

The Geography of Travel to Work in England and Wales: Extracts from the 2011 Census

Alasdair Rae¹ 

Received: 5 July 2015 / Accepted: 8 June 2016 /

Published online: 16 June 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract From a policy point of view, the question of transport connectivity has recently risen up the policy agenda in the UK, and particularly in the North of England where the government are currently seeking to boost economic growth through their ‘Northern Powerhouse’ initiative. Transport is central to this plan. However, existing patterns of connectivity across England and Wales are not always well understood in the policy domain, so this paper attempts to help fill a gap by taking a geovisualisation-based approach to the analysis of 2.4 million individual journey to work flows across England and Wales. The paper builds upon previous research in the field of spatial interaction by exploring patterns associated with different modes of transport. The analysis highlights London’s dominance as a rail commuter destination, relative to major cities in the North of England, in addition to the growth of cycling as a mode of travel to work. The question of ‘error’ in the dataset is then explored, followed by a discussion of possible explanations. The paper ends by reflecting on three key findings and by highlighting opportunities for future research in this field.

Keywords Commuting · GIS · Spatial interaction · Spatial analysis · Travel to work · Census · Geography · Journey to work · Transport

Introduction

The aim of this paper is to provide a detailed account of contemporary travel to work patterns in England and Wales using the most recent Census data and, in so doing, to help inform contemporary debates on transport connectivity, particularly in the North of England. It builds upon a long tradition of spatial interaction studies, including

✉ Alasdair Rae
a.j.rae@sheffield.ac.uk

¹ Department of Urban Studies and Planning, University of Sheffield, S10 2TN, Sheffield, UK

Ravenstein's 'laws of migration' paper in Ravenstein 1885, the seminal Chicago Area Transportation Study of 1959, Tobler's pioneering work on flow mapping in the 1980s (e.g. Tobler 1981, 1987) and Nielsen and Hovgesen's (2005) analysis of 'urban fields' in Denmark. It also draws upon more recent work on commuting in England and Wales by Nielsen and Hovgesen (2008) and Rae (2009). The data are sourced from the 2011 Census' question on mode of travel to work and are reported for interactions between 7201 small areas. This produces a potential spatial interaction data matrix of almost 52 million cells but in reality the dataset contains just over 2.4 million area interaction flows. The challenges associated with the spatial analysis of this volume of data in a desktop computing environment are of course not new (for example, such analyses have been possible since the 1950s in Sweden) but for the purposes of replicability and transparency, they are described in more detail in the second section of the paper.

In order to achieve the aim of the paper there are three objectives. The first is to provide a detailed account of travel to work in England and Wales at the aggregate level, for all modes of transport. The second objective is to explore the spatial patterns of commuting associated with different transport modes from a visual perspective, in a way that has not been attempted to date for this particular dataset. This paper focuses on travel by car, rail, bicycle and on foot since these account for the vast majority of journeys to work. The third objective of the paper is to highlight the value of geovisualisation in the analysis of large spatial interaction data matrices. Such approaches are an effective means of detecting error in a dataset or to identify seemingly anomalous patterns of movement. The policy relevance of the underlying dataset, and associated geovisualisations, is highlighted in the paper with reference to the Government's Northern Transport Strategy (DfT 2015a), which draws upon the analysis presented here, and features one of the geovisualisations on its front cover. Furthermore, the analysis also features in the Department for Transport's recent report on high speed rail, entitled 'Rebalancing Britain' (DfT 2015b). Examples from other nations, such as Sweden, demonstrate that a better understanding of the geography of commuting can be particularly useful in framing debates on regional development more broadly (e.g. Amcoff 2009).

The next section of the paper briefly describes the data and methods, in order to aid replication. It is evident that such 'big' datasets are inherently messy and that even the simple act of opening a file can be challenging in a desktop computing environment. This may seem a trivial point, but it is an important one, since the volume of data described here – whilst not 'big' from a database management point of view, has arguably contributed to its relative under-exposure in the policy domain and, as a result, to a relative lack of understanding about connectivity, infrastructure and investment gaps (cf. Cox and Davies 2013). This section also attempts to make a small methodological contribution. Following this, the paper presents the results of the spatial analysis in visual form. In addition, the data associated with each map is presented and discussed. This section of the paper looks at aggregate commuting patterns and then compares, in turn, travel to work by car and by rail, and travel by bicycle and on foot. Analysis of travel to work patterns by distance is also presented here in order to determine the different journey lengths associated with the daily commute, as in previous studies (e.g. Green et al. 1999; Rapino and Fields 2012). The question of error in the dataset (and whether it is in fact error at all) is then explored with reference to journeys on foot and by bicycle. The conclusion reflects upon the main findings and makes a small number of recommendations for future research in this area.

Data and Methods: from ‘Big’ Data to Spatial Data Analysis

On 25 July 2014, just over three years after the 2011 Census, the Office for National Statistics (ONS) released the first set of detailed travel to work datasets for England and Wales, available as a series of bulk downloads (NOMIS 2014). Following the identification of minor production errors, the data were re-released on 11 August 2014 so that version of the file is used here. These data were also published for the 1991 and 2001 Censuses, but in a less reliable format for small geographies since the former was a 10 % sample and the latter was subject to an obfuscatory ‘small cell adjustment mechanism’, known as SCAM (see Stillwell and Duke-Williams 2007). It would have been preferable to make meaningful comparisons between inter-censal periods in this analysis, but the lack of clarity over the accuracy of small counts for earlier periods renders it impossible.

