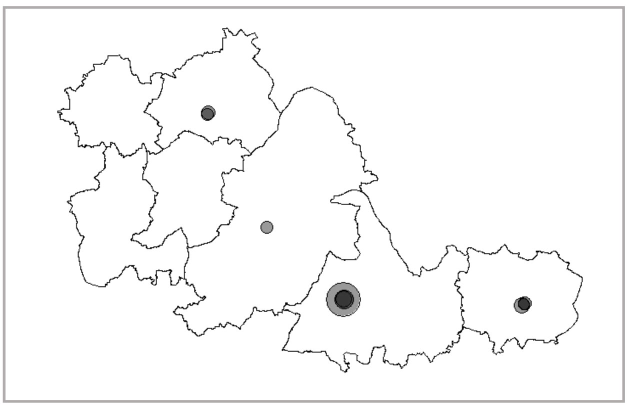

Figure 3 Cluster 1 spatial profile

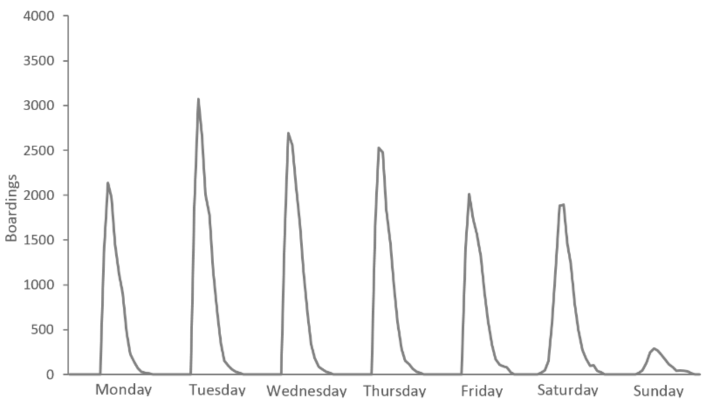

Figure 4 Cluster 1 temporal profile

Figure 5 Cluster 2 spatial profile

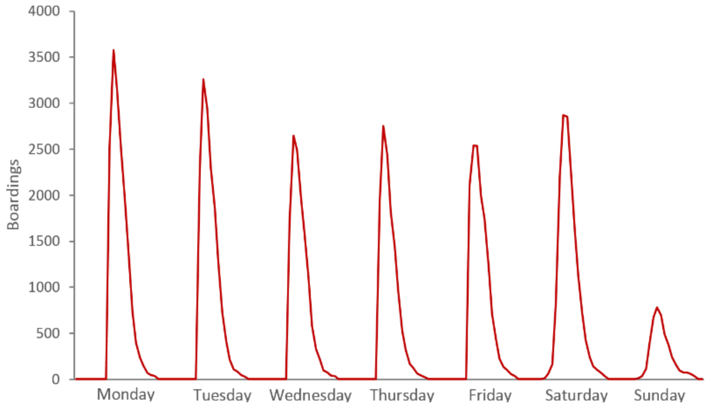

Figure 6 Cluster 2 temporal profile

Figure 7 Cluster 3 spatial profile

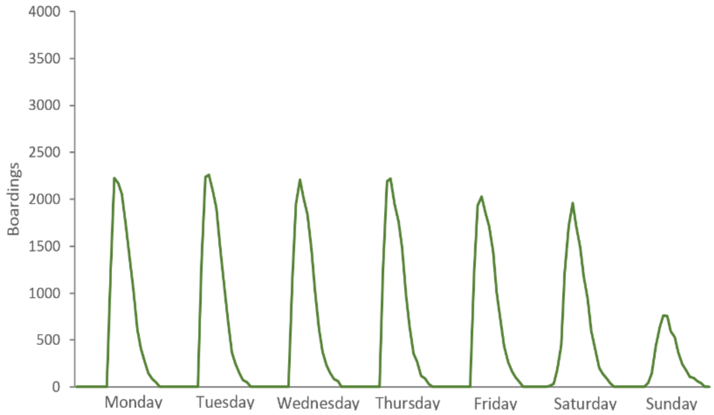

Figure 8 Cluster 3 temporal profile

Figures 3, 5 and 7 show distinct spatial patterns. Clusters 1 and 2 show high levels of activity in the Solihull area, with Cluster 2 also exhibiting high activity in central Coventry. In contrast to this, Cluster 3 shows a distinctive spatial profile different from Clusters 1 and 2, with high bus activity levels in central Birmingham. These insights show how spatial patterns of mobility can be detected through behavioural travel patterns.

