

ONS Methodology Working Paper Series No 5

Comparing travel flows between 2011 Census and Oyster card data

Karen Gask and Susan Williams
June 2015

1. Introduction

The Oyster card is a form of electronic ticketing used on public transport in Greater London. It is promoted by Transport for London (TfL) and is valid on travel modes across London including tubes, buses, Docklands Light Railway, overground trains and trams.

Usage is encouraged by offering substantially cheaper fares in comparison to cash purchases. Passengers touch the card on an electronic reader when entering and leaving the transport system in order to validate it or deduct funds. Therefore the Oyster card system is capable of recording journeys and monitoring the movement of individual Oyster cards over time.

ONS has compared Oyster card data from tube travel in London to travel flows from the 2011 Census to understand their similarities and any limitations in using Oyster card data for travel analysis. This report describes the analysis undertaken.

2. Data

2.1. Oyster card data

TfL has [published](#) Oyster card data on the counts of journeys from an origin tube station (where an Oyster card first enters the network) to a destination tube station (where the same Oyster card leaves the network). The flows selected for comparison with Census data represent the journeys made using an Oyster card on a typical weekday during the peak travel time of 7am to 10am in November 2012. The data is adjusted by TfL to remove any abnormal circumstances that may affect demand such as industrial action or long-term closures. Table 1 shows the first five rows of the data obtained.

Table 1: First five rows of raw Oyster card data

From	To	Early	AM peak	Midday	PM Peak	Evening	Late	Weekday
Start tube name	End tube name	- 7am	7am- 10am	10am- 4pm	4pm- 7pm	7pm- 10pm	10pm+	Total
Acton Town	Alperton	6	43	230	46	0	0	325
Acton Town	Angel	0	0	6	0	6	0	12
Acton Town	Arsenal	0	0	19	0	0	0	19
Acton Town	Baker Street	0	11	0	0	1	1	13
Acton Town	Bank / Monument	15	115	0	114	0	0	244

To enable comparison with Census geographies, each tube station was allocated to a local authority (LA) and a Middle Layer Super Output Area (MSOA, areas containing 5,000-15,000 people) and the corresponding Oyster card flows were aggregated.

2.2. Census data

The 2011 Census asked:

How do you usually travel to work?

Tick one box only.

Tick the box for the longest part, by distance, of your usual journey to work.

- Work mainly at or from home
- Underground, metro, light rail, tram
- Train
- Bus, minibus or coach
- Taxi
- Motorcycle, scooter or moped
- Driving a car or van
- Passenger in a car or van
- Bicycle
- On foot
- Other

The table [WU03UK](#) from the 2011 Census shows the location of usual residence and place of work for workers aged 16 or over travelling most of their commute by underground, metro, light rail or tram. It is not possible to separate those mainly commuting by underground alone. Only those LAs / MSOAs with a tube station are included in the analysis, and some LAs / MSOAs contain more than one tube station.

3. Analysis

Travel flows obtained from the Oyster card data were compared with those from Census data. While the definitions in the datasets are not the same, the key assumption in the Oyster card data is that journeys starting during the rush hour peak of 7am-10am on weekdays are most likely to represent workers, with the start station most likely to represent the area where they live and the end station most likely to represent the area they work.

3.1. Local authorities

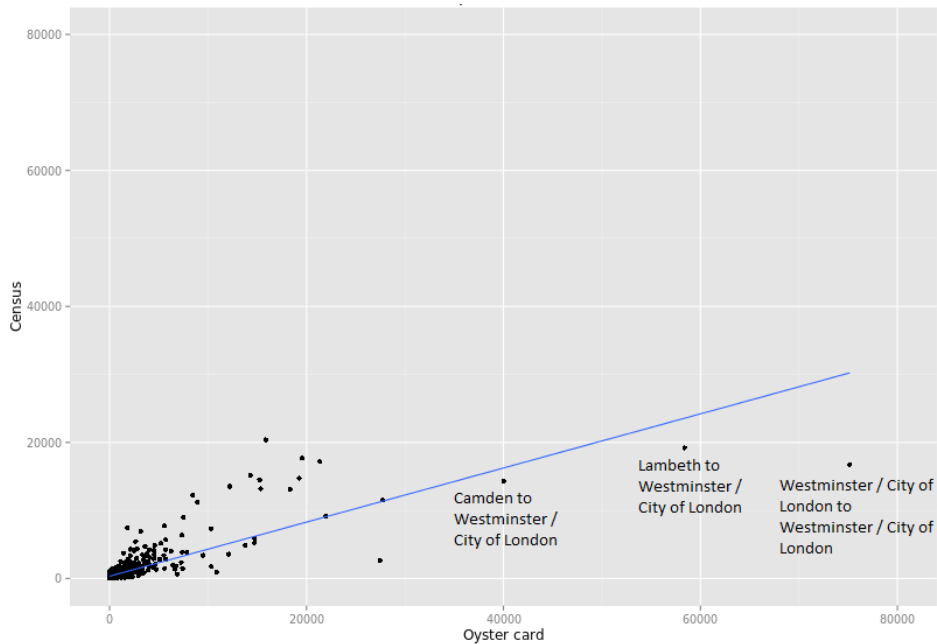
This section provides an overview of the distribution of flows from an origin local authority to a destination local authority using both Census and Oyster card data. Similarities and differences between the data sources are highlighted by the use of scatter plots, deciles and heat maps, and Hackney is used to illustrate how commuters using different modes of transport impact on the analysis.