There is now a large volume of available travel to work data for England and Wales but for the purposes of this study, Table WU03EW was used. This table covers commuting between the 7201 middle layer super output areas (MSOAs) in England and Wales, which had an average population of 7787 at the time of the 2011 Census. In addition to providing the total commuter flow between MSOAs, this table also includes sub-totals for every mode of travel. In this paper, the focus is on total flows, journeys by car, train, bicycle and on foot. These constitute the majority of commuter flows in England and Wales (72 %). Before moving on to present the results of the analysis, this section looks at the overall characteristics of the dataset then describes the analytical method in more detail.

The Dataset: MSA-level Commuting in England and Wales

The MSA-level commuting table WU03EW contains a total of 2,402,201 rows of data, and 12 columns, for a total of over 28,826,412 cells of numeric data. Each row of data contains the alphanumeric code for the origin and destination associated with individual MSA-to-MSA journeys, plus the 12 fields. These are shown below, in addition to the total number of commuters associated with each mode of transport. There were just under 27 million individual commuters accounted for in the dataset and the largest single modal category was ‘driving in a car or van’ with 14.5 million person commutes and 54.4 % of the total. Just over 10 % of the population surveyed in the 2011 Census worked mainly at home. It is worth noting that even those who ‘work at home’ may actually use different transport modes to meet clients or carry out work-related duties, so the ‘work at home’ category might involve some form of travel (ONS 2013). Thus, the second highest discrete modal category was for travel to work by foot with 2.6 million and 9.8 % of the total. This is followed by ‘bus, minibus or coach’ (7.2 %) and ‘passenger in a car or van’ (5.0 %) or train (5.0 %). As we will see later in the paper, however, the spatial distribution of this modal split is very uneven, as one might expect. Although cycling to work has risen by 13.9 % nationally since 2001, and has more than doubled in London during the same time period, it still only accounts for 2.8 % of journeys to work in England and Wales (ONS 2014a).

All Categories: Method of Travel to Work 26,681,568 (100 %)

1. Work mainly at or from home 2,778,019 (10.4 %)
2. Underground, metro, light rail, tram 998,526 (3.7 %)
3. Train 1,327,193 (5.0 %)
4. Bus, minibus or coach 1,916,704 (7.2 %)
5. Taxi 128,433 (0.5 %)
6. Motorcycle, scooter or moped 209,240 (0.8 %)
7. Driving a car or van 14,506,119 (54.4 %)
8. Passenger in a car or van 1,330,855 (5.0 %)
9. Bicycle 744,459 (2.8 %)
10. On foot 2,610,382 (9.8 %)
11. Other method of travel to work 131,638 (0.5 %)

In the above dataset, those who work from home are assigned the MSOA code from their residence address in the origin column and OD0000001 as their destination (2.8 million in total). Similarly, those who work offshore are coded with OD0000002 for their destination (44,000), those with no fixed location as OD0000003 (2.2 million) and those who work outside the UK as OD0000004 (38,097). Therefore, there are over 5 million commuters who appear in the origin column of the dataset but cannot be assigned a precise geographic location for their destination. The majority of these are individuals who work at home and therefore do not have a ‘journey to work’ in the traditional sense even if they may, as noted above, travel during the working day.

Manipulation of the Spatial Interaction Dataset

Despite recent computational advances, such large datasets are inherently ‘messy’ in that they are difficult to manipulate, contain limited geographic information, and include hundreds of seemingly anomalous records. In relation to the latter, however, it is difficult to accurately identify these without first taking a spatial approach. To achieve this, the origin and destination MSOA area codes in the WU03EW dataset were used as the basis for making the data both mappable and meaningful. ONS population-weighted centroid x and y coordinates for MSOAs in England and Wales were obtained, then used with an ONS lookup file for MSOAs with MSOA names, local authority codes and local authority names in order to construct a single file containing MSOA x and y population-weighted centroid coordinates, plus each MSOA’s name, local authority name and local authority code. After performing this lookup operation in Excel, the MSOA attribute table was joined to the England and Wales commuting dataset. These joins were performed in QGIS 2.4, then subsequently saved as a text file and imported and mapped (see Fig. 1 for screenshot). Further details on the method used here can be found in Rae (2014). Similar operations are also possible with proprietary desktop GIS packages such as ArcGIS and MapInfo (see Rae 2011) and other open source tools such as R. However, in terms of simplicity and efficiency, QGIS was preferred here.

Once the file was imported and the flow lines drawn, the new dataset was saved in the ESRI shapefile format and used as the basis for initial exploratory spatial data analysis. It was immediately clear, however, that a structured and logical filtering

approach would be necessary in order to make any sense of the data from a geovisualisation perspective. Therefore, the results presented below follow Rae's (2011) 'principles for the orderly loss of information', which are based on 'expansive inclusion' (include all possible data elements at the outset), 'iterative loss' (filter the data iteratively to derive meaning), 'simplicity from complexity' (aim to identify key patterns) and 'optimal compromise' (recognise that a series of analytical compromises are necessary in spatial data analyses).

The maps presented in the next section, like all carefully constructed geovisualisations, are based on a series of analytical decisions underpinned by the goal of moving from data to information (Longley et al. 2005). The progression from information to knowledge and then wisdom, in relation to the policy implications of the findings, is beyond the scope of this paper. However, it is hoped that the results presented here will serve as useful empirical evidence in relation to national patterns of travel to work in England and Wales. The fact that earlier versions of this analysis have captured the attention and imagination of national policymakers is perhaps testament to the power of a geovisualization approach, as attempted here (DfT 2015a b).

