Beecham, R., Wood, J. & Bowerman, A. (2012). Identifying and explaining inter-peak cycling behaviours within the London Cycle Hire Scheme Conference. Paper presented at the Workshop on Progress in Movement Analysis: Experiences with Real Data, 15 - 16 Nov 2012, Zurich, Switzerland.

CITY UNIVERSITY LONDON
EST 1894

[City Research Online](#)

Identifying and Explaining Inter-Peak Cycling Behaviours within The London Cycle Hire Scheme

R. Beecham[1], J. Wood[2], A. Bowerman[3]

[1]City University London, St. John's Street, London, EC1V 0HB
Email: roger.beecham.1@city.ac.uk

[2]City University London, St. John's Street, London, EC1V 0HB
Email: j.d.wood@city.ac.uk

[3]Transport for London, Palestra, 197 Blackfriars Road, London SE1 8NJ
Email: audreybowerman@tfl.gov.uk

1. Introduction

Since the launch of the London Cycle Hire Scheme (LCHS) in July 2010, over 15 million bike journeys have been made in and around central London. The scheme continues to extend geographically, but for *Transport for London* (TfL), the authority responsible for its implementation, two concerns shape the scheme's future: that it continues to attract new people to cycling in the city; and that it is financially sustainable. Both these ambitions are potentially threatened by the fact LCHS is used ostensibly for commuter travel. Recent analyses of LCHS usage data have found a tidal flow of bikes into and out of central London and the City, which coincide with commuting peaks (Wood et al. 2011, Lathia et al. 2012). To keep the system balanced, bikes are transported across the city at peak times, and in priority areas, docking stations are continually replenished with bikes or bikes continually removed from docking stations. Whilst this level of fleet management is expensive, with limited space to expand in central London, the alternative - increasing capacity at these sites - is not feasible. A desirable solution would be a more natural redistribution by customers using the scheme throughout the working day, and a challenge for policy makers at TfL is to encourage this inter-peak usage.

Working with colleagues at TfL, and with access to LCHS's customer and journeys databases, we attempt to identify and explain inter-peak travel by exploring the context, circumstances and customer characteristics that underpin inter-peak journeys. Using techniques from information visualization and geovizualization (Dykes 1997, Roberts 2004, Wood et al. 2011), we demonstrate how this type of analysis can be achieved using a visual analytics application. After outlining our approach, we present some initial findings and discuss how, in better understanding inter-peak travel behaviours, we aim to provide insights that may directly inform promotional activities and operational decisions around the LCHS's expansion.

2. Dataset and Analysis Techniques

2.1 The LCHS Customer and Journeys Dataset

Our analysis relies on two complementary datasets: a customer database and a complete set of journey records. For every customer registering with the LCHS, that individual's gender and postcode are stored within a customer database and a unique customer identifier is generated. For every journey made, an origin-destination (OD) pair representing the docking station that journey started and finished at, along with timestamps for these instances, are recorded within a journeys database. The journey

data do not provide details about specific routes taken by customers. However, by relating the two datasets - by linking customers with their journeys - we can explore how *individual* members use the scheme. As yet there is no way of identifying returning casual payment users, or non-members, who make around 25-30% of all journeys.

After processing over 120,000 members and 11 million member journeys, a number of derived variables were computed. Assuming the postcode variable represents customers' home address, straight-line distances from customers' postcodes to their nearest docking station were calculated, and a 'distance to docking station' variable created. The postcode variable was then linked to two geodemographic classifiers: the 2001 Census Output Area Classification (OAC) (Vickers & Rees 2006) and the 2010 Indices of Multiple Deprivation (DCLG 2011). To further disaggregate customers, we used recency-frequency (RF) segmentation, a technique used in direct marketing to classify customer purchase behaviours (Kohavi & Parekh 2004, Wood et al. 2010). 'Recency' scores were calculated by identifying customers' most recent journey and assigning discrete scores within five equal frequency bins, from most (score 5) to least (score 1) recent. For 'frequency', the first and last journey appearing in customers' records were identified and the average weekly number of journeys made between these dates computed. Finally, personalised travel time z-scores were calculated, whereby an individual's travel times for the journeys they make are compared to the same journeys made by the total population of bike share customers.

Supplementing our dataset with these geodemographic and behavioural data enables an attribute-rich, as well as longitudinal, record of customer cycling behaviour.

2.2 Visual Analytics Application

We wanted to quickly explore the extent to which these customer variables relate to inter-peak cycling, and built a visual analytics application for querying our combined dataset (Figure 1). We chose to represent this information within four linked views (Roberts 2004, Dykes 1997). In the centre, we attempt to show all journeys by drawing lines between sets of OD journey pairs. Following Wood et al. (2008), we use asymmetric Bezier curves to encode direction, and to communicate flow magnitudes, we vary the transparency, thickness and colour of flow lines according to their frequency. Below the flow map, hourly usage by day of week appears as a cycle plot (Robbins 2005). The left margin of the graphic displays our customer related variables: RF scores are presented within a matrix (Kohavi & Parekh 2004, Wood et al. 2010); the gender and geodemographic variables appear as horizontal bars; and the 'distance to docking station' and travel time z-score variables are shown as histograms.