Scatter plots

Figure 1 compares flows between the Census and Oyster card data, along with the linear regression line which best fits the data. There are 30 LAs containing a tube station, and therefore 900 (30x30) points on the graph in total considering that an LA can be a start or end of a journey. In other words, each point represents a LA to LA flow from Oyster card (on the x-axis) or Census (on the y-axis). The largest flows are labelled. There appears to be a reasonable one to one relationship where Oyster card flows are less than 20,000, but large Oyster card flows appear to dominate the linear regression line and pull it down. For example, 58,000 people travel on the tube from Lambeth to Westminster /

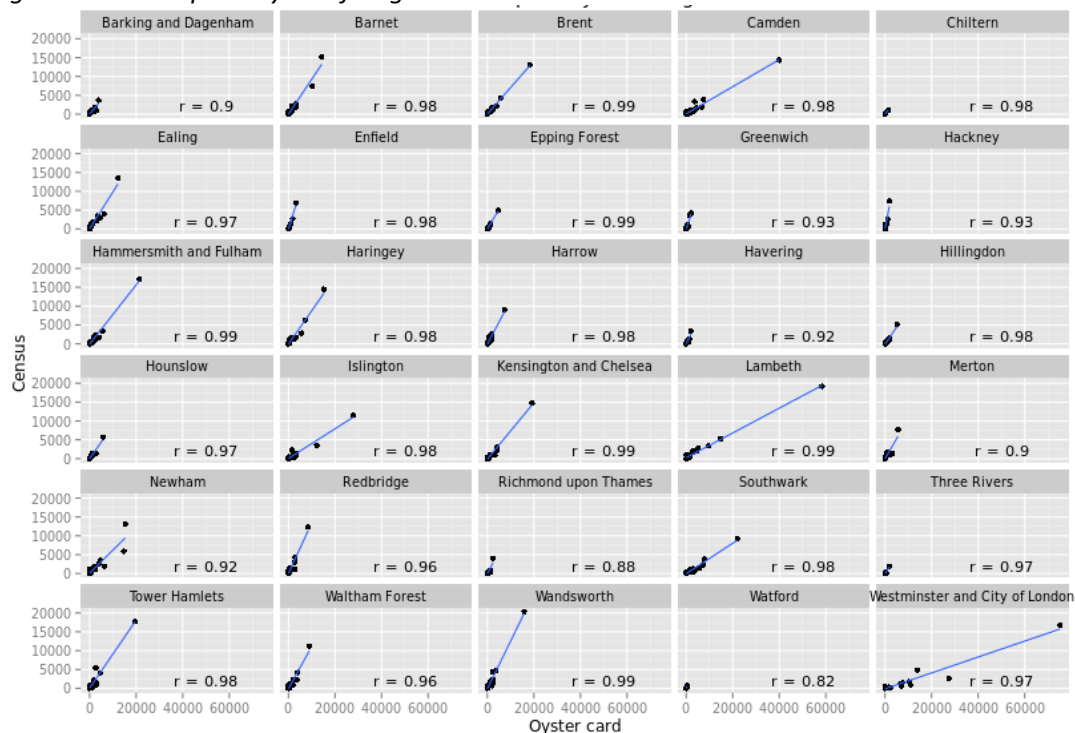
City of London according to Oyster card data compared to only 19,000 tube / light rail commuters according to the 2011 Census.

Figure 1: Scatter plot of Census and Oyster card flows for LAs



Taking each LA as origin in turn, the Oyster card flows from it into each LA were correlated with corresponding Census flows. The correlations generated were generally very good (average of 0.96). However as can be seen in Figure 2, care needs to be taken here as the correlations and regression lines shown tend to be driven by a small number of dominant flows, such as into Westminster / City of London. The axes are the same for each small chart in Figure 2.

Figure 2: Scatter plots by LA of origin



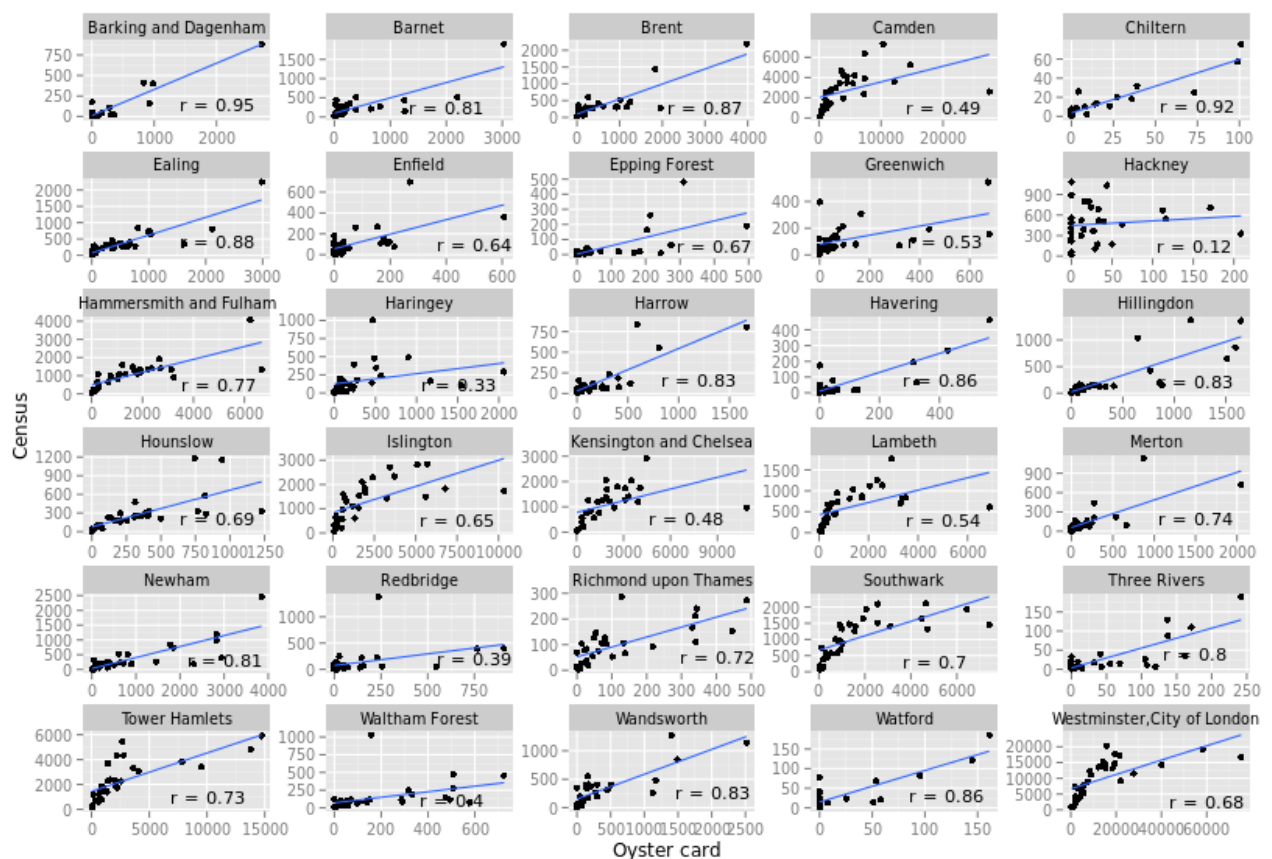
The poorest correlations are in:

