Expanding the Big E-Society: A Contemporary Model of Internet Access, Exclusion and Investment

Dean Riddlesden
Faculty of Science and Engineering
University of Liverpool

PhD August 2016

Acknowledgements

I would like to extend my thanks to the many people and organisations who have contributed both directly and indirectly to the research presented in this thesis.

I would like to thank my funding body, the Economic and Social Research Council, without the support of whom, my PhD would not have been possible.

I express my profound gratitude to my primary academic supervisor, Professor Alex Singleton, for his enthusiasm and unrelenting support throughout this process. Thank you for the countless hours of technical discussions, knowledge sharing, introductions, opportunities and transatlantic travel that made this experience so rich and valuable.

Similarly, I extend my thanks to my secondary supervisors, Professor Chris Brunsdon and Dr Paul Williamson, who have offered their expertise and guidance throughout.

Special thanks go to my colleagues at the University of Liverpool and Walgreens Boots Alliance for their understanding and encouragement whilst I balanced write-up with my career.

Finally, I would like to thank my family for their immeasurable support, which often felt the most important. I dedicate this thesis to Mum and Dad.

Contents

List of Figures **6**

List of Tables **10**

1 Introduction **13**

1.1 Background 13

1.2 Thesis Structure 14

2 Literature Review **18**

2.1 Background 18

2.1.1 Introduction 18

2.1.2 From the Wires and Infrastructure to the Protocols and Content 19

2.1.3 Early Network Technology 20

2.1.4 The 'Broadband' Internet and Landmark Advances in Communications Technology 23

2.1.5 Government Regulation and Local Loop Unbundling 25

2.1.6 From Fixed to Mobile Broadband 26

2.1.7 Coverage and Throughput 28

2.1.8 Mobile Handset Evolution 31

2.2 Social and Economic Implications of the Internet 32

2.2.1 Advantages of the Internet 32

2.2.2 Enhancing Education 33

2.2.3 Broadband and Economic Advantage 34

2.2.4 Social and Economic Capital 35

2.2.5 Entertainment, Lifestyle and Retail 37

2.2.6 Privacy, Surveillance and Unethical uses of the Internet 41

2.3 Broadband, Governance and Civic Engagement 44

2.3.1 Comparisons with the EU5 . . . 48
2.4 Broadband and the Digital Divide . . . 51
2.4.1 Domestic Digital Differentiation . . . 53
2.5 Conclusion . . . 57

3 Exploring the Geography of Access to Fixed Line Broadband Services in England Using Crowd Sourced Speed Check Data . . . 61
3.1 Introduction . . . 61
3.2 Data Characteristics . . . 63
3.2.1 Crowdsourced Speed Test Datasets . . . 63
3.2.2 Dataset of UK Telephone Exchange Locations . . . 65
3.2.3 Data Cleansing and Validation . . . 66
3.2.4 Coverage and Data Organisation . . . 70
3.3 Data Analysis . . . 71
3.3.1 The Effects of Exchange Proximity . . . 71
3.3.2 Speed and Indicators of Socio-spatial Structure . . . 73
3.3.3 Quantifying Spatial Disparities in Broadband Performance . . . 81
3.3.4 The Geography of Next Generation Infrastructure . . . 84
3.3.5 Exploring Temporal Variations in Performance . . . 86
3.4 A Case Study of Local Needs and Private Infrastructure Overhaul, Kings Cliffe, Northamptonshire. . . . 89
3.4.1 Background and Funding . . . 89
3.4.2 Service Implementation . . . 91
3.5 A Consideration of Mobile Broadband: A Substitute or Complementary Infrastructure? 93
3.6 Conclusions . . . 97

4 A Geography of Internet Demand and Engagement . . . 99
4.1 Introduction . . . 99
4.1.1 Objectives . . . 100
4.2 Data Characteristics . . . 101
4.2.1 The 2013 Oxford Internet Survey . . . 101
4.2.2 2011 Census Datasets . . . 104
4.2.3 The National Statistics Socio-economic Classification (NS-SeC) . . . 105
4.3 Selecting Appropriate Questions . . . 106
4.4 Estimation Approach One: Ecological Regression . . . 112
4.5 Estimation Approach Two: Data Mining Using Decision Tree Induction . . . 115

4.6 Results and Profiling 118
4.6.1 Data Profiling by Output Area Classification 126
4.6.2 Material Deprivation and Internet Behaviours 139
4.6.3 Urban Versus Rural Behaviours 143
4.6.4 Visualising Patterns of Engagement and Exclusion 148
4.7 Conclusions 154

5 From Univariate Measures to Multivariate Classification 157

5.1 Introduction 157
5.1.1 Background to Geodemographic Classification 158
5.1.2 Conceptual Framework for Classification of Internet Use and Engagement . . 159
5.2 Geodemographic Specification 161
5.3 Building a Classification of Internet Use and Engagement 166
5.3.1 Selecting Measures 166
5.3.2 Assessing Skew 171
5.3.3 Data Evaluation 172
5.3.4 Transformation and Normalisation 173
5.3.5 Variable Weighting 175
5.3.6 Standardisation 175
5.3.7 Initial Clustering 176
5.3.8 Assessing Variable Influence 178
5.3.9 Construction and Hierarchical Design 180
5.4 Cluster Mapping, Summarisation and Assessment 182
5.4.1 Supergroup 1: E-unengaged 182
5.4.2 Supergroup 2: E-professionals 183
5.4.3 Supergroup 3: Typical Trends 185
5.4.4 Supergroup 4: E-students 186
5.4.5 Supergroup 5: E-rural and fringe 187
5.4.6 Supergroup 1: Group 1a: Too Old to Engage 188
5.4.7 Supergroup 1: Group 1b: E-marginals: Not a Necessity 189
5.4.8 Supergroup 1: Group 1c: E-marginals: Opt Out 190
5.4.9 Supergroup 2: Group 2a: Next Generation Users 191
5.4.10 Supergroup 2: Group 2b: Totally Connected 193
5.4.11 Supergroup 3: Group 3a: Uncommitted and Casual Users 194
5.4.12 Supergroup 3: Group 3b: Young and Mobile 195

5.4.13 Supergroup 4: Group 2c: Students Online 196
5.4.14 Supergroup 5: Group 5a: E-fringe 197
5.4.15 Supergroup 5: Group 5b: Constrained by Infrastructure 199
5.4.16 Supergroup 5: Group 5c: Low Density but High Connectivity 200
5.5 Final IUC Hierarchy 202
5.6 Conclusions 203

6 Geodemographic Application **205**
6.1 Introduction and Background 206
6.2 A Framework to Understand and Measure e-resilience 208
6.3 Demand Factors and Internet Engagement Behaviour 210
6.4 Retail Centre Vulnerability and Supply 212
6.5 Reconciling Supply and Demand 213
6.6 Discussion and Conclusions 219

7 Conclusion **222**

Bibliography **227**

A Applied Geography Paper **247**

B Geoforum Paper **254**

C IUC User Guide **269**

D Example Geodemographic Input Correlations **291**

List of Figures

2.1 Internet Submarine Cables Connecting the United Kingdom . . . 20
2.2 Internet Users per 100 People: UK, USA and Worldwide: 1993 - 2014 . . . 22
2.3 Net Growth of Mobile and Fixed-line Broadband Subscribers in Austria . . . 30
2.4 MapMyRun Mobile App . . . 41
2.5 Sky+ Mobile App . . . 41
2.6 Visual Overview of the UK's Position Relative to the EU5: Coverage . . . 50
2.7 Visual Overview of the UK's Position Relative to the EU5: Take-up . . . 50
2.8 Internet Users per 100 People: Worldwide Divides: 1993 - 2014 . . . 52
2.9 Digital Divide: The Geography of Internet Access . . . 54

3.1 Exchange Locations (WGS84 Projection) . . . 67
3.2 Speed Test Locations (Postcode Centroid) 2010/11 Dataset . . . 70
3.3 Data Distribution: Distance to Exchange . . . 72
3.4 Mean Speed by Distance to Exchange . . . 73
3.5 Speed Tests Per Head of Population: English Districts . . . 75
3.6 Test Rate (Tests per Postcode) by Output Area Classification (OAC) . . . 77
3.7 Mean Download Speed (Kbps) by English District 2010/11 and 2012/13 . . . 81
3.8 Percent Increase in Mean Download Speed by English District 2010/11 to 2012/13 . 82
3.9 Change in Superfast Deployment 2010/11 to 2012/13 . . . 85
3.10 Percentage of FTTC Enabled Exchanges by English District (October 2013) . . . 86
3.11 Hourly Fluctuations in Mean Download Speed . . . 87
3.12 Hourly Fluctuations in Mean Download Speed by Urban/ Rural Index . . . 88
3.13 Installation of Rutland Telecom FTTC Service in Lyddington . . . 90
3.14 A Typical FTTP Installation Showing Coiled Fibre-Optic Cable Ready for Termination in a Customer Premises . . . 92
3.15 Spatial Distribution of Mobile Base Stations in England . . . 95
3.16 Average Distance (m) to Nearest Mobile Base Station, English LSOAs . . . 96

4.1 Example Full QUEST Output: OXIS Question QA1a1 . . . 119
4.2 Example QUEST Output: Nodes: Question QA1a1 . . . 120
4.3 R Code Snippet: OXIS Rate Calculations . . . 122
4.4 Rates of Engagement by OAC Group: Domain 1: Seeking information: QA1a1: Seeking information on local MP through Internet on a desktop/laptop/tablet computer . . . 127
4.5 Rates of Engagement by OAC Group: Domain 1: Seeking information: QA1a2: Seeking information on local MP through Internet on a smartphone . . . 128
4.6 Rates of Engagement by OAC Group: Domain 2: Perceptions: QA3d: Internet regarded with high importance for entertainment . . . 130
4.7 Rates of Engagement by OAC Group: Domain 3: Household Access: QH1a: Household has access at present . . . 131
4.8 Rates of Engagement by OAC Group: Domain 3: Household Access: QH3: Long-term connected households (10 years+) . . . 132
4.9 Rates of Engagement by OAC Group: Domain 4: Mobile Access: QH11: Mobile phone ownership . . . 133
4.10 Rates of Engagement by OAC Group: Domain 4: Mobile Access: QH12k: Use Internet on mobile phone . . . 134
4.11 Rates of Engagement by OAC Group: Domain 5: Access Patterns: QC1b: Access Internet while travelling . . . 135
4.12 Rates of Engagement by OAC Group: Domain 6: Commercial Applications: QC30b: buying online . . . 136
4.13 Rates of Engagement by OAC Group: Determining Rates of Personal Access: Current Internet Users . . . 138
4.14 Rates of Engagement by OAC Group: Determining Rates of Personal Access: Internet Non Users . . . 139
4.15 OXIS Questions QH1a (Estimated % of Households with Internet Access at Present) and QA9 (Estimated % of People Indicating they are Interested in the Internet) by IMD Decile . . . 140
4.16 OXIS Questions QC30b (Estimated % of People Frequently Buying Products Online) and QC30e (Estimated % of People Frequently Using Online Banking) by IMD Decile 141
4.17 OXIS Questions QH13_Current (Estimated % of People Using the Internet) and QH13_Non (Estimated % of People Who Have Never Used the Internet) by IMD Decile . . . 142
4.18 OXIS Estimates: QH13_Current: % of Individuals Using the Internet: Liverpool OAs 149

4.19 OXIS Estimates: QH13_Non: % of Individuals Who Have Never Used the Internet: Liverpool OAs . . . 150
4.20 OXIS Estimates: QH13_Non: % of Individuals Who Have Never Used the Internet: National LSOAs . . . 151
4.21 OXIS Estimates: QC30b: % of Individuals Frequently Buying Products Online: London LSOAs . . . 153
4.22 OXIS Estimates: QC30h: % of Individuals Frequently Ordering food or Groceries Online: London LSOAs . . . 154

5.1 Graphical Representation of an Iterative Allocation-reallocation Method on a Two-dimensional Dataset . . . 164
5.2 Top Down and Bottom Up Clustering Approaches . . . 165
5.3 Box-Cox Power Transformation on a Highly Skewed Variable (QC2a: Frequent Buying of Products Online) Q-Q plots Before and After. . . . 175
5.4 Clustergram of the PCA Weighted Mean of K-means Clusters . . . 181
5.5 Supergroup 1 National Distribution . . . 183
5.6 Supergroup 2 National Distribution . . . 184
5.7 Supergroup 3 National Distribution . . . 185
5.8 Supergroup 4 National Distribution . . . 186
5.9 Supergroup 5 National Distribution . . . 188
5.10 Group 1a: Too Old to Engage: Group Characteristics . . . 189
5.11 Group 1b: E-marginals: Not a Necessity: Group Characteristics . . . 190
5.12 Group 1c: E-marginals: Opt Out: Group Characteristics . . . 191
5.13 Group 2a: Next Generation Users: Group Characteristics . . . 192
5.14 Group 2b: Totally Connected: Group Characteristics . . . 193
5.15 Group 3a: Uncommitted and Casual Users: Group Characteristics . . . 194
5.16 Group 3b: Young and Mobile: Group Characteristics . . . 196
5.17 Group 2c: Students Online: Group Characteristics . . . 197
5.18 Group 5a: E-fringe: Group Characteristics . . . 198
5.19 Group 5b: Constrained by Infrastructure: Group Characteristics . . . 199
5.20 Group 5c: Low Density but High Connectivity: Group Characteristics . . . 201

6.1 A Conceptual Framework of e-resilience . . . 210
6.2 IUC Catchment Profile for Central Milton Keynes Retail Centre . . . 213
6.3 Propensity to Shop Online by IUC Groups (OXIS Question - QC30b 'Frequently buy products online') . . . 214

6.4 Highly Exposed Retail Centres in SE England 215
6.5 Process of Calculating e-resilience Scores 216
6.6 The e-resilience of Town Centres in England 221

List of Tables

2.1 UK Network Coverage 28
2.2 Fixed and Mobile Broadband Capacity: June 2012 29
2.3 Economic Advantages of a Connected Society 45
2.4 Standard Broadband Availability (percentage of households) 46
2.5 Superfast Broadband Availability (percentage of households) 47

3.1 Data Characteristics: Speed Test Databases 64
3.2 Data Summary: Speed Test Datasets 68
3.3 Number of Connections Recorded by Distance to the Nearest Exchange 2012/13 . . 74
3.4 Mean Download Speeds by OAC Sub-group 76
3.5 Mean Download Speeds by Urban/Rural Indicator 78
3.6 Mean Download Speeds by IMD Decile 80
3.7 Largest Ten Increases in Mean Download Speeds by English District 2010/11 to 2012/13 83
3.8 Smallest Ten Increases in Mean Download Speeds by English District 2010/11 to 2012/13 83

4.1 OXIS Target Populations 102
4.2 2011 Census Dataset Selection: Approach 1: Ecological Regression 104
4.3 2011 Census Dataset Selection: Approach 2: Decision Trees 104
4.4 NS-SeC Classes 106
4.5 Selected OXIS Questions: Domain 1: Seeking Information 107
4.6 Selected OXIS Questions: Domain 2: Perceptions 108
4.7 Selected OXIS Questions: Domain 3: Household Access 109
4.8 Selected OXIS Questions: Domain 4: Mobile Access 109
4.9 Selected OXIS Questions: Domain 5: Access Patterns 110
4.10 Selected OXIS Questions: Domain 6: Commercial Applications 111

4.11 Selected OXIS Questions: Determining Rates of Personal Access 111
4.12 Covariates for OXIS Estimations 113
4.13 OXIS Aggregation Structure 114
4.14 NS-SeC to Social Grade Group Mapping 118
4.15 Percentage of Respondents Who Use the Internet: 65+ Age Group by Social Grade: OXIS Question QA1a1 121
4.16 OXIS Estimates: Estimated National Means Vs. OXIS Derived National Means: Domain 1: Seeking Information 123
4.17 OXIS Estimates: Estimated National Means Vs. OXIS Derived National Means: Domain 2: Perceptions 123
4.18 OXIS Estimates: Estimated National Means Vs. OXIS Derived National Means: Domain 3: Household Access 124
4.19 OXIS Estimates: Estimated National Means Vs. OXIS Derived National Means: Domain 4: Mobile Access 124
4.20 OXIS Estimates: Estimated National Means Vs. OXIS Derived National Means: Domain 5: Access Patterns 125
4.21 Estimated National Means Vs. OXIS Derived National Means: Domain 6: Commercial Applications 125
4.22 OXIS Estimates: Estimated National Means Vs. OXIS Derived National Means: Determining Rates of Personal Access 125
4.23 2011 ONS Rural-Urban Classification for Small Area Geographies 143
4.24 OXIS Estimates: QH1a: % of Households with Internet Access at Present by ONS Rural-Urban Classification 144
4.25 OXIS Estimates: QA9: % of Individuals Indicating they are Interested in the Internet by ONS Rural-Urban Classification 145
4.26 OXIS Estimates: QC30b: % of Individuals Frequently Buying Products Online by ONS Rural-Urban Classification 146
4.27 OXIS Estimates: QC30e: % of Individuals Frequently Using Online Banking by ONS Rural-Urban Classification 146
4.28 OXIS Estimates: QH13_Current: % of Individuals Using the Internet by ONS Rural-Urban Classification 147
4.29 OXIS Estimates: QH13_Non: % of Individuals Who Have Never Used the Internet by ONS Rural-Urban Classification 148

5.1 Input Variables Selected for Evaluation 168

5.2 Input Variables Selected for Evaluation (2) 169
5.3 Input Variables Selected for Evaluation (3) 170
5.4 Summary of Input Measure Skewness 171
5.5 Resulting Effects of Transformation Methods on Skewed Variables 174
5.6 Cluster Characteristics: Transformed Data 177
5.7 Cluster Characteristics: Naturally Distributed Data 177
5.8 Variables that Increased Cluster Homogeneity 179
5.9 Variables that Decreased Cluster Homogeneity 179
5.10 Final Classification hierarchy 202

6.1 The 20 Most e-resilient Town Centres 217
6.2 The 20 Least e-resilient Town Centres 218

Chapter 1

Introduction

1.1 Background

Since inception in the mid 1990's, the Internet has become increasingly commonplace, engendering benefits for everyday life across multiple domains, including; information access, education, entertainment, retail and governance. Although such advantages are wide-reaching, access to, and engagement with the Internet is not evenly distributed across physical space or societal groups. Such disparities have been the focus of numerous academic and government studies within the contemporary period, attempting to isolate those physical and socio-economic factors that may impact Internet access and engagement. In the infancy of the Internet, these factors were largely binary, explaining the differences between the 'haves' and 'have-nots'. More recently, and as broader divides have narrowed, these binary concepts have diversified, with researchers adopting the term 'Digital Differentiation' to explore those differences in engagement between societal groups.

Governments in developed nations have increasingly recognised the importance of connected societies. In recent years, the UK government have funded numerous targeted schemes to close remaining divides pertaining to physical access through Broadband Delivery UK (BDUK) and the Rural Community Broadband Fund (RCBF). There have also been blanket approaches to ensure every household in the UK has access to a minimum viable broadband connection, in performance terms, through the Universal Service Commitment (USC). With regards to engagement, there have again been multiple initiatives to educate those who have limited or no engagement with the Internet, including; GoOn UK, UK Online Centres and most recently 'Doteveryone'.

Despite initiatives to close those remaining divides, the wider picture of access and engagement remains somewhat blurred, with research often limited to isolated geographic contexts, and rarely

considering patterns of engagement alongside those of localised infrastructure provision and performance. As such, there exists a requirement for access, performance and engagement patterns to be explored in parallel, as the intersection of these three factors will likely engender a complex geography of Internet use at the aggregate national and local scales.
This thesis aims to explore the plethora of factors that differentiate access to, and engagement with, the Internet, through the use of large quantitative data sources, including novel crowdsourced and open data where applicable to conduct geographically granular analyses. Furthermore, it is anticipated that a series of complex unidimentional outputs of these analyses may be combined to form a composite view of the current geography of Internet access and engagement, thus contributing to the current field of research and engendering benefits for applications where the Internet engagement characteristics of localised populations are of high importance. Such applications would likely include policy formation and delivery, profiling and further public and private sector research.

1.2 Thesis Structure

This thesis is comprised of five empirical chapters and two published papers. The empirical chapters broadly cover; a review of relevant literature, a study of broadband performance and access within the national context, a study of engagement patterns, the creation of a nested typology of Internet use and engagement and an example study demonstrating the application of the nested typology. The following sections provide a more detailed overview of each chapter.

Chapter 2 presents the relevant literature within this domain, beginning with the creation of the modern Internet and the World Wide Web. The underlying technical development are first considered, including ARPANET, one of the most influential developments in computer networking, discussed in terms of its purpose and how associated developments saw the emergence of domestic Internet access. Development of the World Wide Web is also covered, in conjunction with the evolution of consumer technology that enabled widespread adoption of the Internet across much of the mid 1990's. The growth of Internet users worldwide, from this period is also discussed, noting rapid uptake in developed nations, enabled through prioritised infrastructure rollout. The emergence of broadband Internet is discussed, as these technological advances underpin modern day Internet access. An Overview of technologies such as Digital Subscriber Line (DSL) and fibre is presented, noting the benefits which include speed and bandwidth increases. Discussion then moves on to consider the parallel development of mobile broadband networks within the contemporary period, as well as how these compare in terms of coverage and throughput to traditional fixed-line networks. In addition, the evolution of mobile handsets is considered, as these technological developments have enabled increasingly ubiquitous access to the Internet. Furthermore, the social and economic implications of the Internet are considered, and discussion covers the domains of; education, economic advantage, social and economic capital, entertainment, retail, lifestyle, privacy and unethical uses. Within each domain the key developments, benefits and implications are presented. Discussion then diversifies to consider the government perspective, including the importance of connected societies, landmark schemes to promote access and engagement and the UK's aggregate position with reference to other countries within the context of the European Union. The final sections of Chapter 2 consider the topic of the digital divide, beginning with the traditional divides between developed and developing nations, emergent as consumer access to the Internet was in its infancy. Discussion then moves on to consider the concept of digital differentiation and the factors that are reported to correspond with differentiated levels of Internet engagement in the modern context, thus framing the objectives of this thesis.

Chapter 3 is the first of four quantitative studies presented within this thesis and aims to examine spatial disparities in fixed-line broadband services using crowdsourced speed check data. The purpose is to present detailed analyses of the factors that influence aggregate levels of broadband access and performance, using spatially referenced performance data alongside open data sources. Analyses cover the topics of physical performance, first through distance metrics, before expanding this analysis to incorporate measures of socio-spatial structure, the latter aiming to identify the extent to which different areas and societal groups may be experiencing uneven levels of service. Analyses consider influencing factors such as; rurality, deprivation, localised population characteristics and contextual indictors pertaining to the local built environment, as well as temporal attributes, to assemble a comprehensive overview of performance and access disparities. Analysis for the national extent is also presented, giving an overview of Internet performance for two time periods between 2010 and 2013. The spatial disparities of performance increases are then quantified, revealing further uneven spatial distributions. In addition, analysis is conducted to examine the spatial disparities of next generation infrastructure, noting that while superfast connections are becoming more prominent, the rollout of infrastructure to support higher levels of service is constrained in some areas. A case study of private infrastructure overhaul in a rurally isolated conurbation is presented, demonstrating the demand for high speed broadband service in such areas, as well as evidencing alternative funding mechanisms. The final sections of the chapter analyse the provision of mobile base station infrastructure, which in addition to fixed-line infrastructure, exhibits significant spatial disparity between rural and urban areas.

Chapter 4 presents the second quantitative study and is focused on the analysis of Internet demand

and engagement. The chapter utilises data from the Oxford Internet Survey (OXIS) to create local estimates for the national extent for a number of measures of Internet access and engagement. Two methodological approaches are presented, with the first focusing on ecological regression and the second utilising decision tree induction. A bespoke methodology is presented and responses for 42 questions of the OXIS are estimated nationally at Output Area (OA) level, using census derived covariates in the modeling process. The estimates created are profiled by those measures that were reported to differentiate access and performance in Chapter 3, to examine whether similar disparities are apparent when use and engagement characteristics are considered. Analysis suggests that these characteristics are similarly differentiated across indicators of socio-spatial structure, rurality, deprivation and local context, invoking a complex geography, evident through each of the univariate measures created.

Chapter 5 considers how the complex sets of univariate outputs from previous empirical chapters can be combined using multivariate classification to produce a nested typology of Internet users at the small area level. The methodology presented involves the creation of a geodemographic classification, using input measures created through previous analyses. An overview of the typical stages of constructing a geodemographic classification are set out, before revisiting each within the context of the data available. The chapter segments construction of the classification by presenting consideration of; input measures, assessment of skew, exploratory data analysis, data transformation and normalisation, variable influence and hierarchical design. The resulting classification is presented through a series of cluster summarisations and assessments, which describe the prevailing characteristics of each of the clusters identified, forming a nested typology of Internet users referred to as the Internet User Classification (IUC).

Chapter 6 presents a use case of the IUC, examining the e-resilience of retail centres in England, assessing the extent to which retail centres have spatially differentiated vulnerability to the impacts of online consumption. The study is formed around the IUC, which is used to assess the extent to which local populations are engaged with online retailing. Retail centres are profiled by the IUC and demand factors are coupled with catchment models to create a composite index of exposure, engendering a remarkable geography of retail centres that are at high exposure to the effects of online retailing. Measures of exposure are then coupled with measures of supply vulnerability pertaining to the mix of stores within each retail catchment nationally, to create a composite e-resilience score, thus enabling the identification of those centres that are most, and least, e-resilient. The findings of this research contribute considerably to the understanding of the nature and impact that Internet user behavior is having on retail centres nationally, and as such demonstrates high policy relevance and contribution to the wider field of research.

Chapter 7 concludes the thesis by revisiting the main findings of each chapter and their contribution given the overarching objectives of the research. In addition, issues faced during the research and limitations of the work are discussed, noting how these issues were addressed and where limitations of this research may present opportunity for future work.

Chapter 2

Literature Review

2.1 Background

2.1.1 Introduction

"Globalisation, as defined by rich people like us, is a very nice thing... you are talking about the Internet, you are talking about cell phones, you are talking about computers. This doesn't affect two-thirds of the people of the world."

Jimmy Carter, former President of the United States of America

Since the 1980's, information and communications technologies (ICTs) have seen a series of radical advancements, including the development of affordable computing, email and mobile telephony. Arguably the most significant advancement has been the arrival of the Internet, bringing fundamental changes to the way in which populations can communicate, work and interact with governance structures. For those who have access, the Internet allows for near instant communication, giving disparate populations the ability to connect and interact with one another without the constraints of distance. Early commentary suggested this may facilitate the 'death of distance', a concept that argued the rise of ICTs would be associated with a fall in communication, information and transport costs, and as such, the impact of distance would become less important (Cairncross, 1997; Pitt et al., 1999). Despite the convergence of modern ICTs, namely computers, television and telephone enabling fast communication, the death of distance could be realised only by products or services that could be digitised. Digitisation of products such as newspapers, magazines and music, as examples, have had a profound impact on traditional retailing. This has led to the demise or significant diversification of many retailers in this space as access to the Internet has become more commonplace (Dholakia et al., 2010; Carlson et al., 2015; Verhoef et al., 2015). Consumer behaviours have also changed as a result, with an increasing likelihood for Internet users to purchase online given significant savings and added convenience (Calderwood and Freathy, 2014; Beck and Rygl, 2015). Despite reduced constraints on traditional methods of communication, benefits to proximity remain, and clustering of economic activity prevails despite advances in ICTs (Quah, 2000; Nathan and Rosso, 2015). Furthermore, the importance of face-to-face communication is a considerable barrier constraining the use of the Internet in some circumstances (Kaufmann et al., 2003).

Within the contemporary period, advances in telecommunications infrastructure and mobile computing have increasingly enabled the Internet to become omnipresent, which has largely underpinned the development of services such as social networking, streaming media and multi-channel retailing (Wei, 2008; Chang et al., 2015; Shankar et al., 2010). In addition, Internet enabled devices have become so commonplace in everyday life that the term 'the Internet of things' is frequently used to describe the vast networks of coded devices that have the ability to communicate with one another, with or without human interaction (Atzori et al., 2012; Gubbi et al., 2013).

2.1.2 From the Wires and Infrastructure to the Protocols and Content

'Internet' and 'World Wide Web' (WWW or 'Web') are often used interchangeably, but they are not synonymous terms. The Internet and the Web are separate but related developments. The Internet is a global network of networks that connects millions of computers together via physical infrastructure (See Figure 2.1). Assuming two computers are connected to the Internet, communication can be facilitated. The Web is best described as the prevailing paradigm for information dissemination through the Internet (Dikaiakos, 2004). In this respect, the Web is built as a communication architecture underpinned by the physical infrastructure of the Internet. The Web uses the Hypertext Transfer Protocol (HTTP), which is a standardised 'language' used to transmit data between computers. Browsers are used to access and display Web documents known as Web pages. Collections of hyperlinked Web pages are referred to as Web sites. The Web is just one way in which information can be exchanged through the Internet. Email uses the Internet, not the Web, to exchange data and does so using Simple Mail Transfer Protocol (SMTP), Post Office Protocol (POP) and Internet Message Access Protocol (IMAP), different protocols (or languages) to that of HTTP.

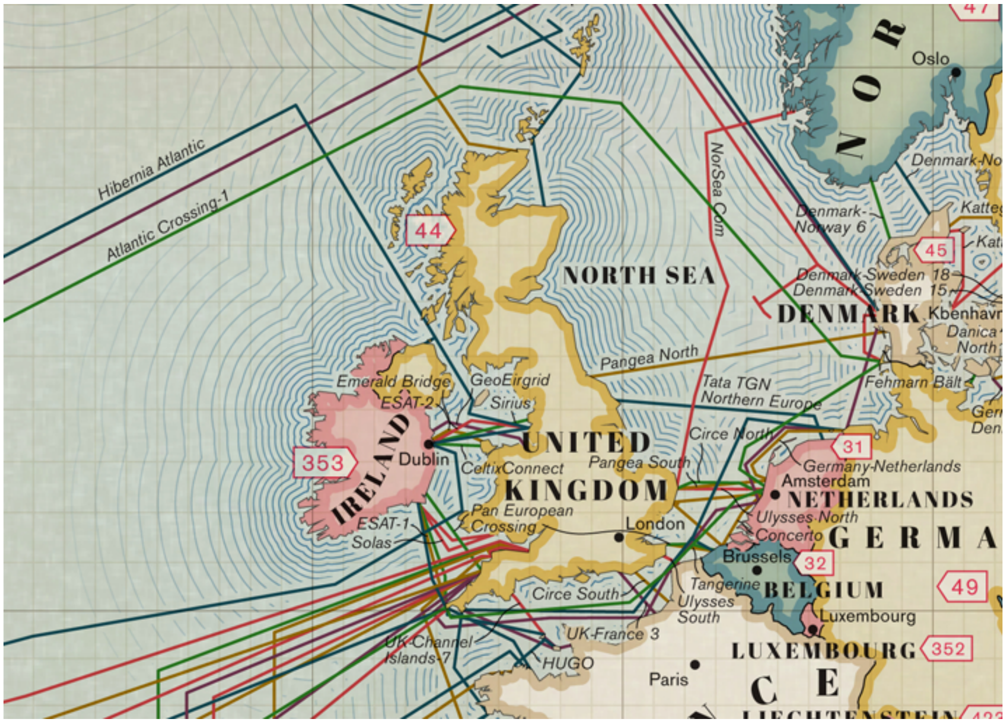


Figure 2.1: Internet Submarine Cables Connecting the United Kingdom.
Source: TeleGeography Cable Map 2013.
Accessed: http://submarine-cable-map-2013.telegeography.com/

2.1.3 Early Network Technology

One of the most influential projects in the development of the Internet was ARPANET, commissioned by the US Department of Defence's Advanced Research Projects Agency (ARPA). In 1969, ARPA funded a project linking together computers at four locations on the West Coast of the USA: the University of California at Santa Barbara, UCLA, Stanford, and the University of Utah (Glowniak, 1998). ARPA was interested in creating a set of telecommunications standards that could be used to interconnect the growing system of military computers worldwide. As the number of academic institutions connecting to ARPANET grew, some began to connect their local area network (LAN) of computers to ARPANET through a router; forming larger networks called wide area networks (WANs). This structure became known as the Internet. In August 1981, 12 years after the creation of ARPANET, there were 213 computers on the Internet. By October 1985, the number had increased to nearly 2,000 (Glowniak, 1998). Alongside the development of physical networks, Transmission Control Protocol and Internet Protocol (TCP/IP) technology was developed in the early 1980's. This was a standardised set of open communication protocols used for the Internet that were freely available. These protocols allowed any computer, regardless of the operating system or type of hardware to connect to the Internet. ARPANET was deactivated in 1990 after NSFNET, a project funded through the US National Science Foundation, took over much of the development of the United States Internet backbone network. Alongside the development of NSFNET, major telecommunications companies had become interested in supplying backbone service through their networks for commercial use. As NSFNET was prohibited from carrying commercial traffic, Advanced Network and Services (ANS), a non-profit organisation set up to manage NSFNET, was privatised and sold to America Online (AOL), which was founded in 1985 and later began to offer domestic Internet connections (Mowery and Simcoe, 2002). The modern Internet consists of a large number of communications networks owned by telecommunications companies, connected by backbone infrastructure.

The WWW was developed between 1991 and 1993 (Mowery and Simcoe, 2002) by a group of physicists at the CERN laboratory in Switzerland. The first development of the Web occurred in 1991 when Tim Berners-Lee and Robert Cailliau developed a document format called HyperText Markup Language (HTML) and the accompanying document retrieval protocol called HTTP. HTML incorporated multimedia capabilities that allowed pictures and graphics to be integrated within text. HTTP and HTML underpinned the development of the WWW, allowing documents to be linked through hypertext, or searched through text strings (Berners-Lee, 1992). Searching could produce a hypertext list of items, each linking to some document which matched the query, much like a modern search engine. Both the format and protocol, although having undergone many revisions (Hoy, 2011), have remained the primary method for publishing and retrieving content on the WWW to date. The launch of early graphical interface browsers such as Mosaic and Cello in 1993 and Microsoft's Internet Explorer in 1995 enabled widespread access to the WWW (Cheung et al., 1998; Norris, 2001), and there are now an estimated 3.6 billion users worldwide (Internet World Stats, 2016). Although the genesis of the Internet was initially fairly slow as a result of limited practical use to the general public, most post-industrial societies experienced sharply accelerating adoption rates in the mid 1990's and early 2000's (See Figure 2.2). This was aided by the introduction and promotion of commercial broadband connections and technological advances in computer hardware, aligned with falling prices (Kyriakidou et al., 2011).

Throughout the early years of the new millennium, the potential benefits at the national, individual and organisational levels contributed to something of a consensual view that broadband should be promoted by governments, as this was becoming increasingly important to the development of business, industry, shopping and trade in developed countries (Xavier, 2003). Similarly, there were perceived benefits regarding governance, as the Internet had the potential to facilitate communications between citizen and the state (Norris, 2001). As such, the rollout of high-speed infrastructure accelerated rapidly. It is estimated that investment in broadband infrastructure accounted for 9.53% of the United Kingdom's GDP growth in the period 2002-2007 (Koutroumpis, 2009).

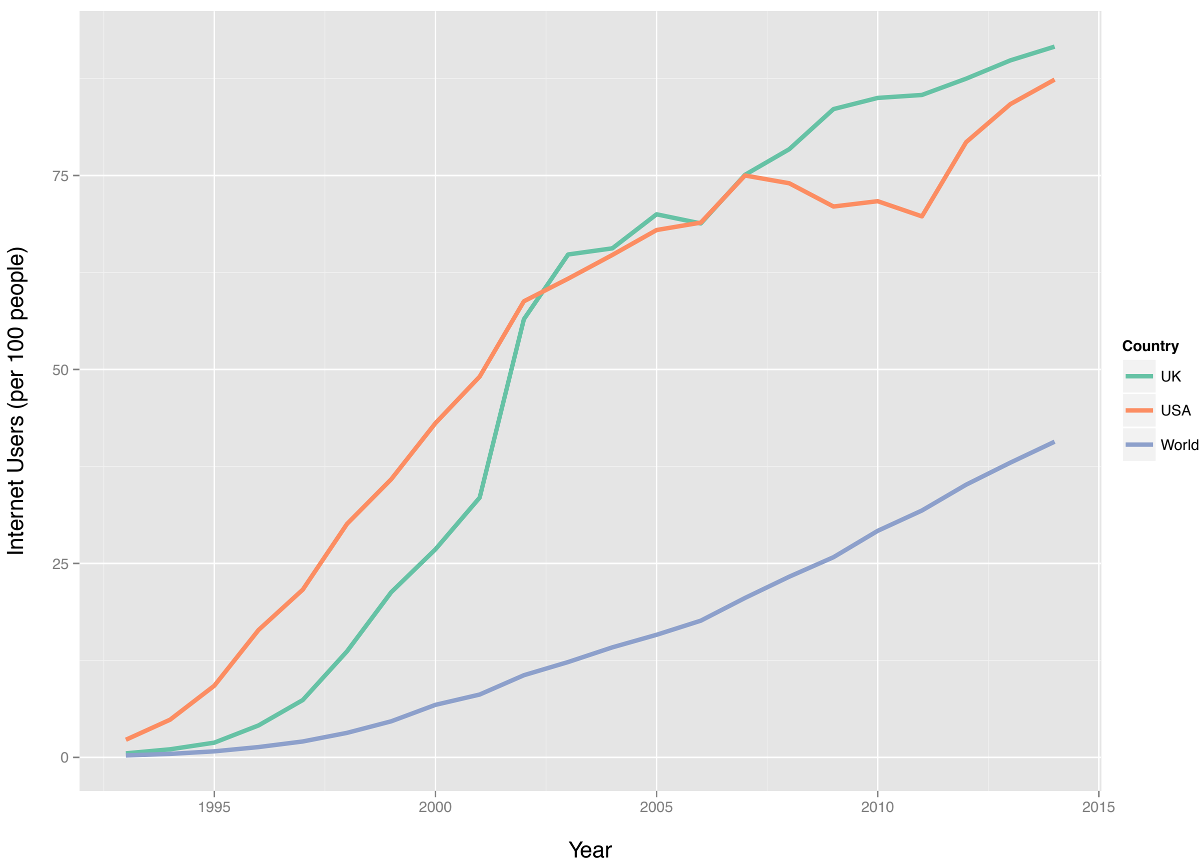


Figure 2.2: Internet Users per 100 People: UK, USA and Worldwide: 1993 - 2014.
Source: Based on World Bank World Development Indicators.
Accessed: http://databank.worldbank.org/data/home.aspx

Internet connections in the early years of access are referred to as 'narrowband'. This describes the bandwidth characteristics of the transmission type and the ability of this technology to carry only one signal type. Dial-up modems were commonly used to connect to the Internet through public switched telephone networks (PSTNs). Computers or routers would encode and decode Internet Protocol packets from the analogue frequencies sent over the telephone line. Dial-up modems had limited performance, offering theoretical maximum transfer speeds of 56Kbps. The first wave of next generation infrastructure, defined as 'broadband' as opposed to 'narrowband', was underpinned by a number of key developments in communications technology in the early 2000's.

2.1.4 The 'Broadband' Internet and Landmark Advances in Communications Technology

The term 'broadband' refers to the wider bandwidth characteristics of the transmission medium used to send and receive data over the Internet. Unlike the procurer of 'narrowband' technology such as dial up which only had the capacity to carry a single signal, broadband allows the transportation of multiple signals and traffic types simultaneously. In real terms, this means a broadband connection can transmit both voice telephone calls and data traffic in parallel. Broadband connections eliminated the requirement for customers to install a second telephone line should they require voice and data services simultaneously. The simultaneous transmission of multiple signals over PSTN was initially made available by the development of Integrated Services Digital Network (ISDN), a set of communication standards for the simultaneous transmission of voice, video, data and other network services over the traditional circuits of public switched telephone networks. Although this technology was available over narrowband and broadband connections, the development of DSL technology, which had the same application, but offered higher speeds, rendered ISDN technology near obsolete by the early 2000's.

The modern technology implemented for the simultaneous transmission of voice and data signals is referred to as 'Digital Subscriber Line' (DSL), and there are a number of variations of this technology that have been available on the commercial market in the UK since the rollout of broadband Internet services in late 2000. The term DSL is often used interchangeably with ADSL, which means 'Asymmetric Digital Subscriber Line'. The term 'asymmetric' describes the unevenness in the distribution of bandwidth frequencies for the transmission of upstream (upload) and downstream (download) signals. ADSL is the most commonly implemented technology by broadband providers as it allows customers to receive higher downstream speeds, which enables faster browsing. The frequency spectrum of an ADSL line is divided to allow for simultaneous transmission of voice data and upstream and downstream data traffic. A splitter commonly referred to as a 'microfilter' installed at each telephone outlet in a customer's premises separates the low frequency voice signals from the higher frequency data signals. Another derivative of DSL technology is 'Symmetric Digital Subscriber Line' (SDSL). SDSL connections have evenly distributed bandwidth frequencies for upstream and downstream traffic. This allows for faster upload speeds than ADSL and is often preferred by businesses that send large amounts of data. It is common for SDSL lines to be 'data only' and not carry analogue voice signals. Individual consumers rarely seek SDSL connections.

Internet structure in the UK is arranged through 'Internet Exchange Points' (IXPs) and 'Points of Presence' (POPs) (Tranos, 2013). POPs represent telephone exchanges where incumbent suppliers such as British Telecom (BT) and third party suppliers provide broadband service to individual premises. The service is routed via street cabinets. These cabinets connect all the local phone lines of small areas to the main distribution lines running from the POP. POPs are connected to higher-level nodes (IXPs), where Internet service providers exchange Internet traffic between their networks, and backbone networks, which run internationally forming the physical structure of the Internet.

Coaxial cable connections such as those provided by Virgin Media in the UK are an example of higher bandwidth networks that have had sustained success in the UK broadband market. Coaxial lines are capable of offering much higher download speeds using 'Very-high-bit-rate Digital Subscriber Line' (VDSL) Technology. VDSL technology is capable of delivering download speeds of up to 100 Mbps by utilising frequencies of up to 30MHz (ITU, 2002), much higher than the frequencies used with ADSL technology. VDSL is currently supplied over both copper wire (usually the link from street cabinet to premises after fibre optic terminations) and coaxial networks. Optical fibre connections to street cabinets known as 'Fibre to the Cabinet' (FTTC) have enabled the rollout of higher bandwidth services across much of the UK since early 2010 (Angelou and Economides, 2013). High-speed services of up to 50Mbps were available over coaxial connections from December 2008 with Virgin Media, although only in those areas that were already cabled. BT, who sell the 'Infinity' service, which offers a series of FTTC and FTTP (Fibre to the Premises) network upgrades, aimed to offer 'Superfast' fibre-optic broadband to around two-thirds of UK homes by the end of 2014 (BT PLC, 2011). Although the term 'Superfast' is recent, its definition is frequently amended to reflect new services and higher speeds available on the commercial broadband market. Ofcom, the regulatory authority for the telecommunications industry, changed its definition of 'Superfast' broadband from 24Mbps in 2011 to 30Mbps in 2012 to reflect market changes (HM Treasury, 2012) whereas BT classified 'Superfast' as speeds exceeding 25Mbps (BT Openreach, 2012).
One factor that has been influential in driving broadband uptake has been the competitive pricing of services (Biggs and Kelly, 2006; Flamm and Chaudhuri, 2007). The price of ADSL services fell throughout the early 2000's; and this pattern is now repeating with the prices of fibre-based services falling steadily. Since rollout in late 2008, 'Superfast' services have been available alongside standard voice and television packages as a single bundle from many providers. This 'bundling' of service offers the consumer access to voice, broadband and television at a reduced rate. Broadband

pricing strategies, especially the growing trend towards flat-rate pricing, aim to transform the revenue streams and expansion of communication services in the future (Biggs and Kelly, 2006).

2.1.5 Government Regulation and Local Loop Unbundling

Since the privatisation of British Telecom in 1984, Local loop Unbundling (LLU), which has been in effect since August 2000, has been one of the most influential government interventions in the telecommunications market. The European Commission argued that suppliers would not be able to duplicate incumbent operators' local infrastructure and unbundling was necessary to promote competition in the telecommunications industry (European Parliament, 2000). LLU represents the process where the incumbent operators (BT and Kingston Communications in the UK) make their local network (the copper cables that run from customers premises to the telephone exchange) available to other companies (Oftel, 2002). These changes were successful, and by 2012, 92% of Households in the UK had a telephone line supplied by an unbundled BT exchange (Yiu and Fink, 2012). In terms of access to broadband services, this meant that consumers were provided with a wider choice of broadband supplier. Competitors with access to the incumbent operator's infrastructure can offer telephone (voice) and data (broadband) service, or just data, with voice service being supplied by the incumbent operator. This increased competition in the marketplace has led to reduced prices and faster rollout of new access technologies across much of the UK (Biggs and Kelly, 2006; Yiu and Fink, 2012). The logistics of LLU vary based on the requirements of the proposed supplier and the limitations of the incumbent operator's exchange facilities. In most instances LLU is facilitated under one of four scenarios. The first requires Physical space within the incumbent operator's exchange, whereby a third party operator's equipment can be installed within a hostel (a room that is built to accommodate a number of operators' equipment in rack bay formations) or in a bespoke arrangement. This area is separate from the incumbent operator's operations. Cables running from customer premises are terminated at the incumbent operator's exchange and made available to third party operators through a Handover Distribution Frame (HDF). The second scenario, referred to as 'distant co-location', is where a third party operator houses its equipment away from the incumbent operator's exchange. A connection is facilitated to connect the incumbent operator's lines and the remote site. The remote site can be a building or a cabinet at the side of the road. The third scenario is 'line sharing', where data services are supplied by a third party operator while voice services remain with the incumbent supplier. Suppliers in the UK who offer broadband packages, but rely on BT for the provision of voice services adopt this type of LLU. The final scenario is 'sub-loop unbundling', which allows third party operators to connect to the local access network at a point between the incumbent operator's exchange and the end user. The connection points in the ‘sub-loop’ are street cabinets, known as PCPs. The closer proximity to the end user allows very high bandwidth (VDSL) services to be delivered. BT’s ‘Infinity’ network is growing rapidly, aiming to reach around two thirds of all UK homes by the end of 2014. Because the network is primarily FTTC-based, sub-loop unbundling will be the only viable option for operators who wish to supply fibre-based services to consumers, without rolling out their own infrastructure.

2.1.6 From Fixed to Mobile Broadband

‘Mobile broadband’ is a term used to describe wireless Internet access through a portable modem, mobile phone or other mobile devices such as tablet computers. Access to wireless Internet through these devices was enabled by the development of second generation (2G) mobile telecommunications, although mobile broadband access is now also available through 3G and 4G technologies. The term ‘broadband’ is used as it describes the capability to transmit voice as well as data signals simultaneously (the same as fixed-line broadband, only through different transmission methods). ‘Mobile Internet’ is a synonymous term which can be used to describe mobile broadband access, or in some cases, the mobile web (websites which have been optimised for viewing on mobile devices). Wireless Internet access can also include Wi-Fi, a technology that allows the wireless exchange of data using radio waves, over a computer network. This technology is generally used to wirelessly broadcast fixed-line broadband DSL connections, both in the home and in public ‘hotspots’. Wi-Fi is a separate development to mobile broadband. Many modern mobile phones and tablet computers do, however, have the ability to access the Internet through both mobile broadband networks, as well as traditional fixed-line networks through Wi-Fi functionality.

Mobile cellular technologies are categorised as First Generation (1G), Second Generation (2G), Third Generation (3G) or Fourth Generation (4G). 1G technologies are rarely discussed in current literature, but refer to the first generation of wireless telephone technology. This voice-only technology was based on analogue, as opposed to digital wireless standards and was introduced to the commercial market in the 1980’s (Ergen, 2009). Mobile telephony was migrated from analogue (1G) to digital (2G) technology as this allowed for compression of signals transferred between mobiles and network infrastructure. As such, this compression reduced load on networks and increased capacity, so that more calls could be handled simultaneously. The ability to carry data and voice simultaneously (mobile Internet) came with the rollout of General Packet Radio Service (GPRS), often referred to as ‘2.5G’. This technology was a development of 2G that occurred in 2000. GPRS was capable of providing data rates from 56Kbps to 115Kbps and could be used for services such as Wireless Application Protocol (WAP) access, essentially a Web browser for mobile phones, Multimedia Messaging Service (MMS) and email. 3G, which first emerged on the commercial market in the UK in 2003, included 3G Universal Mobile Telecommunications System (UMTS) and, later, High Speed Packet Access (HSPA) technologies. These technologies, sometimes referred to as 'wireless standards', provide the majority of 3G mobile broadband access in the UK. EDGE (Enhanced Data Rates for GSM Evolution) is a technology that is considered a sub-development of 2G as opposed to a core development of 3G technologies, and as such, has been largely superseded by the enhanced capabilities of 3G. 3G mobile broadband coverage from all network operators was available (via 3G UMTS or HSPA) to 77.3% of UK premises as of 2012, while 2G (voice only service) was available to 93.6% (see Table 2.1). Average download speeds associated with 3G mobile broadband connections are however, slower than those associated with fixed-line DSL connections due to reduced bandwidth. Independent tests have revealed that although network operators claim speeds of up to 7.2Mbps under optimal conditions, in reality, peak speeds are around 4.4Mbps with average download speeds much lower, at under 2Mbps (BroadbandGenie, 2013). Fourth generation technologies are sometimes referred to as 'Long Term Evolution' (LTE). LTE refers to a standard for wireless communication of high-speed data for mobile phones and data terminal. Generally this standard is marketed as '4G' or '4G LTE'. In terms of speed and usability, 4G LTE service is faster than 3G as it is provided over higher bandwidth, thus enabling faster access and the provision of more data rich services. 4G LTE networks currently offer download speeds of 8 – 12Mbps, similar to that of a fixed-line DSL connection and higher than average 3G download speeds (BroadbandGenie, 2013). Furthermore, 4G LTE technology has the potential to support speeds of up to 100Mbps downstream and 50Mbps upstream under optimal conditions (Ergen, 2009). Although these speeds are theoretical, and as such unlikely to be achieved in the real world, it shows there is scope for faster services in the future, as uptake of 4G LTE increases and network infrastructure is upgraded. As of late 2015 all major mobile operators in the UK offered 4G LTE service, which is available in most urban areas.

WLAN (Wireless Local Area Network) technology, usually implemented through Wi-Fi, is an example of how fixed DSL connections can be broadcast wirelessly to allow for mobile access. This type of mobile access, however, is limited, as the user must be within range of the broadcasting equipment (typically a wireless router). Domestic wireless routers are less powerful than broadcasting equipment used by mobile network operators, and therefore the 'mobility' of the connection is more constrained. Collectively, UK fixed and mobile operators provide over 16,000 Wi-Fi hotspots in locations such as cafes, restaurants, trains, pubs, airports and London Underground stations (Ofcom, 2012). Despite increased availability of public Wi-Fi hotspots, data gathered by Ofcom suggests consumers often prefer to make use of their mobile network data connections. This may be due to the time consuming process of locating an open Wi-Fi network and establishing a connection that may only be used for a short period of time.

2.1.7 Coverage and Throughput

Table 2.1 details the coverage of fixed public switched telephone network, Mobile 2G and 3G infrastructure as of 2012. Fixed-line broadband coverage is included for comparison. It is evident that although PSTN coverage is ubiquitous, broadband service at a standard of 2Mbps or higher is not possible over all of these lines. Similarly there are disparities between 2G and 3G service coverage, particularly by geographic area.

Table 2.1: UK Network Coverage. *Source: Adapted from Ofcom Infrastructure Report 2012 Update*

Fixed Telephony (PSTN)	
Coverage of fixed-line telephony	100% of premises
Fixed Broadband	
Coverage of broadband at 2Mbps or more	89.9% of existing connections
Coverage of superfast broadband	65% of premises
Mobile 2G (outdoor)	
Premises served by all operators	93.6% of premises
Premises not served by any operator	0.3% of premises
Geographic area coverage by all operators	58.8% of land area
Geographic area not served by any operator	12.8% of land area
Mobile 3G (outdoor)	
Premises served by all operators	77.3% of premises
Premises not served by any operator	0.9% of premises
Geographic area coverage by all operators	19.9% of land area
Geographic area not served by any operator	24.3% of land area

Table 2.2 details capacity statistics for fixed-line and mobile broadband connections for the month of June 2012. Comparing the average throughput statistics for fixed-line and mobile broadband connections it is apparent that fixed-line DSL connections carry the majority of data. Mobile throughput accounts for just 3.91% of all throughput (based on June 2012 statistics). Average throughput per connection also confirms that, in the UK, mobile data connections are utilised to a lesser extent than fixed-line. The average data throughput per mobile connection is 0.24GB com-

pared with 23GB for fixed-line connections. This contrast is to be expected as the reduced speed and reliability of mobile connections means they are unlikely to be used as a primary connection by most users, instead, mobile connections handle smaller data transfer tasks when away from a fixed-line connection. The high price and limited allowances of mobile data provided by network operators would typically render mobile broadband an inefficient primary connection method.

Table 2.2: Fixed and Mobile Broadband Capacity: June 2012.
Source: Adapted from Ofcom Infrastructure Report 2012 Update

Fixed Broadband	
Average fixed broadband modem sync speed	12.7Mbps
Total data throughput on residential fixed-lines	484,000,000 GB
Average data throughput per residential connection	23GB
Mobile	
Total number of mobile connections	81.7 million
Total mobile data throughput	19,700,000 GB
Average mobile data throughput per connection	0.24GB

Fixed-line connections presently have two major constraints: time and Place (Cheong and Park, 2005). These limiting factors are removed when broadband access becomes mobile, such as that enabled by the provision of 3G and 4G technologies where speeds of access are becoming comparable to fixed-line products. A key influencing factor in consumer decisions to adopt mobile broadband as a complementary or substitute service is pricing (Cardona et al., 2009). Research into the substitution of fixed-line to mobile telephony has reported an element of ‘true substitution’ in this market; as revenue from traditional fixed-line traffic has declined, it has tended to migrate increasingly into higher bandwidth fixed-line services as well as into mobiles (Albon, 2006). The extent to which mobile broadband is a complementary or substitute service to fixed-line broadband, has received little attention in the research literature. Previous studies such as that by the Organisation for Economic Co-operation and Development (OECD) report that there appears to be some substitution taking place in certain OECD markets, namely those with low prices. In other markets, growth is a function of the demand for mobility and data access (Graham and Dutton, 2014). Although in some markets, mobile broadband has grown as a substitute service, currently within the UK these services are typically inferior to fixed-line broadband in terms of data allowances, price and speed (OECD, 2009). These constraints on mobile broadband will, in most markets, inhibit the potential as a substitute service. Alongside developments in DSL services, which have been fast-paced, mobile broadband has developed less quickly. However, mobile broadband has seen success in some economically developed countries. Lower priced mobile broadband packages in Austria helped subscriptions grow significantly from 216,000 at the end of 2006 to 665,000 at the end of the 1st quarter of 2008 (OECD, 2009). Figure 2.3 shows the Austrian growth of mobile services alongside fixed-line broadband between Q1 2006 to Q1 2008.

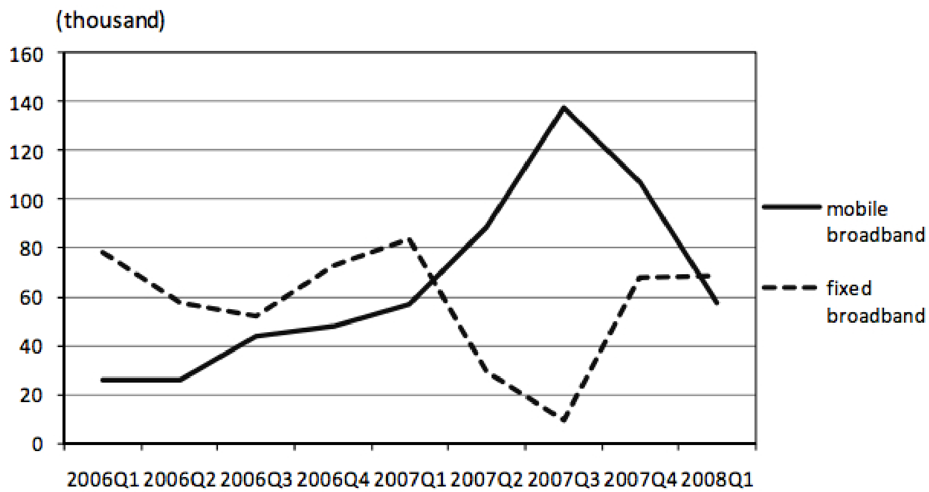


Figure 2.3: Net Growth of Mobile and Fixed-line Broadband Subscribers in Austria. *Source: Mobile Broadband: Pricing and Services (OECD 2009)*

Similar penetration of mobile broadband services has been reported in Sweden, which has one of the highest rates of household Internet access, with 92% household Internet penetration compared with 76% for the EU-27 countries (Eurostat, 2012). Research into this penetration has suggested that mobile broadband has higher price elasticity of demand than fixed-line services, and as such, is more sensitive to changes in pricing. It was also found that substitution of fixed broadband for mobile broadband was occurring in some areas. Adoption rates of mobile broadband were linked with factors such as age, income and type of housing (Srinuan et al., 2012). In the UK however, relatively high cost and low data allowances are viewed as limiting factors to replacing a fixed-line broadband connection with a mobile connection. The relatively small percentage of mobile broadband-based throughput and relatively low throughput per connection suggests that fixed-line connections will remain dominant for the foreseeable future.

2.1.8 Mobile Handset Evolution

Alongside developments in mobile telecommunications infrastructure, mobile handsets and other hardware used to access mobile broadband, such as tablet computers and modems have undergone a series of technical advances in recent years, enabling the delivery of increasingly rich data services over mobile networks. Since the emergence of rudimental mobile Internet browsers, incorporated into the operating systems of early handsets such as the Nokia 7110 (the first to feature a WAP browser), applications that utilise mobile data have increasingly featured in the development of mobile operating systems (West and Mace, 2010). Modern 'smartphones', named primarily because of their ability to be used both as a mobile telephone and as a handheld computer, allow users to install custom applications (Verkasalo et al., 2010). This key feature of modern smartphone operating systems has enabled applications, which were previously governed by the manufacturer, to be developed and distributed through 'application stores' by third party developers. Coupled with advances in mobile hardware, this has resulted in a wider range of mobile phone applications. The two most widely adopted operating systems, in terms of market presence are currently Apple's iOS (iPhone devices) and Google's Android OS (Android devices, which are produced by a number of different manufacturers) (Kenney and Pon, 2011; Müller et al., 2011). Both operating systems enable users to download custom applications through application stores. Geo-location functionality, through the integration of Global Positioning System (GPS) technology in mobile hardware, has also influenced the types of application that are produced for the smartphone market. Location-based services in the context of mobile communication have existed for some time (Kühn, 2004), but have been popularised more recently through practical applications for the mass market. These applications create individualised information to people that is contextually relevant to their present geographic location (Kitchin and Dodge, 2011). As well as delivering contextually rich services, location-based applications are also used for targeted mobile advertising (Li and Du, 2012). With data connections becoming an integral focus of mobile application development, substitution of traditional mobile telecommunications services such as SMS messaging has begun to emerge. According to analysis conducted in early 2013, chat applications that utilise cellular data connections and offer unlimited digital messaging have overtaken traditional SMS messaging. It is estimated that 19 billion chat messages are sent per day, compared to 17.6 billion texts (Informa, 2013). Such applications utilise Voice-over-IP (VoIP) and Internet Protocol (IP) technology to send messages through mobile data connections. These applications are often referred to as 'over the top' messaging, as two users communicate directly with each other through an Internet connection, bypassing the traditional wireless networks used to send standard SMS. Such changes highlight how consumers are beginning to shift to data-based services as opposed to those based on cellular technologies for some aspects of mobile communication.

