Dissertation

Conceptualising and Mapping Multi-Modal Opportunity: An Application in Hertfordshire, UK.

Hannah Gumble

August 2019

This dissertation is submitted in part requirement for the Master of Science in Spatial Data Science and Visualisation in the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Built Environment, UCL.

Submitted:	30th August 2019
Supervisor:	Dr Anahid Basiri
Module Code:	CASA0012
Module Title	MSc Spatial Data Science and Visualisation Dissertation
Student ID:	17127561
Word Count:	11,299

Abstract

Multi-modal travel, where an individual uses more than one mode of travel in a single trip is increasing (Kuhnimhof, 2011). Understanding barriers of multi-modal trips (physical and non-instrumental) is limited, with literature commonly focusing on modes in isolation (Antonio Lindau *et al.* 2014). This study aims to synthesise the main barriers to undertaking multi-modal trips into a composite index of multi-modal opportunity. Multi-modal opportunity is used to encompass the factors that can influence a person's opportunity and want to undertake a multi-modal trip. Hertfordshire is a suitable study area due to plans to encourage multi-modal travel (Hertfordshire County Council, 2018).

The index contains six main domains: demographics, local characteristics, safety, connectivity, railway stations and bus stops. Spatial data are used to measure key components which form the basis of index construction. Principle Component Analysis is undertaken to create area attributes and Analytical Hierarchy Processing is used to weight the area attributes and key components for bus stops and railway stations into single values. Railway station and bus stop catchments are determined using Network Analyst and used to calculate an overall score of multi-modal opportunity.

Mapping provides insight into the variation of multi-modal opportunity across Hertfordshire, with statistically significant clusters of opportunity found in town centres. The importance of bus services was identified, as high index scores are driven by accessibility to frequently serviced bus stops. Sensitivity analysis examined the robustness of the index considering different weighting methods and principle component selection. It is concluded that principle component selection can significantly alter results.

This study provides an analysis framework that allows for the spatial nature of multi-modal opportunity to be understood, building upon previous studies by encompassing non-instrumental trip determinants. Future work should increase the granularity of analysis and understand how changes in the theoretical framework affect the outcome results.

Declaration of Authorship

I, Hannah Gumble, hereby declare that this dissertation is entirely my own original work and that all sources have been acknowledged. This dissertation is 11,299 words in length. Word count processed by Microsoft Word.

Signed:

Date: 30th August 2019

Table of Contents

Figures

Tables

Acknowledgements

Firstly, I would like to thank my supervisor Dr Anahid Basiri for her enthusiasm, guidance and feedback throughout this research. Your enthusiasm is infectious and inspired me to rise to the challenge.

I would like to extend my gratitude to all the staff at UCL's Centre for Advanced Spatial Analysis for inspiring me to challenge myself throughout this degree.

I am grateful to Hertfordshire County Council for allowing me to use their data within this project and my colleagues at AECOM for their support throughout my Masters.

I owe many thanks, and probably some apologies, to my mum, sister and partner for all their support and patience throughout this process – you were always there for me and I would not have been able to do it without you.

1. Introduction

While using public transport is an essential part of many people lives, the need for having low-cost (El-Geneidy *et al.* 2016), environmentally sustainable (Button and Henser, 2009) and accessible transport is still one of the biggest challenges of social equality (El-Geneidy *et al.* 2016). Today society is increasingly mobile, however, access to transport networks, particularly public transport services is inconsistent (Banister, 2019). Furthermore, with increasing focus being placed on sustainability and the environmental cost of travel, it is recognised that there is a need to change the approach transport planners use to tackle these issues (OECD, 2009) and multi-modal transportation is identified as a potential solution (Liao *et al.* 2010).

Increasing focus has been placed on multi-modality with the EU Transport Commissioner Violeta Bulc calling for 2018 to be the “year of multimodality” whereby legislative and policy initiatives focused on encouraging multi-modal transport behaviour and infrastructure (European Commission, 2019). Historically, the approach to accommodating growth and ensuring economic development within the UK has been to facilitate growing car travel, however, after a realisation that road network improvements could not keep pace with traffic growth a more balanced approach between public transport and car travel was promised (House of Commons, 2002). This has led to the government driving forward the importance of multi-modal studies and initiatives (House of Commons, 2002). Although, the need for multi-modal travel initiatives is stressed, the exploration and adoption are limited.

Definitions of multi-modal mobility vary across studies, but it is commonly defined as the use of more than one mode within a trip chain whereby an individual stops and changes mode in between an origin and destination (Buehler and Hamre 2016, Clifton and Muhs, 2012). Literature has identified increasing multi-modal trends, with individuals looking to combine modes within a single journey due to growing environmental awareness (Kuhnimhof *et al.* 2011). However, historically aspects of multi-modal mobility have been neglected in transport literature due to its complexity (Molin *et al*. 2016).

Understanding the barriers which may affect the ability for an individual to travel in a multi-modal manner is limited with literature focusing on a single mode (Antonio Lindau *et al.* 2014) and of those that have encompassed two or more modes in a single trip (O'Sullivan *et al.* 2000) the focus is mainly on connectivity, omitting non-instrumental factors even though there importance is stressed within literature (Stead 2001, Stradling *et al.* 2007 and Molin *et al.* 2016). Given this, there is a need to synthesise trip barriers and determinants into a comprehensive analysis framework.

The overarching aim of this research is to synthesise the main barriers to undertaking multi-modal trips into a composite index and develop an assessment framework examining multi-modal opportunity. Multi-modal opportunity is used within this study to encompass all the factors that can influence a person's opportunity and want to undertake a multi-modal trip.

This research aims to answer three principle research questions:

- What factors encourage multi-modal travel and how can they be measured using spatial data?
- Can the trip determinants be used to develop a composite index for multi-modal opportunity?
- Can this framework provide any insights into spatial variation in multi-modal opportunity in Hertfordshire?

These will be achieved through:

- Undertaking a review of literature to identify key travel determinants and barriers for individual modes that can form part of a multi-modal trip chain and multi-modal journeys as a whole.
- Examining and measuring the trip determinants using spatial data available and defining proxy measurements.
- Creating a composite index of multi-modal opportunity.

- Undertaking spatial analysis to derive insights into multi-modal opportunity within the study area.

This study focuses on the county of Hertfordshire, which as part of local transport plans has identified multi-modal initiatives as key strategy for the county (Hertfordshire County Council, 2018). Key components of multi-modal opportunity are calculated using a mix of publicly available data and data obtained through Hertfordshire County Council. These components form six principle domains: local characteristics, safety, local potential, connectivity, railway stations and bus stops. A high-level flowchart of the methodology undertaken is shown in Figure 1.

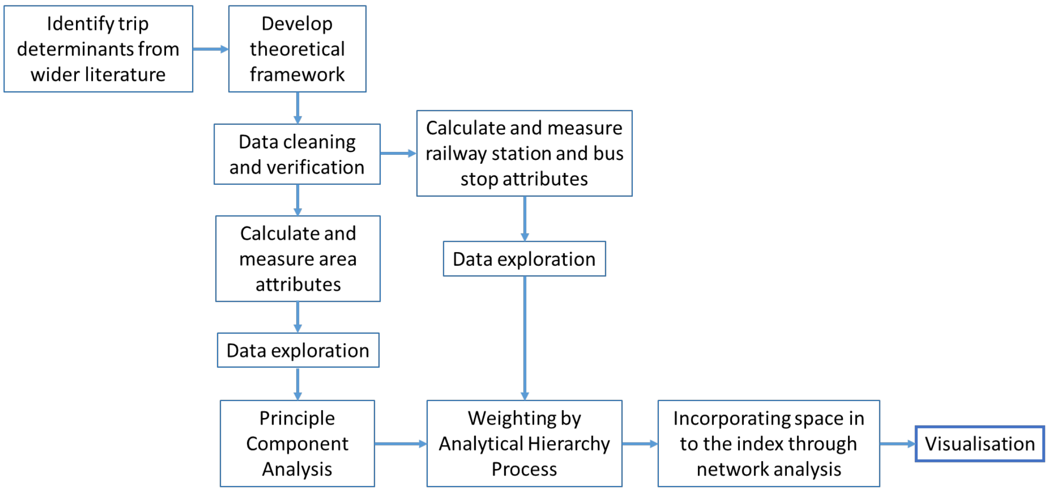


Figure 1: High level study methodology.

This dissertation contains seven chapters and is structured as follows: The second section outlines key literature surrounding multi-modal mobility discussing definitions of multi-modal travel, trip determinants and modelling multi-modal mobility. This allows a gap in the literature to be found and shape the aims this dissertation seeks to address. The third chapter introduces the study area and data sources used. The forth chapter gives a detailed description of the methodology for calculating the indicators and the construction of the composite index, which is an indicator for multi-modal opportunity. Results are then presented alongside some spatial analysis in the fifth chapter. Chapter six interprets the results and the last chapter draws conclusions, discusses the limitations of this work and suggests future work.

2. Literature Review

2.1. Introduction

The implementation of multi-modal transport schemes in urban areas to help deliver long-term sustainable development is widely discussed (Liao *et al.* 2010). Multi-modal mobility has the potential to reduce congestion, ease overcrowded transport corridors and mitigate the environmental impacts of travelling (European Commission, 2013). Despite an agreement on the importance of multi-modal mobility for sustainable cities, there is not much of an agreement on the definition within the literature. Kramers (2014) uses the term 'multi-modal' to describe when different modes of transport are used in different combinations. While Molin *et al.* (2016) use the term to describe when an individual consciously chooses a mode of transport based on trip context rather than habit. Literature also discusses both inter-modal and multi-modal travel rather interchangeably. Nobis (2007) states multi-modality is the use of different modes within a given time frame and uses the term inter-modal to describe individuals who use different modes within a single trip. Within this study the term multi-modal is going to define where an individual changes mode between the origin and destination aligning with Krygsman and Dijst (2001).

The following review outlines literature within the field of multi-modal mobility encompassing trends, travel determinants, accessibility and barriers as well as measuring multi-modal transportation and trip determinants.

2.2. Multi-modal travel

Multi-modal travel behaviour is increasing, with individuals adapting their mode dependant on the trip being undertaken. Kuhnimhof *et al.*(2011) explore trends of young British and German travellers to uncover that 20-29-year olds are less car orientated than previous generations and inhibit growing multi-modal behaviour. This is supported by both the British and German National Travel Surveys which have reported a reduction in private car mileage coupled with a growth of mileage in other modes. It is hypothesised that this is due to increased knowledge and environmental awareness (Kuhnimhof *et al.* 2011).

Inter-modal trips can be broken down in to a minimum of three stages: access, main and egress (Kagerbauer *et al.* 2015). It is widely cited that walking is the most typical access mode followed by cycling, with these modes accounting for a total 80% of access trips. Walking is the most common egress mode as travellers do not commonly have access to private modes at this end of the trip (Krygsman and Dijst, 2001). Kagerbauer *et al.* (2015) identified that in Germany 22% of trips are executed with more than one mode and many have a trip length between 4.5-12kms. Evidence from stated preference surveys and travel diaries suggests walking stages are commonly less than 2.5 km and cycling stages are normally <5km in multi-modal trips (Krygsman and Dijst, 2001). Although there is recognition of growing multi-modal trends, there is a distinct lack of literature that moves beyond focusing on a single mode in a trip chain (Chen *et al.* 2008).

2.3. Trip determinants/barriers

Human activities are spread across time and space, which transportation aims to facilitate. Given this, an individual has specific points they could occupy during each travel episode termed 'potential path-space' (Miller, 1991). However, travel choice is constrained through sequential decision making undertaken by the traveller (Kramers, 2014). Central to understanding how to facilitate multi-modal journeys is the need to underpin factors that enable and impede transport networks and affect an individual's choice (Hägerstand, 1970). Multi-modal journeys place influential space-time constraints on travellers through transfers and increased travel time (Krygsman and Dijst, 2001). Furthermore, Meyer de Freitas *et al.* (2019) recognise that intermodal travel requires greater effort and planning from the traveller coupled with a need for flexibility and commonly lower levels of comfort. Bandland *et al.* (2010) argue that a better understanding of the associations between mode choice, transport networks and local characteristics will help to refine infrastructure provision, increasing the use of sustainable modes.

2.4. Physical Constraints

Physical accessibility is a key determinant influencing modal choice and multi-modal travel. Oostendorp *et al.* (2019) uncovered that intermodal travel behaviour is spatially variable, influenced by local environments and spatial structures, therefore there is a need to understand local context when examining multi-modal travel. Sanchez (2004) outlines five factors which affect the use of public transport:

- Network characteristics,
- Network extensiveness,
- Network connectivity,
- Physical Access; and
- Vehicle Ownership.

Although a network can be well connected, it may not adequately serve the geographic area. Garćia-Palomares *et al.* (2013) acknowledge the effect of ‘distance decay’, whereby the further away an individual is, the less likely they are to use that public transport access point (Sanchez, 2004) and consequently embark on multi-modal journeys. The threshold for decay is acknowledged to be variable between different socio-demographic groups (Garćia-Palomares *et al.* 2013). Distance decay links heavily to street design and network density. The street layout of an area heavily affects routing and connectivity; however, many new developments are characterised by cul-de-sacs offering few route choices and subsequently deterring walking trips (Saelens *et al*. 2003).

Network connectivity including transfers and subsequent waiting times are especially negative for trip satisfaction and can commonly deter travellers (Krygsman and Dijst, 2001). The logistic complexity of determining the best multi-modal travel option, considering potential connectivity problems leads individuals to select mono-modal options (Clauss and Döppe, 2016). König (2002) argues it is not necessarily the waiting but the uncertainty around when the mode will arrive. Furthermore, it has been identified with bus services, users are unwilling to change service unless it is convenient and quick (Beirão and Cabal, 2007). This aligns with Kuhnimhof *et al*. (2011) finding that in Helsinki and Finland 44% of individuals always chooses the fastest mode when travelling. Furthermore, National Academies of Sciences, Engineering, and Medicine (2014) discuss the importance of waiting environment quality including comfort, lighting, weather protection and cleanliness in facilitating modal transfers.

The impact of automobiles on public transportation and multi-modal trips is debated. Distance to the nearest transit stop is a key determinant of car use (Chen *et al.* 2008). However, contrasting to the associations commonly drawn between public transport ridership and car travel, Buehler and Hamre (2016) presented evidence suggesting multi-modal trips are not wholly associated with owning a private vehicle.

As well as access, Badland *et al.* (2010) stress the importance of egress constraints particularly for work based multi-modal trips. This is supported by Krygsman and Dijst (2001) who found a strong negative correlation between increasing egress distances and multi-modal travel, as during trip egresses there is a greater emphasis on walking as private modes are not available.

2.5. Non-Instrumental Constraints

Alongside network characteristics mental barriers and non-instrumental factors also effect travel choice. Stead (2001) uncovered that within Britain socio-economic characteristics play a greater role on mode choice than quality of the built environment. Stradling *et al.* (2007) uses 'non-instrumental' factors to describe, the presence of trees, crowding, air quality and local demographic variables. Knowledge has also been identified as a barrier to choosing public transport, as car drivers typically over estimate journey times by an average of 46% (Molin *et al.* 2016). An added complexity to understanding non-instrumental factors is that they are dependent on individuals, varying by age, gender and socio-demographic background (Susilo and Cats, 2014). An example being safety and uncertainty on service reliability which is a key trip satisfaction determinant for women (Tranter, 1995). Understanding safety not only includes actual but also perceived safety (Tumlin, 2012). The perception of safety on undertaking walking trips has been examined through measuring land use mix, street connectivity, housing density and local crime occurrence (Paulley *et al.* 2006). Paulley *et al*. (2006) stress that an individual's fear of crime is often disproportionate to crime occurrence supporting Scheider and Kitchen's (2007) assertion that they should be examined separately.

Physical and mental barriers are widely discussed, however, Susilo and Cats (2014) reflect that many studies focus on one trip stage, subsequently neglecting multi-modal journeys. Given this, there is limited research on the impact of interchange facilities and transitions between different modes going beyond describing important characteristics. Gaining a clearer understanding of the impact of interchange facilities and transitions is central to understanding and facilitating multi-modal trips.

2.6. Towards modelling trip determinants and multi-modal trips.

Modelling and measuring multi-modal transportation is inherently complex due to the multi-dimensionality and difficulties obtaining data (Fiorenzo-Catalano *et al*. 2004, Clifton and Muhs, 2012). Modelling must encompass different choices, journey legs, modes and interchange facilities. However, as well as understanding multi-modal trip patterns, the opportunity a place provides to undertake multi-modal travel needs to be understood.

Several studies have aimed to evaluate the service quality of multi-modal networks. O'Sullivan *et al.* (2000) created isochrones to combine aggregate accessibility measures creating a representation of how individuals perceive transport opportunity within Glasgow's bus and rail network (Figure 2). The isochrones are calculated based on elapsed travel time however did not encompass other recognised travel determinants due to the complexity. Furthermore, Friederick (2016) measured service quality of multi-modal networks encompassing trip time ratio, direct speed of the service and a detour factor creating functions focused on travel time and speed.

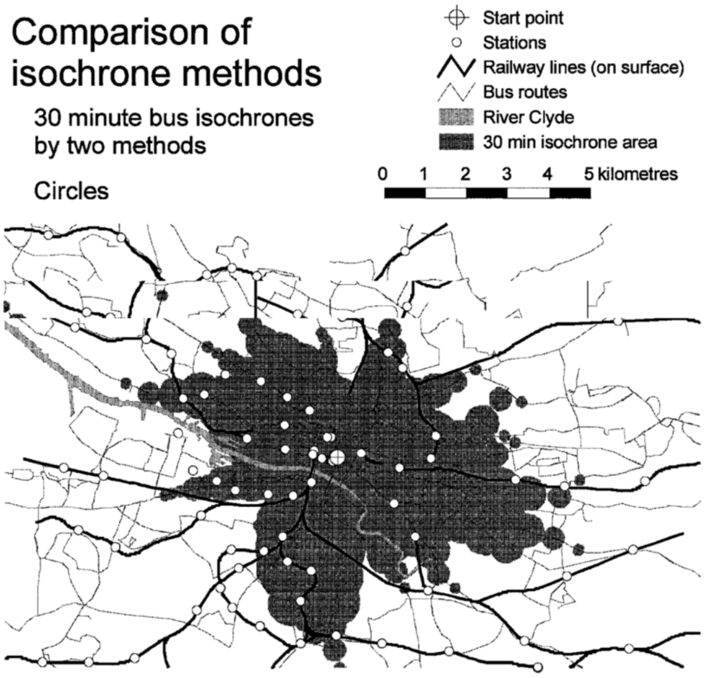


Figure 2: Perceived Transport Opportunity in Glasgow constructed through aggregate accessibility measures (O'Sullivan et al. 2000).

One of the most comprehensive assessment methods for public transport in London is the use of Public Transport Accessibility Levels (PTALS) developed by Transport for London (TFL) shown in Figure 3. Although it does not examine connections and interchanges, it is the most comprehensive analysis framework developed for analysing public transport accessibility. PTALS measure accessibility to transport network points (bus stops, underground stations and rail stations) accounting for service availability and walking time (TFL, 2010). PTALS have been developed using parameters including:

- A maximum walk time of 8 minutes or 640m distance for bus stops,
- A maximum walk time of 12 minutes or 960m distance for railway stops,
- An assumed average walk speed of 4.8km; and
- An average waiting time is calculated based on service frequency and a service reliability factor. (TFL, 2010)

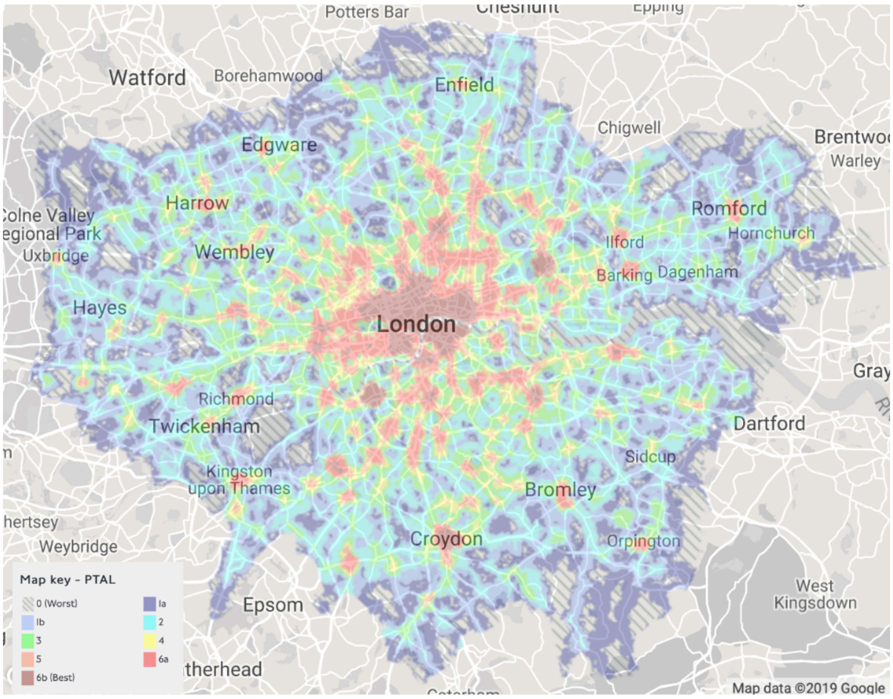


Figure 3: A map illustrating PTAL values for London (TFL, 2019).

In addition to PTALS, TFL have also developed a calculator of Public Transport Access in London (CAPITAL) to assess travel time between zones including access to the transport network within the origin and destination zones (TFL, 2015). CAPITAL uses an assessment of all strategic access and egress points relevant to the origin and destination zone selected and selects the route with the shortest journey time. Travel times are calculated using strategic models and results presented in a choropleth map (Figure 4).

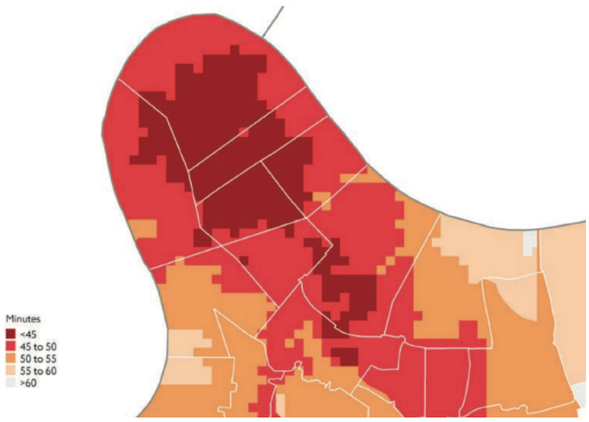


Figure 4: Example CAPITAL Output for Greenwich Peninsula to central London (TFL, 2015).

As an assessment method, PTALS do not include speed or utility of services, and although CAPITAL assesses journey time neither method assesses the ease of interchange, the ability to board services or 'non-instrumental' constraints. Furthermore, the use of PTALS and CAPITAL in areas other than London is somewhat limited, as London is characterised by its dense interconnected public transport system with over 12,000 access points where the speed and utility of services is not as crucial due to its frequency and modal connectivity.

Public transport connectivity has been modelled in Auckland using quality-of-transfer measurements (Hadas and Ranjitkar, 2012). These measurements encompassed travel time attributes: ride time, wait time and walk time but also transfer attributes including non-walk transfers, street-crossing transfers and sidewalk transfers. Through modelling transfers, it allowed for a more comprehensive look at multi-modal networks focusing on interchanges.

Curtis *et al.* (2012) developed a tool to assess the relationship between public transport network performance and land use activities to create a composite indicator of overall transport accessibility named SNAMUTS (Figure 5). The composite indicator includes activity node catchments, travel impediment (a proxy developed from travel time and service frequency) and operational aspects including: nodal connectivity, number of transfers and closeness. Through their work the need for a visual representation of multi-modal accessibility to facilitate understanding was emphasised.

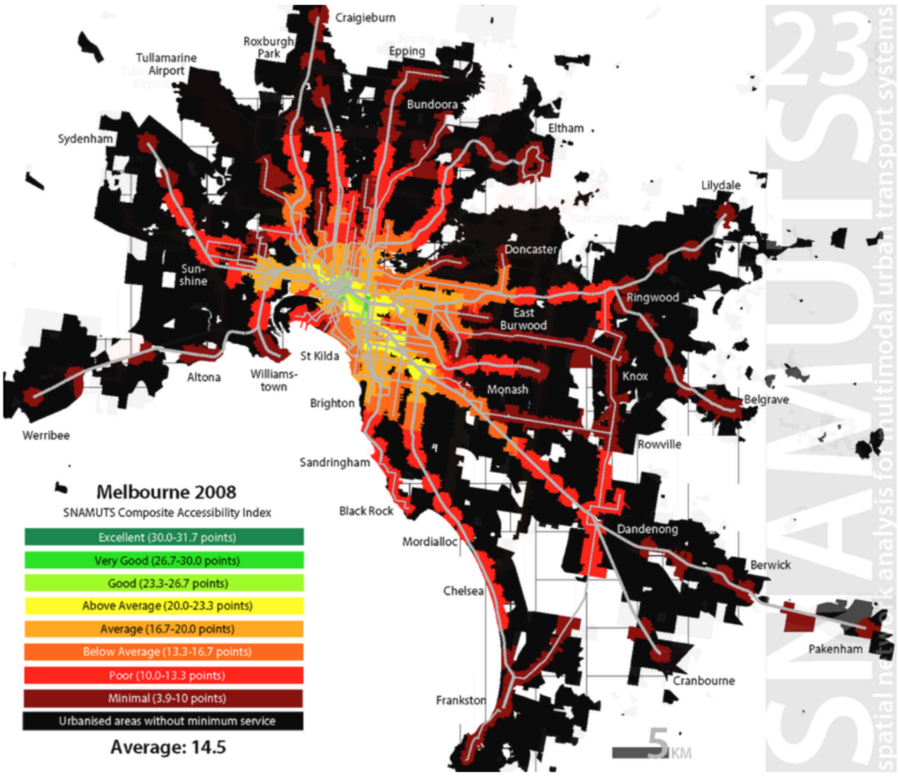


Figure 5: Example SNAMUTS output for Melbourne (Curtis et al. 2012).

Furthermore, work has been undertaken to understand multi-modal routing (Salonen *et al.* 2014). Public Participatory GIS (PPGIS) was used as a tool to understand patterns of multi-modal suburban mobility. PPGGIS is the creation of tools to assemble and collect geographic information provided voluntarily to support the inclusion of the general public into planning decisions (Brown, 2012). PPGIS allowed individual travel chains to be examined and uncovered the importance of flexibility in routing facilitating the user to be adaptable (Salonen *et al.* 2014).

It can be identified that different ways of measuring accessibility, connectivity and service quality of public transport and multi-modal networks have been developed. All methodologies have encompassed different variables with many omitting non-instrumental factors. This highlights a gap in the literature to collate an overall synthesis of multi-modal trip determinants and develop a subsequent analysis framework.