Mapping National Commuting Patterns in England and Wales

Commuting in England and Wales is geographically extensive, multi-modal in nature and dominated by a relatively small number of urban centres, principally Greater London. Almost 27 million individual commuter journeys are accounted for in the dataset, so it is impossible to display every flow. Rather, a series of maps was produced for England and Wales which display all commutes above a carefully considered threshold, then filtered and layered in order to highlight dominant flows. Initially, the

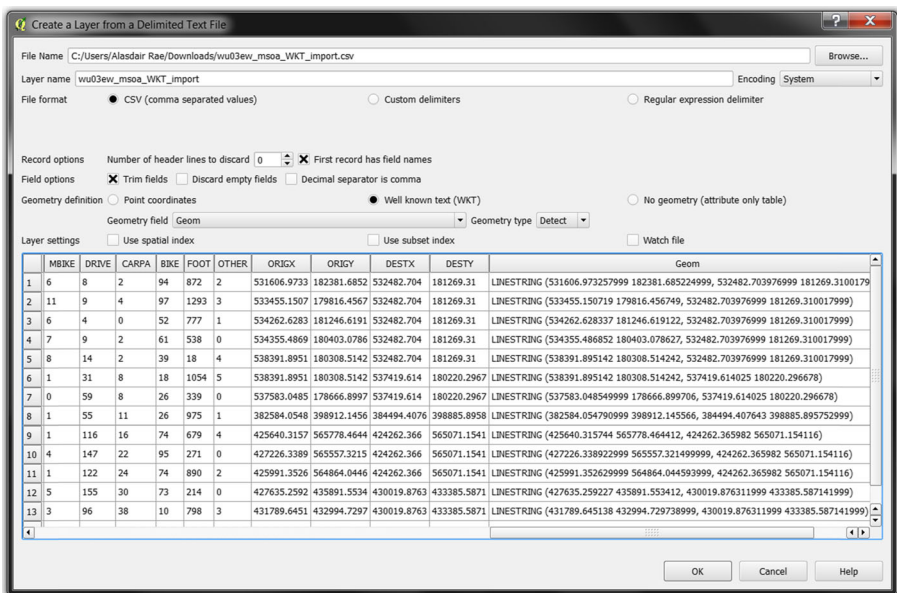


Fig. 1 Importing WKT-formatted flow dataset in QGIS 2.4

focus is on national patterns of commuting in England and Wales in relation to all modes of travel to work, with additional data for local areas to provide some context. The focus then shifts to a comparison of travel to work by car or van compared to journeys by train. Finally, patterns associated with travel to work by bicycle and on foot are examined. This serves as a useful lead-in to the paper's final section, on mapping error. An important point to emphasise is that the contribution here is centred on the visual approach to unlocking spatial patterns in the data, rather than the act of exploring spatial interaction data per se. The latter has, of course, been done many times before (e.g. Nielsen and Hovgesen 2008) but the former has not, to date, been published in this form. Furthermore, commuting data is only one type of interaction data with which we can understand spatial interaction, as Hincks and Wong (2010) demonstrated most effectively in their study of the interaction of housing and labour markets in North West England.

An Overview of Travel to Work in England and Wales

In Fig. 2, all MSOA-to-MSOA travel to work flows of 10 or more in England and Wales are shown. Here, and subsequent maps, the magnitude used as a cut-off for display purposes has been set so that the vast majority of all flows are represented (e.g. 77 % in Fig. 2), whilst maintaining visual coherence. Put simply, the cut-off values were chosen in order to maximise the communicative power of the maps whilst minimising the amount of data loss, in line with Rae's (2009) geovisualization principle of 'optimal compromise'. Individual lines are not symbolised by flow magnitude (since this would cause severe visual occlusion) but instead the QGIS 'addition' blend mode has been used to highlight very dense travel to work clusters. This has the effect of making individual cities 'glow' on the map since it uses a simple pixel value addition function to add together attributes for overlapping flow lines. The cities of Newcastle, Leeds, Birmingham, Southampton and London - to name a few - are easily identifiable as major commuter locations, as one would expect. Figure 2 mirrors to a large extent the underlying settlement patterns of England and Wales, but it provides an insight into the functional economic geography of the nation and, crucially, the degree to which places are connected or not. Also visible in the map are a smaller number of very long distance flow lines, two of which can be seen emanating from Manchester and heading in southwesterly and southeasterly directions. Such apparent anomalies are discussed in more detail later in the paper. At a basic level, then, this provides a useful general overview of commuting patterns in England and Wales but the underlying dataset is much richer, as described below.

Individual summary queries were performed on the travel to work dataset for major cities. Greater London, for example, accounts for just over 3.7 million commuter flows and attracts 789,833 commuters from other English regions, including over 400,000 from the South East and over 300,000 from the East of England. Over 2.9 million journeys to work have both their origin and destination within Greater London. Overall, Greater London was the destination point for 17.9 % of all commuter journeys in England and Wales in 2011, compared to 13.9 % of the working age population.