The weekly temporal profiles for each cluster, shown in Figures 4, 6 and 8, show less distinct differences, with all days except Sundays exhibiting similar boarding numbers within clusters. This uniformity is likely a result of the focus of this study on the older population. As this study only utilises data produced by ‘Over 60 Concession’ smart cards, a high proportion of the population are likely to be retired and are therefore less likely to demonstrate morning and evening weekday peaks often observed in the temporal profiles of the commuting populations (Ma et al. 2017).

4.3 Demographics

The smart card metadata includes the registered LSOAs of cardholders and can therefore be used to understand differences in the socioeconomic and demographic make-up of each cluster. Figure 9 shows the proportion of the eligible population assigned to each cluster, with Figure 10 showing the proportion of each cluster assigned to each Index of Multiple Deprivation (IMD) decile.

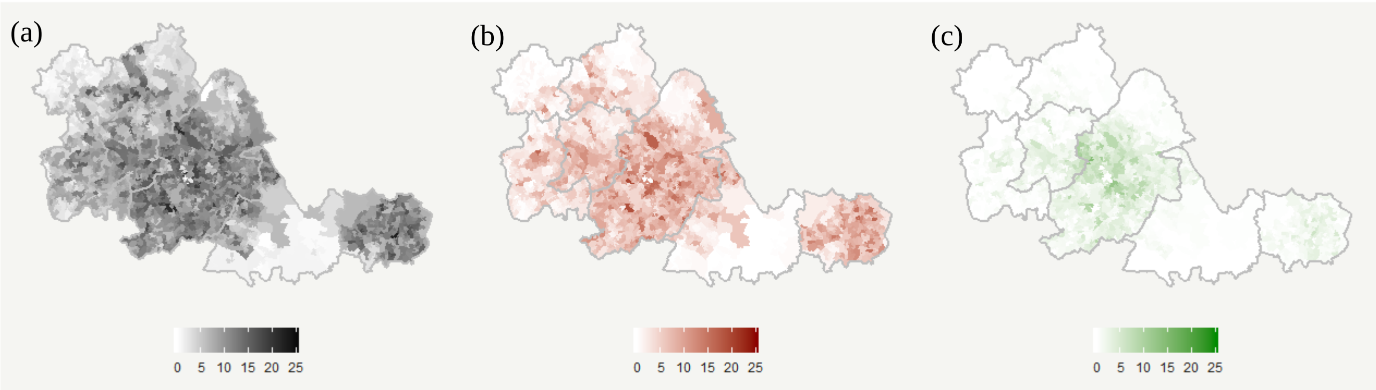

Figure 9 Proportion of eligible population allocated to Cluster 1 (a), Cluster 2 (b) and Cluster 3 (c) (LSOA)

