Exploring gendered cycling behaviours within a large-scale behavioural dataset

R. Beecham[a]; J. Wood[b]

[a,b] giCentre, City University, London, United Kingdom

[a,b] Rm A304, College Building, St John's Street, City University, London, EC1V 0HB

[a] Email: Roger.Beecham.1@city.ac.uk; +44 20 7040 3914 (Corresponding author)

[b] Email: J.D.Wood@city.ac.uk; +44 20 7040 0146

Received 11 March 2013

Revised 5 July 2013

Exploring gendered cycling behaviours within a large-scale behavioural dataset

Analysing over 10 million journeys made by members of London's Cycle Hire Scheme, we find that female customers' usage characteristics are demonstrably different from those of male customers. Usage at weekends and within London's parks characterises women's journeys, whereas for men, a commuting function is more clearly identified. Some of this variation is explained by geodemographic differences and by an atypical period of usage during the first 3 months after the scheme's launch. Controlling for each of these variables brings some convergence between men and women. However, many differences are preserved. Studying the spatio-temporal context under which journeys are made, we find that women's journeys are highly spatially structured. Even when making utilitarian cycle trips, routes that involve large, multi-lane roads are comparatively rare, and instead female cyclists preferentially select areas of the city associated with slower traffic streets and with cycle routes slightly offset from major roads.

Keywords: gender and cycling behaviour; bicycle share schemes; visual analytics; behavioural datasets.

1 Introduction

As access to public or shared transport systems becomes increasingly digitised, new datasets have emerged offering opportunities to research travel behaviour in a continuous, large-scale and non-invasive way (Blythe and Bryan 2007; Froehlich, Neumann, and Oliver 2008; Kusakabe, Iryo, and Asakura 2010; Páez, Trépanier, and Morency 2011; Lathia, Ahmed, and Capra 2012). The data produced by urban bike share schemes can be regarded as a particular instance of these new datasets. In most recent bike share schemes, data on usage are continually reported to central databases. Researchers working within data mining (Froehlich, Neumann, and Oliver 2008; Jensen et al. 2010; Borgnat et al. 2011; Lathia, Ahmed, and Capra 2012) and information visualization (Wood, Slingsby, and Dykes 2011) have processed and then queried these data to identify patterns of usage at various spatial and temporal resolutions. Some of this work has been used by scheme operators to help overcome problems around fleet management, and by policy makers for better understanding usage at particular docking stations. It has nevertheless been constrained by the level of detailed information made easily available (Wood, Slingsby, and Dykes 2011; Lathia, Ahmed, and Capra 2012). In many studies, data were harvested from the web, where local transport authorities publish in real-time the number of available bikes at individual docking stations (Froehlich, Neumann, and Oliver 2008; Lathia, Ahmed, and Capra 2012). Others gained access to journey records, including journey origin-destination (OD) and start and end times, and identified more sophisticated usage characteristics (Jensen et al. 2010; Borgnat et al. 2011; Wood, Slingsby, and Dykes 2011). Without access to a customer database, however, journeys could not be linked back to individual customers: individuals' journey histories, and the context framing those journeys, could not be identified. This limits the extent to which such datasets can be used to study the more complex motivations and barriers that might affect cycle behaviours.

Working collaboratively with Transport for London (TfL), this more detailed information on usage of the London Cycle Hire Scheme (LCHS) has been made available for specific use in this research. A full set of customer records containing customers' gender and the postcode they registered with, along with a database of over 16 million journeys made since the scheme's inception, have been provided. An aim for strategists at TfL is to attract greater numbers of women to the scheme, and numerous empirical studies have found very distinct attitudes towards, and varying experiences of, cycling amongst men and women (Garrard, Rose, and Lo 2008; Emond, Tang, and Handy 2009; Heesch, Sahlqvist, and Garrard 2012). After describing the dataset and our approach to analysis, we explore the extent to which cycling behaviours differ between male and female bike-share members. Our analysis may have wider implications than the LCHS itself, and we conclude by discussing findings in the context of existing research into gender and cycling behaviour.

2 Related work: gender and cycling behaviour

Within the social sciences, research into the barriers and incentives that motivate particularly urban cycling is burgeoning (Pucher and Buehler 2012). A substantial area of research is around gender and cycling behaviour. Whilst in bicycle-friendly cities and countries cycling is seen as a highly inclusive activity open to most demographic groups, in car-oriented cities it is the preserve of largely young and middle-aged men (Garrard, Handy, and Dill 2012). There are various explanations for the gender gap in cycling uptake for low-cycling environments. Detailed qualitative studies have related motivations around cycling amongst women and men to particular personal circumstances and life stages (Bonham and Wilson 2012). Larger survey-based research suggests that these differences between men's and women's uptake might relate to preference: men are more likely than women to agree that they enjoy cycling (Emond, Tang, and Handy 2009). A substantial barrier is that of perceived personal safety. A relatively large survey of 1,862 cyclists in Queensland, Australia found that women are more likely to cycle off-road than men; are less likely to commute by bicycle than men; and that, although factors related to traffic conditions, motorist aggression and safety are concerns for both women and men, women report a far greater number of these constraints (Heesch, Sahlqvist, and Garrard 2012). Similar findings were identified by Tilahun, Levinson, and Krizek (2007) in a study of participants' stated preferences around route choice. Observational studies have also shown these preferences to be expressed in women's *real* cycle behaviours. In Portland, Oregon, a sample of 166 self-selected participants were recruited and their cycling monitored using GPS (Dill and

Gliebe 2008). Compared with male participants, women made a smaller share of their journeys on major roads or routes without bike lanes and more often cycled on low-traffic streets or boulevards (Dill and Gliebe 2008).

Clearly there is a growing empirical basis around the barriers and factors that motivate men and women to cycle. We use this work to validate our own findings discussed in section 5. However, amongst this existing research the scale and scope of observational based studies – studies that consider *real* behaviours – is relatively limited. Dill and Gliebe's (2008) study of 166 participants, for instance, is one of the largest and most comprehensive of its kind. Datasets of such size are clearly problematic where analysis at finer spatial and temporal scales is required. Other than their size, the fact that such studies typically rely on participants remembering to carry and activate their GPS devices whenever a journey is made is a concern. In addition, social-desirability bias, where a participant affects their behaviour simply because they are being observed, might also exist in research contexts where a healthy and socially useful activity like cycling is being monitored (Dill 2006).