At the individual city level beyond London, Birmingham is the destination for 2.0 % of all journeys to work in England and Wales, followed by Leeds (1.7 %), Manchester (1.3 %), Sheffield (1.0 %), Liverpool (1.0 %), Bristol (0.9 %), Bradford (0.8 %) and



Fig. 2 The Daily Commute in England and Wales (flows of 10 or greater). Data Source: Office for National Statistics licensed under the Open Government Licence v.1.0

Cardiff (0.8 %). Across England, 37.6 % of all commuter journeys terminate in Greater London or the metropolitan districts of Greater Manchester, Merseyside, South Yorkshire, West Yorkshire or Tyne and Wear. These patterns of movement, and their associated urban structures, have been the subject of previous studies (e.g. Giuliano and Small 1993) so the analysis conducted thus far provides confirmation from the most recent data of the continued concentration of employment in the urban centres of England and Wales.

What is more revealing, however, is to consider the distances associated with these patterns. After calculating the mean Euclidean (straight line) distance between all 2.4 million MSOA-to-MSOA population-weighted centroids, the distributional breakdown for travel to work distances of up to 5 km, up to 10 km and then for 10 km bands, up to

100 km and above was explored. The results of this analysis are presented in Table 1. According to analysis by the Office for National Statistics (2014b) the mean Euclidean distance travelled to work in England and Wales in 2011 was 15.0 km, up from 13.4 km in 2001. This compares to a calculated mean Euclidean distance of 16.3 km from the MSOA dataset used here. This higher figure is a reflection of the fact that the analysis is undertaken using the larger MSOA geography rather than the full postcode origin and destination approach taken by ONS. The data presented below indicate that 41.2 % of journeys to work in England and Wales were under 5 km, 63.3 % under 10 km and 82.9 % under 20 km. There is a notable decline in the proportion of people commuting after the 30 km cut-off. At the other end of the scale, 2.0 % of commutes were above the 100 km threshold. Since these distances are computed from population-weighted MSOA centroids they have a lower level of precision than official ONS estimates. However, it should also be noted that ONS also use a straight line distance measure, between the postcode of origin and postcode of workplace. Thus, the figures presented here are likely to provide a good indication of the distances associated with travel to work in 2011.

Modal Comparison 1: Car and Van Commuting vs. Train Journeys

In absolute terms, the number of people travelling to work in a car or van was by far the highest, at over 14.5 million (54.4 % of the total). If car or van passenger journeys are added to this, the total traveling by car or van is 59.9 %. This compares to 1.3 million (5.0 %) for those taking the train to work. In Fig. 3 these patterns are presented side by side, for all flows of 5 or higher, using the same mapping methodology as in Fig. 2, where the lighter areas represent the highest flows. Given the much more extensive nature of the road network and far greater number of possible journeys, it is not

Table 1 Distribution of travel to work distances in England and Wales in 2011

Distance (km)	Total commuters	% of total	Cumulative total	Cumulative %
0 to 4.99	8,908,897	41.2	8,908,897	41.2
5 to 9.99	4,779,528	22.1	13,688,425	63.3
10 to 19.99	4,239,091	19.6	17,927,516	82.9
20 to 29.99	1,579,290	7.3	19,506,806	90.2
30 to 39.99	696,132	3.2	20,202,938	93.4
40 to 49.99	388,319	1.8	20,591,257	95.2
50 to 59.99	234,529	1.1	20,825,786	96.3
60 to 69.99	134,719	0.6	20,960,505	96.9
70 to 79.99	102,165	0.5	21,062,670	97.4
80 to 89.99	72,175	0.3	21,134,845	97.7
90 to 99.99	54,074	0.3	21,188,919	98.0
100 and above	433,266	2.0	21,622,185	100.0

These totals exclude those workers for whom no geographical workplace can be specified, such as those with no fixed workplace, and those who work at home and therefore do not 'travel to work' in the traditional sense. This accounts for over 5 million workers in England and Wales

surprising that car and van commutes should look as they do in Fig. 3. The very different picture provided in the commuting by train map demonstrates the dominance of London as a rail commuter destination, with other major English and Welsh cities less dominant from a national perspective. The relatively weak east to west links stand out clearly here, in addition to the apparent monocentric nature of rail commuting in England as a whole, with London dominating the national picture.

Besides the contrasting volume of flows, the major difference between these modes of transport as a method of travel to work is in their spatial distribution. For example, the areas with the highest percentage of journeys to work by car or van are in more rural and suburban locations. The highest figure is for East Dorset on the south coast of England, where 79.5 % of commuter journeys terminating there are by car or van. This is followed by North West Leicestershire (79.1 %) and North Warwickshire (78.0 %) in the Midlands. At the other end of the scale, only 6.1 % of people arriving for work in the London Borough of Westminster drive there, followed by other London Boroughs such as Camden (9.3 %), Kensington and Chelsea (11.6 %) and Islington (12.2 %). Given the level of congestion in central London, this is not surprising. For Greater London as a whole, 64.6 % of all journeys by rail in England and Wales terminate there, with a total of 780,589 rail commuters. By contrast, England's 36 metropolitan districts, with a combined population of more than 11 million (compared to 8.3 million in London), accounted for just 13.3 % of all rail commuter destinations in 2011. The first non-London local authority to appear on the list of train commuting locations is Liverpool, ranked 17th, where 9.6 % of terminating commutes are by train.