2.2 Social and Economic Implications of the Internet

Access to the Internet has become increasingly omnipresent over the last decade (Peter and Valkenburg, 2006), resulting in a number of social and economic implications, related to; education, economy, social capital and lifestyle. The Internet brings benefits at a number of levels, both individual and organisational. As such, there are a number of reasons why governments invest or attempt to stimulate investment in Internet infrastructure (Lai and Brewer, 2006; Picot and Wernick, 2007; Xavier, 2003). As well as the reported benefits to access, there are a growing number of implications concerning online privacy (Fogel and Nehmad, 2009; Paine et al., 2007) , surveillance (Dinev et al., 2008; Wang and Hong, 2010) and unethical uses of the Web (Burden and Palmer, 2003; Hunton, 2009; Marshall and Tompsett, 2002). As such, this section reviews the academic literature regarding the relative importance of Internet infrastructure and Web access against the themes discussed above.

2.2.1 Advantages of the Internet

The development of the Internet has brought a number of benefits through the delivery of e-services, which lift traditional constraints of time and place. Such service delivery through the Internet has been widely researched. Internet banking provides a useful example of how the Internet has changed the business model within the financial services industry. By the end of 2003, more than half of the commercial banks in the United States offered some form of online banking services to their customers (Hernández-Murillo et al., 2010). Today, Internet banking is widely used and is accessible through both mobile and fixed access to the Internet. Similarly, the Internet has enabled fast access to health information and service delivery (Gray et al., 2005; McMullan, 2006; Rice, 2006) as well as educational resources (Butler, 1995; LeBaron et al., 1998). Internet shopping has also had an effect on the way in which we consume, by 2005 online sales accounted for 3.1% of the total retail expenditure in the UK (Weltevreden, 2007). Although unlike e-service delivery, Internet shopping still relies on proximity to some extent, as products must still be delivered, and does not fully realise the concept of 'death of distance'. Further developments underpinned by Internet adoption include social networking through websites such as Facebook and Twitter and voice-over-IP services such as Skype, all of which facilitate near-instant communication. The advantages of a connected world are abundant and appear limitless. However, as is frequently the case with sudden advancements in technology, some have been left behind. Primarily those who are unable or unwilling to engage. As such, it should be considered that the proposed benefits of the Internet discussed here are not universally applicable to populations, instead it is those who are actively engaged who realise the greatest benefits.

Given the advantages enabled by the Internet, researchers have long raised concerns about the impact that inequity of access has across a variety of geographic scales. Much early working concerned those disparities in connectivity between developed and developing countries, and what impact these differences would generate in terms of civic engagement. As such, the term 'Digital Divide' was introduced in the late 1990s in reference to those disparities in access to information that were observed as a result of variable Internet provision (Norris, 2001). Although a valuable concept at the time when the Internet was first developing, more recent discussion has diversified this concept, adopting instead the term 'Digital Differentiation' (Hargittai, 2002; Longley et al., 2008). As broader divides have narrowed (Kyriakidou et al., 2011; Peter and Valkenburg, 2006), the digital differentiation approach aims to depict differences between societal groups in terms of use and engagement patterns, that are evident within those countries with more developed Internet Infrastructure (Longley, 2003).

2.2.2 Enhancing Education

Since inception, the Web has been utilised as a platform for learning. An early study in 2001 by the Pew Internet and American Life Project reported that 94% of youth aged 12-17 and who had access to the Internet, stated that they used it for school research, and 78% believed the Internet helped them with schoolwork (Pew Internet, 2001). The open exchange of information on the Web has had a profound effect on the scale and brevity of resources for education and research (Allison et al., 2012). Today, almost all academic journals and books are available in digital format, as such, enabling the widespread dissemination of academic material. However, in terms of access, disparities exist, as much of this material is not open source (unlike Wikipedia for example) or freely accessible. In this respect, the web may enable access to academic materials, but it does not necessarily enable knowledge exchange, instead it has two separate but related uses in terms of education; as a platform for dissemination and as a platform for interaction and collaboration. The latter of these two uses represents a key area in which the web has enhanced education since adoption. Allison et al. (2012) present two distinct styles of Web-based learning; formal and informal. Formal Web learning is organised around educational standards, institutions, courses, and degree programmes, based mainly in universities, schools and colleges. Formal Web learning uses the Web as a platform for the deployment and delivery of formal education. This can be through 'blended learning' at academic institutions, using online resources alongside traditional delivery methods such as lectures and textbooks. Alternatively, formal Web learning can be facilitated through distance (or remote) degrees or short courses offered by academic institutions. Informal Web learning refers to learning as a side effect of other activities. This can be through online social networking and interaction with other Web sources. Informal Web learning can also represent Web sources such as Wikipedia, which are open source and are not subject to the same stringent review processes that exist in the traditional academic environment. In terms of the Web as a platform for interaction and collaboration, Virtual Learning Environments (VLEs) represent a key development. VLEs are defined as learning management software systems that synthesise the functionality of computer-mediated communications software and online methods of delivering course materials (Britain and Liber, 1999). VLEs include functions such as email, group chat rooms, computer aided assessment and file sharing. Content created by academic staff can be uploaded and viewed by students remotely; in this respect VLEs also enable collaborative working amongst groups of students without the need for face-to-face meetings. Blackboard is an example of a VLE developed privately that has been widely adopted across academic institutions (Liaw, 2008). Similar VLEs, including some which are open source, are used to deliver cost-free distance courses to large numbers of users. These 'mass market' courses are referred to as 'Massively Open Online Courses' (MOOCs) and frequently run in conjunction with major academic institutions (Coursera, 2013).

2.2.3 Broadband and Economic Advantage

Since the emergence of modern ICTs, the 'death of distance' is a concept widely discussed, but not fully realised. Although the Internet and the Web have changed the way in which we do business, the 'death of distance' was realised primarily by those services or goods which were able to be digitised, primarily communications, media and some tertiary service industries. Since the work of Cairncross (1997), there have been numerous studies disputing the predictions that distance would become less important through radical improvements in the cost and efficacy of long distance communications and movement of goods. Research has instead suggested that distance related trade costs have been increasing over time (Berthelon and Freund, 2008) and that variations in communication costs can have significant influences on trade patterns (Fink et al., 2005). Despite the limited application of the argument for the 'death of distance', convergence of modern ICTs such as fixed-line and mobile telephony and broadband Internet have enabled new forms of trade and consumption such as online shopping through the Web. These developments realise the 'death of distance' to some extent for the consumers who do not have to travel for goods and services, but do little to overcome the time and costs associated with traditional methods of transportation. Regardless of these limitations, the Internet has long been reported to stimulate economic growth (Freund and Weinhold, 2004), through the consumption of goods and services online, and facilitation of a global marketplace. In

2012 the Internet economy accounted for over 8% of UK GDP, a higher share than any other country in the G20. This figure was forecast to rise to over 12% by 2016, with the Internet accounting for around a quarter of our economic growth (Yiu and Fink, 2012). As well as economic advantages through online trade, the physical installation and upgrading of broadband infrastructure has been shown to facilitate widespread job creation. A study by the Information Technology and Innovation Foundation estimated that a $10 billion investment in broadband would produce as many as 498,000 new jobs. These include the construction workers and telecommunications technicians who must dig up streets, lay down fibre-optic lines and install wireless towers, as well as the engineers and factory workers at companies that make the fibre, electronics and computer equipment needed to build the networks (Tessler, 2009). With fast-paced rollout of fibre-optic infrastructure upgrades across much of the UK in progress (DCMS, 2010), similar economic benefits are likely to be seen. Recent studies have also reported that jobs dependant on broadband availability, such as computer systems analysts, database administrators and media and communications workers are projected to increase 25% by 2018 (Federal Communications Commission, 2011). In 2008 it was reported that 62% of American workers could be considered 'Networked Workers' who use the Internet or email at their workplace (Madden and Jones, 2008), thus highlighting how importance of broadband in the working environment has grown in the contemporary period. Among those who rely on broadband for work, it is estimated that 14% telecommute or work from home at least one day a week (NTRS, 2009), a practice that has only become possible due to the convergence of modern ICTs. Growing trends in the use of ICTs in the workplace can have adverse effects on those who are not engaged with such technologies. Businesses have come to expect candidates to be conversant with the basic functions involved in the use of personal computers (Longley et al., 2008), as such, general ICT competency is no longer seen as a specific skill. For those who possess no ICT skills, the job market will be limited. In this respect, and particularly so within the UK, government seek to provide generic ICT skills training to those who are unengaged, through the delivery of schemes such as GoOn UK, a digital inclusion scheme whereby those who possess adequate digital skills can volunteer to teach those who do not.

2.2.4 Social and Economic Capital

Social capital is defined as the expected collaborative or economic benefits derived from the preferential treatment and cooperation between individuals and groups (Bourdieu and Wacquant, 1992; Portes, 2014). A social network is described as a social structure made up of a set of social actors such as individuals or organisations and a complex set of ties between these actors, which can be analysed to identify local and global patterns. Economic capital has been discussed at the macro level in the previous section but will be expanded in this discussion to incorporate micro level influences. Economic capital within this context refers to the ability of a consumer who has access to the Internet to benefit financially, either through access to more competitively priced goods or services, or through forms of capital exchange, for example using existing social capital to find new employment. Cultural capital refers to non-financial social assets such as education, intellect and style of speech or dress, which can enhance social mobility beyond economic means alone (Ostrower, 1998; Dumais and Ward, 2010). Social capital and the formation of social networks are topics that have been widely researched long before the inception of the Internet. One of the first to introduce the theme of social 'decapitalisation' was Putnam (1995) who argued that despite the advantages of social capital, namely: better schools, faster economic development, lower crime, and more effective government, there had been a shrinking of America's civic society over recent decades. Possible explanations for this trend were attributed to more women in the workplace, increased mobility of families and changing demographics (Putnam, 1995). Putnam's research was famously focused on the demise of bowling leagues in America and how these types of 'physical' social capital were fading. Conversely Lin (2001) argues that it is not simply a case of whom you know, but also what you know, and that access to ICTs have led to a rise in the formation of social capital. The widespread availability of the Internet and convergence of modern ICTs (much of which occurred after Putnam's research) have changed the way in which individuals can communicate. The Web, for example, is a platform that can support communication between individuals or groups near-instantaneously through the Internet. These developments have been argued to lead to the emergence of 'virtual social capital', a new form of social capital facilitated through interaction online (Shah et al., 2001; Steinfield et al., 2009). Virtual social capital differs from traditional social capital in a number of ways, including its inability to be traced to 'real world' geographies (or material places) that are attributed to the formation of traditional social capital, such as the neighbourhood. As such, more abstract definitions of space have emerged to include macro level typologies such as 'Cyberspace' and 'Cyberplace' falling within the realm of 'virtual geographies' (Batty, 1997).

Online social capital generated through the Web is located in an abstract environment, which is not locatable in physical space and is therefore best defined as 'cyberspace'. The formation of social capital through cyberspaces has attracted much attention since the emergence of the Internet (Nie, 2001; Shah et al., 2001; Steinfield et al., 2009). With the emergence of a range of Web sites such as LinkedIn, orientated towards a work-related context, and Facebook, orientated towards friendship networks, online social networks allow individuals to present themselves, articulate these networks and establish or maintain connections with others (Ellison et al., 2007). The formation and structure of the Internet, and the ability to communicate from anywhere in the world, means that real world and online networks do not necessarily correlate. Some research has, however, indicated that exposure to social networking online enriches real-world relationships (Katz et al., 2001; Ellison et al., 2007). Conversely, others have suggested disparities exist, with those who have good offline relationships benefitting the most from online relationships (Kraut and Kiesler, 2003). Steinfield et al. (2009) who builds on the original work of Putnam (1995) by investigating the use of social network sites in organisations hypothesises that the use of social network sites contributes to social capital within organisations in that users are able to maintain larger networks of heterogeneous contacts. In this respect, the use of the Internet and the Web to maintain global social networks is realised. Despite a large body of research that suggests the Internet and the Web enable social capital, the extent to which this has affected the adoption of these technologies remains unclear. Although a widely cited benefit for those who have access (the UK government see Internet access as a facilitator of socio-economic inclusion (DCMS, 2009)), it may be argued that consumers adopt the Internet for more pragmatic reasons, such as greater access to services, competitive pricing and as a means of communication. Economic Capital at an individual level, as a result of access to the Internet can be enhanced in a number of ways. In a contemporary and quantifiable context, access to the Internet has been shown to facilitate savings at an individual level amounting to as much as £560 per person per year, compared with those who are not engaged online. This represents a cumulative total of around £1 billion per year in the UK alone (Lane-Fox, 2009). These savings are largely attributed to wider consumer choice and access to competitive pricing. In a theoretical context, economic capital can be enhanced through the exchange of other forms of capital accrued through online participation. For example, social capital and the maintaining of large social networks (both professional and informal) may facilitate opportunity for financial gain, either through job offers or collaborative working. Research has suggested that these 'weak' ties developed online possess value in their ability to provide new information and access to disparate networks (Wellman et al., 2001). The influence of the Internet on cultural capital is less easily placed. In one sense, those who may have previously been constrained by their cultural background have the opportunity to research and engage in activities online which may have been previously unavailable to them. In another sense, cultural capital, which is traditionally formed through social structures, has no relevance in cyberspace, which is geographically neutral and highly anonymous. In this sense, a lack of traditional cultural capital may not necessarily inhibit the ability to accrue virtual social capital via the Internet.

2.2.5 Entertainment, Lifestyle and Retail

Widespread adoption of the Internet and the Web, alongside the development of broadband technologies such as DSL, has contributed to changing trends in the way media is distributed and consumed over recent years. Media such as film and music, in particular, which have traditionally been distributed in a range of 'hard formats' such as tape, CD and DVD, have been increasingly distributed in digital format since the inception of the Internet. Much of the early digital distribution of music over the Internet was facilitated through file sharing sites such as Napster. Founded as a pioneering peer-to-peer sharing Internet service in 1999, the site allowed users to share audio files, typically encoded in MP3 format, with one another. The music available at Napster was varied, from the most popular to the hard to find. In some instances, music that had not been released by record companies was easy to find at Napster (Giesler, 2006). Because the network allowed consumers to download music they would otherwise have had to buy, the service ran into legal difficulties over copyright infringement and was shut down in July 2001. At its peak, the service is estimated to have had 26.4 million users worldwide (Jupiter Media Metrix, 2001). Although short-lived, the service highlighted the potential for the Internet as a medium to distribute digital music. In April 2003 the Apple iTunes Music Store opened as a software-based online digital media store operated by Apple Inc. Currently the store offers over 26 million songs, videos, e-books and apps for sale (Apple Inc., 2012) and by February 2013 the store had sold 25 billion songs worldwide (Apple Inc., 2013). The iTunes store is one example of the multiple online media stores that have been created to legally distribute music, video and other digital entertainment. In more recent years a number of music streaming services such as Spotify have emerged as an alternative to purchasing. For a fixed monthly fee, members of the service are free to stream as much music as they like. However, due to the nature of streaming media, a good quality broadband connection is required to use the service without disruption. Both streaming services and music stores are available on mobile platforms which allow users to purchase music through mobile devices such as smartphones and tablet computers, thus highlighting how consumption of these media types has moved away from traditional store purchases to online downloads and streaming through a range of devices.

The growing trend for consumers to purchase digital media online has had a profound impact on traditional retailing (Kim and Park, 2005; Neslin et al., 2006). Stores that were previously reliant on physical media such as book shops and music stores have struggled in recent years given widespread purchasing through mediums such as the iTunes store or streaming through services like Spotify. The extent of these changing consumer behaviours has not been limited to those traditional retailers dealing in digital media (Demirkan and Spohrer, 2014), many others categories have suffered given the financial savings and added convenience that shopping online can offer. Such is the extent of this decline that the theme of resilience is commonplace in contemporary retail research, often referring to the extent to which retailers have adapted, or are resilient to, the impact of online shopping (Wrigley and Dolega, 2011; Singleton and Dolega, 2015). Most large retails now recognise that a multi-channel strategy is an important factor in remaining competitive with online-only

retailers and retaining a base of customers who increasingly demand cheaper prices, convenience and flexible service (Pantano and Viassone, 2015). Responses have been varied, but typically most large retailers will have invested in click and collect services to allow customers to order online and collect at their convenience, or enhanced home delivery services. These will often be supplemented with mobile applications, allowing customers to order on the move, and enhanced online stores to establish a strong Web presence. Such an approach is often referred to as multi-channel or 'omni-channel' retailing, whereby a customer can connect with a retailer through multiple touch points for increased convenience, but has access to the same products, services and support as they would in a physical store (Neslin et al., 2006).
In addition to retailer diversification, research in this field has considered the decline of traditional retailers given the proliferation of online shopping, with the most recent literature noting that the extent to which localised populations are engaged with the Internet could be considered as an influential factor in the attractiveness and success of traditional retail centres (Dolega et al., 2016), thus highlighting a requirement for more quantitative multivariate measures of online engagement at varying geographic scales.

Alongside changes in consumer behaviour, connectivity has evolved rapidly, everyday items such as domestic appliances, handheld tools, sporting equipment, medical devices and children's toys have become embedded with increasingly advanced software (Kitchin and Dodge, 2011). In some instances, the software embedded in these items makes them uniquely addressable, integrating them into the emerging 'Internet of things'. The Internet of things refers to uniquely identifiable objects and their virtual representations in an Internet-like structure. In much the same way that a website can be accessed by a unique domain name, it is envisaged that this infrastructure will facilitate a similar process for any uniquely tagged object (Schoenberger, 2002). Developments in consumer technology have been fast paced in recent decades and many everyday appliances are now considered 'coded objects' (Kitchin and Dodge, 2011). Coded objects, as identified by Kitchin and Dodge, range in their technical abilities, and are divided into six classes:

- Peripherally Coded Objects – Objects in which software has been embedded, but where the software is incidental to the primary function of the object. For example, a digital timer in a domestic oven

- Hard Codejects – Objects embedded with firmware that is essential for their functioning. Functionality is predetermined and fixed. A USB memory stick is an example

- Unitary Codejects – Objects that are programmable to some degree and therefore exhibit some degree of interactivity. Some aspects of functionality can be controlled, for example, as and when operation is required. Examples include washing machines and DVD players

- Logjects – Objects that record their status and usage, and can retain these logs when deactivated and utilise them when reactivated
- Permeable Logjects – Objects that are relatively self-contained such as MP3 players and Presonal Digital Assistants (PDAs), which have the potential to be connected to wider networks
- Networked Logjects – Objects that do not function without continuous access to other technologies and networks because they require constant two-way data exchanges. Examples include satellite and cable television control boxes and mobile telephones

Increasingly, consumer devices such as televisions are incorporating network technology in order to deliver on-demand programming over the Internet. The growth of on-demand video streaming services such as Netflix, and BBC's iPlayer has been fast-paced in recent years and has been enabled by rollout of high-speed broadband service worldwide. In 2012 it was reported that a third of all downstream Internet traffic in North America during peak hours (defined as 9pm to 12am in this instance) was attributed to the Netflix service (CNET News, 2012). As well as the growth of 'smart' appliances such as televisions and set top boxes which integrate networking capabilities, other devices such as mobile phones are increasingly utilising applications that share data through Internet connections.

Examples include 'MapMyRun' [1], an application that utilise the GPS functionality of smartphones to track routes of cyclists and runners, providing detailed split times, elevation and speed measurements. Stats from routes can be shared, via an Internet connection, to social networking sites such as Facebook and Twitter. Similarly, users of the application can view their friends' routes and times. Other examples include applications that 'link' devices remotely, for example the Sky Plus application [2], which allows users to remotely program their set top box to record programming through an Internet connection. The integration of network technologies into many modern 'coded objects' is a reflection of how embedded the Internet has become in modern society. Greenfield (2006) introduced the concept of 'everyware', the idea that computing will become ubiquitous, integrating into every object of significance in our everyday lives. To some extent this has been realised through 'coded objects', which range from the chips in credit cards through to smartphones, which are becoming increasingly more powerful. It may be argued that the Internet has promoted the development of both 'coded objects' and 'everyware' as it is the system that is instrumental in facilitating communication between many modern devices. However, It is important to note that

[1] Source: `http://www.mapmyrun.com/app/`

[2] Source: `https://itunes.apple.com/gb/app/sky+/id301250225?mt=8&ls=1`

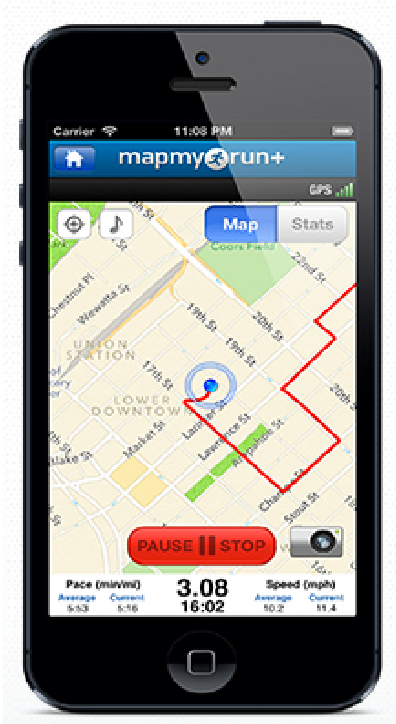


Figure 2.4: MapMyRun Mobile App

Figure 2.5: Sky+ Mobile App

ownership and use of such devices is influenced by many external factors including experience with technology, affluence and infrastructure constraints. As such, differentiated patterns of engagement are likely to exist across demographic groups and physical space.

2.2.6 Privacy, Surveillance and Unethical uses of the Internet

Counter to the benefits of access, there are potential negative implications concerning online privacy (Paine et al., 2007; Fogel and Nehmad, 2009), surveillance (Dinev et al., 2008; Wang and Hong, 2010) and unethical uses of the Web (Marshall and Tompsett, 2002; Burden and Palmer, 2003; Hunton, 2009).

Online privacy is a topic that attracts media and public attention, and is an issue that is closely linked with online surveillance. Academic studies on privacy concerns, such as that by Paine et al. (2007), have indicated that Internet users are concerned about a wide range of privacy issues including the sharing of email and contact information, being tracked when using websites and payment card fraud. Age has been identified as the best predictor of whether people are concerned about their privacy whilst they are using the Internet, with younger age groups less concerned than those who are older. Beyond perceptions of privacy, it was reported that 73% of the sample interviewed take action to protect their privacy online (Paine et al., 2007). Earlier studies such as those by Harris and Associates and Westin (1998) and Harris (2004) had also confirmed similar findings, suggesting that the vast majority of Internet users had concerns over privacy. Of the respondents concerns about privacy online, as reported by Paine et al. (2007), the top two are viruses and spam. Both represent forms of malicious attacks on Internet users recognised as cybercrime (Marshall and Tompsett, 2002; Hunton, 2009). The topic of cybercrime is closely related to online privacy, and as such, encourages many Internet users to adopt measures such as firewalls and anti-virus software packages to protect their personal information (Hunton, 2009). Beyond the less serious, but widely implemented crimes of spamming and virus attacks, cybercrime facilitated through the Internet has grown to include a number of serious offences including fraud, identity theft and theft of intellectual property. Bryant and Bryant (2008) offer some explanation as to the growth of cybercrime, suggesting that using digital networks such as the Internet provides cyber criminals with a simplified, cost effective and repeatable means to conduct rapid and large scale attacks against a global cyber community. Fletcher (2007) expands that the lack of face-to-face contact creates a perception of reduced risk and increased anonymity for the cyber criminal. Marshall and Tompsett (2002) provide a comprehensive classification of computer and Internet crime that details not only the technology used by the criminal but also the nature of the crime being committed. The six broad categories of cybercrime are proposed as:

- Computer assisted – Those crimes which are easily committed using conventional means, but are equally possible using computing equipment. In this respect, they may bear some similarity to legitimate activities which are also made easier by the use of computing equipment. Examples include blackmail

- Computer enabled – Crimes which can be committed using non-computing techniques, but are difficult to perpetrate successfully without technological support. Examples include music piracy

- Computer only – Crimes which cannot be committed without computers being directly involved. Examples include virus propagation

- Internet assisted – Crimes which can be committed easily by conventional means, but are also possible using internet technologies. Examples include stalking

- Internet enabled – Crimes which are difficult, but possible, to commit using conventional means, but are made easy by the application of internet technologies. Examples include money laundering

- Internet only – Crimes which require the existence of the Internet for their perpetration; for example, crimes against virtual entities. These are Internet crimes perpetrated against entities which have no clear existence in the normal corporate sense, and exist only as collections of data held on Internet servers or as streams of data passing between other Internet entities. Examples include denial of service attacks

Cyber criminals are increasingly committing these crimes, in particular, those that rely on the use of the Internet. Professional and organized cyber crime targeting financial institutions and e-commerce infrastructure grew 35 times over during the period 2001 – 2005 (Grow, 2005). A more recent report, conducted in partnership with the Office of Cyber Security and the Cabinet Office estimated the cost of cyber crime to the UK to be £27 billion per annum, with an economic cost to citizens of £3.1 billion. The direct cost to citizens includes an estimated £1.7 billion for identity theft, £1.4 billion for online scams and £30 million for 'scamware' and fake anti-virus software. (Detica, 2011). In addition to increasing concerns over privacy and cyber crime, surveillance is another related topic that has attracted much attention over recent years. 'Coded objects' and instant communication facilitated through Internet infrastructure have made it increasingly possible to locate and monitor people's activity. Surveillance of Internet users by governments, in particular, has become increasingly commonplace since terrorist attacks such as 9/11 (Dinev et al., 2008). After counter-terrorism and counter-intelligence, cyber crime is ranked as the third highest priority of the U.S. Federal Bureau of Investigation (FBI) (Grow, 2005), much of the investigation of these offences is heavily reliant on Internet surveillance and intelligence gathering at both vendor and Internet Service Provider (ISP) level (Grow, 2005), hence growing concerns among Internet users about how their online activity may be monitored. Disclosures of government Internet surveillance programs such as the U.S. National Security Agency (NSA) 'PRISM' program have also fuelled concerns about online privacy and the extent of government surveillance. Recent leaks by Edward Snowden, a former NSA contractor, regarding 'PRISM', a clandestine mass electronic surveillance and data mining program have suggested that the extent of mass data collection is far greater than the public are aware of and incorporates activities that may be dangerous or criminal in their nature (Guardian, 2013). Much like the use of Internet enabled devices, issues of privacy and cyber crime are likely to create differentiation amongst Internet users with varying levels of knowledge and concern. Exploring patterns of such differentiation has attracted little academic research within this domain and highlights a current research gap to consolidate the many factors that influence engagement.

2.3 Broadband, Governance and Civic Engagement

Access to broadband networks and to high-speed Internet are considered a necessary precondition for economic growth and competitiveness (Picot and Wernick, 2007). As an essential infrastructure investment, the deployment of broadband and its importance to governments worldwide has been widely researched in recent years (Picot and Wernick, 2007; Trkman and Turk, 2009; Jameel et al., 2012). In the UK, the importance of a connected society is widely discussed in government reports (DCMS, 2009; HM Treasury, 2012). As discussed earlier, broadband is argued to carry a number of benefits, including savings at an individual and household level as well as socio-economic and E-governance advantages. Table 2.3 summarises a number of potential savings identified from the Digital Britain report (DCMS, 2009) and the wider literature.

In response to the identified benefits of a well-connected society, in 2009 the UK Government set out the Universal Service Commitment (USC), a commitment to provide a universal broadband connection speed of at least 2Mbps to every home in the UK by 2012. In the 2009 Digital Britain Report (DCMS, 2009) it was stated that:

"More than one in 10 households today cannot enjoy a 2Mbps connection. We will correct this by providing universal service by 2012. As such, the UK's Commitment leads Europe. It has a measure of future proofing so that, as the market deploys next-generation broadband, we do not immediately face another problem of exclusion. The USC is also a necessary step if we are to move towards digital switchover in the delivery of more and more of our public services."

The delivery of the USC comprised a number of technologies including DSL, FTTC, wireless and satellite infill. It was funded by £200m direct public funding in addition to five other sources, comprising of:

- Commercial gain through tender contract and design
- Contributions in kind from private partners
- Contributions from other public sector organisations in the nations and regions who benefit from the increased connectivity
- The consumer directly for in-home upgrading
- The value of wider coverage obligations on mobile operators arising from the wider mobile spectrum package

Table 2.3: Economic Advantages of a Connected Society
Adapted from: Digital Britain report (DCMS, 2009)

Benefits	**Reasoning**
Financial savings	It is estimated that households that are offline lose around £560 per year in potential savings that could be made by shopping and paying bills online. A cumulative total of around £1billion (Lane-Fox, 2009)
Greater consumer choice	The Internet allows consumers to access products and services that would have otherwise been unavailable to them. This benefit is of particular importance in rurally isolated areas and is one of the key drivers behind rural broadband provision
Improved access to health and well being services	A connected society makes for easier communication between health services and the general public. Government initiatives such as the NHS 'N3' project also highlight the importance of superfast connections between NHS related sites within the UK. It is also widely argued that connectivity is a key driver in facilitating independent living for the elderly (DCMS, 2009)
Socio-economic inclusion for the physically and socially isolated and the economically excluded	Digital participation breathes new life into cultural understanding, formal and informal learning opportunities, engagement activities and employment opportunities
E-governance	A connected society is important for the delivery of online government services and the 'Big Society' mechanism of the current government. The Internet is seen as an important enabler in the process for citizens to self-govern (Yang and Bergrud, 2008)

The decision to set the minimum standard for broadband access at 2Mbps has come under some scrutiny, particularly in recent years as superfast services have begun to emerge. Independent broadband research groups such as thinkbroadband have commented that 2Mbps represents a minimum, and not a target speed and that evidence suggests the public believe a 2Mbps standard is no longer sufficient. Furthermore, at the time of the USC's adoption, the average broadband speed experienced by customers in the UK was already 4.1Mbps (Ofcom, 2009). It has been argued, however, that Government setting a standard of 2Mbps was an attempt to balance a reasonable connection speed with the associated cost of providing that service and the number of homes to which the service would need to be extended (BIS, 2010).

There has been little evaluation of the success of the USC, however, in the original Digital Britain report (DCMS, 2009) it was stated that 11% of all lines were unable, at the time, of delivering a 2Mbps standard. However, by 2011, the coverage of 2Mbps standard broadband was listed by Ofcom as 86%, which then rose in November 2012 to a figure of 89.9% (thinkbroadband, 2013). Recent reports state that the availability of 2Mbps standard broadband is 95.3% of UK households (as of May 2013). Although the 2Mbps standard set out in the USC may not have been reached by the 2012 deadline initially set, it is apparent that coverage of 'standard broadband' is increasing. Alongside the increased coverage of standard ADSL services, there has also been a significant rollout of superfast services within this timeframe. Tables 2.4 and 2.5 compare the availability of standard and superfast broadband services in the UK by region and measures of rurality.

Table 2.4: Standard Broadband Availability (percentage of households)
Adapted from: The availability of communications services in the UK (Ofcom, 2013a)

	All	Urban	Semi-urban	Rural
UK	95.3	98.3	96.9	80.1
England	95.8	98.5	97.1	80.6
Northern Ireland	87.4	96.0	95.6	66.5
Scotland	95.3	98.0	96.8	85.9
Wales	91.9	96.4	95.2	77.0
East Midlands	94.5	98.5	97.4	76.1
East of England	94.0	97.6	96.9	79.0
Greater London	99.2	99.3	100	87.4
North East England	95.1	97.4	96.1	80.5
North West England	96.3	98.1	97.3	79.0
South East England	95.7	97.1	97.6	85.3
South West England	94.7	98.5	97.9	81.4
West Midlands	96.1	98.8	97.1	79.6
Yorkshire & the Humber	94.5	96.9	95.8	78.9

The percentage of households with superfast availability is likely to increase significantly over the coming years due to infrastructure investment through government initiatives such as Broadband Delivery UK (BDUK). Rurally isolated areas are also likely to benefit from superfast fibre infrastructure in the future as this is the most viable option to supply broadband service to rurally isolated areas. Penetration rates will however, remain highest in urban and semi-urban areas as these have the highest population densities and highest take-up rates of superfast services. Ensuring a minimum standard of connectivity and accelerated rollout of superfast services is particularly relevant to this research as access to connections with sufficient bandwidth would logically influence the types of Internet services an end user engages with. For example, those Internet users in rural areas with poor quality, low bandwidth connections may not be able to use the Internet for media streaming whereas those consumers who live in urban areas may be adopting such applications at a fast pace given better infrastructure. In this sense, inequity of infrastructure provision may begin to differentiate Internet users in addition to traditional factors such as age and experience with technology.

Table 2.5: Superfast Broadband Availability (percentage of households)
Adapted from: The availability of communications services in the UK (Ofcom, 2013a)

	All	Urban	Semi-urban	Rural
UK	67.9	86.0	67.0	21.2
England	70.9	86.7	70.5	19.1
Northern Ireland	96.0	98.4	97.1	92.4
Scotland	47.6	72.3	48.3	6.3
Wales	39.8	90.1	33.8	6.6
East Midlands	67.1	93.5	68.8	17.9
East of England	68.5	94.0	74.4	14.0
Greater London	87.9	87.9	95.0	58.5
North East England	71.2	79.1	74.4	23.4
North West England	72.9	84.3	74.1	18.0
South East England	70.8	90.7	75.8	25.4
South West England	52.3	89.9	46.5	16.0
West Midlands	75.0	86.4	75.9	18.8
Yorkshire & the Humber	63.6	70.9	67.0	20.0

Within the Public Sector there have been a number of advances that have relied on the availability and use of broadband Internet. Broadly speaking these can be summarised into two categories;

broadband for public services and broadband for public sector information. In the UK, central government has developed a number of online information and service portals such as Directgov and Business Link in recent years. These services have since been superseded by GOV.UK, a tool that allows users to find government services and information online (gov.uk, 2013). The site acts as a central hub for key government services and allows users to submit online applications for a range of these, such as driving licenses and jobseekers allowance, and download the necessary forms for others, such as passport applications and managing benefits. The site is an example of how the Internet is being used in a public sector environment to deliver services faster and more efficiently, with reduced administration costs. There are a number of benefits to the end user too, as applications made online can be submitted, tracked and amended from anywhere over the Internet, whereas previously, such applications would require postal services or visits to council or government offices. In addition to central government services being migrated online, the majority of local authorities now provide online services to local residents. These services include systems for paying council tax, business rates, parking fines and other council fees as well as the ability to submit and view planning applications. Public sector information is increasingly seen as a powerful resource and broadband Internet has been instrumental in the availability and use of this information in recent years. The OECD recognise this importance stating that there are two main technological developments that have changed and shaped the role of public sector information and content. These are; technologies that enable the digitisation of public resources and the deployment of broadband technologies as these constitute a means of rapid dissemination (OECD, 2006). In the UK, large volumes of public open data are available online, and in most cases are free to download. There are a range of online tools provided by the Office for National Statistics, as well as academic institutions, that facilitate the dissemination of this data. Examples include the Nomis service for census and labour market statistics, and house price data provided by the land registry. Recent public sector data releases include the police.uk service, which allows the public to generate street-level maps and area statistics about crime and policing. The raw data is also available for download by month and is free and open. Such releases highlight the growing pressure for governments to be transparent about their data. The WWW is the best possible medium for dissemination of this data and has been a key driver in facilitating open data within the UK and worldwide.

2.3.1 Comparisons with the EU5

Across the EU, investment in broadband infrastructure is viewed as a priority from a top-down governance perspective. The EU Directorate Generals for Regional Policy, Agriculture and Rural Development and Information Society and Media, in their 'Guide to Broadband Investment' identify several key regional benefits of broadband infrastructure, noting: contribution to GDP growth and productivity gains, the creation of new jobs or businesses, minimising the digital divide and improving social cohesion (Fikkers, 2013). In 2010 the DCMS report entitled 'Britain's Superfast broadband Future' set out the government's ambition that the UK should have the best superfast broadband network in Europe by 2015 (DCMS, 2010). This development is seen as an important driver of economic growth and recovery from the recession. The government recognises that ubiquitous broadband access has the ability to increase consumer choice, reduce costs and streamline public service delivery. Broadband Delivery UK (BDUK), a team within the Department for Culture, Media and Sport has been responsible for the delivery of the government's broadband strategy, with £530m of funding allocated up to 2015 and the potential for an additional £300m up to 2017 (DCMS, 2011). The funding is to be used to support superfast infrastructure rollout in both urban and rural areas through the development of local broadband plans. Developed by county councils, unitary authorities and Local Enterprise Partnerships, broadband plans set out how broadband access will be provided within an area, and formed the basis of the application for BDUK funds. Once local plans were sufficiently developed, BDUK allocated funding and the infrastructural works were put out to tender to bidding suppliers. As of May 2013, 22 out of 44 BDUK projects had signed contracts with suppliers and were either being deployed or awaiting deployment. The political and financial backing for the BDUK program highlights the importance of the UK's vision to have the best superfast broadband network in Europe by 2015. Alongside the delivery of the government's BDUK program, Ofcom conducted research in the form of 'The European Broadband Scorecard'. The research aimed to identify, collate and publish the best available data relating to a number of indicators set out by BDUK regarding the development of the UK's broadband network, relative to those in other EU countries. The metrics included in the scorecard were: coverage, take-up, speed, price and consumer choice. Figures 2.6 and 2.7 show the rank of each EU5 country for a number of measures. Note that multiple countries can share a rank (i.e. all EU5 countries fall within the same banding for standard broadband coverage and are ranked joint first).

Figure 2.6 shows the position of the UK relative to the EU5 regarding broadband coverage, including fixed-line standard, superfast and mobile broadband services. The UK scores highly in standard fixed-line and mobile broadband availability, but falls below Spain and Germany for superfast services. Given the large number of superfast infrastructure projects funded by BDUK, the UK has the potential to have the best superfast coverage within the EU5 by 2015. Figure 2.7 shows the position of the UK relative to the EU5 regarding broadband take-up per 100 people. The UK scores highly in both mobile and fixed-line take-up, yet sits mid-table for superfast take-up. This is likely due to much of the UK's superfast infrastructure not being fully deployed yet.

With regards to pricing, the UK was found to be the cheapest of the EU5 for broadband and fixed voice services. Similarly, the best offer price for mobile broadband offering a 1GB per month

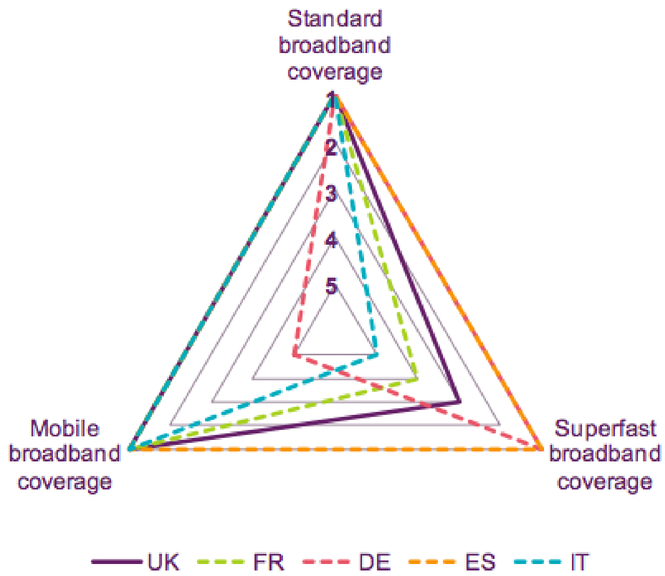


Figure 2.6: Visual Overview of the UK's Position Relative to the EU5: Coverage
Source: The European Broadband Scorecard (Ofcom, 2013b)

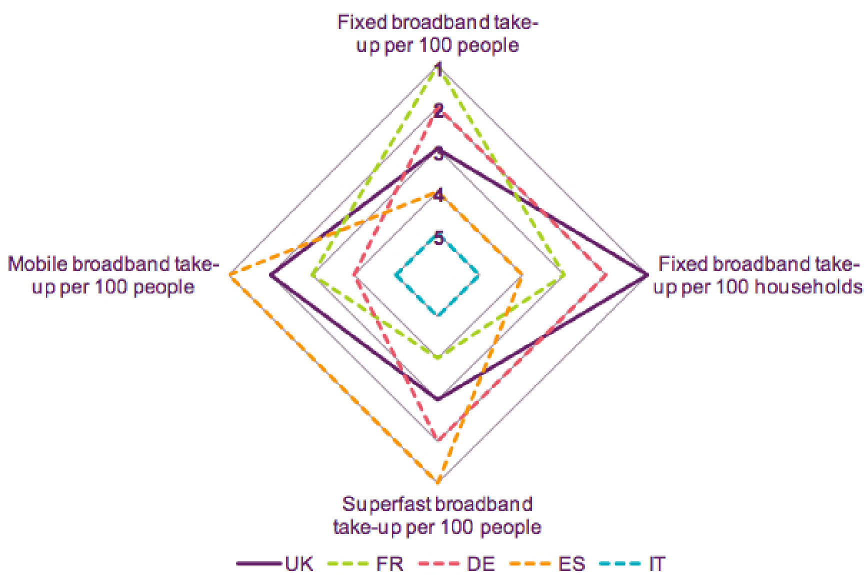


Figure 2.7: Visual Overview of the UK's Position Relative to the EU5: Take-up
Source: The European Broadband Scorecard (Ofcom, 2013b)

download limit was from a UK operator at £6 per month. Packages with higher monthly download limits (3GB and 5GB), however, were cheapest in Italy, placing the UK second overall within the EU5 for mobile broadband pricing. In terms of choice, two proxy measures of consumer choice in the broadband market were used; these were the percentage of fixed broadband lines operated by the incumbent supplier in each EU5 country and the market share of the largest mobile network operator. The UK was shown to have the lowest percentage of fixed broadband lines operated by the incumbent supplier (BT) at 31%, representing the least centralised fixed broadband market. The second least centralised market was France, with 42% of fixed broadband lines operated by the incumbent supplier. The most centralised broadband market was Italy at 53%. In the mobile market, the UK was tied with Germany and France, with the leading mobile network operators (EE in the UK) operating 33% of connections. This represents the least centralised markets. The most centralised markets were found in France and Spain where the leading mobile operators operated 40% and 42% of connections respectively. There was no comparison of broadband speed between the EU5 countries included in the Ofcom study as there were no fixed-line broadband upload speed and mobile broadband download speed datasets that might be considered for publication (Ofcom, 2013b).

2.4 Broadband and the Digital Divide

Previous sections have discussed the constraints and influences on engagement within the context of evolutionary infrastructure; however, disparities in access and engagement have been the focus of research for some time. The term 'digital divide' was introduced to describe those economic and social inequalities between groups, in terms of the access to, and use of information and communication technologies. The term was coined during the first decade of the Internet age and quickly became shorthand for a plethora of disparities within the online community (Norris, 2001). The global expansion in network infrastructure and uptake has been fast paced since the birth of the Internet and inevitably some have been left behind, most notably less developed nations at an international level and areas which are rurally isolated at the national level (Wilson et al., 2003; LaRose et al., 2007). Digital divide research over the past two decades has identified an extensive number of potential limiting factors to engagement, but has moved away from discussion of a more general digital divide (Norris, 2001) to encapsulate a broader range of factors including age (Rice and Katz, 2003), ethnicity (Prieger and Hu, 2008) and gender (Cooper, 2006) within the contemporary period.

The development of network infrastructure and the use of the Internet has not been evenly distributed across the globe. In many cases, investment in digital technologies has been most prominent in countries with advanced economies. Developed nations such as Sweden, Australia, the United Kingdom and the United States, which have invested heavily in communications infrastructure have been at the forefront of the technological revolution and these networked nations have been well placed to pull even further ahead of developing nations through increased economic opportunity (Norris, 2001).

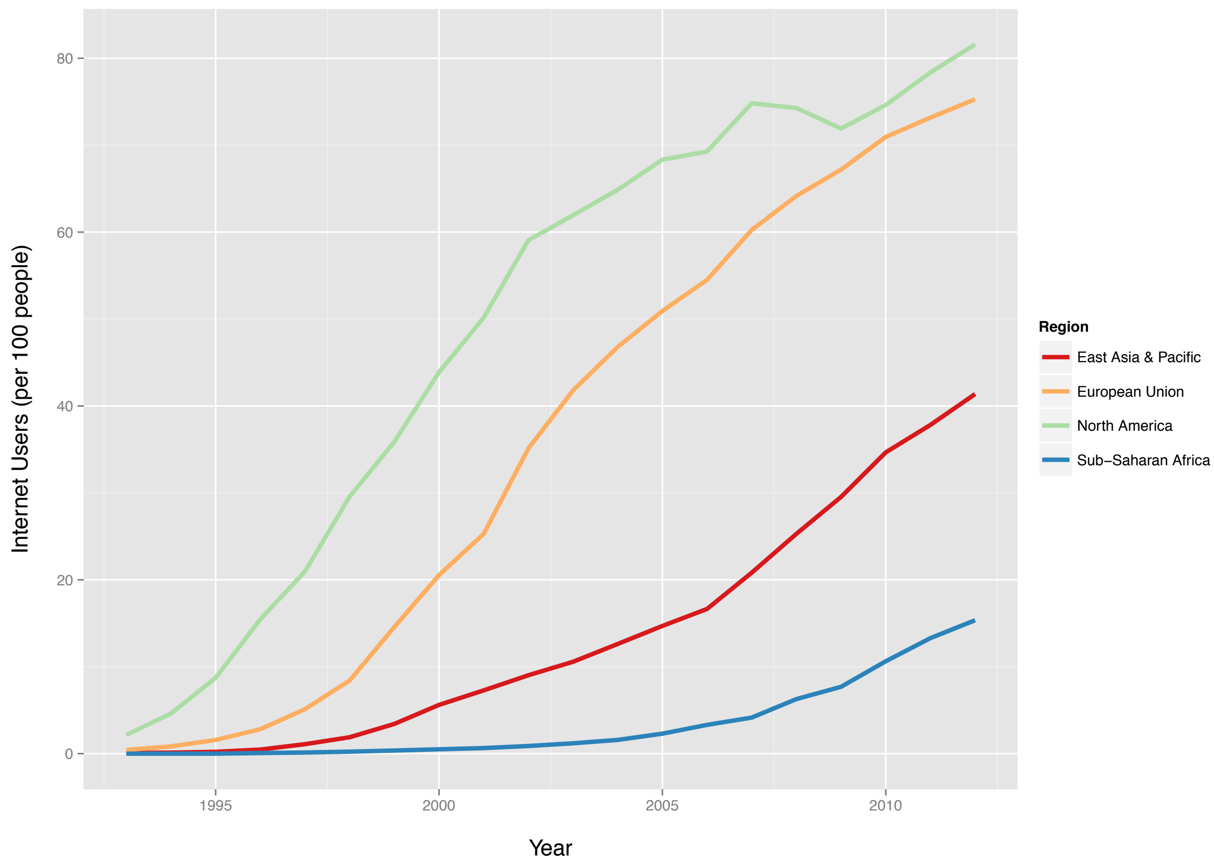


Figure 2.8: Internet Users per 100 People: Worldwide Divides: 1993 - 2014.
Source: Based on World Bank World Development Indicators.
Accessed: http://databank.worldbank.org/data/home.aspx

Figure 2.8 shows the extent of worldwide disparities in Internet access between four key regions between 1993 and 2014. There are clear divides between those regions that are developed, with advanced economies and those which are developing. Mid-level and emerging economies such as Mexico, Bangladesh and South Korea have benefited from global developments in ICTs whilst not being at the forefront of digital developments. Whilst much of the network infrastructure that supports the Internet has been developed and deployed across developed nations, mid-level economies have been instrumental in the manufacture of computer hardware and software, as well as support services. This involvement has encouraged the adoption of modern ICTs in these nations, but poorer economies still trail behind (Norris, 2001; James, 2007). With developed nations benefitting from near-ubiquitous access to the Internet and network technologies, and discussion of 'digital differentiation' as opposed to a more traditional digital divide in these areas; the international digital divide can be defined more simply. On an international scale, the digital divide can be expressed as 'haves' and 'have-nots' or more specifically, those who have access to ICTs and those who do not. The 'have-nots' often being poorer nations that have historically struggled with conflict, disease, famine and political issues. The digital divide at the international scale does, however, have a second dimension. Researchers have suggested 'access' is an ambiguous term when used in this context, often referring to an individual's ability to physically access a computer or other communication technologies. Although individuals in poorer nations may have access to this equipment, they may not necessarily be able to use it. Therefore, any realistic notion of access to ICT must be defined from an individual's perspective (Selwyn, 2004). Recent research has attempted to visualise these worldwide disparities in Internet access. Graham et al. (2012) uses statistics from the World Bank to illustrate the raw numbers of Internet users in each country as well as the percentage of the population with Internet access (see Figure 2.9). Clear disparities between developed and developing countries can be seen.

Academic studies into the development of broadband networks in developing countries are scarce. The most notable research to date has been undertaken by the Oxford Internet Institute (OII) in the form of a collaborative project with the University of Nairobi. The project aims to monitor the development of broadband Internet access in East Africa and commenced in 2010.

2.4.1 Domestic Digital Differentiation

In addition to inter-country divides, there has also been extensive research into those barriers to access that have emerged within developed countries. Widely researched factors include age (Loges and Jung, 2001; van Dijk and Hacker, 2003; Lenhart et al., 2005), gender (Cooper, 2006), rurality (Pigg and Crank, 2005; Prieger, 2013) and ethnicity (Prieger and Hu, 2008). In terms of barriers to access, age is seen as one of the most influential and as such, has attracted a large number of academic studies over recent years. Broadly speaking there has been a long-established digital divide between those who are young and those who are elderly. Lenhart et al. (2005) suggest that as younger generations have grown up with technological advances, they are quicker to learn and adopt new technologies. Older generations however, are generally slower to adopt either through a lack of understanding or a lack of interest. Such age disparities have long been known, and research conducted by the Pew Internet and American Life Project in 2001 detailed that 87% of

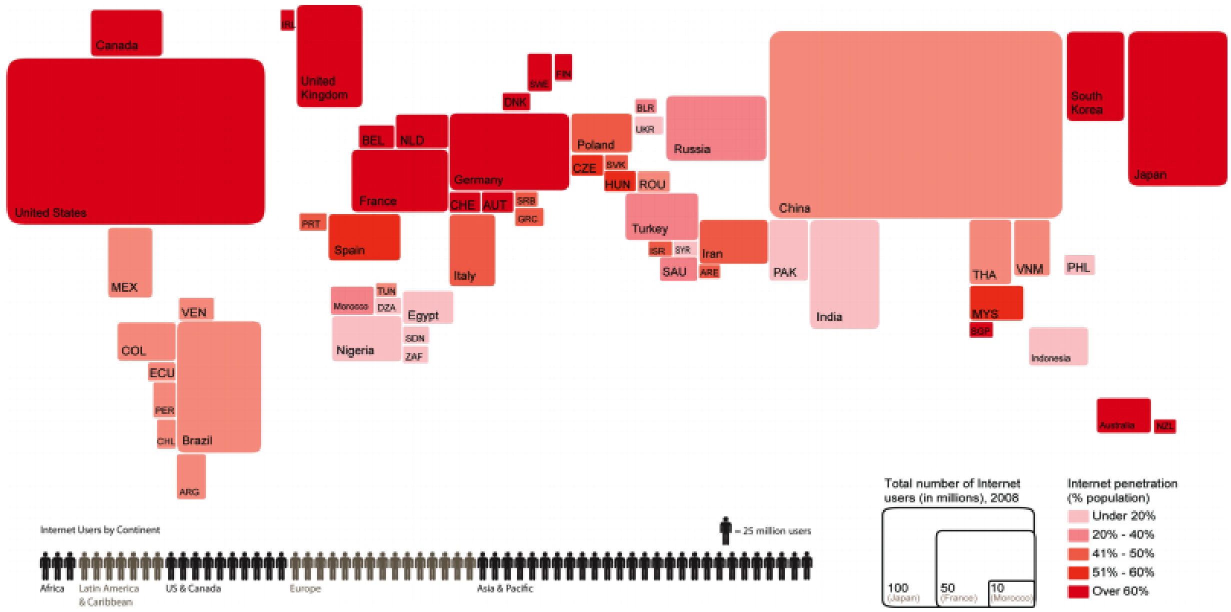


Figure 2.9: Digital Divide: The Geography of Internet Access. *Source: (Graham et al., 2012)*

U.S. teens aged 12-17 used the Internet, up from 73% in 2000. By contrast, just 66% of adults used the Internet, up from 56% in 2000 (Lenhart et al., 2005). This supports the notion that age is a primary limiting factor in an individual's ability to be active online. With ubiquitous broadband deployment in many developed countries, it may be that these figures increase further, however, it may be argued that increased participation amongst younger generations would serve only to widen the participation gap between the young and the old. Other research has considered how age impacts use and engagement with the WWW as a multidimensional concept, including consideration of the scope and intensity of the relationship that people develop with the Internet. Loges and Jung (2001) show age to be significantly associated with access, but also with a tendency to pursue a narrower range of personal goals online. It was also found that older users connect to the Internet from a smaller range of places, suggesting that that the 'mobility' of some broadband connections is utilised more by younger users.

In the UK, overall levels of digital exclusion have declined steadily in recent years (Lane-Fox, 2009). However, a significant proportion of the population remains digitally excluded. In 2009, 10.2 million adults (21% of the UK population) had never accessed the Internet and a further 2 million had not used it for 3 months: 7.8 million households (30% of those in the UK) had no Internet connection at home (Lane-Fox, 2009). In terms of assessing the generational divide in online participation in the UK, it was found that exclusion was not uniform across different ages, with 62% of the adults who had never accessed the Internet (6.4 million) being over the age of 65. A generational divide within the UK is therefore apparent, and as such, government digital inclusion schemes such as GoOnUK have been created to encourage volunteers who possess adequate digital skills to teach others, often the elderly, how to engage online. Despite the elderly generally being less engaged with modern ITCs, there are a small proportion who are digitally enabled, most of whom tend to use electronic communications in a more instrumental manner than younger generations, or in other words, as a tool for achieving other benefits rather than as a form of leisure activity or entertainment (Longley et al., 2008).

The second factor that is widely regarded as impacting an individual's ability to engage online is rurality (Parker, 2000; Wilson et al., 2003; Xavier, 2003; LaRose et al., 2007). In these studies, rurality is a physical barrier to access and refers to the relative remoteness of an area in terms of broadband or mobile service availability. In this sense, and unlike age, rurality can be viewed as a geographic differentiator, as opposed to a personal differentiator, of engagement. Living in a rurally isolated area may inhibit an individual's ability to engage online, but that is not to say that those who live in rurally isolated areas have a lesser desire to participate. Given those technical limitations of broadband infrastructure discussed earlier (the requirement to be in relatively close proximity to a local telephone exchange), rural areas can often suffer from poor service or no service at all. Private sector investments within rural areas have historically been more limited, with profit from rural infrastructure investments marginalised through the higher costs of installation and smaller numbers of end users (Parker, 2000). In the UK, the government has indicated fibre deployment will play a significant role in supplying broadband to the most hard to reach areas (DCMS, 2011; Yiu and Fink, 2012); however, there is a high cost associated with this level of infrastructure investment, hence the delivery of schemes such as BDUK. Previously, some rurally isolated areas have had to rely on satellite broadband that suffers from limited bandwidth and reliability and comes at a high cost, to infill where fixed-line services are not available (Ofcom, 2014). While rural availability of mobile broadband is lower than in urban areas (Ofcom, 2012), recent studies have suggested that this still helps fill in gaps where fixed broadband coverage is limited (Prieger, 2013). However, there are higher costs associated with mobile broadband connections, as well as limited data allowances, which make this an unsuitable substitute for fixed-line service in the UK, particularly as data consumption will likely increase as the Internet becomes increasingly content rich.

Two further factors that have been widely researched within the context of the digital divide are gender and ethnicity. There have been a number of academic research projects that address the gender divide in terms of Internet access (Bimber, 2000; van Dijk and Hacker, 2003). Some researchers trace a gender divide to the idea that computers and the Internet (as well as technology in general) are gendered (Wilson et al., 2003). Early empirical studies, such as that by van Dijk and Hacker suggested that gender was significantly related to digital divides and that the possession of computers and network connections was significantly influenced by an individual's gender (van Dijk and Hacker, 2003). This was however, based on limited data and occurred long before developed countries had ubiquitous access to the Internet. By 2002, reports had infact already suggested that there were no gender imbalances in accessing the Internet, indicating that in 2001, 53.9% of men and 53.8% of women reported using the Internet (NTIA, 2002). These findings were extrapolated from much larger samples, in this instance, a survey of over 137,000 individuals across the United States. Cooper (2006) offered the concept of 'computer anxiety' as an explanation for earlier digital divides based on gender, explaining that the digital divide is fundamentally a problem of computer anxiety whose roots are deep in socialization patterns of boys and girls and that interact with the stereotype of computers as 'toys for boys' (Cooper, 2006).

Similarly to research into gender divides, the impact of race and ethnicity on digital divides has been widely researched with conflicting results. Early research suggested that computer and Internet users are divided along the lines of race, with white households far more likely to have computer access from their homes than black or hispanic households (Wilson et al., 2003). However, research conducted by Prieger and Hu (2008) looked at a geographically fine dataset of DSL subscriptions in the United States and found that gaps in DSL demand for black and hispanic groups did not disappear when income, education and other demographic variables were accounted for (Prieger and Hu, 2008). Conversely, Rice and Katz (2003) suggested that the gap between Internet users and non-users is associated with income and age, but no longer with gender and race. The ethnic and demographic makeup of areas is arguably a key driver behind an individual's ability to become digitally enabled, however, many studies that reveal relationships between ethnicity and digital exclusion have been conducted in a local context. Much of the current research in this area was not only conducted while Internet access was far from universal, but relates almost solely to the United States. Although viable examples of academic study into ethnicity and digital exclusion, they are difficult to transpose to the UK context. Arguably, it is because of contradictory research into these coarse digital divides that the digital differentiation approach began to emerge in the early 2000's. Further divides were found to exist as adoption of the Internet became more commonplace, particularly within developed countries. Early research into a second level digital divide such as that by Peter and Valkenburg (2006) confirmed that the emerging digital differentiation approach described current digital divide phenomena more adequately than the disappearing digital divide approach (Peter and Valkenburg, 2006). The study was based on multivariate analyses of a survey of over 700 Dutch adolescents, whose unequal access to socio-economic and cognitive resources was investigated alongside their use of the Internet. As well as supporting the general notion of a digital differentiation approach, the research concluded that adolescents with greater socio-economic and cognitive resources used the Internet more frequently for information and less often for entertainment than their peers. Similar patterns were also found with regards to adolescents' tendency towards 'ubiquitous Internetting' (using the Internet for multiple applications). Alongside earlier research by Hargittai (2002), which criticised existing digital divide literature as having limited scope, based on its tendency to focus on the binary classification of 'haves' and 'have nots', it is apparent that the digital differentiation approach has become widely adopted. Most recently, discussion has moved beyond differences in skill levels and introduces speed and type of access as factors that differentiate Internet users. Modarres (2011) states that "in the evolving world of mobile computing and an uninterrupted 24/7 world of 4G connections, access to a desktop computer and DSL can be seen as a lighter shade of connectivity". Similar concepts are echoed by Dutton and Blank (2011) who introduce the concept of the 'Next Generation User', as a group who reportedly comprised 42% of Internet users in Britain in 2011, and have been rapidly increasing since 2007. The characteristics of the next generation user are defined by two separate trends; portability and access through multiple devices such as smartphones and tablet computers. Research has also explored the relationship between the next generation Internet user and income, indicating that the two factors are strongly linked, with the proportion of next generation users increasing in line with annual income. Such outcomes would indicate that divides between first generation and next generation Internet users may be underpinned by a more complex mix of socio-economic conditions.

2.5 Conclusion

This chapter has reviewed a wide variety of literature sources that relate to the development and application of the Internet since its inception as a research project in the late 1960's, through to the modern Internet and World Wide Web of the present day. In doing so, a number of themes have been discussed, including; the physical infrastructure of the Internet, development of protocols and content, advances in associated technologies including mobile telephony and wireless Internet access, government regulation and investment in communications technology, social and economic implications and the emergence of digital divides and differentiation amongst users.

To summarise and conclude, this chapter has discussed how the development of ARPANET, a research project commissioned by the US Department of Defence, which sought to link together computers at academic institutions, led to the development of large scale networks of computers and standardised communication protocols. This project would become the foundation of the modern Internet which was rolled out to the commercial market with widespread investment from telecoms providers throughout much of the 1990's (see Sections 2.1.1 - 2.1.3).