The advent of large-scale urban bike-share schemes offers new opportunities for studying observed behaviour. Although there are problems associated with the LCHS dataset, its size is unprecedented: we have a population of over 135,000 cyclists and more than 10 million journeys. Since every journey that an individual makes through the scheme is recorded, we have a complete history of each individual's usage. Moreover, since only journeys taken through the LCHS are considered, we necessarily standardise by variables such as the size and nature of bikes being used, individuals' ability to access bikes (at least at the scheme-wide level), the cost of making a journey and the geography of the city - central London.

3 Datasets

This research relies on two complimentary 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 identifier is generated. For every journey made, an origin-destination (OD) pair representing the bike share docking station that journey started and finished at, along with timestamps for these instances, are recorded within a journeys database. The journeys 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.

Customers

memID	###82
gender	f
postcode	nw5
distance	1.3km
oac	cl
imd	3
recency	3
frequency	4
ttZScore	0.6

Journeys

memID	oTime	dTime	oStation	dStation
###82	18:44:26	18:50:20	61	223
###82	11:06:24	11:15:04	62	223
###82	22:09:24	22:23:19	94	94
###82	20:30:36	20:46:26	94	194
###82	19:00:17	19:04:38	94	269
###82	14:30:38	14:34:17	94	269
###82	07:58:09	08:02:05	94	269
###82	21:05:22	21:11:34	94	269

Figure 1: 'Journeys' and augmented 'Customers' datasets.

After processing over 135,000 LCHS members linked to over 10 million member journeys, a number of derived variables aimed at disaggregating customer behaviours were computed. Assuming the postcode variable represents a home address, straight-line distances from customers' home 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) and the 2010 Indices of Multiple Deprivation (Department for Communities and Local Government 2011).

In order to disaggregate customer behaviours, we used Recency-Frequency (RF) segmentation, a technique used in direct marketing to classify customer purchase behaviours (Kohavi and Parekh 2004). RF segmentation is based on empirical research that finds both Recency (how recently a customer bought or used a product) and Frequency (how often a customer has bought or used a product) to be very good predictors of future purchase (Novo 2004). In a RF analysis customers are given scores for Recency and Frequency on a 5-point-scale. These scores are then concatenated to give 25 customer segments. For the LCHS dataset, '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 each customer's records was identified, and the total number of journeys divided by the time that elapsed between these two dates. Members with the highest scores for both Recency and Frequency (5-5) are the most 'heavy' scheme users; they make journeys regularly and their last journey was very recent. Those within the lowest RF group (1-1) typically used the scheme when first registering, but have made very few, if any, journeys since.

Finally, personalised travel time *z-scores* were calculated whereby an individual's average travel time for each unique journey (OD pair) they make $(\bar{u}_{od})$ is compared to the average travel time for the same journey pair made by the total population of bike share customers $(\bar{p}_{od})$. Following statistical probability theory (Field 2009), we only compute this score where the total frequency for that journey made by the member population is at least 30 $(\bar{f}_{od} >= 30)$:

$$zScore(u_{od}) = \frac{\bar{u}_{od} - \bar{p}_{od(fod>=30)}}{std(p_{od(fod>=30)})}$$

To prevent any errors in this very large journeys dataset from affecting the travel time calculation, a minimum journey time of 3 minutes and a maximum of 4 hours was used when computing the scores. User *z-scores* were averaged to give a single value for each member, and the scores were made to fit a normal distribution by taking the square root of travel times when calculating the *z-scores*.

4 Analysis techniques

We wished to explore the extent to which levels of usage, as defined by the RF segmentation, varied by our geo-demographic and other derived variables. Querying these data, cross-tabulations using the chi-statistic were computed, whereby observed frequencies within each RF segment for a subset of members were compared to modelled (expected) frequencies within that RF segment given the member population as a whole. This exploratory analysis enabled significant differences in the level of scheme usage to be related to customers' gender, geodemographic classification and how far customers apparently live from their nearest docking station. Increasingly, however, it was necessary to explore, test and compare *multiple* combinations of these derived variables simultaneously. Moreover, if our analysis was to attend to the context framing particular behaviours, it was vital that we could also identify the temporal and spatial structure of the journeys we were seeking to analyse.

A visual analytics application for performing these sorts of analytics queries 'on the fly' was developed in Processing, a Java based programming environment used for developing visual analytics software. The application (Figure 2) combines three coordinated and linked views (Dykes 1997; Roberts 2005). In the centre, a spatial overview of journeys is achieved by drawing lines between all possible origin-destination (OD) pairs. This is done using Bezier curves, and following Wood, Slingsby, and Dykes (2011) we encode direction by making these curves asymmetric; the straight end representing journey origin, the curved end journey destination. In

Wood, Slingsby, and Dykes (2011) various techniques for visually depicted large numbers of spatially complex flows are critiqued. To overcome problems of visual clutter and salience bias typical in many flow visualizations, the authors propose a weighting factor that emphasises flow magnitudes. We use this same weighting factor (w_{od}) where, for each unique OD pair, the number of journeys made between that pair of docking stations (f_{od}) is scaled to the most frequently travelled OD pair in the dataset (f_{max}):

$$w_{od} = \left(\frac{f_{od}}{f_{max}}\right)^{exp}$$

The weighting factor determines the thickness, transparency and colour of each flow line, so that there is a direct mapping between flow frequency and visual saliency. Varying the exponent (exp) alters the impact of the weighting factor, and if decreased allows less common flows to be given slightly greater prominence. Finally, to further ensure that less common flows do not occlude more common flows, OD pairs are ordered from least to most frequent and then drawn in reverse frequency order.