2.7. Conclusion

Multi-modal journey planning is often overlooked, focusing on modes individually. Clauss and Döppe (2016) highlight that in urban areas travellers show increasing willingness to combine multiple modes of transport in to one journey, however, options are not always well established. While extensive literature exists on public transport accessibility, the benefits of and the need to facilitate multi-modal journeys there is a lack of defined methodology for establishing the opportunity an area provides an individual to travel in a multi-modal manner. The literature covers a wide range of travel determinants, barriers and facilitators however, it is important to comprehensively synthesize these findings. Through understanding and quantifying variables discussed a succinct framework can be developed to help refine infrastructure provision and policies with the aim of increasing multi-modal travel.

Given this, the overarching aim arising from the literature is to develop a conceptual framework for assessing multi-modal transport opportunity.

3. Study Area and Data

This chapter outlines characteristics of Hertfordshire, UK, the study area for this investigation and the data sources used.

3.1. Study area

The availability of data in Hertfordshire posed a key opportunity to explore multi-modal opportunity. Hertfordshire is comprised of 10 districts with a high number of medium sized towns and dispersed population highlighted through the counties Lower Super Output Area Boundaries (LSOA) (Figure 6).

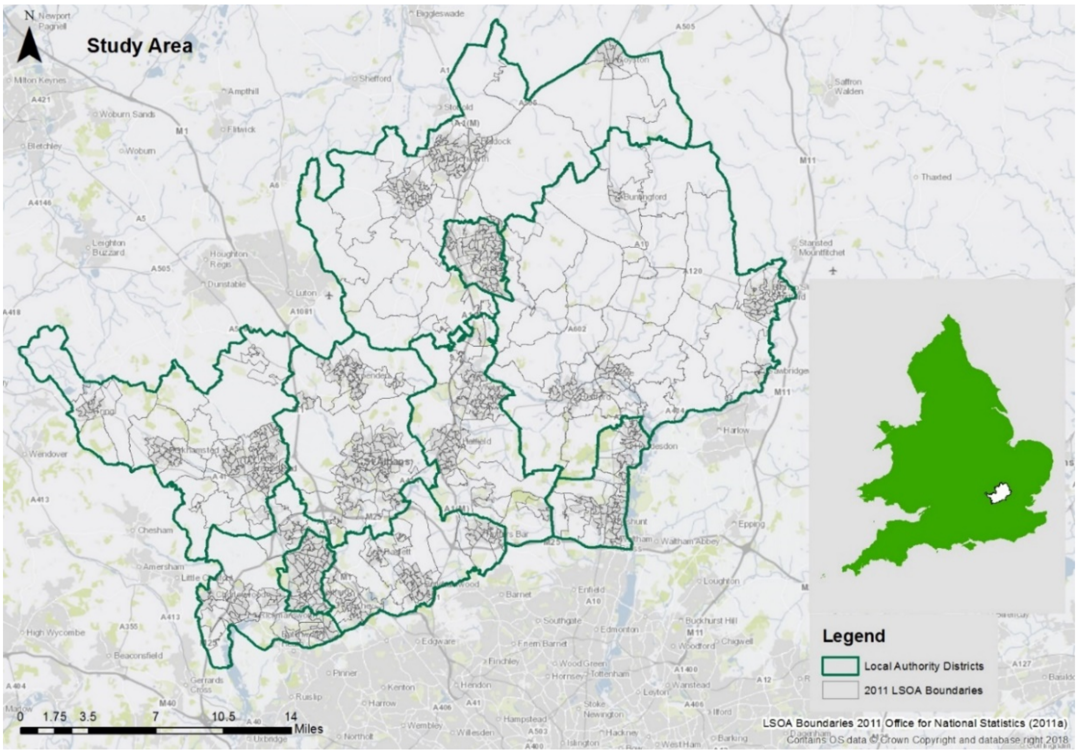


Figure 6: Hertfordshire, UK with associated LSOA and District Boundaries.

The disaggregate nature of the towns creates complex travel movements, facilitated by good North-South transport infrastructure but poor east-west connections. Hertfordshire is characterised by high levels of car ownership and rising numbers of air quality management areas (Hertfordshire County Council, 2018). By 2031 the county is predicted to have up to 18% more car trips in the peak periods and the average travelling speed by car will reduce by 15% (Hertfordshire County Council, 2018). Transportation has a key role to play in facilitating long term economic growth and mitigating worsening environmental conditions. Considering this, Hertfordshire's Local Transport Plan outlines a key challenge for the county is to encourage walking, cycling and passenger transport (Hertfordshire County Council, 2018). This need to encourage greater sustainable travel behaviour has posed a unique opportunity to analyse the multi-modal opportunity in the county and priorities that need to be addressed to encourage multi-modal travel.

3.2. Data Sources

Data used within this project is outlined in Table 1. Public and private datasets have been used, access to private datasets was granted by Hertfordshire County Council and included the use of Ordnance Survey datasets under a Public Sector End User License Agreement.

Table 1: Overview of Data used in this study.

Data	**Source**	**Data Overview**
Lower Super Output Area Boundaries (LSOA)	Office for National Statistics (2011a)	LSOA's are a geographic zoning system developed to be consistent in population size. LSOA's were chosen as the base zoning system due to the availability of statistics and the increased detail in areas with larger populations.
LSOA population weighted centroids	Office for National Statistics (2011b)	LSOA Population Weighted Centroids represent the spatial distribution of the population in each LSOA.
Census data	Office for National Statistics (2011c)	2011 LSOA Census data was extracted from NOMIS for: • Approximated Social Grade (QS611EW) • Highest Level of Qualification (WD501EW) • General Health (WD302EW) • Method of travel to Work (QS701EW) • Population Density (QS102EW) • Car or Van availability (KS404EW)
CCTV cameras	Hertfordshire County Council	A shapefile of CCTV and traffic cameras was provided. Attributes include: location and speed limit.

CCTV cameras	Hertfordshire CCTV Partnership (2019)	CCTV Camera locations were extracted from Hertfordshire CCTV Partnership Ltd online map.
Bus routes	Ordnance Survey	A shapefile was provided of 2,475 bus routes and included the following attributes: operator name, service number, days of operation and start times.
Bus stops	Hertfordshire County Council	Bus stop locations were provided in a shapefile.
Cycle ways	Ordnance Survey	A shapefile containing cycle routes and a breakdown of carriage way types.
Pedestrian and cycle casualties	Hertfordshire County Council	Data from 2013 – 2017 was provided including day of the week, lighting conditions, co-ordinates and severity.
Street Lights	Hertfordshire County Council	A shapefile of street lamps locations.
Road speed limits	Hertfordshire County Council	A shapefile of roads with the enforced speed limit.
Traffic calming measures	Hertfordshire County Council	A point shapefile of traffic calming measures with a type classification.
Pedestrian footpaths	Ordnance Survey	Pedestrian footpaths were provided in a shapefile. Attributes include: path type, name, district and the year the footpath was last surveyed.
Rights of way	Hertfordshire County Council	The right of way network was provided in a shapefile. Attributes include: road type, hierarchy and road class.
National Public Transport Access Nodes (NaPTAN)	Department for Transport (2014)	NaPTAN contains all public transport access points. Railway stations have been used from this dataset.
Crime	Home Office (2019)	Crime data was extracted for the Hertfordshire Constabulary for 2018. Data includes a crime ID, month, latitude and longitude, location, LSOA code, crime type and outcome category.
Land use	Ordnance Survey	Land use classification shapefiles were provided for each district in Hertfordshire.
Fix My Street	FixMyStreet (2019)	Fix My Street is an online platform that allows individuals to report problems in their local area. Data for 2013-2018 has been used.

Rail services	Recent Rail Times (2019)	Rail service data was extracted for all train companies serving the station. Data was extracted for the last four weeks on 8/06/2019 where the service ran more than once within the period. Data includes the average and scheduled arriving time, journey duration and percentage of services cancelled. Services were selected based on frequent final destinations. Data format can be seen in Appendix A.
Bus timetables	Interlink (2019)	Extracted for each route included in this study (please see methodology). Average service frequency by time period was extracted manually for each route from timetable information online.
Railway station facilities	National Rail Enquires (2019)	Station facilities were extracted and stored in a csv. Attributes included: • Cycle parking facilities • CCTV Cameras covering cycle parking • Waiting room facilities • Baby changing facilities • Toilet facilities.

3.3. Ethical Considerations

All data used within this this study has been stored on a password-protected computer, which has only been accessed by the author. Datasets regarding people have been aggregated by data collectors and therefore individuals cannot be identified from the dataset or analysis. A low risk statement of ethics application has been made as Analytical Hierarchical Process (AHP) questionnaires are used in the methodology. This study has been approved by the UCL Research Ethics Committee (Project ID Number: 16169/001). A copy of the participant information sheet and AHP questionnaire can be found in appendices B and C.

4. Methodology

The OECD (2004) defines composite indicators as a measure of a multi-dimensional concept that cannot be captured by one single indicator. Derived from an overarching analysis framework, indicators are identified, measured and combined to reflect the dimensions of the phenomena. In this study, factors which inhibit and affect an individual's opportunity to undertake multi-modal travel are measured and combined to create an index of multi-modal opportunity. The following chapter works through the methodology outlined in Figure 7, which incorporates key stages from the OECD (2008) methodology for creating composite indices.

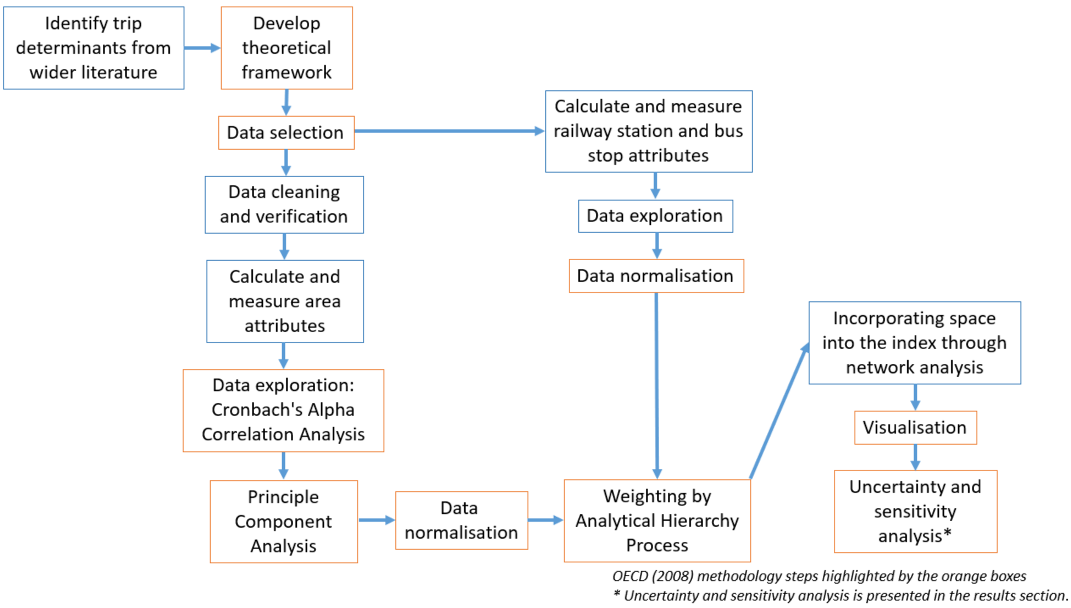


Figure 7: Methodology used to create the composite index of multi-modal opportunity.

4.1. Theoretical framework

The theoretical framework for understanding multi-modal opportunity, developed from the wider literature, is illustrated in Figure 8. Six overarching domains have been identified, which have been subdivided into area attributes and attributes that are only applicable in specific locations for a certain proximity. Area attributes include safety, network connectivity, local demographics and local characteristics. These attributes encompass key components including land use mix, local road speeds and CCTV coverage all of which are measured at an LSOA level. The location of bus stops and train stations determine their impact on multi-modal travel opportunity; therefore, they are not included in the overarching area attributes and treated separately in the location and proximity sub-division.

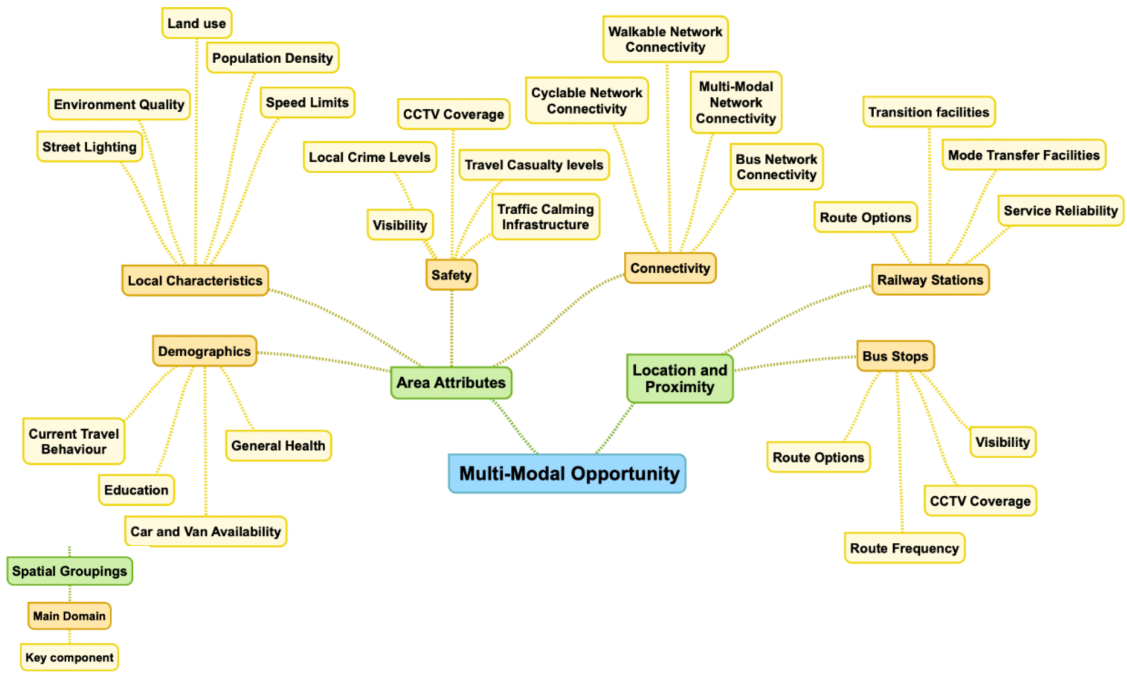


Figure 8: Multi-modal opportunity theoretical framework.

4.2. Data Preparation

To prepare the data for the composite index datasets were cleaned and validated before calculating the key components. The limitations of the data and calculation methods chosen are then reflected upon and outlined in the assumptions.

4.2.1 Data Cleaning

The datasets used within this project were available in several different formats including: ArcGIS shapefiles, XML files, spreadsheets and PDFs. Therefore, cleaning and validation had to be undertaken for each dataset individually. Some datasets used i.e. Census and casualty data needed little cleaning. However, datasets such as the railway data and bus data needed comprehensive cleaning. Data cleaning steps are outlined below.

Train time Data

Train times were extracted manually for each railway station and end of line destination into excel files. Data files were combined based on their daily coverage (Monday to Friday, Saturday and Sunday). Due to the volume of data and the need for data manipulation, data cleaning was undertaken using Pandas 0.23.0 and NumPy 1.14.3 packages. The steps undertaken were:

- Removal of special characters,
- Ensuring data types are correct,
- Identification and removal of missing data and outliers; and
- Identification and removal of any duplicate data.

Bus Route Data

Initial data validation checks on bus routes highlighted several issues including services no longer existing and incorrect routes. Furthermore, service frequency is linked to facilitating multi-modal travel behaviour, given this, all routes in the shapefile were checked and excluded if one of the following criteria was met:

- The bus service only ran specific days of the week,
- The bus route was a school only service,
- The route does not match the route online or;
- No corresponding service information could be found.

As a part of data validation, average waiting times were manually extracted and calculated for the AM, Inter-Peak and PM periods which run from 7am-10am, 10am-4pm and 4pm-7pm respectively. Data for these were extracted from Interlink timetable search (Interlink, 2019) or the individual bus operator’s website where the service did not feature in the Interlink database. Commonly the bus operator noted in the raw data was out of date. In these instances, bus services were included if the route number and route matched the online timetable. A key shortfall of this method is that if a route has been renumbered but the route has remained the same it is not included in this study. For the calculation of average waiting times, the route must have a minimum of two services running in the period and was calculated using the departure time from the first station on the route. An example calculation is shown in the Figure 9:

7	7A	7	7A	7	7A
1030	1130	1245	1330	1500	1550

Route 7A – 11:30, 13:30, 15:50 – Maximum waiting time of 130 minutes.

Route 7 – 10:30, 12:45, 15:00 – Maximum waiting time of 135 minutes.

Figure 9: Example maximum waiting time calculation between services for route 7 and 7A operated by Trustybus (Interlink,2019).

A total of 508 services were provided in the shapefile covering a total of 2,457 routes. Post-cleaning a total of 122 services remain covering 255 different routes.

Bus Stops

In total, the location for 13,604 bus stops was provided. However, due to the cleaning of the bus routes the bus stop data has to be cleaned to remove stations with no associated route. Cleaning of the shapefile was undertaken in ArcGIS. Due to the positions of the bus stops and routes, the two feature classes could not be intersected. A buffer of 12 meters was applied to each bus stop before an intersect was used to capture all the stops along the routes being considered. A 12-meter buffer was used to incorporate the average road width (5.5m-7.3m for a single carriage way with two lanes) and the understanding that bus stops are ideally 3 meters offset from the kerb for safety (TFL, 2017). If no routes intersected the bus stop, it was removed. It is acknowledged that bus services may not stop at every stop along the route, particularly for strategic services. However, no information could be sourced to account for this.

Walkable and Cyclable networks

To calculate the key components walkable and cyclable networks were created. Although, a highway network shapefile was provided, not all roads are suitable for cycling or walking. Pedestrians and cyclists are legally allowed to walk/cycle along any road apart from a motorway or motorway slip road (Department for Transport, 2019). Given this, all motorways were removed and then, the highway network was integrated with pedestrian footpaths to create a ‘walkable’ road network and cycle paths to create a ‘cyclable’ network.

4.2.2 Key Component Processing

From the literature review several determinants of multi-modal travel were uncovered. These have been grouped into the following sub domains: local characteristics, local potential, connectivity and safety and have been calculated at an LSOA level. The indicators considered when forming each domain, and the methodology used to calculate them is presented in Table 2 to Table 5. These methodologies draw upon ArcGIS tools including:

- Buffering – Buffering creates a polygon of a specified radius around a feature (Longley *et al.* 2015).
- Clustering – Density based clustering identifies spatially significant hotspots and outliers based on a feature’s spatial distribution (Esri, 2018a).
- Spatial Join – Spatial join unifies the attributes of two feature classes based on their spatial relationship (Esri, 2018b).
- Dissolve – Dissolve aggregates features based on a specified field (Esri, 2018c).
- Intersect – Calculates where two or more geometric attributes overlay each other (Wade and Sommer, 2006).

Table 2: Local characteristics indicators and methodology.

Indicator	***Rationale***	***Methodology***
Population density	Population density has a strong link with public transport use and service viability (Cooke and Behrens, 2017).	2016 population was divided by the area (in hectares) for each LSOA.
Local road speed limits	Road speed limits affect walkability, cyclability and perceptions of safety (Department for Transport, 2009).	The percentage of highway network where the speed is <40mph was measured.

Land use mix	Land use mix can affect perceptions of safety for walking trips (Paulley *et al.* 2006).	Land use mix was calculated using an Entropy measure. $$\frac{-\sum(A_{ij}LnA_{ij})}{LnN_j}$$ Where: N_j= Is the number of land uses in the LSOA A_{ij} is the percent of the land use *i*, in area *j* (Manaugh and Kreider, 2013). Land uses were reclassified into aggregate groups (appendix D) using the field calculator and dissolve tool before calculating the entropy.
Illumination ratio	Pedestrians generally feel safer where there is adequate street lighting. (Department for Transport, 2009).	Walkable and cyclable network illumination ratios were calculated for each LSOA. Street lamps were buffered by a lighting range of 87m based on the average distance an individual can see under typical luminance (Crabb *et al.* 2008). Total of paths sufficiently lit was divided by the total paths to create a ratio.
Local environment quality	Quality of the local environment affects a person's willingness to travel through that area.	The relevant 'FixMyStreet' variables were filtered (appendix E). Total reports were standardised using the 2016 population for each LSOA and multiplied by 1000 to produce a rate per 1000 individuals.

All indicators used to define local potential (Table 3) were extracted from the 2011 Census. Given this, the indicators were processed using the same two steps, the relevant variables were identified, and values extracted for Hertfordshire LSOA's only.

Table 3: Local potential indicators and methodology.

Indicator	**Rationale**	**Census variable used**
Motorised vehicle availability	Vehicle ownership affects the perception and knowledge of public transport (Molin *et al.* 2016).	Percentage of households without access to a private vehicle.
General health	An individual's health affects their capabilities to access public transport and complete access	Percentage of the population that have good, very good or fair health.

	and egress portions of multi-modal trips.	
Method used to travel to work	Proportion of people not using cars to travel to work are more likely to incorporate a multi-modal lifestyle.	Percentage of people travelling to work who are not car drivers, passengers or unemployed.
Education	Educational attainment has been linked to a greater use of multi-modal travel (Kuhnimhof *et al.* 2011).	Percentage of the population with formal qualifications.

Table 4: Safety indicators and methodology.

Indicator	**Rationale**	**Methodology**
Pedestrian casualties Cycling casualties	Pedestrian and cyclist accidents affect perceptions of safety.	Number of incidences were calculated using a 'one to many' spatial join and dissolving based on the LSOA. This was normalised using the population of each LSOA. Density-based clustering identified clusters of accidents. The total length of each cluster was summed and standardised by the total road network in each LSOA. The outcome of this was multiplied by the number of incidences.
Street lighting	Individuals can feel more secure when transit stops are well lit (Department for Transport, 2009).	Streetlights per LSOA were counted using a 'one to many' Spatial Join and standardised by LSOA area.
Traffic calming measures	Traffic calming measures help prevent casualties and increase safety (Department for Transport, 2009).	The percentage of highway network where there are traffic calming measures was calculated.
Surveillance ratio	CCTV cameras create a sense of safety.	CCTV cameras were buffered by 50m and a spatial join was undertaken to identify cameras likely to cover the network. Filtered CCTV cameras were buffered by a visibility range of 200m as CCTV can commonly identify objects within 300 meters (Juan Park, 2012). While traffic cameras can see up to 2 miles (RAC, 2019) the view is often disrupted due to road geometry and buildings.

		The total paths sufficiently captured divided by the total paths was then used to calculate a ratio.
Crime rate	Crime can deter use of public transport, walking and cycling (Scheider and Kitchen, 2007).	Relevant crime types were filtered. Total crime occurrence was counted and standardised by the 2016 population for each LSOA.

Table 5: Connectivity indicators and methodology.

Indicator	***Rationale***	***Methodology***
Walkable network connectivity	Network connectivity is identified as a key determinant in undertaking multi-modal trips (Sanchez, 2004).	Connectivity was measured using link-node ratios whereby the number of links is divided by the number of nodes within each LSOA (Tresidder, 2005).
Bus network connectivity		
Cycle route connectivity		

The geographic distribution of bus stops, routes and railways stations heavily affects their use and their spatial distribution is not consistent. Therefore, aggregating and analysing their key components at an LSOA level is not appropriate. The key components for railway stations are shown in Table 6 and for bus stops are shown in Table 7. The key components were calculated for each railway station and bus stop individually.

Table 6: Key Components forming the railway station domain.

Indicator	***Rationale***	***Methodology***
Station amenities	Station amenities provide facilities that support a comfortable change of mode (National Academies of Sciences, Engineering, and Medicine, 2014).	Station amenities were integrated into a .csv and recoded into a binary indicator.

Train reliability-cancellations	The uncertainty of when a mode will arrive affects an individual's willingness to undertake a multi-modal journey (König, 2002).	For each service which passes through in the given time period an average delay in minutes is calculated. As services have different frequencies, the average delay was weighted by the service frequency. The following equation is applied to create an average delay factor per station. $\frac{\sum(Count\ of\ service \times Average\ Delay\ (Minutes))}{Total\ number\ of\ services\ that\ stop\ at\ the\ station}$ The same methodology is applied to calculate a service cancelation factor. Script used can be seen in (appendix F).
Train reliability – delays		
Route options	Network extent and route options affects the use of public transport (Sanchez, 2004).	All routes passing through each station were counted.
Bus service connections	Highlights ease of multi-modal transfers (Sanchez, 2004).	Railway stations were buffered by 600 meters. An intersect was performed between the buffer and bus stops to calculate connection options.

Due to the methodology chosen to combine the railway station domains into an overarching score the following equation is applied to reliability indicators to ensure stations with the least cancellations and delays have a factor close to 1.

$$Train\ reliability\ factor\ = 1 - Train\ reliability\ weighting.$$

Table 7: Key components forming the bus stop domain.

Indicator	***Rationale***	***Methodology***
Presence of street lighting	Lighting draws attention to and encourages the use of amenities (Department for Transport, 2009).	Street lamps were buffered by a lighting range of 87m. An intersect was undertaken to identify bus stops that are covered by the lighting range.
Presence of CCTV coverage	CCTV cameras create a greater sense of safety.	CCTV cameras were buffered by a visibility range 200m. An intersect was undertaken to identify bus stops that have a presence of CCTV coverage.
Route options	Route options and service frequency affect bus stop use and the multi-modal opportunity it provides.	Points to line in QGIS identified routes which may service each stop. The count of services was weighted by the frequency.

4.3. Assumptions

There are several assumptions that need to be recognised within this methodology:

- The road speed limits used are the enforced speeds not the speed travelled.
- The co-ordinate system used for all spatial analysis is British National Grid. Where the data has been provided in a different co-ordinate system it has undergone projection/transformations in ArcGIS.
- Distances used throughout analysis are not the true 3D length as British National Grid is planar projection and does not account for changes in gradients.
- Several datasets used are point data, these are subject to problems regarding precision and accuracy.
- Where point-based data has been aggregated and processed at an LSOA level, the indicators may have been influenced through the shape and scale of the aggregation unit and is therefore liable to the Modifiable Area Unit Problem (MAUP) (Openshaw and Taylor, 1997).
- Datasets used were the latest available, however, it must be recognised that there are temporal inconsistences between datasets e.g. 2016 population data has been used for data standardisations as it is the latest population estimate available, however, crime data is from 2018.
- Datasets incorporated have been cleaned and validated at the time of this study.
- Fix my street data has been used as a proxy for environmental quality. As a crowdsourced dataset it may contain demographic and spatial biases (Basiri *et al.* 2019).
- A buffer has been used as a proxy for CCTV vision, however the position and angle of the camera will affect its viewing pane and vision.
- Radius used for street lighting is indicative as it is affected by the height of the lamppost and light intensity given by the bulb (Haans and Kort, 2012). 87m is used for the buffer radius as it is the average distance an individual can see for with the presence of street lighting (Crabb *et al.* 2008).

- Although data are available on bus stop locations and bus routes, no data are available on the stops of each route. Given this, it is possible that some routes may align with bus stop that the bus does not stop at.
- After calculation all data forming the indicators for multi-modal opportunity have been standardised using Z-score data standardisation to ensure data are comparable.