This chapter has also examined how the rapid roll-out and uptake of the Internet was assisted in part by the development of the World Wide Web by Sir Tim Berners-Lee in the early 1990's and the emergence of graphical interface browsers in the same period. The reduced cost of personal computers and release of Web browsers meant that by the mid 1990's many consumers we able to easily access the Internet for the first time. In the years that followed, content available through the World Wide Web grew rapidly and access to the Internet became increasingly omni-present (see Section 2.1.3).

The literature has also highlighted that as access to the Internet became more commonplace, technology that utilised such connectivity developed significantly. Throughout the late 1990's and 2000's mobile handsets were developed with Internet access as a prominent feature (see Section 2.1.8), as were many consumer goods such as personal computers, and later, tablets and smart goods such as televisions and personal video recorders. During the same period, the technology that delivered Internet connectivity through both fixed-line and wireless connections had evolved (see Section 2.1.4). fixed-line connections quickly progressed form narrowband dial-up, to wideband DSL and eventually ultra-high bandwidth cable and fibre. Access to wireless connections developed with similar pace, moving through standard telephony (1G) to narrowband data capable (2G), to high speed data (3G) and modern wideband 4G connections that cover much of the developed world today. Advances in both consumer hardware and the technical underpinnings of the modern Internet have seen access grow rapidly over the last 20 years (see Figure 2.2).

In addition to discussing the development of the modern Internet and associated technologies, the review of literature has presented a number of themes relating to the advantages of Internet access from the perspectives of; economic advantage, education, entertainment and lifestyle applications. In terms of economic advantage, studies have demonstrated how adoption of the Internet and related communications technologies have facilitated, albeit in part, the death of distance. This has been achieved primarily by reducing the costs and constrains of traditional long distance communication and enabling new forms of trade and consumption. In some industries, such as retail, this has had a significant impact, with the Internet supporting the digitisation and distribution of some goods and services. In terms of those products and services that cannot be digitised, widespread adoption of the Internet has at the very least enabled increased competition, choice, access to and reduced prices for a typical consumer (see Section 2.2.5). The Internet has also allowed many businesses

to operate globally, thus stimulating noticeable growth in the GDP of the UK in recent years. Literature currently suggests that the Internet now accounts for around one quarter of the UK's economic growth. (See Section 2.2.3).
Given measurable benefits of widespread Internet access, discussion of the UK government's key policy responses to promote rollout and adoption of the Internet has been presented (see Sections 2.1.5 and 2.3). The promotion of competition amongst broadband service providers through Local Loop Unbundling was a key driver in lowering prices and ensuring more widespread adoption of the Internet throughout much of the early 2000's. More recent government intervention has been user-centric, aiming to promote the use of the Internet amongst populations who may be unengaged or have limited access, thus mitigating the impact of digital divides between those who are computer literate and those who are not. In addition, recent government policy has funded rollout of high bandwidth infrastructure, typically to areas that are rurally isolated, in an attempt to ensure universal minimum thresholds of service.
Alongside increased economic advantage, similar success has been reported with regards to the Internet as a source of educational content (see Section 2.2.2). With the death of distance and rise of communications technology came the opportunity for distance learning and knowledge sharing through Virtual Learning Environments (VLEs), using the World Wide Web as a medium to host virtual courses and content, available to almost anyone with access to personal computer. The use of the Internet to deliver educational courses has evolved rapidly over the past decade to include offerings such as online degrees from top academic institutions to Massively Open Online Courses (MOOCs), which are often free and open to any interested student.
With regards to entertainment and lifestyle applications associated with the development of the Internet, this chapter has discussed media sharing and streaming as examples of how the digitisation of traditional media has changed consumer behaviours. The associated impact of the Internet on traditional retailers has also been examined, noting a significant shift from bricks and mortar retailing to omni-channel digital stores, by the majority of large retailers as a means of adaptation (see Section 2.2.5). The emergence of increasingly advanced software and connectivity in every-day consumer goods has also been discussed as a by-product of widespread Internet access. Such advances in consumer technology allow smart devices to be controlled remotely from any location given there is Internet access. Such is the proliferation of these devices that the term 'the Internet of things' is used to describe the vast networks of everyday objects that are connected to, and communicating with one another via the Internet.

Despite numerous advantages enabled by the development of the Internet, such high levels of connectivity have presented opportunity for unethical or contentious application (see Section 2.2.6).

As access to the Internet has become more commonplace and less constrained by location, issues surrounding privacy and surveillance have emerged. For governments in particular, the Internet, and the vast networks of coded objects connected to it, make it possible to locate and monitor people's activity with relative ease. Similarly, these vast networks of users can be subject to cyber crimes, as the Internet offers a simplified, cost effective and repeatable means to conduct rapid and large scale criminal attacks. The growth of such crimes has been significant in the contemporary period, largely due to the reduced risk and increased anonymity the Internet can provide.
As well as user-centric disadvantages of the Internet, more general issues exist, predominantly focused on the divides between users and non-users, or the 'haves' and 'have-nots' (see Section 2.4). Since widespread adoption of the Internet throughout the 1990's, people, and in some cases nations, have been left behind. This gap was quickly coined 'the Digital Divide' and attracted much attention through academic research in the early years of mainstream Internet adoption. Common factors that were presented as influencing this divide at an individual level in developed nations included age, gender and ethnicity. In the international context, divides were also clearly visible between those countries that were developed and those that were developing. It is noted that on an international scale, these divides are still significant in the present day. On a domestic scale however, divides between users and non-users have narrowed to such an extent that research has moved away from this binary concept, instead focusing on the factors that differentiate Internet users (see Section 2.4.1). The concept of digital differentiation predominantly explores differences in use, perception and access to the Internet in developed nations, with prominent academic studies noting differentiation between those users of different ages, locations and demographic backgrounds.

The following chapters in this thesis aim to build on the key themes developed in this review of literature. In particular, the research aims to explore the current state of infrastructure and service provision within the national context, differences in Internet user perceptions and behaviours, how these are differentiated in a socio-spatial context and how quantitative data sources can be better utilised to explain these differences. In doing so it is anticipated that a better understanding of the issues that influence Internet access and engagement will be achieved. The proposed research will be situated within the modern context of digital differentiation, but will utilise advanced quantitative methods and big data, with the inclusion of crowdsourced data, to explore differentiation, thus closing a current research gap and better framing the topic given the availability of vast quantities of relevant, yet broadly untapped, data sources.

Chapter 3

Exploring the Geography of Access to Fixed Line Broadband Services in England Using Crowd Sourced Speed Check Data

3.1 Introduction

N.B. The majority of the research presented in this chapter was reformatted for publication in: Riddlesden, D. and Singleton, A. D. (2014). Broadband speed equity: A new digital divide? *Applied Geography*, 52(0):25 - 33.

Broadband infrastructure and access to high speed Internet are topics that have become increasingly important in recent decades. Since its inception in the early 1990's, the Internet has grown from a niche form of communication between academic and research institutions to a global network of networks with over 2.4 billion users worldwide. Furthermore, development of the World Wide Web and advances in computing hardware, coupled with a desire for increasingly ubiquitous Internet access has seen access to broadband infrastructure become something of a necessary precondition for economic growth and competitiveness (Picot and Wernick, 2007). Literature examined in Chapter 2 has detailed how increased access to the Internet has become so important to society, influencing economic competitiveness, education, lifestyle and social capital, thus the importance of broadband infrastructure is recognised by many developed and developing nations alike, the world over. The UK is arguably at the forefront of broadband infrastructure deployment, investing heavily in superfast infrastructure roll-out schemes such as Broadband Delivery UK (BDUK) and the Rural Community Broadband Fund (RCBF) in order to maintain a competitive edge over other countries within the EU, and to have the largest and most accessible superfast broadband network in Europe by 2015. The UK are not alone in this widespread investment, similarly, the United States have adopted the 'National Broadband Plan' which aims to promote broadband availability through ensuring robust competition, universal service for all and maximising the benefits of broadband in government influenced sectors.[1]

The purpose of this chapter is to examine the geography of access to fixed-line broadband services in England and to assess the extent to which physical infrastructure constrains are likely to be impacting use and engagement at national, regional and sub-regional scales. Many reports commissioned by government and industry regulators such as Ofcom detail increasingly high nationwide average speeds, justified in part by the rollout of superfast broadband infrastructure, but with obvious spatial inequity (Ofcom, 2010). Although increases in aggregate speeds are a fair indicator that infrastructure is developing, there is little examination of the socio-spatial structure of access to broadband services, both existing and next generation (superfast). Similarly, in an academic context, there have been few studies to date which examine nationwide disparities in broadband performance, most likely given the scarcity of good quality data, and high costs associated with data acquisition from third parties. As such, this chapter examines these disparities in broadband performance relative to rurality, deprivation and other socio-spatial indicators; and additionally identifies spatio-temporal change in broadband performance over a four-year period. The overarching objective is to highlight those areas which are gaining and losing in terms of investment in infrastructure, and provide a holistic view of the current geography of broadband in England. In addition to the quantitative analysis within this chapter, a qualitative case study of rural broadband investment is presented. The purpose of this is to highlight where those investment mechanisms as discussed in Chapter 2, Section 2.3 may be failing to deliver adequate infrastructure upgrades to the most rurally isolated areas. Given a lack of quantitative measures of rural broadband investment, a case study of an infrastructure overhaul project in Kings Cliffe, Northamptonshire is presented to highlight the needs and constraints in rural communities.

[1] Connecting America: The National Broadband Plan `http://www.broadband.gov/`

3.2 Data Characteristics

In order to undertake the analysis outlined above, a number of datasets were required, including:

- A crowdsourced dataset of geo-tagged Internet speed test estimates totaling 3.6 million results, spanning the period 1/1/2010 to 31/5/2013
- A Dataset of UK telephone exchange locations for 2010

Each of the datasets are documented in the following sections.

3.2.1 Crowdsourced Speed Test Datasets

Crowdsourced speed test data were provided by `www.broadbandspeedchecker.co.uk`, which provide a web-based application enabling users to test their Internet connection speeds. When users visit the website, a page is loaded with an embedded testing application that when run provides a small file (of known size) that is automatically downloaded and uploaded, thus enabling speeds to be estimated (i.e. size / time). The data that has been supplied for the purpose of this research consists of those results, which are augmented with a user supplied postcode, which enables georeferencing and allows the speed test estimates to be mapped. In the context of the broadbandspeedchecker website, the user supplied postcodes allow the test outcome to be displayed within the context of other results proximal to the user, thus allowing instant, albeit un-validated, comparison of local speeds. Users who run tests on the website are not required to save their results or supply a valid postcode, these are optional, however, all results derived through the website are stored by Speedchecker Limited as part of their terms of use.

Similar data capture methods are employed across a number of testing websites including `www.speedtest.net` operated by Ookla. Internet speed test databases have become an increasingly utilised resource in recent years with many companies showing interest in broadband performance statistics for marketing, location planning or market research purposes. For example, British property website Rightmove[2] now gives users access to localised aggregate broadband performance statistics for every property listed on its website, highlighting the importance of broadband connectivity to homebuyers and the wider property industry. Ookla, who are currently the largest speed testing service on the World Wide Web report over 3 million speed tests run each day through their service, indicating considerable consumer demand for measuring broadband performance.

Within an academic context, crowdsourced speed test data have not previously been used to examine spatial disparities in broadband performance and access. As such, this research aims to identify

[2] `www.rightmove.co.uk`

suitable methodological approaches in validating and analysing large crowdsourced datasets of this type, as well as reporting empirical findings concerning their spatio-temporal geography.
Internet speed tests performed through websites such as broadbandspeedchecker test three main criteria of a user's broadband connection, these are; ping (or latency), download speed and upload speed. Ping is the reaction time of a connection, or how fast a response is received after a request has been sent. A fast ping results in a more responsive connection, especially in applications where fast response is important, for example, online gaming. The latency of a users connection is also important in calculating download and upload speed estimates. Download speed is how fast data can be transferred from a server to an end user. Most domestic connections are asymmetric DSL (ADSL), which allows greater bandwidth for downstream traffic (See chapter 2, Section 2.1.4). This results in higher download speeds than upload speeds. Because most Internet activity involves downloading as opposed to uploading data, asymmetric bandwidth profiles offer a better overall connection experience. Download speeds are generally recorded in Kilobits per second (Kbps) or Megabits per second (Mbps). The raw data supplied for this research details download and upload speeds in Kbps. Upload speed is how fast data is transferred from a user to another user (or server). Uploading is necessary for file sharing (through email or other applications), VoIP applications (to send your video and audio feeds) and online gaming, amongst other Web-based services.

For the following analysis, the speed test dataset is split into two time periods. The first covers the period 1/01/2010 to 31/01/2011. The second covers the period 1/04/2012 to 31/05/2013. The provision of data covering a relatively long period of time adds a further dimension to this research, allowing for spatio-temporal changes to be identified, a further research gap in this field. Table 3.1 provides an overview of the data collected by the speed test application provided by `www.broadbandspeedchecker.co.uk`.

Table 3.1: Data Characteristics: Speed Test Databases

Variable	Descriptor	Contents
V1	Download	Download speed recorded at the time of the test (results in Kbps)
V2	Upload	Upload speed recorded at the time of the test (results in Kbps)
V3	Date/Time	Date and time stamp of the test
V4	Postcode	User supplied postcode
V5	Provider	Broadband service provider (optional field filled by user)
V6	Package	Broadband package details (optional field filled by user)

It is important to note that these tests differ from 'official' speed tests in the UK, which are collated on behalf of the industry regulator Ofcom (`www.ofcom.org.uk`) by SamKnows (`www.samknows.com`), who are an organisation that provide information about broadband performance, providers and usage. The crucial difference between SamKnows data and the data supplied by Speedchecker Limited is the collection method. Rather than relying on users to run speed tests through a Web application, SamKnows supply hardware in the form of a small testing box that sits between a participant's existing router and the rest of their network. Boxes are supplied to a representative sample of Internet users nationwide. In 2010, Ofcom's UK broadband performance report utilised data collected from 1,506 testing boxes. The boxes automatically ran speed tests on a user's connection, but only when there was no other network activity. Conversely, data supplied by Speedchecker Limited is much larger, but there are no restrictions to prevent users from performing tests when there is other network activity ongoing (e.g. multiple users online within a property, or a background update being downloaded). As such, the data used in this research could be interpreted as those actual speeds people attain when using a service, taking into consideration local constraints related to router configuration, WiFi coverage or coincident household usage. This said, early comparisons between the regulator estimates and the derived Speedchecker Limited estimates revealed very similar figures. The data sample covering 2010 suggested a nationwide average download speed of around 4.8Mbps, close to that estimated by Ofcom/SamKnows at 5.1Mbps (Ofcom, 2010). The data used for this study can be considered as 'Volunteered Geographic Information' (VGI) (Haklay et al., 2008; Goodchild, 2007), which although a useful way to generate large amounts of geographic information, require interpretive caution related to data accuracy, coverage and bias.

3.2.2 Dataset of UK Telephone Exchange Locations

In order to validate the speed test records and calculate the relative distances between each test location and the closest telephone exchange, a georeferenced dataset of all telephone exchange locations was required. This was compiled by running a custom Python script to Web scrape the data from a Google maps API at `www.samknows.com`. Samknows have conducted a large amount of research in recent years, including a partnership with Ofcom in 2011 in which data was provided for the UK fixed-line broadband performance report (Ofcom, 2011). As such, samknows are a reputable source of research reports and data in this sector. The data obtained forms a comprehensive database of all UK telephone exchanges, georeferenced by latitude and longitude. As well as location information, data was retrieved relating to which ISPs were operating in each exchange, what services in terms of ADSL and SDSL were available and the Ofcom market area. Market areas are based on the number of operators who offer service in each exchange, this is used as a proxy for the competitive conditions in different areas. Currently, Ofcom has defined four different geographic markets.[3]

- The Hull area: (0.7% of UK premises)
- Market 1: exchanges where only BT is present (14.2% of premises)
- Market 2: exchanges where two or three POs are present or forecast (13.8% of premises)
- Market 3: exchanges where four or more POs are present or forecast (71.3% of premises)

However, the data revealing the services available in each exchange were not used for analysis, as providers continually upgrade equipment and offer new packages to consumers, therefore the accuracy of service and package data can be compromised very quickly. Similarly, no analysis was conducted using the Ofcom market area segmentation as this changes in line with new services and operators, and is difficult to validate. It is also important to note that exchanges can close or re-open depending on market conditions and at the discretion of the incumbent suppliers (British Telecom and Kingston Communications) however, changes to the locations and operational status of telephone exchanges occur very infrequently.

Because the dataset was Web-scraped, a number of checks were made to assess data quality, this included plotting the coordinates of the exchanges contained within the dataset and ensuring they all fell within the boundaries of the United Kingdom (see Figure 3.1). In addition, the exchange dataset was loaded into a GIS, and attribute information for each exchange was viewed alongside its projected location. Extensive checks were made to ensure that the exchange 'name' field was relevant to the projected location (for example 'Bedford' fell within the geographic extents of Bedford). In all cases, attribute information was valid.

3.2.3 Data Cleansing and Validation

Prior to validation, a series of data cleansing steps were required. The first was to remove any non-geographic postcodes that appeared in the Office for National Statistics Postcode Directory (ONSPD). The ONSPD is an ONS geography product which details all current and terminated postcodes in the UK alongside relevant attribute information such as the administrative, health and other geographic areas in which each postcode falls. The ONSPD was utilised in the analysis for the purpose of validating user supplied postcodes in the speed test datasets. Non-geographic postcodes are generally only used for mail routing and cannot be used for navigation or distance finding. Some companies also have custom postcodes to reflect their names, for example BT group

[3] Review of the wholesale broadband access markets. `http://stakeholders.ofcom.org.uk/consultations/wholesale-broadband-markets/summary`

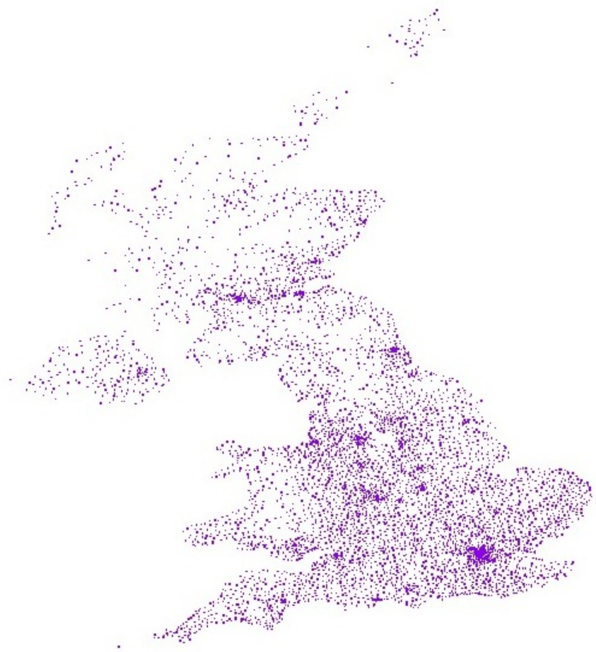


Figure 3.1: Exchange Locations (WGS84 Projection)

use the postcode DH98 1BT. Within the ONSPD, non-geographic postcodes have no geographic coordinates (easting and northing in this instance). Running a short script written in R (a statistical programming environment, which is used for the majority of analysis in this thesis) removed entries without location information from the dataset.

The second step was to calculate the Euclidean distance in metres between every current postcode in England (obtained from the ONSPD) and the closest telephone exchange. The purpose of this analysis was to use distance to the nearest telephone exchange as a key indicator of likely broadband performance (Ofcom, 2012). In order to calculate the Euclidean distance, the ONSPD and exchange location database were loaded into R and the 'FNN' (Find Nearest Neighbour) package was used to calculate distances between the coordinates (easting and northing) of each postcode centroid and the closest exchange. Exchange locations were originally supplied in longitude and latitude format, but were reprojected as easting and northing (British National Grid) to match the format of the ONSPD. In any instance where the closest exchange could not be located (because the postcode centroid was equidistant between two exchange locations) the second closest location was used, although this issue only affected five cases in total. The distance to the closest exchange for each postcode centroid was then bound to the existing ONSPD dataset as a further attribute.

The next stage in the validation process was to merge the newly created ONSPD dataset (with distances to the closest exchange), with the speed test dataset that was provided by broadbandspeedchecker.co.uk. The datasets were merged by postcode, using the postcodes from the speed test dataset as the primary index. This ensured that information from the ONSPD was only joined to the speed test dataset where there was a matching postcode in each dataset.

At this stage it was necessary to set the geographic extents of the research, and because a number of indicators and spatial units that had been earmarked for analysis were attributed only to England, it was deemed that this be the study area to ensure the most robust analysis. It was also deemed necessary as the broadband access and performance statistics for England, Scotland, Wales and Northern Ireland appeared sufficiently differentiated at this point that independent studies would be advisable. Because the ONSPD had been reduced to only contain postcodes within England when the extents of the study were set, this process of joining also allowed any speed test results that did not have a direct match from the ONSPD dataset (because they originated from Scotland, Wales, Northern Ireland or an invalid postcode) to be identified and deleted. The total number of 'live' English postcodes in the ONSPD dataset used at the time of the analysis was 1,441,337. Of these, 328,883 were represented in the 2010/11 speed test dataset, representing a postcode-level coverage of 22.8%. However, data would not be analysed at postcode level as this geographic extent is too fine, instead, data would be aggregated to LSOA and district geographies. At these geographic extents there was coverage of 99.8% and 100% respectively (see Table 3.2).

Table 3.2: Data Summary: Speed Test Datasets

	2010/11 Dataset	2012/13 Dataset
Period covered	1/1/2010 to 31/1/2011	1/4/2012 to 31/5/2013
Initial observations	2,874,959	1,863,627
Results removed through validation	789,723	450,281
Results retained for analysis	2,085,236	1,413,346
Coverage at postcode level	22.82%	20.47%
Coverage at LSOA level	99.79%	99.63%
Coverage at district level	100%	100%

A second stage of data cleansing examined outliers based on the download results recorded in the speed test dataset. As this thesis is examining broadband access and performance, it was necessary to remove any results which fell below the minimum threshold that constitutes broadband Internet. The minimum threshold for broadband is, however, a grey area, with no formal definition.

As such, for the purpose of this analysis the minimum download threshold was set at 512Kbps (half a Megabit per second) as this was the speed of the first broadband package launched in the UK in August 2000.[4] Although many service providers now offer far higher speeds, for those who are extremely rurally isolated, speeds as low as 512Kbps may still be experienced. For the upper speed boundary, the threshold for was set at 102,400Kbps (100 Megabits per second). A small number of packages offered by Virgin Media were providing this level of service to domestic customers as of late 2010, although, achieving the theoretical maximum download speed of any package is unlikely due to signal attenuation. There were a small number of clear outliers recording download speeds of several tens of thousands of Megabits per second which were removed. In applying these thresholds for speed, it was anticipated the results retained would represent a mix of connection types; predominantly traditional ADSL connections, followed by a significant proportion of more recent FTTC and cable connections, and a very limited number of FTTP and satellite connections, thus reflecting the distribution of broadband technologies available in England. As well as identifying speed-based outliers, a distance cap of 9km was applied to the dataset, this removed a small number of results that were recording far higher distances to a local exchange than would be expected. The distance cap of 9km was imposed as connections over 5km from a local exchange are unlikely, but not impossible (Ofcom, 2013a). Because the number of results that recorded distances between 5km and 9km were extremely small (around 0.6% of all valid tests) it was decided that these would have little impact on the analysis, but were likely valid observations. Beyond 9km, a very small number of tests were recorded at sporadic distance measurements and were removed as outliers.

The formatting of the date/time field in the speed test dataset also presented some issues. The data contained within this field was formatted as YYYY:MM:DD:HH:MM:SS (year, month, day, hour, minute, second). This native format was non-standard and was not recognised during exploratory analysis, requiring further manipulation. By specifying a more commonly used date/time format using an R package it was possible to parse the results into a format that could be recognised. This reformatting opened up further opportunities for analysis, as functions to efficiently extract and modify components of date/time data (such as days and hours) were made accessible.

Figure 3.2 visualises the validated 2010/11 dataset of geo-located speed test results by user supplied postcode. Plotting the observations in this way shows the spatial distribution of the dataset, which logically follows that of population density and raises no concerns of an unrepresentative spatial sample.

The same validation process was applied to the 2012/13 speed test dataset with the appropriate ONSPD version updated to reflect the time period in which the speed test records had been collated.

[4] `http://www.thinkbroadband.com/news/54-freeserve-launch-usb-adsl-a-competition.html`

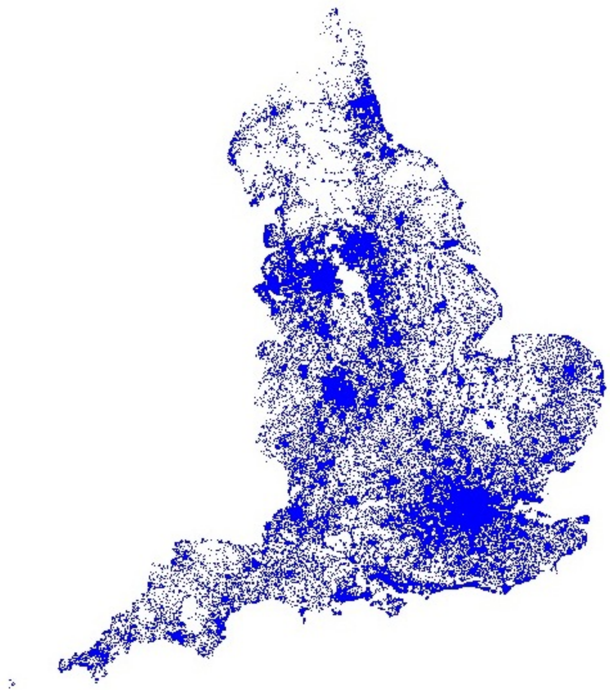


Figure 3.2: Speed Test Locations (Postcode Centroid) 2010/11 Dataset

Table 3.2 details the data summary for both datasets.

3.2.4 Coverage and Data Organisation

Coverage at LSOA level is high across both datasets. The average number of speed test observations per LSOA (after validation) was 64 in the 2010/11 dataset and 44 in the 2012/13 dataset. The raw number of observations collated by broadbandspeedchecker.co.uk are falling, both datasets provided for analysis cover roughly 13 months, however the latter dataset contained around 1 million fewer observations. This may be due in part to Internet speeds increasing over this time period, and users becoming less concerned about speed testing as aggregate connection quality improves. Keeping the two datasets separate in this way has a number of advantages. Firstly, two datasets enabled time series analysis to highlight changes between key indicators of access and performance. The second advantage is that aggregate statistics were not be skewed by merging two datasets covering different periods of time. The two datasets supplied have differences, which reflect changes in broadband infrastructure over recent years, combining the datasets would likely have influenced basic summary statistics. As the majority of government and industry reporting on broadband performance is undertaken on a yearly basis, retaining two annual extracts in this research ensures broad alignment.

3.3 Data Analysis

The following section presents analysis of a number of indicators of the socio-spatial structure linked with broadband access and performance. A number of themes were identified for analysis, including:

- Distribution of speed test results by distance from the closest exchange
- Exploring the relationship between distance and speed
- Indicators of socio-spatial structure and broadband performance
- Speed test locations as a proxy for access across different societal groups
- Broadband speed and indicators of rurality
- Broadband speed and indicators of deprivation
- The geography of English broadband performance by district
- Identifying spatio-temporal changes in broadband performance
- The geography of next generation infrastructure

3.3.1 The Effects of Exchange Proximity

The distribution of speed test results was first compared to the distance from their nearest telephone exchange, which is often used as a proxy for speed (Ofcom, 2012). This analysis showed that the vast majority of speed tests were conducted from postcodes within 2.5km of their nearest telephone exchange. The average distance of a speed test location to the nearest exchange was shown to be around 1.6km in both datasets. These results are to be expected given what is known about the technical limitations of Digital Subscriber Line technology, particularly that long distances result in high levels of signal attenuation and reductions in speed. As such, it would be unlikely to find broadband connections at distances over 5km (Ofcom, 2013a). As Figure 3.3 shows, the distribution of test results by distance is common across both datasets.

Figure 3.4 reports analysis of download speeds by distance from the nearest exchange and reveals that speed tests conducted in close proximity unsurprisingly returned the best download results, particularly those within 1km. Speeds start to deteriorate noticeably at distances of over 2km; this is also to be expected as the average distance from end user to exchange calculated in each dataset was around 1.6km. However, these analyses also highlighted an apparent anomaly. In both datasets, results appeared between 5000m and 7000m from the exchange with increases in mean download speeds; and may represent an increased percentage of speed tests run through fibre or coaxial-based connections. As discussed earlier, such infrastructure allows speeds of up to 76Mbps

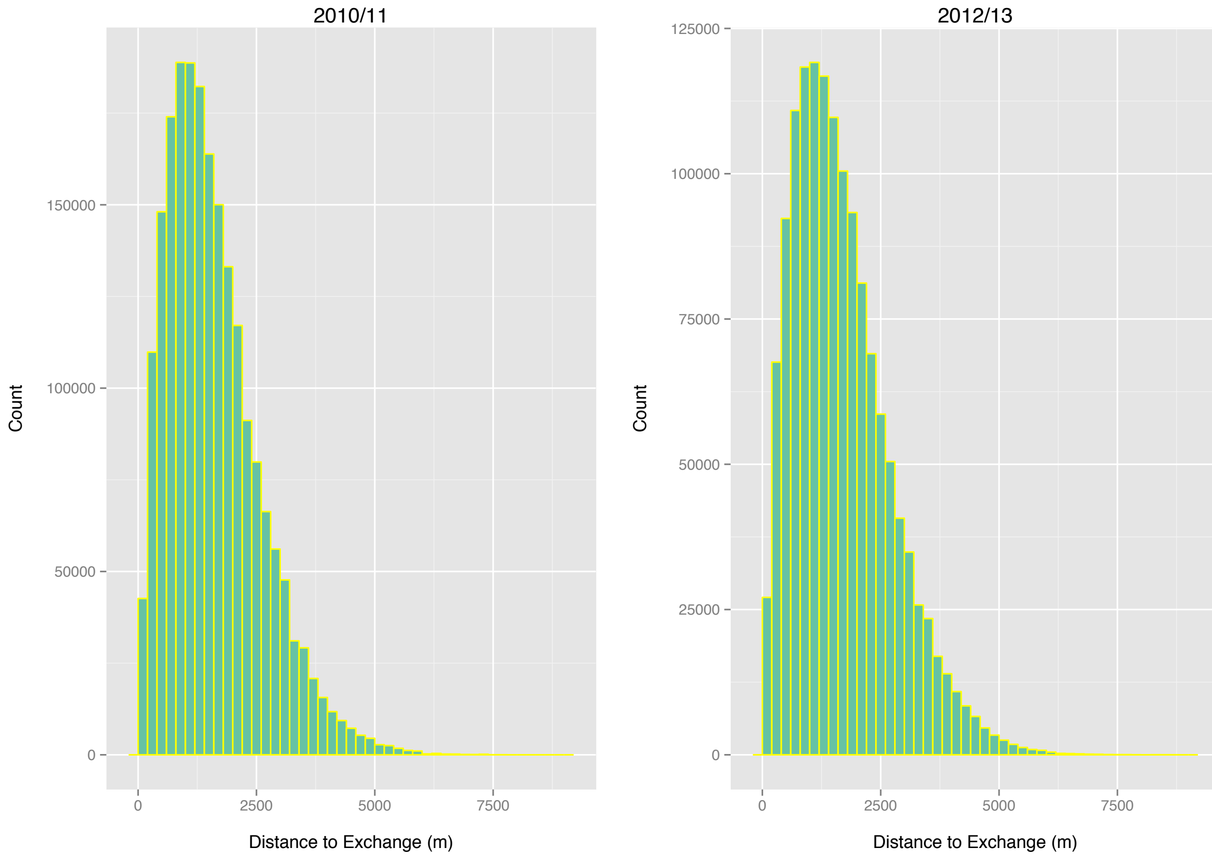


Figure 3.3: Data Distribution: Distance to Exchange

on fibre to the cabinet (FTTC) lines and was being deployed by BT, TalkTalk, BSkyB and a number of other retail providers at the time of data collection (Ofcom, 2012). However, within both datasets only a small percentage (0.5% and 0.6% respectively) of the speed test results were recorded at distances over 5000m from the closest exchange. Table 3.3 presents the raw number of tests recorded by 500m distance intervals for the 2012/13 dataset.

To further explore the geographic distribution of test results, the number of speed tests per head of population was calculated for each English local authority district. For this analysis, both datasets were combined in order to explore geographic bias within the data over the entire period of data collection. Figure 3.5 illustrates the output of this analysis. It is apparent that there are some districts, particularly those that are within predominantly rural areas such as Suffolk, the Lake District and the Cotswolds, which record a higher propensity for speed testing. This may be due to poorer performance within these areas in general due to infrastructure constraints, and a greater tendency for Internet users to seek methods of monitoring their connection performance.

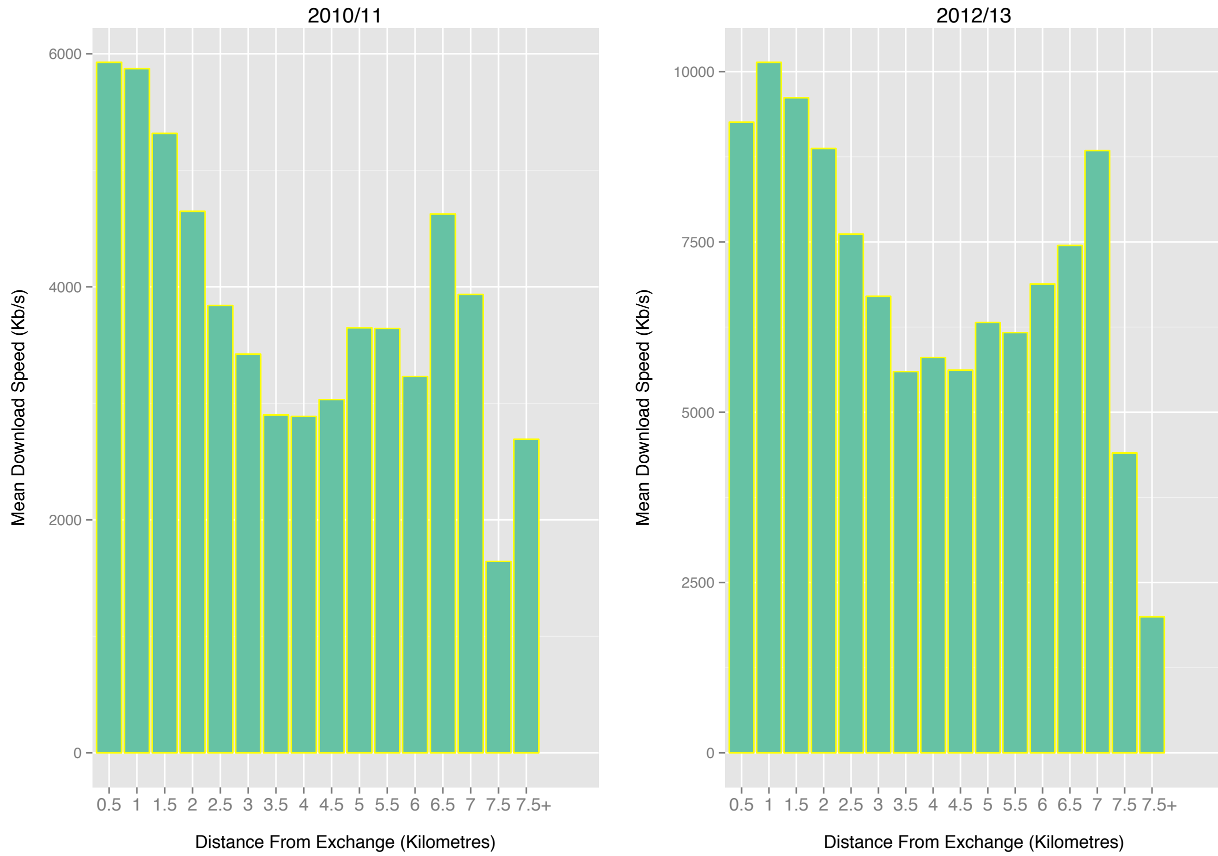


Figure 3.4: Mean Speed by Distance to Exchange

Conversely, predominantly urban centres such as London, Birmingham, Manchester and Liverpool record fewer speed tests per head of population. This may be due to higher average speeds in these areas as a result of better infrastructure. It should, however, be noted that multiple tests cannot be attributed to an individual user given that unit postcodes typically cover multiple addresses. User level attribution would only have been possible if a unique identifier (IP Address, for example) would have been recorded alongside each test result.

3.3.2 Speed and Indicators of Socio-spatial Structure

As discussed in the review of literature, performance of, and access to broadband services can be influenced by factors such as rurality and local socio-economic conditions (see Chapter 2, Section 2.4.1). As such, profiling the speed test datasets by a geodemographic Classification would reveal how areas with differing built environment and socio-economic conditions rank in terms of aggregate performance statistics. Analysis of download speeds by the 2011 Output Area Classification

Table 3.3: Number of Connections Recorded by Distance to the Nearest Exchange 2012/13

Distance from Exchange (m)	Test Count
0 - 500m	136697
500 - 1000m	279502
1000 - 1500m	292328
1500 - 2000m	247114
2000 - 2500m	181033
2500 - 3000m	119005
3000 - 3500m	73380
3500 - 4000m	41663
4000 - 4500m	22775
4500 - 5000m	11150
5000 - 5500m	4812
5500 - 6000m	2458
6000 - 6500m	820
6500 - 7000m	371
7000 - 7500m	150
Over 7500m	56

(OAC) was undertaken to establish how aggregate broadband performance was differentiated between geodemographic clusters. OAC is a geodemographic classification derived using census data that groups together geographic areas according to key characteristics of the population within them. OAC has three tiers of granularity; Supergroups (most coarse tier), Groups (mid-tier) and Subgroups (most granular tier). Table 3.4 shows the mean download speeds by OAC group for both datasets, as well as the percentage increase over the time period 2010/11 to 2012/13. As might be expected, the slowest average download speeds are recorded in groups that represent predominantly rural areas, including 1a: Faming communities, 1b: Rural tenants and 1c: Ageing rural dwellers. Given such low aggregate speeds, it is likely that a significant number of connections may be falling below the Universal Service Commitment (USC) threshold of 2Mbps. As might be expected, the highest average download speeds were recorded in predominantly urban and suburban groups, likely due to better infrastructure provision and higher demand justifying roll-out of next generation services.

Between the two time periods, similar patterns emerge, with more densely populated urban and suburban areas generally recording the largest increases in speed. The largest increases in speed were observed in OAC groups 5a: Urban professionals and families (114.7%), 4a: Rented family living (107.8%), 8d: Migration and churn (102.9%) and 6a: Suburban achievers (102.3%). These

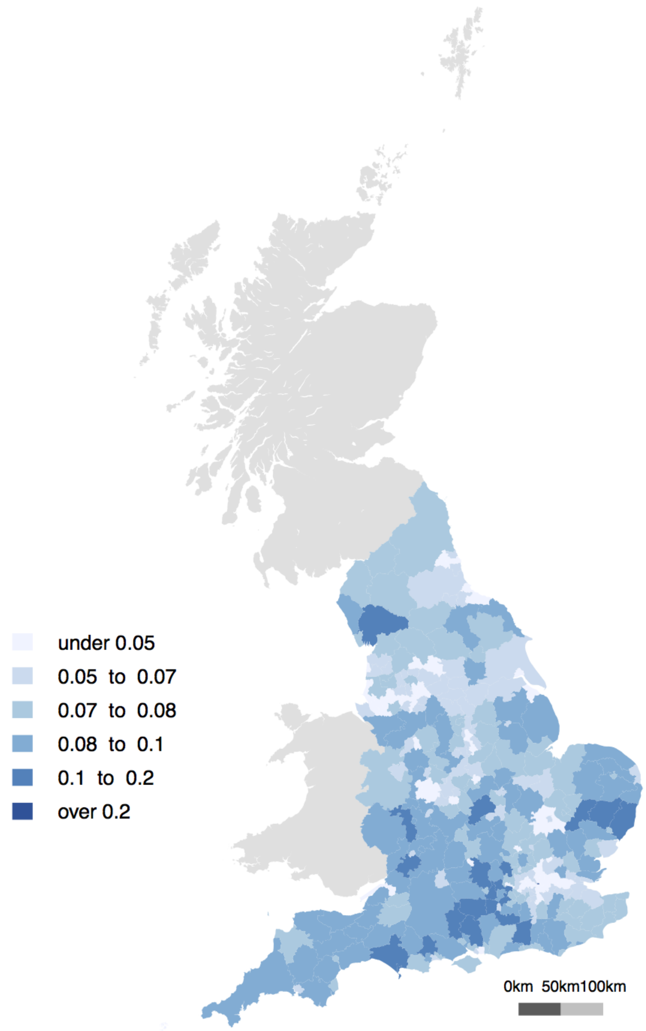


Figure 3.5: Speed Tests Per Head of Population: English Districts

figures suggest that infrastructure investments which enable higher speeds may have been rolled out to these groups before those which are more rurally isolated, for example groups 1a - 1c, which record far lower aggregate increases in performance over the same period. Interestingly, those groups that are predominantly located within inner city areas, or ranked amongst the highest aggregate download performers in the 2010/11 dataset, generally record lower increases in aggregate performance over the time period. This is likely due to high-speed infrastructure already being in place within these areas, potentially combined with slower adoption rates of superfast services through optical fibre or coaxial cable. Given the traditional limitations of distance and signal attenuation have a lesser affect on performance in dense urban settings (Ofcom, 2010), traditional ADSL broadband connections in these areas will often perform above the national average. This may reduce the propensity for a consumer to upgrade their connection given a lack of necessity.

Table 3.4: Mean Download Speeds by OAC Sub-group

OAC Group	Mean Download 2010/11 (Kbps)	Mean Download 2012/13 (Kbps)	% Increase
1a: Farming communities	2690	3688	37.1
1b: Rural tenants	3060	4523	47.8
1c: Ageing rural dwellers	3035	4497	48.2
2a: Students around campus	6448	10788	67.3
2b: Inner city students	5204	8047	54.6
2c: Comfortable cosmopolitan	5788	9808	69.5
2d: Aspiring and affluent	5730	10096	76.2
3a: Ethnic family life	5858	10138	73.1
3b: Endeavouring ethnic mix	5495	10133	84.4
3c: Ethnic dynamics	7090	11018	55.4
3d: Aspirational techies	5907	10804	82.9
4a: Rented family living	5604	11644	107.8
4b: Challenged Asian terraces	6089	10648	74.9
4c: Asian traits	5536	10674	92.8
5a: Urban professionals and families	4788	10280	114.7
5b: Ageing urban living	4923	9189	86.7
6a: Suburban achievers	4306	8711	102.3
6b: Semi-detached suburbia	4536	9068	99.9
7a: Challenged diversity	6018	10346	71.9
7b: Constrained flat dwellers	6338	10902	72.0
7c: White communities	6178	10068	63.0
7d: Ageing city dwellers	5798	10192	75.8
8a: Industrious communities	4815	8815	83.1
8b: Challenged terraced workers	5636	9933	76.2
8c: Hard pressed ageing workers	4769	8574	79.8
8d: Migration and churn	5637	11436	102.9

In addition to analysing download speeds by OAC group, further analysis was undertaken to establish, to what extent, each OAC group was represented in the datasets. In doing so, it would be possible to assess the propensity for broadband speed testing across geodemographic clusters. In order to perform the analysis, the number of current postcodes within each OAC group were calculated from the ONSPD, this was then compared to the known number of speed test results within each OAC group to produce the average number of speed test results per postcode within each OAC cluster, i.e.

$$\textit{Test quotient} = \textit{Cluster test count} / \textit{ Valid postcodes in cluster}$$

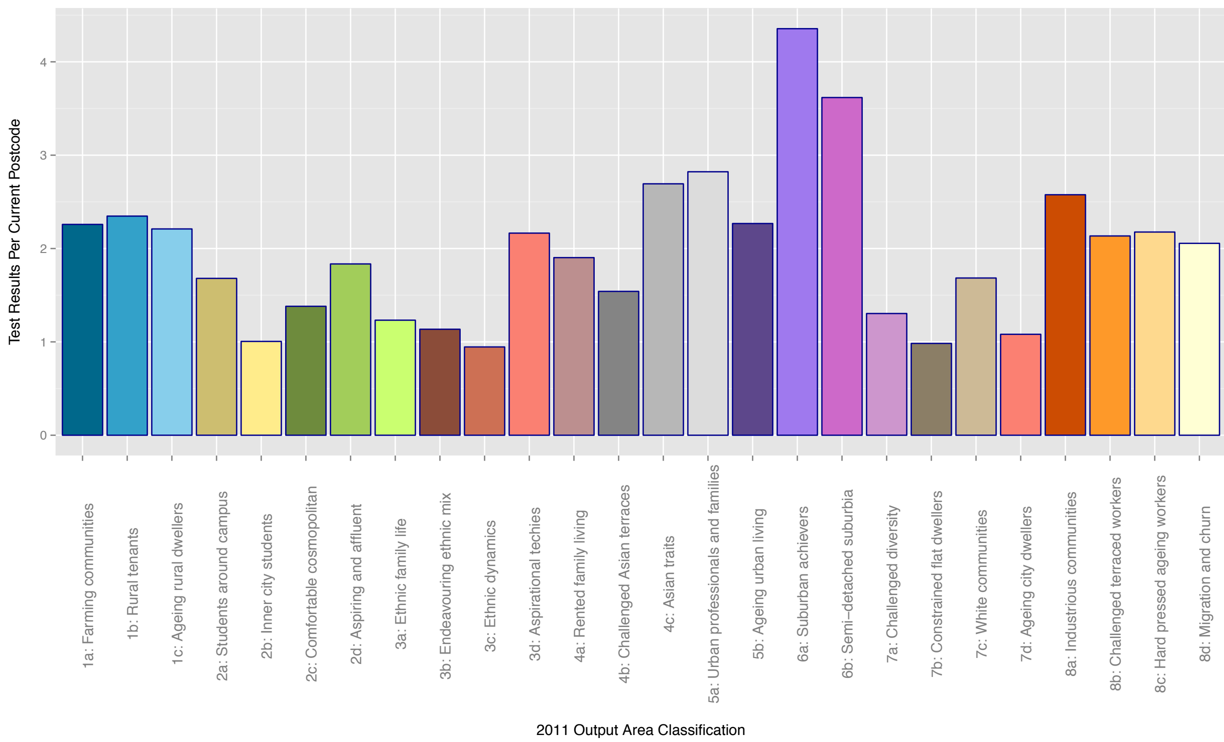


Figure 3.6: Test Rate (Tests per Postcode) by Output Area Classification (OAC)

Figure 3.6 shows the output of this analysis. In this instance the two speed test databases were combined. It is apparent that speed tests were observed variably between OAC groups. Group 6a: Suburban achievers record the highest propensity for speed testing at over 4 test results per current postcode within this group. This is closely followed by groups 6b: Semi-detached suburbia and 5a: Urban professionals and families. As would be expected, groups that represent areas with older residents, who are frequently reported to be less engaged with communications technology such as broadband Internet (see Chapter 2, Section 2.4), are under-represented in the datasets (Group 7d: Ageing city dwellers). Groups that are prominent in deprived areas such as 7b: Constrained flat dwellers, also display a lower propensity for speed testing, perhaps where higher levels of material deprivation have a limiting impact on Internet access. Interestingly, rural OAC groups such as

1b: Rural tenants and 1a: Farming communities show a higher than average propensity for speed testing, possibly due to constrained connections in these areas and an increased motivation to monitor performance. The varying representation of different OAC groups within the datasets broadly supports the notion of digital differentiation across societal groups.

To examine the relationship between speed and rurality more explicitly, the Office for National Statistics (ONS) Urban/Rural classification was appended to each postcode in both datasets. These definitions were created at a Census Output Area (OA) level of geography for England and Wales, and classified areas as 'urban' if the majority of the population of an output area lived within settlements with a population of 10,000 or more. In addition, the classification also categorises output areas based on context; such as whether the wider surrounding area of a given output area is sparsely populated or less sparsely populated (Bibby and Shepherd, 2004).
Table 3.5 shows the average download speed recorded by the Urban/Rural classification alongside the average distance to an exchange for both datasets. Unsurprisingly, results mirror those seen when profiling by OAC, whereby the highest average download speeds occur in highly urbanised areas. The highest average speeds are recorded in those areas classified as 'Urban >= 10K - Less Sparse', essentially areas which are urban and have more densely populated surroundings. Unsurprisingly, the slowest average speeds are recorded in areas classified as 'Hamlets and isolated dwellings' (both 'Sparse' and 'Less Sparse') and villages (again, both 'Sparse' and 'Less sparse'). This is not a surprising distribution when line length is taken into account, given that longer lines are subject to greater signal attenuation, and as such, likely to deliver slower average speeds.

Table 3.5: Mean Download Speeds by Urban/Rural Indicator

Urban/Rural Indicator	Mean Download 2010/11 (Kbps)	Mean Download 2012/13 (Kbps)	% Increase	Mean Distance to Exchange (m)
Urban >= 10K Sparse	4241.23	6650.78	56.81	971
Town and Fringe Sparse	4023.32	6248.33	55.30	736
Village Sparse	2899.43	3380.41	16.59	2272
Hamlet and Isolated Dwelling Sparse	2521.54	3151.11	24.97	2790
Urban >= 10K Less Sparse	5423.51	10176.39	87.63	1496
Town and Fringe Less Sparse	3839.09	6951.65	81.08	1575
Village Less Sparse	2694.63	4042.23	50.01	2375
Hamlet and Isolated Dwelling Less Sparse	2745.07	3895.95	41.93	2477

Between the two time periods it is also apparent that urban and town and fringe areas have seen the largest increases in average download speeds in the period 2010/11 to 2012/13. Areas classified as 'Urban >= 10K - Less Sparse' record the highest speed increases at over 87%, followed by 'Town and Fringe - Less Sparse' at 81%. Rural areas, particularly those that are sparsely populated, record much lower increases in speed, areas classified as 'Hamlet and Isolated Dwelling Sparse' and 'Village Sparse' record speed increases of around 25% and 17% respectively. As a current snapshot of broadband speeds in England, this analysis would suggest that disparities in service exist between urban and rural areas, both in terms of average download speeds and speed increases through investment in infrastructure. Mean distance to exchange (calculated here through combined datasets) in the 'Urban >= 10K', 'Town and Fringe' and 'Village' categories are shorter for the 'Sparse' context than 'Less Sparse'. Initially, this looks anomalous, however, OAs surrounded by less sparsely populated areas would logically be closer to their local exchange, as that exchange would typically be located towards the centre of a populated area in order to supply the best service. In OAs where the surrounding area is classed as 'Less Sparse' (i.e. of higher population density) the logical placement of the exchange may not be as obvious due to land constraints or uneven urban expansion. Therefore, it follows that there could be far more OAs at greater distance from their local exchange than those which happen to be closer, thus increasing the average distance.

Although rurality is an important differentiating factor of broadband speed in aggregate, these patterns interact with other factors. As has been shown elsewhere (Longley and Singleton, 2009b), differences in use and engagement with the Internet occur between and at the intersection of patterns of material deprivation. However, to date, equity in broadband performance has not been explored within this context. To examine the relationship between prevailing levels of material deprivation and broadband speed, the Indices of Multiple Deprivation 2010 (IMD) were converted into ranked deciles and appended to each of the test results.

The English Indices of Deprivation 2010 provide a relative measure of deprivation at Lower Super Output Area (LSOA) level across England. LSOAs represent small area statistical boundaries that capture between 400 and 1,200 households with area populations of between 1,000 and 3,000 persons. Within the IMD, areas are ranked from least deprived to most deprived on seven different dimensions of deprivation and an overall composite measure of multiple deprivation. The seven dimensions include:

- Income deprivation
- Employment deprivation
- Health deprivation and disability
- Education deprivation

- Crime deprivation
- Barriers to housing and services deprivation
- Living environment deprivation

The composite measure, which is a rank of each LSOA in the country, ranging from 1 (most deprived) to 32,482 (least deprived) was selected for this analysis.

Table 3.6: Mean Download Speeds by IMD Decile

IMD Decile	Mean Download 2010/11 (Kbps)	Mean Download 2012/13 (Kbps)	% Increase
Decile 1	6539.38	10356.14	58.37
Decile 2	6163.30	9730.00	57.87
Decile 3	5890.13	8821.72	49.77
Decile 4	5638.22	8072.95	43.18
Decile 5	5269.23	7674.97	45.66
Decile 6	4960.79	7386.53	48.90
Decile 7	4765.86	8067.41	69.28
Decile 8	4731.30	8009.22	69.28
Decile 9	4664.43	8778.90	88.21
Decile 10	4832.39	9325.59	92.98

Average speeds were calculated for each of the IMD deciles and are shown in Table 3.6. These analyses show that average download speeds are higher in those areas ranking as more deprived (particularly so for deciles 1 and 2). This is due to such areas of high deprivation typically being located within more densely populated urban conurbations, and as such, benefiting from more developed network infrastructure that is necessary to support higher speeds. Calculating increases in average download speeds between the 2010/11 and 2012/13 datasets reveals that the most affluent areas have seen the largest increases, particularly deciles 9 and 10, with increases of around 88% and 93% respectively. This would suggest that outside of densely populated urban areas (which generally receive infrastructure investments first), affluent areas are likely being targeted for faster services. These results would suggest that prevailing levels of deprivation in England are not necessarily a barrier to broadband access or to higher speed broadband connections. However, one caveat of this analysis is that although speed tests (based on the user supplied postcode) were relatively evenly distributed across all IMD deciles, the data contains no information regarding the levels of material deprivation that are experienced by each individual test user, only that at the aggregate level of the LSOA.

3.3.3 Quantifying Spatial Disparities in Broadband Performance

The spatial disparities in broadband speed across England are presented through three analyses. The first visualises disparities in average download speeds at district level. The second explores these changes at district level between the 2010/11 and 2012/13 datasets in order to highlight those areas that are likely to be receiving investments in infrastructure. The third provides an overview of superfast service penetration (infrastructure capable of delivering speeds of 24Mbps+), arguing that there are significant disparities in the delivery of these next-generation network upgrades across England.

Disparities in broadband speed across England are visualised in Figure 3.7 by mapping the average download speeds of both datasets by English local authority district. Results for 2010/11 are presented in the map on the left and results for 2012/13 on the right.

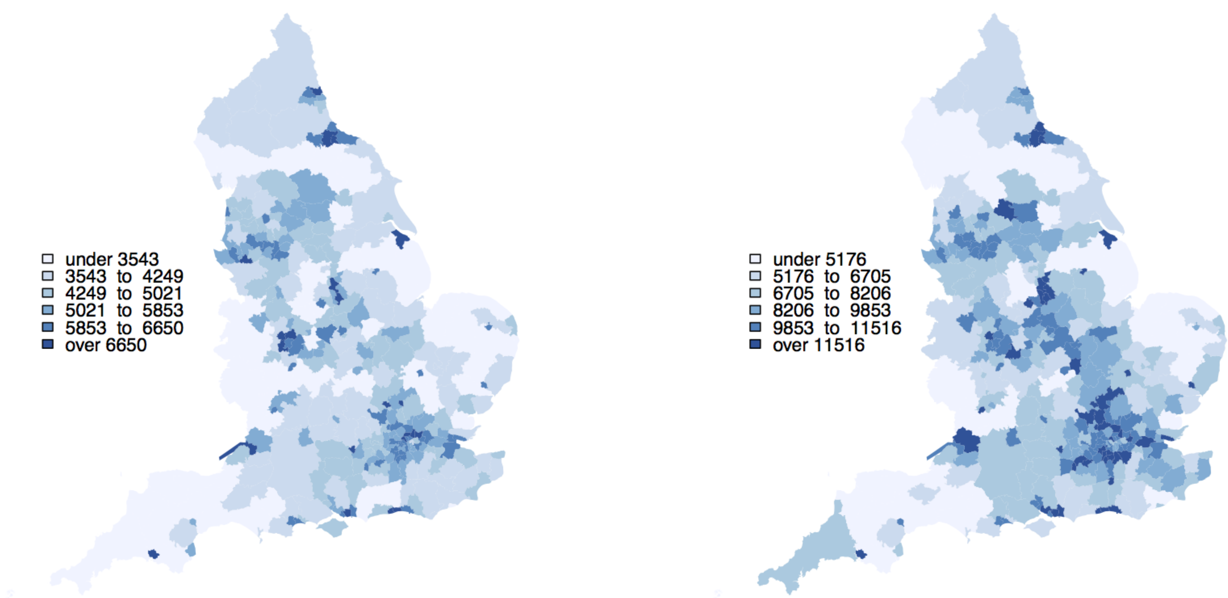


Figure 3.7: Mean Download Speed (Kbps) by English District 2010/11 and 2012/13

There is significant clustering of higher average download speeds around major urban centres such as London, Birmingham, Manchester and Liverpool. As might be expected, predominantly rural regions such as the South West, Norfolk, Yorkshire and Lincolnshire record the majority of the slowest download speeds. On the basis of the averages within the datasets, it was estimated that a large proportion of households, particularly those in rural areas, at the time the data were collected, would be falling below the UK government's Universal Service Commitment threshold of 2Mbps. Of the validated speed test results used for analysis, in the 2010/11 dataset 30.9% fell below the USC threshold, this figure fell to 21.8% in the 2012/13 dataset. However, if these changes are mapped for local authority district geography, there are apparent geographic disparities to these improvements (See Figure 3.8 and Tables 3.7 and 3.8).

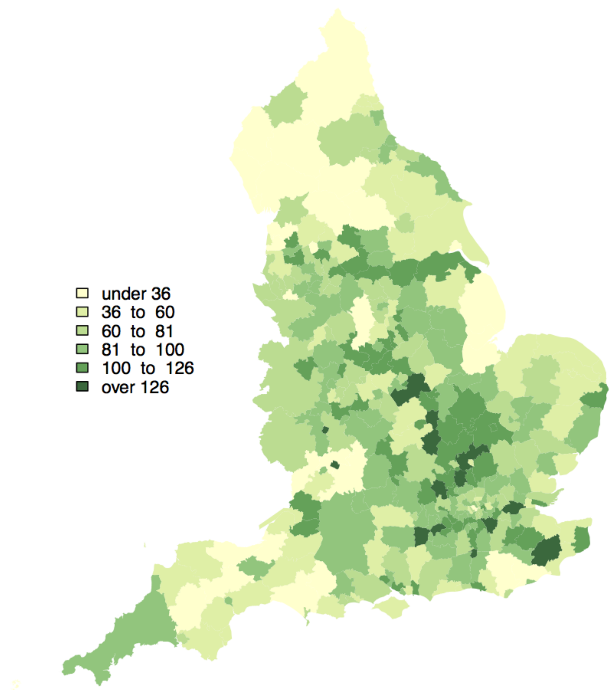


Figure 3.8: Percent Increase in Mean Download Speed by English District 2010/11 to 2012/13

Unlike the previously presented maps of average download speeds (See Figure 3.7), change is less obviously clustered around urban conurbations, and instead, high increases in download speed are more widely dispersed. The lowest increases appear predominantly in rural areas, however, some rural areas such as Norfolk and Cornwall have also seen relatively high speed increases over the time period. It would be safe to assume that large urban conurbations have not seen large percentage increases in average download speeds as the infrastructure necessary to support high speeds has typically long been in place within these areas. The largest increases in average download speed are recorded in predominantly urban areas, but those which would not be considered as major urban centres. In most cases, these represent areas that may not previously have been prioritised for next generation infrastructure upgrades replacing elements of backhaul cabling with fibre. More densely populated urban centres would logically offer the best return on investment for such enhancements given shorter runs of new cabling and the ability to upgrade services for comparitively large numbers of consumers.