The temporal view displays hourly daytime usage by day of week as a cycle plot (Robbins 2005). Journeys made at particular times of day and days of week can be selected by interacting with the temporal view and, using a slider, it is possible to analyse behaviours over varying temporal resolutions. The left margin of the graphic displays the customer related variables. RF scores are presented within a matrix (Kohavi and Parekh 2004); 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 interact with any component of our application. Simply clicking or dragging on particular geo-demographic or behavioural groups, time periods and spatial areas, filters those members and their journeys. When making these interactions, with the exception of the map view, selected subsets (blue) can be compared with the total member population (grey) by overlaying one on top of the other. In the RF view, we show variation from expected frequencies by directly mapping signed residuals from the chi-statistic onto a red-blue colour scheme.

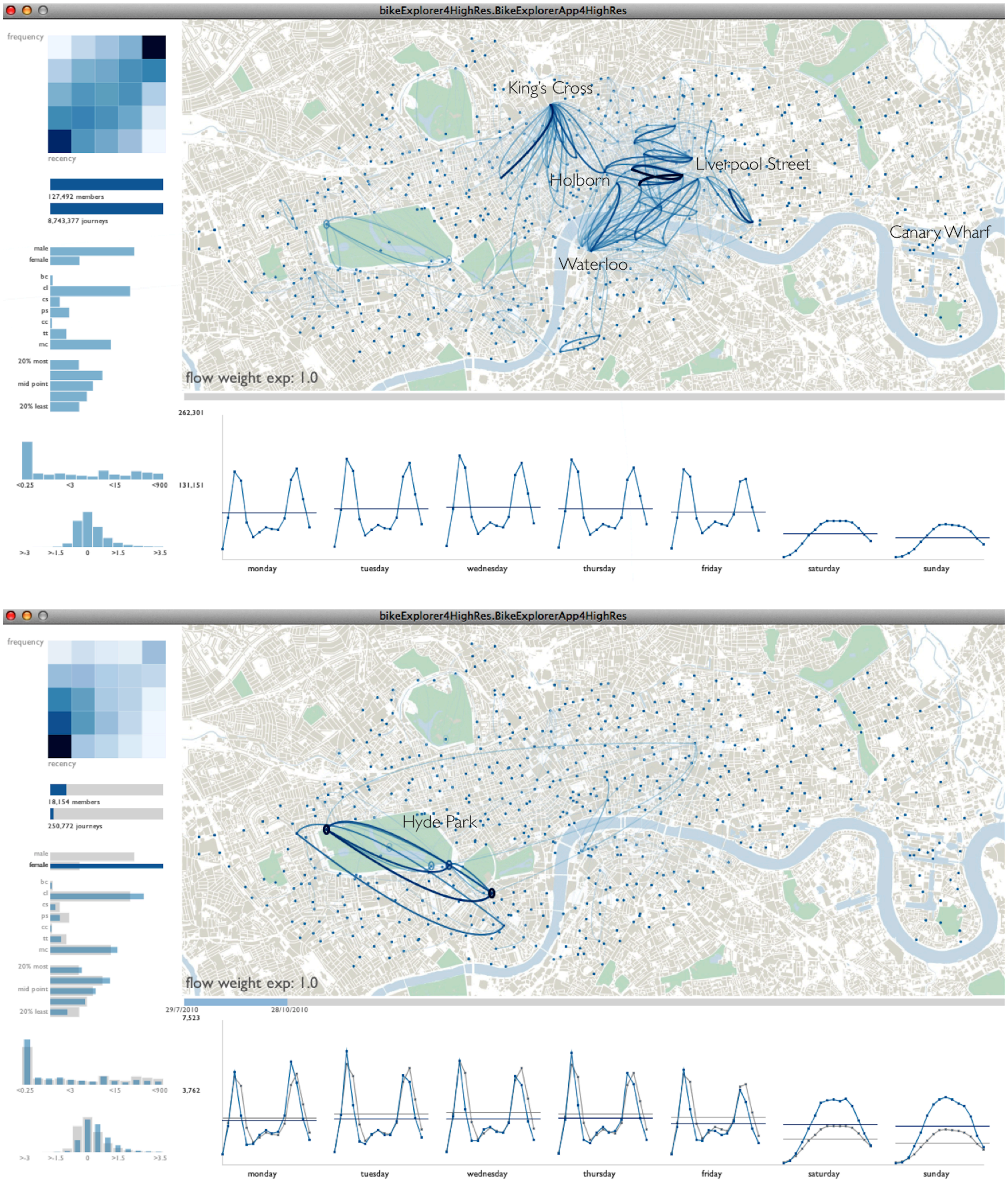

Figure 2: Screenshots of the application. Above: journeys around Liverpool Street Station and Holborn, and between particularly Holborn and London's major commuting rail terminals, King's Cross and Waterloo, are visually salient when all journeys are selected. Below: journeys made by women within the first 3 months of the scheme's

launch are selected, with trips within Hyde Park dominant. Colours are taken from the Brewer ‘Blues’ sequential colour scheme (Harrower and Brewer 2003).

The visual analytics application we describe here enables members’ cycle behaviours to be rapidly explored and related to a combination of spatial, temporal and customer-related variables. The analysis and substantive discussion that follows is a direct result of interactions with the application.

5 Findings

5.1 Contrasting all journeys made by men and women since the scheme’s inception

Women currently make up around a quarter of all LCHS members. They registered with scheme at similar times to men. After a significant surge in interest at the scheme’s launch in July 2010, there were more modest increases in demand amongst both male and female members during January and Summer 2011, and other slight increases in registrations in early Spring and July 2012. Considering our derived variables, substantial differences can be identified in both the geodemographic and behavioural profile of male and female LCHS cyclists. We find higher proportions of women apparently living in urban communities than would be expected given the member population as a whole, and much fewer in affluent, semi-rural communities. Female members also appear to be far less heavy scheme users than men. Whilst they comprise 25-26% of all LCHS members, women make up only 17% of members within the top RF segment - of heavy and recent scheme users - and 34% of members in the bottom RF segment.

Querying male and female journeys within our application, it appears that these differences in usage characteristic have a distinct spatial and temporal expression. For men, flows between London’s major rail terminals and workplaces - between Waterloo, Liverpool Street, central and the City of London (Figure 2) - overwhelmingly dominate the map view, and we find slightly higher than expected flows during weekdays coinciding with commuting peaks. By contrast, for female members, journeys within London’s parks dominate the map view, with round trips - those that finish at the same station they started at - particularly dominant. Weekend journeys also constitute a much larger share of all journeys made by female members: 22% of trips made by women take place at weekends, whilst for men this figure is just 16%.