4.4. Building the Composite Index

Initial data exploration is then undertaken on the area attribute key components to identify patterns and trends. Principle Component Analysis (PCA) is used as a form of dimension reduction for the key components and Analytical Hierarchy Processing is used to determine the weights for the index. Network Analyst is then used to combine the two spatial groupings into the composite index before visualisation is discussed. Table 8 outlines Python libraries used for data exploration and PCA analysis, code extracts can be seen in appendix G.

Table 8: Python Libraries used for analysis.

Python Package	***Version Number***
Pandas	0.23.0
Seaborn	0.8.1
Scikit-learn	0.19.1

4.4.1 Initial Data Exploration

Understanding the data being used to form a composite index is key before undertaking further analysis. Employing multivariate statistics allows for the structure of the data to be understood which can inform the suitability of the dataset and provide guidance for choosing the most appropriate methodology (Nardo *et al.*2005). In this instance, Spearman's Rank has been used to summarise the strength and direction of each pair of variables forming the overarching domain and a scatter plot matrix has been produced to visualise data relationships. The correlation matrix (Figure 10) highlights that population density and highway network illumination ratio have the highest correlation coefficient of 0.79, followed by 0.73 and 0.66 for cycle facilities and cycle route illumination ratio and network speeds < 40mph and highway illumination ratio respectively. The strongest negative correlation is measured between cycle leisure routes and highway network illumination ratio.

	Cycle Route IR	Highway Network IR	Foot Path IR	Environmental Quality	Speed <40mph	Population Density	Landuse Mix	Cycle Facility	Cycle Leisure Route	National Cycle Network
Cycle Route IR	1.0	0.31	0.23	0.00041	0.17	0.31	0.23	0.73	0.016	0.15
Highway Network IR	0.31	1.0	0.34	-0.19	0.66	0.79	0.2	0.27	-0.34	-0.014
Foot Path IR	0.23	0.34	1.0	-0.11	0.28	0.29	0.18	0.11	-0.079	0.014
Environmental Quality	0.00041	-0.19	-0.11	1.0	-0.32	-0.25	-0.027	0.032	0.042	0.15
Speed <40mph	0.17	0.66	0.28	-0.32	1.0	0.69	0.14	0.1	-0.21	-0.046
Population Density	0.31	0.79	0.29	-0.25	0.69	1.0	0.19	0.26	-0.28	-0.047
Landuse Mix	0.23	0.2	0.18	-0.027	0.14	0.19	1.0	0.23	-0.12	0.07
Cycle Facility	0.73	0.27	0.11	0.032	0.1	0.26	0.23	1.0	-0.26	0.093
Cycle Leisure Route	0.016	-0.34	-0.079	0.042	-0.21	-0.28	-0.12	-0.26	1.0	-0.027
National Cycle Network	0.15	-0.014	0.014	0.15	-0.046	-0.047	0.07	0.093	-0.027	1.0

Figure 10: Correlation Matrix of variables used to form the local characteristics domain.

The suitability of using correlated variables to form an index is contested. It is discussed that a high correlation among variables should be corrected, however, others consider high correlations as a feature of composite indicators (Nardo *et al.*2005). For this study, PCA is used to overcome problems of covariance and correlation between input variables.

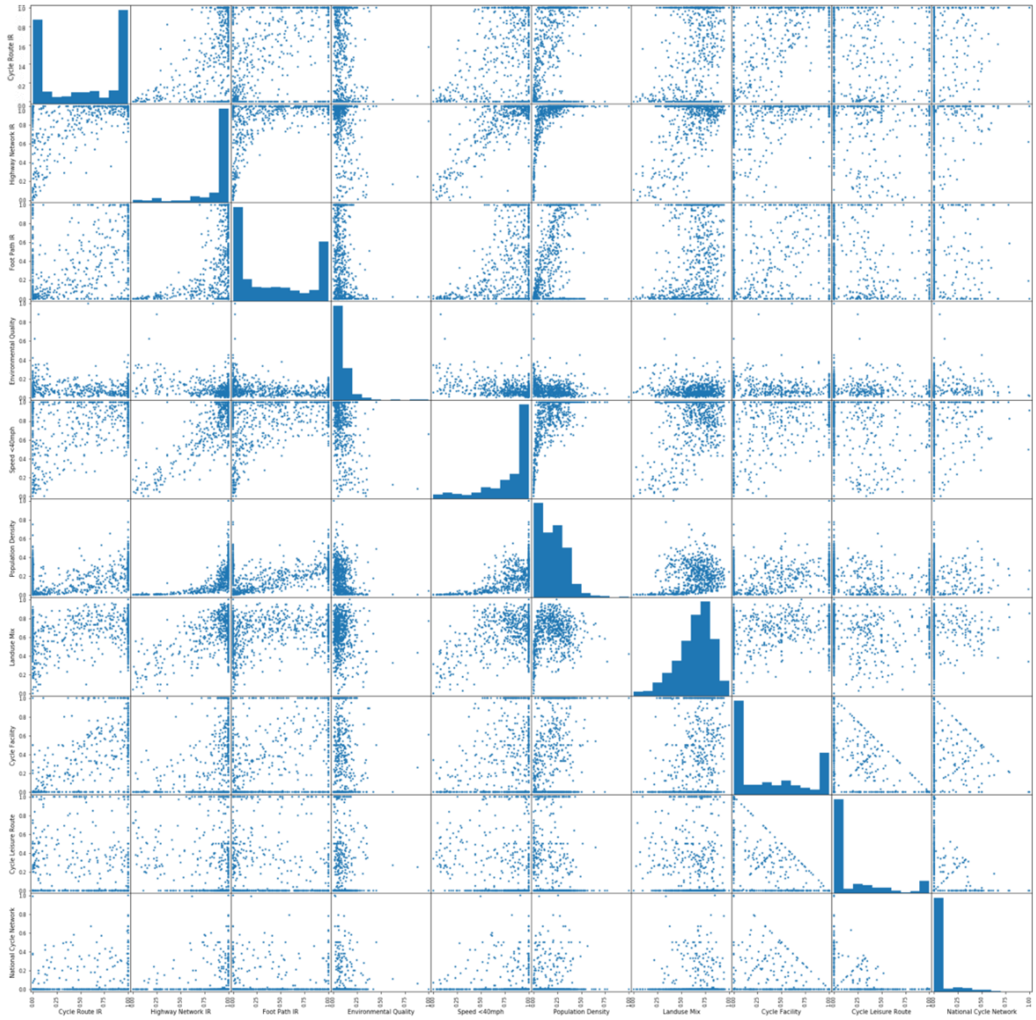


Figure 11: Scatter Plot Matrix of variables contributing to the local characteristics domain.

The scatter plot matrix (Figure 11) shows a pairwise comparison for all variables contributing to the local characteristic's domain. A weak positive relationship can be observed between population density and highway network illumination ratio and population density and road speed <40 mph. No pair of variables exhibits a negative correlation. A summary of variables with high correlations for each domain is presented in Table 9.

Table 9: Summary tables of significant correlation between variables of the same domain.

Domain	***Variable***	***Strong Correlation With***	***Correlation Coefficient***
Local characteristics	Population density	Road illumination ratio	0.79
Local characteristics	Cycling facilities	Cycle path illumination ratio	0.73
Local characteristics	Population density	Local road speed limits	0.69
Connectivity	Walkable highway network connectivity	Cycle route connectivity	0.81
Local Potential	Education	General health	0.85

4.4.2 Internal consistency

Cronbach's alpha (C-Alpha) measures how closely related a set of items are and is considered to be a measure of reliability (Nardo *et al.*2005). Cronbach's alpha is a common tool employed to measure internal consistency (OECD,2008), which can determine if variables forming a domain are statistically balanced (Nardo *et al.*2005). C-Alpha is not a statistical test to measure the reliability but to indicate how well the individual components measure the overarching phenomenon. If C-Alpha measures one, the indicators considered are perfectly correlated, a result of zero highlights all the values are uncorrelated. In its calculation (Figure 12) N is equal to the number of Items, c-bar is the average inter-item covariance among all items and v-bar is the average variance (UCLA, 2018).

$$\alpha = \frac{N \cdot \bar{c}}{\bar{V} + (N-1) \cdot \bar{c}}$$

Figure 12: Cronbach's alpha formula (UCLA, 2018)

Table 10: Cronbach's Alpha results for area domains.

Domain	***C-alpha value***	***Number of Variables***
Safety	0.96	8
Local characteristics	0.95	10
Local potential	0.92	4
Network connectivity	0.99	5

Results of the C-Alpha analysis highlight that all domains have very high C-Alpha values particularly when considering the acceptable limit of 0.6/0.7. It is cited that including a high number of variables in analysis can inflate C-Alpha (UCLA, 2018). However, for the number of variables included in each domain in this study, this is not the case (Table 10).

Helms *et al*. (2006) highlights two conditions that can result in a high C-Alpha:

1. If the variables being used to calculate the C-Alpha is an index i.e. an empirically derived composite or;
2. If there is item redundancy whereby the items that are included exhibit a narrow coverage of the construct being considered.

Variables included in analysis have been calculated from spatial data or extracted from a secondary data source and therefore do not conform to the first condition. Kline (1978) suggests that narrow coverage of a construct can be investigated using correlation analysis, where correlations greater than 0.75 suggest variables are paraphrases of each other. However, the correlation analysis highlighted in Figure 10 does not show evidence of this. Therefore, it is considered that there is a high degree of internal consistency between the variables and little redundancy.

4.4.3 Principle Component Analysis

It is common that although a multitude of variables can be considered when measuring a phenomenon, much of the variance can be captured in a small number of variables (OECD, 2008). Therefore, when constructing composite indexes, a data reduction technique is normally employed. Two common data reduction techniques used are Factor Analysis (FA) (European Commission, 2016) and PCA (Winter and Dodou, 2013).

PCA is the most popular dimension reduction technique due to its simplicity and smaller computational demands (Winter and Dodou, 2013). Preference of which method to use is debated in the literature with some arguing that there is no difference in the outcome between the two methods (Widaman, 1993) and others concluding that the outputs of PCA are closer to the true input variables compared to FA (Winter and Dodou, 2013). A key difference between the two models is the approach. PCA begins from original data to create a hypothetical model, new variables are constructed through creating linear combinations of the original data which aim to capture as much of the data variability as possible (Jolliffe and Cadima, 2016). Whereas, FA works from a hypothetical model to latent variables that explain the variance in the input data (Winter and Dodou, 2013).

PCA is undertaken for each LSOA level domain to act as form of dimension reduction in this study as the aim is to retain as much of the variance as possible. The goal of PCA is to simplify high dimensional data while retaining key patterns and trends through the identification of new orthogonal principle components (PCs) (Abdi and Williams, 2010).

PCA is an unsupervised method and therefore finds patterns without prior knowledge of the data (Lever *et al.* 2017). Scale is important in PCA; therefore, the standard scaler was used from the Sklearn Python package which scales the data so that each indicator has the mean of zero and standard deviation of one (Scikit learn, 2019).

Following data standardisation, the covariance matrix is computed and used to determine the eigen decomposition which is used to drive PCs. The covariance matrix is a symmetric matrix which stores the variance (spread) of each feature and the orientation (covariance) of the data (Alto, 2019). The eigenvectors and eigenvalues of the covariance matrix are then calculated. Eigenvalues have been termed ‘characteristic roots’ for PCs (Weisstein, 2019). The eigenvector can be defined as the orthogonal direction of the data and the eigenvalue as the magnitude of variance captured (Alto, 2019).

To understand the factors that have the greatest influence on the PCs, a correlation matrix was produced between the output PCs and the input variables. For the local characteristics' domain, nine PCs were created. The highest positive correlation can be seen between PC one and cycling facilities which has a correlation of 0.74. The strongest negative correlation of -0.78 can be observed between Population Density and PC zero (Figure 13).

	0	1	2	3	4	5	6	7	8	9
Cycle Route IR	-0.67	0.58	-0.46	0.24	-0.25	-0.058	0.024	-0.051	-0.16	0.16
Highway Network IR	-0.74	-0.1	0.11	-0.17	-0.058	0.2	0.2	-0.054	-0.055	-0.15
Foot Path IR	-0.5	-0.14	-0.13	0.078	-0.26	0.65	-0.53	-0.045	0.12	-0.019
Environmental Quality	0.24	0.44	0.066	0.1	0.27	0.42	0.15	-0.059	-0.023	0.0096
Speed <40mph	-0.62	-0.33	0.12	-0.1	-0.056	0.11	0.28	-0.23	0.23	0.3
Population Density	-0.78	-0.16	0.066	-0.23	-0.13	0.21	0.35	0.25	0.0051	0.0039
Landuse Mix	-0.4	0.075	0.04	0.43	0.64	-0.1	-0.28	0.29	0.042	0.11
Cycle Facility	-0.64	0.74	-0.29	-0.11	-0.071	-0.18	-0.074	-0.0077	0.23	-0.069
Cycle Leisure Route	0.35	-0.24	-0.6	0.57	-0.14	0.038	0.22	0.047	0.057	-0.03
National Cycle Network	-0.041	0.35	0.48	0.48	-0.26	0.015	0.033	0.029	0.077	-0.018

Figure 13: Correlation Matrix between the input variables and principle components for the local characteristics domain.

How to best select the number of PCs to use within the composite index is heavily debated. The Scree method is discussed in the literature as a way to determine the number of PCs to carry forward based on their explained variance (eigenvalues). The eigenvalues are plotted in descending order for each PCs and the identification of an elbow in the plot is used as a determining threshold on whether to keep a PC (Zhu and Ghodsi, 2006). Although this method is widely used, its subjectivity is recognised as it is dependent on individual interpretation (Zhu and Ghodsi, 2006). Scree plots were produced for each domain through plotting sklearns 'explained_variance_ratio'.

Figure 14 shows the corresponding explained variance ratio for each PC calculated for the local characteristics domain. The first PC accounts for 36% of the variation of the input variables and the second explaining 14%. On the other end of the scale, PCs 8 and 9 capture 1.5% and 1% of the variation respectively. It can be determined that 8 and 9 are noise, rather than PC's that contribute significantly to the variance captured in the raw data.

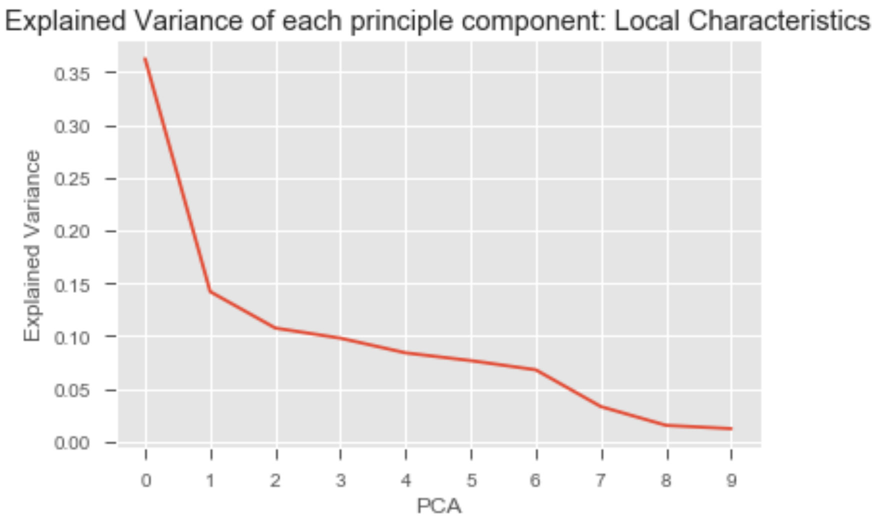


Figure 14: Explained variance ratio for local characteristics domain.

Employing the Scree method to Figure 14 is somewhat difficult. The scree method hypothesises that the explained variance contributed from PCs is greater than random noise. Therefore, an 'elbow' can be identified when there is a transition from meaningful components to those exhibiting random noise (Zhu and Ghodsi, 2006). Interpreting Figure 15, it could be determined that there are two elbows, with varying significance. The first at the second PC where all the preceding PC's explain <10% of the variance and a second at PC six where the where the proceeding PC's explain >5% of the explained variance. Depending on the cut off chosen depends how much variation is captured. Through using the first 'elbow', 60% of the explained variance is captured compared to 81% if the second 'elbow' is chosen. For this domain PCs zero to six were carried forward as after PC six the explained variance diminishes.

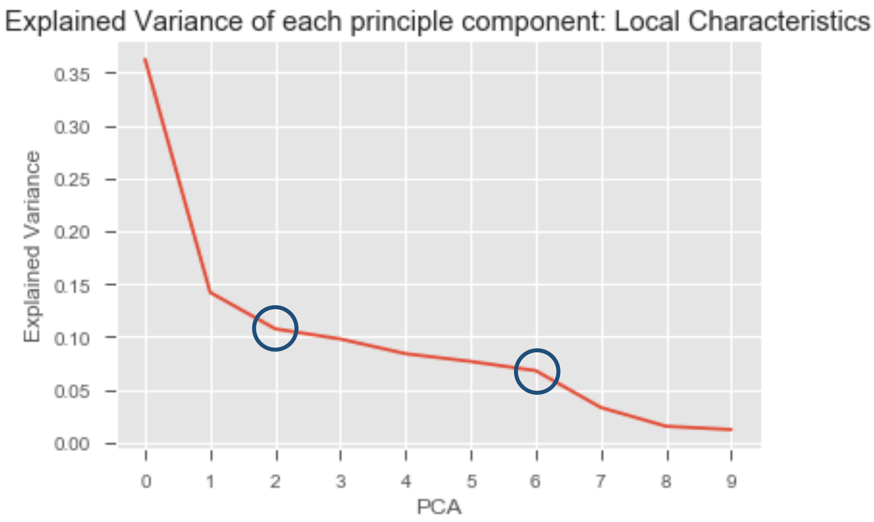


Figure 15: Explained variance ratio for local characteristics domain. Blue circles highlight possible cut off points to determine the number of principle components to carry forward.

The following PCs were chosen to form part of the composite index (Table 11). PCs chosen for local potential explained 95% of the total variance, for safety explained 87% of the total variance and for local characteristics and connectivity the total explained variance from the PCs carried forward was 90% and 91% respectively.

Table 11: Principle Components chosen to form the composite index.

Domain	PC	% Explained Variance	Correlations	Correlation Coefficient
Local Potential	PC0	55%	Motorised vehicle availability	0.58
			General health	-0.94
			Education	-0.95
	PC1	34%	Motorised vehicle availability	0.71
			Method used to travel to work	0.89
	PC2	6%	Education	0.62
Safety	PC0	30%	Street lighting	0.7
			Crime rate	0.65
			Pedestrian casualties	0.62
	PC1	15%	Cyclist casualties	0.32

			Highway network surveillance ratio	-0.5
	PC2	13%	Traffic calming measures	-0.57
			Street lighting	-0.62
	PC3	12%	Highway network surveillance ratio	0.39
	PC4	9%	Crime rate	-0.47
	PC5	8%	Cycle route surveillance ratio	0.39
Local Characteristics	PC0	36%	Cycle network illumination ratio	-0.67
			Highway network illumination ratio	-0.74
			Local Road speed limits	-0.62
			Population density	-0.78
	PC1	14%	Cycling network illumination ratio	0.58
			Proportion of cycle facilities	0.74
	PC2	10%	Proportion of leisure cycle routes	0.48
			Cycling network illumination ratio	-0.46
	PC3	9%	Proportion of leisure cycling routes	0.57
	PC4	8%	Land use mix	0.64
	PC5	7%	Footpath illumination ratio	0.65
	PC6	6%	Footpath illumination ratio	-0.53
Connectivity AM	PC0	47%	Connected cycle node ratio	0.85
			Connected walking node ratio	0.78
			Cycle network link node ratio	0.67
	PC1	25%	Connected walking node ratio	-0.64
	PC2	19%	Bus service connectivity AM	0.95

4.4.4 Weighting and aggregation

Selecting an appropriate weighting method is crucial as it heavily affects the outcome of the composite index (Becker *et al.* 2017, Greco *et al*. 2019). There is no uniformly agreed approach on how to weight the variables, however one of the most common options is to apply an equal weighting. Equal weighting is most appropriate when there are no empirical grounds for choosing a different approach (European Commission, 2016). However, in this instance, PCs have been weighted based on their corresponding eigenvalue (the amount of explained variance it contributes to the domain). Applying an equal weighting was discounted as it could diminish the value of the most explanatory PCs and artificially inflate one of the PCs that contributes less to the domain.

All PCs were normalised using the min-max rescaling method ensure all values lay between 0 and 1 and there were no negative values that could contribute to the PCs cancelling each other out when combined. As the eigenvalues are a measure of spread squared, the square root is taken to ensure it is comparable to the units of the input data. The normalised PCs are then multiplied by the factor and summed to create a single value for each domain.

4.4.5 Analytical Hierarchy Process

The methodology and weighting applied to combine the indicators discussed to form one overall measure of multi-modal opportunity is crucial. Due to the spatial nature of multi-modal mobility and the fundamental role location and proximity play in facilitating multi-modal travel a two-step approach is undertaken to incorporate the indicators.

Analytical Hierarchy Process (AHP) (Saaty, 1980) has been used to weight the LSOA level domains in to one composite indicator and the individual components for each bus stop and railway station separately. Employing a participatory approach, allows the relative societal importance of the indicators to be expressed (European Commission, 2016). AHP was developed to understand complex decision making (Saaty, 1980). AHP uses pairwise comparisons to derive priority scales (Saaty and Vargas, 1987). These comparisons allow for a judgment to be made of how one indicator dominates over another, which can be used to derive weights for the indicators. For this study, AHP was conducted using a survey with 5 participants who were deemed to be experts within the field of transportation. Experts were chosen, rather than public participants to ensure an acceptable consistency ratio can be met (Lee and Chan, 2008). A copy of the AHP survey can be seen in appendix C. The weights from each survey were calculated using the online AHP Priority Calculator (Goepel, 2019). An average weighting was then calculated from five AHP Priority Calculator results and applied to combine the local characteristics, local potential, and safety and connectivity indicators. The area domain weights applied can be seen in Table 12, and the weights applied to the railway station and bus stop components can be seen in Table 13 and Table 14 respectively.

Table 12: Weightings applied to the LSOA level domains.

Domain	***Applied Weight***
Local characteristics	4.7%
Local potential	44.2%
Connectivity	25%
Safety	26.1%

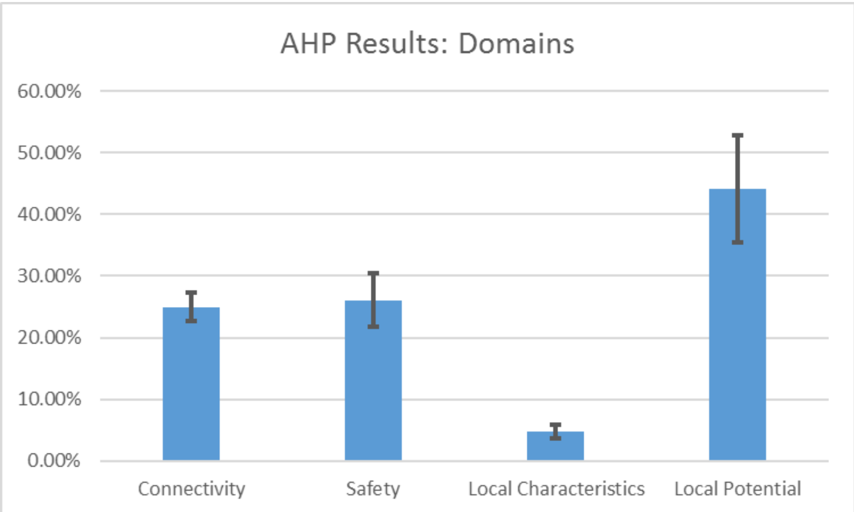


Figure 16: Analytical Hierarchy Results for the main domains.

Results from the AHP survey highlights that in the LSOA level domains local potential is deemed to be the most important with a weighting of 44.2% followed by safety at 26.1% and network connectivity at 25% (Table 12). Calculation of the weights is an iterative process, Figure 16 highlights the weightings of each component along with the maximum and minimum weights calculated in each iteration. Local potential has the highest final weighting but also had the widest variability.

Table 13: Weightings applied to the Railway station key components.

Railway Station Key Component	***Applied Weight***
Waiting room facilities	3.8%
Route options	21.6%
Average rail service delay	14.3%
Average rail service cancellation	13.5%
Cycle hire facilities	1%
Accessible parking	4%
Baby changing facilities	2.4%
Toilet facilities	5.7%
Local bus connection options	29.4%
Presence of CCTV	3%
Cycle parking facilities	1.3%

Railway stations had the most key components to be considered in the AHP survey, with the participant having to make 55 attribute comparisons. Weightings calculated for the railway station key components highlight that local bus service connections and route options are determined to the most important components when understanding a railway stations attributes impact on multi-modal opportunity. Service reliability components are then weighted to be the next important at 14.3% and 13.5% for average railway services running on time and average non-cancelled services respectively. The importance of having a suitable waiting environment to aid multi-modal transfers is discussed in the literature as they increase comfort and provide weather protection (National Academies of Sciences, Engineering, and Medicine, 2014). However, the AHP survey weighting calculated for waiting room facilities was 3.8%, below the importance of having toilet facilities which is weighted at 5.7%.

Table 14: Weightings applied to the Bus stop key components.

Bus Stop Key Component	***Applied Weight***
Presence of street lighting	17.8%
Presence of CCTV coverage	7%
Route options	75.2%

For bus stops, route options and subsequently service provision were weighted to be the most important key component with a weight of 75.2%. Presence of street lighting was weighted to be the second most important at 17.8% followed by CCTV coverage at 7%.

4.4.6 Incorporating space into the Index

Location and proximity play a crucial role in facilitating multi-modal travel (Krygsman and Dijst, 2001). Given this, it is key to ensure this is captured in the index as not all bus stops and railway stations are accessible from a single location.

To understand which areas are within walking and cycling distance of each railway station and bus stop ArcGIS Network Analyst was used. A network dataset was created using the footpaths, cycle ways and roads datasets. The Service Areas Network Analyst extension was used to define polygons based on all walkable and cyclable paths within the given impedance (Esri, 2019a). For this study distance was used as the specified impedance. Railways stations and bus stops were used as facilities and Table 15 highlights the distance thresholds defined. Cycling to a bus stop was not considered in analysis as there was no known information regarding cycling provision at bus stops within the county and many of the bus operators considered do not allow bicycles to be taken on board (Arriva, 2019).

Table 15: Distance Thresholds used for the Service Area Network Analysis in ArcGIS.

	Railway Station	***Bus stop***	***Source***
Walkable distance	1610 meters	800 meters	Wakenshaw and Bunn (2015)
Cyclable distance	3700 meters		Sherwin and Parkhurst (2010)

Network Analyst was used instead of a buffer, as a buffer works on producing polygons based on crow-fly distance not accounting for network detail. An example of the service area created by Network Analyst can be seen in Figure 17. The polygons were then joined with the railway station and bus stop weightings.

