A visual analytics approach to understanding cycling behaviour

Roger Beecham*
City University London

Jo Wood, *Member, IEEE* †
City University London

Audrey Bowerman‡
Transport for London

ABSTRACT

Existing research into cycling behaviours has either relied on detailed ethnographic studies or larger public attitude surveys [1] [9]. Instead, following recent contributions from information visualization [13] and data mining [5] [7], this design study uses visual analytics techniques to identify, describe and explain cycling behaviours within a large and attribute rich transactional dataset. Using data from London's bike share scheme[1], customer level classifications will be created, which consider the regularity of scheme use, journey length and travel times. Monitoring customer usage over time, user classifications will attend to the dynamics of cycling behaviour, asking substantive questions about how behaviours change under varying conditions. The 3-year PhD project will contribute to academic and strategic discussions around sustainable travel policy. A programme of research is outlined, along with an early visual analytics prototype for rapidly querying customer journeys.

1 INTRODUCTION

Despite recent investment into sustainable travel, there is little evidence that levels of cycling in the UK are increasing [9]. Results from the National Travel Survey show that cycling accounted for 2% of all trips of 5 miles or less made in Great Britain in 2010, and that the average number of trips per person made by bike has generally fallen over the last 15 years [4]. A tranche of research aimed at better understanding cycling behaviours has subsequently emerged within the social sciences [9]. Much of this work has been based on empirically collected datasets - attitudinal surveys and deeper ethnographic studies - and has identified distinct but nuanced attitudes towards cycling [9] [1]. Studies focussing on actual rather than self-reported behaviour have generally relied on GPS logs [3] or automated traffic counts at fixed sites [6]. Due to their complexity, the former have been relatively small in scale and the latter, only reporting counts and timestamps at specific locations, too coarse to meaningfully identify behaviour.

Shared-bicycle schemes offer new research possibilities. In many recent schemes, data on usage are continually reported to central databases. Researchers working within data mining [5] [7], and information visualization [13] have queried these data to identify patterns of scheme use, as well as more nuanced space-time journey flows. This analysis has nevertheless been constrained by the level of detailed information made publicly available. Whilst a complete set of journey records, including journey origin-destination (OD) and start and end times was used by Wood *et al.* [13], these data could not be linked back to individual customers. Cyclists' journey histories, and the context framing those journeys, could not be identified. This limits the extent to which such data can be used to engage with the more complex questions around motivations and barriers to cycling [7] [13]. Working collaboratively with Trans-

---

*e-mail: roger.beecham.1@city.ac.uk

†e-mail: j.d.wood@city.ac.uk

‡e-mail: audreybowerman@tfl.gov.uk
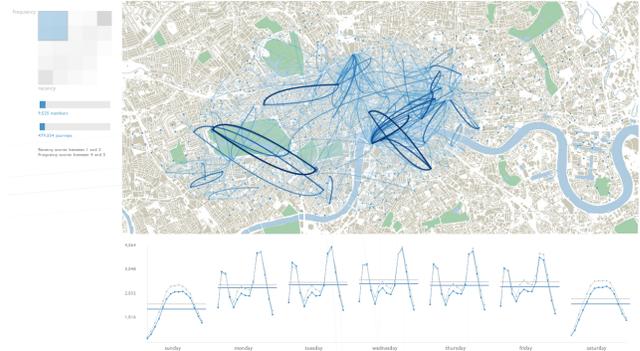
[1]Transport for London's cycle hire website.

Figure 1: An early visual analytics prototype, which links customer segments (top left), with a spatial (centre) and temporal (bottom) view.

port for London (TfL), customer records reporting a unique customer identifier, gender and postcode the customer registered with, have been made available. So too has a complete set of user journeys. Linking with geodemographic and other contextual information, and querying these attribute rich data within a visual analytics application, we attempt to explore and explain cycling behaviour from an individual customer perspective.

2 OBJECTIVES AND APPROACH

The research project aims to:

- Classify bike share customers according to the journeys they make.

- Validate, rewrite and add qualitative descriptions to these classifications paying attention to journey context; by querying the dataset at particular space-times, in response to changes internal and external to the scheme.

- Furnish social scientists, and strategists within TfL, with generalisable insights into the barriers, incentives and conditions that motivate cycling behaviour.

This approach sits comfortably within a visual analytics framework [11]. We will take a large and attribute rich dataset, query it to identify general and distinct space-time journey patterns, before creating modelled data and subsequently querying the models. Attempting to engage with difficult questions around cycling behaviours, it will be necessary to quickly consider many combinations of contextual variables and customer attributes. This might be best achieved through a highly flexible visual analytics application.

3 EARLY ANALYSIS AND FIRST VISUAL ANALYTICS PROTOTYPE

At the time of writing, the dataset contained 114,947 valid customer records, linked to 6,490,479 journeys. After loading data into an SQLite database, relevant derived variables were computed. Linking customers with their journeys, recency-frequency (RF) segmentation, a technique used within direct marketing [8], was performed. Matching customers' postcodes to geographic coordinates,

a straight-line 'distance to nearest docking station' variable was created and customer records were further augmented by linking the postcode variable to a geodemographic classifier.

Access to the customer dataset enables a profile of London bike share members. We found the geodemographic composition of members to reflect relatively closely that of London's residents, that the average (median) 'distance to nearest docking station' for members with valid postcodes is 1.7km, with the middle 50% of members ordered according to nearest docking station living between 0.2 km and 13 km of a docking station. Comparison by RF scores suggests an association between usage and proximity to a bike share docking station. For high RF segments (combined RF scores of 4/5), the median distance to a docking station is 0.4km, whereas for lower segments (combined RF scores of 1/2) the median distance is 2.8km.