After exploring these data within our application - particularly within the first three months after the scheme’s launch (Figure 2) - we suggest that retention rates appear to be particularly poor for women. Many female members appear to be within a group who, living relatively close to the scheme’s boundary, registered with the LCHS when it first launched, but after experimenting with the scheme by making a small number of ‘leisure’ journeys ostensibly within London’s Hyde Park (Figure 2), decided not to use it on a regular basis. The travel behaviours we identify for this group of early ‘detractors’ resonate with the anecdotal and high-level analysis carried out by policy-makers at TfL. Partly due to the LCHS’s high profile at its inception, the first three months of usage they regard as atypical. In order to better understand current behaviours we analyse and segment customers using only the most recent 12 months of data.

5.2 Contrasting male and female journeys made between September 2011-2012

Analysing over 5 million member journeys made between September 2011-2012, then, we discover that many of these differences are preserved. There are very significantly fewer ($p < 0.0001$) female members in the highest RF group than would be expected given the member population as a whole, and women are overrepresented amongst the lowest RF scores. The relative number of weekend journeys is greater for women than it is for men, and we find fewer than expected women amongst the faster travel-time *z-scores*. Exploring journeys within our application, men's cycling behaviours again remain highly regular: journeys between major rail terminals and the City of London are clearly visible. For women, however, cycle behaviours appear more varied. Journeys within Hyde Park and west London are visually salient, but we now also begin to see journeys within parts of central London.

One means of quantitatively testing the prominence of commuter flows made by men is to calculate the total share of journeys involving hub stations. Hubs are generally large docking stations located at two major rail terminals – King's Cross and Waterloo – and at the intersection of the City of London and central London (Holborn, labelled in Figure 2). In order to cope with very high demand at peak times, bikes are continually replenished at, or withdrawn from, these strategically important stations. Compared with women, very significantly more men ($p<0.0001$) make journeys that either start or end at a hub station: 31% of men versus 21% of women. This is also true when comparing numbers of journeys: 10% of men's journeys involve a hub station, whereas this figure for women is 3%.

These insights led us to study in more detail the most common journeys made by men and women. Ranking journeys (OD pairs) between specific docking stations according to their frequency, and plotting these ranks and sizes reveals a power-law distribution (Reed 2001) whereby rank position is inversely related to journey frequency (Figure 3). Whilst both curves for men and women follow this familiar distribution, the gradient on the curve is slightly steeper for women, suggesting that the rank-size effect is severe.

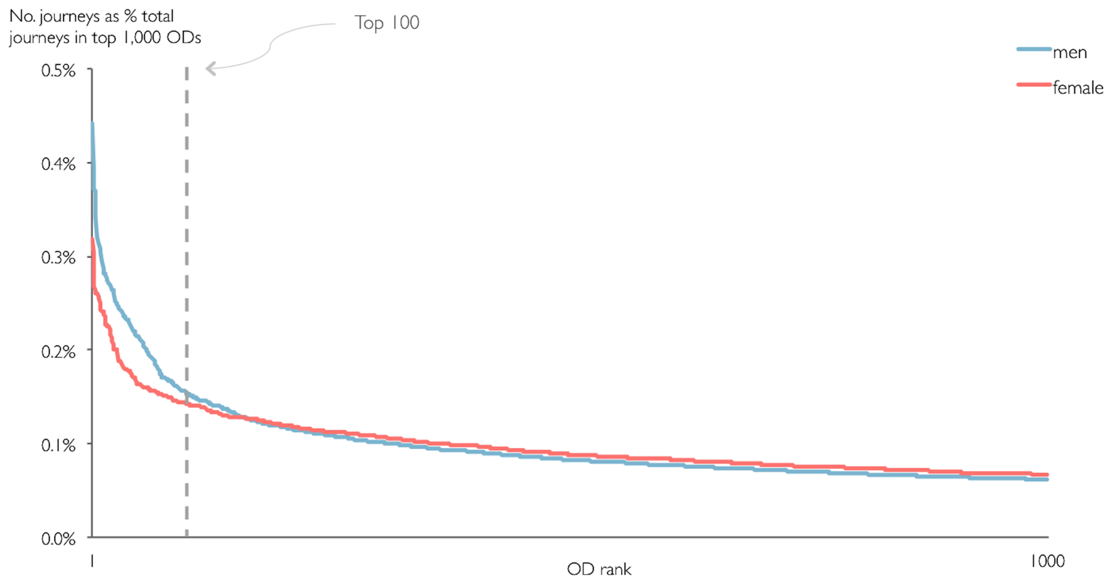

Figure 3: Rank-size distribution of 1,000 most commonly made journeys for male and female members.

Studying journeys within our application, we can explore these heavily repeated journeys and infer more about their context and purpose. Figure 4 shows the 100 most common journeys made by male (top) and female (bottom) members. For men, we immediately find a familiar spatial and temporal pattern, with journeys almost exclusively suggesting a commuter function: weekday journeys between 6am-9am and 4pm-7pm account for 75% of all journeys, with weekends only accounting for 2% of these journeys. When analysing women's top 100 journeys, a large number also coincide with weekday commuting times. This might be expected since these are heavily repeated trips. Notice though, that it is only the morning peaks that are overrepresented. Inspecting all journeys made by 'commuting' female members - those within the high RF segments - this pattern is reinforced: when commuting, female members are more likely than men to make journeys in the morning peak. Unlike the patterns we observe for men, though, weekend journeys are not entirely absent. Around 10% of the top 100 journey combinations for women are made at weekends and, inspecting the map view, 'leisure' journeys within Hyde Park remain visually salient.