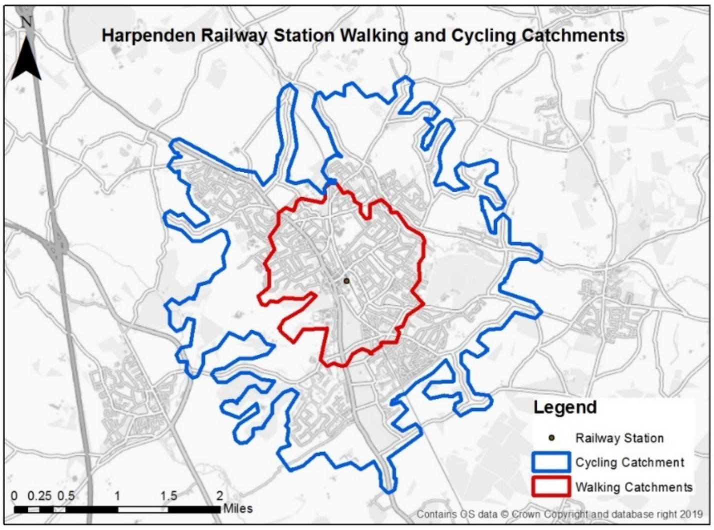


Figure 17: Example Network Analyst Service Areas.

To combine the area attributes with the railway station and bus stop catchments '*Feature to Polygon'* ArcGIS tool was used which creates new polygons based on overlaying segments. The outcome of this was a layer which contained 322,002 polygons.

A spatial join was then employed between the disaggregated polygon layer and the area, railway station scores and bus stop scores separately, with the relationship between the area and the data being made if the centroid (centre point) of the polygon was within the original LSOA or bus stop or railway station catchment area. *Dissolve* was used in between each spatial join to aggregate the scores by catchment area as a '*one to many'* join was used for railway station and bus stop catchments.

4.5. Visualising the Index

Many of the attributes considered for this composite index were calculated at a time period level to capture service variability throughout the day and add greater detail to the output. In light of this, there are three separate composite index maps created for AM, IP and PM. Due to the number of polygons created when using the '*Feature to Polygon*' tool, generalisation was undertaken to increase literacy of the map. In this case, areas < 1000m in area were incorporated into the neighbouring area in which they share the greatest boundary with.

Alongside the time period composite index maps, maps have been created outlining each spatial grouping for each time period. The time periods presented are AM (7am-9am), inter-peak (10am-3pm) and PM (4pm to 6pm) representing an average weekday, and Saturday and Sunday which is represented by an average hour. A key consideration when visualising the index was the symbology and selection of class intervals. Brewer (2006) argues that different class intervals can portray different stories with the same information, altering the overall message of the map. A total of 7 class intervals have been employed with a divergent colour scale. A divergent colour scale was used as having more than five class intervals with the same colouring can lead to colours being difficult to differentiate (Heywood *et al.* 2006). Alongside the ArcGIS outputs shown here, an interactive map has also been produced using Mapbox (Figure 18) and can be accessed through this link: https://hannahg1250.github.io/MMOIndex.github.io/ . Code extracts for the interactive map can be seen in appendix h.

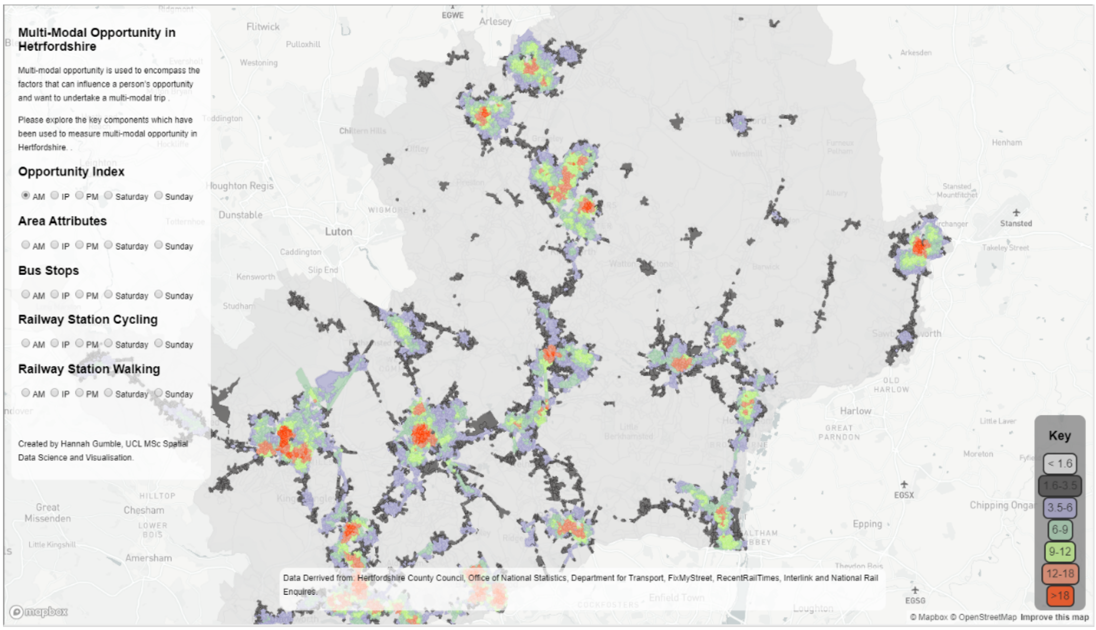


Figure 18: Index of Multi-modal opportunity online map.

5. Results

This chapter outlines the results of this study. Results of the composite index are discussed first followed by maps associated for each spatial grouping. Further analysis is then presented to explore to spatial patterns and the sensitivity of the index with respect to the weighting method applied and the number of PCs selected.

5.1. The Multi-modal Opportunity Index

The multi-modal opportunity index represents factors that can influence a person's opportunity and want to undertake a multi-modal trip, with a higher score exhibiting greater multi-modal opportunity. Due to the methodological choice to account for the specific locations of bus stops and railway stations and subsequent catchment areas, there is no maximum score, as a single area could be captured in numerous station and bus stop catchments. Figure 19 to Figure 21 highlight the output multi-modal opportunity index results for weekday AM, IP and PM respectively.

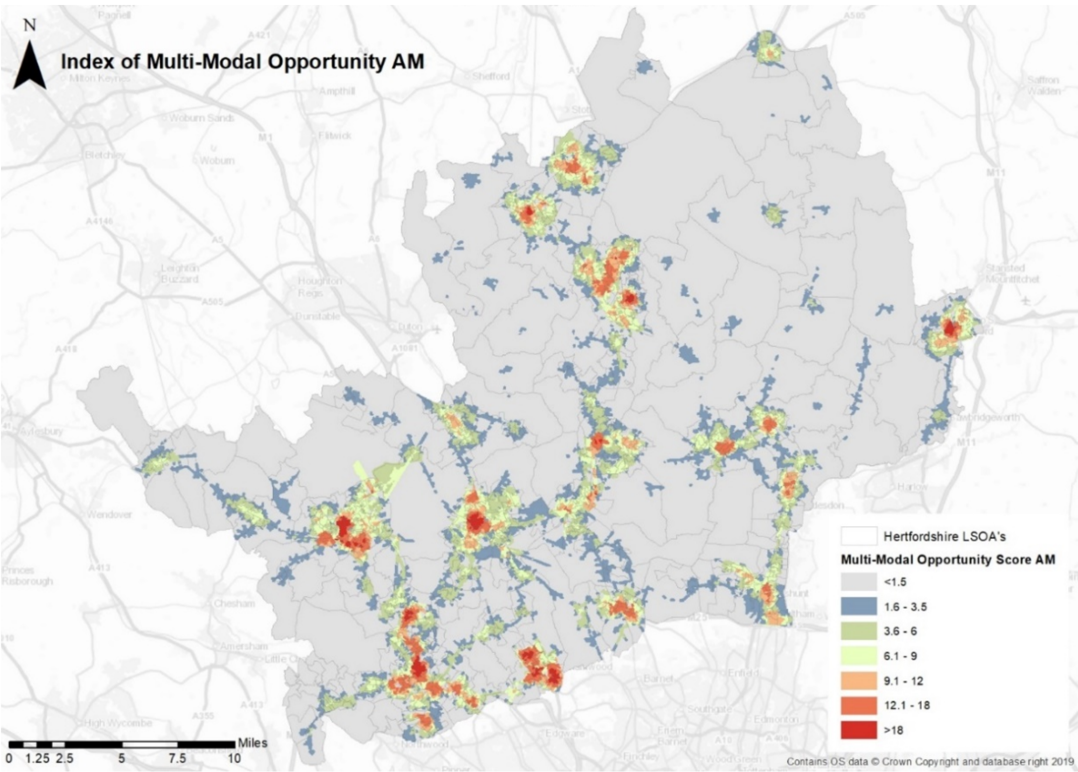


Figure 19: Index of Multi-modal Opportunity AM Results.

Looking at the multi-modal index results (Figure 19) higher multi-modal scores can be seen in town centres. Key movement corridors can also be seen particularly for North to South movements in the South-West of the county. As seen in Figure 20, for many large LSOA's particularly in the North-East of the county, opportunity is not spatially consistent and dominated by low scores for multi-modal opportunity.

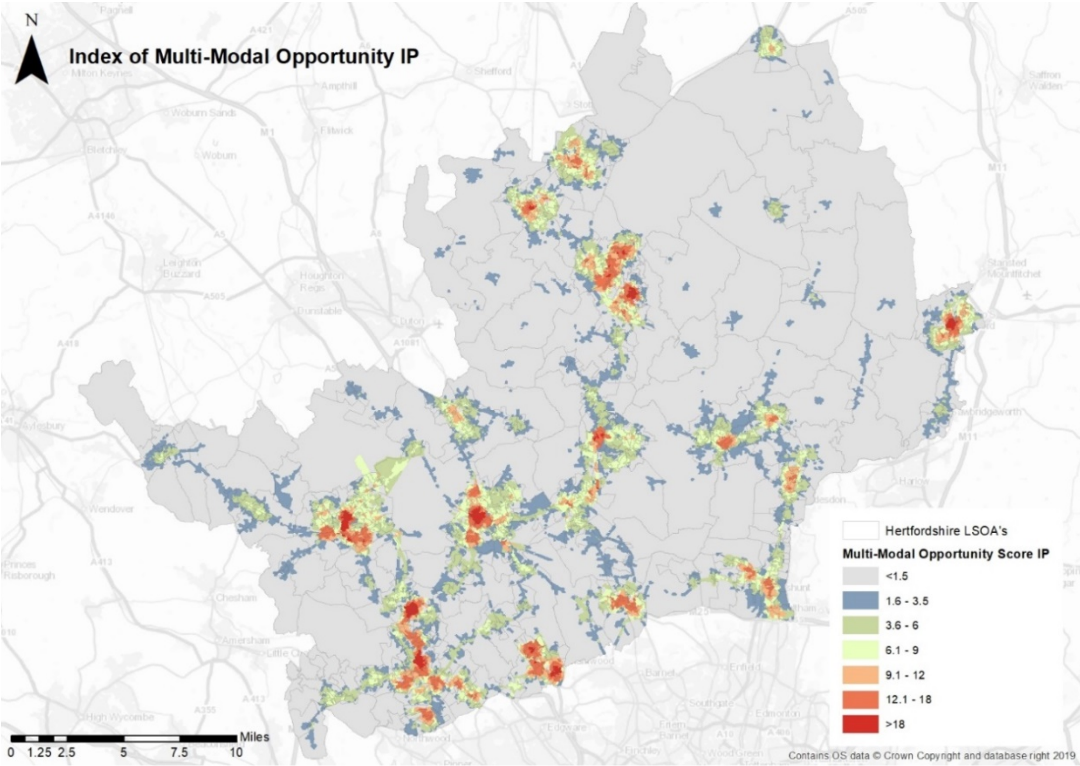


Figure 20: Index of Multi-modal opportunity IP results.

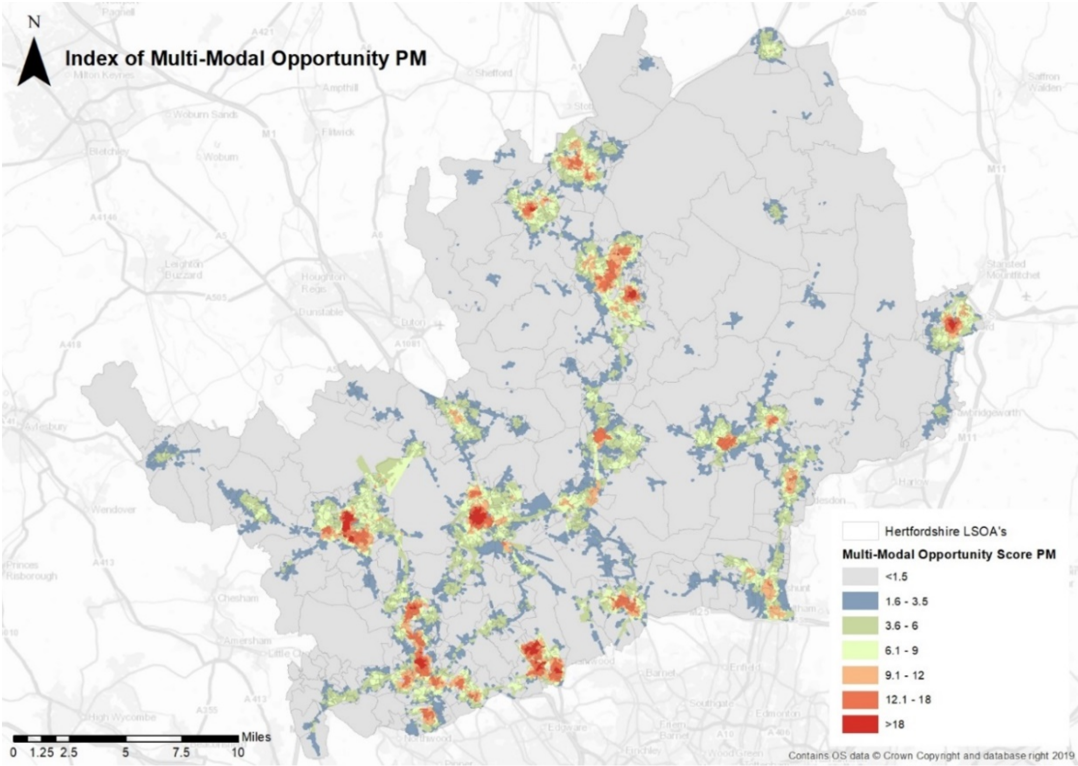


Figure 21: Index of Multi-modal opportunity PM results.

Comparing the results by time period, it can be seen that the pattern, particularly of the low multi-modal opportunity scores is consistent. The main fluctuations can be seen in areas with higher opportunity. This is further highlighted in Figure 22 which compares the time period scores for the district of St Albans.

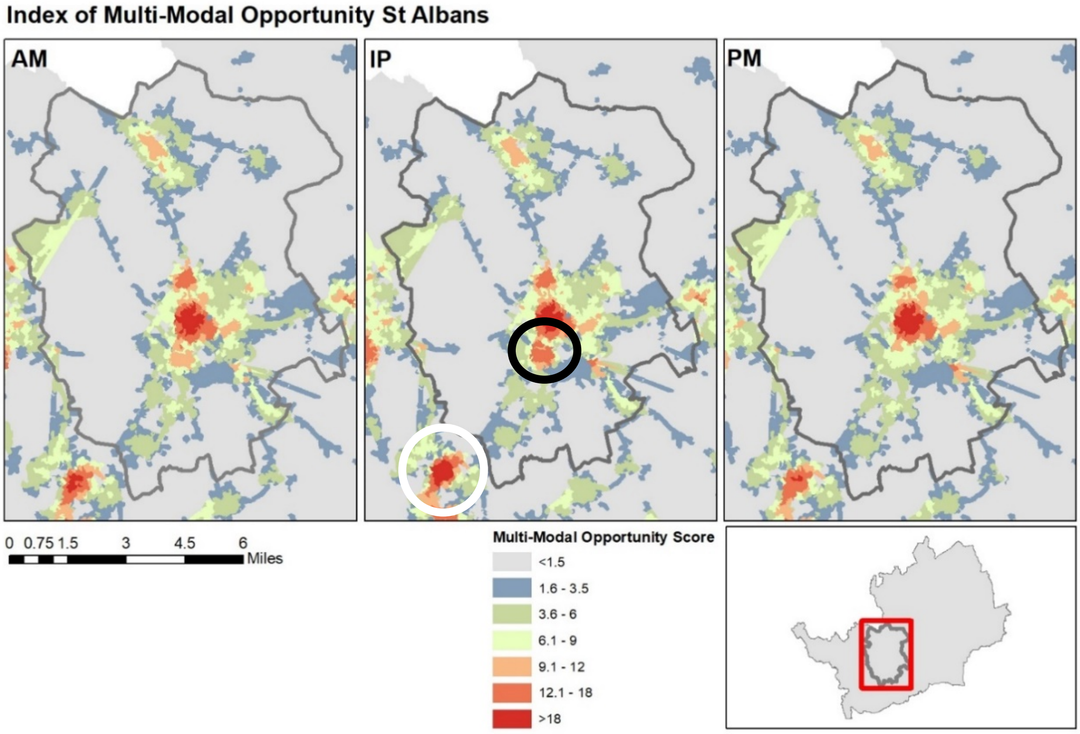


Figure 22: Index of Multi-modal opportunity time period comparison for St Albans.

The black circle on Figure 22 highlights an area where the multi-modal opportunity score becomes distinctly higher in the IP, particularly compared to the PM where the area shows lower multi-modal opportunity between 6.1 and 9. The white circle (Figure 22) demonstrates a very similar pattern with the highest multi-modal opportunity score being observed in the IP followed by the AM and PM respectively.

Stevenage, another district in Hertfordshire (Figure 23) exhibits a very similar pattern, with higher scores for most areas occurring in the IP. There are two clusters of high multi-modal opportunity highlighted through the black ellipses. This spatial pattern highlights a clear lack of multi-modal opportunity for the housing areas to the North-East and South-East of the district.

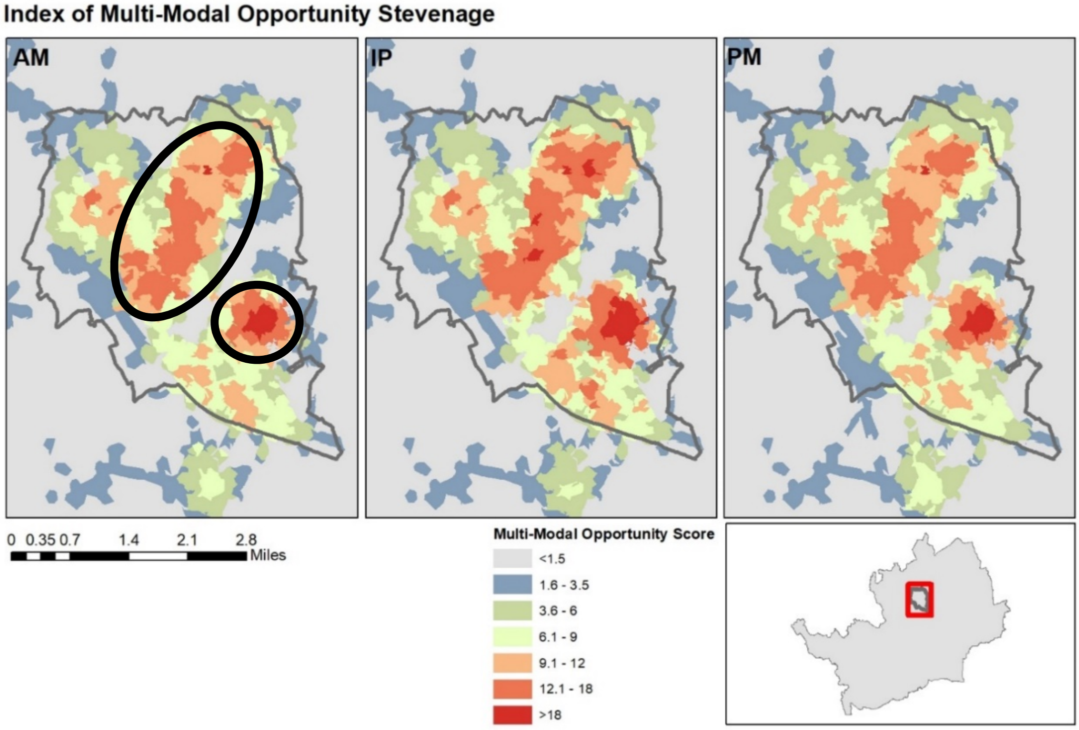


Figure 23: Index of Multi-modal opportunity time period comparison for Stevenage.

The calculation of multi-modal opportunity is driven by the location and provision of bus services. This is particularly visible at an aggregate level. The score for each bus and railway catchment is added to the base area scores. The number of bus stops throughout the county is significantly higher than railway stations, given this, the output result is driven by the opportunity that the bus service provides. Therefore, when the multi-modal opportunity score is calculated at an aggerate level, the area scores are over-shadowed and cannot be distinguished.

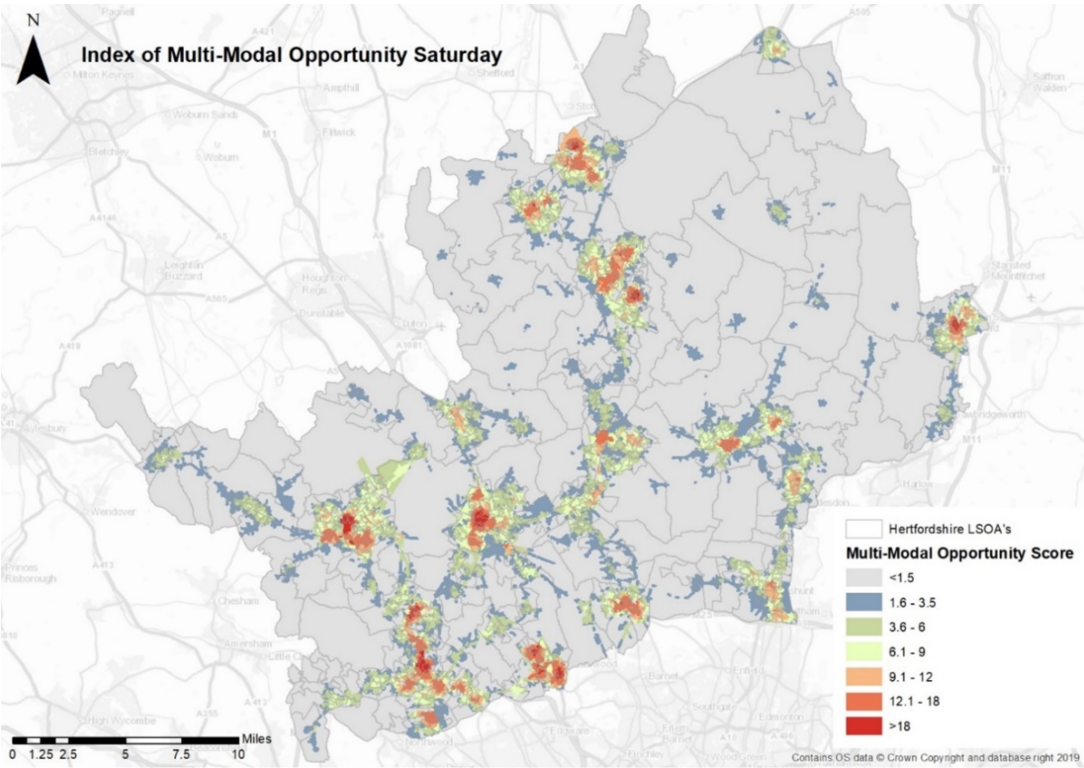


Figure 24: Index of Multi-modal Opportunity Results Saturday.

Index results for Saturday (Figure 24) and Sunday (Figure 25), highlight that on Sunday there is a reduction of areas with scores >18, indicated through the black circles on Figure 25. This highlights the day-to-day variation in opportunity throughout the county, supporting findings of lower service provision on weekends due to lower demand as identified by Buehler and Pucher (2012).

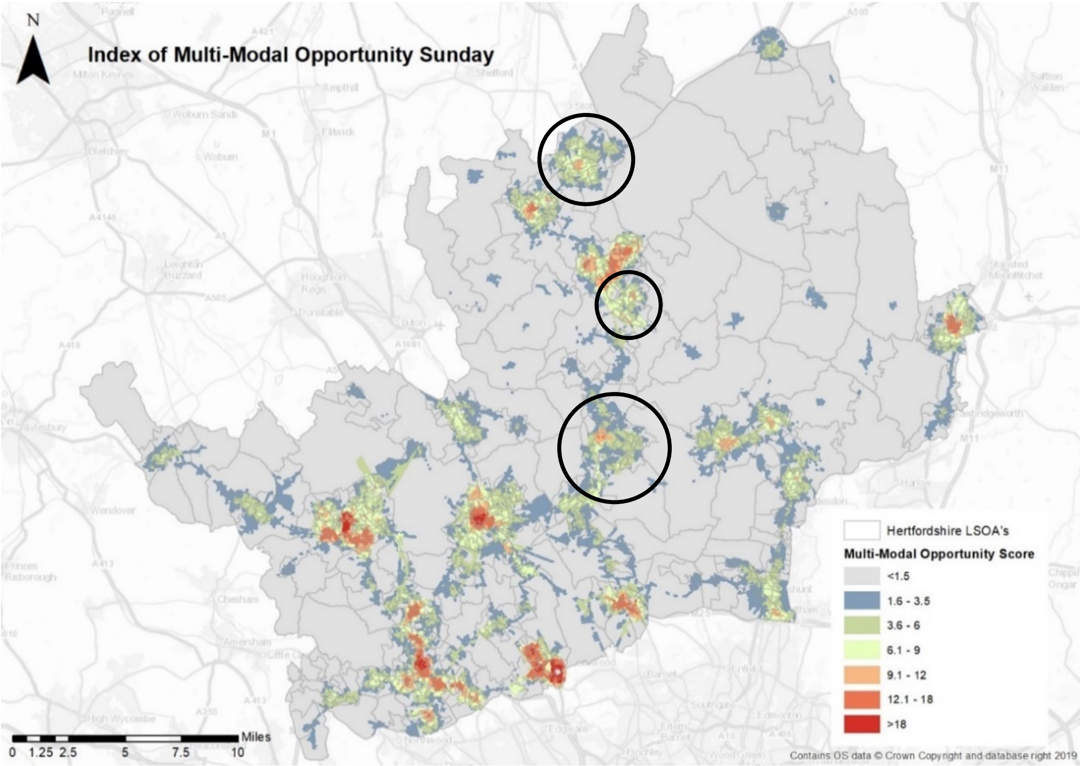


Figure 25: Index of Multi-modal Opportunity Results Sunday.

Area scores for the AM are shown in Figure 26, results for IP and PM can be seen in appendices I and J. When examining area level indicators, it is important to recognise that domains did not carry an equal weighting, reflecting the explained variance of each PC. The explained variance captured by the PCs explains a specific amount of multi-modal opportunity, therefore it was important that this was captured. Throughout all time periods many LSOA's have a multi-modal opportunity score between 0.21 and 0.4.