It is possible to brush (Becker & Cleveland 1987) each of these views simultaneously and therefore explore particular combinations of contextual variables. When making such selections, we use colour to delineate between selected customers (blue) and the total (grey) member population. Within the RF matrix we show the extent to which RF characteristics of selected customers deviate from the total population by displaying signed residuals from the chi-statistic within a diverging colour scheme (Figure 1).
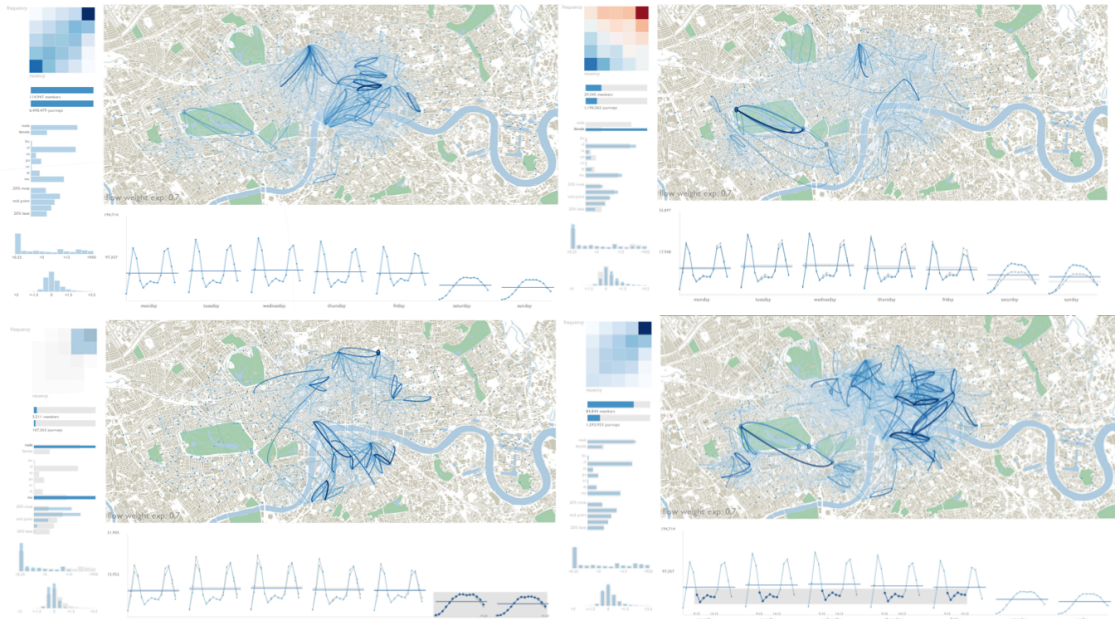
Figure 1. Screenshots from VA application. Clockwise from top left: default view without variable selection; female members and signed chi-residuals appear in RF matrix (negative residuals are red, positive residuals are blue); weekday interpeak journeys; weekend journeys made by male high RF members living in OAC classification *multicultural*.

3. Findings

We aim to evaluate all journeys that take place outside of peak weekday travel times. First, then, we filter journeys made at weekends. Considering the total member population, OD flows that start and end around London's major parks dominate, and in terms of customer variables, female members, those living within London, and members with slightly slower than average travel times are overrepresented. Querying these journeys further, we find that weekend journeys appear more utilitarian in nature as the scheme matures, and with particular geodeomographic and behavioural characteristics: London members, who use the scheme regularly and who live in the OAC classification *multicultural* - inner-city communities experiencing moderate-to-high levels of deprivation - are most likely to make these utilitarian journeys. Such journeys tend to be concentrated around more peripheral docking station locations, but in relatively busy and popular parts of the city: Angel, Clerkenwell, Shoreditch, Borough, Kennington and Vauxhall.

Filtering on journeys that take place during the working day and outside of peak travel times, we again find a number that are made within London's parks, but also significant flows between central London and the City of London. Flows to and from major rail terminals are particularly prominent. Exploring these journeys further, for high RF customers, inter-peak journeys in Hyde Park and Regent's Park become much less salient, and instead journeys within the City of London dominate. Importantly, though, many of these journeys appear reciprocal, suggesting that in these instances bikes are not disproportionately transported from one part of the city to another.