- Watford (0.82)
- Richmond upon Thames (0.88)
- Barking and Dagenham (0.90)
- Merton (0.90).

These areas are all in outer London and residents of these areas are more likely to take the car to work than the tube¹. In Watford and Richmond upon Thames, only 5% and 12% of workers respectively take the tube to work. More generally the availability of alternative quicker or cheaper transport, such as buses in south London, is likely to influence travel behaviour.

A different story arises in Figure 3 when taking each LA as a destination in turn and looking at the correlation across the origin LAs. The first point to note is that a small number of LAs have large flows (Westminster and City of London, Camden, Kensington and Chelsea, and Tower Hamlets), while the rest have small flows. The high flows into LAs such as Westminster and City of London mean that it is difficult to illustrate the patterns using the same axes for all charts. As a result, the axes are different for each small chart in Figure 3.

Figure 3: Scatter plots by LA of destination



Looking at the scatter plot for the destination of Westminster and City of London in particular, there are a small number of origin LAs where Oyster card flows are more than 20,000 and are much larger than their corresponding Census flows. These flows have the effect of reducing the correlation between the two sets of data heavily. These same origin LAs have similarly high Oyster card flows into other destination LAs leading to lower correlations by destination than by origin across most LAs (an average of 0.68).

¹ Source: 2011 Census table QS701EW (Method of travel to work)

Deciles

Deciles are created separately in both sets of data by ordering the 900 LA to LA flows from smallest to largest before splitting the sorted data into ten equal groups, each representing 90 flows of increasing magnitude. The largest flow value in each part represents the decile value.

Therefore the first decile value represents the 90th smallest flow in each dataset, the second decile the 180th smallest flow and so on right up to the tenth decile which is the maximum flow overall.

The distribution of the deciles of Oyster card and Census data is compared in Table 2. For example, decile 2 illustrates that 20% of flows in the Census data are made by 23 people or less compared with 8 people or less in Oyster card data.

Table 2: Deciles of 2011 Census data and Oyster card data for LAs

Decile	Census	Oyster card
Decile 1	10	0
Decile 2	23	8
Decile 3	55	31
Decile 4	93	67
Decile 5	155	142
Decile 6	251	264
Decile 7	447	508
Decile 8	877	1,124
Decile 9	1,394	2,566
Decile 10	20,385	75,172

For LAs the distribution of Oyster card flows tends to be similar to Census flows up to around deciles 6 or 7 where they become increasingly large and are skewed with some very high values, especially flows into Westminster and City of London. For example, the Oyster card flow for journeys originating and ending in Westminster / City of London is over 75,000 whilst Census data shows that less than 17,000 people both live and work there (and take the tube for commuting). This can be clearly seen in Figure 1.

Heat maps

Heat maps showing the distribution of flows in both the Census and Oyster card data for LAs are shown in Figure 4 and Figure 5. In both Figures, the format is an origin-destination grid of flows between LAs, with each flow shaded according to its relative magnitude across all the flows: Smallest flows are not coloured and the largest flows are coloured the darkest. The same legend is used for both heat maps for ease of comparison.

The heat maps appear similar and show that commuters living in all areas are more likely to work in destinations towards the bottom of chart; in central London. In fact, both heat maps show that there are seven local authorities to which people travel most:

- Westminster, City of London
- Camden
- Tower Hamlets
- Islington

- Kensington and Chelsea
- Hammersmith and Fulham, and
- Southwark.

One key observation is that flows into and out of Hackney LA are much smaller in Oyster card data than in Census data. This is examined further in the following section.