We also see a significantly greater number ($p<0.0001$) of apparently utilitarian journeys between King's Cross and the Bloomsbury area of London (highlighted in Figure 4): 19% of women's top 100 journeys are made within this area, whereas for men this figure is 8%. There is a sense here that, even when making utilitarian journeys, female members may preferentially select more cycle-friendly parts of the city. Journeys between docking stations at either side of the River Thames - routes that generally involve relatively large, multi-lane roads and busy junctions - are rare. Instead at peak times, journeys around the Bloomsbury area (Figure 4), where roads are narrower, a number of traffic calming measures have been introduced and cycle lanes slightly offset from major roads, are more common.

We can further quantitatively test the finding that women make fewer journeys that involve a river crossing by filtering only those journeys. Whilst 50% of men have made journeys that involve a river crossing, this figure for women is 41%, a very significant difference ($p<0.0001$). These differences are even preserved when controlling for how heavily members use the scheme. Sixty-seven percent of high RF men have made journeys that involve a river crossing, and these journeys represent 21% of all journeys taken by high RF men. These figures for women are 62% and 16% respectively, and differences both between the number of people making journeys and actual journeys being made are again very significant ($p<0.0001$).

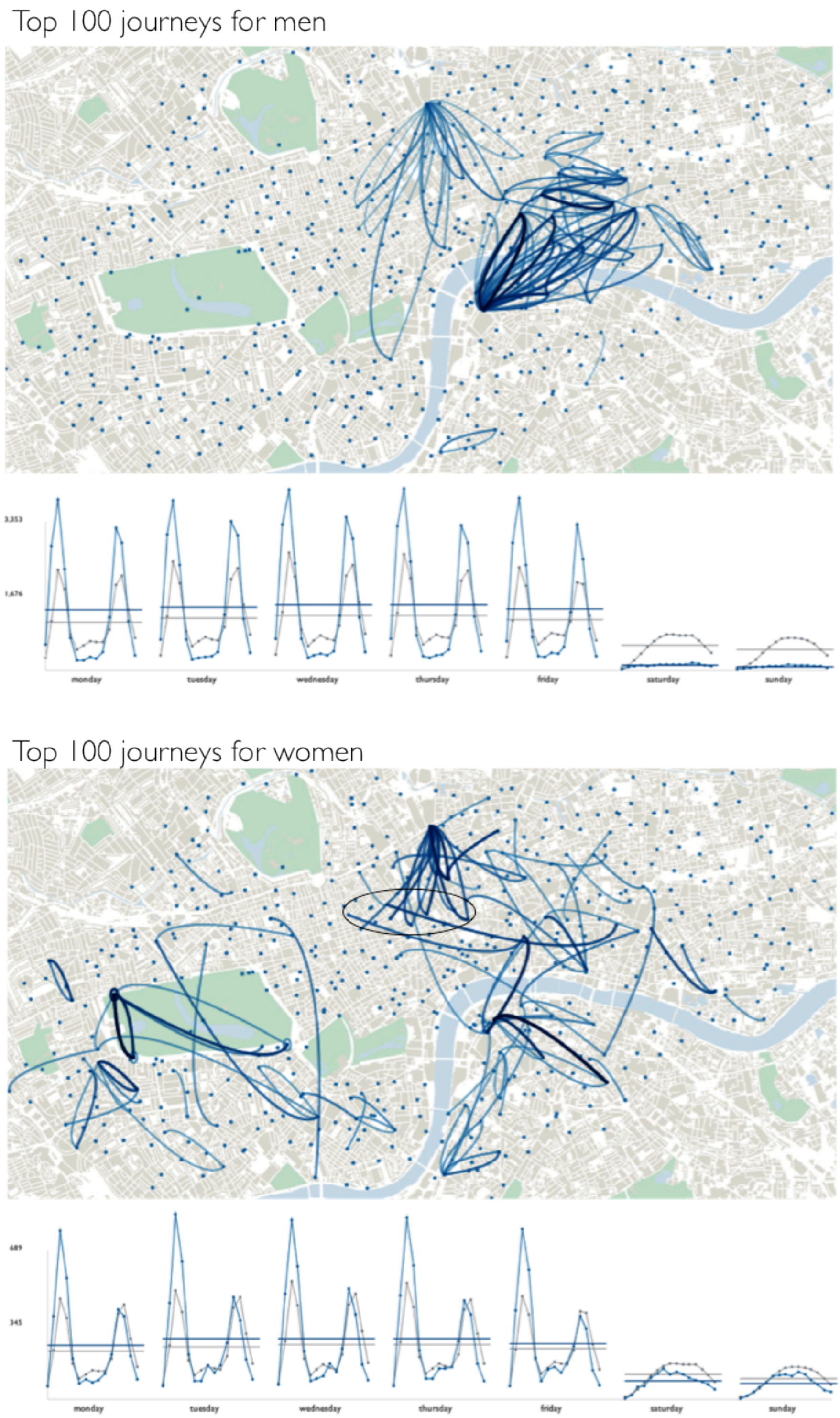

Figure 4: Top 100 journey pairs made between September 2011-2012 by male (above) and female (below) members. Docking stations within Bloomsbury area are highlighted.

5.3 Controlling for geodemographic variations between men and women

Although these findings are true of the total member population, we should be cautious in treating these gendered differences as essential. The dominant pattern when querying male users is of commuter travel. Highly visible amongst these journeys is a group of users apparently living in semi-rural and suburban communities who, after commuting into London on a train, routinely use the scheme to make these highly regular journeys. Women are very significantly underrepresented ($p<0.0001$) amongst the non-London member population. They represent just 15% of all members living more than 15km from a docking station, and are therefore underrepresented amongst this group of very heavy scheme users. In evaluating male and female members, then, we are comparing two different populations. We can control for these differences within our application by selecting only members who apparently live less than 5km from a docking station. This subset represents over 50,000 members and 3.2 millions journeys made between September 2011-2012.

Immediately we find greater convergence between male and female members. Though to a lesser extent than for women, men living less than 5km from a docking station become slightly overrepresented amongst weekend journeys. The spatial patterns of men's journeys are now far less regular, with journeys between Waterloo and the City of London no longer dominating the map view. Flows within Hyde Park and west London can also be identified and journeys within the recent eastern expansion area (opened in March 2012) are now visible. We also find a very diverse set of journeys extending into the semi-residential areas of the east and south east of the city.