Finally, brushing short journeys that take place in central London (Figure 2), inter-peak weekday journeys become overrepresented. Those within the highest RF segment

- members who have recently used the scheme and use it regularly - make almost half of these journeys, and of this subset, male members are more likely to make such journeys within the working day than are female members.
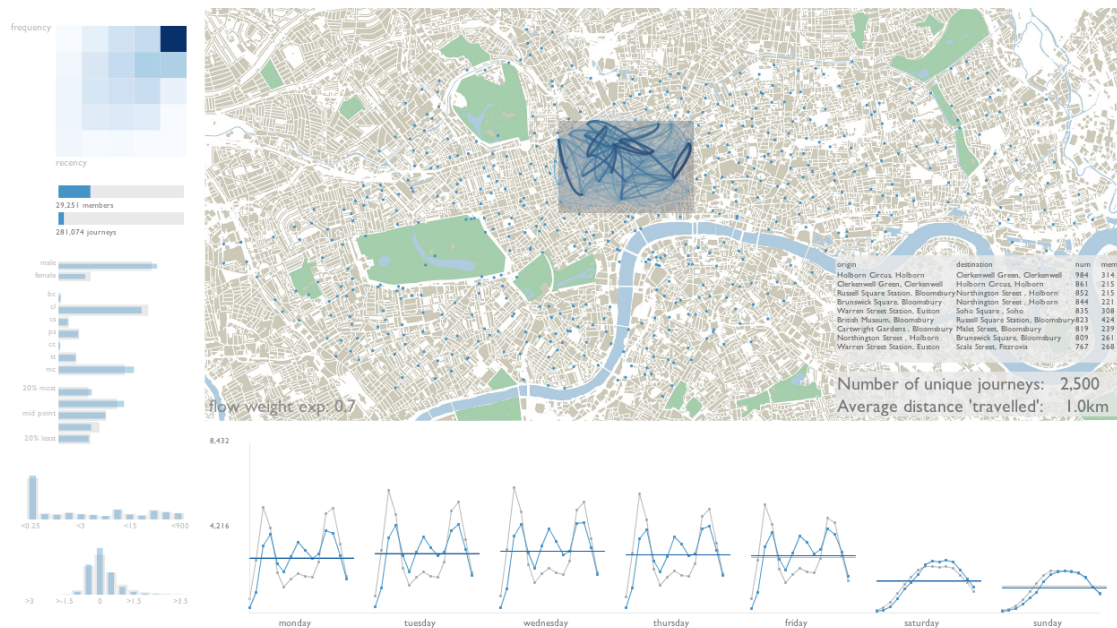

Figure 2. Spatial selection on short journeys made within Central London.

4. Discussion and future work

The outcomes from this work may have direct implications to the scheme's expansion and continuing operation. In March 2012, the LCHS boundary was extended east to Canary Wharf, and a further phase of expansion is planned for the residential boroughs of Wandsworth, Hammersmith and Fulham, Lambeth and Kensington and Chelsea (TfL 2012a). That we find increases in utilitarian weekend travel within more residential, but popular parts of the city, and that we can relate these journeys to specific sub-sets of the user population, is therefore encouraging. So too is the fact that we identify inter-peak utilitarian journeys made by heavy scheme users within the working day, and that these journeys do not appear to disproportionately transfer bikes into particular sections of the city. By analysing these journeys over time, and against a wider set of external variables, we hope to classify inter-peak weekday journeys more formally, and make firmer inferences about journey purpose. Such insights might be used by LCHS policy makers in considering promotional activity targeted at, for example, universities or employers. A further research priority will be to mine the journey dataset to identify and describe instances of 'group cycling' at inter-peak periods. Efforts are currently being made to encourage more extended leisure cycles (TfL 2012b), and we speculate that such journeys might be most likely within groups.

With the increasing availability of similarly large and continuous sensor data, this work is of relevance to the wider research community. Our approach - designing and building a tailored visual analytics application - enabled us as analysts to explore complex social-spatial phenomena in an efficient and iterative way. More importantly, the immediacy of the interactions, and the clarity of our design, meant that policy makers and strategists could fully participate in this analysis. We feel this assemblage of analyst and policy maker significantly improves the research process and may facilitate deeper, more substantive research outcomes.

Acknowledgements

References

Becker, R.A. & Cleveland, W.S., 1987. Brushing Scatterplots. *Technometrics*, 29(2): 127–142.

Department for Communities and Local (DCLG), 2011. *The English Indices of Deprivation 2010: Technical Report*, Available at: http://www.communities.gov.uk/publications/corporate/statistics/indices2010technicalreport [Accessed September 5, 2012].

Dykes, J.A., 1997. Exploring spatial data representation with dynamic graphics. *Computers & Geosciences*, 23(4): 345–370.

Kohavi, R. & Parekh, R., 2004. Visualizing RFM Segmentation. In *Proceedings of the 4th SIAM International Conference on Data Mining*. Florida, USA.

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

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

Roberts, J.C., 2004. Exploratory Visualization with Multiple Linked Views. In *Exploring Geovisualization*. Amsterdam: Elseviers. Available at: http://kar.kent.ac.uk/14046/ [Accessed September 5, 2012].

Transport for London, 2012a. Barclays Cycle Hire Expansion. Available at: http://www.tfl.gov.uk/corporate/projectsandschemes/18339.aspx [Accessed September 5, 2012].

Transport for London, W.H., 2012b. Cycling Leisure Routes. Available at: http://www.tfl.gov.uk/roadusers/cycling/23069.aspx [Accessed September 5, 2012].

Vickers, D. & Rees, P., 2006. Introducing the area classification of output areas. *Population trends*, (125): 15–29.

Wood, J., Radburn, R. & Dykes, J., 2010. 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, April 2010.

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