Figure 4: Heat map of LA travel patterns using Census data, 2011

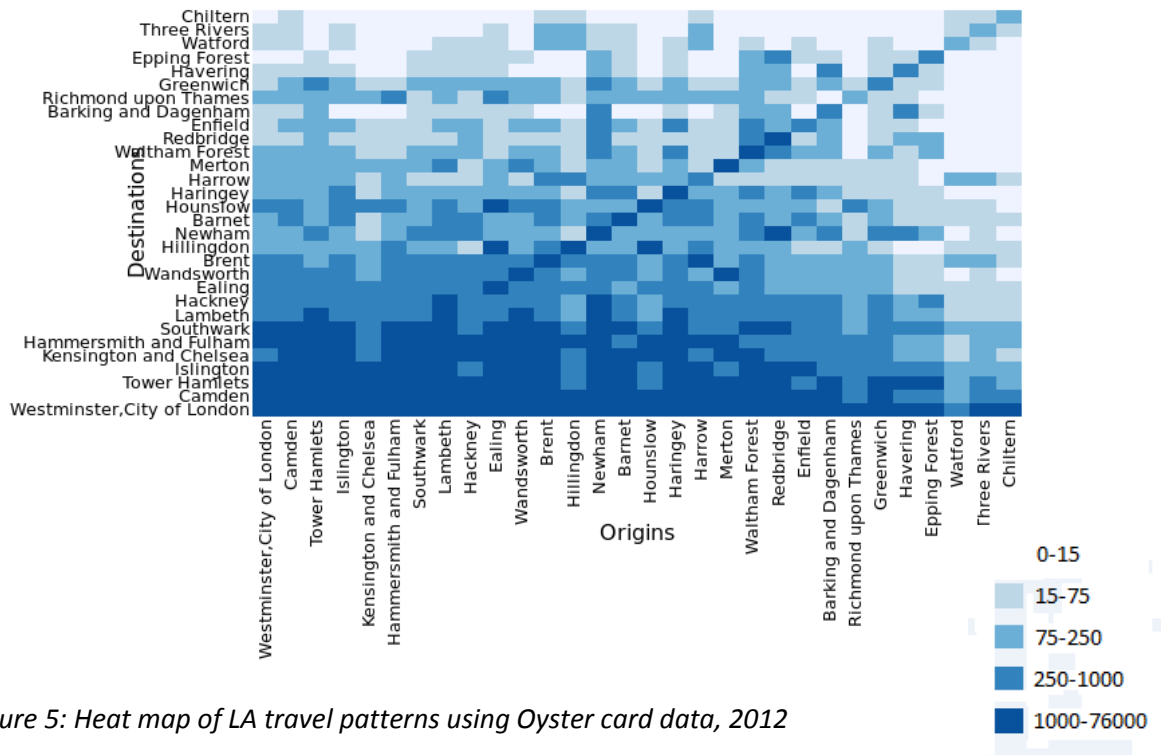
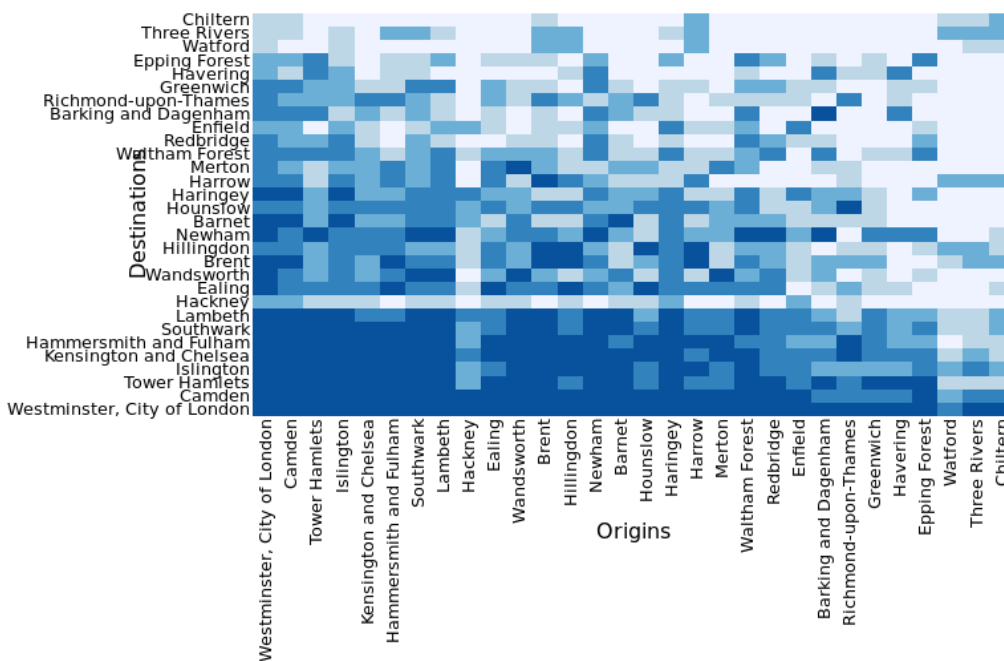


Figure 5: Heat map of LA travel patterns using Oyster card data, 2012



Investigation into Hackney LA

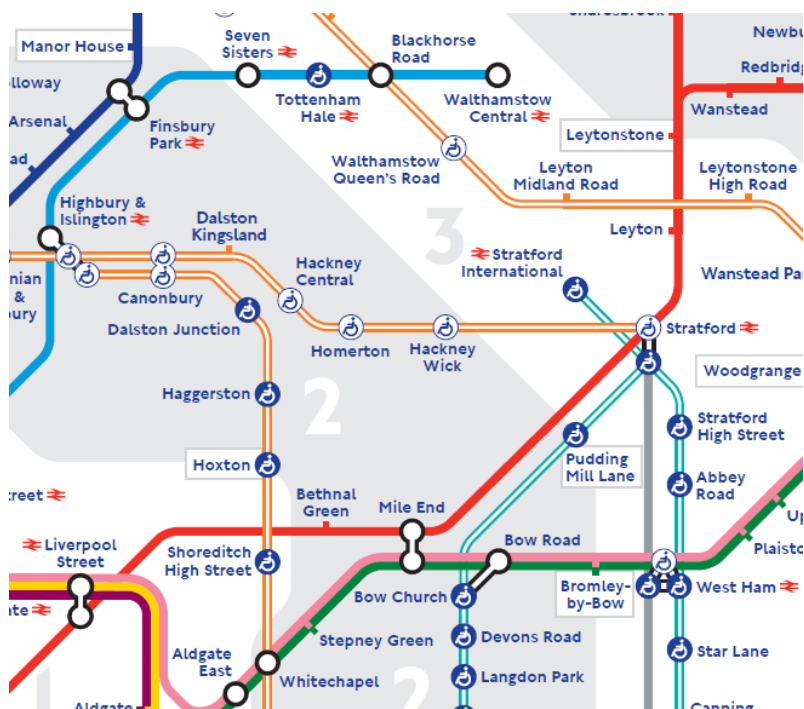
Census data shows that 18,900 Hackney residents mainly take the underground, metro, light rail or tram to work whilst Oyster card data shows that only 5,700 start their underground journey at the only tube station in Hackney, Manor House. Similarly, Census data shows there are 14,100 commuters from other LAs who work in Hackney whereas Oyster card data has only 1,100 tube journeys ending there.

