Revealing and Informing Transport Behaviour from Bicycle Sharing Systems

James Cheshire[1], Oliver O'Brien[1]

[1]UCL Centre for Advanced Spatial Analysis, UCL, Gower Street, London, WC1E 6BT
Tel. +44 (0) 20 3108 3914
o.obrien@ucl.ac.uk, james.cheshire@ucl.ac.uk

KEYWORDS: bike-sharing; cycling; cities; health; smart transport

1. Introduction

Bicycle sharing systems (bike-shares) are becoming increasingly popular in towns and cities around the world. They offer a cheap, efficient, and healthy means of navigating dense urban environments. This presentation draws from the analysis of data from 38 systems located in Europe, the Middle East, Asia, Australasia and the Americas. The work has required the creation of an extensive database containing the geographical location and occupancy of each docking station within a particular system over a number of months to chart their usage and provide a consistent basis on which to compare and classify them. Analysis of the variation of cycle docking station occupancy rates over time, and comparison across the system's extent, infers the likely demographics and intentions of user groups. A classification of bike-shares, based on the geographical footprint and diurnal, day-of-week and spatial variations in occupancy rates is presented and offers insights into the general characteristics of system usage. In addition, some system operators release individual journey information that can be used to infer specific travel behaviours. The knowledge of such patterns and characteristics identifiable from the dataset has a range of both practical and research applications, including informing the maintenance of a suitable balance of bikes throughout the system area (a nontrivial problem for many bike-shares), the location of new docking stations and cycle lanes, and better targeting of promotional materials to encourage new users. We demonstrate this with a case study that combines the journey information with pollution data to identify areas of high usage and high likely exposure. Such work has been used as inputs to health impact models with a view to informing travel behaviours. As this presentation will seek to demonstrate, bike-shares offer a wealth of insights into transport behaviours at the international and urban scales.

There are approximately 450 bike-shares operating around the world (DeMaio and Meddin 2012) and analysing the relatively freely available data generates insights into the habits of their users and, by proxy, movements within urban areas (Padgham 2012). The schemes are often single systems located in and around the commercial or business centres of their host towns and cities. Exceptions include suburb-based systems such as Mexico City (Kazis 2012) or Taipei (Tso 2009), or those extending well beyond the city core, as is the case for Barcelona. Current bike-shares make use of technology to operate on a largely automated basis and it is this that generates the data to be analysed and also enables users to monitor cycle availability and docking station spaces via near real-time online maps. These websites often specify and supply an applications programming interface (API) for external software developers to access the underlying data. In addition, a number of system operators release datasets pertaining to individual journeys made over a particular time period. As will be demonstrated, both types of data offer insights in the usage of particular bike-shares and provide a ready basis for utilisation in transport research.

2. Characterising Global Bike-share Usage

The data used here were collected from operator-run websites and include locations, capacity and current load factor of docking stations. A script, written in the Python programming language, and customised for each system, is run on a regular basis to access the bike-share's docking station data online. The load factor is the proportion of docks in each docking station that currently have a bicycle available to hire. It is normally calculated from the number of bicycles and the number of free spaces in each docking station. Systems that do not make this information available online – a key metric for users trying to discover bicycles or free spaces in their vicinity – are not included in this study.

The data can be typically collected every two minutes, except where the system's server is slow to respond, in which case the data is collected every 10 or 20 minutes. This frequency is sufficient to accurately show the activity and availability changes throughout the day, highlighting commuter "rush hours" and other features. The database covers a period of up to 2 years and over 80 cities. For this study we focus on data collected throughout September 2012.

Of the 38 bike-shares included here, 16 fall in Europe & the Middle East, 11 in Asia, 9 in the Americas and 2 in Australasia. Looking at the relative locations of docking stations in each system, and the diurnal and weekly variations in the aggregated load factor a number of characteristics can be compared and contrasted. Temporal characteristics provide our focus and include the variations in load factor on an intraday and weekday/weekend basis. The measures are obtained by regularly counting up the number of full and empty docking points, within each docking station. For the load factor measure, these are simply aggregated across the system.