Of course, none of the above is really surprising, since people will generally choose the most convenient mode of travel, regardless of where they are. In relation to recent policy developments in the North of England (DfT 2015a), and transport strategy in particular, understanding the modal split between car and train commuting could

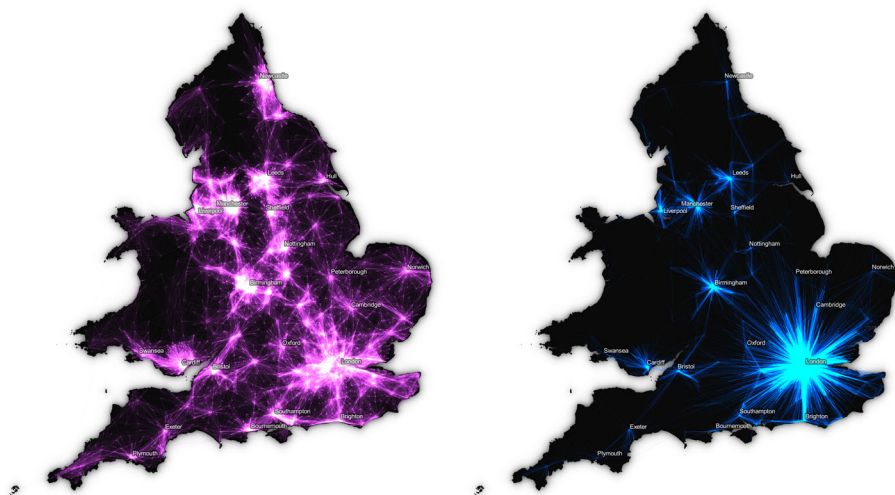


Fig. 3 Travel to work by car (*left*) vs train (*right*) in 2011, all flows ≥ 5 . Data Source: Office for National Statistics licensed under the Open Government Licence v.1.0

provide an insight into potential connectivity gaps between major cities. To explore this proposition in a little more detail, four inter-city pairings (based on local authority boundaries) are explored here in relation to travel to work by train versus travel to work in an automobile (i.e. either driving or as a passenger). The results are shown in Table 2. The city pairings of Liverpool and Manchester (56 km apart), Leeds and Sheffield (56 km), Manchester and Sheffield (61 km), and Manchester and Leeds (71 km) were explored in relation to bi-directional travel to work flows. These locations are critically important to the current government's 'Northern Powerhouse' initiative (DfT 2015a), which talks of 'transformative connectivity' (p. 3) and the particular need to transform 'city to city rail connectivity' (p. 5). However, the evidence suggests that the total volume of commuting between these major Northern cities is low and – more importantly – connectivity is in some cases dominated by auto travel where, with sufficient investment in infrastructure, a major modal shift to rail would be viable.

In modal share terms, London clearly dominates rail travel and this is to be expected given its economic power, traffic congestion, location and existing infrastructure. However, it is also somewhat surprising that the main regional cities of England do not have a higher share of rail commuting, and particularly so in the case of the analysis of city pairings presented in Table 2 where, for example, six times as many people commute from Leeds to Sheffield by automobile than by train. This may provide useful evidence in relation to the effects of relative under-investment in infrastructure in the North of England of the kind identified by Cox and Davies (2013). To base any firm infrastructural investment decisions solely on the data presented here would be a leap of faith, but the evidence does suggest that there is potential for significant modal shift between car and rail travel in the major English cities outside London. This is further reinforced by evidence from the Department for Transport (2013) which noted that Sheffield had the highest peak overcrowding figures outside London and that Leeds and Manchester also suffer from significant peak hour overcrowding.

Modal Comparison 2: Cycling vs. Travel on Foot

Between 2001 and 2011 the percentage increase in people cycling to work in England and Wales was 13.9 %, with over 740,000 more people now riding their bicycle to

Table 2 Inter-city modal share comparisons: train vs automobile, 2011

Inter-city links	Total commuters	Train (%)	Auto (%)	Ratio
Liverpool to Manchester	1706	30.5	58.1	1.9
Manchester to Liverpool	666	21.0	67.9	3.2
Leeds to Sheffield	1154	12.9	79.4	6.1
Sheffield to Leeds	2477	16.4	73.5	4.5
Manchester to Sheffield	236	31.8	51.7	1.6
Sheffield to Manchester	527	44.8	43.3	1.0
Leeds to Manchester	718	31.6	54.0	1.7
Manchester to Leeds	530	28.7	51.7	1.8

Source: 2011 Census. Calculations are author's own

work. This figure masks significant variation at the local level, since research by ONS (2014a) shows that the number of people living in London who cycled to work more than doubled over ten years (100.8 % increase) with significant increases in other areas such as Brighton and Hove (109.4 %), Bristol (94.5 %), Manchester (82.8 %), Newcastle (81.0 %), Sheffield (80.4 %) and Cardiff (64.8 %). The growth in cycling as a mode of commuting in major cities is one of the key messages to emerge from the 2011 travel to work data since it has placed increasing demands upon many local authorities as they seek to deliver the kind of infrastructure necessary to support this change (Chatterjee et al. 2013).

Despite recent growth, cycling across England and Wales accounts for just 2.8 % of all journeys to work. By contrast, travel to work on foot accounted for 9.8 % of journeys in 2011 (2.6 million). The spatial distribution of these patterns is shown in Fig. 4, which follows the mapping method shown previously. Given the highly localised nature of cycling and travel on foot as a mode of travel to work, these maps most closely resemble the urban footprint of England, yet there are some notable exceptions, discussed on the next section of the paper. In both maps, central London stands out as a walking and cycling commuter hotspot, and in the cycling map Cambridge and Oxford do as well. These local authorities have the highest proportion of workers cycling to work, at 21.7 % and 14.6 % respectively. If we include commutes originating in these areas, with a destination inside *or* outside Cambridge and Oxford, these figures rise to 29.0 % and 17.1 %. In addition to Cambridge and Oxford, only Gosport on the south coast and York in the North of England saw 10 % or more of their workers arrive by bicycle. The highest placed London Borough was Hackney, ranked 6th with 8.3 % of workers arriving by bicycle.