Table 3.7: Largest Ten Increases in Mean Download Speeds by English District 2010/11 to 2012/13

District	Mean d/load 2010/11 (Kbps)	Mean d/load 2012/13 (Kbps)	% Increase
Harborough	3763	10366	175
Ashford	3725	9363	151
Chiltern	4070	10193	150
Wellingborough	4635	11601	150
North Hertfordshire	5302	13112	147
Worcester	3135	7638	144
Surrey Heath	5265	12600	139
Bromley	5471	12915	136
Thurrock	5114	12013	135
Milton Keynes	3474	8128	134

Table 3.8: Smallest Ten Increases in Mean Download Speeds by English District 2010/11 to 2012/13

District	Mean d/load 2010/11 (Kbps)	Mean d/load 2012/13 (Kbps)	% Increase
Copeland	3467	3491	0.67
Allerdale	3634	3726	2.53
Craven	4275	4470	4.56
Mid Devon	3248	3667	12.92
Barrow-in-Furness	3702	4226	14.17
West Somerset	3149	3711	17.83
Richmondshire	3332	3932	18.01
Boston	3279	3870	18.02
Cotswold	3599	4279	18.90
Isles of Scilly	2558	3044	18.99

As would be expected, the smallest increases in average download speeds were recorded in those districts that are predominantly rural. Districts such as Copeland, Boston and the Isles of Scilly, in particular, are rurally isolated and sparsely populated which constrains the viability of infrastructure upgrades. Areas such as these would likely require government intervention to stimulate the required investment in infrastructure. One such initiative in the UK has been the Rural Community Broadband Fund (RCBF[5]), jointly funded by Defra and Broadband Delivery UK (BDUK[6]), which is aiming to deliver improvements in broadband infrastructure to the most rurally isolated areas.

3.3.4 The Geography of Next Generation Infrastructure

Supplying broadband over fibre optic or high bandwidth coaxial connections has the greatest impact on performance, particularly so for those consumers who are rurally isolated, as fibre optic connections are less susceptible to signal attenuation and can therefore supply service over much longer distances. As such, the government has indicated fibre deployment will play a significant role in supplying broadband to the most hard to reach areas (DCMS, 2011; Yiu and Fink, 2012) using funds such as RCBF and BDUK to enable delivery. Because such government initiatives were in progress during the period of data collection for this study, it was possible to expand the analysis to focus on the spatial disparities of those connections considered to be 'superfast'. Speed tests were limited so that only those connections above the threshold for superfast (>= 24Mbps) were present in the datasets. The locations of these connections were then plotted to represent where superfast access has the highest penetration within England. Applying this visualisation technique across both datasets (Figure 3.9) reveals how superfast broadband availability has developed between the two time periods.

Results show clearly that the majority of superfast connections in 2010/11 were recorded around large urban centres; London, Birmingham, Manchester and Liverpool are all easily identifiable in the visualisation. There are far fewer superfast connections recorded in rural areas such as Norfolk, Cumbria and Cornwall. Visualisation of the 2012/13 dataset reveals how these superfast connections have spread over a relatively short period of time. Major urban centres remain clearly visible but some rurally isolated areas record higher numbers of superfast connections, in particular, Cornwall, which has become more clearly visible in the latter dataset.

The relationship between infrastructure investment over these two time periods and broadband speeds can be illustrated further by exploring the geography of fibre enabled local exchanges. For a consumer to access a fibre enhanced broadband service, the local telephone exchange serving a property would require this provision to be enabled. In order to investigate the geography of next generation network upgrades, a comprehensive list of exchanges that were supplying FTTC

[5] `www.gov.uk/government/statistics/rural-community-broadband-fund-rcbf-output-indicator`

[6] `www.gov.uk/guidance/broadband-delivery-uk`

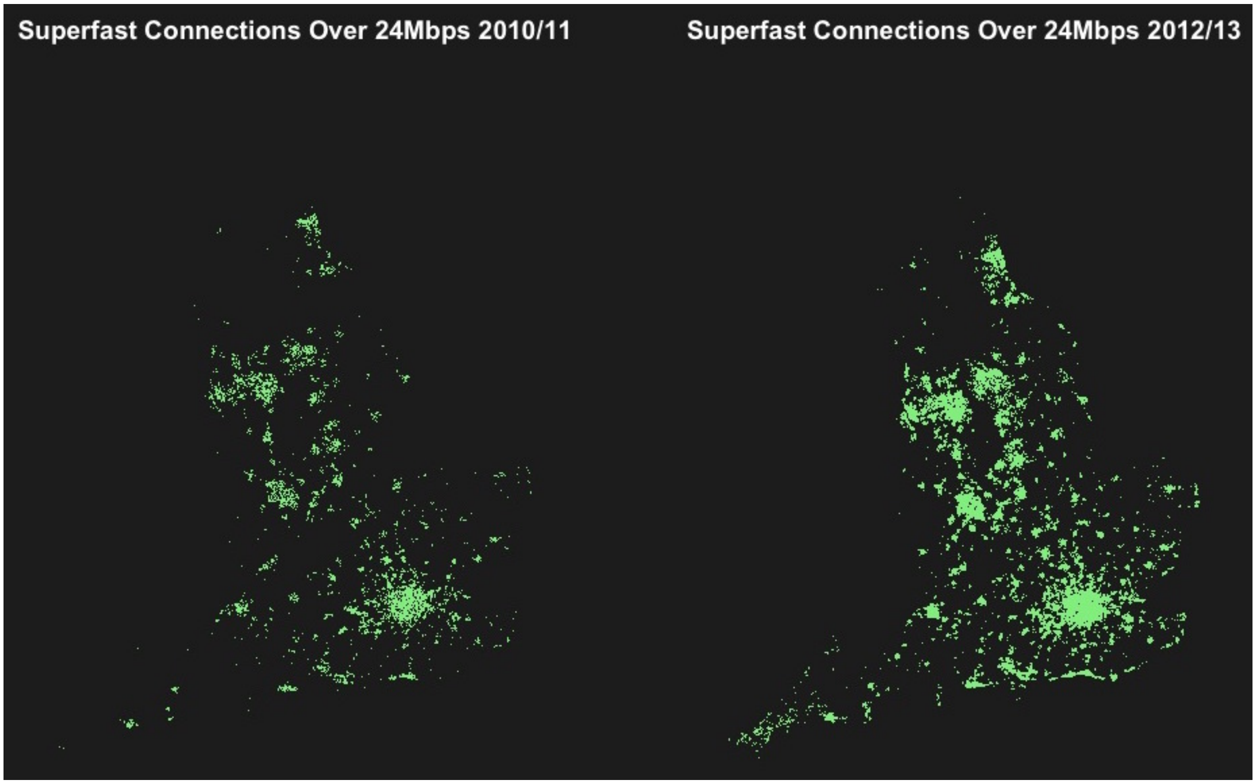


Figure 3.9: Change in Superfast Deployment 2010/11 to 2012/13

connections was obtained from BT. The dataset contained the name and a code of each exchange in the UK that was supplying BT's 'Infinity' fibre services. These data were geo-tagged by matching attribute information from the original exchange dataset obtained from SamKnows. The percentage of fibre enabled exchanges within each district was then calculated relative to the total number of exchanges in that district. Figure 3.10 shows this analysis for the national extent.

Broadly speaking, the geography of fibre enabled exchanges matches that of broadband speeds, with more FTTC enabled exchanges in predominantly urban areas. London and Manchester in particular have high percentages of FTTC enabled exchanges; however, it is apparent that this is also the case in some predominantly rural areas such as Cornwall. Cornwall has been promoting the expansion of fibreoptic broadband through the Superfast Cornwall Program, a £132m partnership funded by the European Regional Development Fund Convergence Programme, BT and Cornwall Council. This has aimed to upgrade existing broadband infrastructure and enable fibre access to 95% of homes and businesses in Cornwall by 2015. Many other rural areas such as North Yorkshire and Cumbria also have programs in place that make use of BDUK funds to supply superfast broadband to the most hard to reach areas. It is encouraging to see that such nuances have been identified

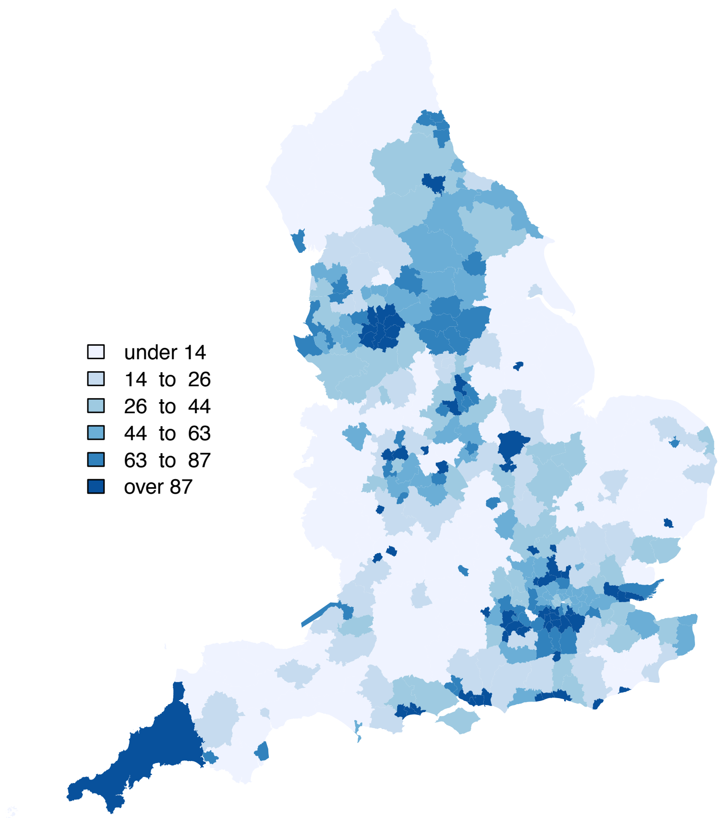


Figure 3.10: Percentage of FTTC Enabled Exchanges by English District (October 2013)

through data analysis, broadly validating the robustness of the data sources collated for use in this research.

3.3.5 Exploring Temporal Variations in Performance

In addition to examining spatial disparities in performance, the datasets were analysed for temporal fluctuations in speed, as there was no reason to assume that aggregate speeds were static throughout the day. Local factors such as changes in resident population density associated with commuting patterns were likely to impact performance and assist in identifying times of peak usage. In order to explore temporal variations in broadband performance, both datasets of speed test results were first analysed by the hour of the day in which each test was conducted. This analysis was conducted on a national scale and mean download speeds were calculated across both datasets (see Figure 3.11).

There are clear variations in average download speeds throughout the day, with the highest average

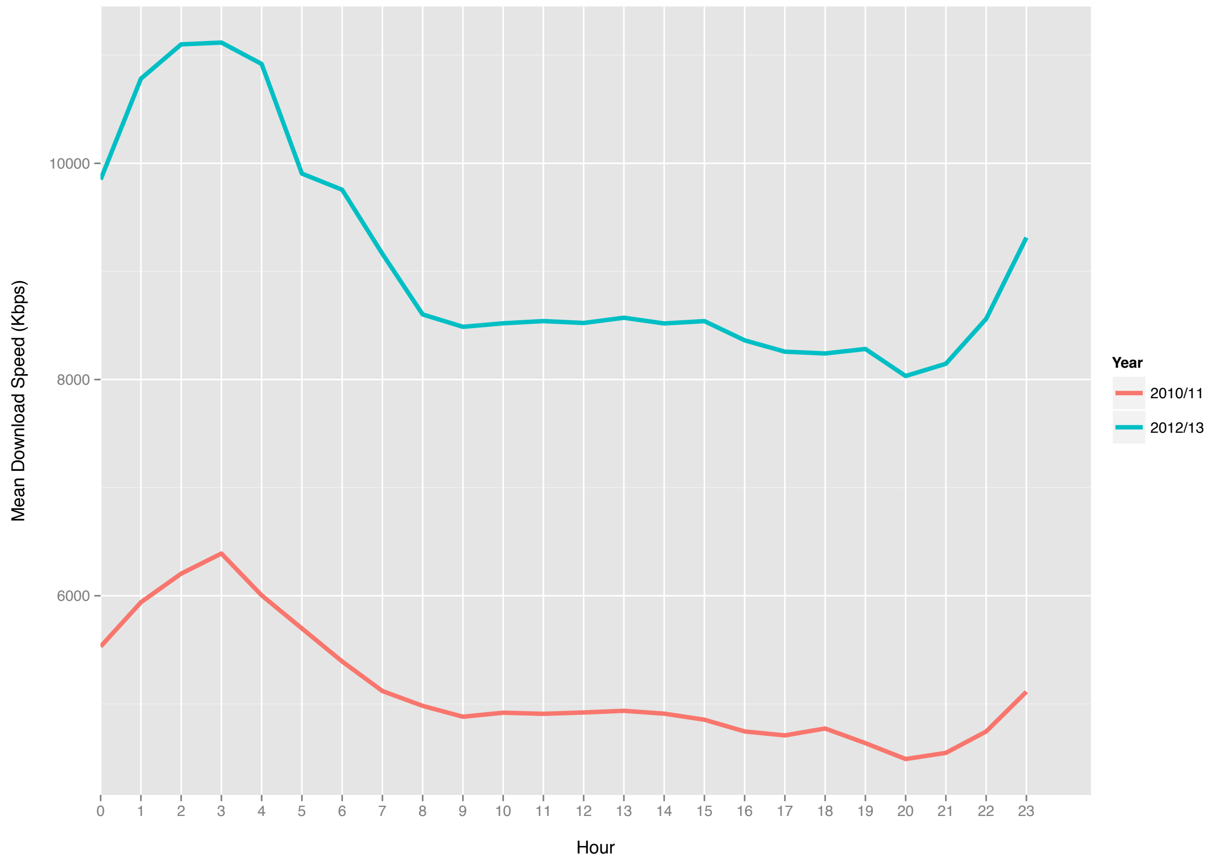


Figure 3.11: Hourly Fluctuations in Mean Download Speed

speeds appearing at 3am in both datasets. This is likely due to fewer users being online at this time of the day, which generally allows for faster download speeds and less 'bottlenecking' (a term used to refer to the slowing of a network in times of high use) of data traffic. Conversely, the slowest average speeds are recorded at 8pm in both datasets, this is widely regarded as peak time for Internet use, and as such, aggregate speeds tend to slow. There is significant slowing of average speeds after the 3am peak that is apparent in both datasets. As more users come online, average download speeds fall before stabilising at around 9am. Between the hours of 9am and 6pm (working hours), speeds remain fairly consistent before falling again around peak evening hours. In both datasets, speeds begin to increase again after 8pm as Internet traffic likely decreases. Although average speeds are significantly higher in the 2012/13 dataset, similar patterns emerge in terms of speed fluctuations. As well as analysing fluctuations in download speed throughout the day at aggregate national scale, analysis was modified to investigate these patterns when measures of local context are introduced. To achieve this, both datasets were combined to limit the effects of low counts and analysis was repeated, grouping speed tests by the relevant urban/ rural index class of the postcode from which they originated. This analysis was designed to highlight how speed fluctuates daily in different types of urban and rural conurbation. The output of this analysis is presented in Figure 3.12.

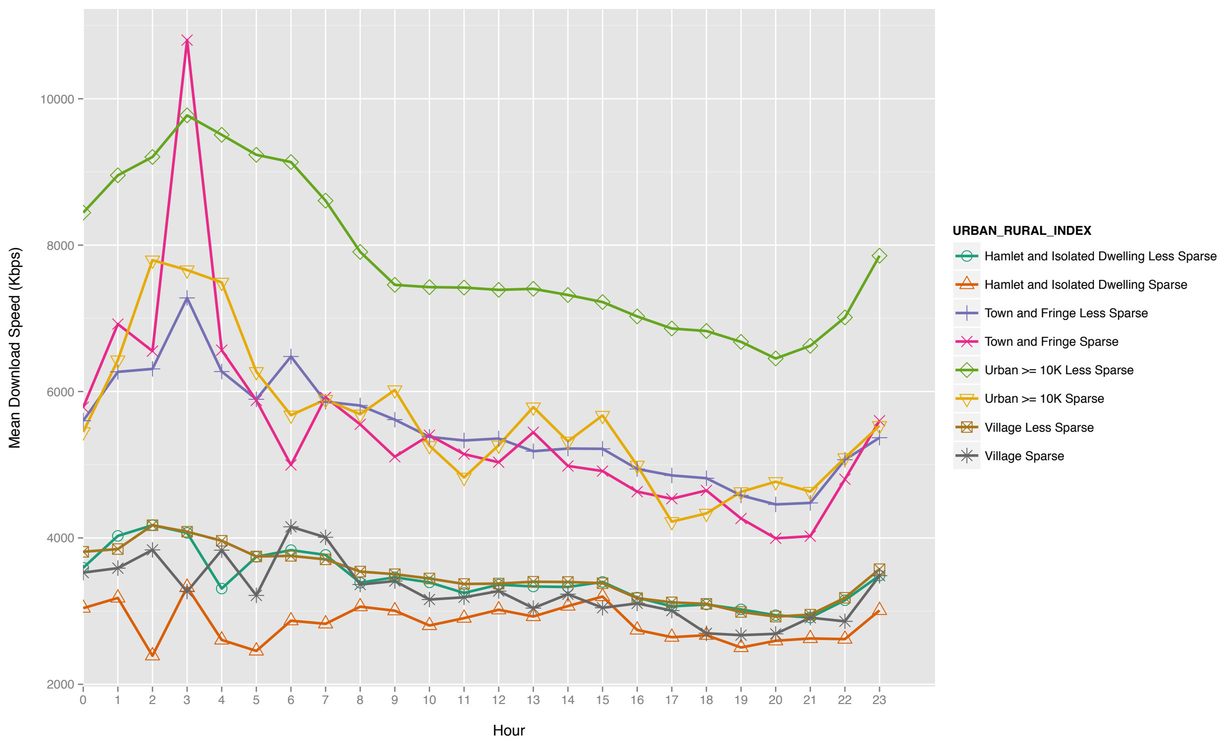


Figure 3.12: Hourly Fluctuations in Mean Download Speed by Urban/ Rural Index

Densely populated urban areas classified as 'Urban >= 10K Less Sparse' record the highest average download speeds throughout the day, with 'Town and Fringe' and sparsely populated urban areas ranking mid-table. 'Villages' and 'Hamlets and Isolated Dwellings', both sparse and less sparse record the lowest average download speeds throughout the day. In terms of performance variation, there appear to be larger fluctuations in urban areas that are likely due to higher population densities, and more prevalent bottlenecking of data traffic when there are a large number of users connected to the Internet. Within predominantly rural areas, fluctuations in download speed appear to be smaller, although speeds are generally much slower throughout the day. Despite delivering slower speeds, infrastructure in rural areas appears to be less susceptible to bottlenecking, and is most likely due to their composite smaller populations, and as such, lower demand. In all Urban/Rural categories the effects of peak and off peak hours are apparent, with average speeds spiking in the early hours of the morning when there are fewer users online, allowing for faster download speeds. Although we would expect these increases to be measurable, the large spike at 3am in the 'Town and Fringe Sparse' class represents noise in this particular subset of the data. Despite over 3 million observations used in this analysis, sub-setting the data into 192 groups (8 Urban/ Rural classes x 24 hours of the day) resulted in a small number of low group counts. Conversely, the slowest average speeds are recorded between 6pm and 9pm in most areas, which is likely a peak time for Internet use, and as such, aggregate speeds tend to slow. As well as detailing performance fluctuations, this analysis again highlights the large disparities that exist between urban and rural areas in terms of aggregate broadband speeds.

3.4 A Case Study of Local Needs and Private Infrastructure Overhaul, Kings Cliffe, Northamptonshire.

Despite increased rollout of high-speed broadband services (as discussed in this chapter and the review of literature (see Sections 2.1.4 - 2.1.5)), there are areas that are still suffering from inadequate legacy infrastructure, mostly due to rurality and limited potential uptake. Such factors mean that the supply of next-generation service to these areas is viewed as a costly option to incumbent suppliers and the limited return on investment means it is unlikely that existing infrastructure would be updated. This has led to some communities in need of enhanced service seeking out private investment. One such example is the village of Kings Cliffe, which is situated around 9 miles northeast of Corby in East Northamptonshire, with a population of 1,200 people. Kings Cliffe represents a strong case study for alternative infrastructure investment as in August 2015 a dedicated, privately funded fibre optic broadband network began serving residents in the village with high-speed Internet service. On the 16th September 2015 I met with Charles Tomalin, representative of the Kings Cliffe Fibre Broadband Group, to discuss the implementation. The following sections are adapted from the interview and present a snapshot view of how private infrastructure overhaul can be achieved in the absence of a willing incumbent supplier.

3.4.1 Background and Funding

The project was initiated by a private firm called Gigaclear [7] who specialise in the delivery of high-speed fibre optic networks to provide broadband services to rurally isolated areas. In the instance of the Kings Cliffe project, Gigaclear approached local residents to gauge interest having completed similar projects in rurally isolated areas in Oxfordshire. Gigaclear had acquired another business,

[7] `www.gigaclear.com`

Rutland Telecom before approaching residents in Kings Cliffe as part of an expansion to the business. Rutland Telecom had completed a similar project delivering fibre in nearby Lyddington in the district of Rutland, East Midlands, so it is understood the opportunity to deploy a similar scheme in Kings Cliffe would be financially viable given the presence in the area and support from residents.

The service in Lyddington supplied by Rutland Telecom went live in March 2010 and was hailed as a much-needed investment to future proof broadband service in the area. The project was initially a test for rollout of high-speed services in rurally isolated villages and required a minimum threshold of 40 resident registrations to commence. After initial consultations in June 2008, the number of registrations had exceeded the target by February 2009. Some private investors were also able to back the project and £37,000 was raised to initiate roll out, these investors are now receiving a 10% return as the service is active. The service delivery was realised by deploying FTTC connectivity to cabinets in the village, this method has been discussed earlier in this chapter and is widely viewed as the best means through which to supply fibre connectivity to consumers who are located at long distances from a local telephone exchange. Cabinets were installed at locations throughout Lyddington next to those of the incumbent supplier (BT) to terminate the backhaul fibre links to the closest exchange in Uppingham (around 3 miles (4.8 km) north of Lyddington), the existing copper connections to residents homes provided by the incumbent supplier were then used to distribute the new service.

Figure 3.13: Installation of Rutland Telecom FTTC Service in Lyddington

The project in Kings Cliffe followed a similar process, with Gigaclear approaching residents in summer 2014 to gauge interest. The project required a minimum threshold of 30% of households in the village to sign up before commencement and this figure was achieved after 3 months. The nearby area of Apethorpe was also included as an extension to this project and achieved its sign up target shortly after Kings Cliffe.

It was noted that Northants County Council had displayed a commitment to ‘pushing and backing’ providers of high speed broadband in the area previously, but in reality the only incumbent supplier able to deliver the necessary infrastructure overhaul were BT, who had expressed no Interest in extending fibre optic service to Kings Cliffe, and as such, private investment was deemed a viable solution. A local business had also contacted BT directly to discuss the provision of better service, but the response was that only a bespoke cabling solution could provide this, at a cost of tens of thousands of pounds. A survey of residents was also conducted in 2010 to assess needs in terms of broadband service, although the village demographic at this time was sufficiently different that little interest was shown. Since the 2010 survey, however, new developments in the village including a large housing estate had seen an influx of younger professional families more engaged with the Internet. It was suggested that the increased population led to bottlenecking of existing broadband services and the new lines required to serve new properties had exceeded the capacity of BT’s cabinets in the area.

3.4.2 Service Implementation

The implementation of the Gigaclear infrastructure was completed in late August 2015, although unlike the implementation in Lyddington, which was FTTC, the deployment in Kings Cliffe is FTTP (Fibre to the Premises). This means higher connection speeds can be achieved and connections are synchronous as opposed to asynchronous which enables equal upload and download speeds. Such connections are not widely available through FTTC infrastructure as bandwidth is generally partitioned in favour of downstream data transfer enabling faster browsing and downloads. Because FTTP connectivity was deployed by Gigaclear, the service offerings and associated speeds are much higher than those achieved through FTTC solutions.

Gigaclear offers residents of Kings Cliffe four service packages; base level 50Mbps, 100Mbps, 200Mbps and 1000Mbps (all are synchronous connections). As such, the speeds achievable in the area far exceed offerings in many urban areas. It was noted that the highest achievable connection speed before the Gigaclear implementation was around 6Mbps, but speeds would frequently drop to around 2Mbps or lower in peak times, below the government’s Universal Service Commitment threshold. Gigaclear estimate that the capital costs of the trenching and cabling will be offset over 5 years, although this will be dependent on on-going subscription amongst residents.

Initial reports suggest that the new service offerings are attracting remaining residents away from

Figure 3.14: A Typical FTTP Installation Showing Coiled Fibre-Optic Cable Ready for Termination in a Customer Premises

the incumbent supplier as the Gigaclear network enables residents to use services which rely on high-bandwidth connections such as high definition video streaming, online gaming and video calling. In some instances, residents have transferred to the Gigaclear service as it enables them to work from home or run businesses in the locality.

The case studies of Lyddington and Kings Cliffe have demonstrated a requirement for high-speed connectivity in areas that are still overlooked by schemes such as BDUK and the RCBF, or because the cost of infrastructure overhaul is viewed as unprofitable by an incumbent supplier. Although some existing residents may be unlikely to realise the benefits of the Gigaclear network in Kings Cliffe, the expanding population in the area highlights the need for infrastructure that can cope. Benefits such as high-speed broadband access are likely to become an increasingly important factor in the decision to purchase new homes in rural areas, particularly so for those who rely on high-speed access for their jobs and lifestyle. As such, it may be argued that private networks such as those deployed by Gigaclear become increasingly commonplace in rurally isolated areas.

3.5 A Consideration of Mobile Broadband: A Substitute or Complementary Infrastructure?

Although peripheral to the scope of the research, this section considers the geography of mobile broadband and its potential as an alternative to fixed-line infrastructure in the domestic context. In line with recent infrastructure developments to improve fixed-line broadband performance, mobile broadband technologies have also been overhauled. Much of this improvement has enabled higher mobile data transfer speeds through the rollout of 4G LTE connectivity across major urban centres throughout the UK. 4G technology (sometimes referred to as Long Term Evolution or LTE) has the potential to support speeds of up to 100Mbps downstream under optimum conditions and could be considered a viable alternative to fixed-line connections in some areas. For a comprehensive breakdown of the various mobile broadband technologies that are operated in the UK, see Section 2.1.6 of the literature review.

Unlike the speed test and exchange location datasets that have been analysed in previous sections of this chapter, there are limited data resources available to assess access to, and performance of, mobile broadband networks. Recently, broadbandspeedchecker.co.uk began to collate data relating to mobile broadband performance through a custom mobile application designed specifically for testing cellular data transfer speeds. Unfortunately, this data was not made available for this research and there are limited numbers of businesses crowdsourcing mobile broadband performance statistics, as such, this data generally commands a high premium. In addition, there have also been issues regarding the sensitivity of mobile infrastructure data. Unlike the dataset of UK telephone exchanges obtained freely from SamKnows, access to mobile infrastructure data has been restricted in recent years. One such dataset that has been heavily guarded is the Ofcom 'Sitefinder', a large dataset of the locations and operational characteristics of mobile cellular transmitters in the UK.

Sitefinder was set up as a result of recommendations of the Stewart Report in 2000 and represents a voluntary scheme under which mobile network operators make information available on the location and operating characteristics of individual base stations, so that people who wish to inform themselves about this can do so. The Stewart Report was commissioned by the UK Government and conducted by the Independent Expert Group on Mobile Phones (IEGMP) to consider the then concerns about the possible health effects from the use of mobile phones, base stations and transmitters. The Report conducted a rigorous assessment of existing research and gave advice based on the state of knowledge at the time, as well as recommendations on further work that should be carried out to improve the basis for sound advice. The report concluded that there was no general risk to the health of people living near to mobile base stations (IEGMP, 2000).

Ofcom hosts the Sitefinder tool on behalf of Government, which can be searched for the location and details of mobile phone base station sites around specific locations. The data within Sitefinder is owned by the mobile network operators, who supplied it on a voluntary basis, and as such, Ofcom makes no corrections to the data supplied by the operators. The last update to the Sitefinder dataset was in May 2012, although some operators ceased providing information as early as 2005 given growing pressure from both the Information Commissioner and the general public to make the database publically available. The dataset was made publically available in 2012 after a Supreme Court ruling ended over six years of legal debate over the sensitivity of the data. Initially Ofcom had denied public access to the full dataset over concerns for national security. The dataset used for analysis in this section reflects the final update from May 2012 which, although dated, remains the most comprehensive publically available dataset relating to mobile broadband and cellular infrastructure.

Unlike earlier analysis of broadband speed tests, no data is available within the Sitefinder dataset to assess the performance of mobile broadband connections. Instead, the locations and spatial density of mobile base stations are used as a proxy for likely levels of service. This analysis is by no means comprehensive, but is designed to visualise the geography of access to mobile broadband, as well assess the feasibility of mobile broadband substitution in the domestic context.

Figure 3.15 shows the spatial distribution of mobile base stations across England, extracted directly from the Ofcom Sitefinder dataset. It is apparent that mobile base stations tend to cluster around densely populated urban conurbations. Major urban centres such as London, Birmingham, Bristol, Manchester, Liverpool and Newcastle are all clearly identifiable in the visualisation. In addition to telephone exchanges, which supply fixed-line Internet connectivity, urban centres would logically require higher densities of mobile base stations to support cellular telephony and mobile data services for comparatively large populations.

In addition to following aggregate population distributions, the locations of mobile base stations also follow a clear pattern of major transport infrastructure, clustering along road networks, where large numbers of people are frequently in transit and require cellular service. Some major roads can be identified, such as the A1 from London, northwards through the Home Counties, Lincolnshire, West Yorkshire and dispersing around Newcastle, where the area to the north (Northumberland and the Scottish Borders) becomes increasingly remote. In the South West, the M4 between London and Bristol, A303 through Dorset and Somerset and M5 between Bristol and Exeter are clearly visible. In the North West, the M6 is visible between Merseyside and the Scottish Borders. More aggregate spatial interpretation of the dataset reveals large disparities between the supply of mobile base stations in urban and rural areas. These apparent disparities mirror those of the supply of telephone exchanges as shown in Figure 3.1 and highlight why the supply of mobile data services, as well as voice-only services, are often heavily constrained in rural areas, as base stations must

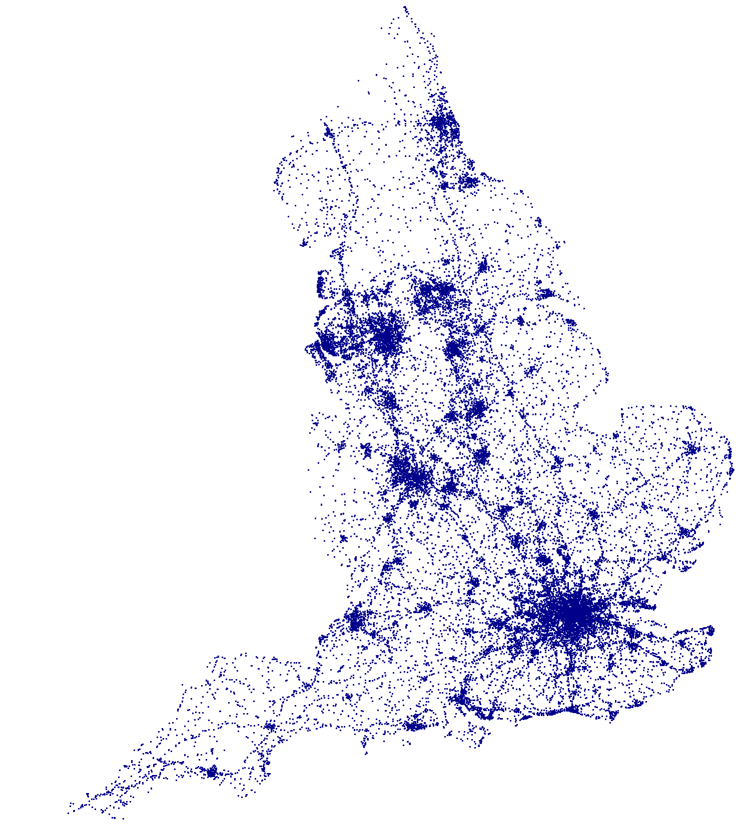


Figure 3.15: Spatial Distribution of Mobile Base Stations in England

cover large areas and disparate populations. Such areas would be unlikely to receive infrastructure investments given the high costs of roll out and limited return on investment for mobile network operators. In some scenarios, mobile broadband has been suggested as an alternative where the feasibility of a fixed-line connection is heavily constrained (Cardona et al., 2009; Prieger, 2013). However, in the context of England, obvious disparities in the supply of mobile infrastructure would suggest that this is an unrealistic alternative for many remote areas and indeed the government has identified fixed-line fibreoptic connections as the most viable means of supplying broadband connectivity to the most rural areas (DCMS, 2009, 2010).

To further highlight the apparent disparities in access to mobile base stations, the average distance to a mobile base station per LSOA was calculated and is presented in Figure 3.16. LSOAs were used as the spatial unit for this calculation, as opposed to districts, which have been utilised previously in this chapter. LSOAs better reflect the geographic extents in which a mobile base station can operate in terms of traffic from local populations and the spacing of such infrastructure. Base stations are typically spaced around 0.2-0.5 km apart in urban areas and 2-5 km apart in rural areas (Ofcom, 2015). The centroids of each LSOA in England were calculated in a GIS and the average distance between these points and any base stations that fell within the bounds of each LSOA were used to calculate a mean distance measurement.

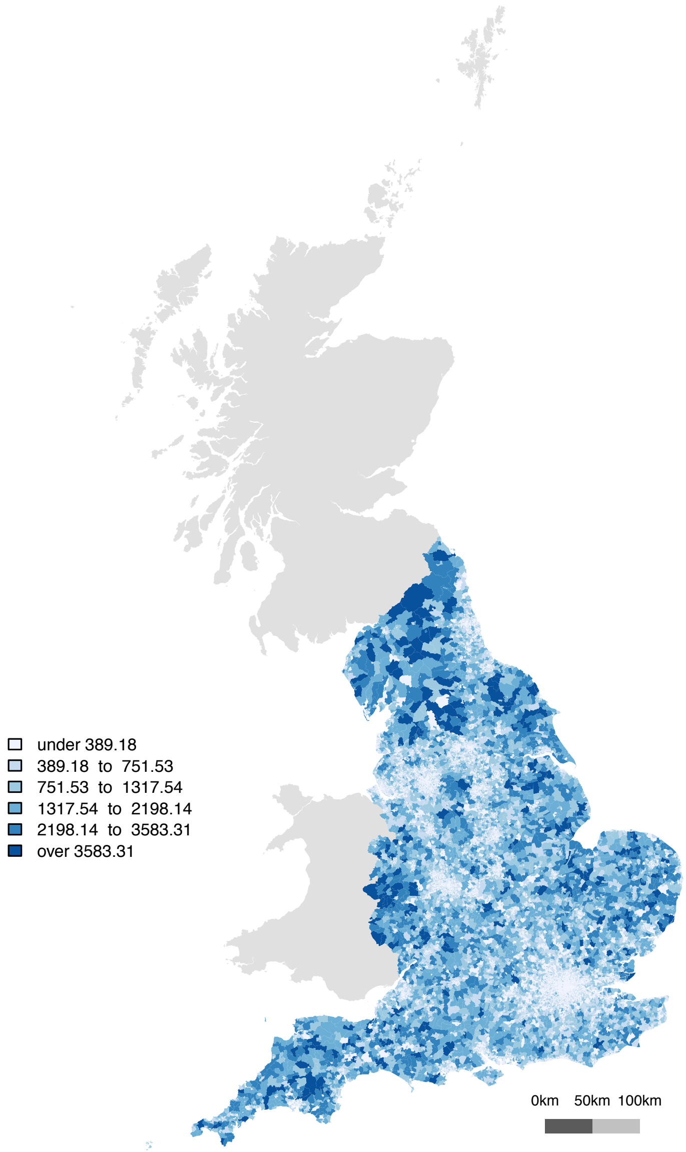


Figure 3.16: Average Distance (m) to Nearest Mobile Base Station, English LSOAs

It is apparent that mobile base stations are more sparsely distributed in predominantly rural areas, and more densely distributed in urban areas. Major urban centres are again clearly visible in Figure 3.16 and the lower average distances to mobile base stations in these areas would suggest higher levels of service. For rural areas the higher average distances would suggest poorer levels of service or no service at all, depending on exact location. One caveat of this analysis is that the Sitefinder dataset holds no reliable information on the services distributed by each cell tower (i.e. voice only or voice and data capable). In this respect, it is not possible to ascertain whether base stations in rural areas are distributing both data and voice services, or voice service alone. Although the same caveat is true of those cell towers located in urban areas, the operating models of telecoms suppliers mean these areas are generally prioritised to receive enhanced services first, as such, it would be expected that the majority of cell towers located in urban areas are voice and data capable. Such sparse distribution of mobile base stations in the most rural areas, which record average distances of over 3.5km to the nearest tower in many cases, supports the government's decision that only fibre broadband should be viewed as a viable long term solution to connect the most isolated areas. With fixed-line broadband connections (without a fibre backhaul) unlikely at distances of over 5km (Ofcom, 2013a) it is even less likely that cellular data service would be a reliable or feasible alternative in these areas.

3.6 Conclusions

This chapter has examined the current geography of access to fixed-line broadband services in England, with a short consideration of mobile infrastructure. Firstly, through analysis of the distribution of distance to the closest exchange it was revealed that the majority of speed tests were ran within relatively close proximity to a local exchange and that large distances between the end user were observed as leading to poor performance, likely as a result of poorer infrastructure or signal attenuation, supporting the literature presented (see Chapter 2, Sections 2.3 - 2.4). Performance was also found to be unevenly distributed when speeds were profiled by indicators of socio-spatial structure such as the Output Area Classification (OAC). The results suggest that those areas which are more densely populated or are within close proximity to a major urban centre are most likely to have access to fast connections. When we consider how broadband speeds are increasing over time, it is evident that the highest increases appear in urban and semi-urban areas, while rural areas see far smaller performance increases despite consistently recording low average speeds. This would suggest that domestic disparities in broadband speed are widening as rural areas are left behind. With regards to access, as opposed to performance, the raw number of speed tests per OAC group suggested that differentiation exists between socio-spatial groups in terms of use of the Internet, mirroring findings from the review of literature. OAC groups that represent areas with older resident populations, who are frequently reported to be less engaged with the Internet (see Chapter 2, Section 2.4.1) are are under-represented in the datasets. Profiling speeds by indicators of deprivation has revealed that the most deprived areas generally see the highest aggregate speeds, as they are commonly located in large urban areas, with the most advanced infrastructure. Conversely, analysis of change in speeds between the datasets used has suggested that it is affluent areas that are seeing the largest increases in performance, most likely due to infrastructure upgrades. After major urban centres, it is likely that that affluent areas represent the best markets for delivery of next generation services. In visualising a 'snapshot' of average broadband speeds at district level, this research has identified areas which benefit from strong performance, as well as those areas (predominantly rural) which may require investment to reduce disparities. Despite average speeds at district level increasing, analysis suggests there may be some areas which are still falling below the USC threshold of 2Mbps. Within England, the largest speed increases appear to be in districts which are predominantly urban, but are not major urban centres. The smallest increases in speed are generally recorded in rurally isolated areas. Similar spatial disparities exist when profiling only those connections which are considered to be 'superfast', with urban areas displaying far higher superfast penetration than those which are rural. Examining how average download speeds fluctuate throughout the day has revealed that at peak hours, aggregate speeds decrease. Conversely, in the early hours of the morning when fewer users are online, average speeds increase. When these daily fluctuations are profiled by measures of rurality, it is apparent that the largest speed fluctuations occur in urban areas, with smaller fluctuations in rural areas, however these areas consistently record much lower average speeds. This analysis has also highlighted the large performance disparities that exist between urban and rural areas. In addition, a short case study of private infrastructure investment has been presented, highlighting how some rural populations have a strong requirement for high speed broadband but are overlooked by incumbent suppliers due to the high costs associated with infrastructure overhaul. Such private investment may become increasingly commonplace if Internet access becomes an important factor in the decision to build or purchase homes in rural areas. Finally, the geography of mobile infrastructure has been examined using a dataset of mobile operator base station locations. Through analysis it is apparent that the geography of mobile infrastructure broadly mirrors that of fixed-line, with better infrastructure provision in densely populated urban areas. It is also apparent that mobile infrastructure clusters around major roads and transport links. In terms of viability as a substitute service, the low densities of mobile infrastructure in the most rurally isolated areas would suggest that this may not be feasible, supporting the government's view that only fixed-line connections through fibre are viable to connect the most remote areas (DCMS, 2009, 2010).

Chapter 4

A Geography of Internet Demand and Engagement

4.1 Introduction

In order to fully understand the geography of Internet use and demand it is necessary to gain insight into people's attitudes and behaviours towards the Internet. While previous chapters in this thesis have examined the influences of infrastructure and access to fixed-line and mobile broadband services, this chapter is focused around the variable geographies of demand for the Internet, the services it provides, and how disparities may be occurring within and between different societal groups and locations. The empirical work contained within this chapter is based on analysis of the 2013 Oxford Internet Survey (OXIS), a survey of the general public commissioned by the Oxford Internet Institute (OII) every two years since 2003, which aims to gather information about Internet access, use, attitudes and the difference this makes for everyday life in Britain. The sample used for the 2013 OXIS is representative of the UK population, but is small in size. As such, an innovative method for synthetic data estimation has been implemented to extrapolate the survey results to national small area coverage. Such analysis falls within the scope of 'Small Area Estimation', a statistical and computational technique involving the estimation of parameters for small populations; either a geographic area or socio-demographic group. The estimations are 'indirect', in that they borrow strength by using values of the variables of interest from related areas. These values are brought into the estimation process through a model (either implicit or explicit) that provides a link to related areas (domains) through the use of supplementary information related to the variables of interest, such as recent census counts and current administrative records

(Rao, 2003). In the most basic sense, it is possible to predict results for unsampled areas by using data from sampled areas. Profiling the relationship between age structure and Internet usage within a sampled geography, for example, and subsequently using the results to predict rates for an unsampled geography where no survey data is available but the age structure is know (I.e. from a recent census of population). It is however, a far more intricate procedure in practice, especially so when there are a number of influences (or explanatory variables) which are shown to affect a given outcome. Imputation, although similar to synthetic estimation in its methods, represents a different area of analysis. Imputation is best described as the replacement of a missing attribute for an individual in some form of survey or census with a 'best guess' value, typically a sub-group mean. Synthetic estimates are estimates that are directly derived from the analysis of sub-groups of a dataset. In the case of the OXIS, estimates are not required to replace missing data, but are instead being created at area level, where the area is a form of administrative geography that is composed of many combinations of sub-groups of the original dataset. Because the OXIS is complete (i.e. there is no missing data) and it is representative, strictly speaking, no imputation work is required. The methods of synthetic estimation presented in this chapter should therefore not be confused with imputation methods. Although the methods presented in this chapter are used for the purpose of small area estimation, they also fall within the scope of 'Data Mining', particularly so for the second example presented, which uses a Quick Unbiased Effective Statistical Tree (QUEST) algorithm to identify response rates amongst sub-populations of the OXIS sample. Data mining is an emerging interdisciplinary sub-field of computer science and statistics, which has grown in popularity amongst academics and industry in recent years (Bordea et al., 2015; Shakshuki et al., 2014; Uzunboylu et al., 2012). The purpose of data mining is to discover hidden knowledge in large datasets and techniques have been used across industries such as banking and financial services, retail, healthcare, telecommunications and counter-terrorism (Xiao and Fan, 2014). Moreover, decision tree algorithms are commonly used in data mining and have been shown to yield highly accurate results (Yu et al., 2010).

4.1.1 Objectives

The main objectives of this chapter are twofold; the first involves designing a robust methodology through which estimates of responses to a set of questions from the OXIS can be calculated. In designing such a methodology, a detailed investigation of the factors that influence people's use of the Internet is presented. Estimates required national coverage and to be calculated at a 'small area' level, for which census covariates to be used in the modelling process can be obtained. In this chapter two modelling approaches are presented; small area estimation through ecological regression and small area estimation through decision tree induction. The latter of the two methods was used to calculate the final estimates, but a technical account of both methods is presented. The ecological regression approach sought to calculate estimates at the aggregate level of the Lower Super Output Area (LSOA) and the decision tree method sought to calculate estimates at the respondent sub-group level which were then extrapolated to the Output Area (OA) level. Due to limitations of the ecological regression approach, this analysis could not be conducted at the respondent level, as respondent level census statistics would have been required to make predictions, and these data are restricted. The sub-groups for which estimates were created in the decision tree approach could be re-created using Output Area level census data, and as such, rates by sub-group could be used to calculate area level estimates. Whilst this represents a caveat in that the two approaches are not directly comparable, the purpose of this section is to present the methods that were explored, as opposed to drawing explicit comparisons between different modelling approaches. The second objective concerns the modelling and subsequent profiling of the estimates. Analysis aims to test the strength of the estimates through profiling by indicators of socio-spatial structure, in addition to investigating geographic variability and patterns. Given the research that has been presented regarding geographic disparities in broadband infrastructure and access in previous chapters of this thesis, it would be expected that similar patters be apparent when Internet use and application is explored. Comparisons of estimated rates against observed national rates derived from the OXIS are also presented for questions that were modelled, aiming to highlight any large deviations that may signify sporadic predictions.

4.2 Data Characteristics

The estimation and profiling of rates for OXIS questions required attributes from a number of datasets. These include:

- The 2013 Oxford Internet Survey (c. 2660 respondents)
- 2011 Census datasets at the OA and LSOA level
- The National Statistics Socioeconomic Classification (NS-SeC)

4.2.1 The 2013 Oxford Internet Survey

The Oxford Internet Survey (OXIS) was launched by the Oxford Internet Institute (OXII) in 2003. Subsequent surveys have been conducted every two years. Each survey uses a multi-stage national probability sample of 2000 people in Britain, which enables the projection of estimates to Britain as a whole. In 2013 the OXII received funding from dot.rural (`www.dotrural.ac.uk`) for an additional

600 rural responses. dot.rural is the Research Councils UK Digital Economy Hub focusing on research into the rural digital economy. The inclusion of the additional dot.rural funded responses raises the total number of respondents in 2013 to 2,657. The comprehensive 2013 OXIS dataset was supplied directly from the OXII for use in this thesis. It is important to note at this stage that the OXIS covers England, Wales and Scotland, however, data pertaining to Scotland were removed before national estimates (for England) were produced. This was due to some datasets used in the predictive models not being available for Scotland, there are also different levels of geographic aggregation used in Scotland which limits the ability to use consistent variables. Similar issues arose during the previous analysis of speed check data and are highlighted in Chapter 3, Section 3.2.3. Despite the objective being to create estimates for England only, data pertaining to Wales were retained alongside England for the estimation models as this provided a larger sample to work with. Equally, the census datasets earmarked for analysis covered both England and Wales at common geographic resolutions so there were no issues regarding data inconsistencies. A number of checks were made to ensure that the OXIS sample remained broadly representative of urban and rural areas and of socio-spatial groups when Scotland was removed form the dataset. In all cases the sample did not appear to become unbalanced by either measure. The 2013 OXIS was compiled through interviews, conducted face-to-face in respondent's homes using traditional pen and paper methods and taking place between 2nd February and 14th April 2013. The sample covers three target populations; shallow rural, deep rural and urban and data are weighted to the profile of each target population. Target populations were defined through the Office for National Statistics Urban/rural definitions at Output Area level. Table 4.1 shows how the appropriate Urban/rural definitions map to the OXIS target populations.

Table 4.1: OXIS Target Populations

OXIS Target Population	Urban/rural Definition
Urban	1. Urban - less sparse
Urban	2. Urban - sparse
Shallow Rural	3. Town & fringe - less sparse
Deep Rural	4. Town & fringe - sparse
Shallow Rural	5. Village, hamlet & isolated dwelling - less sparse
Deep Rural	6. Village, hamlet & isolated dwelling - sparse

The Output Area is the Primary Sampling Unit (PSU) of the 2013 OXIS. Selection of OAs was based on a stratified sampling methodology for each sample; urban, deep rural and shallow rural.

Within each OA a random sample of 20 addresses were identified from the Postcode Address File (PAF) for visits. A further 10 addresses were identified as reserves for each OA, which would be issued to the survey teams upon request. A more detailed sampling design is available from the OXII [1]. The OXIS covers a varied range of topics that are both directly and indirectly associated with people's use of the Internet. The minimum respondent age in the OXIS is 14 and there is no upper limit, as such, the results are representative of the population as a whole (as opposed to a subset of the general population, i.e. those who are economically active or of working age, who may be more likely to engage with the Internet). The survey itself is split into four sections, one initial section which is completed by all respondents, followed by three subsequent sections, one which is completed by current users of the Internet, one which is completed by ex-users and one which is completed by non-users. The questions in each section cover an extensive range of topics including, but not limited to:

- Information and trust
- Attitudes towards technology
- Access in the home
- Access patterns
- Mobile phone use
- Political preferences
- Education

Given the vast number of questions that are available for analysis from the OXIS (there are over 500 potential lines of enquiry) and the complex nature of small area estimation, it was necessary to identify a smaller subset of questions relating to key dimensions of Internet use, behaviours and attitudes. Section 4.3 presents a detailed explanation of themes, question sets and the significance of these choices. The OXIS dataset is geocoded at number of different levels of administrative geography, and therefore offers flexibility as a source of geographic information. Because the Primary Sampling Unit (PSU) of the OXIS is the Output Area, the smallest administrative geography in England and Wales, aggregating data to higher levels of geography is straightforward, as is appending additional datasets. A further benefit of the PSU being at the most disaggregate level is that maximum variability in the data is retained. Each Output Area displays different characteristics in terms of age structure, economic activity, deprivation and other key indicators that are known to influence people's engagement with the Internet (see Chapter 2, Section 2.4.1). However, those output areas that are closer together tend to display more similar characteristics than those that

[1] http://oxis.oii.ox.ac.uk/sites/oxis.oii.ox.ac.uk/files/content/files/publications/OxIS_2013.pdf

are far apart. This echoes the first law of geography; that "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). If the PSU would have been a higher level of administrative geography such as the Lower Super Output Area (LSOA), and data would have been coded at this level, the subtle differences between the Output Areas that make up each LSOA may have been lost through aggregation procedures.

4.2.2 2011 Census Datasets

In order to estimate rates for the 2013 OXIS a number of census datasets were required to be used as covariates in the modelling process. In all cases, these datasets were acquired through Nomis[2], a Web-based service operated by the Office for National Statistics (ONS) for the dissemination of national census data. Tables 4.2 and 4.3 detail which datasets were acquired for each of the two approaches. Datasets for the ecological regression approach were obtained at the level of the LSOA. Datasets for the decision tree approach were obtained at the level of the OA. In all cases, the necessary area codes were appended to allow for data joins.

Table 4.2: 2011 Census Dataset Selection: Approach 1: Ecological Regression

Code	Descriptor
KS102EW	Age structure: All usual residents
QS601EW	Economic activity: All usual residents aged 16 to 74
QS102EW	Population density: All usual residents
QS613EW	Approximated social grade: All usual residents

Table 4.3: 2011 Census Dataset Selection: Approach 2: Decision Trees

Code	Descriptor
LC6114EW	NS-SeC by age: All usual residents aged 16 and over
QS102EW	Population density: All usual residents

The datasets above represent those that would be used to build covariates for the modelling process. They were selected as they represent attributes frequently noted to differentiate levels of engagement with the Internet, mainly; age, socio-economic status, employment type and measures of rurality

[2] http://www.nomisweb.co.uk/

(see Chapter 2, Section 2.4) It is important to note that only those census variables that were already appended to the individual OXIS records could be used for the purpose of modelling at the individual level. Any census derived information that was not appended to OXIS records (for example - car ownership) would not be used as a covariate in the modelling, as there is no way of appending this data retrospectively at an individual level. At area level this method is possible, as aggregate statistics can be appended to individual records using area codes. However, due to the relatively small number of sampled areas (both OAs and LSOAs) in the OXIS, only census datasets that had complete representation (i.e. all age bands/groups/classes had coverage) in the OXIS were deemed appropriate.

4.2.3 The National Statistics Socio-economic Classification (NS-SeC)

The NS-SeC has been available for use in all official statistics and surveys since 2001 and represents the replacement of two older socio-economic classifications; Social Class based on occupation (SC) and Socio-economic Groups (SEG) (ONS, 2014).The NS-SeC has been constructed to measure the employment relations and conditions of occupations, as conceptually, these are central to showing the structure of socio-economic positions in modern societies and help to explain variations in social behaviour and other social phenomena (ONS, 2014). There are a number of different versions of the classification available for use; this research uses the analytic version comprised of eight classes (excluding 'unclassified'). These classes can be further subdivided into operational categories and sub-categories, however, because the classification is required as a cross tabulation with age structure at the level of the OA for the purpose of estimation, only the eight category classification is currently available. Table 4.4 shows the eight analytic classes of the NS-SeC, the 'Unclassified' category generally includes full-time students, occupations which have not been stated or have been inadequately described and those which are unclassifiable for other reasons. The composition of the 'Unclassified' category can vary according to the source of the data.

The NS-SeC dataset that was used as part of the estimation process for the OXIS pertained to a cross tabulation of the NS-SeC by age (Census Table Reference LC6114EW). The dataset records the number of all usual residents aged 16 and over in each NS-SeC group for each OA in England and Wales. Age is banded as opposed to being presented by single year, this is required given the fine geographic resolution of the zones and likely small counts in some areas. Availability of the NS-SeC at this geography by single year of age is not available, most likely because it would conflict with census anonymisation procedures. The age groups used in the dataset are; 16-24, 25-34, 35-49, 50-64 and 65+. Although these are fixed, OXIS response data could be recoded to match this format, as respondent age was reported by single year in the dataset. A more detailed account of how the NS-SeC dataset was used in the estimation of OXIS responses is presented in

Table 4.4: NS-SeC Classes

NS-SeC Class	Descriptor
1	Higher managerial, administrative and professional occupations
2	Lower managerial, administrative and professional occupations
3	Intermediate occupations
4	Small employers and own account workers
5	Lower supervisory and technical occupations
6	Semi-routine occupations
7	Routine occupations
8	Never worked and long-term unemployed
Un	Unclassified

subsequent sections of this chapter.

4.3 Selecting Appropriate Questions

The OXIS covers a broad range of topics that are directly and indirectly linked to people's perceptions and use of the Internet. As such, for practicality it was necessary to identify a smaller subset of questions for which nationwide estimates could be calculated. Since a typology of questions is not provided with the OXIS, a classification of six domains was constructed to aid in the selection of appropriate questions. These domains were constructed and named by examining the context and content of each of the OXIS questions and grouped them as follows:

- Domain 1: Seeking Information
- Domain 2: Respondent Perceptions
- Domain 3: Household Access
- Domain 4: Mobile Access
- Domain 5: Access Patterns
- Domain 6: Commercial Applications

Within these newly constructed domains, questions were ranked according to their viability in identifying potential differences in engagement patterns. The aim was to identify questions in each domain for which the response rates (i.e. percentage of respondents who fell into each category) would vary in line with a set of key variables know to be good predictors of engagement. For example, responses to OXIS question QH7i ("Does your household have a TV with a built-in connection to the Internet?") were found to be significantly differentiated by all variables earmarked for use in the estimation process (i.e. age structure, economic activity, social grade/ NS-SeC and population density), as such this question was selected for estimation at national scale, given it would likely reveal interesting patterns of engagement. The following tables detail the questions that were selected for estimation by each domain. A summary of the reasons why each question was deemed a useful indicator is also included.

Table 4.5: Selected OXIS Questions: Domain 1: Seeking Information

Code	Question
QA1a1	Where would you go first, if you were looking for information on local MP? - The Internet
QA1a2	Where would you go first, if you were looking for information on local MP? - The Internet on a smartphone
QA1b1	Where would you go first, if you were looking for information on council tax? - The Internet
QA1b2	Where would you go first, if you were looking for information on council tax? - The Internet on a smartphone
QA1c1	Where would you go first, if you were looking for information on planning a journey or holiday? - The Internet
QA1c2	Where would you go first, if you were looking for information on planning a journey or holiday? - The Internet on a smartphone
QA1f1	Where would you go first, if you were looking for information on A topic or issue for a professional, school or personal project? - The Internet
QA1f2	Where would you go first, if you were looking for information on A topic or issue for a professional, school or personal project? - The Internet on a smartphone

The questions in domain 1 (see Table 4.5) aim to estimate the percentage of people who would use either the Internet through fixed-line or the Internet through a smartphone to seek information on the topics listed. Including estimations for both fixed-line and mobile users is deliberate and ultimately aims to highlight areas where mobiles may be the predominant form of access to the Internet, this may support the notions of substitution or infill as discussed in the review of literature (see Chapter 2, Section 2.1.7). It would be expected that certain respondent groups, for example those who are elderly, may seek the information listed above through means other than the Internet and would record low rates for these questions (supporting the notion of age as a differentiator of access and engagement, see Chapter 2, Section 2.4.1) Conversely, younger groups who are reported to have higher levels of engagement with the Internet would be expected to record higher rates.

Table 4.6: Selected OXIS Questions: Domain 2: Perceptions

Code	Question
QA2c	For you, personally, how important is each of the following as a source of information? - The Internet
QA3d	How important is each of the following for your entertainment? - The Internet
QA9	How interested are you in the Internet?

The questions in domain 2 (see Table 4.6) aim to highlight differentiation between groups of respondents in the OXIS in terms of their perceived utility of the Internet. In particular, the importance of the Internet for information and entertainment is considered. Estimating the proportion of people who are actively interested in the Internet, and see it as a valuable resource, is key in identifying groups of users who have varying levels of engagement online as studies considered in the review of literature have highlighted links between the use of the Internet for entertainment purposes, age (Longley et al., 2008) and socio-economic and cognitive resource (Peter and Valkenburg, 2006).

The questions in domain 3 (see Table 4.7) have been chosen to highlight differentials in the levels and types of access within the home. Identifying the proportion of households with Internet access in the home at a small area level, such as the OA, would assist in revealing disparities in access nationwide, as well as identifying the extent of any remaining 'binary' digital divides between the 'haves' and 'have-nots' (Norris, 2001). The questions relating to ownership of hardware such as tablets, smart TVs, games consoles and e-readers assists in building an image of digital differentiation amongst user groups (Longley et al., 2008) and access through multiple devices (Dutton and Blank, 2011, 2013). It would be expected that in areas where ownership of several of these devices is widespread, engagement with the Internet would generally be high. In areas where fewer devices are owned, this may be an indicator of more casual engagement patterns. Identifying long-term users (10 years plus) would also assist in understanding patterns of digital differentiation and identifying potential next generation users.

Domain four (see Table 4.8) aims to depict patterns of differentiation amongst mobile phone users.

Table 4.7: Selected OXIS Questions: Domain 3: Household Access

Code	Question
QH1a	Does this household have access to the Internet? - Yes, have access at present
QH1b	Does this household have access to the Internet? - Don't have access now but have had in the past
QH1c	Does this household have access to the Internet? - No, never had access
QH3	How long has your household had an Internet connection? - More than 10 years
QH5	Do you have wireless access in your household? such as through wifi - Yes
QH6a	Is your Internet connection fast enough to do what you want online? - Too slow
QH6b	Is your Internet connection fast enough to do what you want online? - Fast enough
QH7f	Does your household have: A hand held tablet with a touch screen? (eg iPad) - Yes
QH7g	Does your household have: A hand held reader for books/mags? (Kindle/Nook) - Yes
QH7h	Does your household have: Games machine? (Xbox, Wii, Playstation) - Yes
QH7i	Does your household have: A TV with a built-in connection to the Internet? - Yes

Table 4.8: Selected OXIS Questions: Domain 4: Mobile Access

Code	Question
QH11	Do you yourself have a mobile phone? - Yes
QH12b	Do you use your mobile phone for email? - Yes
QH12e	Do you use your mobile phone for posting photos or videos online? - Yes
QH12h	Do you use your mobile phone for finding directions or locations? - Yes
QH12i	Do you use your mobile phone for social network sites? - Yes
QH12j	Do you use your mobile phone for apps? - Yes
QH12k	Do you use your mobile phone for browsing the Internet? - Yes

Access through multiple devices is a key trait of next generation users, and patterns of advanced mobile application use, such as Internet, email and location-based services, would logically reflect those patterns of high engagement in the home. The estimates generated from theses questions may also, highlight areas where mobile Internet usage is high, but fixed line access in the home is low, possibly due to infrastructure constrains or as a result of rurality. Conversely, data may reveal areas of high levels of in-home access but low levels of mobile Internet use, patterns such as this, particularly access through a single device in the home, are common traits of first generation users, or those who are engaged with the Internet to a lesser extent (Dutton and Blank, 2013).