Hackney only contains one tube station but has seven overground train stations and many bus routes. The stations in Hackney are:

- Manor House (tube station)
- Dalston Junction (overground station)
- Dalston Kingsland (overground station)
- Hackney Wick (overground station)
- Hackney Central (overground station)
- Haggerston (overground station)
- Homerton (overground station)
- Hoxton (overground station).

From Hackney, it is a short journey via overground train to connect to the tube network at Whitechapel or Stratford, or to connect to the Docklands Light Railway (DLR) at Stratford. For example, 7,400 residents of Hackney work in Westminster or the City of London according to the Census, but Oyster card data shows that 1,800 people start their tube journey in Hackney and end in Westminster or the City of London. To get to these areas from Hackney, Figure 6 shows that commuters may be more likely to either take the overground train to Stratford, then use the tube's Central line, or take the overground train to Whitechapel and then use the tube's Hammersmith & City or District lines.

Figure 6: Section of tube map showing Hackney



Therefore those living or working in Hackney may commute taking the tube or DLR for the main part of their journey, as well as a short distance via overground train or bus. As a result, they may have said that they predominately take the ‘underground, metro, light rail or tram’ to work on their Census form. This is partly borne out by examining the Census commuter flows out of Hackney: Table 3 shows that only 8% of Hackney residents predominately commute using the train compared with 21% who mainly use the ‘underground, metro, light rail or tram’.

Table 3: Percentage of residents aged 16-74 in employment by method of travel to work in Hackney and London, 2011

Method of travel to work	Hackney	London
Underground, metro, light rail, tram	21	24
Train	8	14
Bus, minibus or coach	28	15
Car or van	13	31
Bicycle or by foot	28	14

Note: Other methods of travel to work have been excluded.

The published Oyster card data only contains journeys which both start and end at a tube station. Any overground or bus journeys, such as might occur regularly into and out of Hackney, would not be included. This might explain why Census data on commuters mainly using the underground is much higher than Oyster card data for flows into and out of Manor House tube station.

Comparing Hackney between both data sources illustrates the difficulties which are observed when commuters take more than one mode of transport to work. This is further illustrated later in the report in areas which contain a mainline train station.

3.2. Middle Layer Super Output Areas

The analysis so far has concentrated on LAs, and now it turns to analysing patterns for Middle Layer Super Output Areas (MSOAs). These geographies are smaller than LAs and contain between 5,000 and 15,000 people.

Similar heat maps to Figure 4 and Figure 5, showing the distribution of flows in both the Census and Oyster card data for MSOAs are now shown in Figure 7 and Figure 8. The same legend is used for both heat maps.

Over 60% of the MSA to MSA flows in Figure 7 and Figure 8 are zero. This is unsurprising as 83% of MSAs contain only one tube station.

Again, the heat maps show that commuters are more likely to work in destinations towards the bottom of chart, in central London. However it is more evident that Oyster card flows are generally much larger than Census flows as the colours displayed are darker. The Census heat map appears to be more graduated with darker counts appearing at the bottom, whereas darker Oyster card counts appear to be more scattered around the heat map.

The diagonal in the heat maps indicate counts of those who both live and work in the same MSA and take the tube to work. The diagonal in the Census heat map appears to be darker than that in the Oyster card map. In the MSAs with only one tube station, Oyster card data naturally records no

people entering and exiting, while Census data records a very small number of people. The Census heat map highlights these MSOAs. Reasons why Census has small numbers here may be attributable to inaccurate responses to the Census question or individuals living in an MSOA whose nearest tube station is in a neighbouring MSOA, and who takes the tube to the underground station at the other end of the MSOA in which they live.

Another explanation for some MSOAs can be illustrated by Tower Hamlets 033, where 74 people lived and worked in the same MSOA and used the tube, metro, light rail or tram to get to work. However no Oyster card journeys were made between tube stations in this area. This is because this MSOA contains only one tube station (Canary Wharf), but four stations for the DLR. Therefore Oyster card counts are zero because there is only one tube station here but Census counts are larger as commuters are likely to have used the DLR rather than the tube.