The local authority with the highest proportion of arrivals to work on foot in 2011 was Scarborough, with 25.9 %, followed by Brighton and Hove at 24.4 %. However, when we look at mode of travel from the perspective of resident origin, 48.4 % of those

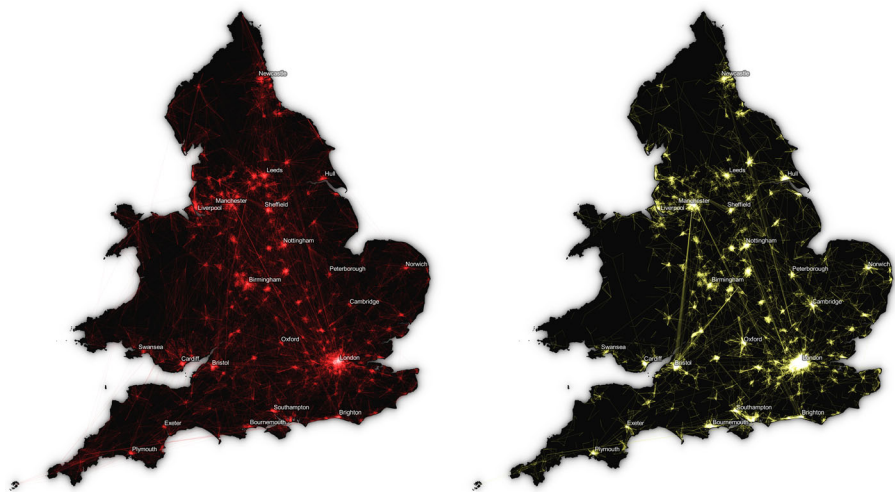


Fig. 4 Travel to work by foot (left) vs bicycle (right) in 2011, all flows ≥ 3 . Data Source: Office for National Statistics licensed under the Open Government Licence v.1.0

living in the City of London travel to work on foot. Again, given the high housing density and level of traffic congestion in central London, this is not surprising. What is particularly surprising, when we look closely at Fig. 4, is the number of apparently very long journeys either on foot or by bicycle. In order to explain this, we need to go back to the original question, from the 2011 Census, which was asked in the following way:

Q41 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

Clearly, given the fact that the question asks respondents to identify their commuting mode by ‘the longest part, by distance’ it is simply not conceivable that the longer commuting flows in Fig. 4 represent daily travel to work journeys. Even the most determined cyclist or seasoned runner could not commute between Bristol and Manchester (275 km), as the maps would suggest. Thus, a visual approach to flow data analysis can serve as an invaluable source of error detection, and this is the subject of the final empirical section of the paper.

Mapping Error?

The extent of apparent ‘errors’ associated with anomalous long distance on foot and cycling journeys to work can be more clearly assessed through some further empirical analysis based on a simple distance variable. For the purposes of this analysis, a high cut-off of 20 km each way for journeys on foot and 100 km each way for journeys by bicycle was used. This very large cut-off was selected as the point beyond which there will be no regular commuting by these modes. In the case of those who travel to work on foot, 204,179 individuals appear to commute more than 20 km and for bicycle journeys over 100 km the figure is 9405. These patterns are displayed in Fig. 5. It is theoretically possible that there could be a small number of dedicated ultra-endurance athletes undertaking such journeys but for the purposes of this analysis it is assumed that journeys beyond this distance are not in fact point-to-point cycling or walking journeys and in this sense are errors. These errors may have arisen in one of two ways. The first is ‘modal misassignment’ and the second is a ‘location misassignment’. In the first type of error, there is the possibility of respondents answering question 41



Fig. 5 Mapping Error? Long-distance on foot (over 20 km, *left*) and bicycle flows (over 100 km, *right*). Data Source: Office for National Statistics licensed under the Open Government Licence v.1.0

incorrectly, or dishonestly. For example, if a person travels a long distance to work on a train with a fold up bicycle, and then travels the last 5 km by bicycle, they may have decided that they travel to work by bicycle and answered the question in that way. If their residence postcode is a long way from their workplace it would result in this kind of modal misassignment error. The same situation also applies to journeys to work on foot. Such errors could also arise when respondents inadvertently tick the wrong box on the Census form.

A location misassignment, on the other hand, may be caused due to a data entry error whereby in question 40 of the 2011 Census form the respondent enters a head office location for their company rather than the location of the specific workplace where they normally work. This could be caused where a commuter works at a branch office location of a national firm but enters the corporate headquarters address in the form for question 40. Such errors could also arise at the data processing stage for the Census, if workplace locations were incorrectly geocoded. In the lead up to the 2001 question, Rees (1998) conducted some research into user needs for the Census and identified concern around the ‘mode of travel to work’ question but these concerns did not highlight either of these sources of error. The 2011 Census Quality Survey (ONS 2014c) does, however, provide additional insight into the accuracy of these data since they compare self-reported responses from the Census to responses to the same questions asked through face-to-face interviews with over 5000 households. The ‘agreement rate’ for the ‘address of workplace’ question in 2011 was 82.2 % and for method of travel to work it was 85.5 % (ONS 2014c, p. 31). This compares to an agreement rate of 99.7 % for ‘sex’ and 60.4 % for ‘national identity’ and 55.0 % for ‘year last worked’. These data provide some useful additional context on the confidence we can have in individual patterns of travel reported in the Census.