Table 4.9: Selected OXIS Questions: Domain 5: Access Patterns

Code	Question
QC1b	Do you currently access the Internet while travelling: mobile/wireless dongle? - Yes
QC2a	Overall, when you go online, do you mostly use your mobile phone or mostly use some other device like a desktop, laptop or tablet computer? - Mostly use mobile phone

The questions selected within domain 5 (see Table 4.9) have been chosen to identify groups of OXIS respondents who stay connected to the Internet whilst on the move. Alongside data pertaining to connectivity through multiple devices, information on access patterns may be helpful in identifying groups of users who's access to the Internet is near ubiquitous. Similarly, question QC2a, alongside question QH12k from the Mobile Access domain will likely aid identification of groups of Internet users who may be predominantly mobile-focused.

Domain 6 aims to depict the extent to which common tasks are performed online amongst OXIS respondents. The common tasks listed in Table 4.10 have been chosen to identify groups of OXIS respondents who use the Internet for those tasks increasingly being migrated online as a result of reduced cost, increased consumer savings and convenience (see Chapter 2, Section 2.2). Analysis of estimates may highlight areas with high rates of engagement with these online applications, and would likely be regarded as Internet savvy. Areas displaying lower rates of engagement with these activities may favour of more traditional methods such as paying bills by post or telephone, shopping on the high street and banking in branch, choices which may be strongly differentiated by indicators of socio-spatial structure. The demand for some online services may also be differentiated across measures of rurality, which would engender a complex geography of online service preferences. Those users who are rurally isolated yet digitally engaged may display higher levels of engagement with these services online given constrains in accessing them locally. Conversely, urban

Table 4.10: Selected OXIS Questions: Domain 6: Commercial Applications

Code	Question
QC22a	Have you ever found a job through the Internet? - Yes
QC22b	Have you ever saved money buying something online? - Yes
QC30b	How often do you buy products online? - Frequent
QC30d	How often do you pay bills online? - Frequent
QC30e	How often do you use online banking? - Frequent
QC30f	How often do you compare products and prices online? - Frequent
QC30h	How often do you order groceries or food online? - Frequent
QC30i	How often do you sell things online? - Frequent

populations may show lower levels of engagement with services such as online shopping and online banking as these are more readily accessible in local retail centres.

In addition to domain-specific questions, three additional questions were selected for the estimation of personal Internet use at a national scale (i.e. using the respondent's own answer as opposed to household level access as seen in questions QH1a/b/c). The questions selected for these estimations are listed in Table 4.11.

Table 4.11: Selected OXIS Questions: Determining Rates of Personal Access

Code	Question
QH13_current	Do you, yourself, personally use the Internet? - Yes. Current user
QH13_ex	Do you, yourself, personally use the Internet? - No. But used it in the past
QH13_non	Do you, yourself, personally use the Internet? - Never used the Internet

In particular, the estimation of non-users is of interest. As access to the Internet has become more commonplace over the last decade, non-users represent an increasingly marginal group. Given recent government responses to closing digital divides through initiatives such as UK Online Centres[3], Go On UK[4] and most recently digitalskills.com[5], it would be beneficial to examine the characteristics

[3] `http://www.ukonlinecentres.com/`
[4] `http://www.go-on.co.uk/`
[5] `http://www.digitalskills.com/`

of non-users. In addition, the geography of non-use could be explored to reveal the nationwide distributions of the digitally unengaged. Given the nature of government schemes to target those areas that are thought to be unengaged (typically those areas which show high levels of material and social deprivation), the availability of spatially disaggregate estimates could yield significant policy applications.

4.4 Estimation Approach One: Ecological Regression

The term 'Ecological Regression' refers to a statistical technique within the field of small area estimation whereby estimates are compiled at the level of the 'ecological unit' (sometimes referred to as the Primary Sampling Unit) using aggregate data. The ecological unit can, in reality, be any level of geography, but would typically relate to an administrative small area such as a postcode district, census output area or ward in the domestic context. The 'Regression' element refers to the statistical modelling technique of regression, which is used to calculate the relationship between the variable that is being predicted (dependant variable) and a predictor (independent variable). In simple regression an outcome variable is predicted from a single predictor variable (i.e. age vs. mobile phone ownership) whereas in multiple regression a series of predictors are used, for example, mobile phone ownership rates may be estimated using age, gender and social grade. In either instance, regression analysis is a useful tool as it allows for estimation beyond the collected data. As will be discussed later, this approach was ultimately deemed redundant for the purpose of estimating rates for the national extent, however, the method is presented in order to maintain a complete account of the development of the final methodology.

In the first estimation approach, OXIS data to be used in the regression models were aggregated to the level of the Lower Super Output Area (LSOA) for each of the questions of interest. This aggregation was performed on the dataset supplied directly from the OXII and resulted in a series of output tables detailing the LSOAs present in the OXIS sample and the associated rates for each of the questions of interest (for example, percentage of households with access to the Internet by LSOA). In addition to preparation of the OXIS data at the aggregate level of the LSOA, census covariates (independent variables) to be used in the regression models were also prepared at the same scale. In all cases these covariates were acquired from Nomis and represent 2011 census key statistics. Table 4.12 builds upon the information presented in Table 4.2 to detail covariates prepared from each of the census datasets.

The next stage involved the preparation of a dataset of English LSOAs present in the OXIS, the associated rates attributed to these areas for each of the selected OXIS questions and census-derived covariates detailed in Table 4.12. Once this dataset was assembled, it was imported into R for the

Table 4.12: Covariates for OXIS Estimations

Census Dataset	Descriptor	Variables Derrived (LSOA Level)
KS102EW	Age structure: All usual residents	% Aged 10-14
		% Aged 15-17
		% Aged 18-19
		% Aged 20-24
		% Aged 25-29
		% Aged 30-44
		% Aged 45-59
		% Aged 60-64
		% Aged 65-74
		% Aged 75-84
		% Aged 85-89
		% Aged 90 plus
		Mean Age
QS601EW	Economic activity: All usual residents aged 16 to 74	% Economically Active
		% Long Term Unemployed
QS102EW	Population density: All usual residents	Persons per hectare
QS613EW	Approximated social grade: All usual residents	% Social Grade AB
		% Social Grade C1
		% Social Grade C2
		% Social Grade DE

regression analysis. The analysis run was a multiple regression using a stepwise method. The stepwise method was selected to assess the influence of each predictor that was entered into the regression model. The direction of the model was specified as 'backward' in that all predictors are placed into the model in the first instance before removing predictors one by one to attempt to reduce remaining unexplained variance and improve (lower) the Akaike Information Criterion (AIC). The AIC is an assessment of the relative quality of a statistical model and provides a means for model selection. When the removal of remaining predictors can lower the AIC no further, the model is stopped. The backward direction was chosen for the stepwise as this limits the occurrence of suppressor effects (when a predictor has an effect on a model, but only when other predictors are present). Forward selection is more susceptible to excluding predictors involved in suppressor effects as it adds predictors in sequence as opposed to removing them from an initial model that contains all predictors. Predictors involved in suppressor effects will not enter the model in a forward method unless all predictors independently improve the AIC.

Results from the regression analysis were mixed but generally poor. Many models for the questions identified from the OXIS were of a poor fit and explained little variation in the data. None of the models derived through the stepwise regression yielded an overall an R^2 value exceeding 0.21. The ability to explain only a small amount of variance in the data was deemed insignificant and unlikely to yield accurate estimates if the models were used to predict rates at national coverage. There are a number of reasons why the models may have yielded poor fit. Firstly, this method sought to create estimates at an aggregate level, in this case the LSOA. Although this is a convenient geographic resolution, it is apparent that aggregating respondent data to this level reduces variability (or granularity) in the dataset; in essence, the data are smoothed and it becomes more difficult to identify relationships between independent and response variables through modelling. In order to improve estimates, maximum variability (more independent cases) in the data needs to be retained. Profiling at an aggregate level reduces the number of cases (these can be considered as data points or PSUs) to be used in an estimation model, reducing the number of cases generally yields less accurate results and makes estimation beyond the sample more difficult. Table 4.13 details how the OXIS sample diminishes in size when profiled at varying levels of aggregation.

Table 4.13: OXIS Aggregation Structure

Aggregation	Cases (n)	Notes
Base OXIS Sample	2657	Total number of individual responses available for profiling
OA	260	Total number of output area sampling units for which aggregate rates can be compiled. Comprised from 130 OA pairs.
LSOA	194	Total number of lower super output area sampling units for which aggregate rates can be compiled

Given there are 34,753 LSOAs and 181,408 OAs across England and Wales, the sample of these aggregate geographies in the OXIS is small. In addition, the sampling method of the OXIS, in particular, the selection of OAs, was constrained by cost. OAs to be sampled were paired, this meant that a random sample of 130 OAs was identified in the first instance, then each OA was paired with an adjoining OA to boost the sample. Adjacent OAs would likely show more similar characteristics than a pair of OAs that are more distant. Given the small size of an OA; on average containing around 125 households, socio-economic and Internet engagement characteristics are less likely to be significantly differentiated across two adjoining OAs. Again this echoes Tobler's first law of geography; that "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). This paired sampling method was adopted to reduce the travel costs of the survey teams conducting the OXIS interviews. Although the sample is known to be representative, it may be argued that selecting multiple OAs in such close proximity added little variability to the data collected (although the OXIS team do not state this was the intention), instead the pairing only acts as a means of boosting the sample.

Given the the poor results achieved using ecological regression, it was apparent that the strongest results would likely be derived by profiling the OXIS at the respondent level, i.e. 'within' the survey and without any form of aggregation. This would allow relationships to be profiled amongst several thousand OXIS respondents by those census attributes recorded as part of the survey. It was deemed this may yield stronger results than profiling a much smaller sample of PSUs such as the LSOA or OA. As such, mining the OXIS data for relationships that could predict engagement patterns was identified as the next methodological approach to be tested.

4.5 Estimation Approach Two: Data Mining Using Decision Tree Induction

The second methodology adopted for the estimation of rates for the selected OXIS questions falls within the scope of 'Data Mining'. Data mining represents an interdisciplinary subfield of computer science and statistics and is concerned with knowledge discovery in large datasets (Han et al., 2012). The aim of data mining is to extract nascent information from a dataset and interpret it such that it can be applied to further scenarios. Currently there exists a large range of data mining methods, these are firmly grounded in statistics and include; classification, clustering, regression, neural networks and structured prediction. In the use of data mining for estimation, the OXIS sample is approached from a different perspective, profiling at the respondent level, with no aggregation to administrative geographies. This retains the largest possible sample for analysis. The aim is to identify relationships and measure statistical significance in engagement patterns by multiple respondent groups. These groups would be formed by factors that have previously been shown to be good predictors of engagement, for example age, rurality and socio-economic position. Rates could then be fitted at a small area geography such as the OA, by looking at the structure of each OA by the identified groups, and estimating an overall rate based on this structure. For example, if 85% of people aged under 60 use the Internet and 40% of those over 60 use the Internet, it is possible to estimate an overall rate for an output area where 75% of the population is aged under 60 and 25% of the population is aged over 60 as follows:

$$Estimate = (85/100 * 75) + (40/100 * 25)$$

In this example the estimated % of Internet users in this output area would be **73.75%**

Of course, this example has been oversimplified and calculation by a single predictor would be unlikely to yield reliable estimates. In reality, respondents are grouped by many factors. Analysis structured by five age groups, five measures of socio-economic status and four measures of rurality could produce as many as one hundred independent sub group estimates (5x5x4 = 100) to be used in an OA level estimation. Fundamentally, this method is no different to taking a complex weighted average, using the proportion OA residents who fall into each of the sub groups identified through the analysis as the weights, and the rates reported through survey response as the figures to be weighted.

The data mining method applied to the OXIS data falls under the umbrella term of Decision Tree Induction (DTI). DTI is defined as the learning of decision trees from class-labelled training tuples. A decision tree is a flowchart-like tree structure, where each internal node (non-leaf node) denotes a test on an attribute, each branch represents an outcome of the test and each leaf node (or terminal node) holds a class label or value (Han et al., 2012). In this instance it is the value (the percentage of respondents in each leaf node) who gave a particular response to the OXIS question being modelled that is of interest. Decision trees are a common predictive modelling approach used in data mining given their visual representation. There are a large number of decision tree algorithms available for use in most statistical computing packages including; ID3 (Iterative Dichotomiser 3), CART (Classification and Regression Tree), CHAID (Chi-squared Automatic Interaction Detector) and QUEST (Quick Unbiased Efficient Statistical Trees). The algorithm selected for use in this chapter was QUEST, due to its unbiased variable selection techniques (some algorithms are biased in favour of producing splits with categorical predictors), computational efficiency and classification accuracy compared with other algorithms (Loh and Shih, 1997; Lim et al., 2000; Rokach and Maimon, 2010). The algorithm works by applying statistical tests between each input attribute and the target attribute to calculate univariate or linear combination splits (nodes). The tests are dependant on the attributes and include ANOVA F-tests or Levene's tests (for ordinal and continuous attributes) or Pearson's chi-square (for nominal attributes) (Rokach and Maimon, 2010).

The QUEST algorithm was applied to the OXIS data in SPSS (a statistical software package) because this produced the clearest and most easily manipulated visual output for the tree diagrams.

A QUEST analysis was run on each of the questions selected for estimation. In each instance a core set of predictor variables were used, which have previously been shown to be linked to Internet use and engagement patterns (see Chapter 2, Section 2.4). The predictor variables used for the QUEST analysis were age, social grade and population density quintile, the latter being used as a proxy for rurality. Population density was calculated using OA level data from the 2011 census and was appended to the OXIS sample before the QUEST analysis was run, as such, this represents external data, i.e. it was not recorded as part of the survey but would be used for estimation. OXIS respondents were first binned into five age groups; 16-24, 25-34, 35-49, 50-64 and 65plus. These age groups match those in Census Table Reference LC6114EW (NS-SeC by Age), which forms part of the target data frame to which the models built by the QUEST analysis would be fitted. It is important to note that the structure of the training data (OXIS sample) and target data (all OAs nationally for rates to be estimated) had to be identical in terms of age bands, social grade and indicators of population density otherwise the models would not fit the target data. For example, if the QUEST algorithm calculates a rate for age group 16-24/social grade AB/density quintile 4, and there is no such group in the target dataset (detailing the % of the OA comprised of this subgroup), applying a rate would not be possible.

Because the OXIS sample includes respondents aged 14 and 15, and the census derived cross tabulation of NS-SeC by age begins at age 16, OXIS respondents who were aged 14 or 15 were assigned to the 16-24 age group. Before this decision was made, profiling of 14 and 15 year old respondents was undertaken alongside profiling of the 16-24 age group as a whole. It was found that there were no significant differences in engagement patterns between the 14 and 15 year old respondents and the age group to which they would be re-assigned. As such, it was deemed re assigning these respondents (9 in total) would have no significant impact on the results of the QUEST analysis.

The NS-SeC is used in place of the Social Grade classification for fitting the models to the target data, this is because Social Grade is not available as a cross tabulation with age at the level of the OA (or any level of administrative geography). The Social Grade groups were re-mapped to their NS-SeC equivalents using a series of paired T-tests to assess similarity (using the percentage of persons per OA in each NS-Sec/Social Grade group on a national scale). There were no statistically significant differences found between the two classifications when the mapping detailed in Table 4.14 was used.

Table 4.14: NS-SeC to Social Grade Group Mapping

OXIS Recorded Social Grade	Equivalent NS-SeC Group
AB	1. Higher managerial, administrative and professional occupations 2. Lower managerial, administrative and professional occupations
C1	3. Intermediate occupations 4. Small employers and own account workers
C2	5. Lower supervisory and technical occupations 6. Semi-routine occupations
DE	7. Routine occupations 8. Never worked and long-term unemployed Un. Unclassified (including students)

As well as age, social grade and rurality, gender was considered as a potential explanatory variable. However, this proved insignificant in revealing any predictive influence in the models and was excluded from the analysis, perhaps suggesting that those gender divides discussed in relevant literature (see Chapter 2, Section 2.4) are no longer applicable. Similarly, Government Office Region (GOR) and marital status were also poor predictors. Initially, GOR was included in the QUEST analysis as an attempt to tease out geographic variability in the data, however, a measure of population density proved to be a better predictor (producing more frequent splits, with higher significance) and GOR was removed to limit risk of over-fitting. More generally it would be logical to assume that the rurality of an individual's surroundings is more likely to affect their use and perceptions of the Internet than the GOR in which they reside. Access and performance of broadband Internet is near ubiquitous across all GORs, but more heavily differentiated across different scales of rurality, as demonstrated in Chapter 3 of this thesis.

4.6 Results and Profiling

A QUEST analysis was run on each of the OXIS questions that were selected for estimation. The predictors used were kept constant (as the QUEST model would automatically remove any covariate that did not produce statistically significant splits) and included age (5 categories), social grade (4 categories) and population density (5 categories), this allowed for a theoretical maximum of 100 sub group estimates to be calculated and fitted at the OA level. Figures 4.1 and 4.2 show a typical output from the QUEST analysis, here results for OXIS question QA1a1 (Where would you go first, if you were looking for information on your local MP? - The Internet) are displayed.

Figure 4.1: Example Full QUEST Output: OXIS Question QA1a1

Figure 4.1 shows the full tree structure for question QA1a1. The internal nodes, branches and leaf nodes are clearly visible. Each leaf node details a series of figures, identifying what percentage of respondents in that sub-group gave the answers listed. A more detailed image of the left hand branches of this tree are shown in Figure 4.2. Here it can be seen, for example, that 28.8% of respondents who are aged 65 or over and fall into social grade category C1 (re-mapped to NS-SeC group 3/4) would use the Internet on a desktop/laptop/tablet to seek information on their local MP. As would be expected, in this particular group, respondents are more inclined to seek this information through more traditional means, the dominant method in this sub-group is to use the telephone. This is most likely due to the age of the respondents, however, those who are in higher social grade groups have a tendency to use the Internet. Table 4.15 highlights this relationship.

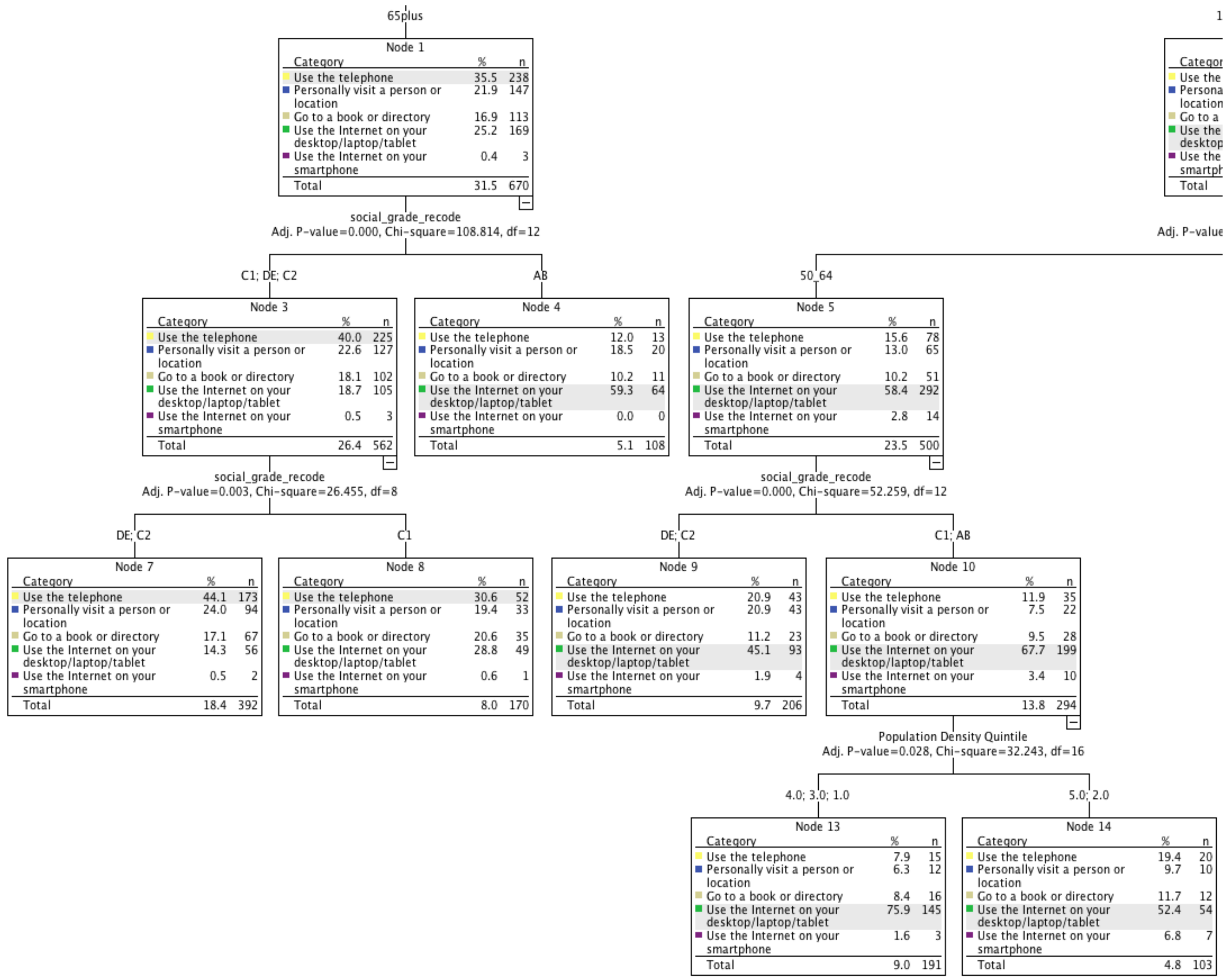


Figure 4.2: Example QUEST Output: Nodes: Question QA1a1

This relationship is common across many of the OXIS questions that were profiled using the QUEST analysis. It is evident that those respondents in higher social grade classes have a tendency to use the Internet by default, perhaps attributed to increased exposure to IT in the workplace as a result of more technical occupations and higher levels of educational attainment. This relationship is not mutually exclusive to the 65+ age group, it is evident across all age groups in the sample and echoes findings of those studies discussed in Chapter 2, Section 2.4.1. In particular, those presented by Rice and Katz (2003), Peter and Valkenburg (2006) and Longley et al. (2008), who make reference to socio-economic factors such as income, cognitive ability and exposure to technology as differentiators of engagement.

Table 4.15: Percentage of Respondents Who Use the Internet: 65+ Age Group by Social Grade: OXIS Question QA1a1

Social Grade (Within 65+ Age Group)	% of Respondents Who Use Internet
AB	59.3
C1	28.8
C2	14.3
DE	14.3

In cases where the QUEST trees were overwhelmingly large or complex, tree pruning methods were implemented to improve the efficiency of the output. Tree pruning is a method of addressing the potential of over fitting in DTI and is implemented through either a 'prepruning' or 'postpruning' approach. Prepruning involves setting a pre-defined threshold for significance and if a split in the tree construction falls below this threshold, tree growing is halted early. There are a number of issues with this method, most notably, setting an appropriate threshold, as high thresholds could produce oversimplified trees and low thresholds could result in very little simplification (Han et al., 2012). Postpruning involves the removal of subtrees from a 'fully grown' tree where results appear to reflect noise in the data. This method was implemented in the analysis of the OXIS questions in a small number of cases where multiple leaf nodes stemming from a common branch were too similar to reflect any meaningful knowledge gain from the sample. As each tree was built, a record of all the subgroups (leaf nodes) and the values of the target variable (OXIS question of interest) was kept. The values of the subgroups were then fed into a custom script written in R, which used these values, alongside the target data (containing the proportions of each OA nationally comprised of these subgroups (weights)) to estimate an overall rate for each OA nationally. For example, if a QUEST analysis had indicated, as above, that 28.8% of people aged 65+ and in social grade group C1 would use the Internet for a given task, and the target data (derived from the census) suggests that this group accounts for 10% of the population of a given OA, then the contribution of this group to the final OA estimate is 2.88% (28.8/10). This calculation is repeated for each subgroup and all rates are summed to give a final OA rate. This process was automated to repeat for every OA in the country, and for every OXIS question of interest. In instances where population density had provided a split in the QUEST analysis, OAs were subset by population density quintile, and the above sequence of calculations was applied to each subset in turn.

Figure 4.3 shows a code snippet, which calculates rates for OXIS question QA1a1 for each OA

```
imp$imp50_64_C1<-0
#where density quintile = 1 and social grade = C1
x<- which(imp$Density_quintile == 1)
imp$imp50_64_C1[x]<-75.9/100*imp$PCT_50_64_3_4_C1[x]
#where density quintile = 2 and social grade = C1
x<- which(imp$Density_quintile == 2)
imp$imp50_64_C1[x]<-52.4/100*imp$PCT_50_64_3_4_C1[x]
#where density quintile = 3 and social grade = C1
x<- which(imp$Density_quintile == 3)
imp$imp50_64_C1[x]<-75.9/100*imp$PCT_50_64_3_4_C1[x]
#where density quintile = 4 and social grade = C1
x<- which(imp$Density_quintile == 4)
imp$imp50_64_C1[x]<-75.9/100*imp$PCT_50_64_3_4_C1[x]
#where density quintile = 5 and social grade = C1
x<- which(imp$Density_quintile == 5)
imp$imp50_64_C1[x]<-52.4/100*imp$PCT_50_64_3_4_C1[x]

#----
imp$imp50_64_C2<-45.1/100*imp$PCT_50_64_5_6_C2
imp$imp50_64_DE<-45.1/100*imp$PCT_50_64_7_8_un_DE

#65plus group
imp$imp65plus_AB<-59.3/100*imp$PCT_65plus_1_2_AB
imp$imp65plus_C1<-28.8/100*imp$PCT_65plus_3_4_C1
imp$imp65plus_C2<-14.3/100*imp$PCT_65plus_5_6_C2
imp$imp65plus_DE<-14.3/100*imp$PCT_65plus_7_8_un_DE

# Final rate for Question QA1a1
imp$QA1a1 <-rowSums(imp[,26:45])
```

Figure 4.3: R Code Snippet: OXIS Rate Calculations

nationally, based on the figures derived from the QUEST analysis. In total, 42 OXIS questions were estimated at a national scale, covering each of the 181,408 Output Areas in England and Wales, reflecting a total of 7,619,136 small area estimates. Tables 4.16 - 4.22 present summary statistics for each of the OXIS questions that were extrapolated to the national extent. The national mean rate for each OXIS question (prior to any estimation) is also included for reference. If the estimated data are robust, it would be logical to assume that the mean rates calculated from these datasets should not deviate significantly from the mean rates calculated from the original OXIS datasets. OXIS question codes in the following tables relate to those questions discussed in Section 4.3.

Table 4.16: OXIS Estimates: Estimated National Means Vs. OXIS Derived National Means: Domain 1: Seeking Information

OXIS Question	Est. National Min %	Est. National Max %	Est. National Mean %	OXIS National Mean %	Deviation from OXIS National Mean
QA1a1	21.3	73.0	59.6	56.1	+3.5
QA1a2	0.5	24.4	8.5	8.7	-0.2
QA1b1	15.8	65.9	48.5	47.7	+0.8
QA1b2	0.5	20.1	5.5	6.0	-0.5
QA1c1	23.6	76.6	61.5	59.5	+2.0
QA1c2	0.9	15.0	6.3	7.4	-1.1
QA1f1	28.8	81.2	66.1	59.7	+6.4
QA1f2	1.1	17.0	5.0	5.5	-0.5

Table 4.17: OXIS Estimates: Estimated National Means Vs. OXIS Derived National Means: Domain 2: Perceptions

OXIS Question	Est. National Min %	Est. National Max %	Est. National Mean %	OXIS National Mean %	Deviation from OXIS National Mean
QA2c	32.5	95.3	77.7	77.3	+0.4
QA3d	26.8	94.5	67.7	68.6	-0.9
QA9	30.1	88.6	72.6	72.8	-0.2

Table 4.18: OXIS Estimates: Estimated National Means Vs. OXIS Derived National Means: Domain 3: Household Access

OXIS Question	Est. National Min %	Est. National Max %	Est. National Mean %	OXIS National Mean %	Deviation from OXIS National Mean
QH1a	37.8	94.6	79.2	80.6	-2.0
QH1b	1.2	6.6	3.8	3.3	+0.5
QH1c	4.0	56.8	17.0	15.4	+1.6
QH3	38.3	62.3	48.8	48.5	+0.3
QH5	85.6	96.1	93.2	95.4	-2.2
QH6a	3.9	26.5	11.6	9.2	+2.4
QH6b	50.2	72.5	63.1	63.6	-0.5
QH7f	8.49	72.5	35.2	36.8	-1.6
QH7g	9.26	41.3	27.2	27.0	+0.2
QH7h	6.2	65.4	46.0	49.5	-3.5
QH7i	2.7	28.8	21.2	21.8	-0.6

Table 4.19: OXIS Estimates: Estimated National Means Vs. OXIS Derived National Means: Domain 4: Mobile Access

OXIS Question	Est. National Min %	Est. National Max %	Est. National Mean %	OXIS National Mean %	Deviation from OXIS National Mean
QH11	65.0	97.4	91.9	91.1	+0.8
QH12b	10.6	78.5	49.3	48.2	+1.1
QH12e	6.7	64.6	35.0	34.8	+0.2
QH12h	7.2	68.8	39.2	39.6	-0.4
QH12i	3.9	71.5	38.2	37.1	+1.1
QH12j	6.7	71.5	42.1	41.3	+0.8
QH12k	8.4	79.2	47.0	45.2	+1.8

Table 4.20: OXIS Estimates: Estimated National Means Vs. OXIS Derived National Means: Domain 5: Access Patterns

OXIS Question	Est. National Min %	Est. National Max %	Est. National Mean %	OXIS National Mean %	Deviation from OXIS National Mean
QC1b	11.3	78.4	48.6	43.9	+4.7
QC2a	3.0	38.9	14.3	13.7	+0.6

Table 4.21: Estimated National Means Vs. OXIS Derived National Means: Domain 6: Commercial Applications

OXIS Question	Est. National Min %	Est. National Max %	Est. National Mean %	OXIS National Mean %	Deviation from OXIS National Mean
QC22a	2.7	50.1	25.1	26.3	-1.2
QC22b	61.4	82.3	71.9	73.8	-1.9
QC30b	37.3	66.0	53.5	55.9	-2.4
QC30d	26.6	52.5	41.5	41.9	-0.4
QC30e	40.0	63.8	52.9	54.8	-1.9
QC30f	48.6	72.2	60.8	61.3	-0.5
QC30h	8.6	27.1	17.9	18.6	-0.7
QC30i	4.0	20.9	15.1	16.8	-1.7

Table 4.22: OXIS Estimates: Estimated National Means Vs. OXIS Derived National Means: Determining Rates of Personal Access

OXIS Question	Est. National Min %	Est. National Max %	Est. National Mean %	OXIS National Mean %	Deviation from OXIS National Mean
QH13_current	32.6	95.3	78.4	78.8	-0.4
QH13_ex	1.0	6.1	3.5	3.4	-0.1
QH13_non	3.6	61.5	18.1	17.9	+0.2

Concurrent analysis of the estimated OXIS questions suggests the average deviation between mean rates of the estimated data, and mean rates of the original OXIS sample is +0.09%. This suggests the estimated dataset is broadly representative, as national means are comparable to those of the original data. Furthermore, it would appear that the estimation method is not skewing the output such that it is unrepresentative of the sample it was built from. Vastly different average rates between the estimated and original data at this stage would have flagged potential problems with estimation methods.

The next logical stage in validating the estimated rates involved profiling geographically, to examine if variability pertained to patterns that would be expected. The kinds of geographic variability in data linked to engagement with the Internet have been shown in previous chapters of this thesis and elsewhere (Parker, 2000; Wilson et al., 2003; LaRose et al., 2007; Longley et al., 2008; Riddlesden and Singleton, 2014). For example; higher levels of engagement clustering around urban centres, where the necessary infrastructure to support online engagement is most prevalent, levels of material deprivation influencing engagement or socio-economic indicators highlighting differentiated use of Internet applications. As such, data estimated from the OXIS were profiled by a number of indicators including:

- Output Area Classification
- Urban/Rural Classification
- Indices of Multiple Deprivation (IMD)
- Mapping of national patterns
- Mapping of local patterns

The following sections present the results of this profiling for a number of the OXIS questions that were estimated at the national extent.

4.6.1 Data Profiling by Output Area Classification

The ONS Output Area Classification (OAC) has been used in previous chapters of this thesis for the purpose of profiling and represents a useful tool to analyse socio-spatial patterns in large datasets. In this section, measures of perception and use of the Internet are profiled by OAC using the national estimates derived from the OXIS.

Domain 1: Seeking Information

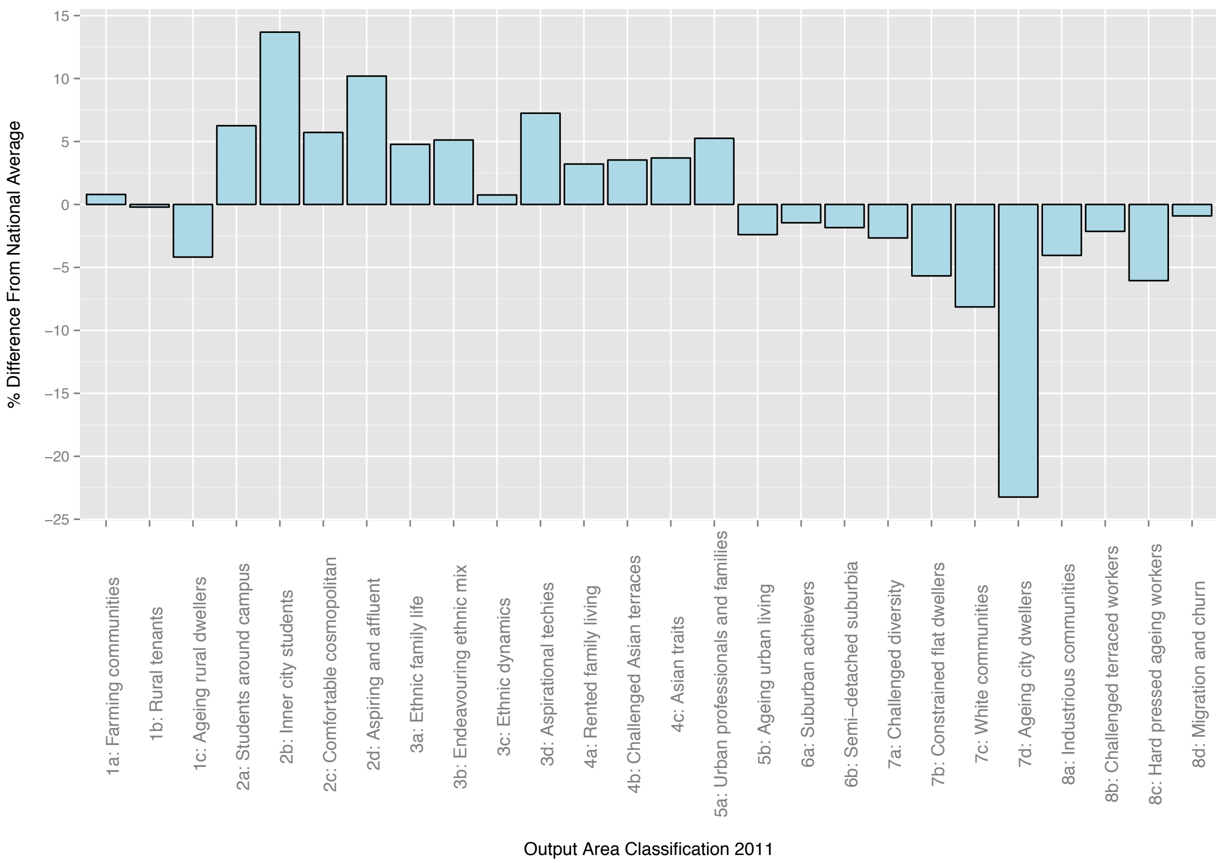


Figure 4.4: Rates of Engagement by OAC Group: Domain 1: Seeking information: QA1a1: Seeking information on local MP through Internet on a desktop/laptop/tablet computer

For this question there are clear differences between OAC groups and their constituent population's varying levels of engagement with the Internet. Figure 4.4 shows the extent of these differences for OXIS question QA1a1 (detailing the percentage of people who would seek information on their local MP through the Internet on a laptop, desktop or tablet computer). The national mean for this question is 59.5% and Figure 4.4 shows the deviation of each OAC group from this mean. There is a tendency for those groups which are predominantly comprised of younger urban populations such as 2b: Inner City Students and 2d: Aspiring and Affluent to show higher levels of engagement. As might be expected, OAC groups with older demographics show much lower rates of engagement when compared to the national average. Group 7d: Ageing City Dwellers records the lowest engagement of any OAC group for this question at 23.2% below the national average. Analysis of the original OXIS sample suggests respondents who are elderly would be more likely to seek information through traditional means such as telephone or face-to-face contact in most instances, which may be influencing the rates recorded by this OAC group. Groups that display higher levels of material deprivation, such as 7a: Challenged Diversity, 7b: Constrained Flat Dwellers and 8c: Hard Pressed Ageing Workers display lower than average levels of engagement. This would suggest that despite those areas that are materially deprived typically having better infrastructure due to urban bias (as discussed in Chapter 3, Section 3.3.2), engagement with the Internet may be more limited. Groups that are predominantly rural recorded mixed levels of engagement with groups 1a: Farming Communities and 1b: Rural Tenants close to the national average, but group 1c: Ageing Rural Dwellers around 4% below, possibly due to the more elderly age profile of this group.

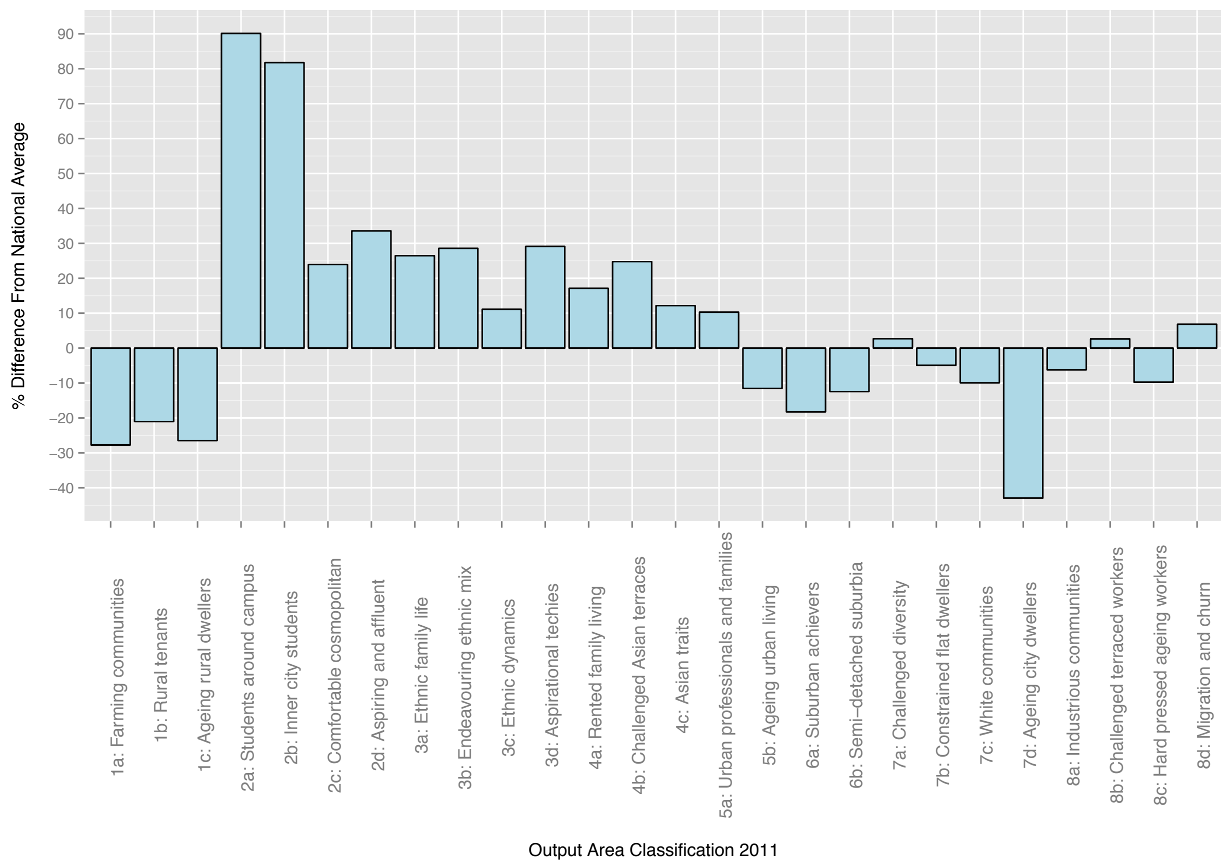


Figure 4.5: Rates of Engagement by OAC Group: Domain 1: Seeking information: QA1a2: Seeking information on local MP through Internet on a smartphone

Figure 4.5 Shows rates for OXIS question QA1a2, which is identical to QA1a1, but here rates are for those who would access information on their local MP using the Internet on a smartphone.

The inclusion of identical questions, but different methods of access was deliberate in this instance and attempts to identify which groups have higher levels of engagement with mobile technology. It has been reported for a number of years that younger users show higher levels of engagement with mobile services (Rice and Katz, 2003; Lenhart et al., 2005; Dutton and Blank, 2011) and as such, some level of differentiation may evident in the estimated data.

It is evident that some OAC groups have far higher propensity to seek information through smartphones than others. OAC groups with prevalent younger residents such as 2a: Students Around Campus and 2b: Inner City Students are up to 90% above the national average rate (around 8.5%) for this type of engagement. This is in line with trends highlighted in previous studies and suggests a good degree of correspondence with these. As would be expected, predominantly elderly groups and those containing higher proportions of retirees, such as 5b: Ageing Urban Living and 7d: Ageing City Dwellers, fall below the national average. Similarly, those groups which are rurally isolated record significantly lower than average rates of engagement, which may result from infrastructure constraints in rural areas and more limited availability of mobile broadband services (see Chapter 3, Section 3.5). Urban groups in general (see groups 2, 3 and 4) record higher propensity for access through smartphones, most likely because these correspond to geographic areas that have the infrastructure necessary to support enhanced mobile services.

Domain 2: Attitudes and Perceptions

To profile the data estimated from domain 2, which focuses on questions pertaining to respondent's thoughts, perceptions and general interest in the Internet, OXIS question QA3d (detailing the percentage of people who report the Internet is of high importance for their entertainment) was profiled by OAC. The results show obvious differentiation between some OAC groups, with those comprised of younger populations in urban settings showing the highest engagement with the Internet for entertainment purposes. The most engaged groups are 2b: Inner City Students, 2a: Students Around Campus and 2d: Aspiring and Affluent, which record rates of 26%, 17% and 16.5% respectively above the national mean of 67.7%. Again, predominantly rural groups record lower rates, likely due to infrastructure constraints in these areas, and the consistently slower speeds of fixed-line broadband services which would be required for applications such as streaming media (as presented in Chapter 3). Elderly groups also appear to be under engaged, although the primary limiting factor here is less likely to be infrastructure. Elderly populations are frequently reported to be less engaged with new technologies that underpin online entertainment (Rice and Katz, 2003; Longley, 2003; Dutton and Blank, 2013). Analysis of the original OXIS sample suggests elderly respondents prefer more traditional means of entertainment such as television, radio and spending time with other people, which may be influencing the lower rates recorded amongst these groups.

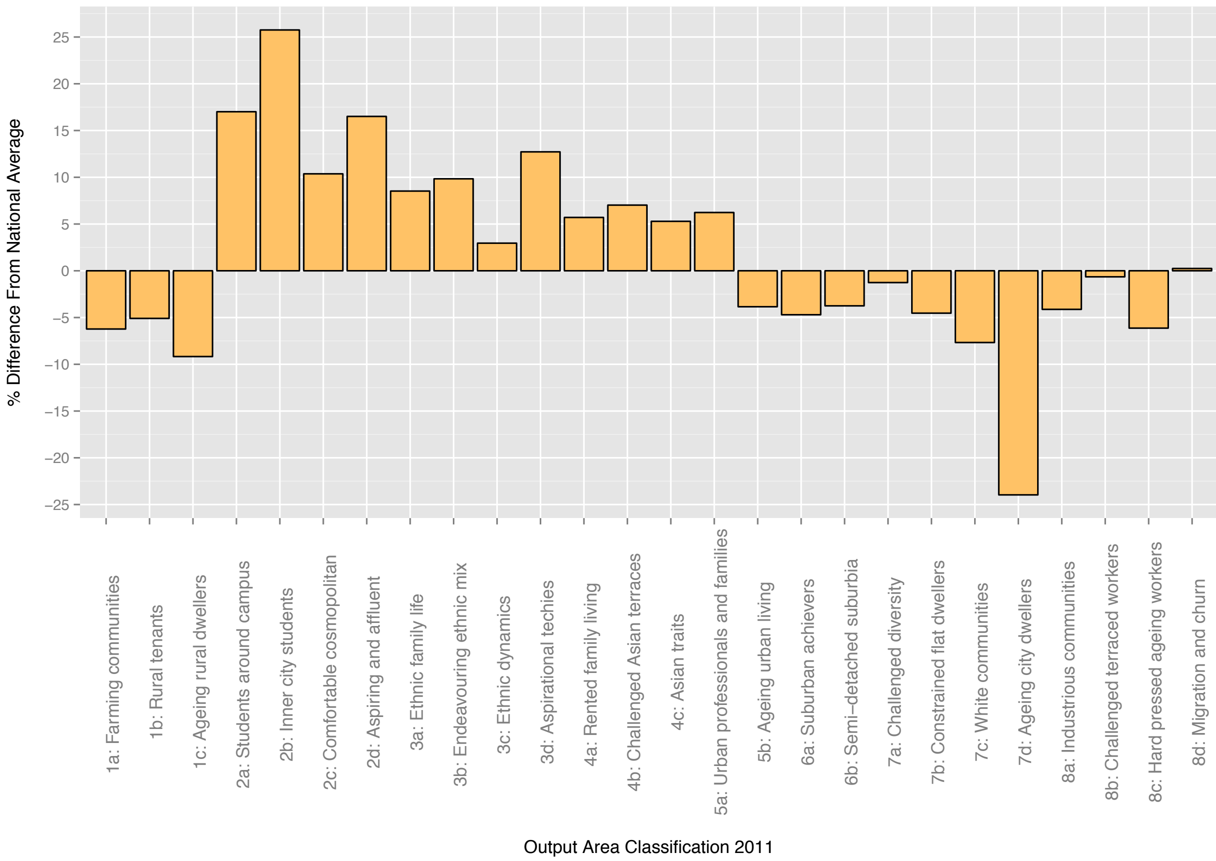


Figure 4.6: Rates of Engagement by OAC Group: Domain 2: Perceptions: QA3d: Internet regarded with high importance for entertainment

Domain 3: Household Access

To examine trends in household access between measures of socio-spatial structure, two questions have been profiled by OAC. OXIS question QH1a details the percentage of households with access to the Internet at present and question QH3 details the percentage of those households who are long term Internet users, having had access for 10 years or more. Profiling the percentage of households with access at present (QH1a) reveals varying levels of differentiation between OAC groups. The output of this analysis is shown in Figure 4.7.

It is apparent that those groups who are constrained by material deprivation (such as 8c: Hard Pressed Ageing Workers and 7b: Constrained Flat Dwellers), which likely acts as a barrier to access because of the cost of equipment and service, show lower than average rates of access in the home. The lower rates observed are likely due to these factors, as opposed to availability, as

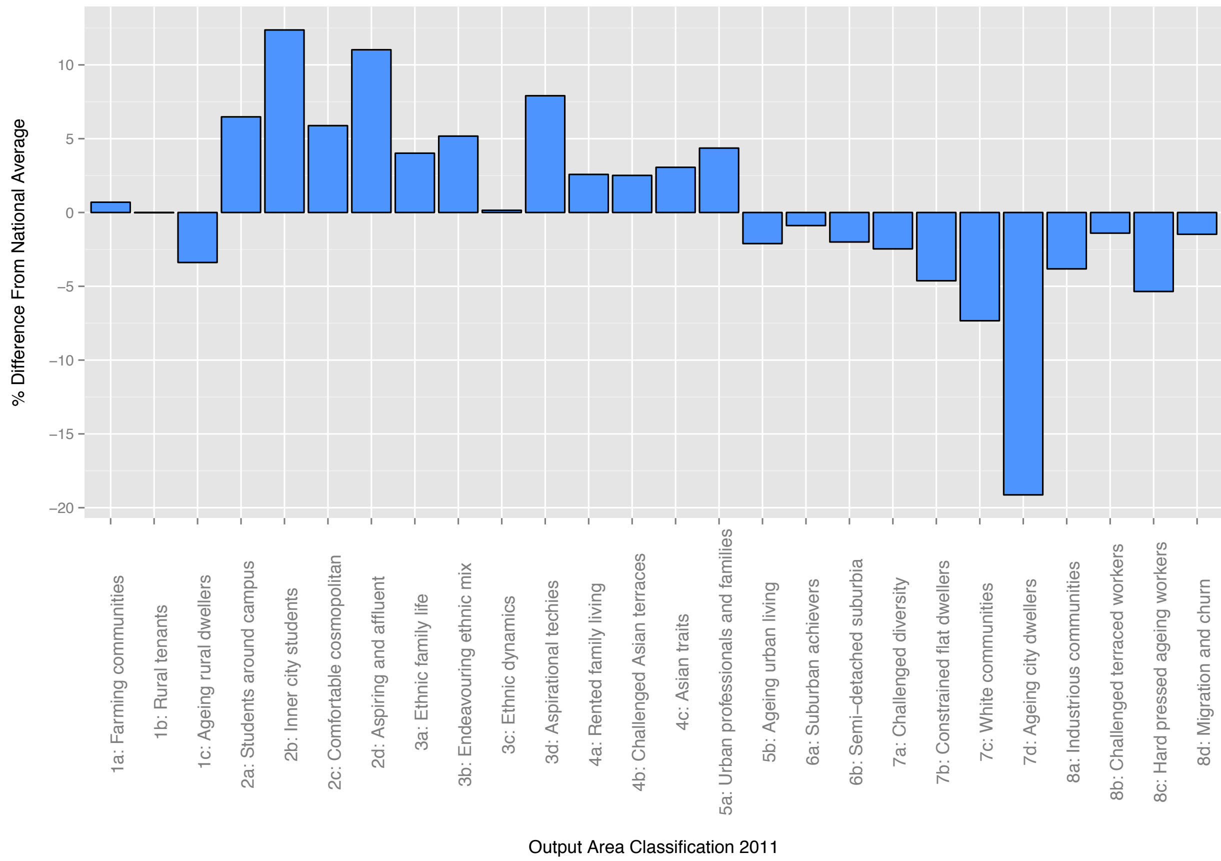


Figure 4.7: Rates of Engagement by OAC Group: Domain 3: Household Access: QH1a: Household has access at present

availability has previously been shown to be high in those areas that are ranked as more deprived, in part due to urban skew (see Chapter 3, Section 3.3.2). The national average in this case is 79.2% and the Hard Pressed Ageing Workers and Constrained Flat Dwellers groups fall 5.4% and 4.6% respectively below this. The lowest rates of household level access are, however, recorded by group 7d: Ageing City Dwellers. This group's constituent populations have a significant elderly bias, and as such, this may be a limiting factor to aggregate rates of household access. As would be expected, groups that are predominantly urban and comprised of younger, highly educated professionals and students record the highest rates of present household access. Interestingly, those groups that are predominantly rural record rates close to that of the national average, with the exception of group 1c: Ageing Rural Dwellers, where again the bias towards elderly resident populations may be influencing access.

Figure 4.8 shows differentiation between OAC groups in terms of households that have been con-

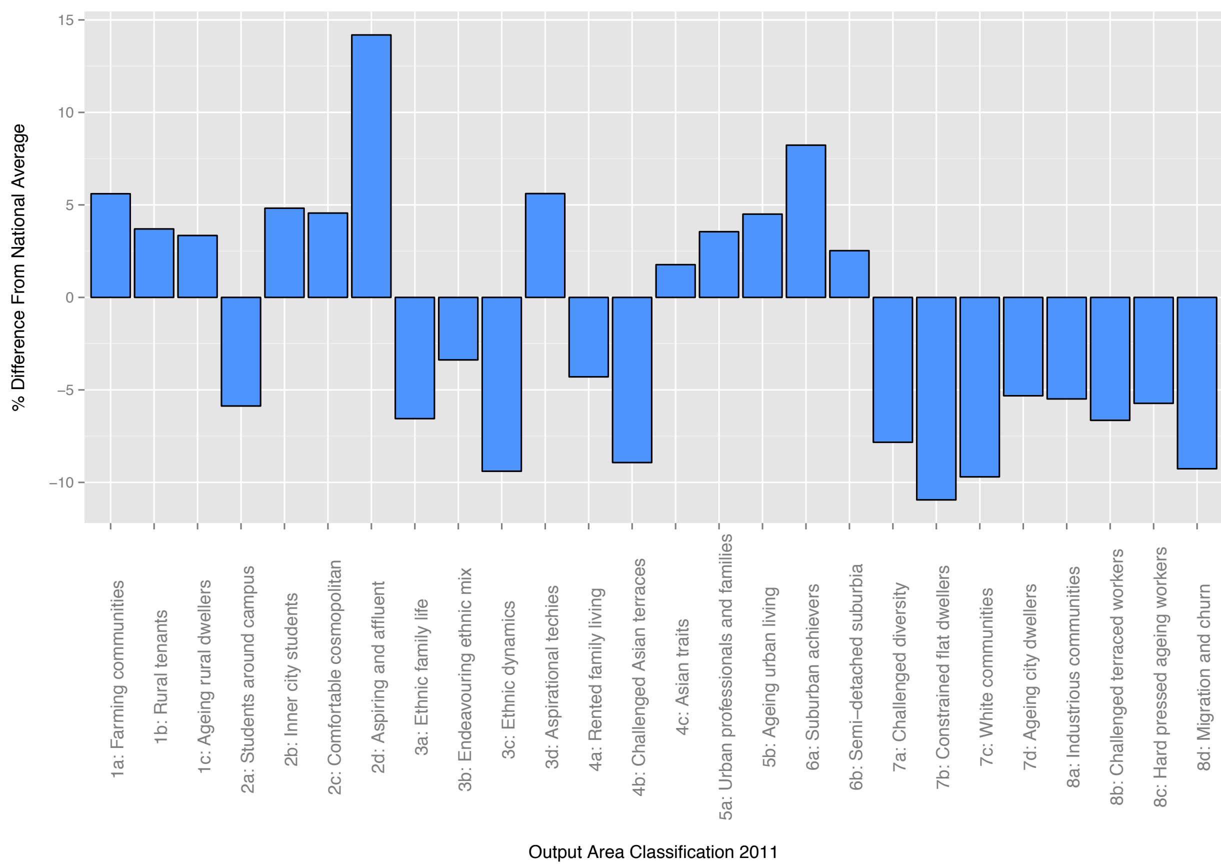


Figure 4.8: Rates of Engagement by OAC Group: Domain 3: Household Access: QH3: Long-term connected households (10 years+)

nected to the Internet for 10 years or more. The national average for this OXIS question is 48.8% and it is evident that some OAC groups show significantly higher rates of long-term access than others. For this question, the observed patterns appear to follow broad trends in age structure and prevailing levels of deprivation as might be expected, but also appear to be influenced by population turnover. In those groups that likely experience higher rates of population turnover, such as 2a: Students Around Campus and 8d: Migration and Churn, rates of long-term access are significantly below the national average. Conversely, groups that typically see lower rates of population turnover, such as 6a: Suburban Achievers and 5b: Ageing Urban Living, record above average rates for long term access, most likely because of the demographic composition of these groups and the reduced likelihood of moving home frequently, which is more typical of younger professionals and students. Interestingly, rural groups record rates of long-term household access higher than that of the national average. Given these groups typically record average or below average rates for current household access (see Figure 4.7) it is apparent that aggregate characteristics are somewhat mixed, engendering the possibility of a complex geography of engagement, alongside exclusion, in rural areas.

Domain 4: Mobile Access

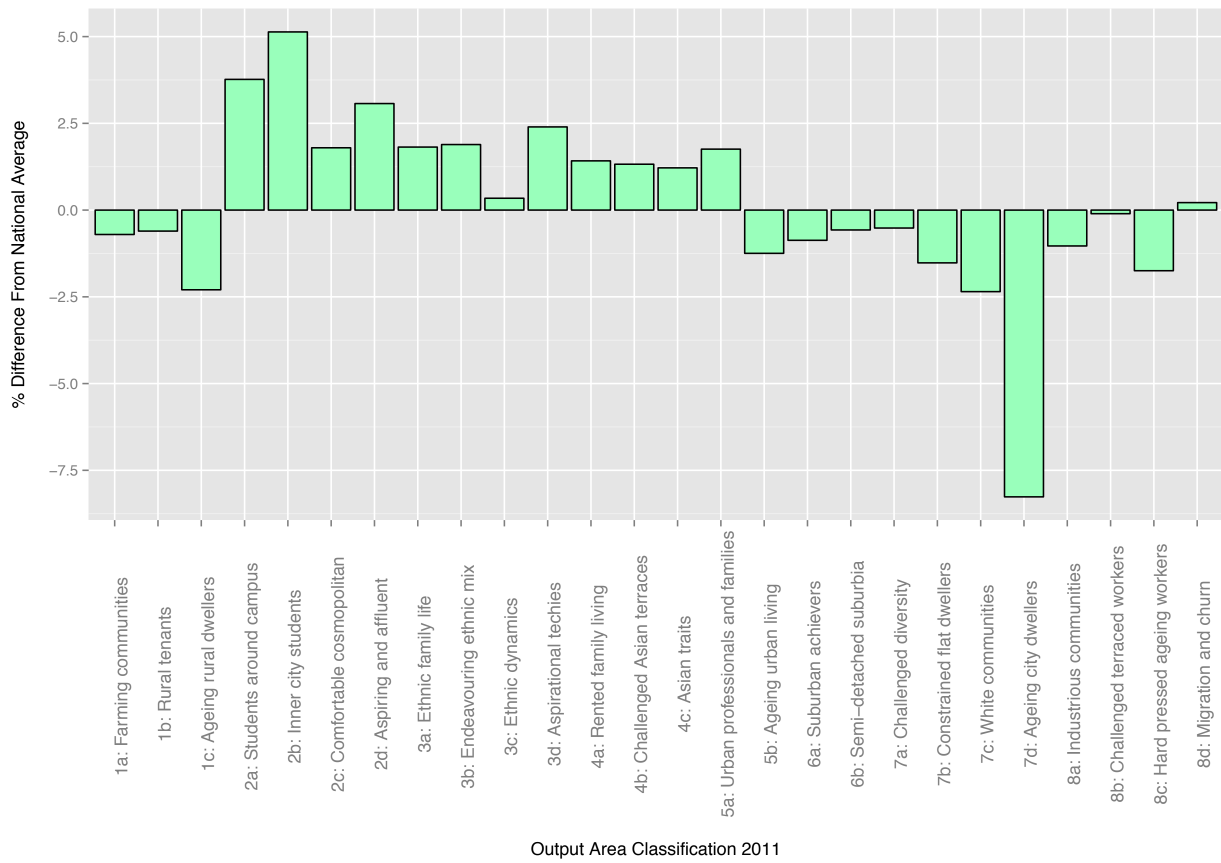


Figure 4.9: Rates of Engagement by OAC Group: Domain 4: Mobile Access: QH11: Mobile phone ownership

Figure 4.9 presents rates for OXIS question QH11 (mobile phone ownership). National rates of mobile phone ownership are generally very high, with a mean ownership rate of 91.9%. However, when profiled by OAC, differences are still apparent, albeit these are subtle, with the majority of OAC groups staying close to the national mean. As would be expected, younger populations in urban settings record higher than average rates of mobile phone ownership. Conversely, groups that are predominantly comprised of elderly or retired populations record rates below the national average. As expected, the profile by OAC for this question also broadly reflects that of seeking information through a smartphone (see Figure 4.5), as it would be unlikely for those groups who show a higher propensity to seek information through a smartphone to display lower levels of mobile phone ownership and vice versa. Rurality also appears to interact with ownership rates, with ownership rates across all rural OAC groups below that of the national average. This is likely due to limited infrastructure to support reliable cellular service in some rural areas, which has previously been discussed as a potential limiting factor (see Chapter 3, Section 3.5).