Figure 7: Heat map of MSOA travel patterns using Census data, 2011

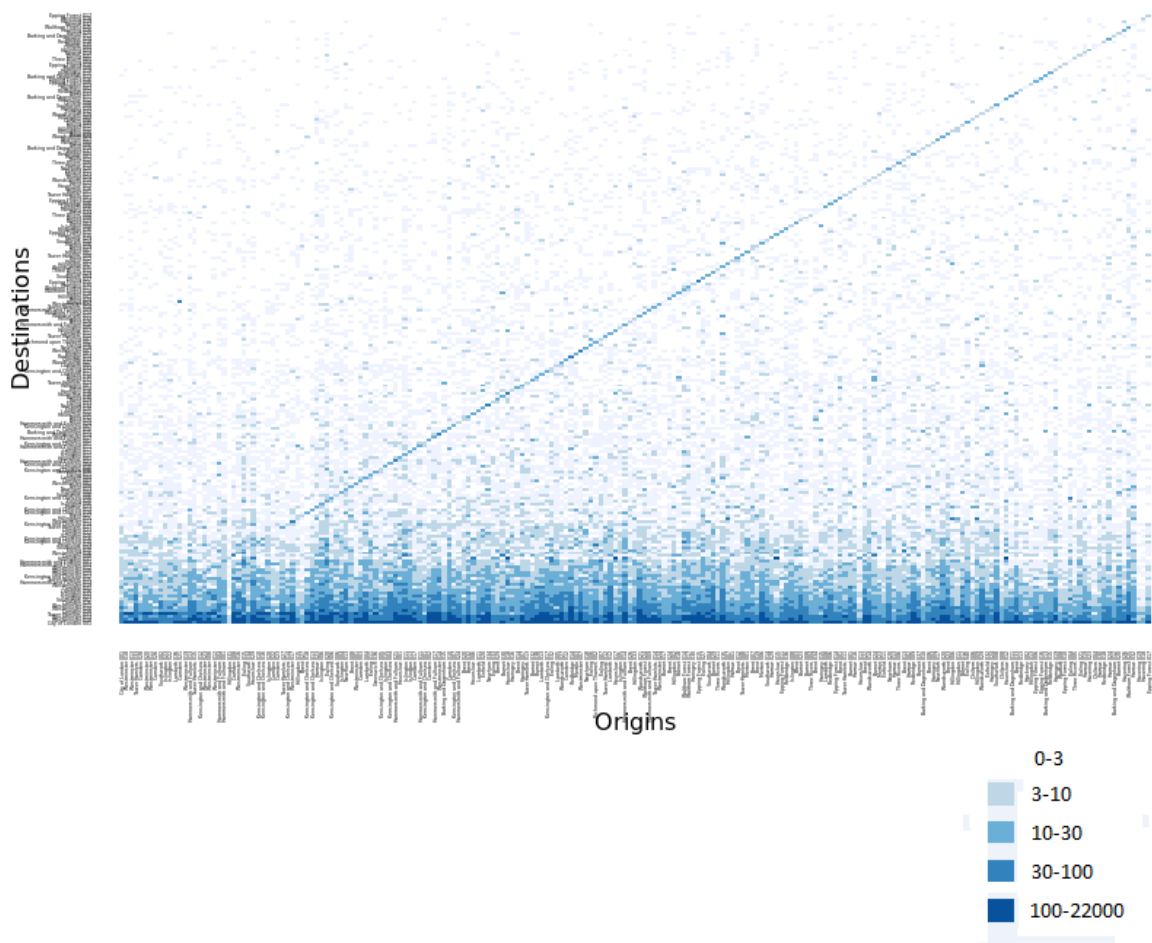
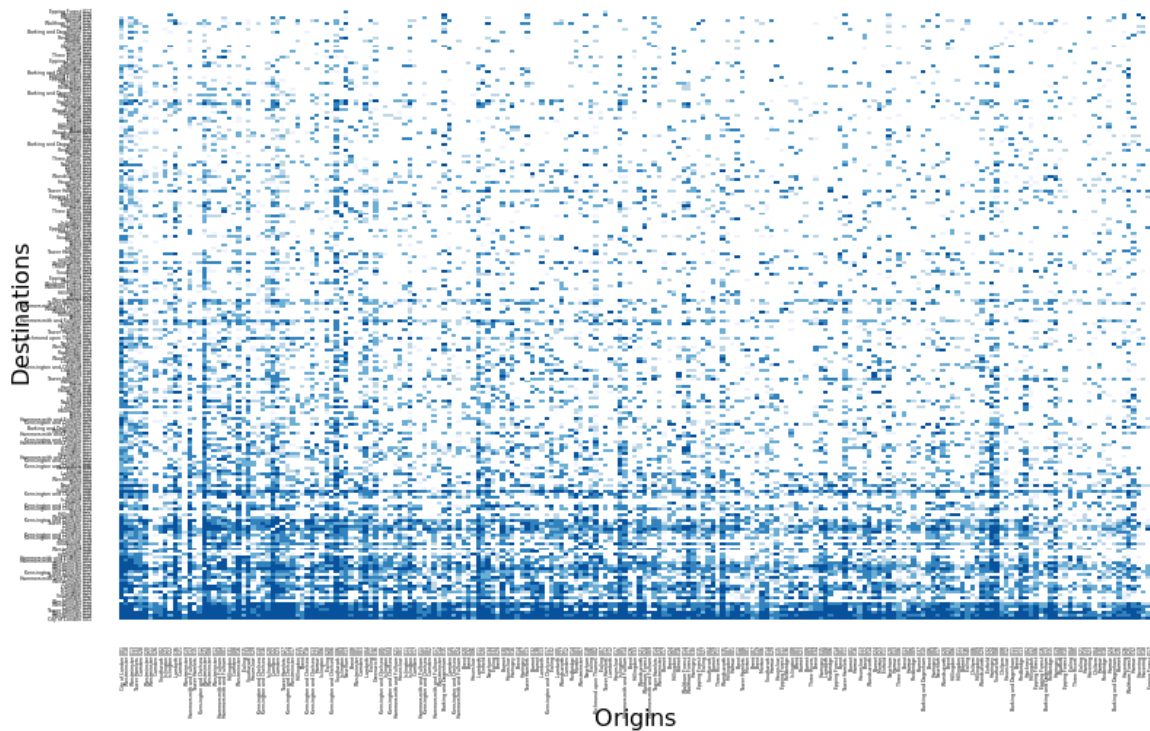


Figure 8: Heat map of MSOA travel patterns using Oyster card data, 2012



A possible explanation for larger flows in Oyster card data than Census data can be illustrated by an example: 611 people live in the MSOA “City of London 001” and take the tube to work (according to the Census), while 55,948 people start their journey on the tube from there according to the Oyster card data. This large difference (a ratio of 1:92) is partly due to Liverpool Street train station being in this MSOA, suggesting that a large number of people may commute from outside London to this station, then use an Oyster card to commute a further distance using the tube.

Table 4 illustrates the ratios for MSOAs containing nine mainline train stations, Heathrow and a summary of other MSOAs. It can clearly be seen that larger ratios are evident in all MSOAs which contain a mainline station, with the largest ratios in the MSOAs which contain Liverpool Street, Waterloo and Kings Cross St Pancras.