Other sources of complexity, rather than error, also exist in the dataset. Where a respondent works in a different location each day or week and those locations are geographically disparate, any resultant flow line is unlikely to be a true representation

of their journey to work. Equally, residents who live in one place during the week and return home at weekends may be the source of seemingly impossible commutes when plotted on a map. Therefore, the majority - or even all - of the very long commutes identified in Fig. 5 may not be errors at all. They may simply reflect the complex interaction between residential and labour market mobilities in England and Wales at the time of the 2011 Census.

It is perfectly feasible that individuals could live in Brighton and work in Manchester during the week, and travel to work on foot or by bicycle when they are in Manchester. Further complexity is introduced when we consider what Hardill and Green (2003) refer to as ‘new movements and mobilities’; essentially, the increasingly diffuse and diverse relationships between where people work and where people live (see also Green 1997). They cite the example of an ‘economically motivated’ commuter who works in London but whose main family home is in Cumbria, in the North of England (over 400 km away), in addition to a number of other scenarios.

The extent to which the long-distance commuting patterns identified in Fig. 5 are related to an increasingly complex spatiality within the labour market, or, are related to sources of error in the underlying data is something it is not possible to identify here. Crucially, it is also not possible to distinguish between those long-distance ‘journeys’ which represent a true link between a residence and workplace (even if it is not a daily link) and those which do not (e.g. location misassignment errors). In the context of a dataset with over 2.4 million individual interactions, these problems are relatively small but some further analysis on the issue by the Office for National Statistics would be particularly welcome. The value of the geovisualisation approach taken here, then, is in the simple way it can help identify such ‘errors’ and help us understand their significance and volume. They are important in a small number of cases but not in terms of their overall volume. Therefore, the belief here is that these errors do not undermine the validity of the dataset as a whole, nor do they invalidate analyses of travel to work flows associated with individual towns or cities.

The final points on error here relate to the issues of ‘field of application’ and the severity of errors. Depending upon the end use of the data - that is, what field it is being applied to and at what spatial scale, any errors will have different degrees of severity. A good example of this can be found with reference to the ‘Quality Note’ on 2011 Census Origin and Destination statistics released by ONS in 2014, some time after the initial release of the data (ONS 2014d). The note explains that subsequent to the publication of the data in July 2014, a number of anomalies were observed by users, including flows of people between local authorities which would not be expected (e.g. because they are a long way apart). At one end of the error scale, the data suggested a flow of 1193 commuters between Sheffield (residence) and Bury (workplace). These settlements are nearly 80 km apart and as such we would expect a much lower level of interaction between them. The impact of ignoring this potentially quite large error could be quite significant, were it to be taken into account in transport planning, for example. At the other end of the scale, an erroneous flow value of 119 between Conwy in North Wales and Liverpool (90 km apart)

is much less likely to have a negative impact in practice. In fact, as the ONS note recognises, these are only ‘potential errors’ so any analysis of this kind requires more in-depth analysis of individual anomalous flows. For such a task, as discussed above, a visual approach can be invaluable.

Conclusions

From the foregoing analysis, the complexity and volume of the daily commute in England and Wales is evident, so the paper ends by reflecting on key points and by making some recommendations for future research. The first obvious conclusion to be drawn from the analysis is that the majority of commuting remains quite localised, with over 40 % of journeys to work within 5 km or less and over 60 % within 10 km or less. However, this is likely to vary significantly by area and by mode of transport, so future research which can untangle this more complex nexus of travel to work linkages would be particularly welcome. This is particularly true in the North of England, with the current ‘Northern Powerhouse’ strategy explicitly incorporating inter-city connectivity gaps as a policy problem to be solved.

A second conclusion, directly relevant to current policy developments, is that there would appear to be significant potential for growth in commuting by rail in the North of England, given the wide disparity between flows into London and those into the major regional cities, such as Manchester, Newcastle and Leeds (DfT 2015a). However, this would require significant investment in infrastructure and rolling stock to increase capacity and at the present time this seems unlikely. Indeed, at the time of writing, the new Conservative government have just delayed plans for major rail infrastructure upgrades in the North of England. For positive changes to take place in this area it would require a reappraisal of public transport’s role (Banister and Gallent 1998) and the kind of long-term commuting sustainability approach taken in Sweden in relation to Stockholm’s satellites (Cervero 1995). It is hoped that the evidence presented above could be used to highlight the need to re-think public transport investment within England and Wales and to begin a conversation around the sustainability of current commuting patterns more generally. A useful avenue for future research in this area would therefore be an empirical study which models the modal shifts associated with increased rail capacity in the North of England, particularly relating to east-west connectivity. Clearly, modal choice goes beyond utilitarian concerns about time and cost (Mann and Abraham 2006) but the evidence presented in Table 2 on inter-city modal share suggests that further investigation of this topic, in relation to infrastructure investment options, is warranted.