Railway station and bus stop catchment scores are illustrated in Figure 27 and Figure 28, respectively. The railway station with the highest multi-modal opportunity is Stevenage. There is a clear spatial pattern to the bus weightings with rural bus stops exhibiting lower multi-modal opportunity highlighting a lack of service provision. Hertfordshire County Council (2018) recognise that the county has greater provision of North-South transport links than East-West links, this is supported by the bus stop results as higher multi-modal opportunity can be seen in the North-South corridors compared to the East-West corridors.

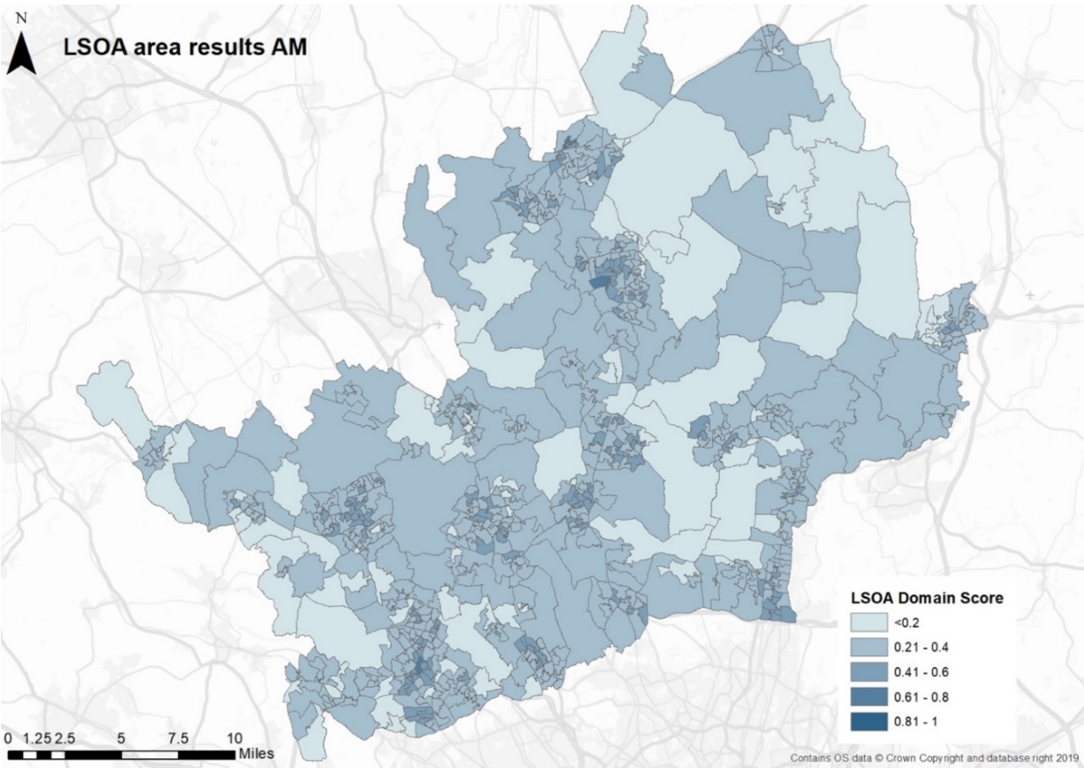


Figure 26: LSOA area scores AM.

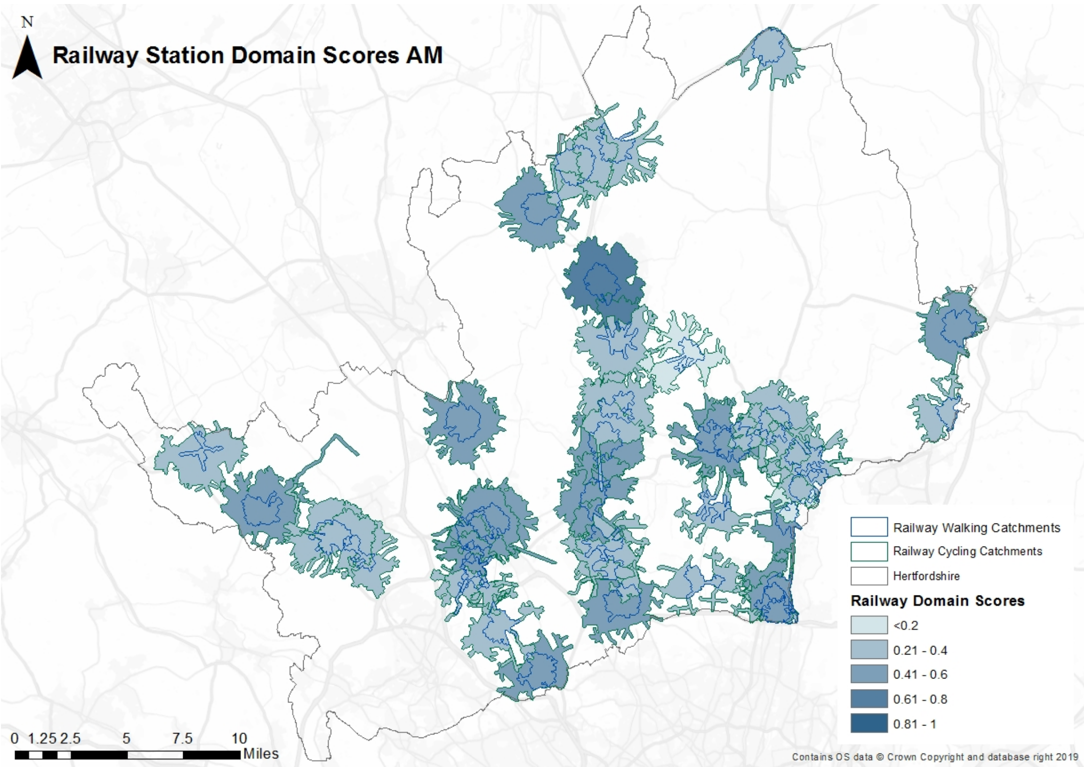


Figure 27: Railway station domain score AM.

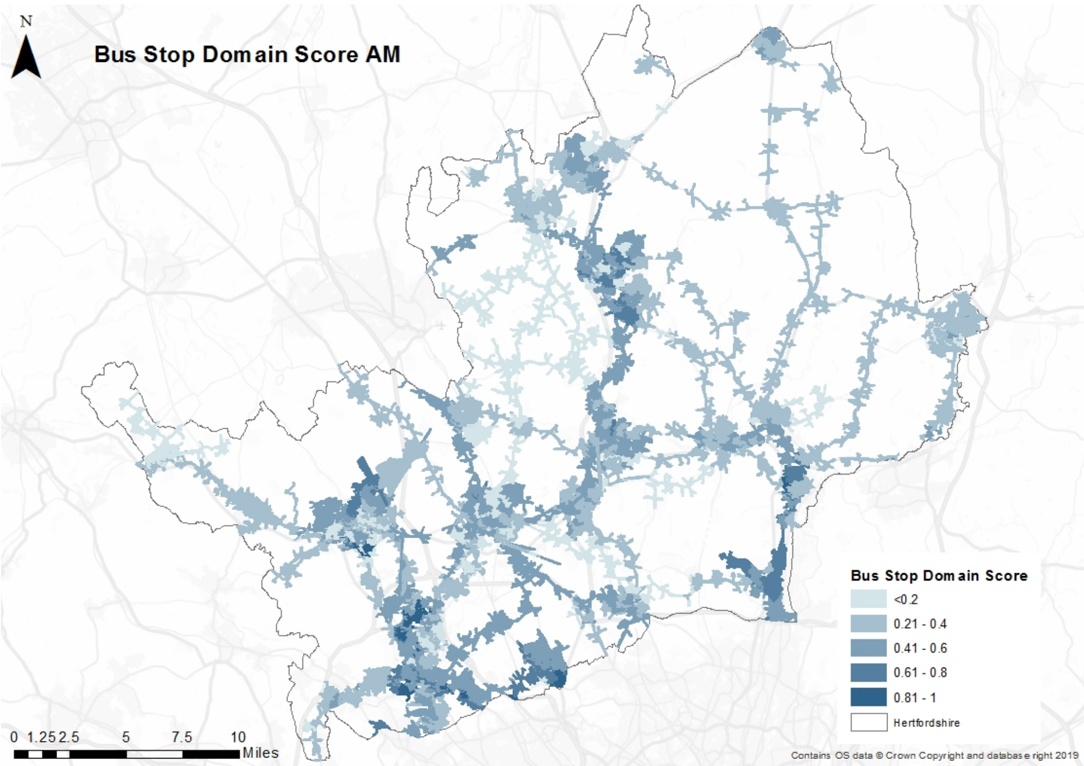


Figure 28: Bus stop domain score AM.

5.2. Spatial Analysis

5.2.1 Spatial Autocorrelation

Having the multi-modal opportunity results for the county allows for comparisons and spatial patterns to be observed. However, observation alone cannot be used to determine if the clusters of score seen are statistically significant. Moran's I spatial autocorrelation statistic has been used to determine if multi-modal opportunity is dispersed, clustered or random throughout the county (ESRI, 2018). The results presented here have been interpreted with the null hypothesis that multi-modal opportunity is randomly distributed.

Table 16: Spatial Autocorrelation Moran's I Result.

Time Period	***Moran's Index***	***Z-Score***	***P-Value***
AM	0.48	381.14	0.00001
IP	0.47	378.05	0.00008
PM	0.48	382.32	0.00000

Moran's I was undertaken on all time periods independently, with the spatial relationship set to inverse distance. This means that areas closest to the area in question have a greater influence in the calculation. As the P-Value is significant i.e. smaller than 0.05 and the Z-Score is positive, it can be concluded that the distribution of high and low scores is spatially clustered. Results for all time periods indicate that there was a less than 1% likelihood that the multi-modal opportunity scores were as a result of random chance and therefore the null hypothesis is rejected.

Due to the recognition that at an aggregate level the multi-modal opportunity index is driven by the spatial locations of bus stops and railway stations, spatial autocorrelation was also undertaken to examine the spatial nature of the underlying area attributes. Again, the null hypothesis is that considering area attributes alone, multi-modal opportunity is randomly distributed. The spatial relationship was defined as 'contiguity edges and corners'. Similarly, to the overall index, it can be concluded that the distribution of high and low area scores is spatially clustered.

Table 17: Moran's I Spatial Autocorrelation result considering the area attributes.

Time Period	***Moran's Index***	***Z-Score***	***P-Value***
AM	0.38	17.18	0.00001
IP	0.38	16.98	0.00000
PM	0.39	17.38	0.00003

5.2.2 Hotspot Analysis

Hotspot analysis was undertaken to understand if there were any statistically significant hot or cold areas of multi-modal opportunity. Although a high value can be observed, it may not be statistically significant. Hotspots are calculated using the Getis-Ord Gi statistic for each feature in the dataset. The local sum of features captured by the defined spatial relationship (in this instance ‘contiguity edges and corners’) is calculated and compared proportionally to the sum of all the features (ESRI, 2019c). Hotspots are designated where the results are significant cannot have been caused by random chance. Results of the AM hotspot analysis (Figure 29) highlight that all hotspots are within town centres, however, there are some densely populated areas where no hotspots are present. No cold spots (statistically significant areas of low multi-modal opportunity) were identified, this highlights that there is no one area where multi-modal opportunity is distinctly worse compared to the rest of the county. Similar results were observed in the IP and PM, these can be seen in appendix K.

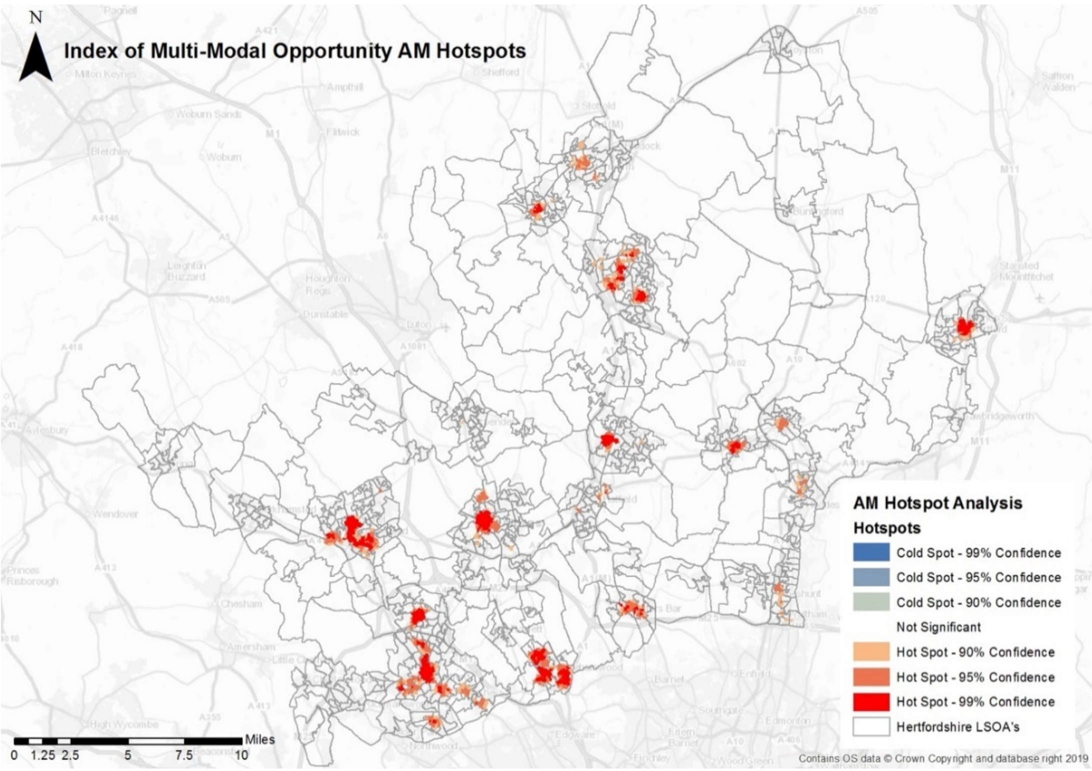


Figure 29: Index of Multi-modal Opportunity AM hotspot analysis.

5.3. Sensitivity Analysis

This section analyses the sensitivity of the index with respect to PC selection and weighting method chosen.

5.3.1 Principle Component Selection

PCA is a common methodology used when creating a composite index, however, as discussed in subsection 4.4.3 selecting the components to carry forward is not mathematically defined. For the local characteristics domain an argument could be made to select 4 or 7 PCs. To understand the impact of the number of components selected further analysis was undertaken on the area attributes score. The resulting score for the local characteristics domain was mapped to understand the change in values when more PCs were used in the study.

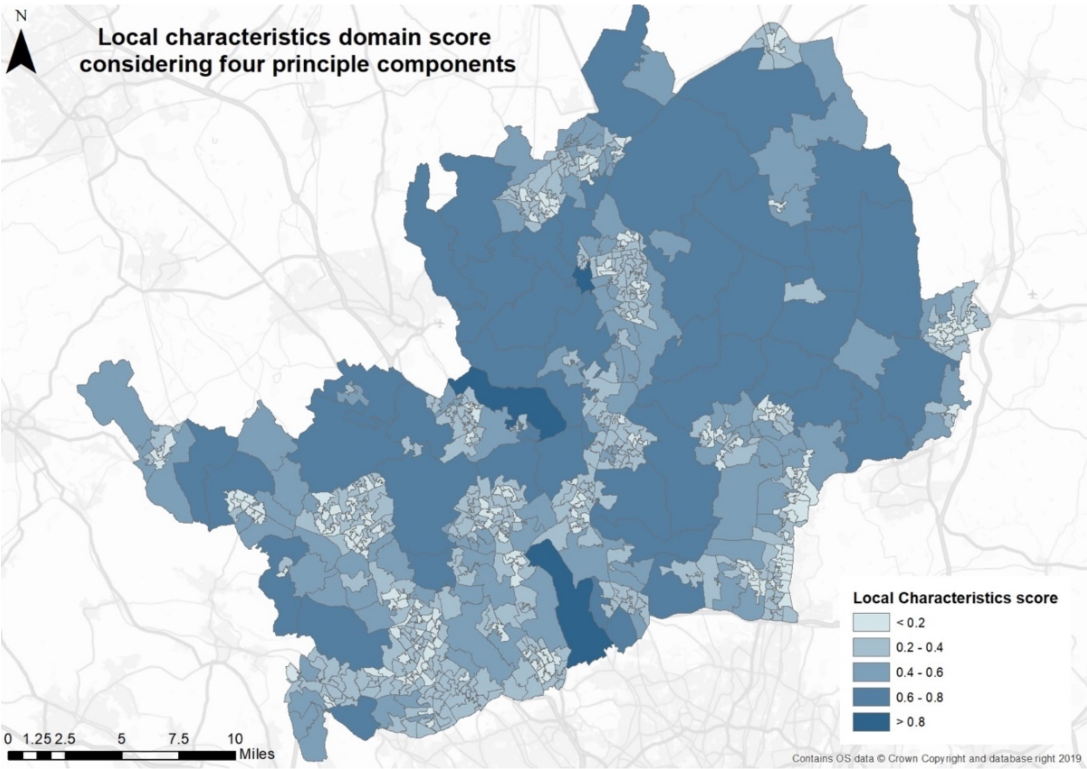


Figure 30: Local Characteristics domain score when considering four principle components.

Comparing Figure 30 and Figure 31, there is a significant change in areas with higher scores when using four PCs weighted by their explained variance compared to seven. The greatest impact is on the larger rural LSOA's with many having a score > 0.61 reducing to

0.2-0.6. This highlights the importance of choosing a suitable number of PCs as the effect of using more or less to form the domain is not even spatially. Furthermore, this highlights the correlation between different types of areas i.e. urban vs rural, and the PCs considered. Clear differences in travel behaviour, patterns, practices and even policies within urban and rural areas have been studied widely (Abenoza *et al.* 2017), and it can be interpreted that some PCs reflect these spatial differences.

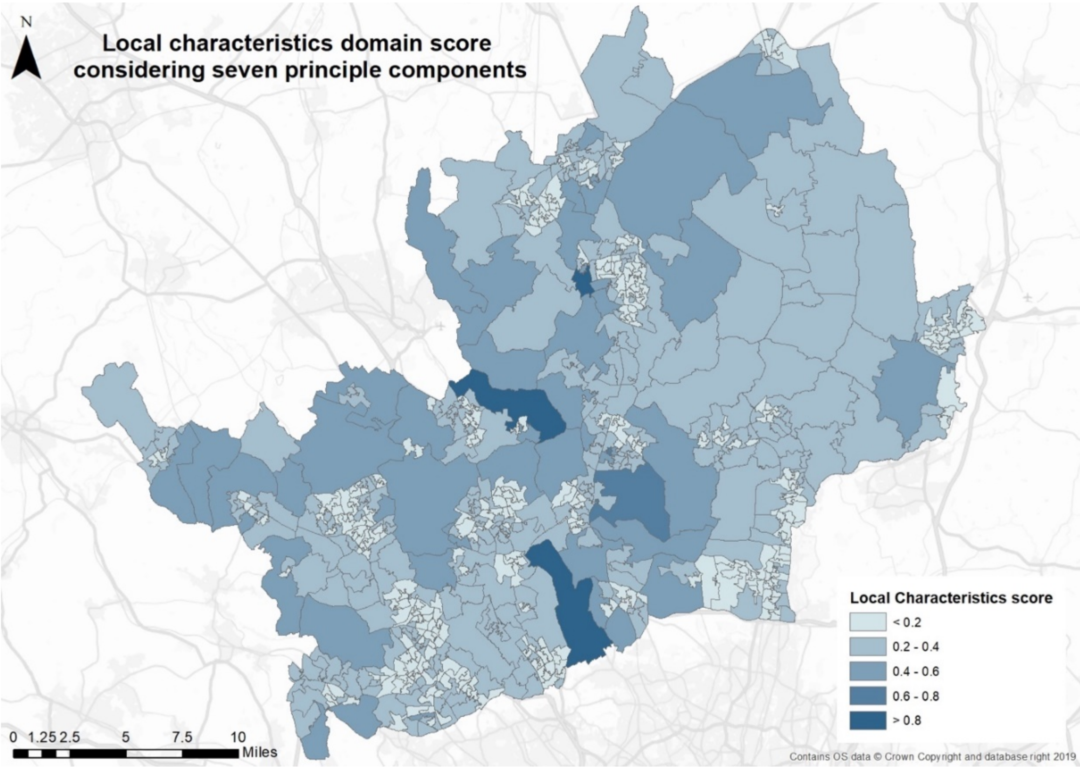


Figure 31: Local Characteristics domain score when considering seven principle components.

5.3.2 Importance of weighting the Principle Components

The importance of weighting was stressed throughout the literature (Becker *et al.* 2017, Greco *et al*. 2019). For this study PCs were weighted by the square root of their variance. To understand the impact of the weighting when combining the PCs sensitivity analysis is undertaken. Alternative weightings for the local characteristics domain were generated from a Dirichlet distribution (Wang *et al.* 2011) using the following code:

```
import scipy.stats as sps
number_of_components = 3
random_weights = sps.dirichlet.rvs([1]*number_of_components)
```

These were then incorporated into the area level weightings and the resulting scores were analysed to understand the impact on the area score. The weightings used can be seen in Table 18.

Table 18: Weighting used for the principle component weighting investigation.

Weighting	***PC 0***	***PC 1***	***PC 2***	***PC 3***	***PC 4***	***PC 5***	***PC 6***
1	0.12	0.36	0.02	0.10	0.08	0.10	0.22
2	0.22	0.01	0.16	0.00	0.41	0.11	0.08
3	0.04	0.02	0.12	0.11	0.12	0.50	0.08
4	0.39	0.07	0.06	0.03	0.32	0.03	0.10
5	0.12	0.05	0.19	0.10	0.19	0.03	0.34
6	0.04	0.04	0.09	0.41	0.14	0.15	0.13
7	0.10	0.29	0.27	0.11	0.00	0.18	0.04
8	0.40	0.14	0.11	0.10	0.19	0.05	0.01
9	0.24	0.12	0.12	0.01	0.02	0.23	0.26
10	0.22	0.52	0.12	0.02	0.06	0.02	0.03

To analyse the impact of different weightings, the range of scores was examined using the minimum and maximum score (Table 19). The weighting used has little impact on the range of scores (i.e. +/- 0.007 for the minimum and +/-0.013 for the maximum). This analysis has been undertaken on the local characteristics domain which only accounts for 4.7% of the main domain. Therefore, the impact of weighting may not be as subtle when considering a domain which was weighted higher by the AHP survey.

Table 19: Range of scores with different weights applied (Weighting applied in the composite index is highlighted).

Weighting	***Min***	***Max***
1	0.034	0.756
2	0.035	0.754
3	0.035	0.750
4	0.035	0.754
5	0.033	0.757
6	0.038	0.763
7	0.034	0.759
8	0.037	0.757
9	0.031	0.750
10	0.037	0.760
Explained variance	0.036	0.757

The LSOA domain scores for three LSOA's were monitored to understand the impact of the weighting method chosen (Table 20). In all instances weighting by explained variance gives a higher score than the random weightings applied, although the magnitude is not the same as LSOA 23780 is distinctly higher.

Table 20: LSOA Domain score comparison for three LSOA's.

LSOA ID	Explained Variance	Weight 1	Weight 2	Weight 3	Weight 4	Weight 5	Weight 6	Weight 7	Weight 8	Weight 9	Weight 10
23322	0.311	0.252	0.258	0.252	0.263	0.254	0.259	0.251	0.266	0.254	0.256
23780	0.570	0.449	0.447	0.442	0.304	0.450	0.449	0.444	0.447	0.444	0.451
33609	0.265	0.218	0.224	0.215	0.231	0.219	0.214	0.220	0.234	0.223	0.227

The variation observed for these three LSOA's reinstates the importance of the weighting method chosen especially as the domain tested here only accounted for 4.7% of the overarching domain. Even though the range of scores does not change significantly, the distribution of areas within that range did shift.

6. Discussion

Several questions were posed at the beginning of this study, which are discussed in this chapter in the context of the results. This chapter also reflects on the limitations and challenges faced.

6.1. What factors affect an individual undertaking multi-modal travel and how can they be measured using spatial data?

The factors that affect an individual's opportunity and desire to undertake multi-modal travel were identified from wider literature. In total, 26 variables were considered in the proposed framework covering both instrumental and non-instrumental factors. AHP weightings allowed for the relative importance of trip determinants to be understood. It is concluded that non-instrumental factors, particularly railway service reliability, local demographics and perceptions of safety are a critical determinant of multi-modal travel. This supports assertions that non-instrumental trip determinants need to be examined when considering travel mode choice (Stradling *et al.* 2007, Tumlin, 2012 and Molin *et al,* 2016). This is reinforced by the finding that local demographic variables and perceptions of safety were identified as more important than network connectivity, however, non-instrumental variables are somewhat ambiguous and do not have a consistent effect on individuals (Susilo and Cats, 2014).

Variables were measured from numerous datasets which were subject to different understandings of data reliability. Furthermore, the measurement of some variables was conceptualised as part of this study and reliant on geographic data. Geographic data are inherently uncertain, as it is a conceptualisation of the real world at a given time (Longely *et al.* 2015). Longely *et al.* (2015) hypothesise accurate representation of real-world attributes in GIS is impossible due to conceptualisation being a matter of judgement, subject to measurement errors, generalisation and temporal inconsistences.

Although much of the geographic data was provided by Hertfordshire County Council, little was known regarding the accuracy of data to that specific point in time. The calculation of variables such as illumination ratios and the use of Network Analyst is heavily dependent on locational accuracy. However, when 'real world objects' are represented in GIS, uncertainty is introduced through approximation and the co-ordinate system used (Dixon and Uddameri, 2015). Precautions were taken to clean and verify geographic data sources i.e. removing bus routes no longer in service and projecting all datasets to British National Grid, however, issues of locational accuracy would still be present. This may have affected the calculation of variables particularly where the data being measured was near the boundary of two LSOA's.

LOSA's were used as the zoning system to measure area variables. Given this, variables are subject to the modifiable area unit problem (MAUP), whereby using a different spatial zoning system could lead to a different result of a measured value (Wong, 2004). LSOA's were chosen as Census statistics were available at this level, however, if a different zoning system was chosen i.e. a raster grid, results would be different. Furthermore, through using LSOA's the concepts measured were constrained by the zone's crisp boundary. However, in real life, many spatial datasets do not have defined boundaries but are 'fuzzy' (Zadeh, 1997). Although in the visualisation of the index there appears to be harsh boundaries between two areas, in reality multi-modal opportunity is not as distinct.

In summary, non-instrumental factors have been identified as a key determinant of multi-modal travel in the AHP results. The composite index was successful in accounting for non-instrumental factors, which were not incorporated in Transport for London's PTAL's assessment tool (TFL, 2015) or Curtis *et al.* (2012) composite index SNAMUTS. However, there are key limitations that must be recognised.

6.2. Can the trip determinants identified be used to develop a composite index for multi-modal opportunity?

Based on the evaluation presented, a composite index was defined. However, the reliability and representation of the index could be challenged. Indicator selection is inherently subjective even if there is an overarching framework (Dobbie and Dail, 2013). Dobbie and Dali (2013) state that variables should be selected based on their relevance and importance, through selecting the variables from existing literature it ensured that all were relevant to the use of public transport and multi-modal travel. However, the index may be affected by omitted variable bias (Joavis *et al.* 2011) whereby a key attribute affecting multi-modal opportunity was not identified and not included in the construction of the composite index.