Table 4: Ratios between Oyster card (start tube) and Census (residence) data (Oyster card divided by Census), for MSOAs containing a mainline station

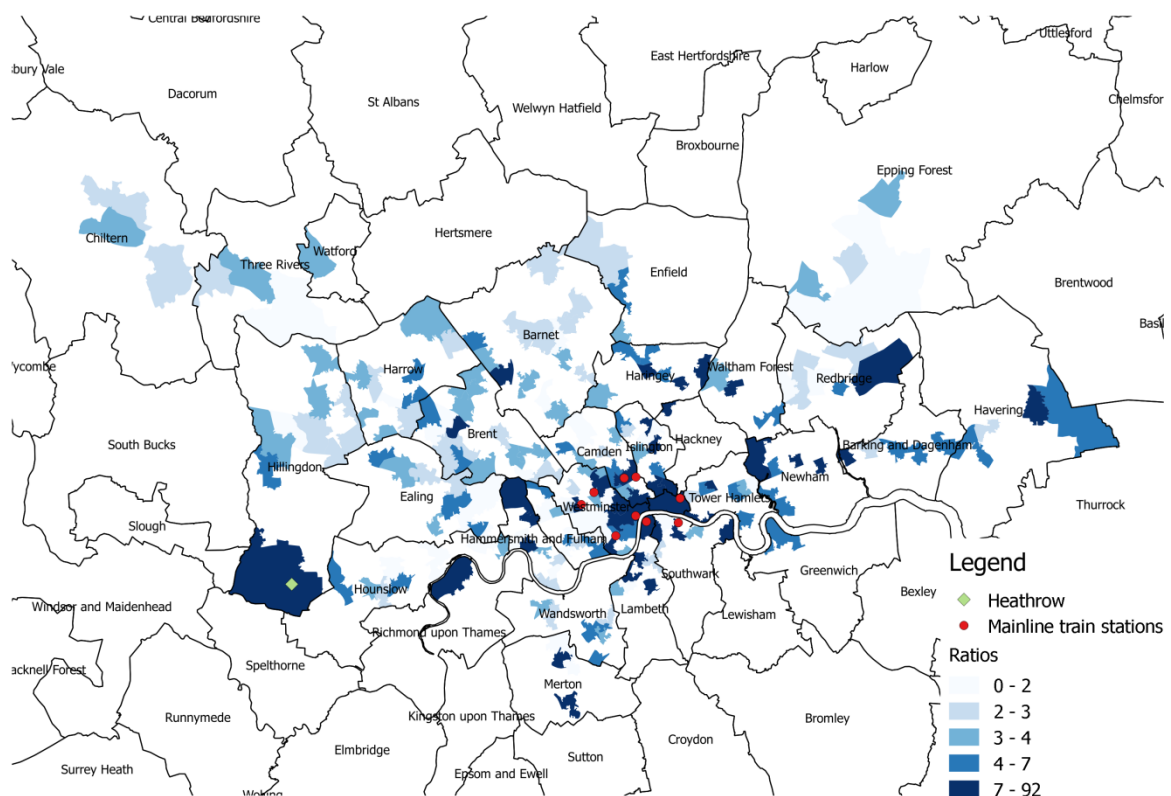
MSOA	Mainline station in this MSOA	Number of people starting at this station (Oyster card)	Number of workers who live here who commute by tube (Census)	Ratio
City of London 001	Liverpool Street	55,948	611	92
Lambeth 036	Waterloo	61,348	720	85
Camden 022	Kings Cross St Pancras	29,125	383	76
Westminster 023	Victoria	32,621	805	41
Camden 023	Euston	16,258	427	38
Southwark 006	London Bridge	30,302	797	38
Westminster 018	Charing Cross	12,182	650	19
Westminster 016	Paddington	24,082	1,716	14
Westminster 008	Marylebone	12,837	1,246	10
Hillingdon 031 (containing Heathrow)		3,142	74	42

All MSOAs containing a mainline train station	274,703	7,355	37
All MSOAs containing a small train station	249,507	28,384	9
All remaining MSOAs	559,992	159,058	4

A large ratio (42) between Oyster card and Census data is observed in Hillingdon 031 (which contains Heathrow), indicating that people fly to Heathrow then use their Oyster card to get into central London from there. Even having a smaller train station in an MSOA appears to have an effect on the ratio generated; on average for all MSOAs containing a small train station there is a ratio of 1:9 between Oyster card and Census data. Again, this indicates that people may take the train or bus for part of their journey before changing to the tube where their Oyster card journey begins.

Figure 9 provides a map of these ratios. In general, higher ratios (dark blue) can be found in central London while lower ratios are more common in outer London. Hillingdon 031 containing Heathrow is the large dark area to the west of London. Further examination of dark blue areas with high ratios indicates that not all contain a train station or airport; some contain a bus terminus, overground train station, large car park, end of a tube line, or boundary of different fare zones. This further indicates that commuters may be using a different mode of transport to get to a tube station, and then start their Oyster card journey from a different MSOA to that in which they live.