A load factor of 45-50% seems typical of bike-shares, with Europe's average of 45% being noticeably lower, and America's average of 50% being higher, than the worldwide average of 48%. The London system, for example, was specified with a docking point ratio of 1.7:1, equating to an unusually high maximum load factor of 59% (Transport for London 2009, p10) and launched with a load factor of 54%. It then gradually reduced it to around 50% as the overall system size increased over the first year of operation – the number of bicycles in the system being in fact reduced slightly. During September 2012 it was 45%, with some sign of a slight increase in bikes available during the wintertime, viable because of lower usage, and a corresponding decrease in summertime.

Figure 1 highlights obvious differences between systems, but it is also remarkable that many the systems show common traits, despite substantial differences in geographical footprint, system size and density. In particular, a double-trough weekday usage and a wide single-trough weekend usage is a characteristic shared in a substantial number of the systems we have studied here.

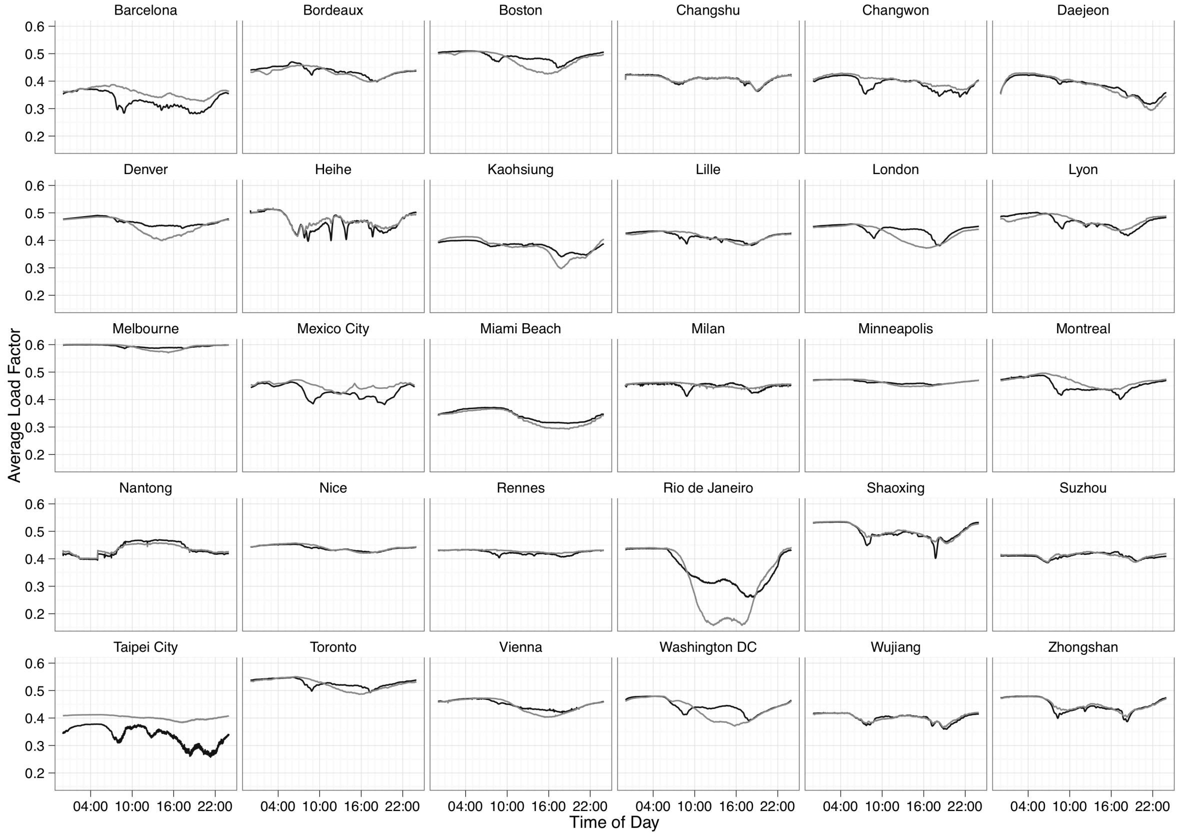

Figure 1: Load factor variation across 24 hours, for weekdays (black lines) and weekend days (grey lines), the data being averaged across April-September 2012, for 30 of the 38 cities in this study. For two systems here (Melbourne and Rio) this was across wintertime. Some data cleaning was employed to remove fluctuations caused by unplanned outages and collecting errors. Denver's system is closed from midnight to 6am. Taipei City's data is from September 2012 only and shows some noise due to the smaller number of days it is averaged across. The time of day is local to each system.