As discussed by Greco *et al.* (2019), weighting of a composite index can heavily affect the result. Weighting was applied to the composite index at an aggregate level, however, it is likely that the importance of the variables may change when considering different trip purposes and travel times. Different trip purposes and population subgroups have different values of travel time (Gunn, 2001) which affect travel mode choice and the importance of attributes such as delay. Furthermore, the importance of variables may vary over weekends which has not been accounted for.

PCA was undertaken as a form of dimension reduction within this study and choosing PCs to carry forward is a subjective choice. The results seen in 5.3.1 suggest that taking more PCs would affect the outcome result, specifically as they were weighted by their corresponding explained variance. Lever *et al.* (2017) argue that PCA does not always find the important patterns. The PCs are built off of the domains defined. It is possible that there is overlapping covariance between variables of different domains which is not captured, and no analysis was undertaken on the effect of changing which variables makeup the key domains.

Using AHP for the weighting was appropriate as it captures a ‘real world’ viewpoint. However, the number of variables considered affects the maximum weighting that can be given to the most important variable. The weight given to a variable is a function of the maximum importance possible on the AHP weighting scale (M) and the number of criteria (n) which can be expressed as:

$W_{max} = M/(n+M-1)$ (Goepel, 2014)

Goepel (2014) determined that when 10 variables are considered, the maximum weight that can be given to a single variable is 50%, even if it is given the maximum preference.

Therefore, as AHP was used to weight the area level indicators and bus stops and railway stations separately, the weighting given to one variable maybe lower within its sub-group, even if it has greater importance against attributes of another group. Therefore, a weighing of 30% for a bus stop attribute is not equal to a 30% weighting in another key component.

6.3. Can the framework devised provide any insights into the spatial variation in multi-modal opportunity in Hertfordshire?

The framework presented here derives key insights into multi-modal opportunity. This study shows there is a greater amount of multi-modal opportunity in urban centres compared to rural areas. The index created highlights some key disparities in multi-modal opportunity within urban areas, which may not be identified without the index. Multi-modal opportunity appears to cluster spatially, highlighting that opportunity is a spatial problem that needs to be understood and addressed to encourage multi-modal travel.

The index can be used to support understanding of gaps within public transport services to improve continuity in service provision and multi-modal opportunity. Hotspot analysis (section 5.2.2) showed there were no statistically significant cold spots for multi-modal opportunity throughout the county but highlighted clear areas of low multi-modal opportunity. As part of the Hertfordshire's 2018 Local Transport Plan Hertfordshire County Council identified eight multi-modal movement corridors (Figure 32) which will be targeted to improve connectivity throughout the county.

Corridor 1: Aylesbury – Watford – London	**Corridor 2: London – Watford – Luton – Milton Keynes**
Corridor 3: London-Stevenage – Peterborough	**Corridor 4: London – Harlow – Stansted – Cambridge**
Corridor 5: Hemel & Watford – St Albans – Harlow	**Corridor 6: Luton – Stevenage**
Corridor 7: Stevenage – Cambridge	**Corridor 8: Stevenage - Stansted**

Figure 32: Multi-modal movement corridors identified as part of Hertfordshire's 2018 Local Transport plan (Hertfordshire County Council, 2018).

When these corridors are digitised and overlaid upon the multi-modal opportunity index (Figure 33) key areas for increased public transport provision can be identified i.e. between the urban centres in corridor eight. It also highlights that some corridors, particularly corridor three, have greater opportunity for individuals to adopt multi-modal travel behaviour, with the opportunity following a similar spatial direction to the corridor. Providing a visual representation of current multi-modal transport systems is key for enhancing understanding (Curtis *et al.* 2012) and mapping multi-modal opportunity can help improve understanding of the spatial variation of opportunity and key areas for future transport initiatives to support multi-modal travel behaviour.

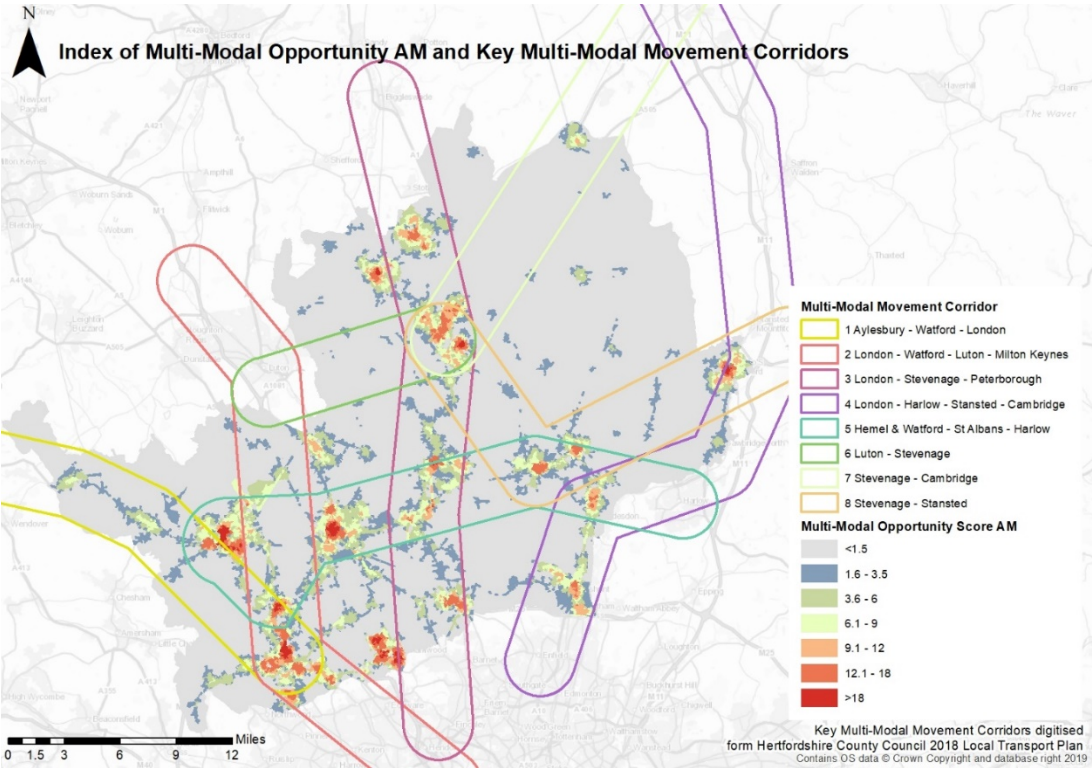


Figure 33: Hertfordshire County Council Key Multi-modal Movement Corridors and Index of Multi-modal Opportunity Results AM.

Key data limitations exist, specifically around bus stops and their corresponding routes which would affect the results. Elements of travel demand including: trip origin, trip destination and purpose were not accounted for, even though they form key interdependencies between a traveller and their mode choice (Sanchez, 2004). The index also does not incorporate any information regarding intermodal connections beyond facilities that can aid the comfort of a connections/waiting periods.

7. Conclusion

This study synthesised multi-modal trip barriers and determinants into a comprehensive analysis framework to improve understanding of the spatial nature of multi-modal travel opportunities. Understanding multi-modal opportunity allows infrastructure provision to be visually understood and targeted to encourage greener multi-modal travel.

This composite index was constructed using 26 variables found in the literature that affect multi-modal travel behaviour, including both physical and non-instrumental travel determinants. Conceptualisation of non-instrumental attributes was undertaken with the spatial data available; however, a different approach could be considered.

Analysis aimed to understand if mapping and spatial analysis offered greater insight into multi-modal opportunity in Hertfordshire. Mapping the multi-modal index highlighted a clear spatial bias with clusters of high and low scores. In urban areas, there is a greater level of multi-modal opportunity, however, it is clear that this opportunity declines with distance from the towns centre. Key travel corridors are observed, reflecting Hertfordshire County Councils (2018) understanding that there are better North-South travel links through the county. Variation in multi-modal opportunity by time period and day of the week is also highlighted. The greatest opportunity exists in the IP and the least multi-modal opportunity exists on Sundays. Results highlight the importance of bus service provision, particularly given the current infrastructure provision in Hertfordshire. In this case study high multi-modal opportunity scores were driven through areas having greater accessibility to bus stops with higher service frequencies.

AHP, a simple pair-wise comparison method was used to capture expert's opinion and weight the variables considered. This refined suitability of the index, helping to conceptualise how individuals perceive multi-modal opportunity and the importance of the domains considered. Analysis on the PC selection and weighting method employed shows the number of components considered clearly impacts the results.

This research helps understand multi-modal opportunity and sought to address the gap in the variables considered when evaluating multi-modal transport systems and how they can be conceptualised into an analysis framework. The framework presented here is one of many ways to conceptualise multi-modal opportunity. Future research should seek to address the limitations in this study including the data limitations and the granularity of analysis. The methodology used to calculate the non-instrumental factors should also be assessed to understand the effect of conceptualisation and approach on the outcome result.

8. References

Abdi, H and Williams, L, J. (2010) Principle Component Analysis, *WIREs Computational Statistics,* 2, Pp. 433-459.

Abenoza, R, F., Cats, O and Susilo, Y, O. (2017) Travel satisfaction with public transport: Determinants, user classes and regional disparities and their evolution, *Transportation Research Part A: Policy and Practise,* 95, Pp.64-84.

Alto, V (2019) *Towards Data Science: PCA: Eigenvectors and Eigenvalues,* [Online] Available at: https://towardsdatascience.com/pca-eigenvectors-and-eigenvalues-1f968bc6777a [Accessed 29th July 2019].

Antonio Lindau, L., Hidalgo, D and de Almedia Lobo, A. (2014) Barriers to planning and implementing Bus Rapid Transit systems, *Research in Transportation Economics,* 48, Pp. 9-15.

Arriva (2019) *Arriva UK Bus Conditions of Carriage (Last Updated February 2019),* England. [Online] Available at: https://www.arrivabus.co.uk/globalassets/documents/global-pages/about-us/conditions-of-carriage---february-2019.pdf [accesses 15/07/2019].

Badland, H, M, Garrett, N., Schofield, G M. (2010) How Does Car Parking Availability and Public Transport Accessibility Influence Work-Related Travel Behaviours? *Sustainability,* 2, Pp.576-590.

Banister, D (2019) Transport for all, *Transport Reviews* 39(3): 289-292.

Basiri, A., Hacklay, M., Foody, G and Mooney, P (2019) Crowdsourced geospatial data quality: challenges and future directions, *International Journal of Geographic Information Sciences,* 33(8), Pp. 1588-1593.

Becker, W., Saisana, M., Paruolo, P and Vandecasteele, I (2017) Weights and importance in composite indicators: Closing the gap, *Ecological Indicators,* 80, Pp.12-22.

Beirão G. and Sarsfield Cabral, J, A (2007) Understanding attitudes towards public transport and private car: A qualitative study, *Transport Policy,* 14, Pp.478-489.

Brewer, C. (2006) Basic Mapping Principles for Visualising Cancer Data Using Geographic Information Systems (GIS), *American Journal of Preventative Medicine,* 30(2), Pp.25-36.

Brown, G. (2012) Public Participation GIS (PPGIS) for Regional and Environmental Planning: Reflections on a Decade of Empirical Research, *Journal of The Urban & Regional Information Systems Association,* 25(2), Pp. 7-18

Buehler, R., Hamre, A (2016) An Examination of recent trends in multimodal travel behaviour among American Motorists, *International Journal of Sustainable Transportation,* 10(4), Pp.354-364.

Buehler, R and Puncher, J. (2012) Demand for Public Transport in Germany and the USA: An Analysis of Rider Characteristics, *Transport Reviews,* 32(5), Pp 541-567.

Button, K, J. and Hensher, D. (2009) *Transport Planning: Handbook of Transport systems and Traffic Contol,* Emerald: UK.

Chen, C., Gong, H and Paaswell, R (2008) Role of the built environment on mode choice decisions: additional evidence on the impact of density, *Transportation,* 35, Pp. 285-299.

Clauss, T and Döppe, S. (2016) Why do urban travellers select multimodal travel options: A repertory grid analysis, *Transportation Research Part A: Policy and Practice,* 93, Pp.93-116.

Clifton, K and Murs, C, D. (2012) Capturing and Representing Multimodal Trips in Travel Surveys: Review of Practice, *Transportation Research Record: Journal of the Transportation Research Board,* 2285(1), Pp.74-83.

Cooke, S and Behres, R. (2017) Correlation or Cause? The limitations of population density as an indicator for public transport viability in the context of a rapidly growing developing city, *Transportation research Procedia,* 25, Pp.3003-3016.

Crabb, G, I., Beaumont, R and Webster, D (2008) *Review of the class and quality of street lighting,* Transport Research Laboratory, [Online] Available at: https://www.theilp.org.uk/documents/css-sl1-class-and-quality-of-street-lighting/ [Accessed 1st July 2019].

Curtis, C., Scheuer, J and Mellor, R (2012) *Spatial Network Analysis for Multimodal Urban Transport Systems (SNAMUTS)*, Perth, Australia. [Online] Available at: http://www.accessibilityplanning.eu/wp-content/uploads/2013/01/1-SNAMUTS-R1.pdf [Accesses 15th May 2019].

Department for Transport (2009) *Manual for Streets,* London: Thomas Telford publishing. [Online] Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/341513/pdfmanforstreets.pdf [Accessed 20th June 2019].

Department for Transport (2014) *National Public Transport Access Nodes (NaPTAN),* [Online] Available at: https://data.gov.uk/dataset/ff93ffc1-6656-47d8-9155-85ea0b8f2251/national-public-transport-access-nodes-naptan [Accessed 22nd May 2019].

Department for Transport (2019) The Highway Code. [Online] Available at: https://www.highwaycodeuk.co.uk/# [Accessed 20th June 2019].

Dixon, B and Uddameri, V (2015) *GIS and Geo-computation for Water Resource Science and Engineering.* John Wiley & Sons: Chichester.

Dobbie, M, J and Dali, D. (2013) Robustness and Sensitivity of weighting and aggregation in constructing composite indices. *Ecological Indicator,* 29, Pp.270-277.

El-Geneidy, A. Levinson, D. Diab, E., Boisjoly, G. Verbich, D and Loong, C. (2016) The cost of equity: Assessing transit accessibility and social disparity using total travel cost, *Transportation Research Part A: Policy and Practice,* 91, Pp.302-316.

Esri (2018a) *Density Based Clustering*, [Online] Available at:https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/densitybasedclustering.htm [Accessed: 20th June 2019].

Esri (2018b) *Spatial Join* [Online] Available at: https://pro.arcgis.com/en/pro-app/tool-reference/analysis/spatial-join.htm [Accessed: 20th June 2019].

Esri (2018c) *Dissolve.* [Online] Available at:https://pro.arcgis.com/en/pro-app/tool-reference/data-management/dissolve.htm [Accessed 20th June 2019].

Esri (2018d) *How Spatial Autocorrelation (Global Morans I) Works,* [Online] Available at: http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm [Accessed 1st August 2019].

Esri (2019a) *Feature To Polygon,* [Online] Available at: https://pro.arcgis.com/en/pro-app/tool-reference/data-management/feature-to-polygon.htm [Accessed 29th July 2019].

Esri (2019b) *How Hot Spot Analysis (Getis-Ord-GI*) Works,* [Online] Available at: https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm [Accessed 1st August 2019].

European Commission (2013) *Key to Innovation Integrated Solution Multimodal personal mobility,* [Online] Available at: https://eu-smartcities.eu/sites/default/files/2017-10/Multimodal%20personal%20mobility%20january.pdf [Accessed 1st August 2019].

European Commission (2016) *Competence Centre on Composite Indicators and Scoreboards, Step 6: Weighting,* [Online] Available at: https://ec.europa.eu/jrc/en/coin/10-step-guide/step-6 [Accessed 18th July 2019].

European Commission (2019) *2018 the Year of Multimodality*, [Online] Available at:https://ec.europa.eu/transport/themes/logistics-and-multimodal-transport/2018-year-multimodality_en [Accessed 12th August 2019].

Fiorenzo-Catalano, S., Van Nes, R and Bovy, P, H, L (2004) Choice Set Generation of Multi-modal Travel Analysis, *European Journal of Transport and Infrastructure Research,* 4(2), Pp.195-209.

FixMyStreet (2019) *FixMyStreet Geographic Count Data,* [Online] Available at: https://github.com/mysociety/fms_geographic_data [Accessed 15th May 2019].

Friedrick, M. (2016) Evaluating the service quality in multi-modal transport networks, *Transportation Research Procedia,* 15, Pp.100-112.

García-Palomares, J, C., Gutiérrez, J and Cardozo, O, D. (2013) Walking accessibility to public transport: an analysis based on microdata and GIS, *Environment and Planning B: Planning and Design,* 40, Pp.1087-1102.

Goepel, K, D (2014) *AHP and the Magical Number Seven Plus or Minus Two,* [Online] Available at: https://bpmsg.com/tag/ahp-weight-calculation/ [Accessed 22nd July 2019].

Goepel. K, D. (2019) *AHP Priority Calculator,* [Online] Available at: https://bpmsg.com/ahp/ahp-calc.php [Accessed 15th July 2019].

Greco, S. Ishizaka, A. Tasiou, M and Torrisi, G. (2019) On the Methodological Framework of Composite Indices: A review of the Issues of Weighting, Aggregation and Robustness, *Social Indicators Research,* 141(1), Pp.61-94.

Gunn, H. (2001) Spatial and temporal transferability of relationships between travel demand, trip cost and travel time, *Transportation Research Part E: Logistics and Transportation Review,* 37(2-3), Pp.163-189.

Haans, A and Kort, Y, A, W. (2012) Light distribution in dynamic street lighting: Two experimental studies on its effects on perceived safety, prospect, concealment and escape. *Journal of Environmental Psychology,* 32, Pp.342-352.

Hadas, Y and Ranjitkar, P. (2012) Modelling public-transit connectivity with spatial quality-of-transfer measurements, *Journal of Transport Geography,* 22, Pp.137-147.

Hägerstand, T. (1970) What about people in regional science, *Papers of the Regional Science Association,* 24, Pp.7-21.

Helms, J, E. Hanze, K, T. Sass, T, L and Mifsud, V, A. (2006) Treating Cronbach's Alpha Reliability Coefficients as data in Counselling Research, *The Counselling Psychologist,* 34(5), Pp.630-660.

Hertfordshire County Council (2018) *Hertfordshire's Local Plan* [Online] Hertfordshire: Hertfordshire County Council. Available at: https://www.hertfordshire.gov.uk/media-library/documents/about-the-council/consultations/ltp4-local-transport-plan-4-complete.pdf [Accessed 27th May 2019].

Hertfordshire Partnership CCTV (2019) *Public Space CCTV Coverage* [Online] Available at: http://www.hertfordshirecctv.co.uk/public-space-cctv-coverage/ [Accessed 22nd May 2019].

Heywood, I., Cornelius, S and Carver, S (2006) *An Introduction to Geographical Information Systems & Science,* 3rd ed. Wiley: USA.

Home Office (2019) *Data Downloads,* [Online] Available at: https://data.police.uk/data/ [Accessed 22nd May 2019].

House of Commons (2002) *Jam Tomorrow?: The multi modal study investment plans.* London: House of Commons.

Interlink (2019) Timetable Search. [Online] Hertfordshire: Hertfordshire County Council. Available from: < https://www.intalink.org.uk/timetables/ > [Accessed: 20th June 2019].

Jolliffe, I, T. and Cadima, J (2016) Principle Component Analysis: a review and recent developments, *Philosophical Transactions A, Mathematical, Physical and Engineering Sciences,* 374(2065), Pp.1-16.

Jovanis, P, P. Aguero-Valverde, J, A. Wu, K and Shankar, V. (2011) Analysis of Naturalistic Driving Event Data Omitted-Variable Bias and Multilevel Modelling Approaches, *Transportation Research Record,* 2236, Pp.49-57.

Jun Park, S (2012) ' *CCTV Evaluation in Cincinnati within GIS Environment for Crime Prevention'*, PhD Thesis, Michigan State University, Cincinnati. Available at: https://www.semanticscholar.org/paper/CCTV-Evaluation-in-Cincinnati-within-GIS-for-Crime-Park/758323ee715297fd98e651fe2455c97083baccbd [Accessed 15th July 2019].

Kagerbauer, M., Hilgert, T., Schroeder, O and Vortish, P. (2015) Household travel survey of intermodal trips – Approach, challenges and comparison, *Transportation Research Procedia,* 11, Pp.330-339.

Kline, P. (1978) *Psychometric theory,* (3rd ed.) McGraw-Hill, New York.

König, A (2002) 'The Reliability of the Transport System and its Influence on the Choice Behaviour', *2nd Swiss Transport Research Conference,* Monte Verità/ Ascona, 20-22 March.

Kramers, A. (2014) Designing next generation multimodal traveller information systems to support sustainability-orientated decisions, Environmental Modelling and Software, 56, Pp.83-93.

Krygsman, S and Dijst, M (2001) Multimodal Trips in the Netherlands- microlevel Individual Attributes and Residential Context, *Transportation Research Record,* 01-3102, Pp.11-19.

Kuhnimhof, T., Buehler, R. and Dargay, J (2011) A New Generation: Travel Trends for Young Germans and Britons, Transportation Research Record: Journal of the Transportation Research Board, 2230(1), Pp.58-67.

Lee, G, K, L. and Chan, E, H, W. (2008) The Analytic Hierarchy Process (AHP) Approach for Assessment of Urban Renewal Proposals, *Social Indicators Research,* 89(1), Pp.155-168.

Lever, J., Krzywinski, M. and Altman, N. (2017) Principle Component Analysis, *Nature Methods*, 14, Pp.641-642.

Liao, F., Arentza, T and Timmermas, H. (2010) Supernetwork Approach for Multimodal and Multiactivity Travel Planning, *Transportation Research Record: Journal of Transportation Research Board,* 2175(1), Pp.38-46.

Longley, P, A., Goodchild, M, F., Maguire, D, J and Rhind, D, W. (2015) *Geographic Information Systems and Science.* Wiley, United States.

Manaugh. K and Kreider, T (2013) What is mixed use? Presenting an interaction method for measuring land use mix, *The Journal of Transport and Land Use,* 6(1), Pp.63-72.

Meyer de Freitas, L., Becker, H., Zimmermann, M and Axhausen, K, W. (2019) Modelling intermodal travel in Switzerland: A recursive logit approach, *Transportation research Part A: Policy and Practice,* 119, Pp.200-213.

Miller, H, J. (1991) Modelling accessibility using space-time prism concepts within geographical information systems, *International Journal of Geographical Information System,* 5(3), Pp.287-301.

Molin, E., Mokhtarian, P., Kroesen, M (2016) Multimodal travel groups and attitudes: a latent class cluster analysis of Dutch travellers, *Transportation Research Part A,* 83, Pp.14-29.

Nardo, M., Saisana, M., Saltelli, A and Tarantola, S. (2005) *Tools for composite indicator building – joint research centre European commission,* [Online] European Commission: Ispra. Available from: http://publications.jrc.ec.europa.eu/repository/bitstream/JRC31473/EUR%2021682%20EN.pdf [Accessed 1st August 2019].

National Academies of Sciences, Engineering, and Medicine (2014). *Estimating Bicycling and Walking for Planning and Project Development: A Guidebook*. Washington, DC: The National Academies Press. Available at: https://doi.org/10.17226/22330 [Accessed 15th May 2019].

Nobis, C (2007) Multimodality: Facets and Causes of Sustainable Mobility, *Transportation Research Record: Journal of Transportation Research Board,* 2010(1), Pp.35-44.

O'sullivan, D., Morrison, A and Shearer, J (2000) Using desktop GIS for the investigation of accessibility by public transport: an isochrone approach, *International Journal of Geographic Information Science,* 14(1), Pp.85-104.

OECD (2004) "The OECD-JRC Handbook on Practices for Developing Composite Indicators", paper presented at the OECD Committee on Statistics, 7-8 June 2004, OECD Publishing, Paris.

OECD (2008) *Handbook on constructing composite indicators: methodology and user guide.* OECD Publishing, Paris.

OECD. (2009) *OECD Economic Surveys: United Kingdom 2009,* Paris:OECD. [Online] Available at: https://www.oecd-ilibrary.org/economics/oecd-economic-surveys-united-kingdom-2009_eco_surveys-gbr-2009-en [accessed 9th July, 2019].

Office for National Statistics (2011a) *Lower Layer Super Output Areas (December 2011) Full Extent Boundaries in England and Wales,* [Online] Available at: http://geoportal.statistics.gov.uk/datasets/da831f80764346889837c72508f046fa_1 [Accessed 10th May 2019].

Office for National Statistics (2011b) *Lower Layer Super Output Areas (December 2011) Population Weighted Centroids,* [Online] Available at: http://geoportal.statistics.gov.uk/datasets/b7c49538f0464f748dd7137247bbc41c_0 [Accessed: 10th May 2019].

Office for National Statistics (2011c) *Census 2011 Quick Statistics,* [Online] available at: https://www.nomisweb.co.uk/query/select/getdatasetbytheme.asp?theme=75&subgrp=Quick+Statistics [Accessed 1st June 2019].

Oostendorp, R., Krajzewicz, D., Gebhardt, L and Heinrichs, D. (2019) Intermodal mobilities in cities and its contribution to accessibility, *Applied Mobilities,* 4(2), Pp.183-199.

Openshaw, S and Taylor, P. J. (1997) A Million or so Correlation Coefficients: Three Experiments on the Modifiable Area Unit Problem. In Wrigley, N. (eds) *Statistical Applications in the Spatial Sciences,* Pp.127-144. Pion, London.

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J.M., Wardman, M.R., Shires, J.D. and White, P. (2006) The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport Policy*, 13(4), Pp.295-306.

RAC (2019) *Speed Cameras – Our Guide to Staying Safe and Legal,* [Online] Available at:https://www.rac.co.uk/drive/advice/cameras/speed-cameras/ [Accessed 1st July 2019].

Recent Rail Times (2019) *Recent Train Times Search Results,* [Online] Available at: https://www.recenttraintimes.co.uk/ [Accessed 10th June 2018].

Saaty, T, L. (1980) *The Analytic Hierarchy Process*. McGraw Hill, New York.

Saaty, T, L. and Vargas, L, G. (1987) Uncertainty and rank order in the analytic hierarchy process, *European Journal of Operational Research,* 23(1), Pp.107-117.

Saelens, B, E. Sallis, J, F and Frank, L, D. (2003) Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures, *Annals of Behavioural Medicine,* 25(2), Pp.80-91.

Salonen, M., Broberg, A., Kytta, M and Toivonen, T. (2014) Do suburban residents prefer the fastest or low-carbon travel modes? Combining public participation GIS and multimodal travel time analysis for daily mobility research, *Applied Geography,* 53, Pp.438-448.

Sanchez, T (2004) 'Connecting Mass Transit and Employment', In Hensher, D, A. Button, K, J. Haynes, K, E and Stopher, P, R (eds) *Handbook of Transport Geography and Spatial Systems,* Emerald, Bingley, Pp. 111-124.