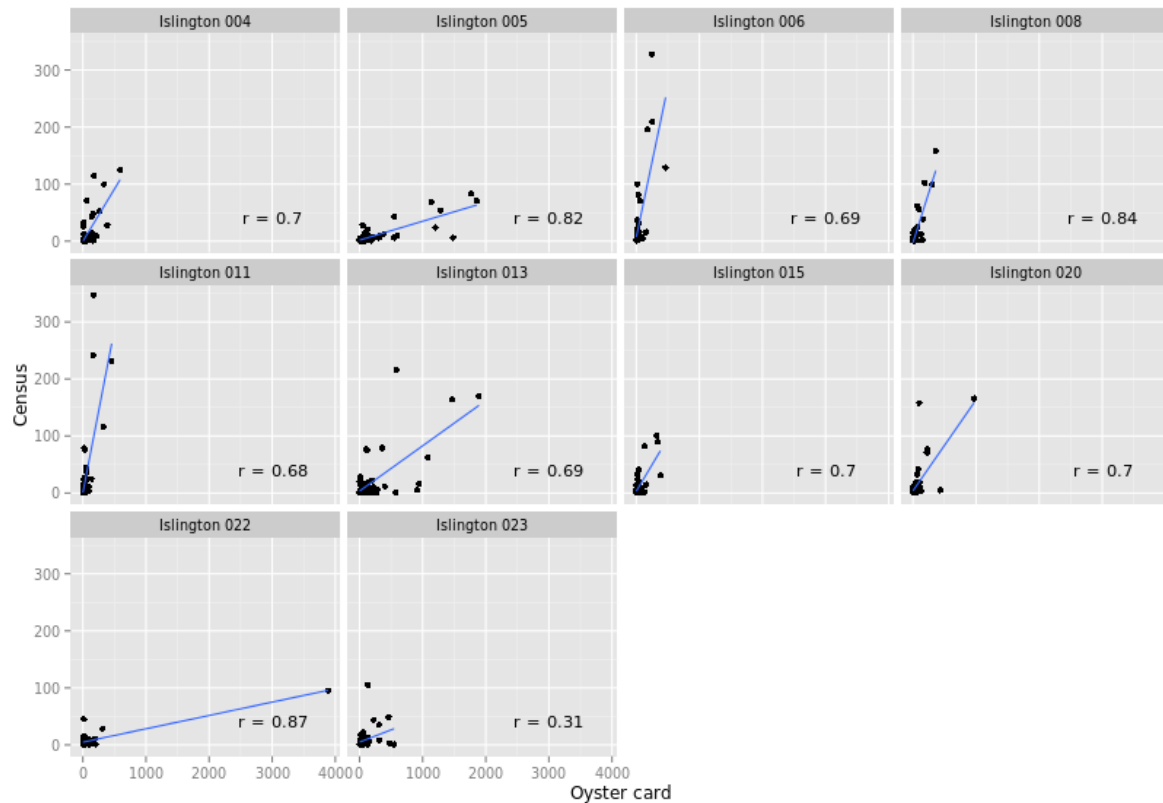
Figure 9: Map showing ratios between Oyster card (start tube) and Census (residence) data (Oyster card divided by Census)



As for LAs, each origin MSOA was taken in turn, and the Oyster card flows from it into each MSOA were correlated with corresponding Census flows. The correlations were poorer in general than those for LAs (average of 0.71 for MSOAs compared with 0.96 for LAs), which would be expected

given the higher degree of variability associated with analysis using smaller areas. However as can be seen in Figure 10, the better correlations tended to be driven by a smaller number of dominant flows. Figure 10 shows scatter plots of Oyster card and Census data for each MSOA in Islington containing a tube station. Most of the MSOAs are dominated by four or five large flows and have correlations around 0.7, with the exception of Islington 022 which has a good correlation (0.87) dominated by one large flow (to City of London 001 which contains nine tube stations).

Figure 10: Scatter plots by MSOA of origin in Islington



It is worth noting that some commuters are excluded from the analysis undertaken. For example some commuters may not live in an MSOA containing a tube station but may still take the tube to work. Such commuters would be included in Oyster card data but excluded from Census data in the analysis undertaken.

Further, the Census statistics include those who commute by underground, metro, light rail or tram. Therefore comparability problems may arise in east London where those who commute using the Docklands Light Railway would be included in Census data but excluded from Oyster card data.

Other comparability problems may occur as the Census counts workers aged 16 or over, while Oyster card data may include those under 16 taking the tube to school, students who regularly take the tube to college or university, or anyone else who starts their regular tube journey between 7am and 10am.

3.3. Factors affecting usefulness of Oyster card data

There are many factors which affect how well Oyster card tube data reflects true commuting patterns. On one hand, the data is very timely and is owned by TfL, a governmental body. Additionally, the data provide complete coverage of the tube network in London.

However, the data do not accurately count commuters. Flows are particularly distorted in areas with mainline train stations, where commuters from outside London arrive and start their tube journeys using their Oyster cards. The availability of alternative transport modes such as overground trains (in Hackney) and cars for commuting (in outer London) also distorts the true picture. Other factors which may affect Oyster card flows are tourists and visitors, although as the Oyster card data used in this analysis represent the number of people travelling on an average weekday between 7am and 10am in November 2012, tourists and visitors may have a smaller impact on the data than the other factors noted.

If this data were to be used to help produce or quality assure existing statistics, its continuity would need to be considered. For example, TfL has started promoting the use of contactless payment cards to pay for travel around London, which can be used instead of an Oyster card for the same price. By using a contactless payment card, TfL will have less information about a traveller than currently with an Oyster photo card (as the photo card application process asks for name, address, date of birth and phone number). However the data used in this analysis did not use this level of detail.

4. Conclusion

This analysis compares 2011 Census data of people taking the tube to work, by their home and work locations, with Oyster card data of commuters' start and end tube stations. The two sources of data correlate reasonably well, although these correlations are driven by a small number of flows to central London.

Oyster card data are more skewed than Census data, with a lower number of small flows and a higher number of large flows. Some of these larger Oyster card flows are distorted by train and airport commuters from outside London starting their tube journeys at mainline train stations in central London, and by those commuters who take more than one mode of transport to work, such as those in Hackney.

In order to further verify the patterns observed, a next step would be to obtain further data from TfL regarding sequences of journeys by the same commuters on different modes of transport, such as those who start their journey at an overground train station in Hackney and then transfer to a tube line later. However, comparability problems may still exist for rail commuters from outside London who use their Oyster card in the tube once they arrive in London.

However in general, the good correlation between the sources raises the possibility that other similar sources of data (such as data from bus and train tickets) could be used to extend this research across the whole country.