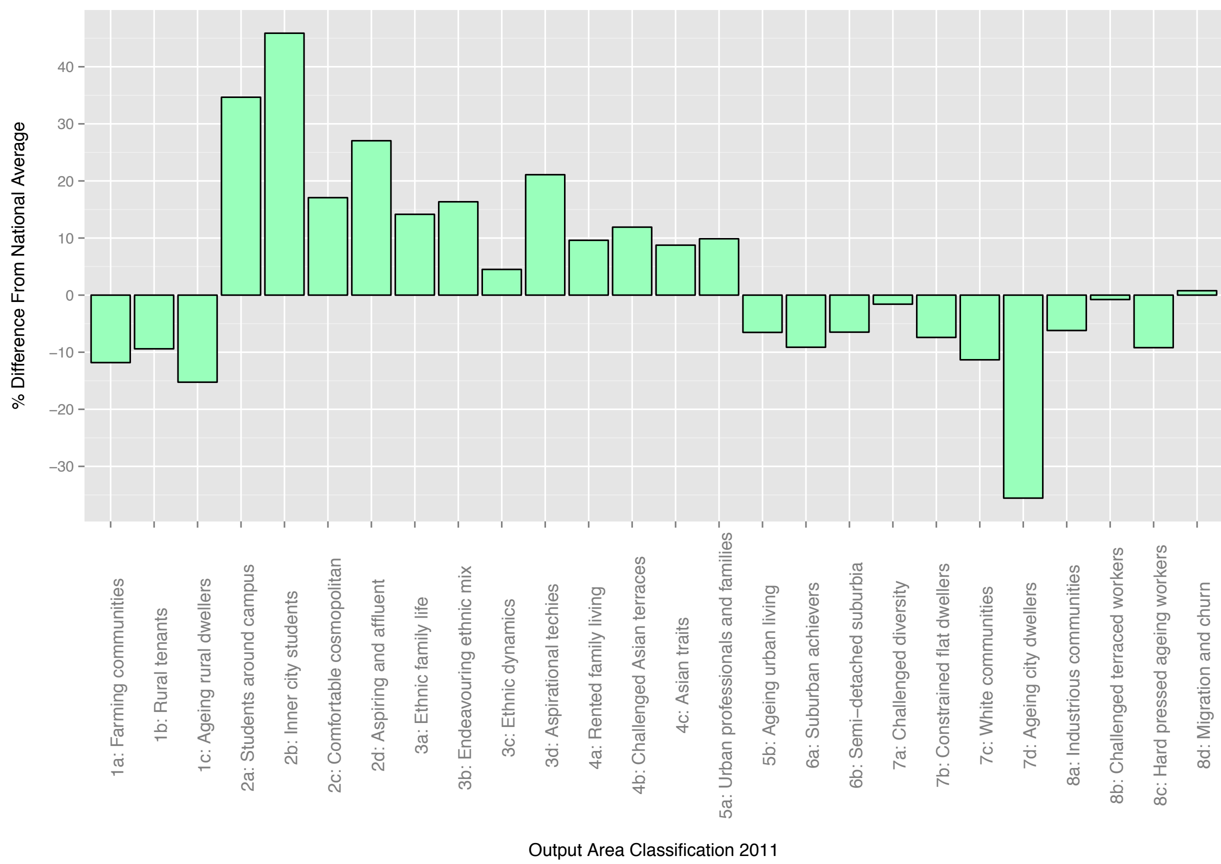


Figure 4.10: Rates of Engagement by OAC Group: Domain 4: Mobile Access: QH12k: Use Internet on mobile phone

In addition, OXIS question QH12k (do you use your mobile phone for Internet browsing) was profiled in order to establish the differences between mobile phone ownership and the use of mobile phones for mobile broadband applications (see Figure 4.10). The national average for this question is 47%. It is apparent that the use of mobile handsets for Internet browsing is most prolific among those groups associated with higher levels of exposure to next generation technology, mainly areas with students, professionals and younger populations more generally. Rurality again appears to have an influence, with urban groups generally displaying higher rates than rural groups. It is also apparent that although rates of mobile phone ownership are generally high across all OAC groups, more advanced usage, in this case mobile Internet browsing, is more exclusively attributed to those groups who display characteristics (such as younger age profiles, lower levels of deprivation and more professional careers) linked with higher levels of exposure to technology and the Internet more generally (Hargittai, 2002; Longley et al., 2008; Dutton and Blank, 2011; Modarres, 2011; Dutton and Blank, 2013).

Domain 5: Access Patterns

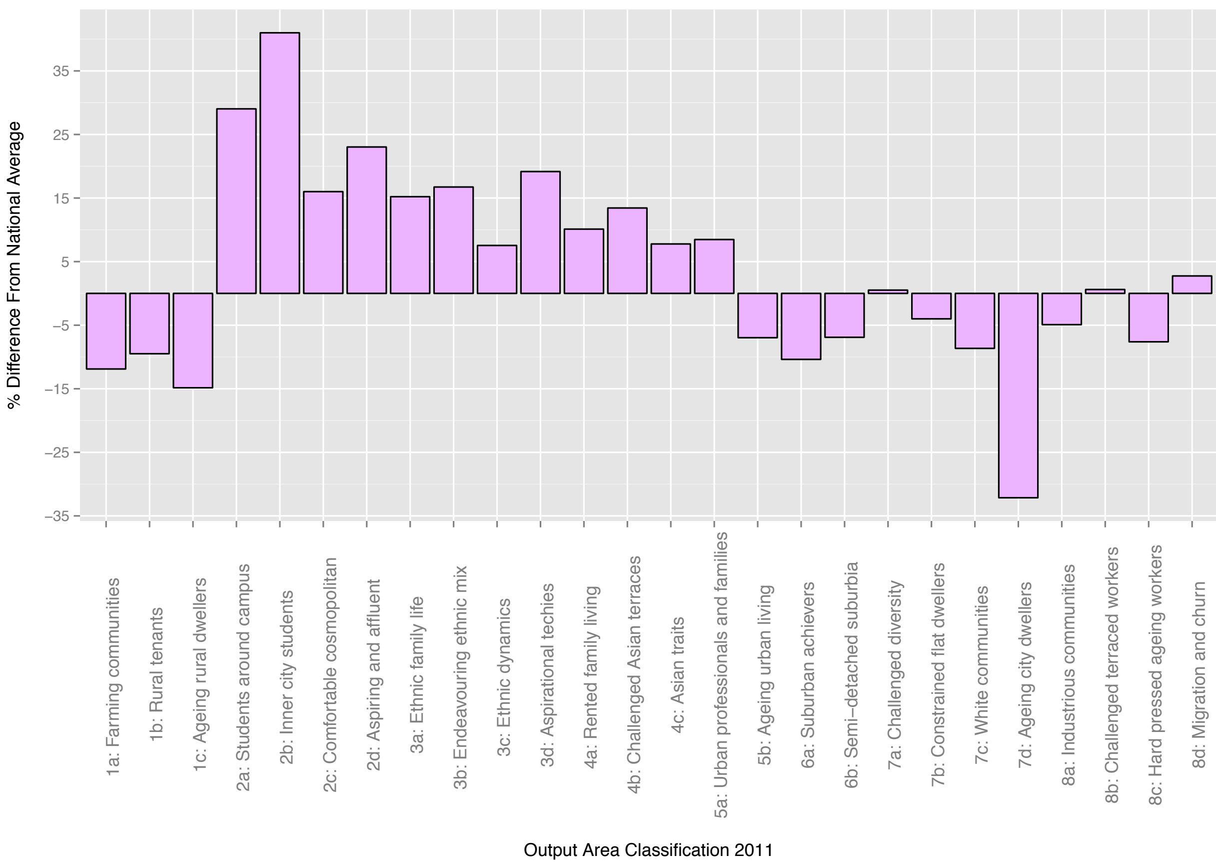


Figure 4.11: Rates of Engagement by OAC Group: Domain 5: Access Patterns: QC1b: Access Internet while travelling

To examine trends in access patterns, OXIS question QC1b (do you currently access the Internet while travelling, through a mobile or wireless dongle) was profiled by OAC. Figure 4.11 shows the output of this analysis. The national average for this question is 48.6%. Results broadly mirror those of mobile Internet use (QH12k), likely because most users who access the Internet on the move do so through a smartphone as opposed to a wireless dongle or modem. As would be expected, groups comprised mainly of students and young professionals display above average rates of engagement in this domain. Rural groups fall below the national average, likely due to poorer network coverage in these areas. Groups comprised mainly of elderly or retired populations also record lower than average rates, however, this is to be expected as these groups are frequently reported to be less engaged with next generation technology.

Domain 6: Commercial Applications

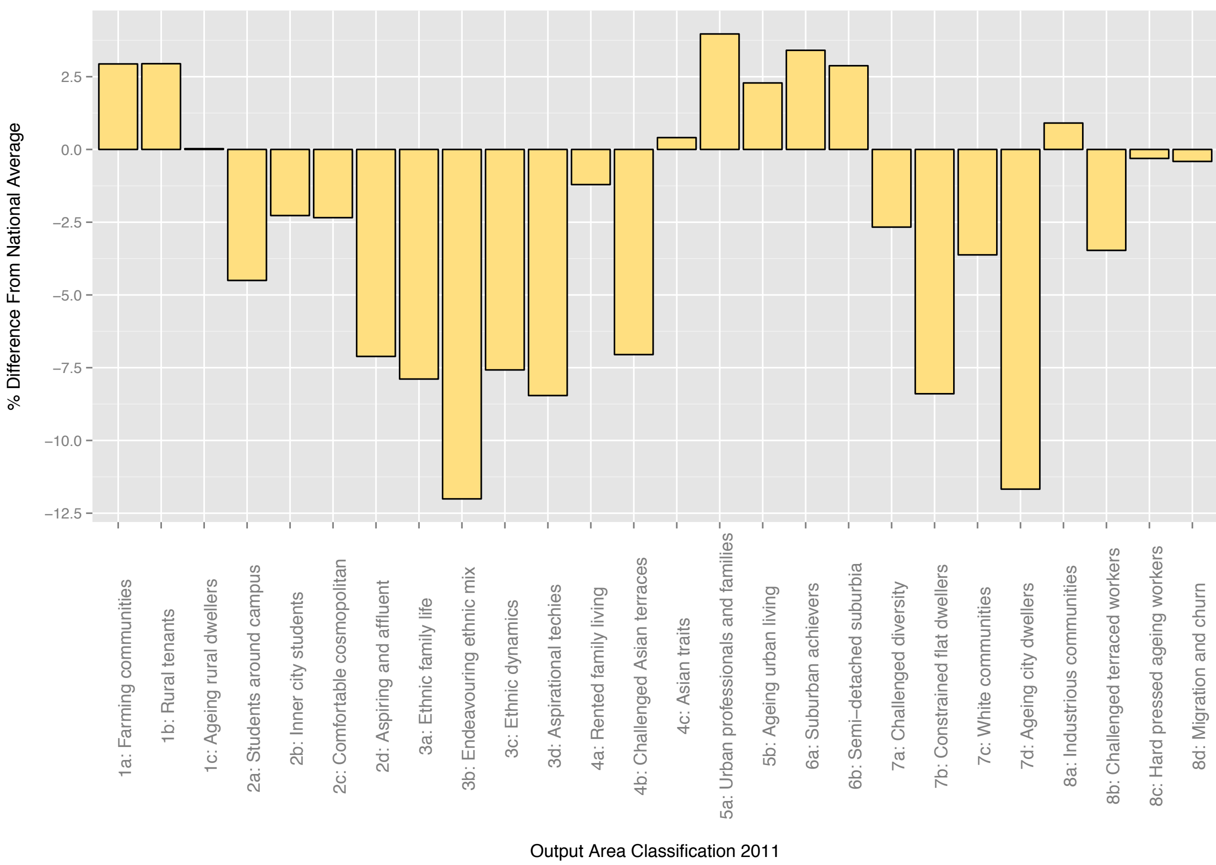


Figure 4.12: Rates of Engagement by OAC Group: Domain 6: Commercial Applications: QC30b: buying online

Levels of engagement in online commercial applications were profiled by OAC to establish the propensity of different groups for the ordering of goods and services online (see Figure 4.12). OXIS question QC30b (buying online) was chosen for profiling. Rates reflect those estimated to be ‘frequent’ users (i.e. buying online at least monthly), this suggests that 53.5% of Internet users are classed as frequent online buyers and as such, analysis shows how OAC groups deviate from this national average. It is apparent that predominantly urban groups (in particular those with higher proportions of ethnic minority residents), record lower than average rates for frequent online shopping. This may be due to the increased provision, choice and convenience of stores in the local area or perhaps where ethnicity is a limiting factor to engagement due to a lack of trust or understanding. While this would broadly support some literature presented in Chapter 2, Section 2.4.1, mainly those studies by Wilson et al. (2003) and Prieger and Hu (2008), it should not be implied that the aggregate relationships presented here are directly applicable at an individual level, as that would represent an ecological fallacy, and there are likely to be exceptions to these observations. In addition, groups with prominent student populations also display lower rates of engagement with online ordering, again this may be due in part to the increased provision of local retail, or may extend to encompass factors such as constrained income. Again these observations may not be directly applicable beyond the aggregate level. Conversely, some rural groups show higher than average rates of engagement, most likely due to a more limited provision of local retail and a greater requirement to travel to retail centres. Internet shopping would, for these groups, logically provide a more cost effective means of purchasing goods, echoing those benefits of rural broadband access described by the Government (DCMS, 2009; HM Treasury, 2012). Suburban areas also display some of the highest rates, in particular, groups; 5a: Urban Professionals and Families, 6a: Suburban Achievers and 6b: Semi-detached Suburbia. likely a result of these groups being comprised mainly of affluent, educated professionals who are more likely to be experienced Internet users.

Determining Rates of Personal Access

In terms of revealing rates of personal-level use of the Internet across OAC groups (as opposed to those measure of household-level access as presented earlier), OXIS question QH13 (detailing current, ex, and non users) was selected. Figures 4.13 and 4.14 detail the rates of current Internet users and non-users respectively by OAC group. In terms of current Internet users it is apparent that those groups comprised mainly of younger, educated professionals and student populations in urban settings record the highest rates at up to 14.8% above the national average (78.4% for current Internet users). Rurality appears to affect rates of personal use, with those groups that are predominantly rural recording lower than average rates in all instances, up to 5.8% below the national average in group 1c: Ageing Rural Dwellers, where age is likely interacting with rurality to constrain Internet usage. Below average rates are also apparent in those groups comprised mainly of elderly populations, as has been a recurring theme through this analysis. Group 7d: Ageing

City Dwellers records the lowest rate of personal Internet use at 21% below the national average, further elucidating a correspondence between areas with ageing populations and aggregate levels of engagement with the Internet and associated technologies.

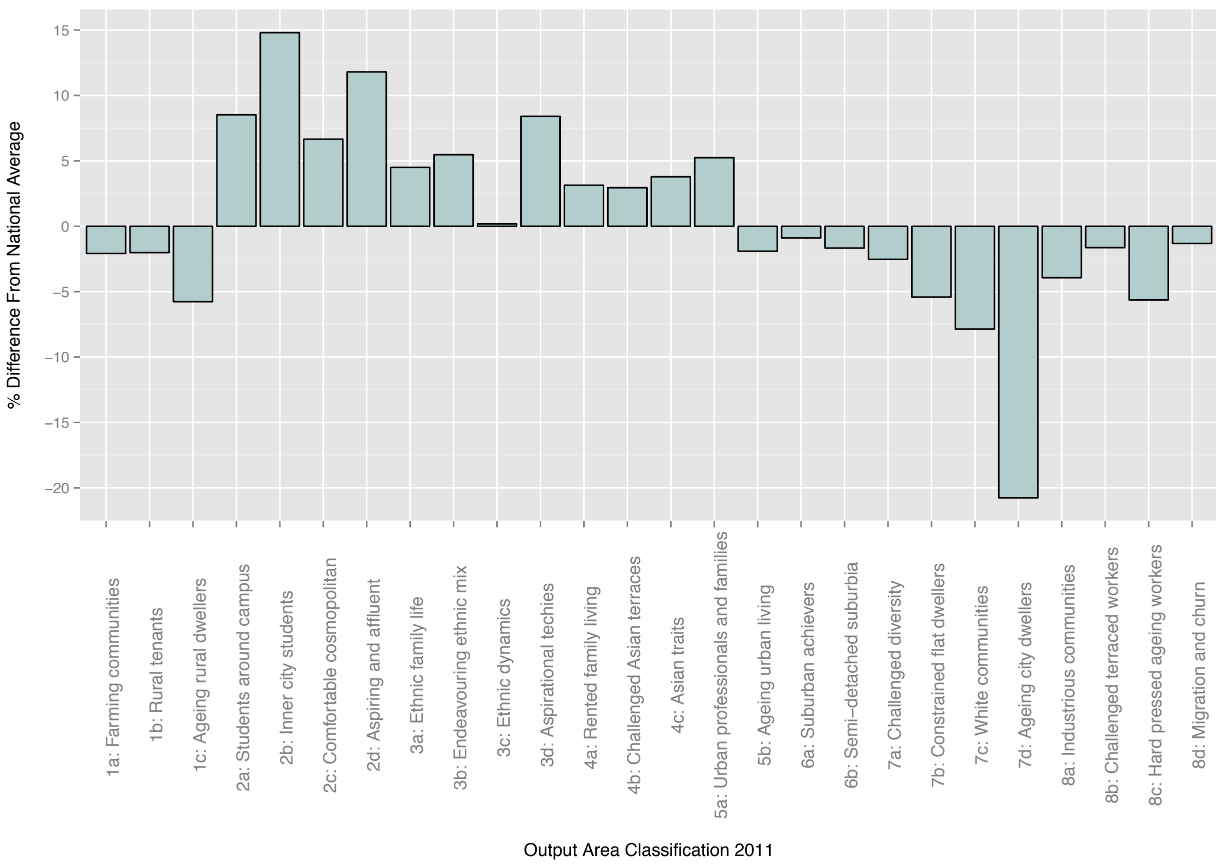


Figure 4.13: Rates of Engagement by OAC Group: Determining Rates of Personal Access: Current Internet Users

As would be expected, rates of Internet non use by OAC groups are broadly the inverse of rates of current use, with those groups that have been shown to be less engaged through profiling generally displaying higher rates of non-use than the national average (18.1% in this case). Those groups that have more elderly constituent populations unsurprisingly record the highest rates of Internet non-use. Group 7d: Ageing City Dwellers, record the highest rate at 84% above the national average, again suggesting correspondence between areas with ageing populations and limited engagement online. Interestingly, groups that exhibit higher levels of material deprivation and groups that have lower proportions of professionals, generally display higher rates of Internet non use, perhaps where there is a lesser necessity for access or where the costs associated with access are a limiting factor, such correspondence would broadly support the wider literature, in particular those studies presented by Longley (2003), Longley et al. (2008) and Longley and Singleton (2009b).

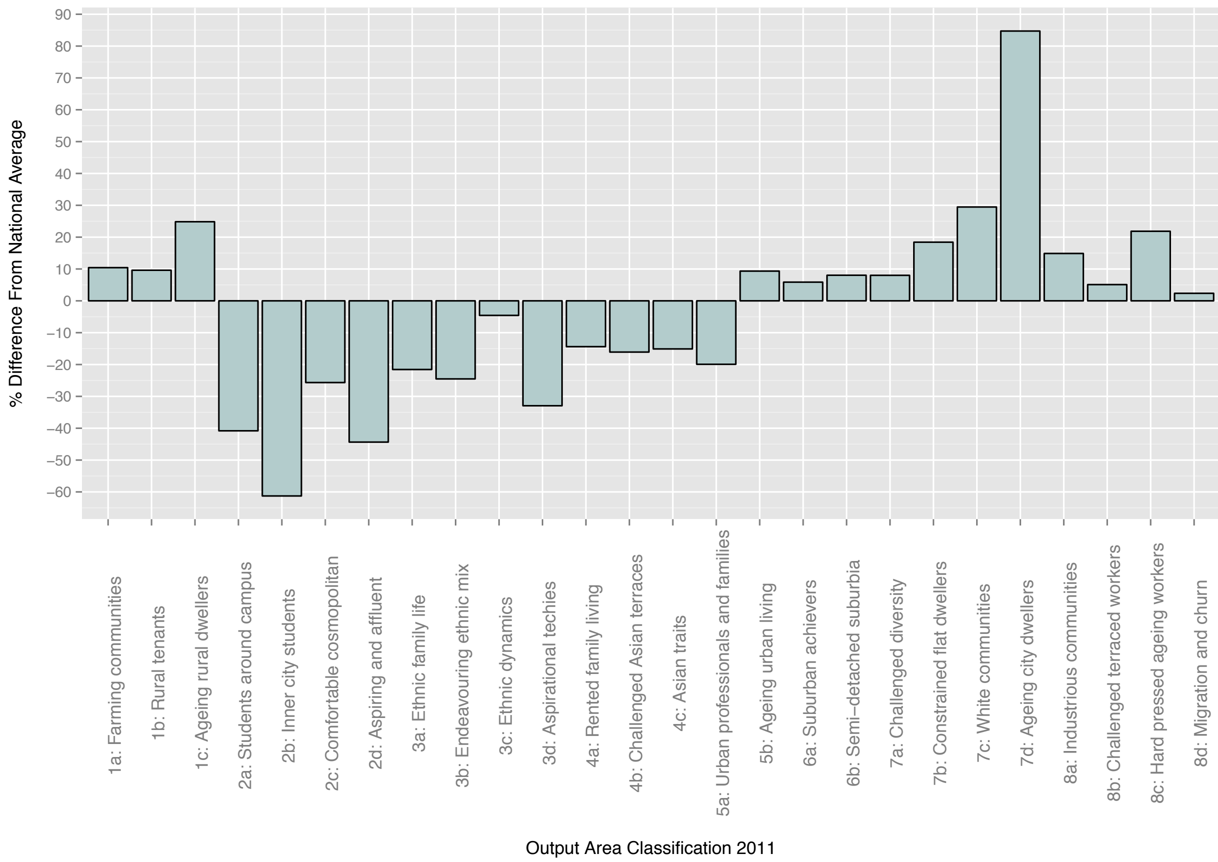
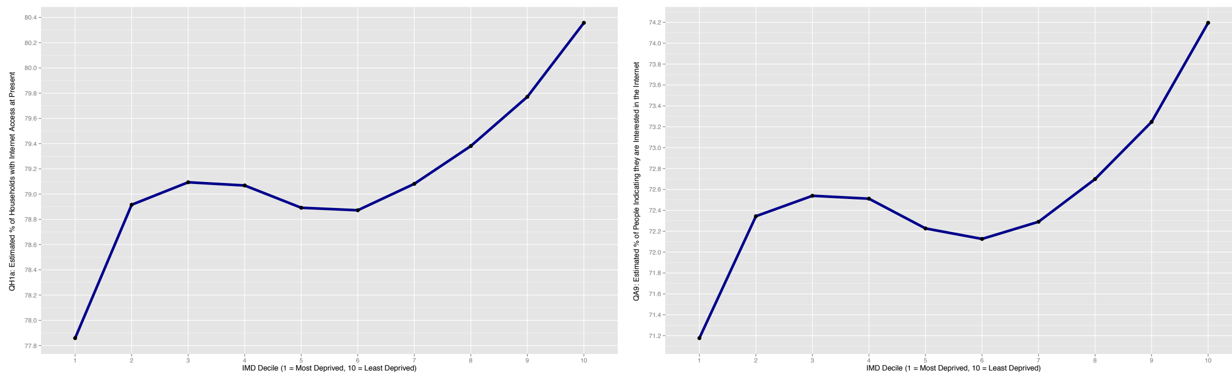


Figure 4.14: Rates of Engagement by OAC Group: Determining Rates of Personal Access: Internet Non Users

4.6.2 Material Deprivation and Internet Behaviours

High levels of material deprivation and the component factors such as low educational attainment, low income or unemployment and limiting long-term illness have previously been shown to limit engagement with the Internet (Selwyn, 2006; James, 2007; Longley and Singleton, 2009b) As such, when profiling rates of use and engagement with the Internet it would be logical to expect more ubiquitous interaction in those areas that are less materially deprived. In order to identify such correspondence, estimates were profiled by the Indices of Multiple Deprivation (IMD) to investigate how rates of engagement differ across IMD deciles. IMD has been used for profiling earlier in this thesis and represents a relative measure of deprivation at Lower Super Output Area (LSOA) level across England (see Chapter 3, Section 3.3.2). In line with previous analysis, the composite measure, which is a rank of each LSOA in the country, ranging from 1 (most deprived) to 32,482 (least deprived) was used for profiling. For this analysis, a subset of OXIS questions for which national estimates were produced was profiled by deciles of the IMD. In the following analysis decile 1 represents the most deprived areas and decile 10 the least deprived.

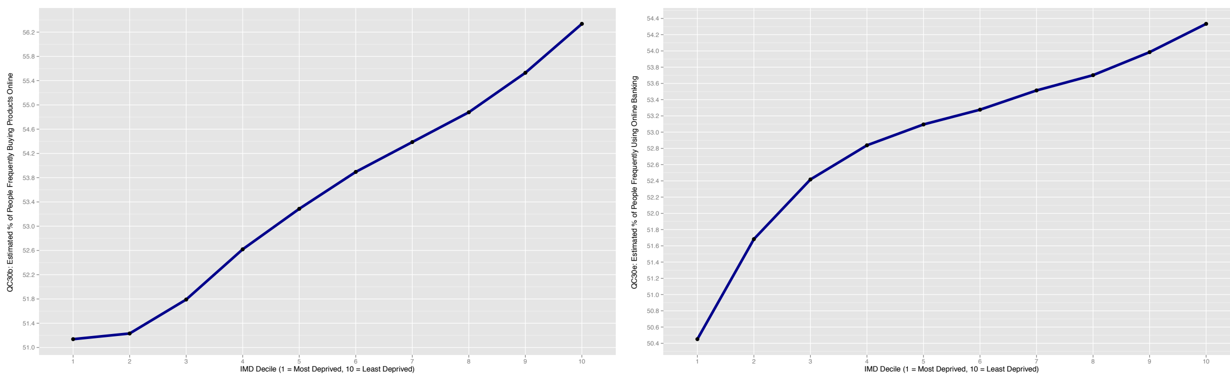


(a) QH1a

(b) QA9

Figure 4.15: OXIS Questions QH1a (Estimated % of Households with Internet Access at Present) and QA9 (Estimated % of People Indicating they are Interested in the Internet) by IMD Decile

Profiling the estimated percentage of households with Internet access at present by IMD decile reveals that those areas which are more deprived show, on aggregate, lower levels of household Internet access. Conversely, those areas that are less deprived (higher deciles of the IMD) show higher rates of household Internet access. Such patterns would be expected given deprivation has previously been discussed as a limiting factor in engagement with the Internet. It should, however, be noted that although there are differences between deciles of the IMD, these are relatively minor at this aggregate scale. Each decile of the IMD is comprised of around 3200 LSOAs, and when averages are calculated within these aggregate groups, patterns can be significantly smoothed. Figure 4.15 shows OXIS questions relating to household access (QH1a) and interest (QA9) profiled by IMD decile. Both show very similar patterns, which would be expected given it is unlikely for an individual or household to pay for broadband access in the home, if there is no interest in accessing the Internet. It is particularly notable that areas ranked in deciles 1 to 3 of the IMD, which are typically located in deprived (but well connected) urban areas, still display lower rates of engagement (In terms of both access and interest) than the most affluent deciles, which are more mixed in terms of urban/rural bias. This would suggest that despite the higher availability of infrastructure, deprivation may still prevail as a limiting factor to engagement. Rates appear to dip slightly across the middle deciles of the IMD, possibly because these areas are comprised of more mixed population structures and socio-economic groups that may display less homogenous patterns of engagement. The differences in rates between the most deprived and the least deprived deciles are 2.5% for household Internet access (QH1a) and 3% for interest in the Internet (QA9). At such high levels of aggregation it is encouraging to see this level of differentiation.

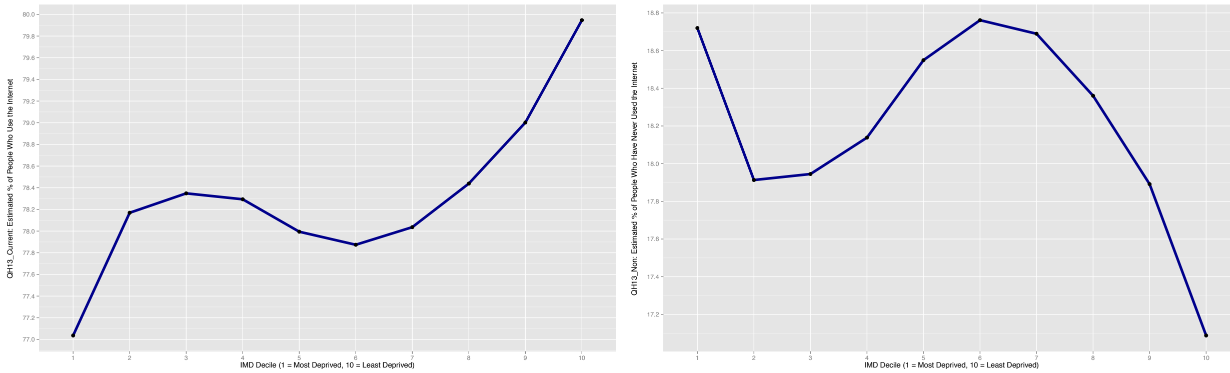


(a) QC30b

(b) QC30e

Figure 4.16: OXIS Questions QC30b (Estimated % of People Frequently Buying Products Online) and QC30e (Estimated % of People Frequently Using Online Banking) by IMD Decile

In terms of rates of engagement with common Internet tasks such as online shopping and Internet banking, more linear correspondence between the proportion of people engaged and levels of deprivation is apparent. Figure 4.16 shows rates for OXIS questions QC30b (buying products online) and QC30e (use of Internet banking). It is apparent that rates of both online shopping and Internet banking are lower in those deciles which rank highly in terms of deprivation, rising steadily across deciles as prevailing levels of deprivation decrease. In terms of buying products online, rates of engagement are around 5.2% lower in the most deprived decile that the least deprived decile. Similar correspondence is evident when profiling engagement with online banking; here rates of engagement in the most deprived decile are around 3.9% lower than the least deprived. It is important to consider, however, that the low rates of engagement with these applications in the most deprived deciles are unlikely to be entirely the result of prevailing levels of deprivation. As mentioned previously, highly deprived areas tend to be located in cities and other large conurbations where the availability of 'bricks and mortar' retailing and services such as in-branch banking is higher, thus offering populations a feasible alternative to online equivalents. Such availability is typically less present in rural and fringe areas, which typically score lower in terms of deprivation, which may influence rates of online engagement. Despite evidence to support interaction of multiple influencing factors, there still remains a clear association between online engagement and prevailing levels of deprivation at the aggregate scale.



(a) QH13_Current (b) QH13_Non

Figure 4.17: OXIS Questions QH13_Current (Estimated % of People Using the Internet) and QH13_Non (Estimated % of People Who Have Never Used the Internet) by IMD Decile

Rates of current Internet users are lowest in the most deprived deciles of the IMD and highest in the least deprived (see Figure 4.17 (a)), with a difference between decile 1 and decile 10 of around 2.9%. However, this relationship is not as linear as those presented previously. It is evident that rates decline again across deciles 5, 6 and 7 before increasing across the least deprived deciles (8, 9 and 10). Deciles 3 and 4 record rates for current Internet users closest to the national average (78.3% nationally, as reported in Figure 4.13). It may be that rates are high, even across more deprived deciles of the IMD, as these areas are typically located within major urban conurbations and are typically populated with a higher proportion of young professionals and students, who have been shown through previous analysis to be heavily engaged online. The decline in rates of engagement across the middle deciles of the IMD is again possibly due to these areas being comprised of more mixed population structures and socio-economic groups that may display less homogenous patterns of engagement.

In terms of non use (see Figure 4.17 (b)) rates are broadly the inverse of those for current users, i.e. high in the most deprived decile before falling across the first third of the IMD, likely because these areas are typically urban and associated with more ubiquitous Internet access. The highest rate of Internet non-use in this instance is not found in decile 1 but in decile 6 of the IMD. Here the prevailing rate of Internet non-use is 18.76%, marginally higher than decile 1 (18.72%) and around 0.6% above the national average of 18.1%. Between deciles 6 and 10, as prevailing levels of deprivation decrease, so do rates of non-use. Decile 10 records the lowest rate of Internet non-use at 17.09%, around 1% below the national average and 1.63% below the most deprived decile.

4.6.3 Urban Versus Rural Behaviours

Rural-Urban classification has been used in previous chapters of this thesis for the purpose of profiling and represents a useful tool to analyse aggregate spatial context in large datasets (see Chapter 3, Section 3.3.2 where the consolidated 8-class version for LSOAs is utilised). In this section, data relating to Internet engagement, estimated at the national extent from the OXIS, are profiled by these measures to reveal any underlying trends. Given performance has been shown previously to be differentiated across different Rural/Urban classes, the purpose of this analysis was to see if similar disparities are apparent when measures of Internet use and engagement were profiled. The ONS Rural-Urban Classification (RUC) was used to profile estimates across a measure of built environment and population density characteristics. The RUC used here is the 2011 version for small area geographies which categorises a range of statistical and administrative units on the basis of physical settlement and related characteristics, here the 10-class version for Output Areas is used. The OA level classification is the most disaggregate version of the RUC and its structure is presented in Table 4.23.

Table 4.23: 2011 ONS Rural-Urban Classification for Small Area Geographies

Urban/Rural	Conurbation/Context	Class	% of OAs Nationally
Urban	Major Conurbation	A1	32.6
Urban	Minor Conurbation	B1	3.5
Urban	City and Town	C1	44.7
Urban	City and Town in a Sparse Setting	C2	0.3
Rural	Town and Fringe	D1	8.7
Rural	Town and Fringe in a Sparse Setting	D2	0.6
Rural	Village	E1	5.3
Rural	Village in a Sparse Setting	E2	0.6
Rural	Hamlets and Isolated Dwellings	F1	3.3
Rural	Hamlets and Isolated Dwellings in a Sparse Setting	F2	0.5

In terms of household Internet access (see Table 4.24) it is apparent that rates are among the highest in major urban conurbations, here rates are around 1.3% above the national average of 79.2%, most likely as a result of strong infrastructure provision and higher proportions of younger professional and student populations who typically show higher levels of engagement online. Rates drop slightly below the national average for minor urban and urban city and town conurbations, particularly so for those that are sparsely populated. Sparsely populated rural town and fringe conurbations record the lowest rate of household Internet access at 75%, around 4.2% below the national average, however, this class accounts for just 1,044 OAs (0.6%) nationally. Interestingly, the most isolated classes of rural hamlets and isolated dwellings record above average rates for household Internet access, at up to 1.9% above the national average for the non-sparse variant, which accounts for 3.3% of all OAs nationally. These higher rates may reflect Internet access being more of a necessity in rural households given the likely constraints of the locality. In these scenarios Internet access could save rurally isolated households travel, time and money by engendering benefits in terms of online ordering, banking, bill and account management, communication and increased consumer choice, broadly supporting the case for rural broadband access as set out by the Government (DCMS, 2009; HM Treasury, 2012).

Table 4.24: OXIS Estimates: QH1a: % of Households with Internet Access at Present by ONS Rural-Urban Classification

Output Area Class	Rate (%)	Deviation from National Average (%)
Urban Major Conurbation	80.5	+1.3
Urban Minor Conurbation	77.7	-1.5
Urban City and Town	78.5	-0.7
Urban City and Town in a Sparse Setting	75.5	-3.7
Rural Town and Fringe	77.5	-1.7
Rural Town and Fringe in a Sparse Setting	75.0	-4.2
Rural Village	79.8	-0.6
Rural Village in a Sparse Setting	77.8	-1.4
Rural Hamlets and Isolated Dwellings	81.1	+1.9
Rural Hamlets and Isolated Dwellings in a Sparse Setting	79.7	+0.5

These rates may suggest that although a performance divide exits between urban and rural areas (see Chapter 3, Section 3.3.2, Table 3.5), an access divide is less apparent, with rates of household Internet access in some rural areas surpassing those of some urban counterparts. This echoes findings presented previously, whereby rural OAC groups were shown to have a higher propensity for testing their broadband connection speed than some urban OAC groups (see Chapter 3, Section 3.3.2, Figure 3.6). Such evidence supports the notion that Internet access in rural areas is of high importance.

Table 4.25 details the rates of individuals who would likely indicate that they are interested in the Internet by Rural-Urban Classification. Here the only areas to exceed the indicated national average of 72.6% are major urban conurbations. The average rate for these areas is 1.6% above the national average and this class accounts for 32.6% of all OAs nationally. The higher rates are, again likely a result of younger professional and student populations in these areas coupled with

Table 4.25: OXIS Estimates: QA9: % of Individuals Indicating they are Interested in the Internet by ONS Rural-Urban Classification

Output Area Class	Rate (%)	Deviation from National Average (%)
Urban Major Conurbation	74.2	+1.6
Urban Minor Conurbation	71.4	-1.2
Urban City and Town	72.2	-0.4
Urban City and Town in a Sparse Setting	68.9	-3.7
Rural Town and Fringe	70.6	-2.0
Rural Town and Fringe in a Sparse Setting	67.9	-4.7
Rural Village	71.5	-1.1
Rural Village in a Sparse Setting	69.4	-3.2
Rural Hamlets and Isolated Dwellings	72.5	-0.1
Rural Hamlets and Isolated Dwellings in a Sparse Setting	70.9	-1.7

better infrastructure. As would be expected, RUC classes with lower levels of household access also show lower levels of interest. The exception to this rule, however, is the rural hamlets and isolated dwellings classes, both sparse and non-sparse, as these classes display higher than average household access, but lower than average interest, although they have not deviated as far as other, predominantly urban or rural town and fringe areas from the national average. As with rates of household access, all sparse classes record lower rates than their non-sparse counterparts, indicating that infrastructure provision (which is likely to be poorer in sparsely populated areas) shows correspondence with uptake and engagement.

To analyse engagement with commercial applications, OXIS questions QC30b (rates of online purchasing (Table 4.26)) and QC30e (rates of online banking (Table 4.27)) are also analysed by RUC. Profiling rates of online purchasing reveals a number of interesting patterns, most notably that rates in major urban conurbations are lower than the national average, and alongside the sparse urban city and town group, they are the lowest rates when profiling by RUC. Lower than average rates in major urban conurbations may likely be a result of the increased supply and convenience of local retail centres, which would be a feasible alternative to online shopping. Areas that are rurally isolated such as villages, hamlets and isolated dwellings, both in sparse and non-sparse contexts display higher than average rates for frequent online purchasing. The highest rate is recorded in the rural hamlets and isolated dwellings class at 2.6% above the national average of 53.5%. It would be logical to assume that this is a result of the Internet offering increased consumer choice and more competitive pricing than those retail centres available locally. Delivery of items also saves rurally isolated populations the expense of travel, which may be higher (given longer travel times) or inconvenient (if private transport is not available, or the provision of public transport is poor).

Table 4.26: OXIS Estimates: QC30b: % of Individuals Frequently Buying Products Online by ONS Rural-Urban Classification

Output Area Class	Rate (%)	Deviation from National Average (%)
Urban Major Conurbation	52.3	-1.2
Urban Minor Conurbation	53.5	0.0
Urban City and Town	53.9	+0.4
Urban City and Town in a Sparse Setting	52.3	-1.2
Rural Town and Fringe	54.6	+1.1
Rural Town and Fringe in a Sparse Setting	52.8	-0.7
Rural Village	55.4	+1.9
Rural Village in a Sparse Setting	53.9	+0.4
Rural Hamlets and Isolated Dwellings	56.1	+2.6
Rural Hamlets and Isolated Dwellings in a Sparse Setting	55.0	+1.5

Table 4.27: OXIS Estimates: QC30e: % of Individuals Frequently Using Online Banking by ONS Rural-Urban Classification

Output Area Class	Rate (%)	Deviation from National Average (%)
Urban Major Conurbation	53.2	+0.3
Urban Minor Conurbation	51.9	-1.0
Urban City and Town	52.7	-0.2
Urban City and Town in a Sparse Setting	50.8	-2.1
Rural Town and Fringe	52.8	-0.1
Rural Town and Fringe in a Sparse Setting	51.2	-1.7
Rural Village	53.4	+0.5
Rural Village in a Sparse Setting	52.2	-0.7
Rural Hamlets and Isolated Dwellings	53.7	+0.8
Rural Hamlets and Isolated Dwellings in a Sparse Setting	53.1	+0.2

In terms of online banking, rates of engagement are again above average in the most rurally isolated areas with the exception of rural town and fringe classes, which fall below the national average, likely due to an adequate provision of these services in local retail centres. Major urban conurbations are slightly above the national average in terms of engagement with online banking, possibly as a result of constituent populations who are more engaged with the Internet and better infrastructure, particularly mobile, which would support online banking through smartphones and cellular enabled tablets. For the rural classes that display above average rates of engagement, this may also relate to poorer service provision within these areas (i.e. fewer high street branches). Again, the pattern of lower rates of engagement in those classes categorised as sparse is evident when compared to non-sparse equivalents.

Table 4.28: OXIS Estimates: QH13_Current: % of Individuals Using the Internet by ONS Rural-Urban Classification

Output Area Class	Rate (%)	Deviation from National Average (%)
Urban Major Conurbation	80.1	+1.7
Urban Minor Conurbation	77.1	-1.3
Urban City and Town	77.9	-0.5
Urban City and Town in a Sparse Setting	74.4	-4.0
Rural Town and Fringe	76.4	-2.0
Rural Town and Fringe in a Sparse Setting	73.5	-4.9
Rural Village	77.3	-1.1
Rural Village in a Sparse Setting	75.1	-3.3
Rural Hamlets and Isolated Dwellings	78.2	-0.2
Rural Hamlets and Isolated Dwellings in a Sparse Setting	76.6	-1.8

In terms of rates of Internet use (Table 4.28), it is unsurprising to find that major urban areas have the highest proportion of current Internet users at 80.1%, 1.7% above the national average of 78.4%. This is likely a result of younger, heavily engaged professional and student populations in these areas alongside increased infrastructure provision that enables more ubiquitous access. Through profiling it is apparent that all rural areas fall below the national average in terms of current Internet users, with those areas that are categorised as rural and sparse recording significantly lower than average rates. The lowest rate of current Internet users is recorded in the sparsely populated rural town and fringe class at 4.9% below the national average and is likely attributed to the demographic composition and possible infrastructure limitations of these areas. It is interesting to note that although hamlets and isolated dwellings (both sparse and non-sparse) have been shown to display higher than average levels of household Internet access (see Table 4.24), these areas fall below the national average in terms of current Internet users. This suggests correspondence between household access and personal use of the Internet in these areas is more complex, and perhaps indicative of instances where households have Internet access, but not all members of the household are Internet users.

Profiling non-use by RUC (Table 4.29) broadly reveals the inverse of the patterns described above.

Table 4.29: OXIS Estimates: QH13_Non: % of Individuals Who Have Never Used the Internet by ONS Rural-Urban Classification

Output Area Class	Rate (%)	Deviation from National Average (%)
Urban Major Conurbation	16.5	-1.6
Urban Minor Conurbation	19.1	+1.0
Urban City and Town	18.5	+0.4
Urban City and Town in a Sparse Setting	21.7	+3.6
Rural Town and Fringe	20.2	+2.1
Rural Town and Fringe in a Sparse Setting	22.8	+4.7
Rural Village	19.5	+1.4
Rural Village in a Sparse Setting	21.4	+3.3
Rural Hamlets and Isolated Dwellings	18.6	+0.5
Rural Hamlets and Isolated Dwellings in a Sparse Setting	20.1	+2.0

Major urban areas record the lowest rate of Internet non-users at 16.5%, 1.6% below the national average of 18.1%. Those areas that are rural, and in particular, those that are sparsely populated, record higher than average rates of Internet non-use, the highest average rates being recorded in sparsely populated rural town and fringe areas at 22.8%, 4.7% above the national average. Although these rates are high, this class accounts for a relatively small proportion of all OAs nationally at just 0.6%. The RUC class that accounts for the largest proportion of OAs nationally (urban city and town at 44.7%) also records a higher than average rate of non-users, albeit significantly closer to the national average at 18.5%. The results of profiling rates of current and non-users of the Internet in this way provide evidence to support the notion that rurality affects engagement.

4.6.4 Visualising Patterns of Engagement and Exclusion

In addition to assessing the distribution of estimates by measures of socio-spatial structure, deprivation and rurality, estimates for a subset of OXIS questions were visualised to explore geographic patterns. In this section, estimates for three OXIS questions are profiled; QH13_Current (current Internet users), QH13_Non (Internet non-users) and QC30b (online shopping). Rates of current and non-use are visualised at two different geographic extents, firstly, the city of Liverpool to explore geographic disparities in a local context and secondly at the national scale to assess the current geography of digital exclusion. Finally, rates of online shopping are visualised for London, to highlight how engagement patterns can differ across a major urban conurbation, and expand on analysis presented in previous sections.

Mapping the estimated percentage of current Internet users by Output Area in the local context of

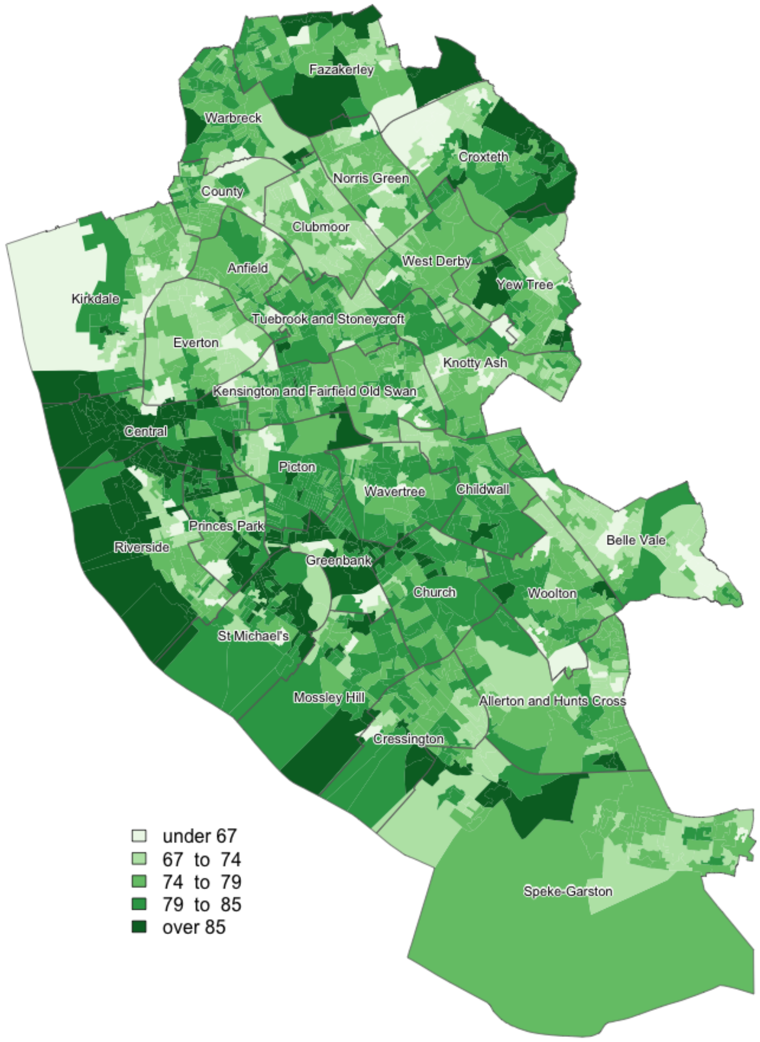


Figure 4.18: OXIS Estimates: QH13_Current: % of Individuals Using the Internet: Liverpool OAs

Liverpool (see Figure 4.18) reveals a number of interesting patterns that broadly reflect the socio-economic geography of the city. It is evident that the city centre wards of Central and Riverside have large proportions of current Internet users, most likely because these areas are typically inhabited by young professionals and student populations, who generally show higher levels of engagement with most aspects of the Internet. High proportions of current Internet users are also apparent across the wards of St Michaels and Mossley Hill, both are affluent suburbs located to the south of the city and inhabited by a mix of young and established professionals, families and the city's wealthy elite, groups who have been shown to be more heavily engaged with technology and the Internet through previous analyses. Higher rates are also evident across the wards of Picton and Greenbank; these areas are predominantly student communities and previous analyses have shown student groups to be heavily engaged across most aspects of engagement with the Internet. Lower proportions of current Internet users are evident across some of the city's more deprived areas such as Everton, one of the most deprived wards in the country, and an area typically representative of the city's more constrained resident populations. Other areas that rank highly in terms of deprivation such as Kirkdale, Anfield and Clubmoor, which have been the subjects of regeneration schemes in recent years, display more mixed patterns of engagement, but rates are generally lower than those areas in the more prosperous south of the city.

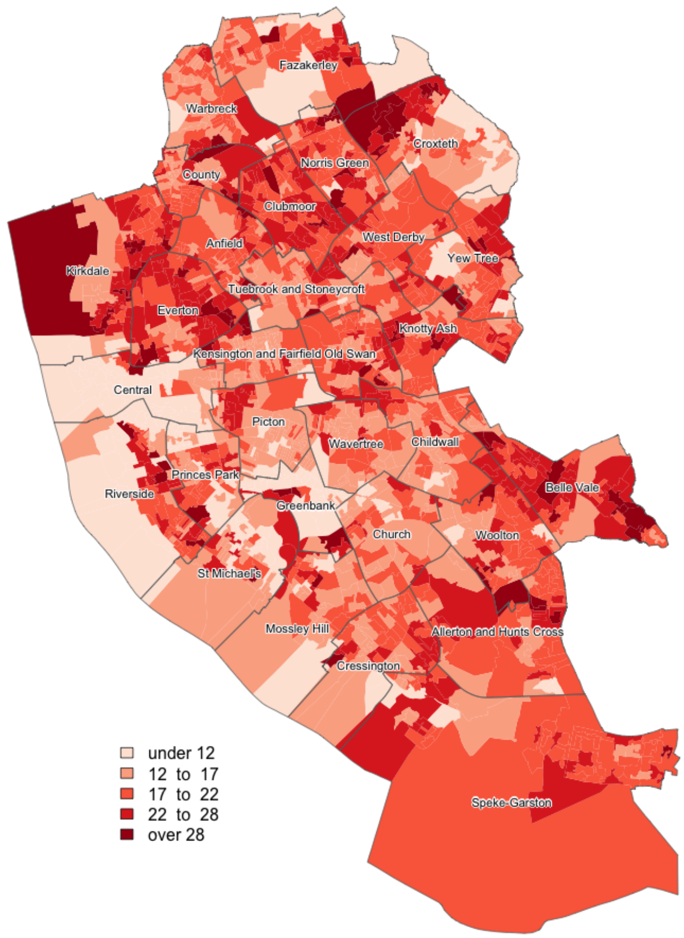


Figure 4.19: OXIS Estimates: QH13_Non: % of Individuals Who Have Never Used the Internet: Liverpool OAs

Comparing the geographic distribution of current Internet users to that of Internet non-users, patterns are broadly inverted (see Figure 4.19). Higher proportions of non-uses are evident in

Everton, for example, with some OAs recording over 28% of the population having never used the Internet, likely manifested through higher levels of material deprivation. Some wards, including Riverside and St Michaels, which generally display high proportions of current users, also display some pockets of high non-use. These results may appear anomalous but in fact represent small areas of high deprivation in the case of Riverside, clustering around the neighbourhoods of Toxteth, an area that has historically been extremely deprived. In the case of St Michaels, profiling by census age statistics suggested the pockets of high non-use represent areas with higher proportions of elderly residents, who generally show lower levels of engagement online.

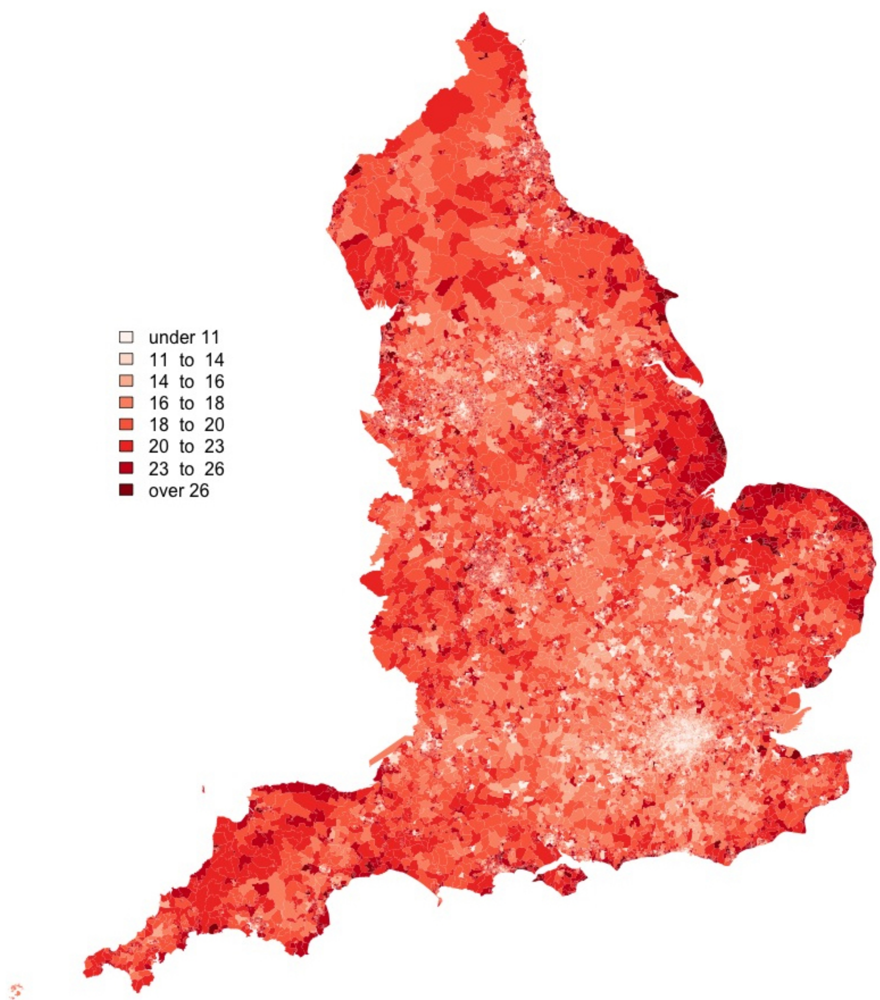


Figure 4.20: OXIS Estimates: QH13_Non: % of Individuals Who Have Never Used the Internet: National LSOAs

Exploring the geographic distribution of non-users at the national extent also reveals a number of distinct patterns (see Figure 4.20). Here rates are calculated at the aggregate level of the LSOA to aid visualisation. It is apparent that high levels of Internet non-use cluster around coastal areas such as the Wash, North Norfolk and much of the Lincolnshire coastline which show particularly high rates. Other coastal areas in predominantly rural regions such as Cornwall and the Lake District also record high rates of Internet non-use, possibly as a result of rurality becoming a barrier to access. However, in the case of Cornwall in particular, recent rollout of fibre optic services across the region (see Chapter 3, Section 3.3.4) may provide the necessary preconditions to see high rates of non-use decline in coming years. Although the areas highlighted are predominantly rural, coastal areas tend to have higher proportions of elderly and retired residents, which would likely affect prevailing levels of engagement with the Internet. Other predominantly rural, but non-coastal regions such as the Welsh and Scottish borders also record high levels of non-use, possibly as a result of rurality and the socio-economic characteristics of these areas' constituent populations, which are less likely to include high proportions of students or younger professionals. In terms of low rates of non-use, clusters appear around major urban conurbations, some major English cities are identifiable within the visualisation, London in particular, given high concentrations of low Internet non-use. These clusters of low non-use are expected, as previous sections have shown how urban areas benefit from enhanced infrastucture to enable more ubiquitous access, alongside younger, professional populations who typicallly show higher levels of engagement with the Internet.

Figure 4.21 visualises rates of online shopping in the Context of London. This analysis reveals a number of interesting patterns; most notably that large proportions of the inner city display comparatively lower propensity for online shopping than those areas on the outskirts. Although this may look anomalous given inner-city populations are generally highly engaged with the Internet, the rates may reflect, to some extent, the convenience, choice and availability of local retail. In locations such as central London it may be a lesser requirement to frequently shop online given local retail is convenient, competitive and easily accessed by an inexpensive, ubiquitous public transport network. Rates of online shopping do start to increase, however, further outside of central London. These increases are likely attributed to higher proportions of suburban population that have been shown through previous analysis to display higher levels of engagement with online shopping (see Figure 4.12). Although city centre retail is likely still accessible from these areas, it may not represent the most convenient option, with many individuals opting to shop online, possibly to reduce the need for travel. Figure 4.22 visualises rates of online food and grocery ordering. Here patterns are broadly the inverse of Figure 4.21, with the inner urban areas displaying comparatively higher rates than those on the suburban fringes of the city. This is likely due to the increased convenience of having groceries delivered in highly urban areas, where the provision of large supermarkets is more constrained, and private transportation is utilised to a lesser extent. The higher rates in the suburban fringe areas are likely attributed to greater availability of large supermarkets and higher rates of car ownership.