Schneider, R, H and Kitchen, T (2007) *Crime prevention and the built environment,* Routledge, New York,

Scikit Learn (2019) *Sklearn Preprocessing Standard Scaler* [Online] Available at:https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html [Accessed: 25th July 2019].

Sherwin, H and Parkhurst, G. (2010) The promotion of bicycle access to the railway network as a way of making better use of the existing network and reducing car dependence. *European Transport Conference, Glasgow October 2010.*

Stead, D. (2001) Relationships between Land Use, Socioeconomic Factors and Travel Patterns in Britain, *Environment and Planning B: Urban Analytics and City Science,* 28(4), Pp.499-528.

Stradling, S, G., Anable, J and Carreno, M. (2007) Performance importance and user disgruntlement: A Six-step method for measuring satisfaction with travel modes, *Transportation Research Part A,* 41, Pp.98-106.

Susilo, Y, O. and Cats, O. (2014) Exploring key determinants of travel satisfaction for multi-modal trips by different traveller groups, *Transportation Research Part A,* 67, Pp. 366-380.

Transport for London (2010) *Measuring Public Transport Accessibility Levels PTAL's Summary,* [Online] Available at: https://files.datapress.com/london/dataset/public-transport-accessibility-levels/PTAL-methodology.pdf [Accessed 15th May 2019].

Transport for London (2015) *Assessing transport connectivity in London,* [Online] Available at: http://content.tfl.gov.uk/connectivity-assessment-guide.pdf [Accessed: 15th May 2019].

Transport for London (2017) *Accessible Bus Stop Design Guidance.* London: Transport for London. [Online] Available at: http://content.tfl.gov.uk/bus-stop-design-guidance.pdf [Accessed 20th June 2019].

Transport for London (2019) *WEBCAT,* [Online] Available at: https://tfl.gov.uk/info-for/urban-planning-and-construction/planning-with-webcat/webcat [Accessed 15th May 2019].

Tranter, P. (1995) 'Behind closed doors: women, girls and mobility', *Women and Transport Conference*, Adelaide.

Tresidder, M (2005) *Using GIS to Measure Connectivity: An Exploration of Issues.* [Online] Available at: http://www.web.pdx.edu/~jdill/Tresidder_Using_GIS_to_Measure_Connectivity.pdf [Accessed 20th June 2019].

Tumlin, J (2012) *Sustainable Transportation Planning,* Wiley, New Jersey.

UCLA (2018) *What does Cronbach's Alpha Mean?* [Online] Available at: https://stats.idre.ucla.edu/spss/faq/what-does-cronbachs-alpha-mean/ [Accessed 25th July 2019].

Wade, T and Sommer, S (2006) *A to Z GIS.* ESRI, California.

Wakenshaw, G and Bunn, N. (2015) *How far do people walk?* [Online] Available at: https://www.wyg.com/uploads/files/news/WYG_how-far-do-people-walk.pdf [Accessed 20th July 2019].

Wang Ng, K., Tian, G and Tang, M (2011) *Dirichlet and Related Distributions: Theory, Methods and Applications,* Wiley, Sussex.

Weisstein, E, W. (2019) *Eigenvalue,* [Online] Available at: http://mathworld.wolfram.com/Eigenvalue.html [Accessed: 23rd July 2019].

Widaman, K, F. (1993) Common Factor Analysis versus Principle Component analysis: Differential Bias in Representing Model Parameters? *Multivariate Behavioral research,* 28(3), Pp.263-331.

Winter, J, C, F. and Dodou, D. (2013) Common Factor Analysis versus Principle Component analysis: A Comparison of Loadings by Means of simulations, *Communications in Statistics and Computation,* 45(1), Pp.299-321.

Wong, D. (2004) The Modifiable Areal Unit Problem. In Donald, J, Warf, B and Hansen, K (eds) *WorldMinds: Geographical Perspectives on 100 Problems: Commemorating the 100th Anniversary of the Association of America Geographers 1904-2004,* Springer Science & Business Media, California, Pp.571-575.

Zadeh, L, A. (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems,* 90, Pp.111-127.

Zhu, M and Ghodsi, A. (2006) Automatic dimensionality selection from the scree plot via the use of profile likelihood, *Computational Statistics and Data Analysis,* 51, Pp.918-930.

9. Appendix

Appendix A: Recent Rail Times raw data format

An extract from the Recent Rail Times raw data can be seen below. Data was extracted into a PDF for all stations in Hertfordshire for all services to main end of line stops.

TOC	Scheduled Times			% Svc	% Arrivals								Actual Arrival Time i	
	d BAY	a HFN	Dur	CANC/NR	On Time	RT-2L	RT-5L	RT-10L	RT-15L	RT-20L	RT-30L	RT-60L	Average	Yesterda
TL	00:01	00:06	5m	0%	67%	67%	67%	67%	100%	100%	100%	100%	**00:09** 3½L	
TL	00:06	00:11	5m	7%	80%	93%	93%	93%	93%	93%	93%	93%	**00:12** 1L	00:09 RT
TL	00:26	00:31	5m	0%	80%	80%	87%	93%	100%	100%	100%	100%	**00:32** 1½L	00:30 RT
TL	00:30	00:35	5m	0%	75%	75%	75%	75%	100%	100%	100%	100%	**00:38** 3½L	
TL	00:36	00:42	6m	7%	73%	80%	80%	87%	93%	93%	93%	93%	**00:44** 2L	00:42 RT
TL	00:56	01:02	6m	7%	80%	93%	93%	93%	93%	93%	93%	93%	**01:03** 1L	01:02 RT
TL	01:04	01:10	6m	0%	67%	67%	67%	67%	100%	100%	100%	100%	**01:14** 4½L	
TL	01:07	01:11	4m	0%	58%	67%	83%	83%	83%	92%	100%	100%	**01:14** 3½L	01:11 RT
TL	05:43	05:49	6m	0%	85%	90%	90%	100%	100%	100%	100%	100%	**05:49** ½L	05:50 1L

Appendix B: Analytical Hierarchy Questionnaire Participant Information Sheet

Exploring the Multi-modal opportunity of Hertfordshire, England

UCL Department: Centre for Advanced Spatial Analysis, The Bartlett School of Planning.

Multi-modal trips, where an individual uses more than one mode of transport within a single trip, is discussed as way to combat growing traffic congestion and environmental problems. However, there are key barriers which affects the ability for an individual to undertake a multi-modal trip. In order to assess these, an analysis framework is being developed to understand an area's multi-modal opportunity. Multi-modal opportunity is used within this study to encompass all the factors that determine how easy it is for an individual to undertake a multi-modal trip.

To understand how important each trip determinant is within an area, a weighting needs to be developed to create an overall 'multi-modal opportunity' score. This survey is being undertaken to develop a weighting. Results from the survey will be used to weight the factors identified from wider literature that affect an areas multi-modal opportunity. The weighting from 10 participants will be aggregated before use, allowing no participant to be individually identified.

Participants are being selected based on their knowledge on transport accessibility and planning. If you choose to participate in the survey, it is expected that the survey will take 15 minutes to complete. Participation within this survey is voluntary and can discontinue at any point of the study through using the contact details below.

The researcher is currently studying for a masters in Spatial Data Science and Visualisation at University College London. This research is being used to support the development of a master's dissertation entitled: Exploring the Multi-modal opportunity of Hertfordshire, England. The outcome of this research will be a dissertation and interactive web map. Copies of this will be distributed to the participants upon request.

Thank you for taking the time to read the participant information sheet. This study has been approved by the UCL Research Ethics Committee Project ID Number: 16169/001. Should you wish to complain, please see the contact details of the researcher, principle investigator and UCL Ethics Committee below.

Researcher Contact Details:
Hannah Gumble, hannah.gumble.17@ucl.ac.uk

Principle Researcher Contact Details:

Anahid Basiri, a.basiri@ucl.ac.uk

UCL Ethics Committee: ethics@ucl.ac.uk

Appendix C: Analytical Hierarchy Questionnaire

Exploring the Multi-modal opportunity of Hertfordshire, England

Thank you for taking the time to complete this survey. Please contact hannah.gumble.17@ucl.ac.uk for any enquires.

Multi-modal trips are discussed as way to combat growing traffic congestion and environmental problems, however, there are key barriers when considering multi-modal mobility. In order to understand these barriers a framework is being developed to assess multi-modal opportunity. Please can you select out of each pair the most important attribute to you when choosing whether to undertake a multi-modal trip and associate it with a relative importance. Relative importance weights can be seen in the table below and definitions of indicators can be found on the next page.

Importance Definitions:

Importance Intensity	Definition
1	Equal Importance
2	Weak or Slight
3	Moderate Importance
4	Moderate Plus
5	Strong importance
6	Strong Plus
7	Very Strong
8	Very, very strong
9	Extreme Importance

Example:

Indicator A	Indicator B	Most important indicator	Relative Importance intensity
Safety	Cleanliness	A	6

Indicator A	Indicator B	Most important indicator	Relative Importance intensity
Safety	Local Environment Characteristics		
Safety	Local Demographics		
Safety	Network Connectivity		
Local Environment Characteristics	Local Demographics		
Local Environment Characteristics	Network Connectivity		
Network Connectivity	Local Demographics		

*Definitions of the indicators can be found on the next page.

Indicator A	Indicator B	Most important indicator	Relative Importance intensity
Number of Cycle Parking Spaces	Bicycle parking CCTV		
Number of Cycle Parking Spaces	Cycle Hire		
Number of Cycle Parking Spaces	Accessible Parking		
Number of Cycle Parking Spaces	Waiting Room facilities		
Number of Cycle Parking Spaces	Baby Changing Facilities		
Number of Cycle Parking Spaces	Toilet Facilities		
Number of Cycle Parking Spaces	Local bus connections		
Number of Cycle Parking Spaces	Average Service Delay		
Number of Cycle Parking Spaces	Average Service Cancellation		
Number of Cycle Parking Spaces	Route options		
Bicycle parking CCTV	Cycle Hire		
Bicycle parking CCTV	Accessible Parking		
Bicycle parking CCTV	Waiting Room facilities		
Bicycle parking CCTV	Baby Changing Facilities		
Bicycle parking CCTV	Toilet Facilities		
Bicycle parking CCTV	Local bus connections		
Bicycle parking CCTV	Average Service Delay		
Bicycle parking CCTV	Average Service Cancellation		
Bicycle parking CCTV	Route options		
Cycle Hire	Accessible Parking		
Cycle Hire	Waiting Room facilities		
Cycle Hire	Baby Changing Facilities		
Cycle Hire	Toilet Facilities		
Cycle Hire	Local bus connections		
Cycle Hire	Average Service Delay		
Cycle Hire	Average Service Cancellation		
Cycle Hire	Route options		
Accessible Parking	Waiting Room facilities		
Accessible Parking	Baby Changing Facilities		
Accessible Parking	Toilet Facilities		
Accessible Parking	Local bus connections		
Accessible Parking	Average Service Delay		
Accessible Parking	Average Service Cancellation		
Accessible Parking	Route options		
Waiting Room facilities	Baby Changing Facilities		
Waiting Room facilities	Toilet Facilities		
Waiting Room facilities	Local bus connections		
Waiting Room facilities	Average Service Delay		
Waiting Room facilities	Average Service Cancellation		
Waiting Room facilities	Route options		
Baby Changing Facilities	Toilet Facilities		
Baby Changing Facilities	Local bus connections		
Baby Changing Facilities	Average Service Delay		
Baby Changing Facilities	Average Service Cancellation		
Baby Changing Facilities	Route options		

Toilet Facilities	Local bus connections		
Toilet Facilities	Average Service Delay		
Toilet Facilities	Average Service Cancellation		
Toilet Facilities	Route options		
Local bus connections	Average Service Delay		
Local bus connections	Average Service Cancellation		
Local bus connections	Route options		
Average Service Delay	Average Service Cancellation		
Average Service Delay	Route options		
Average Service Cancellation	Route options		

Indicator A	Indicator B	Most important indicator	Relative Importance intensity
Route Options	Surveillance Coverage		
Route Options	Lighting		
Surveillance Coverage	Lighting		

Indicator Definitions

Safety
Safety is used to encompass perceived safety on different modes and in different environments.

Local Characteristics
Local Characteristics encompasses variables that describe the local environment including land use, local environmental quality and street lighting.

Local Demographics and potential
Local demographics encompasses variables that would affect an individual's willingness to undertake a multi-modal trip.

Railway Station Service and Route Choice
Railway service and route choice is used to encompass the availability of services, service frequency, reliability and

Bus Stop Environment and Route Choice
Bus stop service and route choice is used to encompass the availability of services, service frequency and the quality of the bus stop weighting environment.

Network Connectivity
Network connectivity is used to encompass the connectivity of each transport modes network and the intermodal connectivity options.

Appendix D: Land use Data Reclassifications

The aggregation groups used for land use reclassification can be seen below:

Original Land Use Classification	Index Classification Group
Agricultural Land	Agriculture
Agricultural Buildings	Agriculture
Community Buildings	Community facilities
Defence	Industrial
Forestry/Woodland	Parks and open space
Rough grassland	Parks and open space
Highways and roads	Infrastructure
Industry	Industrial
Offices	Offices
Retail	Commercial
Leisure (indoor)	Community facilities
Minerals and Mining	Industrial
Natural Land	Parks and open space
Outdoor Recreation	Community facilities
Communal Accommodation	Community facilities
Residential	Residential
Residential Gardens	Residential
Storage and Warehousing	Industrial
Transport	Infrastructure
Utilities	Industrial
Vacant Land	Industrial
Water	Parks and open space
Undeveloped Land	Parks and open space
Landfill and Waste Disposal	Industrial
Unidentified building	Agriculture
Unidentified general manmade surface (not roadside)	Industrial
Unidentified structure	Industrial
Unknown surface type with no classification	Industrial

Appendix E: Fix My Street variable selection

Variable Chosen
Broken Kerbs
Countryside RoW
Dead Animals
Debris/Spillage
Discarded Syringes
Accumulated Litter
Dog Fouling
Dog Litter Bin
Fallen/Dangerous Trees
Flooding
Fly-tipping
Gates/Stiles RoW
Graffiti
Grass/Weed Control
Gritting
Barriers/Bollards on RoW
Highway Condition
Ice & Snow
Leafing
Litter/Litter Bin
Benches/Bicycle Racks
Obstruction
Open Spaces/Parks
Pavement /Footway Defects
Potholes
Public Toilets
Right of Way
Road Surface Defects
Road/Pavement Defects
Street Furniture
Street Lights
Street Signs
Trees and Hedges RoW
Trees on RoW
Trip Hazard

Appendix F: Python script to process Recent Railway times Data

The following code was used to produce the railway station delay indicators. The same structure was used to process the cancellation factors.

```
# Set up the working environment
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import plotly
import seaborn as sns

pd.set_option('display.max_rows',200)

Test = pd.read_csv('sat-FileList.csv',encoding = "ISO-8859-1")

File=[]
for i, row in Test.iterrows():
    text = Test.File[i]
    File.append(text)

MaxNum = len(File)

#set up output dataframe
Output = pd.DataFrame(columns=['Origin', 'Destination', 'AM count', 'AM Avg Delay', 'IP
count', 'IP Avg Delay', 'PM count',
                        'PM Avg Delay','OP count','OP Avg Delay'])

#loop through all the .csv's
i=0
while i < MaxNum-1:
    temp = File[i]
    #Define origin station from file name
    Ori = temp[4:7]
    #Define destination from file name
    Dest = temp[8:11]
    #Read in data
    RailData = pd.read_csv(File[i],encoding = "ISO-8859-1")
    #rename columns to make consistent
    RailData.columns = ['TOC', 'Origin', 'Destination', 'Dur', 'CANC', 'OnTime','2L', '5L', '10L',
'15L', '20L', '30L', '60L', 'AverageArrival', 'Delay']
    RailData
    #Fill null values with 0
    RailData.fillna(0)
    #create a column to record the time period
    RailData['TP']= 'default value'
    #Define the timeperiod from the train departing time
    RailData.loc[RailData.Origin <= '06:59', 'TP'] = 'OP'
    RailData.loc[(RailData.Origin >= '07:00')&(RailData.Origin <= '09:59'), 'TP'] = 'AM'
    RailData.loc[(RailData.Origin >= '10:00')&(RailData.Origin <= '15:59'), 'TP'] = 'IP'
    RailData.loc[(RailData.Origin >= '16:00')&(RailData.Origin <= '18:59'), 'TP'] = 'PM'
    RailData.loc[RailData.Origin >= '19:00', 'TP'] = 'OP'
    RailData
    # fill output data frame. counting the number of trains in each O-D pair and the average
delay factor for trains in each time period.
    Output = Output.append({'Origin': Ori, 'Destination': Dest,
                    'AM count': RailData['Dur'][RailData['TP'] == 'AM'].count(),
                    'AM Avg Delay':RailData['Delay'][RailData['TP'] == 'AM'].mean(),
```

```
                    'IP count': RailData['Dur'][RailData['TP'] == 'IP'].count(),
                    'IP Avg Delay':RailData['Delay'][RailData['TP'] == 'IP'].mean(),
                    'PM count': RailData['Dur'][RailData['TP'] == 'PM'].count(),
                    'PM Avg Delay':RailData['Delay'][RailData['TP'] == 'PM'].mean(),
                    'OP count': RailData['Dur'][RailData['TP'] == 'OP'].count(),
                    'OP Avg Delay':RailData['Delay'][RailData['TP'] ==
'OP'].mean(),},ignore_index=True)
   i=i+1
   print(File[i])

#Calculate the weighted average delay factor.
Output['AM Fac'] = Output.apply(lambda row: row['AM count'] * row['AM Avg Delay'],
axis=1)
Output['IP Fac'] = Output.apply(lambda row: row['IP count'] * row['IP Avg Delay'], axis=1)
Output['PM Fac'] = Output.apply(lambda row: row['PM count'] * row['PM Avg Delay'],
axis=1)
Output['OP Fac'] = Output.apply(lambda row: row['OP count'] * row['OP Avg Delay'],
axis=1)

# replace any null values in the data frame with 0.
Output = Output.replace(np.nan, 0)

#Create a dataframe with the total number of trains originating from each station in each
timeperiod.
Temp = Output.groupby(['Origin']).sum()
TotalTrain=Temp.drop(columns=['AM Avg Delay','IP Avg Delay','PM Avg Delay','OP Avg
Delay','AM Fac','IP Fac','PM Fac','OP Fac'])
TotalTrain

#Created Weighted Delay dataframe.
WeightedDelay = Temp.drop(columns=['AM Avg Delay','IP Avg Delay','PM Avg Delay','OP
Avg Delay'])

#Calculate the Weighted Average Delay factors
WeightedDelay['AMDelayFac']= WeightedDelay.apply(lambda row: row['AM Fac']/ row['AM
count'],axis=1)
WeightedDelay['IPDelayFac']= WeightedDelay.apply(lambda row: row['IP Fac']/ row['IP
count'],axis=1)
WeightedDelay['PMDelayFac']= WeightedDelay.apply(lambda row: row['PM Fac']/ row['PM
count'],axis=1)
WeightedDelay['OPDelayFac']= WeightedDelay.apply(lambda row: row['OP Fac']/ row['OP
count'],axis=1)

#View final dataframe
WeightedDelay

#Output Weighted delay factor
WeightedDelay.to_csv(r'WeightedDelay.csv',index=True)
```

Appendix G: Indicator Processing Code Sample

The following code was used to process the area attribute domains and run principle component analysis.

```
# import libraries, and set pd as the pandas alias
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import plotly
import seaborn as sns
# Run to allow more data to be displayed
pd.set_option('display.max_rows', 200)
# ensure graphs display properly in the notebook
%matplotlib inline

# Read in Indicator Data
LC = pd.read_csv('Localcharacteristics/LocalCharacteristics_Raw.csv')
#Print the first 5 rows to understand data format
LC.head(3)

#Import the Min Max Package package
from sklearn.preprocessing import MinMaxScaler
# create Min Max class object
scale = MinMaxScaler(feature_range=(0, 1))
# Min Max rescailer
LC[['IR_cycling','IR_HWY','IR_FTP','Environmental Quality','Walkabe speed','Population Density','Landuse Mix','Cycle Facility / Route Standardised','Leisure Route Route Standardised','National Cycle Network Route Standardised']] = scale.fit_transform(LC[['IR_cycling','IR_HWY','IR_FTP','Environmental Quality','Walkabe speed','Population Density','Landuse Mix','Cycle Facility / Route Standardised','Leisure Route Route Standardised','National Cycle Network Route Standardised']])

#Check the Standard Scailer worked
LC.min()
LC.max()

# Create the Correlation Matrix
LC.corr(method='spearman').style.format("{:.2}").background_gradient(cmap=plt.get_cmap('coolwarm'))

# Understand the distribution of the data
ScatterPlot = pd.plotting.scatter_matrix(LC, alpha=1, figsize=(30, 30), diagonal='hist')

# Check the Internal Consistency using Cronbach's Alpha
# split data table into data X and class labels y
X = LC.iloc[:,1:10].values
y = LC.iloc[:,0].values
# Preview
X
Y

# Calculate variables for Cronbach Alpha
# Convert input to anarray
itemvalues = np.asarray(X)
# Calculate variance of array elements
itemvarience = itemvalues.var(axis=1, ddof=1)
# Calculate the score by summing along the coloumns
scores = itemvalues.sum(axis=0)
# Count the total variables in the array
```

```
TotalVal = len(itemvalues)

# Calculate CronbachAlpha
CronbachAlphaCalc = TotalVal / (TotalVal-1.) * (1 - itemvarience.sum() /
scores.var(ddof=1))
# Prints the outcome score
print("Cronbach's Alpha = ", CronbachAlphaCalc)

# Prepare the data for Principle Component Analysis
# Remove the LOSA Coloumn
PCAData = LC.drop(['lsoa11cd'], axis=1)
PCAData.head(3)

#Standardise data
from sklearn.preprocessing import StandardScaler
SScaler = StandardScaler()
PCADataScaled = SScaler.fit_transform(PCAData)
PCADataScaled

#Import the PCA package
from sklearn.decomposition import PCA
# create PCA class object
pca = PCA()
# run PCA and write components back to array at_decomp
at_decomp = pca.fit_transform(PCADataScaled)

# See how many principle components have been created.
PCsCreated = pca.components_
PCA_All = pd.DataFrame(list(PCsCreated))
PCA_All.head(3)

# Look at the explained variance of each principle component
Exp_Var = pca.explained_variance_
print("Explained Variance = ", Exp_Var)

# Examine the explained variance of each principle component as a ratio
Exp_Var_ratio =pca.explained_variance_ratio_
print("Explained Variance Ratio = ", Exp_Var_ratio)

# Sets the seaborn plot type
sns.set(style="whitegrid")
sns.set_style("ticks")
plt.style.use('ggplot')
# Produce plot of explained variance
explained_variance = pca.explained_variance_ratio_
plt.plot(pca.explained_variance_ratio_)
# Clearly, our first few components looks useful, as they capture much of the variance
plt.title("Explained Variance of each principle component: Local Characteristics")
plt.xlabel("PCA")
plt.ylabel("Explained Variance")
plt.xticks(np.arange(0,10, step=1))
plt.show()

# Extract the first principle components to be carried forward and save to final_sel
final_at = at_decomp
final_sel=[]
for i in range(len(final_at)):
   final_sel.append(final_at[i][:10])
len(final_sel)
final_sel
```

```
# Add the list of PCAs selected to a data frame.
PCA_Sel = pd.DataFrame(list(final_sel))
# Uncomment to see the dataframe
PCA_Sel.head(3)

#Merge principle components and Attributes in to one dataframe
df_row = pd.merge(LC, PCA_Sel, left_index=True, right_index=True)
df_row.head(3)

# Understand the Relationship between the principle components and attributes
pd.concat([LC, PCA_Sel], axis=1,keys=['LC', 'PCA_Sel']).corr(method='spearman').loc['LC',
'PCA_Sel'].style.format("{:.2}").background_gradient(cmap=plt.get_cmap('coolwarm'))

# Normalise the Principle Components ready for aggregation
# create Min Max class object
scale = MinMaxScaler(feature_range=(0, 1))
# Scale the principle components
PCA_Sel_Scaled = scale.fit_transform(PCA_Sel)
PCA_Sel_Scaled
PCA_Sel_Scaled_df = pd.DataFrame(list(PCA_Sel_Scaled))
PCA_Sel_Scaled_df

# Relate the principle components to the corresponding LSOA and add to data frame
LSOA = LC[['lsoa11cd']]
FinalAtributes = pd.merge(LSOA, PCA_Sel_Scaled_df, left_index=True, right_index=True)
FinalAtributes.head(3)

#Export the final Principle Components
FinalAtributes.to_csv(r'LChar_PCA_All.csv',index=False)

# Save the explained variance to a dataframe and calculate the squareroot.
ExplainedVar = pd.DataFrame(list(Exp_Var))
ExplainedVar['Var Sqrt'] = np.sqrt((ExplainedVar[0]))
ExplainedVar

#Output the Factors.
ExplainedVar.to_csv(r'LChar Factors.csv',index= True)
```

Appendix H: Mapbox Code Extracts

The following code was used for the online map. It has been trimmed due to the repetitive nature of the code. Full code can be seen on GitHub repository.

```
<!DOCTYPE html>
<html>
  <head>
    <meta charset='utf-8' />
    <title>Multi-Modal Opportunity of Hertfordshire</title>
    <meta name='viewport' content='initial-scale=1,maximum-scale=1,user-scalable=no' />
    <script src='https://api.tiles.mapbox.com/mapbox-gl-js/v1.2.0/mapbox-gl.js'></script>
    <link href='https://api.tiles.mapbox.com/mapbox-gl-js/v1.2.0/mapbox-gl.css'
rel='stylesheet' />
    <style>
      body {
        margin: 0;
        padding: 0;
      }
       #map {
        position: absolute;
        top: 0;
        bottom: 0;
        width: 100%;
      }
  .map-overlay {
      font: 12px/20px 'Helvetica Neue', Arial, Helvetica,sans-serif;
      position: fixed;
      width: 280px;
      top: 0;
      left: 0;
      padding: 10px;
            z-index: +1;
}
 .map-overlay-inner {
      position: absolute;
      top: 5px;
      left: 5;
      background-color: rgba(255, 255, 255, 0.7);
      box-shadow:0 1px 2px rgba(255, 255, 255, 0.20);
```

```
     border-radius: 10px;
     padding: 10px;
     margin-bottom: 10px;
     color: black;
    z-index: -1;
  }
  .key_fig {
    background-color: #f7fcf0;
    border-style: solid
    padding: 10px 24px;
    border-radius: 8px;
    color: black;
    padding: 5px 5px;
    text-align: center;
    text-decoration: none;
    display: inline-block;
    font-size: 16px;
    opacity: 0.7;
    display: block;
    border-color: black;
}
.key_fig2 {background-color: #e0f3db;}
.key_fig3 {background-color: #ccebc5;}
.key_fig4 {background-color: #a8ddb5;}
.key_fig5 {background-color: #7bccc4;}
.key_fig6 {background-color: #4eb3d3;}
.key_fig7 {background-color: #2b8cbe;}
.key_fig8 {background-color: #0868ac;}
.key_fig9 {background-color: #084081;}
.key_fig10 {background-color: #DFDFDF;}
.key_fig11 {background-color: #3E3C3C;}
.key_fig12 {background-color: #A3A1CA;}
.key_fig13 {background-color: #A1CAAA;}
.key_fig14 {background-color: #BEF184;}
.key_fig15 {background-color: #E88463;}
.key_fig16 {background-color: #FF3F00;}
.Key {
     font: 12px/20px 'Helvetica Neue', Arial, Helvetica, sans-serif;
     position: absolute;
     display: block;
```

```
      bottom: 0;
      right: 0;
      background-color: rgba(136,136,136, 0.80);
      box-shadow:0 1px 2px rgba(255, 255, 255, 0.20);
      border-radius: 10px;
      padding: 5px;
      margin-right: 10px;
      margin-bottom: 20px;
      color: black;}
.Description-Box {
      font: 12px/20px 'Helvetica Neue', Arial, Helvetica, sans-serif;
      position: absolute;
      bottom: 0;
      left: 400px;
      right: 300px;
      background-color: rgba(255, 255, 255, 0.70);
      box-shadow:0 1px 2px rgba(255, 255, 255, 0.20);
      border-radius: 10px;
      padding: 5px;
      margin-bottom: 20px;
      color: black;}
    </style>
  </head>
  <body>
  <div id='map'></div>
<div class='map-overlay top'>
    <div class='map-overlay-inner'>
        <h2>Multi-Modal Opportunity in Hetrfordshire</h2>
     <p class="defintion"> Multi-modal opportunity is used to encompass the factors that can
influence a person’s opportunity and want to undertake a multi-modal trip .</p>
    <p class="defintion"> Please explore the key components which have been used to
measure multi-modal opportunity in Hertfordshire. .</p>
      <h2>Opportunity Index</h2>
      <input type="radio" name="layers" id="layer1" value="Total_AM"
checked><label>AM</label>
      <input type="radio" name="layers" id="layer2" value="Total_IP"><label>IP </label>
      <input type="radio" name="layers" id="layer3" value="Total_PM"><label>PM </label>
      <input type="radio" name="layers" id="layer19"
value="Saturday"><label>Saturday</label>
      <input type="radio" name="layers" id="layer20"
value="Sunday"><label>Sunday</label>
```