Whilst the summary statistics are instructive, without very detailed analysis it is difficult to describe the temporal and spatial extent of particular member journeys. We therefore propose a visual analytics prototype for rapidly exploring space-time patterns of customer cycling behaviour (Figure 1). Three integrated views are presented. In the centre, a spatial overview of journeys is achieved by drawing lines between sets of OD pairs. For a given pair of bike share docking stations A and B, flows from A to B are made distinct from those between B to A. Following Wood *et al.* [13], this was implemented using Bezier curves made to be asymmetric; the straight end of an OD pair represents an origin, the curved end a destination. To emphasise flow magnitudes, and to partially overcome problems of salience bias, a weighting factor determines the transparency, thickness and colour of flow lines [13]. Existing analysis into the London bike share scheme has found that usage is highly temporally regular [7], and that journeys made within these regular usage intervals have a distinct spatial expression [13]. The temporal view therefore displays hourly usage by day of week as a cycle plot [10]. Finally, to the top left, a customer segment view appears as a Recency (column) Frequency (row) matrix (RF matrix) [12] [8]; the most frequent and recent customers can be found in the top right cell. It is possible to brush [2] values from the RF view and therefore describe the spatial pattern of journeys made by particular RF segments, compare their temporal structure (blue) to that of the total customer population (grey) and contrast the relative number of journeys made by the selected segment with the relative number of customers that segment comprises (bars below the RF view).

Brushing the RF view first on high RF scores, the most recent and frequent scheme users, a dominant commuting function can be identified. Within the temporal view, higher than expected peaks appear during weekdays and, compared to all users, these members make relatively few journeys during the weekends. Within the spatial view, journeys between major rail stations, central London and the City of London also become more prominent. The corollary is true of low RF segments: the weekend pattern is reinforced and journeys within Hyde Park (centre left) and west London become more salient. Selecting on high frequency, low recency scores, the view that appears in Figure 1, reveals a more complicated pattern. The temporal structure of journeys is broadly comparable to that of all journeys. Flows originating from inner city areas south of the river, however, become more prominent, and these journeys generally extend over greater distances. Journeys associated with both leisure and commuter cycling also appear equally significant. Finally, when brushing low to high recency scores the spatial and temporal structure of journeys throughout the scheme remains broadly consistent. Members making more recent journeys, however, tend to travel between a wider set of docking station locations; and this is irrespective of how frequently they use the scheme. For those that have not recently used the scheme (lower recency scores), journeys are ostensibly between the larger and most visible docking stations, and we speculate that as the scheme matures, members become more confident in finding, or at least are made better aware of, the full set of bike share docking stations available to them.

The second iteration of the application will support greater levels of interaction. Brushing of both the temporal and spatial views and allowing filtering by gender, geodemographic group, 'distance to nearest docking station' and user travel times, will allow substantive exploratory potential. If more difficult questions around motivations and barriers to cycling are to be engaged with, then the application will also need to more fully attend to journey context, supporting flexible querying at various spatial and temporal resolutions.

4 CONCLUSION

We outline a programme of research and an early visual analytics prototype for interrogating customer cycling behaviour within a large, transactional dataset. Our approach enables space-time patterns of bike share cycling behaviour to be associated with customers' usage characteristics. Due to its scale and complexity, we feel the dataset and our approach can offer new insights into understanding the barriers and incentives that motivate cycling behaviour. Future work will consider alternative customer classifications and more fully attend to the context framing these classifications.

REFERENCES

[1] R. Aldred. On the outside: constructing cycling citizenship. *Social & Cultural Geography*, 11:35–52, 2010.

[2] R. Becker and W. Cleveland. Brushing scatterplots. *Technometrics*, 29(2):127–142, 1987.

[3] J. Dill and J. Gliebe. Understanding and measuring bicycling behavior: A focus on travel time and route choice. *Bicycling*, 29:1 – 70, 2008.

[4] *Department for Transport*. National Travel Survey 2010. Technical report, *Department for Transport*, London, UK, 2011.

[5] J. Froehlich, J. Neumann, and N. Oliver. Measuring the pulse of the city through shared bicycle programs. In *Proceedings of the International Workshop on Urban, Community, and Social Applications of Networked Sensing Systems*, pp.16–20, Raleigh, North Carolina, USA, Nov. 2008.

[6] G. Gordon. Developing methodological approaches to analysing single point bicycle counts. In *Proceedings of the Universities Transport Study Group 45th Annual Conference*, University of Aberdeen, UK, Jan. 2012.

[7] N. Lathia, S. Ahmed, and L. Capra. Measuring the impact of opening the London shared bicycle scheme to casual users. *Transportation Research Part C: Emerging Technologies*, 22:88–102, 2012.

[8] R. Parekh and R. Kohavi. Visualizing RFM segmentation. In *Proceedings of the 4th SIAM International Conference on Data Mining*, Florida, USA, Apr. 2004.

[9] C. Pooley, M. Tight, D. Horton, G. Scheldeman, A. Jopson, C. Mullen, and A. Chrisholm. Understanding walking and cycling: Summary of key findings and recommendations. Technical report, *Lancaster University Environment Centre*, Lancaster University, UK, 2011.

[10] N. B. Robbins. *Creating More Effective Graphs*. Wiley-Blackwell, New Jersey, USA, 2005.

[11] J. Thomas and K. Cook. A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26(1):10–13, 2006.

[12] J. Wood, R. Radburn, and J. Dykes. vizLib: Using the seven stages of visualization to explore population trends and processes in Local Authority research. In *Proceedings of the GIS Research UK 18th Annual Conference*, University College London, UK, Apr. 2010.

[13] J. Wood, A. Slingsby, and J. Dykes. Visualizing the dynamics of London's bicycle hire scheme. *Cartographica*, 46(4):239 – 251, 2011.