3. Cycle Hire and Pollution

In the final section of this abstract we demonstrate the utility of combining the individual-level journey information with other spatial datasets to better understand transport behaviours and their impacts. To date only a few systems, such as London's, offer historical journey data and these are normally provided by the operator on a bulk-load basis, rather than being queryable from an API or map. These data contain the origins and destinations of each cycle hire journey and their start and end times. From this information we can derive the probable routes taken by each cyclist using the Routino (www.routino.org) routing software. The software can be calibrated to favour cycle friendly routes between docking stations to reflect the fact that cyclists tend to avoid busy main roads in favour of quieter roads with cycle lanes, factors such as cyclist-only turnings are also considered. Once calculated the routes can be spatially joined to gridded pollution data, in this case particulate matter (PM 2.5) values for London, to indicate the likely exposures along each route. This is shown in Figure 2 below.

Figure 2: Modelled cycle hire journeys combined with PM 2.5 values. Thicker lines represent more cycle journeys and redder lines represent higher pollution values.

The map in Figure 2 shows that many high pollution areas, such as the junctions to the north of many of the bridges over the Thames, remain unavoidable for cyclists leading to relatively high exposure values for large numbers of them. Other common routes, however, such as within Hyde Park have relatively low levels of exposure. Such routes are popular with leisure users who favour scenic journeys on cycle only paths over the busy roads endured by commuters using the bikes to get from major stations, such as Waterloo and Kings Cross, to their workplaces.

4. Conclusions

In conclusion, the above abstract and proposed presentation seek to demonstrate the insights offered by the straightforward collection of bicycle sharing system data. This work stands to benefit operators, researchers of urban behaviours and patterns, and users themselves. The usage of such systems is quite well ordered, with a relatively high degree of predictability. With the creation of new systems and increased public availability of individual level origin-destination data for some systems (currently London, Washington DC, Minneapolis and Boston), the opportunities and applications of studying spatial, temporal and journey data associated with bike-sharing will continue to expand. Bicycle sharing systems play an important part in increasing sustainable transport options in cities, and understanding their potential across many diverse types of cities, and multiple user types, is will increase in importance.

5. Software

All the data collection, storage and analysis for this project were undertaken with Open Source Software and data. Accessing the data feeds was done with *Python* and the data were loaded into a *PostgreSQL* database. The routing utilised *OpenStreetMap* for its network and *Routino* for the analysis. *PostGIS* was used for spatial joins and *R* was used for additional calculations and visualisation.

6. References

DeMaio P and Meddin R (2012) *The Bike-sharing World Map*. http://bit.ly/K9pKmO

Padgham, M. (2012) Human Movement is Both Diffusive and Directed. *PLoS ONE* 7(5): e37754. doi:10.1371/journal.pone.0037754

Kazis N (2012) With a Boost From Bike-Share, Cycling Surges on Mexico City's Mean Streets. http://www.streetsblog.org/2012/03/22/with-a-boost-from-bike-share-cycling-surges-on-mexico-citys-mean-streets/

Tso N (2009) Taiwan Goes Green with Bike Sharing. http://www.time.com/time/world/article/0,8599,1898152,00.html

Biography

Oliver O'Brien is a researcher and software developer at UCL's Centre for Advanced Spatial Analysis (CASA). He specialises in web cartography and creating spatial data visualisations. He recently launched CityDashboard, an at-a-glance aggregator of sensor, social and official data sources for cities, and co-edits MappingLondon.co.uk.

James Cheshire is a lecturer at the UCL Centre for Advanced Spatial Analysis. His interests lie in the spatial analysis and visualisation of large urban datasets relating to populations and transport. James features work on spatialanalysis.co.uk and also as co-edits MappingLondon.co.uk.