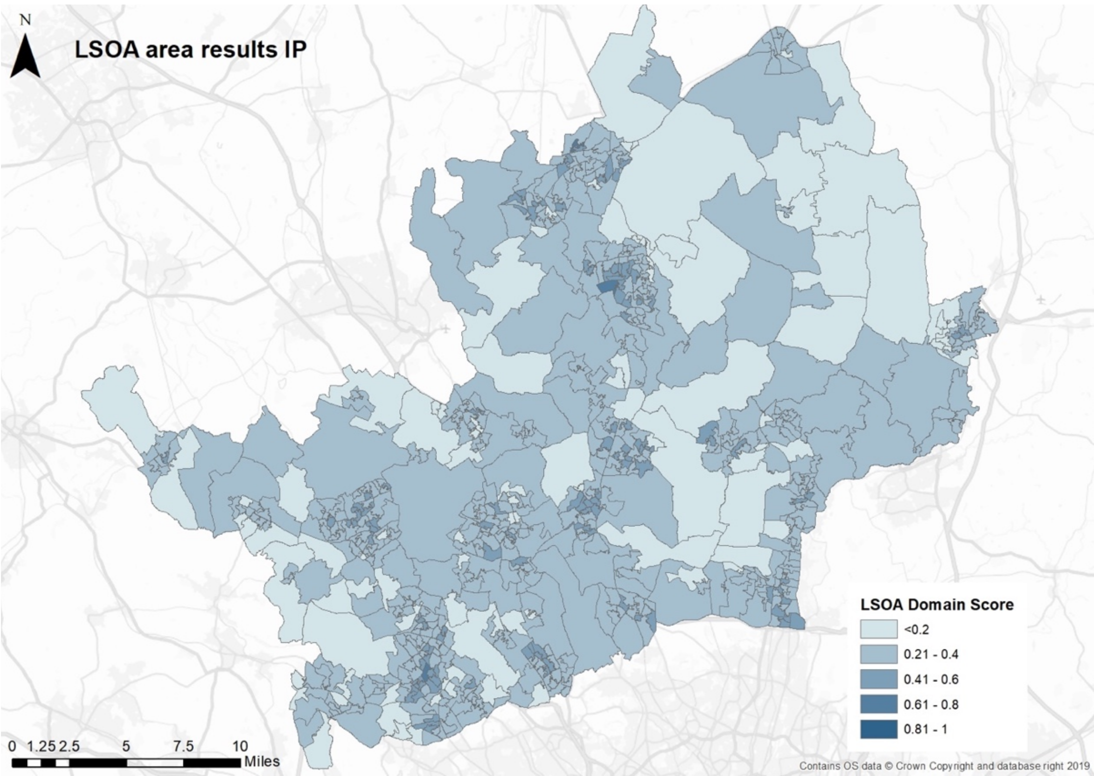
```
     <br>
     <h2>Area Attributes</h2>
     <input type="radio" name="layers" id="layer13" value="Total_AM"><label>AM</label>
     <input type="radio" name="layers" id="layer14" value="Total_IP"><label>IP</label>
     <input type="radio" name="layers" id="layer15" value="Total_PM"><label>PM</label>
     <input type="radio" name="layers" id="layer22"
value="Saturday"><label>Saturday</label>
     <input type="radio" name="layers" id="layer21"
value="Sunday"><label>Sunday</label>
     <br>
    <h2>Bus Stops</h2>
    <input type="radio" name="layers" id="layer7" value="Total_AM"><label>AM</label>
    <input type="radio" name="layers" id="layer8" value="Total_IP"><label>IP</label>
    <input type="radio" name="layers" id="layer9" value="Total_PM"><label>PM</label>
    <input type="radio" name="layers" id="layer23"
value="Saturday"><label>Saturday</label>
    <input type="radio" name="layers" id="layer24" value="Sunday"><label>Sunday</label>
    <h2>Railway Station Cycling</h2>
    <input type="radio" name="layers" id="layer10" value="Total_AM"><label>AM</label>
    <input type="radio" name="layers" id="layer11" value="Total_IP"><label>IP</label>
    <input type="radio" name="layers" id="layer12" value="Total_PM"><label>PM</label>
    <input type="radio" name="layers" id="layer27"
value="Saturday"><label>Saturday</label>
    <input type="radio" name="layers" id="layer28" value="Sunday"><label>Sunday</label>
    <h2>Railway Station Walking </h2>
    <input type="radio" name="layers" id="layer16" value="Total_AM"><label>AM</label>
    <input type="radio" name="layers" id="layer17" value="Total_IP"><label>IP</label>
    <input type="radio" name="layers" id="layer18" value="Total_PM"><label>PM</label>
    <input type="radio" name="layers" id="layer25"
value="Saturday"><label>Saturday</label>
    <input type="radio" name="layers" id="layer26" value="Sunday"><label>Sunday</label>
    <br>
    <br>
    <br>
       <p class="credit">Created by Hannah Gumble, UCL MSc Spatial Data Science and
Visualisation.</p>
</div>
</div>
</div>
<div id = "myDIV" class='Key'>
      <center><h2>Key</h2></center>
   <center><button class='key_fig'> < 0.1 </button></center>
```

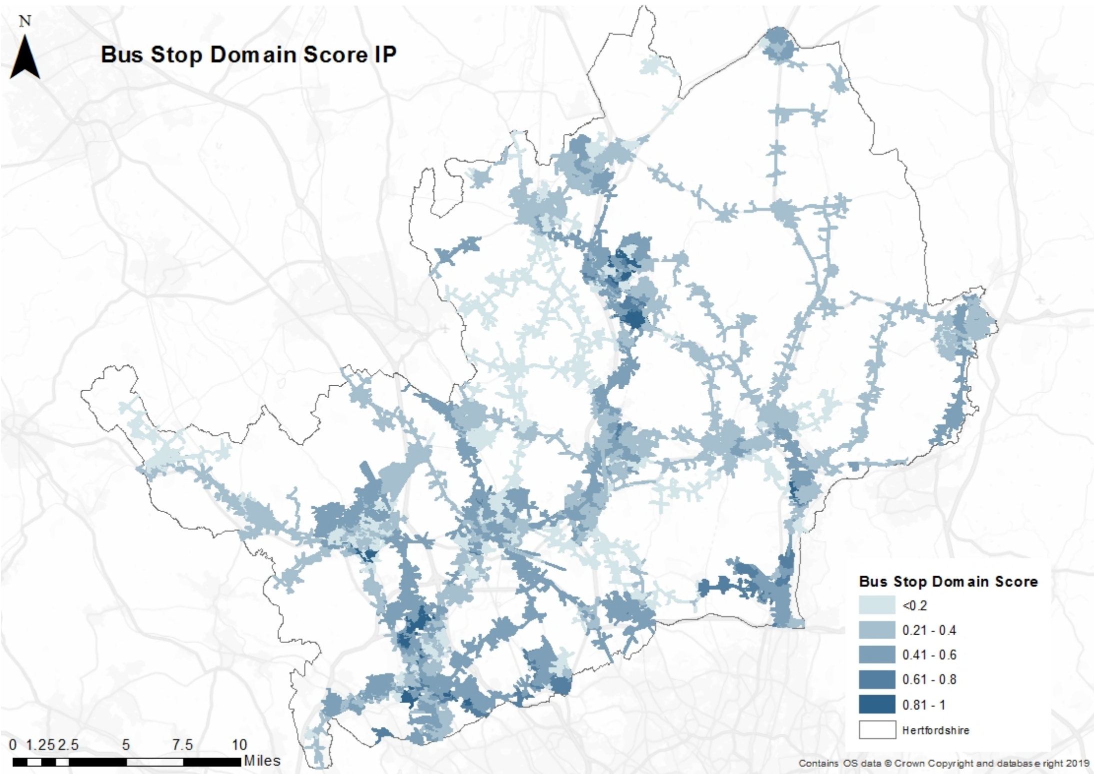
```
    <center><button class='key_fig key_fig2'> 0.11-0.2  </button></center>
    <center><button class='key_fig key_fig3'> 0.21-0.3</button></center>
    <center><button class='key_fig key_fig4'> 0.31-0.4 </button></center>
    <center><button class='key_fig key_fig5'> 0.41-0.5 </button></center>
    <center><button class='key_fig key_fig6'> 0.51-0.6 </button></center>
    <center><button class='key_fig key_fig7'> 0.61-0.7 </button></center>
    <center><button class='key_fig key_fig8'> 0.71-0.8 </button></center>
    <center><button class='key_fig key_fig9'> > 0.81 </button></center>
    </div>
<div id = "myDIV2" class='Key'>
        <center><h2>Key</h2></center>
    <center><button class='key_fig key_fig10'> < 1.6 </button></center>
    <center><button class='key_fig key_fig11'> 1.6-3.5  </button></center>
    <center><button class='key_fig key_fig12'> 3.5-6</button></center>
    <center><button class='key_fig key_fig13'> 6-9 </button></center>
    <center><button class='key_fig key_fig14'> 9-12 </button></center>
    <center><button class='key_fig key_fig15'> 12-18 </button></center>
    <center><button class='key_fig key_fig16'> >18 </button></center>
    </div>
    <div class='Description-Box'>
        <p style="margin-top:0px;"> <label id='narrative'>Data Derrived from: Hertfordshire
County Council, Office of National Statistics, Department for Transport, FixMyStreet,
RecentRailTimes, Interlink and National Rail Enquires.</p></label>
    </div>
    </div>
  <script>
  document.getElementById("myDIV").style.display = "none";
  mapboxgl.accessToken =
'pk.eyJ1IjoiaGFubmFoZ3VtYmxlIiwiYSI6ImNqY2oweWtibDJzbmszM3JxemlncXo5ZDAifQ.h
-gteOJJaIq4GxkctofuOg';
  var map = new mapboxgl.Map({
    container: 'map', // container id
    style: 'mapbox://styles/hannahgumble/cjz2kmqyn03rp1dns8kvqk2b7', // stylesheet
location
    center: [-0.209, 51.827], // starting position [lng, lat]
    zoom: 10 // starting zoom
  });
  map.on('load', function() {
    map.addLayer({
      id: 'Total_AM',
      type: 'fill',
```

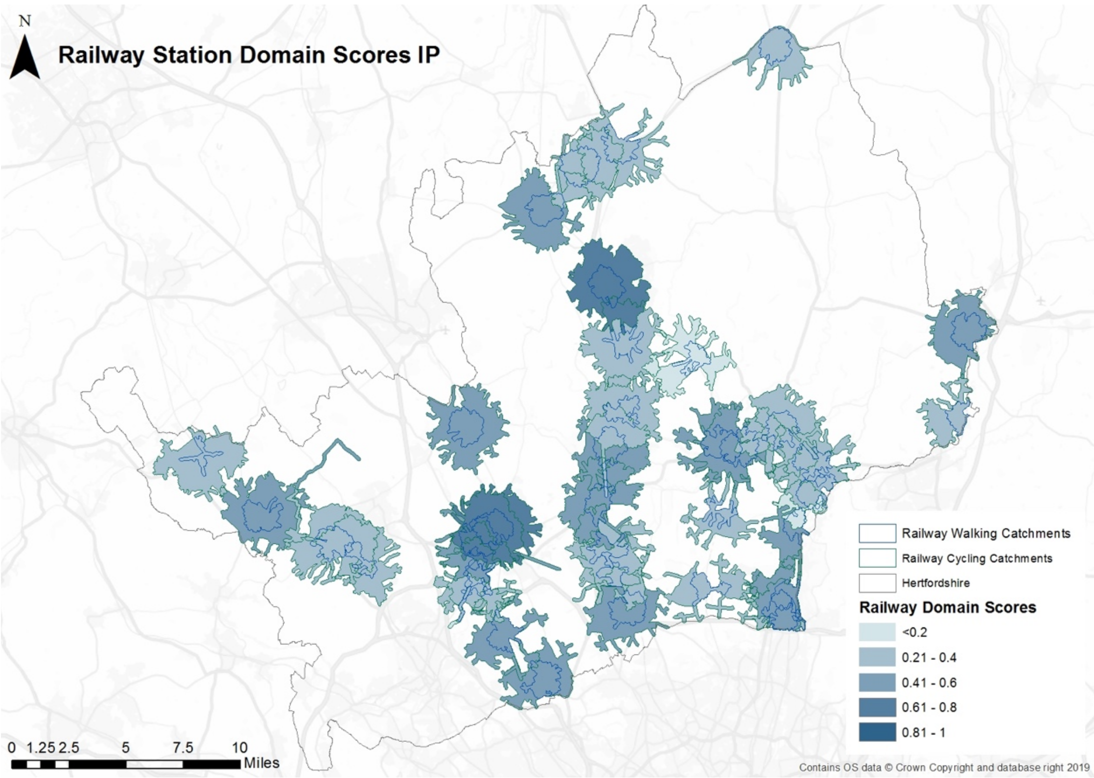
```
  source: {
    type: 'vector',
    url: 'mapbox://hannahgumble.bo4lroz6' },
  'source-layer': 'MapboxLayer-2w0no8',
          'paint': {
        'fill-color':{
        property: 'Total_AM',
        type: 'interval',
        stops: [
          [0.0, '#DFDFDF'],
          [1.6, '#3E3C3C'],
          [3.5, '#A3A1CA'],
          [6, '#A1CAAA'],
          [9, '#BEF184'],
          [12, '#E88463'],
          [18, '#FF3F00'],
          ]
          },
        'fill-opacity': 0.7,
        }
});
    map.addLayer({
  id: 'Cyc_R_AM',
  type: 'fill',
  source: {
    type: 'vector',
    url: 'mapbox://hannahgumble.bo4lroz6'
  },
  'source-layer': 'MapboxLayer-2w0no8',
          'paint': {
        'fill-color':{
        property: 'Cyc_R_AM',
        type: 'interval',
        stops: [
          [0.1, '#F2FFED'],
          [0.2, '#e0f3db'],
          [0.3, '#ccebc5'],
          [0.4, '#a8ddb5'],
          [0.5, '#7bccc4'],
          [0.6, '#4eb3d3'],
```

```
                [0.7, '#2b8cbe'],
                [0.8, '#0868ac'],
                [0.9, '#084081'],
                ]},
            'fill-opacity': 0.0, }
    });
//Example opacity change when layer is selected.
document.getElementById("layer1").addEventListener("click", function(){
map.setPaintProperty('Total_AM','fill-opacity',0.7);
map.setPaintProperty('Total_IP','fill-opacity',0);
map.setPaintProperty('Total_PM','fill-opacity',0);
map.setPaintProperty('Bus_AM','fill-opacity',0);
map.setPaintProperty('Bus_IP','fill-opacity',0);
map.setPaintProperty('Bus_PM','fill-opacity',0);
map.setPaintProperty('Cyc_R_AM','fill-opacity',0);
map.setPaintProperty('Cyc_R_IP','fill-opacity',0);
map.setPaintProperty('Cyc_R_PM','fill-opacity',0);
map.setPaintProperty('MD_R_AM','fill-opacity',0);
map.setPaintProperty('MD_R_IP','fill-opacity',0);
map.setPaintProperty('MD_R_PM','fill-opacity',0);
map.setPaintProperty('Wal_R_AM','fill-opacity',0);
map.setPaintProperty('Wal_R_IP','fill-opacity',0);
map.setPaintProperty('Wal_R_PM','fill-opacity',0);
map.setPaintProperty('Total_MM_S','fill-opacity',0);
map.setPaintProperty('total_MM_1','fill-opacity',0);
map.setPaintProperty('MAX_MAX__1','fill-opacity',0);
map.setPaintProperty('MAX_MAX_MD','fill-opacity',0);
map.setPaintProperty('Total_Sat','fill-opacity',0);
map.setPaintProperty('Total_Sun','fill-opacity',0);
map.setPaintProperty('SUM_Total_','fill-opacity',0);
map.setPaintProperty('SUM_Total1','fill-opacity',0);
map.setPaintProperty('MAX_SUM_To','fill-opacity',0);
map.setPaintProperty('MAX_SUM__1','fill-opacity',0);
document.getElementById("myDIV2").style.display = "block";
document.getElementById("myDIV").style.display = "none";
});
  });
  </script>
  </body>
</html>
```

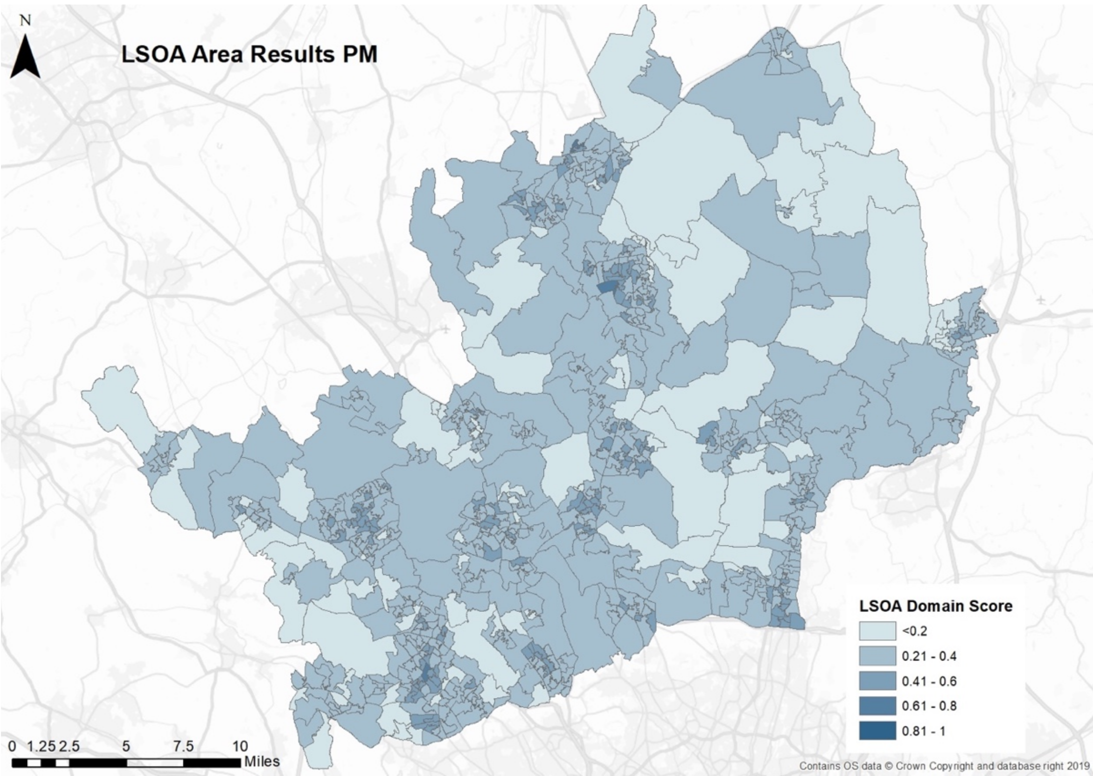
Appendix I: IP Index Domain results

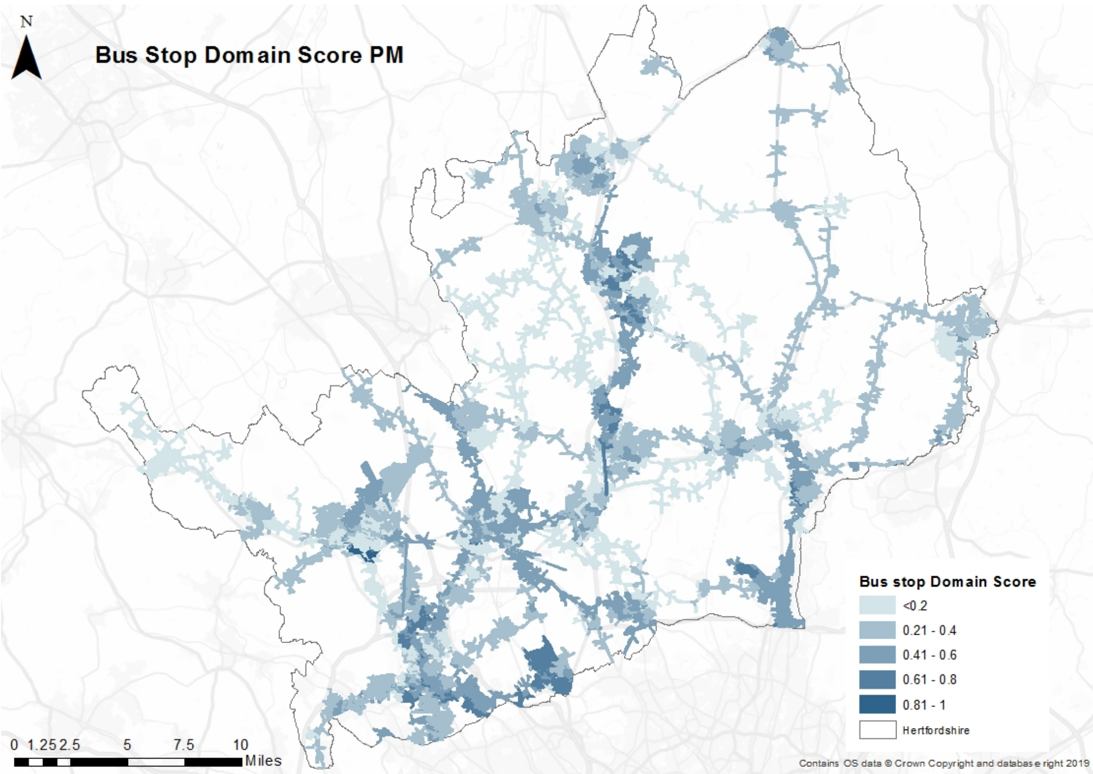


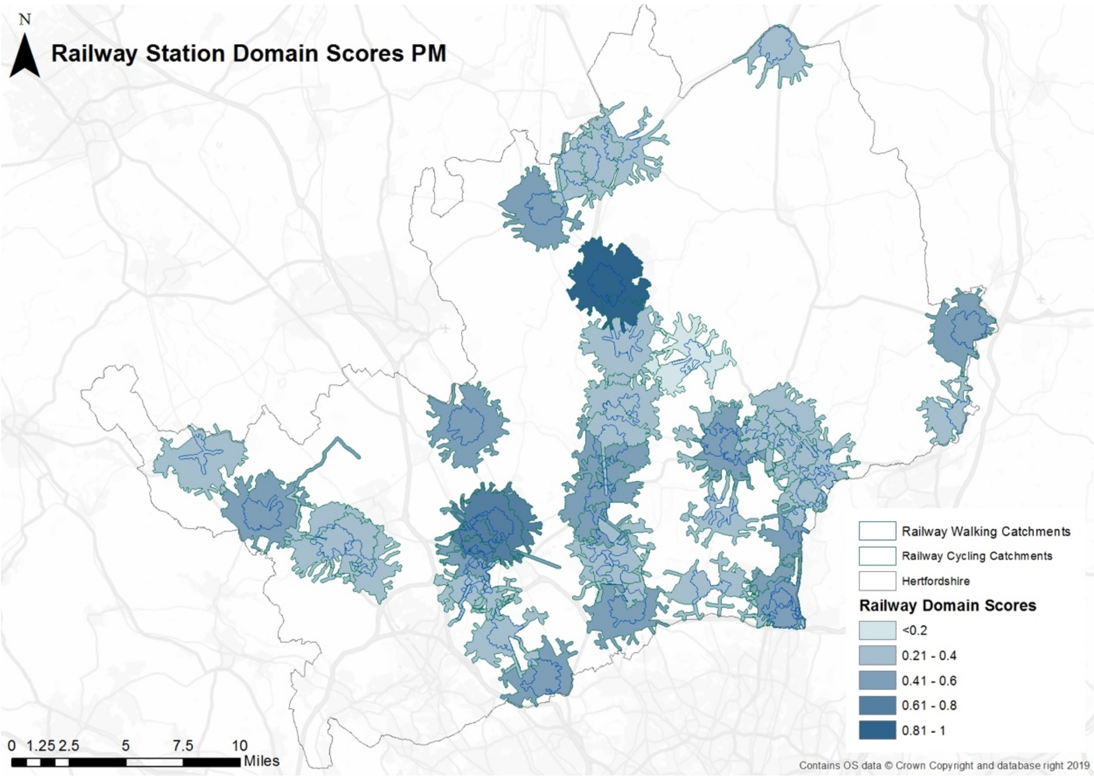



N
Railway Station Domain Scores IP
Railway Walking Catchments
Railway Cycling Catchments
Hertfordshire
Railway Domain Scores
<0.2
0.21 - 0.4
0.41 - 0.6
0.61 - 0.8
0.81 - 1
0 1.25 2.5 5 7.5 10
Miles
Contains OS data © Crown Copyright and database right 2019

Appendix J: PM Domain results






N
Railway Station Domain Scores PM
Railway Walking Catchments
Railway Cycling Catchments
Hertfordshire
Railway Domain Scores
<0.2
0.21 - 0.4
0.41 - 0.6
0.61 - 0.8
0.81 - 1
0 1.25 2.5 5 7.5 10
Miles
Contains OS data © Crown Copyright and database right 2019

Appendix K: Hotspot Results