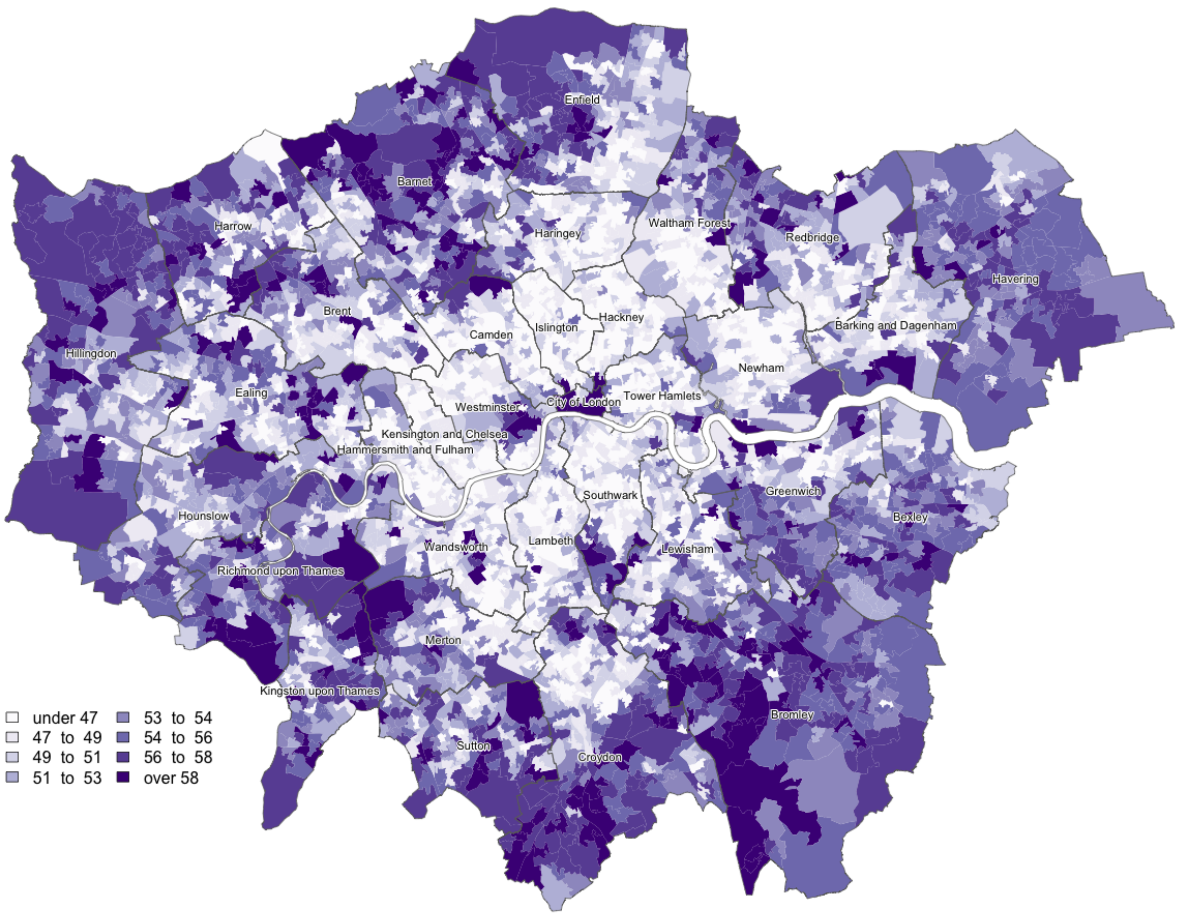


Figure 4.21: OXIS Estimates: QC30b: % of Individuals Frequently Buying Products Online: London LSOAs

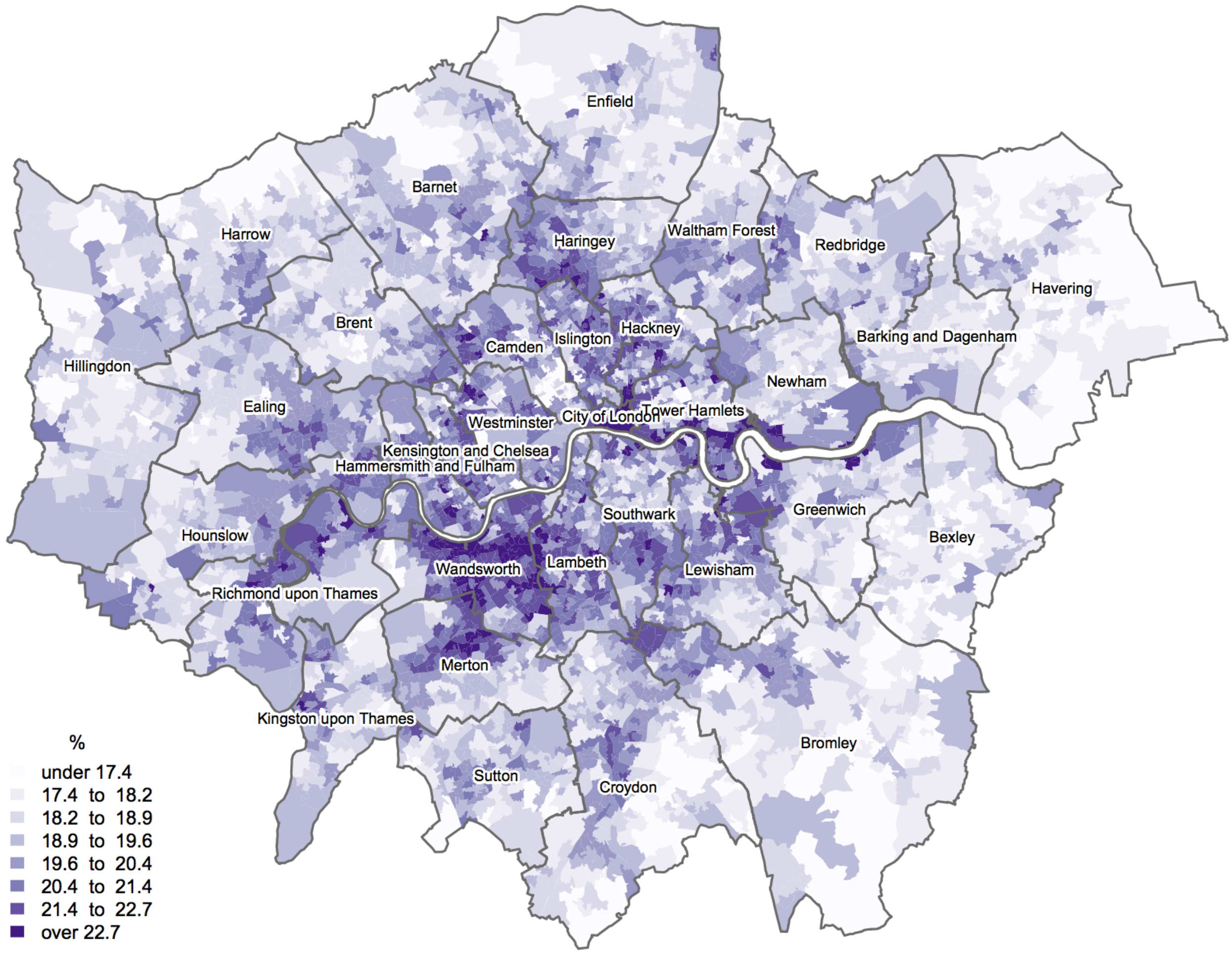


Figure 4.22: OXIS Estimates: QC30h: % of Individuals Frequently Ordering food or Groceries Online: London LSOAs

It should be noted, however, that these patterns are apparent though analysis of estimates modelled using domicile population statistics, and different patterns may occur if estimates for workday populations were able to be produced. Currently there are no Internet survey data (such as the OXIS) available by workday population, and as such, this analysis is not currently feasible.

4.7 Conclusions

This chapter has examined the current geography of Internet demand and Engagement through analysis and extrapolation of the 2013 Oxford Internet Survey (OXIS). Two methods through which estimates for the national extent were produced have been presented and the results have been profiled by indicators of socio-spatial structure, measures of rurality and levels of material deprivation. Results have also been visualised at the local and national extents, thus highlighting the complex geographies and differentiated levels of engagement across these univariate measures. While the OXIS is a valuable source of data for research in this domain, it does not have coverage at the national extent, instead, a representative sample of respondents are selected for interviews. Because the sample is representative of the general population and the results pertaining to each respondent are coded at the area level, it is possible to extrapolate the sample using different modelling techniques to obtain estimates of response for those areas that were unsampled in the original survey. The first method presented in this chapter sought to create estimates through ecological regression at the aggregate level of the LSOA, using census derived attributes as covariates in a series of stepwise regression models. It was, however, found that the aggregation of respondent data to this geographic level, and the small number of LSOAs that were sampled, were limiting the feasibility of this approach. Most of the models created using this method were of poor fit and as such this method was not used for the estimation of national rates. The second method used a data mining technique called Decision Tree Induction (DTI). This method was used to profile the OXIS at the respondent level, meaning no aggregation was necessary. Because the respondent data was not aggregated before analysis, the maximum level of granularity was retained. As with the ecological regression method, census derived attributes were used as covariates in the modelling process. The DTI method created sub-groups of the respondents using the covariates and calculated response rates for each. The sub-groups were then recreated using OA level Census data and the appropriate rates were fitted. Each sub-group rate was then weighted by the proportion of the target OA that sub-group represented in terms of the population. Finally, all the weighted estimates were summed to give an overall OA level estimate. The process was repeated for a total of 42 questions of interest from the OXIS. The estimates were then profiled to assess correspondence with a number of measures of socio-spatial structure, deprivation and rurality, as these methods had been shown to highlight interesting patterns in previous chapters of this thesis.

Profiling by OAC revealed significant correspondence between socio-spatial groups and prevailing levels of engagement with different domains of the Internet, including; seeking information, attitudes and perceptions, household access, mobile access, access patterns and commercial applications. In most cases it was found that those groups whose constituent populations are younger, urban, and have professional backgrounds, displayed the highest levels of engagement with most aspects of the Internet. In contrast, those groups whose constituent populations are elderly, constrained by deprivation, or working in employment sectors that would not require significant exposure to ICT, were generally less engaged across the multiple domains. OAC groups that are predominantly rural displayed more mixed engagement characteristics, with higher engagement rates in some domains, for example, online ordering and household access, but lower rates in domains such as mobile access and interest in the Internet, highlighting a potentially complex geography of engagement in rural areas.

Profiling estimates by measures of deprivation, in this case the Indices of Multiple Deprivation (IMD), revealed apparent correspondence between rates of engagement and prevailing levels of deprivation. Areas that displayed higher levels of deprivation typically recorded lower rates of engagement with the Internet. It was, however, noted that such correspondence was not always linear, with areas that experience more average levels of deprivation displaying more variable rates of engagement. This may be linked to more heavily differentiated population structures and socio-economic groups in these areas, which may result in less homogenous patterns of engagement at the aggregate level. In terms of rurality, higher rates of engagement generally corresponded to those areas that are within major urban conurbations. However, for some domains, correspondence between engagement and rurality was more heavily differentiated, for example shopping online appears to be more prevalent in areas that are rurally isolated and less prevalent in major urban conurbations. This may be linked to external factors, such as the provision of local retail, but again suggested a complex geography of engagement across different levels of rurality. Visualising the estimates at the local and national scales has highlighted the extent of such geographic disparities. With rates of Internet non-use, for example, clearly clustering around those areas that are predominantly rural or have more elderly constituent populations. In the local context of Liverpool, visualising levels of engagement revealed correspondence with areas of known high deprivation and local population structures, for example, areas with prominent student populations displaying higher rates of Internet users. Combined with the outputs of Chapter 4, the estimates created through this research represent a large number of complex univariate measures of Internet access, performance, demand and engagement that can be utilised for further analysis.

Chapter 5

From Univariate Measures to Multivariate Classification

5.1 Introduction

So far, this thesis has focused on the background and development of modern broadband Internet networks and how the geographies, access, and performance characteristics of these physical networks are differentiated by socio-spatial and spatio-temporal measures. In addition, disparities in engagement with the Internet have been examined in Chapter 4. The results of analyses have suggested that across each of the univariate measures examined so far, there are geographic and socio-economic disparities, that when collated, engender a complex geography of Internet access, performance, perception and engagement. Whilst the research presented in each of the empirical studies has made use of relevant and innovative sources of data, the outcomes currently remain independent. As such, there is an opportunity to create a holistic view of Internet access and engagement by combining the univariate outcomes of previous analyses to create a composite summarisation. Doing so would likely create a more simplistic overview of how these univariate measures may interact, to form natural groupings of access, performance, and engagement attributes by small area geography. Given the large amounts of data that have been utilised in previous chapters, the majority of which are geographically referenced at small area resolutions such as the OA or LSOA, geodemographic classification represents a method through which a composite measure of Internet use and engagement could be created, as these summarisation techniques are frequently developed for small area geographies to produce nested categorical typologies of complex socio-spatial characteristics (Harris et al., 2005). Furthermore, geodemographic classification in the form of the Output

Area Classification (OAC), has already been shown to be applicable in this research, assisting in the profiling of large datasets to reveal salient socio-spatial patterns.

5.1.1 Background to Geodemographic Classification

In the most basic sense, geodemographic classification can be described as 'the analysis of people by where they live' (Sleight, 1997), and involves the analysis of multidimensional attributes relating to the socio-economic and built environment characteristics of small geographic areas. Typically, clustering is used to form a typology that captures and summarises these multidimensional characteristics (Harris et al., 2005). Area classification has been a research theme for some time, within the UK, one of the first applications were Charles Booth's poverty maps of London in the early 1900s, which attempted to develop a socio-spatial segregation of the capital's neighbourhoods according to prevailing physical conditions and population characteristics. However, it was not until the late 1920's that area classification began to gather pace as a field of research. Significant studies include those by Burgess (1925), who presented the concentric zone model, a theoretical model to explain urban social structures, and Shevky and Bell (1955), who introduced the concept of 'Social Area Analysis'. Both studies are early examples of the use of census data for area analysis and formed the theoretical basis on which contemporary geodemographic research would develop. Research later diversified to incorporate larger volumes of census derived information and, in the domestic context, contemporary area classifications began to emerge such as the classification of census enumeration districts for inner London (Howard, 1969). The use of small area classification was also adopted by local authorities within the same period, with Liverpool City Council presenting a classification of the concentration of social problems as part of the Liverpool Social Malaise Study in 1969 (Liverpool City Council, 1969). During the 1970s, national classifications began to emerge at varying geographic scales such as the ward, parish and local authority, thus enabling the clustering of spatial units, primarily census enumeration districts, to build socio-spatial typologies (Webber, 1977). At this stage, commercial interest in geodemographics began to emerge, with permutations of those classifications built by academics acquired for the purposes of neighbourhood profiling and targeting (Batey and Brown, 1995). Within the contemporary period, geodeomographic classifications have continued to develop within both the commercial and academic sectors, and research has broadened to international application, with classifications built for many countries including the United States (Singleton and Spielman, 2013), Japan (Asai and Yano, 2001), Italy (Willis et al., 2010) and the Philippines (Ojo et al., 2013). A number of geodemographic classifications now dominate the commercial marketplace including CACI's 'ACORN'[1] and Experian's 'MOSAIC'[2]

[1] `http://acorn.caci.co.uk/`

[2] `http://www.experian.co.uk/marketing-services/products/mosaic-uk.html`

which have seen considerable success (Harris et al., 2005). Alongside the development of commercial classifications, a number of open source equivalents have been developed, including OAC, in the context of the UK, provided by the Office for National Statistics (Vickers and Rees, 2007). Despite widespread adoption of geodemographic classification, the methodological underpinnings have attracted criticism as being simplistic and ambiguous (Harris et al., 2005). The underlying theory on which the conceptual framework has developed relies heavily on the notion of homophily, a principal that assumes people tend to be similar to their friends, manifesting spatially as a general tendency for people to live in places with similar people (Alexiou and Singleton, 2015). This echoes Tobler's first law of geography in that 'Everything is related to everything else, but near things are more related than distant things' (Tobler, 1970), although it may be argued that in aggregating (often spatially disparate) small areas into clusters, it is assumed there are no contextual differences between them, and that a final cluster within a classification assumes that areas within it display the same characteristics, which may not be true given more granular analysis, representing an ecological fallacy. Despite such criticisms, the application of geodemographic classification remains widespread and contemporary application has extended to the domain of engagement with information and communication technologies. Longley et al. (2008) present a typology of the e-society, which aimed to classify each adult in Great Britain according to his or her likely level and manner of engagement with electronic technologies. In the current context, the classification is somewhat dated, given more recent technological advances, such as high speed broadband access and the increased presence of Internet enabled devices. Equally, the classification incorporated no contextual information relating to local infrastructure, instead relying on supplementary contextual data to infer where infrastructure constraints may be influencing behaviours. The study does, however, demonstrate that geodemographic classification can be an effective tool in summarising complex multi-dimensional measures related to engagement with technology. As such, this supports the case for a bespoke classification utilising the data available through research presented in this thesis. It is anticipated this would further enable the understanding of the complex geographic disparities embedded within access to, and engagement with, the Internet. To reiterate the views of Harris et al. (2005); 'Identifying geographic patterns or trends within societies is an important step towards understanding the process and phenomena that gave rise to those patterns in the first place'.

5.1.2 Conceptual Framework for Classification of Internet Use and Engagement

The multi-dimensionality of this issue is important, as analyses presented in previous chapters have identified a range of univariate outcomes that, if combined and summarised effectively, could represent a holistic, spatially disaggregate typology of Internet use and engagement. Reviewing the key analyses of previous chapters, it is apparent that there is a breadth of spatially referenced data that could be utilised for classification. Broadly, these data cover the domains of:

- Fixed-line infrastructure access
- Fixed-line broadband Internet performance
- Mobile infrastructure access
- Perceptions
- Access patterns
- Commercial applications
- Household level access
- Mobile access and usage
- Current Users, ex-users and non-users
- Demographic and contextual attributes

In terms of fixed-line infrastructure access, analysis has suggested disparities exist between different socio-spatial groups. Profiling by OAC revealed speed tests were not evenly distributed across all groups, suggesting that access in this respect is likely influenced by a number of factors, and forming a basis to support a conceptual framework for classification. Further analysis revealed that, as well as measures of access, the performance of broadband connections was unevenly distributed. In this case, performance was shown to have significant correspondence with measures of socio-spatial structure, rurality, deprivation and geographic context. Such correspondence was not limited to a snapshot of data, instead it was apparent that performance increases were also geographically disparate, broadly suggesting inequity in the prioritisation of infrastructure upgrades. In addition to fixed-line infrastructure, mobile infrastructure was also found to be unevenly distributed, with higher concentrations of base stations located in densely populated urban areas. It has also been noted that in some rural areas, where the average distances to the nearest mobile base station are very large, this would likely be a limiting factor to engagement. If mobile service is poor or unavailable in some rural areas, this will likely limit device ownership and this correspondence has been highlighted to some extent through OAC profiling, as those groups that are predominantly rural, recorded lower rates of mobile handset ownership and Internet use through mobile phones than those groups that are predominantly urban (see Chapter 4, Section 4.6.1). In terms of the data estimated from the OXIS, there are a significant number of univariate measures pertaining to; perceptions, access patterns, commercial applications, household access, mobile access and usage and personal access that, if summarised alongside measures of local Internet access and performance, would likely provide a rich typology of Internet use and engagement. Attributes across all domains extrapolated from the OXIS have been shown to display differentiation across measures of socio-spatial structure, deprivation, rurality and physical space. In addition to those attributes that represent analysis outputs, there are further contextual attributes, derrived from the national Census, that may be utilised to enhance a typology of Internet use and engagement. Attributes pertaining to; age, population density, higher education, levels of qualification and occupation have been previously discussed as having links to levels of engagement. The inclusion of such attributes would likely assist in building a typology by introducing measures of demographic characteristics, to supplement attitudinal, access and performance data, and as such, creating a broader view.

5.2 Geodemographic Specification

The construction of geodemographic classifications are often described as part art and part science (Singleton and Spielman, 2013; Harris et al., 2005; O'Neil and Schutt, 2013). As such, there are numerous different approaches to constructing a bespoke classification. Recent developments in open source geodemographic classifications (Vickers and Rees, 2007; Singleton and Spielman, 2013; Singleton, 2014) with more transparent methodological underpinnings can be argued as aiding understanding within the field. However there still remains significant variation in methods employed, which is perhaps to be expected, given a particular set of methods or procedures suitable for one classification may not be suitable for another. Harris et al. (2005) present the most comprehensive overview of the various stages involved in constructing a geodemographic classification. As such, this section will follow a similarly segmented approach, presenting a short overview of the common stages of:

- Selecting potential measures
- Data Evaluation
- Transformations and normalisation
- Weighting
- Standardisation
- Clustering
- Arranging clusters into a hierarchy
- Providing textual and visual summaries

The first stage in building a geodemographic classification is to evaluate potential measures. Typically, K-means clustering is used when building a classification, and as such, these measures must be numeric and continuous (measured) as opposed to discrete or categorical in order for K-means to work effectively. Other clustering algorithms exist, which can handle incomplete or categorical variables (Expectation-Maximisation or hierarchical clustering, for example), although the literature would suggest that these are less commonly used (Harris et al., 2005; Vickers and Rees, 2007; Singleton and Spielman, 2013; Singleton, 2014). Data inputs are generally measured continuously (attributed to some question or observation) and aggregated to a predefined geographic resolution. The geographic resolution at which classifications can be built in a UK context is variable, but common aggregate geographies include the postcode boundary, Output Area (OA), Lower Layer Super Output Area (LSOA) or ward. The data structure prior to clustering is typically a table whereby rows represent small geographic areas and columns represent attribute information on which they will be clustered.

The next stage typically involves an evaluation of input variables through Exploratory Data Analysis (EDA) to examine issues such as correlation, assess distributions and skew and more generally, gain an understanding of relationships between variables within the data. At this stage a classification builder may assess which data inputs would be best to include and which may represent duplicate information (typically those variables that show high correlations with others), although there are no firm rules, and choices can be largely subjective.

Following assessment of input data, the next stage typically involves data normalisation, which aims to limit skew. In an ideal scenario, all input variables would display normal (Gaussian) distributions as some clustering algorithms (K-means for example) are optimised to find spherical clusters, which can be problematic when using skewed inputs. In practice, however, there may be a number of input variables to be included that display skewed distributions. There are a number of common data normalisation methods that can be implemented, including; log10, Box-Cox and inverse hyperbolic sine transformations (Gale et al., 2016). Whether or not skewed data should be transformed at all is a subject that has generated academic debate. There are advantages to transforming variables in that cluster formation is less likely to be adversely affected by skewed distributions. Conversely, there are disadvantages in that outliers are important in the formation of distinctive clusters, and global normalisation measures reduce the impact of these, which can lead to the smoothing or disappearance of interesting geographic patterns (Singleton and Spielman, 2013). In some commercial classifications this problem is overcome by assigning weights to reduce the impact of skewed data on cluster formation (Harris et al., 2005), although how weights are derived is typically subjective and open to criticism.

The next stage is generally the last before clustering the data and involves standardising the inputs so that they share a common scale. This is important because input variables will typically display different distributions, and in order to assess how large or small a particular area's variance is from the norm (mean), and to draw comparisons between variables, a common scale must be used (Harris et al., 2005). A common method is to transform the data values to z-scores. z-scores are calculated by subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation.

$$z = \frac{x - \mu}{\sigma}$$

Where μ is the mean of the population and σ is the standard deviation of the population.

This results in a set of scores that are positive if they fall above the mean and negative if they fall below, i.e. all standardized variables will have an adjusted population mean of 0. Using z-scores can be problematic in some instances, for example if an input variable is highly skewed with many outliers, the resulting z-score may be large enough to influence an area's cluster membership regardless of the area's other attributes. Again, weighting and variable normalisation techniques can be utilised to alleviate this issue.

Following standardisation across the dataset of final variables, the next stage is to run a cluster analysis. Typically an iterative allocation-reallocation method (K-means) is used, although other methods such as hierarchical also exist. The hierarchical method essentially treats each area as a separate cluster in the first instance and merges these 'clusters' based on measures of similarity. After similar clusters are merged, average values for the new clusters are computed and the process repeats until convergence, where an appropriate number of clusters (that exhibit minimum intra-cluster variance and maximum inter-cluster variance) are found. Although methodologically simplistic, this method can be computationally exhaustive, taking comparatively longer than an iterative allocation-reallocation method to run due to the assessment and re-assessment of cluster pairs. This can be particularly problematic when datasets are extremely large.
An iterative allocation-reallocation method uses a different technique to compute cluster assignments. A K-means algorithm works by setting seeds, which can be pre-defined or random observations within a dataset. The number of initial seeds is equal to the user-defined value for k (the number of clusters to be output). The algorithm then begins to assign observations to each seed based on proximity, typically measured by Euclidean distance. This initial allocation represents the first iteration of the algorithm. The centroids of the newly formed clusters are then calculated and become the seeds for the next iteration of assignments. The algorithm aims to minimise the Within Cluster Sum of Squares (WCSS), which is the cumulative sum of all the squared Euclidean

distances from observations to cluster centroids. Smaller WCSS values represent more homogenous (or similar) clusters. The algorithm repeats for many iterations until convergence, when assignments no longer change and WCSS values have been minimised. Figure 5.1 illustrates how the algorithm works on a two dimensional dataset. Beginning in the upper left of the image, a set of data are observed with an initial seed selection shown by crosshairs and cluster assignments shown by colour. Multiple iterations of the algorithm are plotted to show how cluster assignments change as the algorithm cycles.

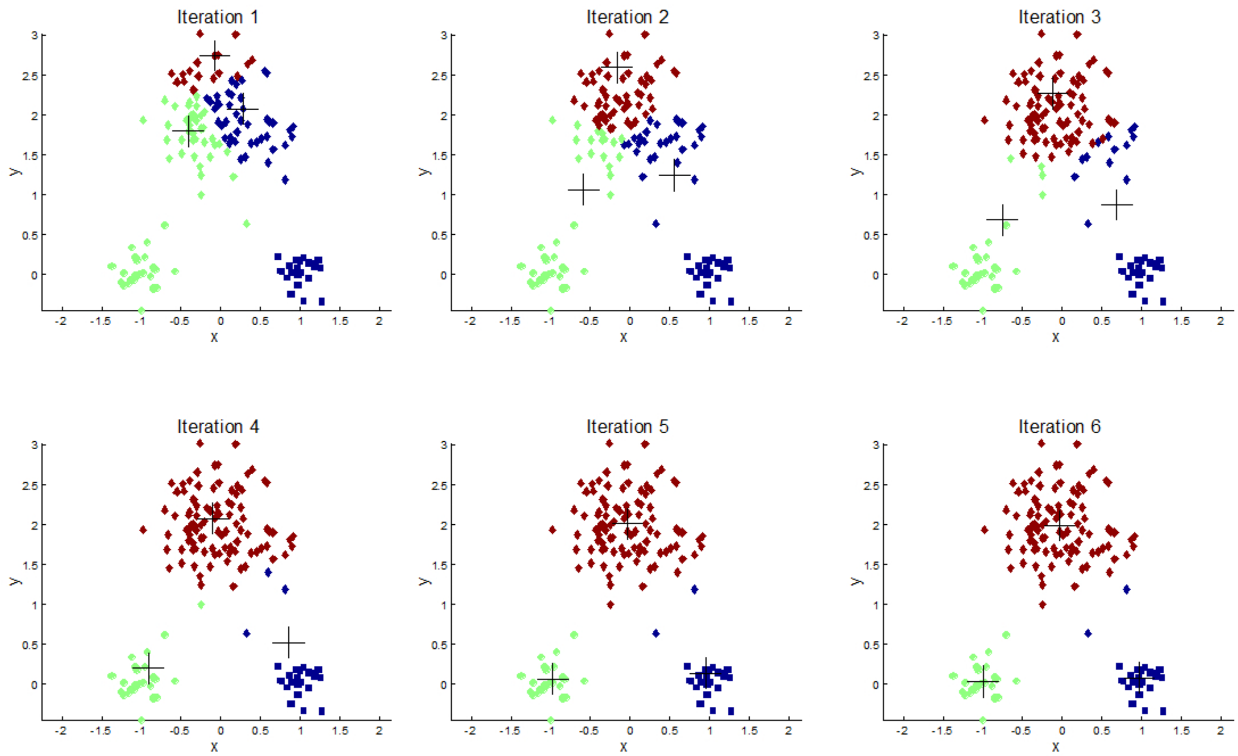


Figure 5.1: Graphical Representation of an Iterative Allocation-reallocation Method on a Two-dimensional Dataset. *Source: http://www.practicaldb.com/data-visualization-consulting/cluster-analysis/*

After initial clustering, it is possible to to design a hierarchical structure to the classification (see Figure 5.2). This can be completed in one of two ways, top-down or bottom-up. A bottom-up hierarchy involves clustering the data into k clusters, which can then be merged to form a higher tier within the hierarchy. For example, the lowest (or most granular) tier of a three-tier classification may be referred to as 'type'. If the data is capable of supporting 50 'types' (essentially 50 distinct clusters) it can be clustered to this extent in the first instance. Similar 'types' may then be merged to form a less granular mid-level tier in the classification, call these 'groups'. Similarly, these 'groups' could be merged, again based on similarity, to form a coarse top tier of the classification, call these 'supergroups'. The number of tiers, initial number of clusters and grouping of clusters is, however, largely decided by trial and error (Harris et al., 2005; Singleton, 2014). Even in large datasets, cluster sizes and characteristics can vary significantly, as such, it may be that different groups have uneven numbers of types assigned to them and so on.

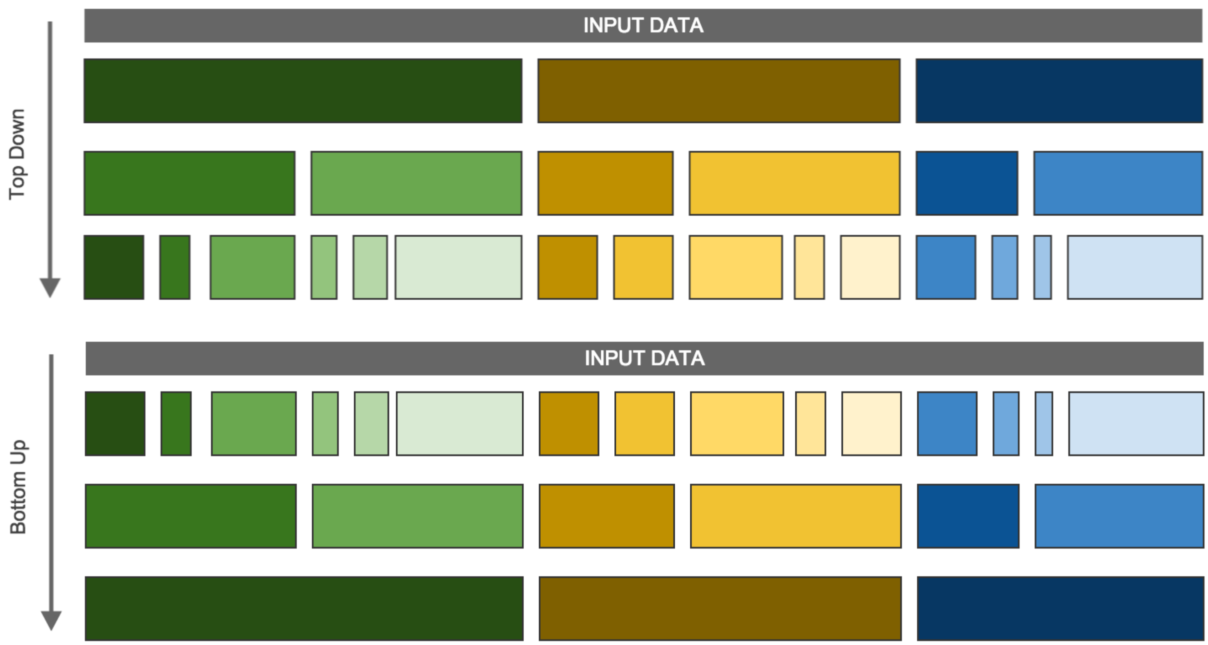


Figure 5.2: Top Down and Bottom Up Clustering Approaches

The second approach is top-down and it has been noted in literature that this method of processing helps to structure expectations about incoming data and emphasises differences between pieces of information (Lindsay and Norman, 1977). The process involves clustering input data into a pre-defined number clusters that will form the most aggregate (highest tier) of the resulting classification. Finding the optimum number of initial clusters is largely down to trial and error, although methods exist to assist with evaluation. One such method involves the use of 'Clustergrams', which are visualisations of the assignment and re-assignment of observations to clusters across a range of values for k. This method can assist in the selection of an optimum k value as it is possible to identify which clusters split to form new clusters and assess similarity or 'closeness' of newly formed clusters. After initial clustering to form a top tier, each of these newly formed clusters can be extracted from the dataset and re-clustered to form a second tier. This method can be repeated as many times as desired, although it would be logical to stop when sub-clusters begin to show no obvious differences from parent clusters. Identifying the number of sub-clusters an initial cluster can support is, again, largely dictated by trial and error, as are the number of tiers to a final classification.

After clusters have been defined and arranged into a hierarchy, the next stage is to make sense of the resulting output. A typical method is to explore the mean values for each input attribute in a cluster and compare these scores to those of other clusters, thus building an understanding of that cluster's key characteristics. Alternatively, index scores can be calculated by dividing cluster means for each attribute by the population mean for each attribute and multiplying by 100. The resulting output is a score where a value of 100 represents the population mean. Therefore, a score of 200 indicates a value of twice the population mean and a value of 50 half the population mean. As cluster characteristics are identified and compared, it is possible to compile textual summaries for each cluster in such a way that an end user with no prior knowledge of the data is able to understand the key characteristics or defining features. These textual summaries are referred to as 'Pen Portraits' and typically average 200 to 1000 words (Harris et al., 2005). The process is complex and may be argued as subjective, as different classification builders would likely develop different textual descriptions of the quantitative information. Often these textual summaries are accompanied by imagery (photos or graphs) to visualise important cluster characteristics. After textual descriptors are finalised, clusters can be named. Typically, cluster names are short and easily communicable.

5.3 Building a Classification of Internet Use and Engagement

5.3.1 Selecting Measures

Tables 5.1 - 5.3 provide a summary of the measures that were considered for inclusion. The majority of these measures are derived from the OXIS, and the procedure through which estimates were calculated has been presented in Chapter 4. A total of 76 measures were considered for classification, representing a number of domains relating to infrastructure, access, engagement and contextual information. The measures were compiled at the aggregate level of the LSOA. This was necessary, as the coverage of some measures, particularly those relating to infrastructure, did not have full coverage at the OA level (see Chapter 3, Section 3.2.3). There were a small number (0.37%) of cases where LSOAs had no data pertaining to local broadband speed. In these instances, data were imputed by assigning the average download speed of the district in which the LSOA fell, as there was 100% data coverage at the district level. Contextual indicators were obtained from the census and broadly represent attributes that are known to show correspondence with levels of engagement with the Internet, including; age, levels of qualification, occupation type, prevalence of student populations and population density (see Chapter 2, Section 2.4. Also Chapter 4, Section 4.6). Cumulatively, the data held totalled 2,496,220 LSOA level observations. It was apparent that a number of the measures displayed skewed distributions. There are a number of methods to deal with skewed distributions, which are discussed in subsequent section of this chapter.

Table 5.1: Input Variables Selected for Evaluation

Variable ID	Descriptor	Domain	Distribution	Skew
QA1a1	Seeking info MP - Internet	Seeking Information	Approximately Symmetric	-0.21
QA1a2	Seeking info MP - Smartphone	Seeking Information	High Positive Skew	+1.30
QA1b1	Seeking info council tax - Internet	Seeking Information	Approximately Symmetric	+0.27
QA1b2	Seeking info council tax - Smartphone	Seeking Information	High Positive Skew	+1.80
QA1c1	Seeking info holiday/journey - Internet	Seeking Information	Approximately Symmetric	+0.08
QA1c2	Seeking info holiday/journey - Smartphone	Seeking Information	High Positive Skew	+1.09
QA1f1	Seeking info topic/professional project - Internet	Seeking Information	Approximately Symmetric	+0.09
QA1f2	Seeking info topic/professional project - Smartphone	Seeking Information	High Positive Skew	+2.68
QA2c	Internet important for information	Perceptions	Approximately Symmetric	+0.19
QA3d	Internet important for entertainment	Perceptions	Approximately Symmetric	+0.43
QA9	Interested in the Internet	Perceptions	Approximately Symmetric	+0.14
QC1b	Use Internet while travelling - mobile/dongle	Access Patterns	Approximately Symmetric	+0.50
QC2a	Mostly use mobile phone for Internet	Access Patterns	High Positive Skew	+1.37
QC22a	Have found a job through the Internet	Commercial Applications	Moderate Positive Skew	+0.60
QC22b	Have saved money buying online	Commercial Applications	Approximately Symmetric	+0.31
QC30b	Frequently buy products online	Commercial Applications	Moderate Negative Skew	-0.52
QC30d	Frequently pay bills online	Commercial Applications	Approximately Symmetric	-0.46
QC30e	Frequently use online banking	Commercial Applications	Approximately Symmetric	+0.17
QC30f	Frequently compare prices online	Commercial Applications	High Negative Skew	-1.27
QC30h	Frequently order food or groceries online	Commercial Applications	Moderate Positive Skew	+0.59
QC30i	Frequently sell things online	Commercial Applications	Approximately Symmetric	-0.06
QH1a	Households that have Internet access at present	Household Access	Symmetric	0.00
QH1b	Households that don't have Internet access but have had in past	Household Access	Approximately Symmetric	-0.22
QH1c	Households that have never had Internet access	Household Access	Approximately Symmetric	+0.07
QH3	Households that have had Internet access for ten years or more	Household Access	Approximately Symmetric	+0.05
QH5	Households with wireless access in home through Wi-Fi	Household Access	Approximately Symmetric	-0.24

Table 5.2: Input Variables Selected for Evaluation (2)

Variable ID	Descriptor	Domain	Distribution	Skew
QH7f	Households with a tablet computer	Household Access	Approximately Symmetric	+0.41
QH7g	Households with e reader	Household Access	Approximately Symmetric	+0.50
QH7h	Households with games console	Household Access	Approximately Symmetric	+0.05
QH7i	Households with a smart TV	Household Access	Approximately Symmetric	-0.35
QH11	Mobile phone ownership	Mobile Access	Approximately Symmetric	-0.21
QH12b	Use mobile phone for email	Mobile Access	Moderate Positive Skew	+0.54
QH12e	Use mobile for posting videos and photos online	Mobile Access	Approximately Symmetric	+0.42
QH12h	Use mobile phone for navigation	Mobile Access	Moderate Positive Skew	+0.63
QH12i	Use mobile phone for social networking	Mobile Access	Approximately Symmetric	+0.45
QH12j	Use mobile phone for apps	Mobile Access	Approximately Symmetric	+0.44
QH12k	Use mobile phone for browsing the Internet	Mobile Access	Moderate Positive Skew	+0.60
QH13_Current	Current Internet users	Personal Access	Approximately Symmetric	+0.08
QH13_Ex	Ex Internet users	Personal Access	Approximately Symmetric	+0.07
QH13_Non	Internet non users	Personal Access	Approximately Symmetric	-0.04
No_qual	Persons with no qualifications	Demographic/ Contextual	Approximately Symmetric	+0.41
L_one_qual	Persons with level one qualifications	Demographic/ Contextual	Approximately Symmetric	-0.39
L_two_qual	Persons with level two qualifications	Demographic/ Contextual	Moderate Negative Skew	-0.78
Apprent	Persons who are apprentices	Demographic/ Contextual	Approximately Symmetric	+0.27
L_three_qual	Persons with level three qualifications	Demographic/ Contextual	High Positive Skew	+5.92
L_four_qual	Persons with level four qualifications	Demographic/ Contextual	Moderate Positive Skew	+0.80
Oth_qual	Persons with other qualifications	Demographic/ Contextual	High Positive Skew	+2.09
MDSO_one	Managers, directors and senior officials	Demographic/ Contextual	High Positive Skew	+1.09
PO_two	Professional occupations	Demographic/ Contextual	Moderate Positive Skew	+0.85
APTO_three	Associate professional and technical occupations	Demographic/ Contextual	High Positive Skew	+1.70
ASO_four	Administrative and secretarial occupations	Demographic/ Contextual	Approximately Symmetric	+0.30
STO_five	Skilled trades occupations	Demographic/ Contextual	Symmetric	0.00

Table 5.3: Input Variables Selected for Evaluation (3)

Variable ID	Descriptor	Domain	Distribution	Skew
CLOSO_six	Caring, leisure and other service occupations	Demographic/ Contextual	Approximately Symmetric	+0.32
SCSO_seven	Sales and customer service occupations	Demographic/ Contextual	Moderate Positive Skew	+0.67
PPMO_eight	Process, plant and machine operatives	Demographic/ Contextual	Moderate Positive Skew	+0.77
EO_nine	Elementary occupations	Demographic/ Contextual	Moderate Positive Skew	+0.85
FT_Students_PCT	Full time students	Demographic/ Contextual	High Positive Skew	+6.53
X0_4_PCT	Persons aged 0 to 4	Demographic/ Contextual	Moderate Positive Skew	+0.72
X5_7_PCT	Persons aged 5 to 7	Demographic/ Contextual	Moderate Positive Skew	+0.51
X8_9_PCT	Persons aged 8 to 9	Demographic/ Contextual	Approximately Symmetric	+0.35
X10_15_PCT	Persons aged 10 to 15	Demographic/ Contextual	Approximately Symmetric	+0.18
X16_17_PCT	Persons aged 16 to 17	Demographic/ Contextual	High Positive Skew	+3.82
X18_19_PCT	Persons aged 18 to 19	Demographic/ Contextual	High Positive Skew	+10.93
X20_24_PCT	Persons aged 20 to 24	Demographic/ Contextual	High Positive Skew	+5.73
X25_29_PCT	Persons aged 25 to 29	Demographic/ Contextual	High Positive Skew	+2.04
X30_44_PCT	Persons aged 30 to 44	Demographic/ Contextual	Moderate Positive Skew	+0.78
X45_59_PCT	Persons aged 45 to 59	Demographic/ Contextual	Moderate Negative Skew	-0.59
X60_64_PCT	Persons aged 60 to 64	Demographic/ Contextual	Approximately Symmetric	+0.19
X65_74_PCT	Persons aged 65 to 74	Demographic/ Contextual	Approximately Symmetric	+0.48
X75_84_PCT	Persons aged 75 to 84	Demographic/ Contextual	Moderate Positive Skew	+0.84
X85_89_PCT	Persons aged 85 to 89	Demographic/ Contextual	High Positive Skew	+1.53
X90_PLUS	Persons aged 90 plus	Demographic/ Contextual	High Positive Skew	+2.37
Den_pph	Population density persons per hectare	Demographic/ Contextual	High Positive Skew	+2.28
dist_to_ex	Distance to closest telephone exchange	Local Infrastructure	High Positive Skew	+1.25
DOWNLOAD	Local download speed 2012/13	Local Infrastructure	High Positive Skew	+1.54
dist_to_mast	Distance to closest mobile base station	Local Infrastructure	High Positive Skew	+3.22

5.3.2 Assessing Skew

After input measures had been joined and formatted a number of tests were applied to ascertain the normality of the distribution of each variable in turn. The results of these tests are recorded in tables 5.1 - 5.3. A test of skewness was applied in favour of visually inspecting histograms, as this outputs a quantitative measure for comparative purposes, reducing the likelihood of interpretation error. Skewness is a measure of the asymmetry of the distribution of a random variable about its mean. Generally, a measure of skewness is either positive (where the right tail of the distribution is longer and the mean is greater than the median) or negative (where the left tail of the distribution is longer and the mean is less than the median) (Papoulis and Pillai, 2002). Three rules for interpreting skewness apply:

- If a skewness value is less than -1 or greater than 1, the distribution is thought to be highly skewed
- If a skewness value is between -1 and -0.5 or between 0.5 and 1, the distribution is thought to be moderately skewed
- If a skewness value is between -0.5 and 0.5, the distribution is thought to be approximately symmetric

Table 5.4 summarises the number of variables deemed to fall into each category of skewness, summarising information presented in Tables 5.1 - 5.3.

Table 5.4: Summary of Input Measure Skewness

Skew	Count	Percentage
Approximately Symmetric	38	50%
Moderate Negative	3	3.9%
Moderate Positive	14	18.4%
High Negative	1	1.3%
High Positive	20	26.3%

The majority of the measures that were assessed were approximately symmetric in their distributions and a small number were moderately skewed. A single variable (frequent comparison of prices online) was deemed to be have high negative skew, although the actual value of -1.27 falls close to the boundary of moderate skew and as the only highly negatively skewed variable, may be considered unlikely to impact cluster assignments. In total, 20 of the 76 measures (around 26%) were deemed to be highly positively skewed. However, these 20 measures would be expected to display skewed distributions given the domain. For example, the distribution of elderly populations and students is not even, as particular areas attract these groups. Similarly, indicators of infrastructure performance and prevalence would be expected to be skewed towards urban areas, as these contain larger populations and generally have the highest performing network infrastructure. No measures were eliminated as a result of uneven distributions. It is argued that measures exhibiting skew would either be normalised using power transformations to reduce skewness, or used regardless of skew, as the outliers within these measures may assist in producing distinct clusters, see Singleton and Spielman (2013).

5.3.3 Data Evaluation

In addition to assessing input measures based on skewness, Exploratory Data Analysis (EDA) was used to assess the extent to which measures were correlated with one another. Linear modelling was used to assess the relationship between key input measures in addition to producing a correlation matrix for the entire dataset. The inclusion of highly correlated measures in a geodemographic classification is generally discouraged as this can result in 'duplicate information' (Harris et al., 2005) whereby multiple measures adequately capture the same relationship. Multiple highly correlated measures can either be omitted, keeping a single variable that is correlated with the largest number of other measures in an attempt to ensure robustness. Alternatively, multiple highly correlated measures can be included in a cluster analysis either with or without applying weights. Weighting can be problematic as the process of selecting weights for individual measures can be argued as subjective (Harris et al., 2005). The correlation matrix that was generated is too large to visualise in this thesis (76x76), however an extract is included as an appendix. In addition it is possible to summarise the most notable correlations observed between sets of input measures. In general:

- Measures related to seeking information online and general Internet use were correlated with one another

- Measures related to Internet use were correlated with age profiles. This relationship was positive for those areas with younger age profiles and negative for those with older age profiles

- Measures related to device ownership were correlated to one another. Suggesting in those areas where there is a high rate of ownership of one device, there will likely be high rates of ownership for other devices

- Measures relating to mobile Internet use were correlated with age profiles, specifically younger age profiles. Similarly, mobile Internet use had a positive correlation with areas containing large student populations

- Measures relating to infrastructure prevalence and performance were strongly correlated to measures of population density

Broadly speaking, these relationships are not unexpected and have been observed through previous literature and analyses within this thesis. As such, despite some instances of high correlation between measures, it was deemed that no measures would be removed as a result. This decision was made on the basis that removing correlated variables based on the analysis of global statistical relationships, which may mask local variation, could lead to the smoothing of important non-linear patterns at a more granular level, as discussed by Singleton and Spielman (2013); Singleton (2014).

5.3.4 Transformation and Normalisation

The next stage was consideration of transformation and normalisation procedures. Variable normalisation methods are, in simplistic terms, adjustments where the intention is to bring distributions into alignment, effectively reducing skew. There exist two competing arguments, normalise to minimise outlier effects or embrace 'natural' distributions and allow outliers to influence cluster formation. Two variable normalisation methods were tested before a decision was taken, Box-Cox and log10, as these are frequently referenced in geodemographic literature (Alexiou and Singleton, 2015; Gale et al., 2016). Both methods require values to be positive and greater than 1. Because a number of variables had values that fell below 1, a constant of 1 was added to each value so that transformations could take effect. This is a common method in this scenario (Osbourne, 2002). log10 transformations, whilst reducing skew in most instances, apply a globally standard method of normalisation across a dataset. This does not optimise sensitive to the distribution of each variable in isolation, instead working by compressing the differences between large values and increasing the differences between small values to artificially reduce variance.
The Box-Cox method uses an exponent (lambda (λ)), to transform a variable (y) and normalise its distribution. Multiple lambda values are tested and the best (that which results in the most normal distribution) is used for the power transformation. This means that the extent to which a variable is transformed is dependent on its level of skew. The Box-Cox method can be expressed as:

$$y_i(\lambda) = \begin{cases} y_i^{\lambda} - 1)/\lambda & (\lambda \neq 0) \\ log(y) & (\lambda = 0) \end{cases}$$

The two methods of reducing skew were compared and Table 5.5 details the results for a small subset of input variables. An initial clustering run was then completed on a transformed set of variables, alongside the naturally distributed (non-manipulated) variables to observe the effects of normalisation on cluster assignments.

Table 5.5: Resulting Effects of Transformation Methods on Skewed Variables

Variable	Descriptor	Skew (Raw)	Skew (Box-Cox)	Skew (log10)
QA1a2	Seeking info (mobile)	+1.30	+0.01	+0.35
QC2a	Mostly use mobile Internet	+1.37	+0.03	+0.44
QC30b	Freq. buy products online	-0.52	-0.36	-0.68
Den_pph	Population density	+2.28	-0.58	-0.82
PO_two	Professional occupations	+0.85	-0.02	-0.28

It is evident that the Box-Cox method significantly outperformed log10 in reducing variable skew, generally reducing all but the most highly skewed variables to approximately symmetric distributions. As such, the Box-Cox method was applied to all skewed variables in the dataset. Two separate datasets were output from this evaluation, one transformed to reduce variable skew, and one where natural distributions were observed. An example visualisation of the reduction in variable skew is presented in Figure 5.3. Here the skewness of an unevenly distributed variable (QC2a: Frequent Buying of Products Online) is plotted 'pre' and 'post' transformation using a Quantile-Quantile Plot. The Quantile-Quantile (Q-Q) plot is a graphical technique for determining if two datasets come from populations with a common distribution. A 45-degree reference line is plotted, and if the two datasets come from a population with the same distribution, the points should fall approximately along this reference line. The greater the departure from this reference line, the more likely it is that the two data sets have come from populations with different distributions. In this instance the measure being assessed was compared with a randomly generated variable of normal distribution.

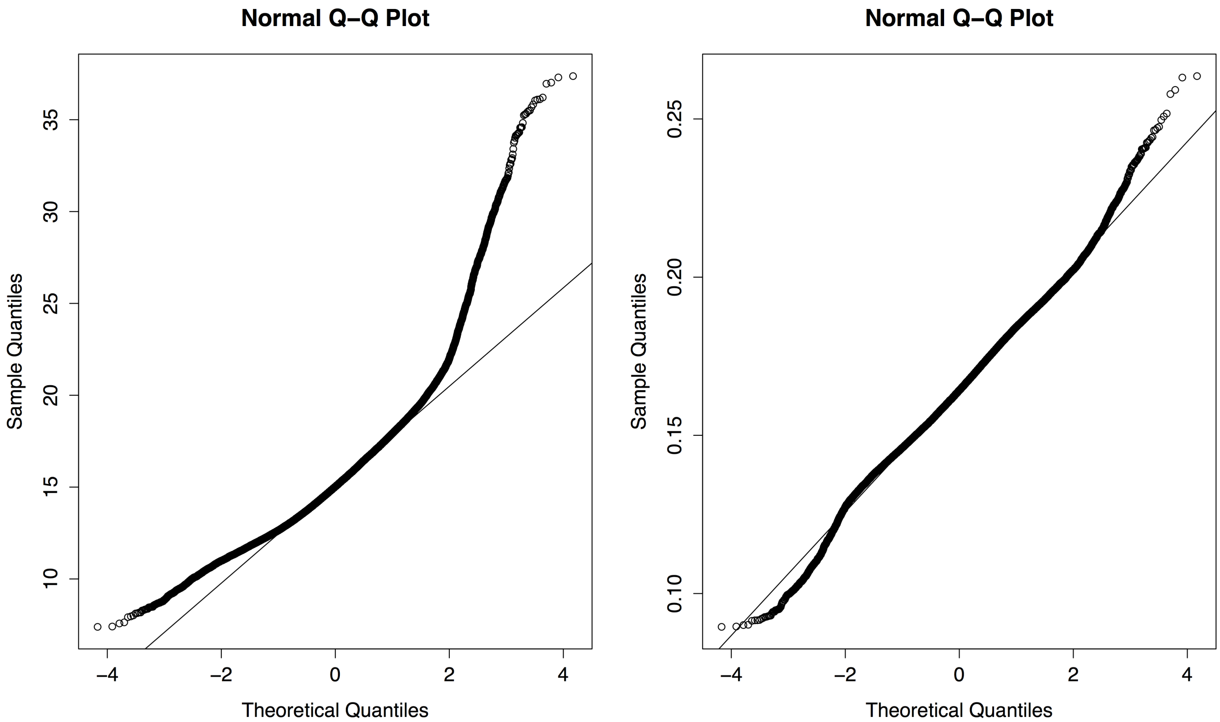


Figure 5.3: Box-Cox Power Transformation on a Highly Skewed Variable (QC2a: Frequent Buying of Products Online) Q-Q plots Before and After.

5.3.5 Variable Weighting

An alternative method to normalisation is weighting. In essence, applying weights lessens the impact of heavily skewed variables in cluster formation by reducing influence. The extent to which a variable should be down (or up) weighted is subjective and heavily reliant on the preferences and objectives of the classification builder. As such, this method has been criticised in academic literature (Harris et al., 2005; Singleton and Spielman, 2013). Given the subjective nature of variable weighting, this method was not assessed for implementation.

5.3.6 Standardisation

Before any cluster analysis could be conducted, it was necessary to standardise input variables so that they all fall on a common scale. As such, both the transformed and naturally distributed datasets were standardised using z-scores. As mentioned previously, z-scores are the most common method for data standardisation and are calculated by subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation.

$$z = \frac{x - \mu}{\sigma}$$

Where μ is the mean of the population and σ is the standard deviation of the population.

This results in a set of scores that are positive if they fall above the mean and negative if they fall below, i.e. all standardized variables will have an adjusted population mean of 0.

5.3.7 Initial Clustering

The next stage in the evaluation was to cluster both the transformed and naturally distributed inputs to assess the impact these processes had on the cluster outputs. At this stage a full classification was not built, instead an initial run of k =5 clusters was completed, the value for k was chosen arbitrarily in this instance. Initial clustering of the two datasets revealed apparent differences in terms of cluster sizes, aggregate characteristics, geographic distribution and interpretability. Tables 5.6 and 5.7 present a summary of the key observations from the clustering process on the transformed and naturally distributed datasets respectively. These summaries are based on visual interpretation of the outputs of the clustering process and are significantly condensed such that they can be presented in a short table. As has been noted in previous sections, cluster interpretation can be viewed as a subjective process.

Initial interpretation of the cluster outputs suggested that the naturally distributed data produced the most homogenous and interpretable assignments. Although much of the output has been summarised in tables 5.6 and 5.7, the aggregate characteristics support this point. In the unedited dataset where natural distributions were observed, the clusters formed were more distinctive. Based even on the aggregated representation it is apparent that:

- Cluster 1 is predominantly elderly and unengaged
- Cluster 2 represents highly engaged professionals in high density areas
- Cluster 3 represents those users with engagement characteristics close to the national average
- Cluster 4 is comprised of highly engaged students
- Cluster 5 is predominantly rural with poor infrastructure but average usage

Table 5.6: Cluster Characteristics: Transformed Data

Cluster	Internet Use	Device Ownership	Mobile Access	Age	Students	Employment Sector	Qualifications	Infrastructure Performance	Population Density
1	Low	Low	Low	High	Low	Trades/ Service Occupations	Low	Low	Low
2	High	High	High	Mixed	Avg.	Professionals	High	High	High
3	High	High	High	Low	Avg.	Mixed	Mixed	Avg.	Avg.
4	Mixed	Mixed	Mixed	Mixed	Avg.	Mixed	Mixed	Avg.	Avg.
5	Mixed	Avg.	Low	High	Low	Managers/ Directors	High	Low	Low

Table 5.7: Cluster Characteristics: Naturally Distributed Data

Cluster	Internet Use	Device Ownership	Mobile Access	Age	Students	Employment Sector	Qualifications	Infrastructure Performance	Population Density
1	Low	Low	Low	High	Low	Trades/ Service Occupations	Low	Low	Low
2	High	High	High	Low	Avg.	Professionals	High	High	High
3	Avg.	Mixed	Avg.	Low	Avg.	Elementary/ Service Occupations	Low	High	Avg.
4	High	High	High	Low	High	Students/ Sales Occupations	Mixed	High	High
5	Avg.	Avg.	Low	High	Low	Managers/ Directors	High	Low	Low

In comparison, the transformed data did not form such easily interpretable clusters. Although a similarly unengaged and elderly cluster (cluster 1) is apparent, as is a highly engaged professional cluster (cluster 2), much of the remainder of the cluster assignments are not easily distinguishable and are heavily mixed. This may be due in part to power transformations, which appear to suppress potentially interesting patterns within the data. The resulting output is a mix of some homogenous clusters, alongside some that are heavily mixed or close to average across all measures, which hinders interpretability. The transformed dataset did, however, produce more evenly sized clusters than the dataset that retained natural distributions. Cluster four of the naturally distributed dataset is comprised almost entirely of student populations, but is small in comparison to other clusters, equating to around 2% of all LSOAs nationally. In the transformed data, the majority of these same areas appeared to fall into clusters 2 and 3, resulting in more evenly sized, but less homogenous clusters overall. Because more homogenous clusters offer greater interpretability and ultimately result in a more defined classification, it was deemed that the data without normalisation would be used to build the final classification.

5.3.8 Assessing Variable Influence

Following consideration of normalisation, the variables that influenced the homogeneity of the clusters were assessed. A ranking was accomplished by noting the Total Within Sum of Squares (TWSS) statistic after an initial clustering run of $k = 5$ on the naturally distributed data. One variable was then removed in turn, the clustering run was repeated, and the resulting TWSS statistic was recorded. It was then possible to assess the effects of each variable on cluster homogeneity by calculating the extent to which each variable increased or decreased the TWSS statistic relative to the number of variables in the model. A recorded increase in the TWSS when a variable was omitted from the model indicates that the inclusion of that variable reduces the TWSS and increases cluster homogeneity. Conversely, a recorded decrease in the TWSS when a variable was omitted from the model indicates that the inclusion of that variable increases the TWSS and reduces cluster homogeneity.

Table 5.8 shows the variables that increased cluster homogeneity the most (a resulting increase in TWSS when they were omitted). Table 5.9 shows the variables that decreased cluster homogeneity the most (a resulting decrease in TWSS when they were omitted). The percentage change in TWSS is calculated per variable, relative to the number of variables that were in the model at that time. The top ten variables that increased cluster homogeneity were derived from the OXIS, which indicates that measures of use and engagement characteristics, as opposed to census and infrastructure attributes, have most impact on the classification structure. In addition, the top ten variables

Table 5.8: Variables that Increased Cluster Homogeneity

Variable Omitted	Descriptor	% Change in TWSS
QH12b	Use mobile phone for email	+0.644
QA3d	Internet important for entertainment	+0.638
QH12k	Use mobile phone for browsing the Internet	+0.638
QH7f	Households with a tablet computer	+0.626
QC1b	Use Internet while travelling - mobile/dongle	+0.607
QA2c	Internet important for information	+0.591
QH13_Non	Internet non users	+0.586
QH12j	Use mobile phone for apps	+0.583
QA9	Interested in the Internet	+0.579
QA1c1	Seeking info holiday/journey - Internet	+0.575

that increased cluster homogeneity were not heavily skewed, instead they exhibited approximately symmetric or moderately skewed distributions in all cases. This may suggest that highly skewed variables in the dataset were not significantly influencing cluster formation. The variables that increase cluster homogeneity also represent a balanced mix across the domains of seeking information, household access, access patterns and perceptions.

Table 5.9: Variables that Decreased Cluster Homogeneity

Variable Omitted	Descriptor	% Change in TWSS
DOWNLOAD	Local download speed	-1.366
dist_to_ex	Distance to closest telephone exchange	-1.301
ASO_four	Administrative and secretarial occupations	-1.201
dist_to_mast	Distance to closest mobile base station	-1.149
X16_17_PCT	Persons aged 16 to 17	-1.077
X8_9_PCT	Persons aged 8 to 9	-0.917
QC30f	Frequently compare prices online	-0.886
X90_PLUS	Persons aged 90 plus	-0.879
X10_15_PCT	Persons aged 10 to 15	-0.837
L_two_qual	Persons with level two qualifications	-0.830

The top ten variables that reduced cluster homogeneity were derived from a mix of the OXIS, census and infrastructure data and exhibited mixed levels of skew. Interestingly, all of the three infrastructure measures (local download speed, distance to closest telephone exchange and distance to closest mobile base station) reduced cluster homogeneity. This may be due to there being a comparatively small number of infrastructure measures compared with larger amounts of OXIS and census data. These variables are, however, highly skewed and theoretically have the power to influence cluster assignments. The output suggests that even within clusters, local infrastructure performance may vary dramatically, a logical assumption given how varied performance can be, even within relatively small geographic areas as a result of uneven rollout of next generation services. Working with this assumption, these metrics might logically decrease cluster homogeneity given a cluster may represent several thousand LSOAs at this level of the classification.

5.3.9 Construction and Hierarchical Design

The final stage of construction was to cluster the input measures to form a conclusive classification. One method of selecting the number of initial clusters is to use 'Clustergrams', which are plots that visualise the distribution and re-distribution of observations between clusters for a range of values for k. These visualisations allow an optimum value for k to be selected by enabling identification of assignment and re-assignment and also by visualising how closely related new clusters are to existing ones as the values for k increase. Figure 5.4 shows a Clustergram where the value of k was tested from 2 through 10. For each cluster iteration the method works by multiplying the cluster centres by the first loading of the principal components of the original data. Thus offering a weighted mean of each cluster's centre dimensions.