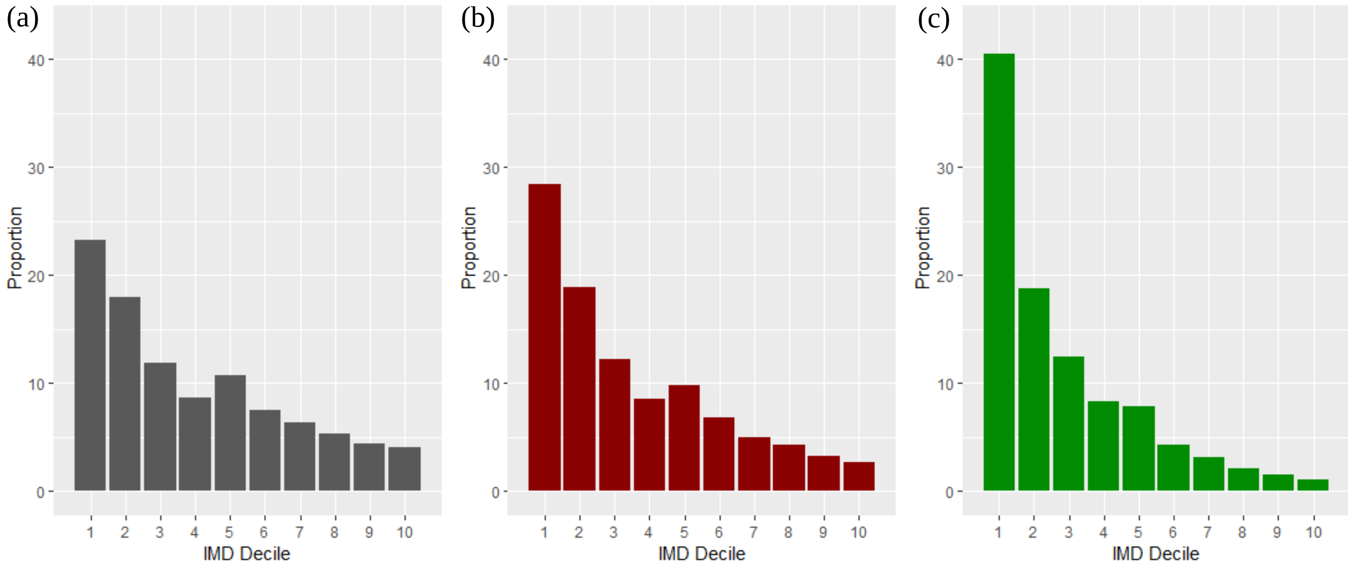

Figure 10 Proportion of population in Cluster 1 (a), Cluster 2 (b) and Cluster 3 (c) assigned to each IMD decile (1 = most deprived, 10 = least deprived)

A significant difference in the deprivation of each cluster can be seen in Figure 10. Cluster 3 contains a large proportion of cardholders residing in areas of IMD Decile 1, characterised as being the most deprived. In contrast, proportions are significantly lower in Clusters 1 and 2. This suggests a relationship between high bus usage and deprivation; likely a reflection of the socioeconomic make-up of the cluster populations. To explore this further, demographic analyses using 2011 Census data were conducted. These analyses concluded that in comparison to Cluster 3, cardholders in Cluster 1 were more likely to have access to a car and to be from a higher socioeconomic background, suggesting that demographics may have a significant influence on mobility and travel behaviour.

5. Conclusions and Future Research

This study forms one area of research being undertaken into how smart card data can be utilised to better understand mobility patterns. Through the clustering of cardholders, distinct patterns of mobility were able to be detected. The differences in the travel behaviours and demographics of each cluster supports the theory that the mobility patterns of the older population are not homogenous, and in order for transport authorities and policy makers to meet the older population's future mobility needs, they should not be treated as such.

Future work will build on the insights of this study, substantiating the demographic analyses and implementing similar methods over a longer study period to detect changes in travel behaviour and bus activity over the past several years.

6. Acknowledgements

This research is funded by the Economic and Social Research Council (ESRC). Data was provided by Transport for West Midlands (TfWM) through the Consumer Data Research Centre (CDRC), based at the Department of Geography, UCL.

References

Ma X, Liu C, Wen H, Wang Y, Wu Y. (2017). *Understanding commuting patterns using transit smart card data.* Journal of Transport Geography, 58, 135-145.

Mackett R. (2014). *Has the policy of concessionary bus travel for older people in Britain been successful?* Case Studies on Transport Policy, 2(2), 81-88.

Mohamed K, Come E, Baro J, Oukhellou L. (2014). *Understanding passenger patterns in public transit through smartcard and socioeconomic data*. UrbComp' 2014. New York City, NY, USA.

Spinney J, Scott D, Newbold B. (2009). *Transport mobility benefits and quality of life: A time-use perspective of elderly Canadians*. Transport Policy, 16(1), 1-11.

Biographies

Ffion Carney is a third year PhD student at the Consumer Data Research Centre, UCL. Her research aims to understand the uses of Big Data for gaining insights into the urban mobility and activity patterns of the older population.

Paul Longley is Professor of Geographic Information Science at UCL, where he directs the Consumer Data Research Centre. His research interests are grouped around socioeconomic applications of Geographic Information Science and Systems.

Jens Kandt is a lecturer in Urban Geography and Data Science at the Bartlett Centre for Advanced Spatial Analysis, UCL. His research interests focus on linking social theory and quantitative geography to understand how social and urban processes shape inequalities in health and well-being.