The final conclusion relates to the power of geovisualisation to identify what appear to be errors in the travel to work dataset. As discussed above, these may not in fact be errors at all but from an origin-destination flow perspective we can safely assume that daily commutes of 250 km each way are not the norm. The mapping methodology adopted here is therefore a very useful entry point to a deeper understanding of the increasing complexity of travel to work in England and Wales, of the kind described by Hardill and Green (2003). However, with the current release of data from ONS it does not appear possible

to identify which of these apparently anomalous flows are associated with misassignment errors, which are associated with complex household arrangements, and which might be caused by other factors. Nonetheless, the next logical avenue of research in this area is to explore in more detail the locations, modes and volumes associated with the relatively small percentage of travel to work flows that appear not to fit the standard model of the journey to work between a known origin and destination.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Amcoff, J. (2009). Rapid regional enlargement in Sweden: a phenomenon missing an explanation. *Geografiska Annaler: Series B, Human Geography*, 91(3), 275–287.
- Banister, C., & Gallent, N. (1998). Trends in commuting in England and Wales—becoming less sustainable? *Area*, 30(4), 331–341.
- Cervero, R. (1995). Sustainable new towns: Stockholm's rail-served satellites. *Cities*, 12(1), 41–51.
- Chatterjee, K., Sherwin, H., & Jain, J. (2013). Triggers for changes in cycling: the role of life events and modifications to the external environment. *Journal of Transport Geography*, 30, 183–193.
- Chicago Area Transportation Study (1959). Final report, Survey Findings (Vol. I, pp. 96–99). Chicago.
- Cox, E., & Davies, B. (2013). *Still on the wrong track: an updated analysis of transport infrastructure spending*. Newcastle upon Tyne: IPPR North.
- Department for Transport (2013). *Rail passenger numbers and crowding on weekdays in major cities in England and Wales: 2013*. DfT, London: Statistical Release.
- Department for Transport (2015a). *The Northern Powerhouse: One Agenda, One Economy*. DfT, London: One North. A report on the Northern Transport Strategy.
- Department for Transport (2015b). *Rebalancing Britain: From HS2 towards a national transport strategy*. London: DfT.
- Green, A. E. (1997). A question of compromise? Case study evidence on the location and mobility strategies of dual career households. *Regional Studies*, 31(7), 641–657.
- Green, A. E., Hogarth, T., & Shackleton, R. (1999). Longer distance commuting as a substitute for migration in Britain: a review of trends, issues and implications. *International Journal of Population Geography*, 5(1), 49.
- Giuliano, G., & Small, K. A. (1993). Is the journey to work explained by urban structure? *Urban Studies*, 30(9), 1485–1500.
- Hardill, I., & Green, A. (2003). Remote working—altering the spatial contours of work and home in the new economy. *New Technology Work and Employment*, 18(3), 212–222.
- Hincks, S., & Wong, C. (2010). The spatial interaction of housing and labour markets: commuting flow analysis of north West England. *Urban Studies*, 47(3), 620–649.
- Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). *Geographic information systems and science* (Vol. 2). Chichester: John Wiley and Sons.
- Mann, E., & Abraham, C. (2006). The role of affect in UK commuters' travel mode choices: an interpretative phenomenological analysis. *British Journal of Psychology*, 97(2), 155–176.
- Nielsen, T. A. S., & Hovgesen, H. H. (2005). Urban fields in the making: new evidence from a Danish context. *Tijdschrift voor Economische en Sociale Geografie*, 96(5), 515–528.
- Nielsen, T. A. S., & Hovgesen, H. H. (2008). Exploratory mapping of commuter flows in England and Wales. *Journal of Transport Geography*, 16(2), 90–99.
- NOMIS (2014) Origin-Destination individual tables - part 1, available from <https://www.nomisweb.co.uk/census/2011/bulk/rOD1>, accessed 30 June 2015.
- ONS (2013) 2011 Census Analysis - Method of Travel to Work in England and Wales Report http://www.ons.gov.uk/ons/dcp171766_299766.pdf, accessed 22 December 2015.
- ONS (2014a). *2011 census analysis - cycling to work*. Titchfield, UK: Office for National Statistics.
- ONS (2014b). *2011 Census Analysis - Distance Travelled to Work*. UK: Titchfield.

- ONS (2014c). *2011 census quality survey*. UK: Titchfield.
- ONS (2014d). *Incorrect flows within 2011 census origin and destination statistics*. Titchfield, UK: Quality Note.
- Rae, A. (2009). From spatial interaction data to spatial interaction information? Geovisualisation and spatial structures of migration from the 2001 UK census. *Computers, Environment and Urban Systems*, 33(3), 161–178.
- Rae, A. (2011). Flow-data analysis with geographical information systems: a visual approach. *Environment and Planning-Part B*, 38(5), 776.
- Rae, A. (2014) Flow mapping with QGIS, available at: <http://www.undertheraedar.com/2014/10/flow-mapping-with-qgis.html>, accessed 24 December 2015.
- Rapino, A. & Fields, M. (2012) *Mega commuting in the US: time and distance in defining long commutes using the 2006-2010 American Community Survey*. Washington, DC: US Census Bureau.
- Ravenstein, E. G. (1885). The laws of migration. *Journal of the Statistical Society of London*, 167–235.
- Rees, P. (1998). What do you want from the 2001 census? Results of an ESRC/JISC survey of user views. *Environment and Planning A*, 30(10), 1775–1796.
- Stillwell, J., & Duke-Williams, O. (2007). Understanding the 2001 UK census migration and commuting data: the effect of small cell adjustment and problems of comparison with 1991. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2), 425–445.
- Tobler, W. R. (1981). A model of geographical movement. *Geographical Analysis*, 13(1), 1–20.
- Tobler, W. R. (1987). Experiments in migration mapping by computer. *The American Cartographer*, 14(2), 155–163.