Although the spatial and temporal pattern of journeys made by men has changed, a number the differences previously identified remain. Women are underrepresented amongst the high RF scores, amongst the faster travel time *z-scores*, and for those who apparently commute, amongst the morning, rather than evening peaks. Women are also less likely to make journeys inter-peak: journeys taken within the working day are less common for this subset of female members. Importantly, when selecting on these inter-peak journeys we see substantial differences between the *types* of journeys being made. For men, many inter-peak journeys take place within Hyde Park, but we also find a diverse set of very short journeys in other parts of the city. Some of these flows suggest leisure cycles, with various journeys from east to west along the popular south side of the river - the Southbank area - easily identifiable. Others appear more utilitarian in nature, with many short journeys made within central and the City of London. For female members, however, these inter-peak journeys are highly spatially concentrated in the more leafy parts of the city - around west London and Hyde Park - and we speculate that very few might be regarded as utilitarian.

By analysing the same geodemographic subset of male and female members, then, men's journeys become far less predictable. Women's journeys appear, if anything, appear to be more regular than do men's and, importantly, this regularity can be seen in the spatial patterns of women's journeys. Altering the way we choose to represent journeys within our map view enables us to better articulate this point. Figure 5 shows all journeys taken by male (top) and female (bottom) members living less than 5km from a docking station. Rather than colouring the flow lines by journey frequency, however, we colour and order journeys according to the number of unique members making them. In order to make a fairer comparison between male and female users, we control by level of usage, and select members only within the top RF segments.

Comparing the two graphics in Figure 5, we again find that this sample of nevertheless frequent female scheme users select very particular parts of the city. Journeys within west London and Hyde Park dominate the map view; we see relatively few members completing journeys within the City of London, involving hub stations (just 51% of women but 81% of men); and journeys that require crossing the river are again underrepresented (62% of women and 67% of men). For men, while most members have made journeys within Hyde Park, we find that many also make journeys across central London, the City of London and further east (Figure 5). Remember, we are weighting flow lines not according to the total number of journeys being made, but by the number of unique members making those journeys. We are also controlling for geodemographics, as well as how often and recently members apparently use the scheme.

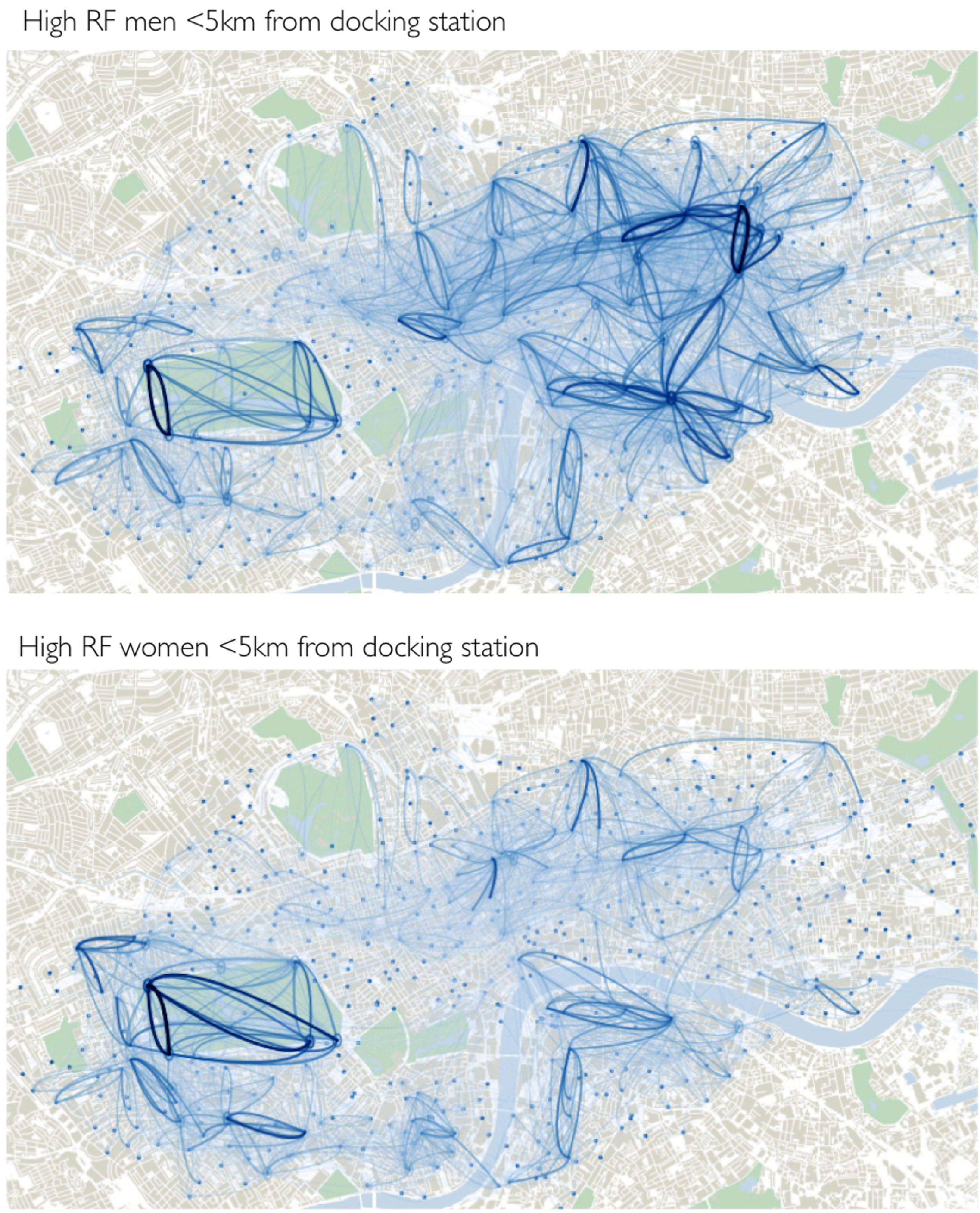

Figure 5: High RF men and women living <5km from a docking station. Flows are coloured by the number of unique members making them.

6 Discussion

This study has identified distinct cycling behaviours amongst male and female members, and related these different behaviours to their spatial and temporal context, and to the personal characteristics of the members making them. Exploring over 10 million journeys made by more than 135,000 members, we feel the research has implications for transport policy-makers and also for the wider study of gender and cycling behaviour.

Our first main finding was that scheme usage amongst men is highly regular, suggesting a strong commuter function, whilst leisure orientated journeys appear to be more dominant for female members. Although some of this variation can be explained by the different geodemographic characteristics of the member population, that comparatively few women are found within a substantial group of non-London-resident commuting members is instructive. Elsewhere, survey based studies into claimed cycle behaviours have found women are generally less likely than men to cycle for commuting purposes (Heesch, Sahlqvist, and Garrard 2012). This has also been confirmed by detailed analysis of observed behaviours, albeit based on a much smaller dataset (Dill and Gliebe 2008). Studying these findings alongside Census travel-to-work data, it might be possible to speculate further about why women are underrepresented amongst this group of regular users.

Secondly, even after controlling for variations in the geodemographic and behavioural characteristics of LCHS members, many important differences can be identified. Women are consistently overrepresented amongst the least heavy LCHS users; are routinely underrepresented amongst the faster travel time *z-scores*; and the temporal structure of women's journeys suggests slightly greater levels of cycling at weekends. These findings again appear consistent with related research, which finds that women generally cycle at slower speeds than men (Dill and Gliebe 2008) and express a strong preference for recreational cycling (Heesch, Sahlqvist, and Garrard 2012).

A more substantial insight is that women's journeys appear to be highly spatially structured. Since female members are more likely to make weekend journeys, it is not surprising that west London and Hyde Park, relatively leafy parts of the city, dominate when querying their journeys. However, there is a sense that women preferentially select very particular parts of the city, even when making utilitarian trips. Journeys around the Bloomsbury area, where roads are narrower, a number of traffic calming measures have been introduced, and cycle lanes slightly offset from major roads, are amongst the most common utilitarian journeys made by women; and journeys between docking stations either side of the river, that generally involve relatively large, multi-lane roads and busy junctions, are comparatively rare.

We should nevertheless be cautious when reporting, and particularly generalising, these findings. Not enough is known about the specific circumstances that underpin our member population. It may be the case that a high proportion of female members are students, with the focus in Bloomsbury indicative of the fact the area contains several large universities. Moreover, the distinct spatial differences may also reflect the geography of employment opportunities for men and women in the capital. Data modelled from 2011 employment figures show that men fill the majority of all jobs (67%) located in the City of London, where we see comparatively few journeys made by female members, whilst in Kensington & Chelsea, and where female customers' journeys are dominant, only 46% of jobs are filled by men (Greater London Authority 2011). Whilst these findings are not simply a function of motivations or barriers, then, various survey (Tilahun, Levinson, and Krizek 2007) and observation-based studies (Dill and Gliebe 2008), including a census of cyclists (Garrard, Rose, and Lo 2008), have found a preference amongst female commuter cyclists for low traffic streets, with routes offering the maximum separation from motorised traffic found to be a particularly high priority (Garrard, Rose, and Lo 2008). Our exploratory analysis appears to support this, but we can also add two new discoveries here: that female commuters using the LCHS are more likely than men to make commuting journeys in the morning than the evening peaks, and that female members are less likely than men to make utilitarian journeys during the working day, irrespective of how often they use the scheme.

Explaining these characteristics at an individual level is more problematic. Related studies have found that normative attitudes, social influence, as well as subjective fears or anxieties around cycling variously affect behaviour (Emond, Tang, and Handy 2009). Additionally, some of the patterns we observe may be an artefact of shared cycling schemes themselves, rather than of broader cycle behaviour. As an example, the LCHS has been designed such that short journeys are encouraged – the first 30 minutes of a journey is free – and we rarely see journeys extend beyond this threshold. By conducting survey research on a sample of LCHS customers and relating claimed attitudes to actual cycling behaviours, it may be possible to attempt at greater explanation. This would, however, inevitably reduce our sample size, making detailed but important spatial-temporal queries challenging.

Since the theme of route choice is substantial, with obvious implications for the provision of cycling infrastructure, a more detailed spatial analysis of men's and women's journeys would also be desirable. For example, using cycling routing algorithms, it may be possible to identify the likely path that members cycled. Although there would necessarily be concerns around measurement validity using such an approach, investigating these routes at a scheme-wide level would enable more confirmatory findings around the nature of men's and women's journeys: the busyness of routes, their elevation and their challenge in terms of negotiating junctions.

Finally, we have yet to explore the extent to which bad weather, failure in other shared transport systems, changes to the LCHS's access policy and the scheme's recent expansion, might affect individual members' behaviours. A further research priority, then, will be to analyse LCHS usage within a wider set of contextual data, and to study the extent to which factors both internal and external to the scheme might variously influence behaviour.

7 Conclusion

In this research, we explore the cycling behaviours of a large population of urban bike share members. Our approach to analysis - designing tailored interactive graphics to identify the spatio-temporal context of cyclists' journeys - may be of relevance to others interested in analysing similarly large behavioural datasets. That the project's findings are highly consistent with existing and recent empirical research into gender and cycle behaviour, carried out on very different and smaller datasets, is highly instructive. We have a large, empirical basis for observing a preference amongst female cyclists, also identified in smaller-scale studies, for leisure-oriented cycling made around comparatively slow and low traffic routes. We also add two new discoveries that may be true outside of our dataset: that female commuters are more likely than men to make commuting journeys in the morning than the evening peaks, and that female LCHS users are less likely than male LCHS users to make utilitarian journeys during the working day. Although the research ambitions for this project were initially exploratory in nature, and notwithstanding some concerns around conflating bike share usage with wider cycling, we provide sufficiently detailed findings to argue that analysis of such large behavioural datasets may significantly advance the study of cycle behaviour.

Word count: 5,724; 148 abstract.

Acknowledgements

We are grateful to colleagues at Transport for London for enabling access to the dataset, supporting research questions and discussing research findings.

References

Blythe, P., and H. Bryan. 2007. “Understanding Behaviour through Smartcard Data Analysis.” *Proceedings of the ICE - Transport* 160 (4): 173–177. doi:10.1680/tran.2007.160.4.173.

Bonham, J., and A. Wilson. 2012. “Bicycling and the Life Course: The Start-Stop-Start Experiences of Women Cycling.” *International Journal of Sustainable Transportation* 6 (4): 195–213. doi:10.1080/15568318.2011.585219.

Borgnat, P., P. Abry, P. Flandrin, C. Robardet, J-B. Rouquier, and E. Fleury. 2011. “Shared Bicycles in a City: a Signal Processing and Data Analysis Perspective.” *Advances in Complex Systems* 14: 415-439. doi:10.1142/S0219525911002950.

Department for Communities and Local Government. 2011. “The English Indices of Deprivation 2010: Technical Report”. Publication (Reports and summaries).

Dill, J. 2006. “Evaluating a New Urbanist Neighborhood.” *Berkeley Planning Journal* 19 (1): 59-78.

Dill, J., and J. Gliebe. 2008. “Understanding and Measuring Bicycling Behavior: A Focus on Travel Time and Route Choice.” *Bicycling* 29 (December): 1 – 70. doi:10.1177/1091581810385956.

Dykes, J. 1997. “Exploring Spatial Data Representation with Dynamic Graphics.” *Computers & Geosciences* 23 (4): 345–370. doi:10.1016/S0098-3004(97)00009-5.

Emond, C. R., W. Tang, and S. L. Handy. 2009. “Explaining Gender Difference in Bicycling Behavior.” *Transportation Research Record: Journal of the Transportation Research Board* 2125 (1): 16–25. doi:10.3141/2125-03.

Field, A.. 2009. *Discovering Statistics Using SPSS*. Third Edition. London: Sage.

Froehlich, J., J. Neumann, and N. Oliver. 2008. “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*, 16–20. Raleigh, North Carolina, USA.

Garrard, J., S. Handy, and J. Dill. 2012. “Women and Cycling.” In *City Cycling*, edited by J. Pucher and R. Buehler, 211–235. London: MIT Press.

Garrard, J., G. Rose, and S. Kai Lo. 2008. “Promoting Transportation Cycling for Women: The Role of Bicycle Infrastructure.” *Preventive Medicine* 46 (1): 55–59. doi:10.1016/j.ypmed.2007.07.010.

Greater London Authority. 2011. “Workplace Employment by Sex and Status, Borough.” Accessed Jan 23 2013. http://data.london.gov.uk/datastore/package/workplace-employment-sex-and-status-borough.

Harrower, M., and C. Brewer. 2003. “ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps.” *The Cartographic Journal* 40 (1): 27–37. doi:10.1179/000870403235002042.

Heesch, K. C., S. Sahlqvist, and J. Garrard. 2012. “Gender Differences in Recreational and Transport Cycling: a Cross-sectional Mixed-methods Comparison of Cycling Patterns, Motivators, and Constraints.” *International Journal of Behavioral Nutrition and Physical Activity* 9 (1): 106-118. doi:10.1186/1479-5868-9-106.

Jensen, P., J-B. Rouquier, N. Ovtracht, and C. Robardet. 2010. “Characterizing the Speed and Paths of Shared Bicycles in Lyon.” *Transportation Research Part D:* Transport and Environment 15(8):522 - 524, doi: 10.1016/j.trd.2010.07.002.

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

Kusakabe, T., T. Iryo, and Y. Asakura. 2010. “Estimation Method for Railway Passengers’ Train Choice Behavior with Smart Card Transaction Data.” *Transportation* 37 (5): 731–749. doi:10.1007/s11116-010-9290-0.

Lathia, N., S. Ahmed, and L. Capra. 2012. “Measuring the Impact of Opening the London Shared Bicycle Scheme to Casual Users.” *Transportation Research Part C: Emerging Technologies* 22: 88–102.

Novo, J. 2004. *Drilling Down: Turning Customer Data into Profits with a Spreadsheet.* Third Edition. Florida, USA: Jim Novo.

Páez, A., M. Trépanier, and C. Morency. 2011. “Geodemographic Analysis and the Identification of Potential Business Partnerships Enabled by Transit Smart Cards.” *Transportation Research Part A: Policy and Practice* 45 (7): 640–652. doi:10.1016/j.tra.2011.04.002.

Pucher, J., and R. Buehler. 2012. *City Cycling*. London: MIT Press.

Reed, W. J. 2001. “The Pareto, Zipf and Other Power Laws.” *Economics Letters* 74 (1): 15–19. doi:10.1016/S0165-1765(01)00524-9.

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

Roberts, J.C. 2005. “Exploratory Visualization with Multiple Linked Views.” In *Exploring Geovisualization*, edited by J. Dykes, A. M. MacEachren, and M-J. Kraak, 159–180. Oxford: Elsevier.

Tilahun, N.Y., D.M. Levinson, and K.J. Krizek. 2007. “Trails, Lanes, or Traffic: Valuing Bicycle Facilities with an Adaptive Stated Preference Survey.” *Transportation Research Part A: Policy and Practice* 41 (4): 287–301. doi:10.1016/j.tra.2006.09.007.

Wood, J., A. Slingsby, and J. Dykes. 2011. “Visualizing the Dynamics of London’s Bicycle Hire Scheme.” *Cartographica* 46 (4): 239 – 251.

Figure 1: ‘Journeys’ and augmented ‘Customers’ datasets.

Figure 2: Screenshots of the application. Above: journeys around Liverpool Street Station and Holborn, and between particularly Holborn and London’s major commuting rail terminals, King’s Cross and Waterloo, are visually salient when all journeys are selected. Below: journeys made by women within the first 3 months of the scheme’s launch are selected, showing that trips within Hyde Park. Colours are taken from the Brewer ‘Blues’ sequential colour scheme (Harrower and Brewer 2003).

Figure 3: Rank-size distribution of 1,000 most commonly made journeys for male and female members.

Figure 4: Top 100 journey pairs made between September 2011-2012 by male (above) and female (below) members. Docking stations within Bloomsbury area are highlighted.

Figure 5: High RF men and women living <5km from a docking station. Flows are coloured by the number of unique members making them.