It is apparent that when k = 2, observation assignments are relatively even and the two clusters are well spaced. The spacing suggests that the two clusters are sufficiently differentiated in terms of their characteristics that they are easily distinguishable from one another. In this context, they would likely represent two very different groups of Internet users. As the number of clusters (value for k) is increased to 3, it is possible to track the reassignment of observations. In this example, a number of observations from the upper and lower clusters are reassigned to form a middle cluster. Some observations from the lower cluster are also reassigned to the higher cluster. At k = 4 a second mid level cluster is formed, taking observations from all three previous clusters. The two mid level clusters at this stage have, however, become very close, which may impact their interpretability. At k = 5, the cluster assignments are well spaced. At k = 6 and k = 7, cluster assignments become less evenly spaced, suggesting the optimum value for k has been exceeded.

Based on the interpretation of Figure 5.4, the optimum number of clusters was set at 5. Once the optimal value for k was selected, a clustering run was repeated for 10,000 iterations to ensure an optimal result. The clusters output were almost identical to those detailed in Table 5.7.

Each of the five initial clusters were then separated and re-clustered in an attempt to form a second tier within the classification, as this would likely result in a more granular nested typology.

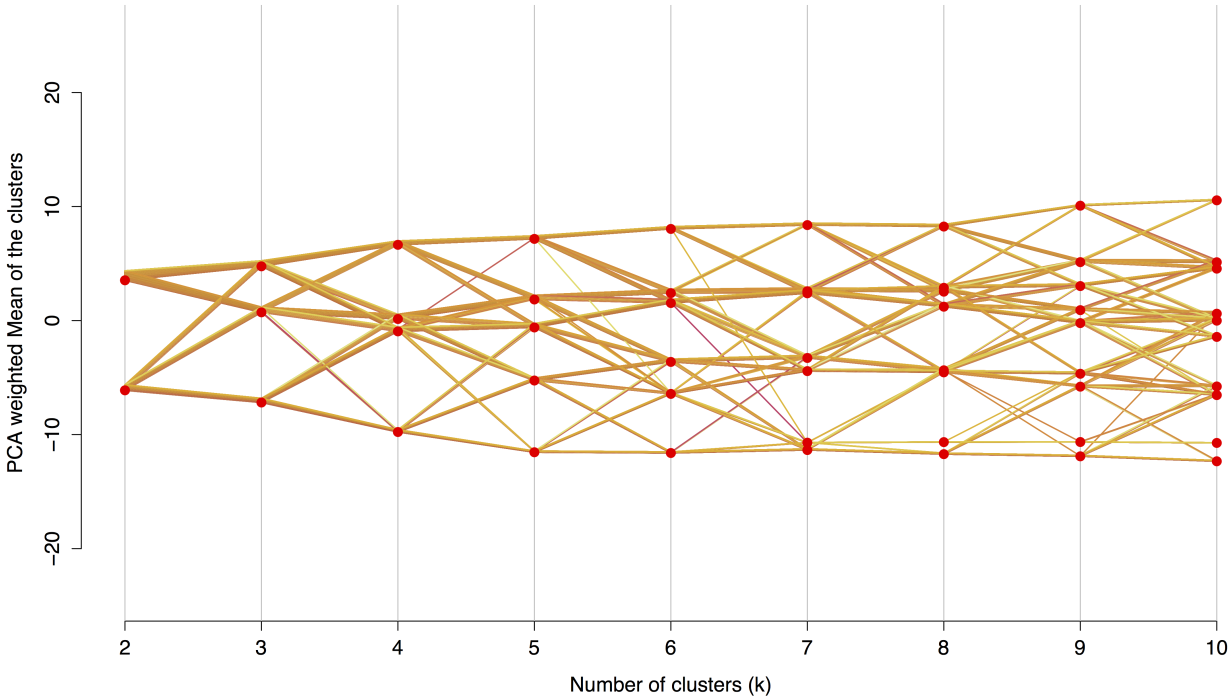


Figure 5.4: Clustergram of the PCA Weighted Mean of K-means Clusters

The method through which each initial cluster was assessed (in terms of how many sub-clusters it could support) was the same as outlined above, using Clustergrams to find optimum values for k, and examining whether outputs were sufficiently differentiated. The latter was completed through calculating cluster means for the input attributes and assessing the differences. The initial clusters were re-clustered to form a second tier of the classification as follows:

- Cluster 1 was re-clustered and formed 3 sub-clusters
- Cluster 2 was re-clustered and formed 2 sub-clusters
- Cluster 3 was re-clustered and formed 2 sub-clusters
- Cluster 4 was deemed too small to support multiple sub-clusters
- Cluster 5 was re-clustered and formed 3 sub-clusters

The creation of a third tier was experimented with, but ultimately it was decided that the sub-clusters could be split no further, with profiling at this level suggesting differentiation between clusters was insignificant. After the final clusters had been formed, the resulting classification was a two-tier nested typology containing 5 clusters and 11 sub-clusters. For the purpose of this classification, clusters and sub-clusters we renamed 'Supergroups' and 'Groups'. The next stage was to study the characteristics of each Supergroup and Group and translate this information into a set of textual and visual summaries.

5.4 Cluster Mapping, Summarisation and Assessment

The process of summarising Supergroup and Group characteristics was achieved using a number of methods. Firstly, these were mapped using a Geographic Information System (GIS) to reveal their geographic distributions. Both Supergroups and Groups were assessed to see if they were predominantly rural or urban, furthermore, were they located in deep rural, rural fringe, major urban, towns or urban fringe areas. Secondly, Supergroup and Group means for each input attribute were recorded and analysed closely. Key characteristics were then recorded through these combined observations and used to create textual summaries ('Pen Portraits'). At this stage, preliminary Supergroup and Group names were created. Graphs were also created to visualise how each cluster differed from the national average across a range of input measures. These visualisations were also created for the purpose of communicating cluster attributes to an end user through a user guide. Combined, pen portraits, cluster mapping and cluster attribute graphs, alongside background information, were used to form a comprehensive user guide, which is attached to this thesis as an appendix. The following sections of this chapter present descriptions of the clusters that emerged from the classification process. Each Supergroup and Group cluster is considered in turn. Cluster attributes are interpreted and pen portraits are presented. For the purpose of creating pen portraits and a user guide, the classification from this point forward is referred to as the Internet User Classification or 'IUC'.

5.4.1 Supergroup 1: E-unengaged

Figure 5.5 shows the national distribution of the first Supergroup within the classification. It is apparent that this Supergroup is concentrated around coastal and rural areas that are associated with elderly populations. This Supergroup does not cluster around major urban centres or metropolitan areas, which typically attract younger populations. Analysis of the cluster means resulted in the Supergroup name "Supergroup 1: E-unengaged" and the following Pen Portrait:

"The E-unengaged Supergroup display apparent low levels of engagement with Internet applications across all measures including; seeking information online, purchasing online, device ownership, general interest and mobile access. The age structure of the E-unengaged Supergroup is significantly skewed towards the elderly, with members most likely to be 60 plus. This Supergroup also has

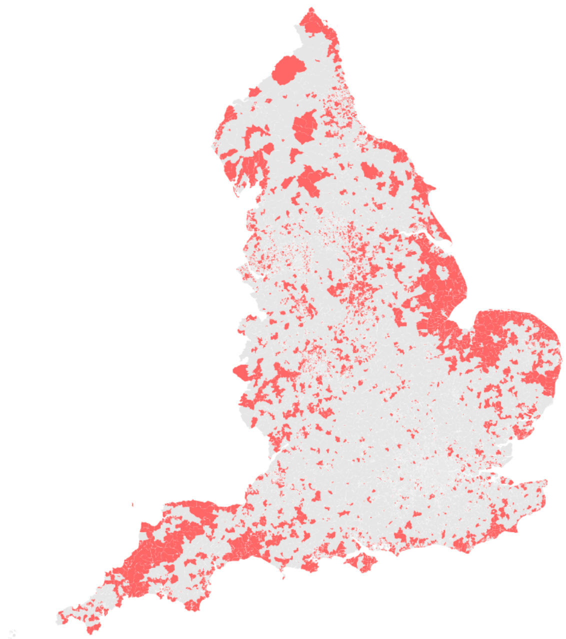


Figure 5.5: Supergroup 1 National Distribution

the highest proportion of residents aged 75 plus, of any Supergroup in the IUC. Members of this Supergroup generally favour traditional means of communication such as telephone and newspapers over their online equivalents. As such, device ownership including smartphones, tablets, e-readers, smart TVs and games consoles is far lower than the national average and the lowest of all Supergroups in the IUC. Infrastructure provision is in line with the national average, although usage falls far below, with Internet non-use significantly higher than the national average. Members of this Supergroup are most likely to be retired or to work in skilled trades or service occupations. Rates of higher-level (level three and above) qualifications are below the national average, most likely due to elderly populations. Geographically, this Supergroup tends to cluster around rural and coastal areas that attract elderly populations, although it is not uncommon for this Supergroup to also appear in urban areas, typically long-established suburbs as opposed to city centre areas. The E-unengaged Supergroup accounts for 24.8% of all Lower Super Output Areas nationally."

5.4.2 Supergroup 2: E-professionals

Figure 5.6 shows the national distribution of the second Supergroup within the classification. Unlike the first Supergroup, it is apparent that this Supergroup is concentrated around major urban conurbations. This supergroup does not cluster around rural areas, or those associated with un-

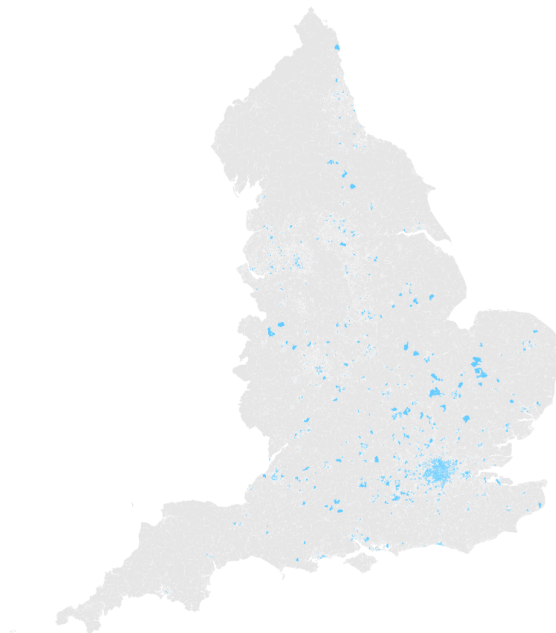


Figure 5.6: Supergroup 2 National Distribution

engaged or elderly populations. Analysis of the cluster means resulted in the Supergroup name "Supergroup 2: E-professionals" and the following Pen Portrait:

"The E-professionals Supergroup display very high levels of engagement with Internet applications across all measures. This Supergroup are characterised by access through multiple devices, favouring access across mobile and fixed line Internet connections to ensure 'always online' connectivity. Seeking information is 'online by default' for this Supergroup, as are most everyday tasks such as banking, account and bill payments and food and grocery shopping. General interest in the Internet for information and entertainment is significantly higher than the national average. Device ownership in this Supergroup is also higher than the national average and the highest of all Supergroups in the IUC, most likely to support such ubiquitous access. This Supergroup are likely to own several Internet enabled devices such as tablet computers, smart TVs, e-readers and networked games consoles. Mobile phone ownership is high, with most users favouring smartphones to support email, social networking, navigation, mobile Internet access and third party apps. As would be expected, current Internet use is higher than the national average within this Supergroup and the highest of any Supergroup in the IUC. Conversely, Internet non-use is far below the national average. The usage characteristics of this Supergroup are strongly underpinned by the socio-economic characteristics of local populations, who are predominantly younger, typically aged between 25 and 44 and highly qualified, in most cases to degree or higher-degree level. In addition, large proportions of this Supergroup have found employment through the Internet. Geographically, this Supergroup tends to cluster around more densely populated urban centres that are well connected in terms of Internet infrastructure and have above average broadband performance. The e-professionals Supergroup accounts for 15% of all Lower Super Output Areas nationally."

5.4.3 Supergroup 3: Typical Trends

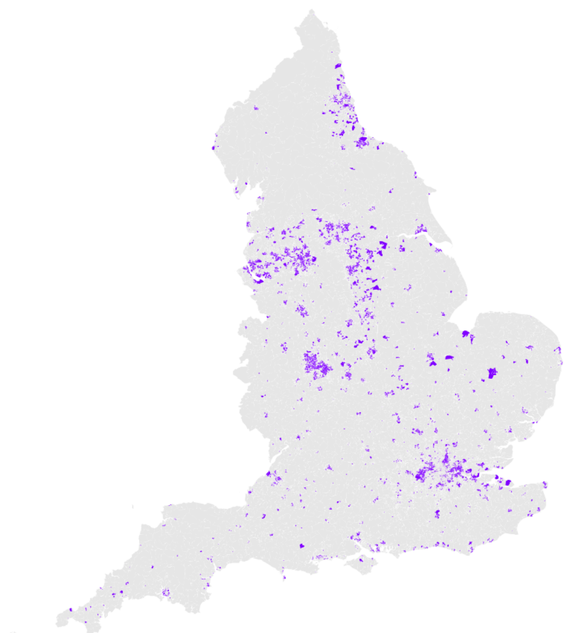


Figure 5.7: Supergroup 3 National Distribution

Figure 5.7 shows the national distribution of the third Supergroup within the classification. This Supergroup has a tendency to cluster within and around urban areas, which in many cases display higher than average levels of material deprivation. There are particularly high concentrations of this Supergroup in traditionally working class areas such as Manchester, Liverpool, Birmingham and Newcastle. Analysis of the cluster means resulted in the Supergroup name "Supergroup 3: Typical Trends" and the following Pen Portrait:

"The Typical Trends Supergroup displays levels of engagement that are the closest to the national average. Individuals are characterised by average engagement in terms of seeking information, device ownership and general interest in the Internet. Use of commercial applications such as online shopping, online banking and online bill payments are slightly below the national average and the use of mobile devices for Internet access is above the national average, in part because the younger individuals within this Supergroup favour mobile use. The Supergroup contains more individuals aged 10 to 17 of any Supergroup in the IUC. Members of this Supergroup who are of a typical working age are not highly qualified, and generally work in elementary or service occupations. Geographically, this Supergroup is clustered within and around urban areas, which in many cases also have higher than average levels of material deprivation. These areas are, however, well connected in terms of Internet infrastructure, and have above average broadband performance. The Typical Trends Supergroup accounts for 27% of all Lower Super Output Areas nationally."

5.4.4 Supergroup 4: E-students

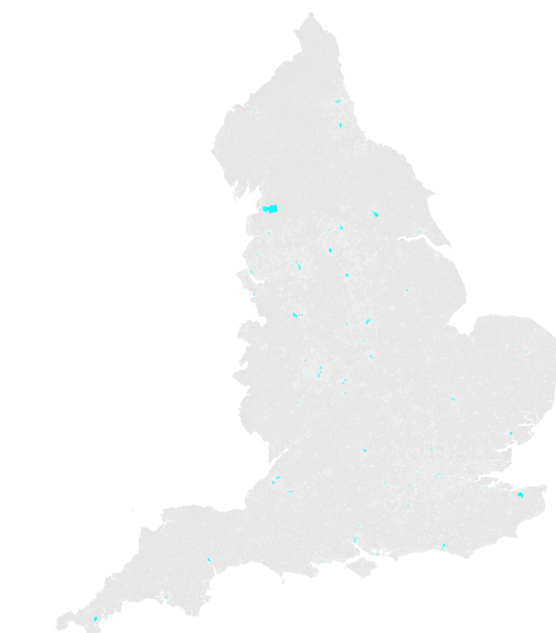


Figure 5.8: Supergroup 4 National Distribution

Figure 5.8 shows the national distribution of the fourth Supergroup within the classification. This Supergroup clusters within and around areas that have high-density student populations. These are both university campus halls of residence and urban areas that traditionally attract large numbers of students, for example, the large area highlighted in the North West which encompases the Lancaster Univerity campus. Analysis of the cluster means resulted in the Supergroup name "Supergroup 4: E-students" and the following Pen Portrait:

"E-Students represents a small but very distinct Supergroup that is comprised almost entirely of student areas. The Supergroup is characterised by very high levels of Internet usage, particularly through mobile devices such as smartphones and tablet computers. Smartphones are the device of choice for electronic communication and are used for a wide range of applications including email, social networking, third party applications, web browsing and sharing photos and videos. Members of this Supergroup are typically aged between 18 and 24 and are registered as full time students. Interest in the Internet for information and entertainment is above the national average, and a higher than average proportion of the local population is likely to have found, or to be seeking, employment through the Internet. With very high proportions of students in this Supergroup, most members are likely to possess Level Three qualifications or above. Employment across all sectors is below the national average with the exception of sales and customer service roles, in which some students choose to work, most likely on a part-time basis to support their studies. Geographically, this Supergroup is often found in the major urban conurbations, usually within city centres and university campus areas where there are highly concentrated student populations. Infrastructure provision and connection performance is above the national average in these areas. The E-Students Supergroup accounts for 1.7% of all Lower Super Output Areas nationally."

5.4.5 Supergroup 5: E-rural and fringe

Figure 5.9 shows the national distribution of the fifth Supergroup within the classification. This Supergroup distinctly clusters within and around areas that are low density and predominantly rural, with constrained infrastructure supply and performance. Analysis of the cluster means resulted in the Supergroup name "Supergroup 5: E-rural and fringe" and the following Pen Portrait:

"Use of the Internet by members of the E-rural and Fringe Supergroup is constrained by poor infrastructure provision, typically because of their predominantly rural locations. Although engagement with Internet applications is only around the national average, this Supergroup display higher than average use of online shopping for products and groceries, online banking, online bill and account payments and price comparisons. This may in part arise because of the limited provision of these services locally. Fixed line broadband connections are used more than mobile broadband. Although mobile phone ownership is in line with the national average, the use of smartphones for data-dependent applications is significantly lower, given poorer infrastructure in these predominantly rural areas. The age structure within this Supergroup is middle aged to elderly with residents most likely to be aged between 45 and 75. Members of this Supergroup are generally well qualified and

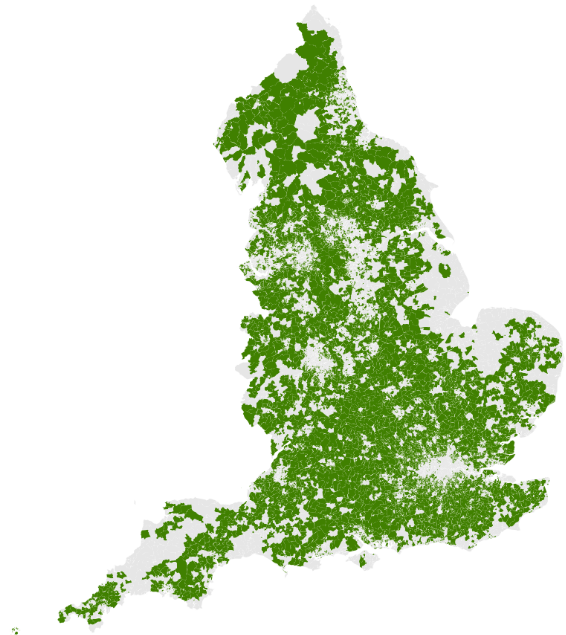


Figure 5.9: Supergroup 5 National Distribution

likely to work in managerial, professional or technical occupations. Device ownership is close to the national average, although devices such as e-readers are favoured over games consoles, consistent with the age structure of these areas. Performance of local broadband connections is below the national average, and the most constrained of all Supergroups in the IUC. The E-rural and Fringe Supergroup accounts for 31.5% of all Lower Super Output Areas nationally."

5.4.6 Supergroup 1: Group 1a: Too Old to Engage

The first of the three Groups formed from the re-clustering of Supergroup 1 was predominantly elderly and displayed very low levels of engagement with Internet applications across all measures. Figure 5.10 shows how this group differs from the national average across a range of engagement measures. Interpretation of the cluster means resulted in the cluster name "Group 1a: Too Old to Engage" and the following Pen Portrait:

"The Too Old to Engage Group is characterised by large elderly populations who show little or no engagement with the Internet across all applications. The proportion of residents aged 75 plus is higher than any Group in the IUC. As a result, Internet enabled device ownership is lower than the Supergroup average, and the lowest of any Group in the IUC. Abstinence from Internet use is

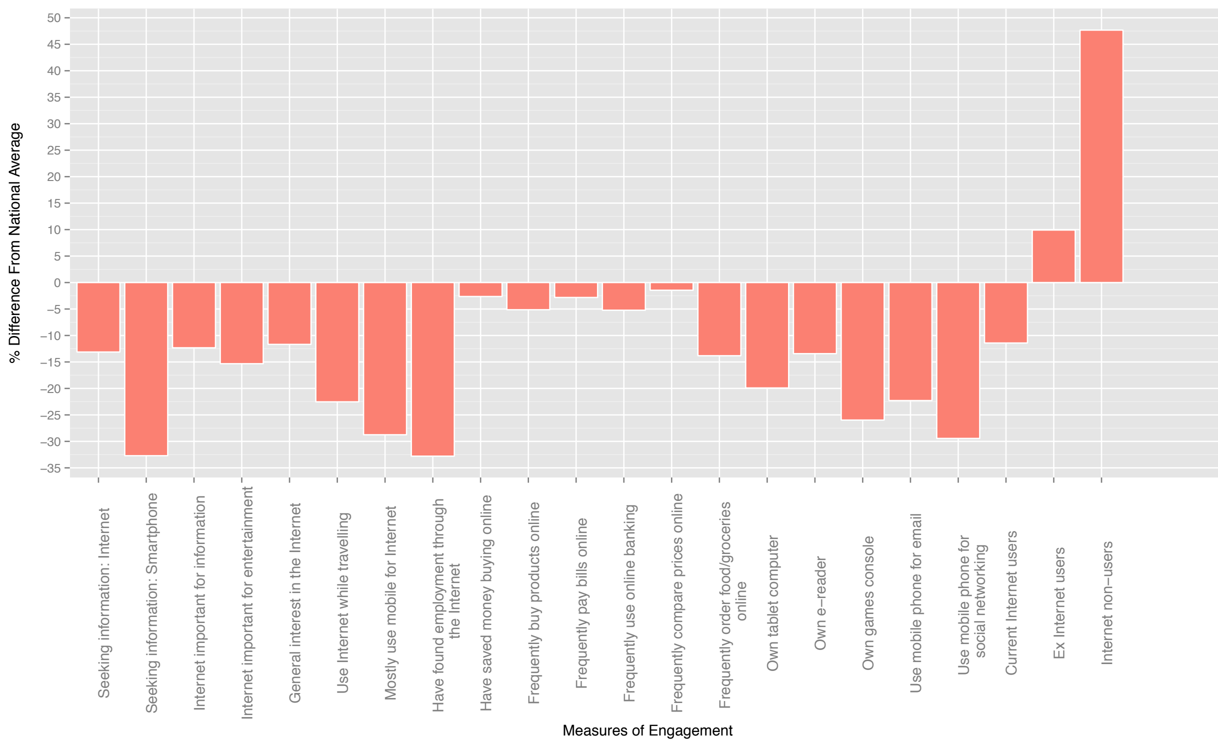


Figure 5.10: Group 1a: Too Old to Engage: Group Characteristics

higher than the Supergroup average and far above the national average. Enclaves of this Group are found in coastal and lower density rural areas that serve as retirement destinations. Infrastructure provision and performance is typically slightly below the national average. The Too Old to Engage Group accounts for 4% of all Lower Super Output Areas nationally."

5.4.7 Supergroup 1: Group 1b: E-marginals: Not a Necessity

The second of the three Groups formed from the re-clustering of Supergroup 1 was characterized by low levels of engagement with Internet applications, low levels of qualifications and higher levels of employment in non-technical occupations. Figure 5.11 shows how this group differs from the national average across a range of engagement measures. Interpretation of the cluster means resulted in the cluster name "Group 1b: E-marginals: Not a Necessity" and the following Pen Portrait:

"Members of the E-marginals: Not a Necessity Group typically have low engagement with Internet applications, lower than average qualifications and higher than average rates of employment in blue collar occupations that are not heavily reliant on digital skills. Of those that do access

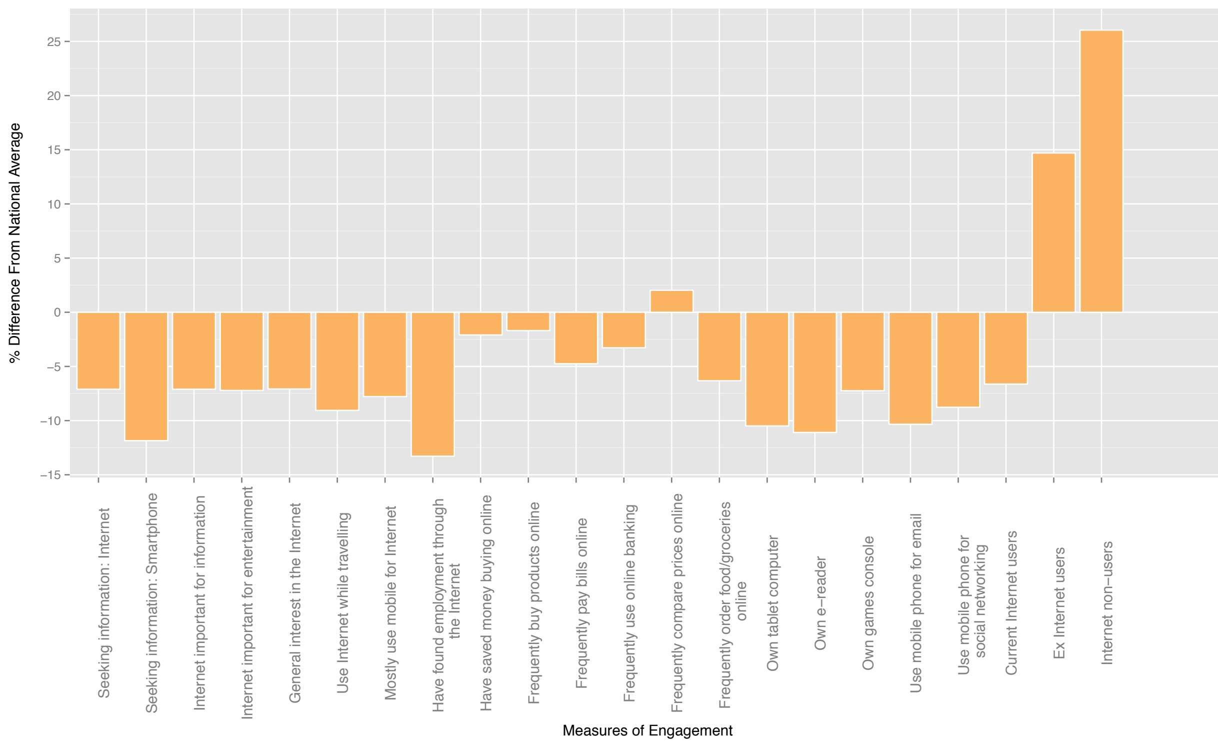


Figure 5.11: Group 1b: E-marginals: Not a Necessity: Group Characteristics

the Internet, many do so using a smartphone. Residents of this Group tend to be found within urban areas characterised by high levels of material deprivation, although infrastructure provision and performance are in line with the national average. The E-marginals: Not a Necessity Group accounts for 10.4% of all Lower Super Output Areas nationally.

5.4.8 Supergroup 1: Group 1c: E-marginals: Opt Out

The last of the three Groups formed from the re-clustering of Supergroup 1 was characterized by affluent rural and fringe populations who are highly qualified professionals or retirees. Internet use within this Group is not a necessity, but an option that is less favorable than traditional media such as newspapers and television. Figure 5.12 shows how this group differs from the national average across a range of engagement measures. Interpretation of the cluster means resulted in the cluster name "Group 1c: E-marginals: Opt Out" and the following Pen Portrait:

"The E-marginals: Opt Out Group are characterised by low levels of engagement with the Internet for applications such as seeking information and entertainment, preferring instead more traditional

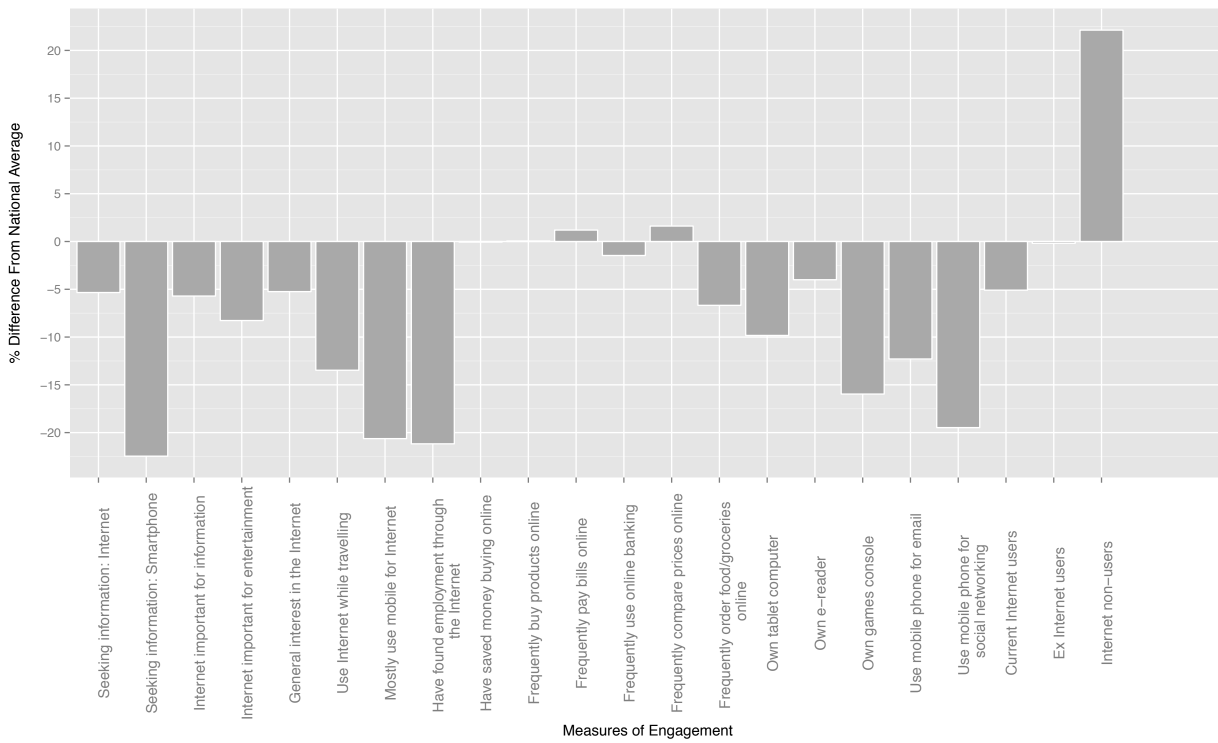


Figure 5.12: Group 1c: E-marginals: Opt Out: Group Characteristics

media such as newspapers and television, in part reflecting the elderly demographic of this Group. Typically this Group is aged 60 plus, with significantly higher than average incidence of those aged 65 to 84. Geographically, this Group tends to be found in affluent rural and fringe areas that are more sparsely populated and where infrastructure provision and performance is below the national average. Access to the Internet through mobile devices is below the national average. Those who do choose to use the Internet tend to use it for price comparison and occasional online shopping. Levels of qualifications are generally above the national average, and those members who are not retired will typically be employed in senior managerial, professional or skilled trade occupations. Abstinence is significantly higher than the national average, but the lowest within the Supergroup. The E-marginals: Opt Out Group accounts for 10.4% of all Lower Super Output Areas nationally."

5.4.9 Supergroup 2: Group 2a: Next Generation Users

The first of the two Groups formed from the re-clustering of Supergroup 2 was characterized by high levels of Internet use across a range of next generation applications. Device ownership within this group is also high, and enables access on the move. Figure 5.13 shows how this group differs from

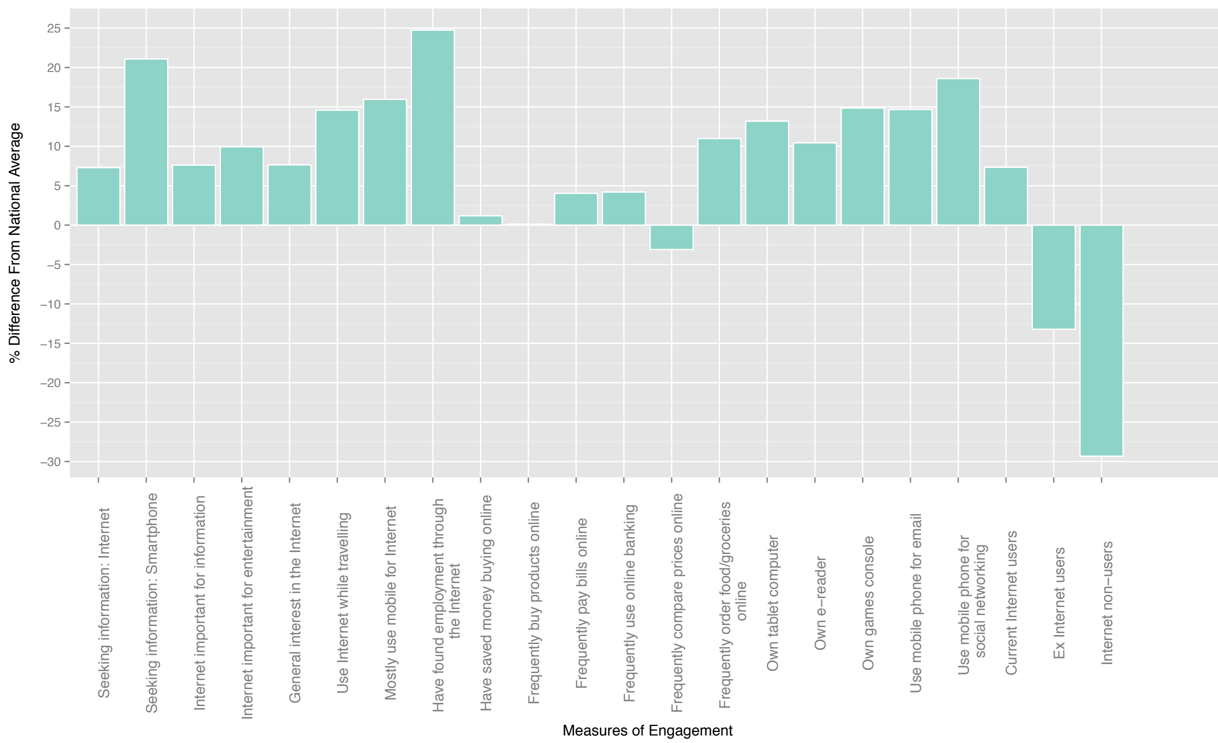


Figure 5.13: Group 2a: Next Generation Users: Group Characteristics

the national average across a range of engagement measures. Interpretation of the cluster means resulted in the cluster name "Group 2a: Next Generation Users" and the following Pen Portrait:

"The Next Generation Users Group is characterised by high levels of engagement across all applications of the Internet. Members of this Group are heavy smartphone users and typically access the Internet on the move and for applications such as email, social networking and navigation. However, they favour fixed line connections for most other tasks such as general browsing and seeking information. Device ownership is higher than the national average, and members of this Group are likely to own several Internet enabled devices, such as tablet computers, e-readers and smart TVs. Levels of qualification are high within this Group, with higher than average rates of degree and higher degree level qualifications. The age structure is young to middle aged, with members of this Group most likely aged between 25 and 44, and in some cases with young children. Employment tends to be in managerial, professional and technical occupations. General interest in the Internet is above the national average. Members of this Group are found in affluent, higher density suburban and city fringe areas where infrastructure provision and performance is above the national average.

Next Generation Users are the second most heavily engaged Group within the IUC, behind Group 2b: Totally Connected and account for 10.2% of all Lower Super Output Areas nationally."

5.4.10 Supergroup 2: Group 2b: Totally Connected

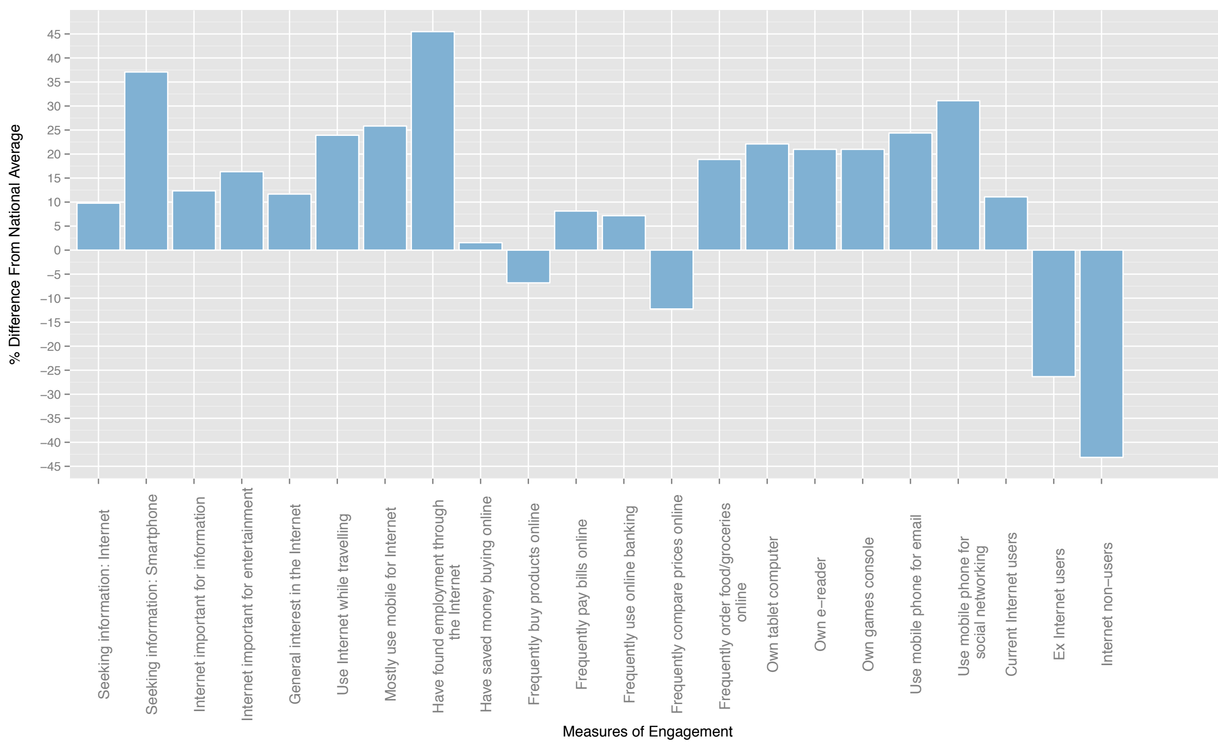


Figure 5.14: Group 2b: Totally Connected: Group Characteristics

The second of the two Groups formed from the re-clustering of Supergroup 2 was characterized by high levels of Internet use across all applications, with the exception of online shopping. This is due to this group being located in high-density urban centres where the provision of local retail is high. Figure 5.14 shows how this group differs from the national average across a range of engagement measures. Interpretation of the cluster means resulted in the cluster name "Group 2b: Totally Connected" and the following Pen Portrait:

"The Totally Connected Group is characterised by the highest levels of engagement within the IUC and score higher than the Supergroup and national averages for most measures of engagement. This Group displays a clear preference to use the Internet by default for almost all applications. Members of this Group access the Internet through multiple devices, whilst on the move and in the home to ensure seamless connectivity. As such, device ownership is significantly higher than the national and Supergroup averages and members of this Group own a wide range of Internet enabled hardware. Levels of qualification are significantly higher than the national average. Professional occupations are most prevalent, with the age structure of residents being young to middle aged, sometimes with young children. Geographically, this Group tends to be found in affluent city centre and city fringe areas that are densely populated and where infrastructure provision and performance is above the national average. Members of this Group show below average rates of online shopping, perhaps given good local retail choice. However, rates of online shopping for food and groceries are significantly above the national and Supergroup averages as this enables wider choice and convenience in highly populated areas. Totally Connected are the most heavily engaged Group within the IUC and account for 4.8% of all Lower Super Output Areas nationally."

5.4.11 Supergroup 3: Group 3a: Uncommitted and Casual Users

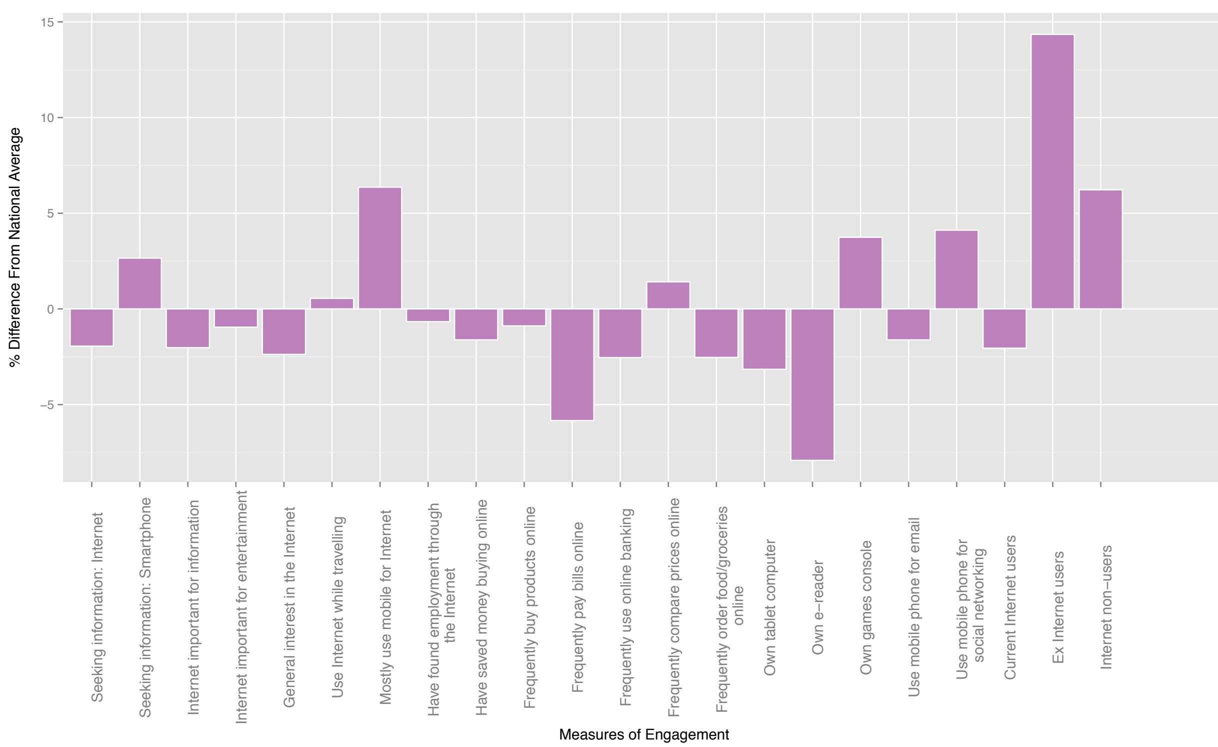


Figure 5.15: Group 3a: Uncommitted and Casual Users: Group Characteristics

The first of the two Groups formed from the re-clustering of Supergroup 3 was characterized by mixed levels of engagement with Internet applications. The group clusters around major urban and city fringe areas where levels of material deprivation are higher than average. The group also includes larger numbers of ex Internet users. Figure 5.15 shows how this group differs from the national average across a range of engagement measures. Interpretation of the cluster means resulted in the cluster name "Group 3a: Uncommitted and Casual Users" and the following Pen Portrait:

"The Uncommitted and Casual Users Group are characterised by mixed levels of engagement with the Internet. Access to the Internet through smartphones is marginally above the national average and access through fixed-line connections falls marginally below. Members of this Group show below average rates for purchasing online but above average rates for price comparison and selling online. Age structure is generally young to middle aged, with higher than average proportions of young and teenage children. Qualifications tend to be of a lower level and members of this Group are most likely to work in service, sales and elementary occupations. Overall, abstinence from Internet use is marginally higher than the national average and general interest in the Internet falls shy of the national average. This Group also contains higher than average numbers of lapsed Internet users. Geographically, this Group tends to be found in major urban and city fringe areas that suffer higher levels of material deprivation, but where infrastructure provision and performance is above the national average. The Uncommitted and Casual Users Group accounts for 15.5% of all Lower Super Output Areas nationally."

5.4.12 Supergroup 3: Group 3b: Young and Mobile

The second of the two Groups formed from the re-clustering of Supergroup 3 was again characterized by mixed levels of engagement with Internet applications. This group however, show a higher propensity for mobile Internet use and access through smartphones for many common applications. The age demographic is generally younger, with large numbers of teenagers and young adults. The group mainly clusters around urban areas with mixed levels of deprivation. Figure 5.16 shows how this group differs from the national average across a range of engagement measures. Interpretation of the cluster means resulted in the cluster name "Group 3b: Young and Mobile" and the following Pen Portrait:

"The Young and Mobile Group is predominantly young and has a tendency to access the Internet using mobile devices rather than fixed line connections. This Group is found in major urban conurbations where population density is above average and infrastructure provision is sufficient to support heavy mobile broadband usage. These areas are typically inner city or city fringe and experience mixed levels of material deprivation. As a Group there are higher than average proportions of young and teenage children and adults aged 25 to 44. Conversely, the proportion

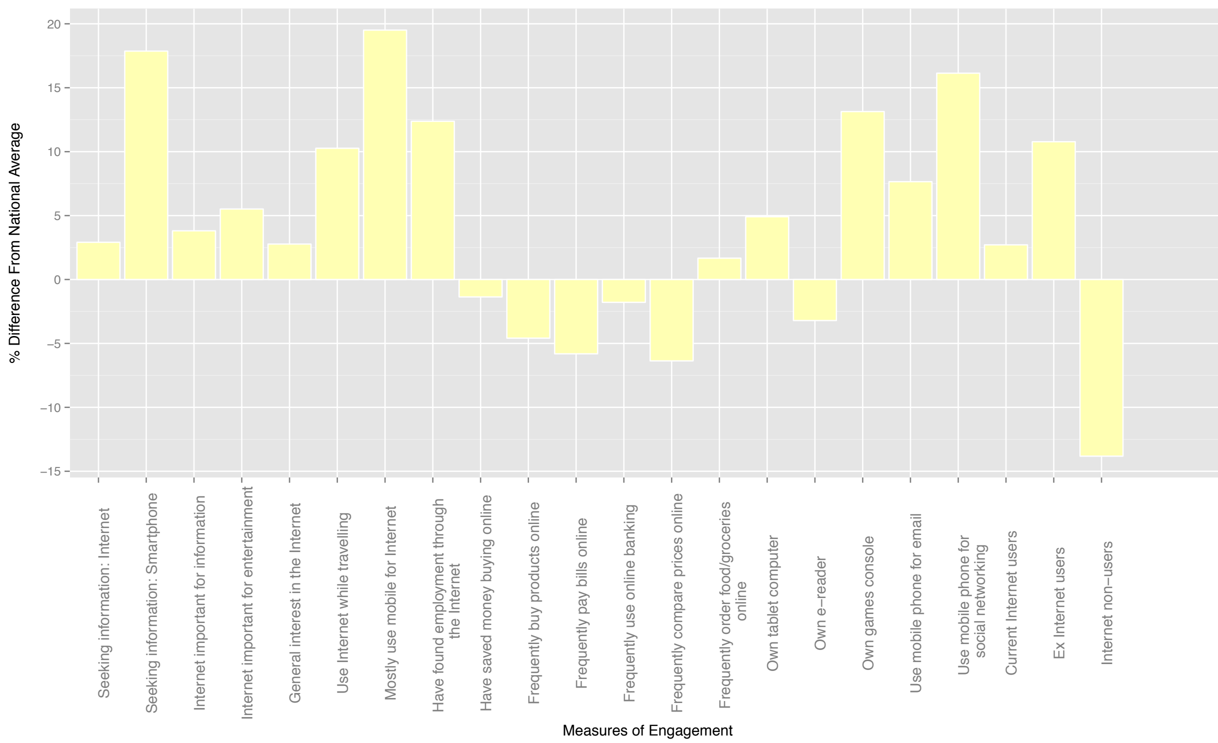


Figure 5.16: Group 3b: Young and Mobile: Group Characteristics

of adults aged over 45 falls below the national and Supergroup averages. All levels of qualification are below the national average and those who work are likely to be employed in elementary, sales or service occupations. Interest in the Internet for entertainment and information is above the national average, most likely reflecting the prevailing age structure. This Group displays a lower than average tendency to purchase online, and would be expected to shop locally in most cases. The Young and Mobile Group accounts for 11.5% of all Lower Super Output Areas nationally."

5.4.13 Supergroup 4: Group 2c: Students Online

Because the size of Supergroup 4 (E-students) was small, at around 1.7% of all LSOAs nationally, it was not re-clustered to form groups. Instead it was decided that this cluster would be merged with Supergroup 2 to form an additional group (2c), as in terms of use, demographics, qualifications and device ownership, these clusters displayed the most similar characteristics. Although originally classified as a Supergroup cluster, it is more appropriate to consider E-students as a group, given the relatively small size. With this in consideration, merging this cluster was deemed to be the most logical solution to avoid having a Supergroup within the classification that was too small to

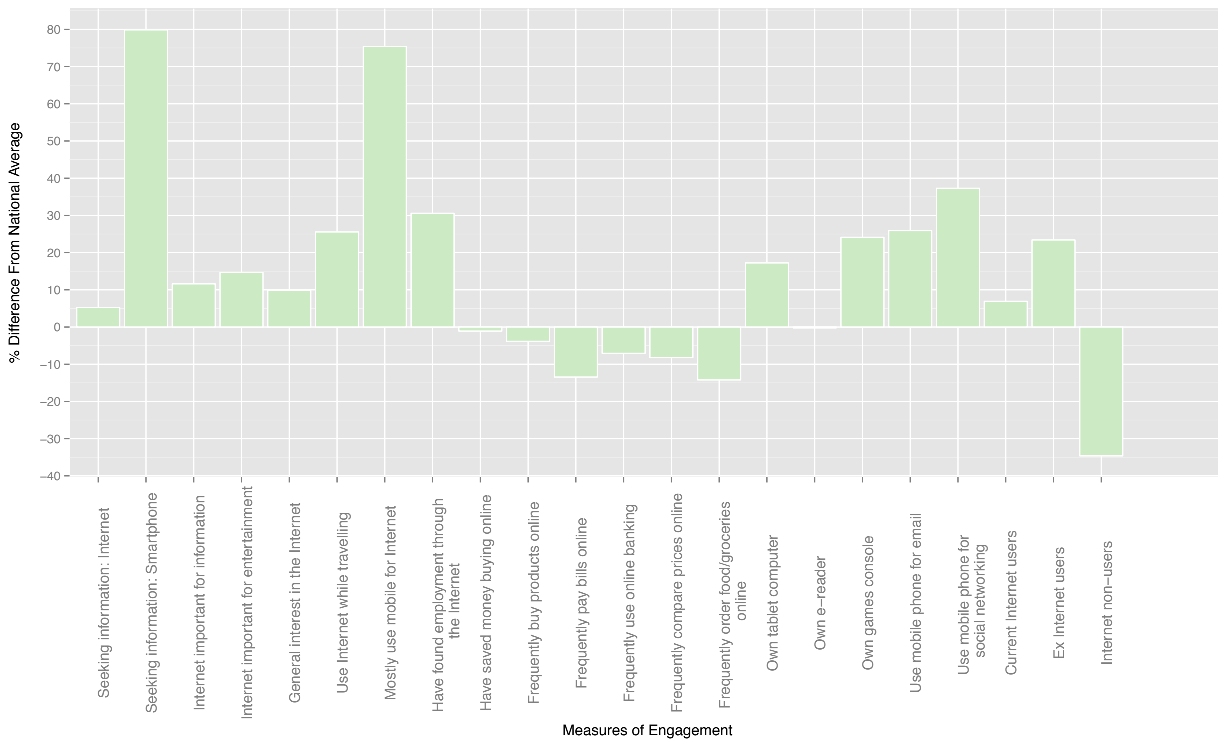


Figure 5.17: Group 2c: Students Online: Group Characteristics

support any number of groups.

The Pen Portrait for this newly created group remained unchanged from the original Supergroup classification as the cluster had not been modified, however, the cluster name was updated and assignment code changed to reflect its new Supergroup membership. At this stage the original Supergroup 4: E-students had been reclassified to Group 2c: Students Online. Figure 5.17 shows how the Students Online group differs from the national average across a range of engagement measures.

5.4.14 Supergroup 5: Group 5a: E-fringe

The first of the three Groups formed from the re-clustering of Supergroup 5 was characterized by its location in low density or semi-rural fringe areas. The group displays lower than average general interest in the Internet, but a higher than average propensity for purchasing, paying bills, banking and comparing prices online, services which save travel to local retail centres. Figure 5.18 shows how this group differs from the national average across a range of engagement measures. Interpretation of the cluster means resulted in the cluster name "Group 5a: E-fringe" and the following Pen

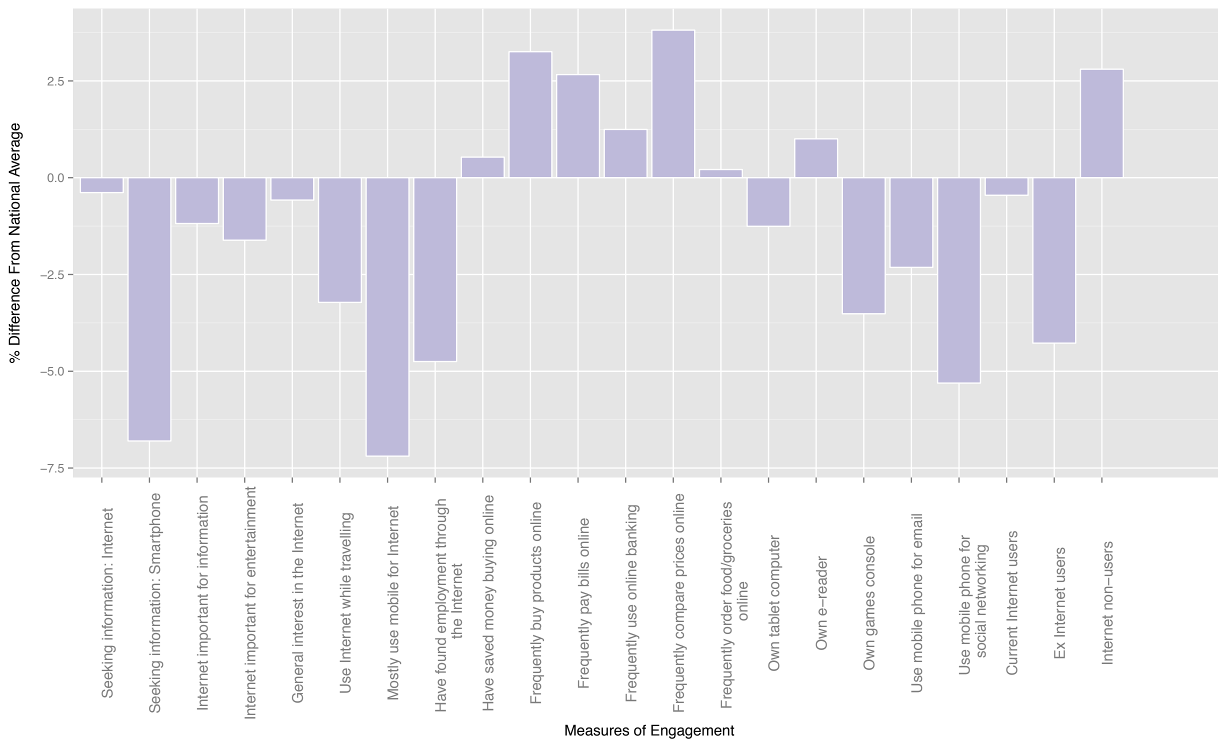


Figure 5.18: Group 5a: E-fringe: Group Characteristics

Portrait:

"The E-fringe Group is distinguished by its location around the fringes of urban areas that are typically low density or semi-rural. Age structure is middle aged to elderly and there are fewer than average numbers of young adults aged 18-29, a group who are likely to have moved to more major urban conurbations. General interest in the Internet within this Group is slightly below the national average and the lowest within the Supergroup, rates of current Internet users are also below average and numbers of Internet non-users are above the national average. Members of this Group generally have mixed levels of qualifications, rates of level 1 to level 3 qualifications are all above the national average and incidence of persons with no qualifications is below average. Members are most likely to work in administrative and secretarial or skilled trade occupations. The most common uses of the Internet within this Group are paying bills and banking online, comparing prices and buying products, which score above the national average. Below average rates are recorded for seeking information and entertainment purposes, consistent with the age profile of this Group. Equally, ownership of Internet enabled devices is below average, with the exception of e-readers, which are popular amongst this Group. Infrastructure provision and performance is marginally below the national average but would be unlikely to limit access. The E-fringe Group accounts for 11.1% of all Lower Super Output Areas nationally."

5.4.15 Supergroup 5: Group 5b: Constrained by Infrastructure

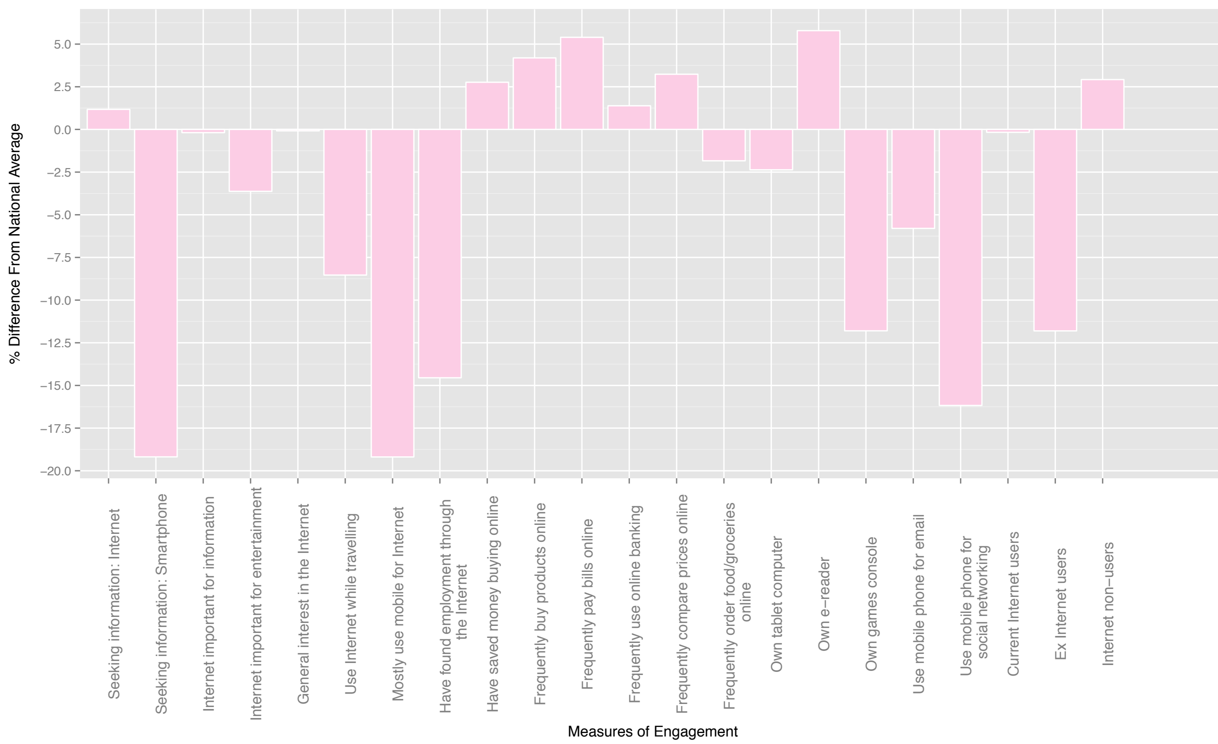


Figure 5.19: Group 5b: Constrained by Infrastructure: Group Characteristics

The second of the three Groups formed from the re-clustering of Supergroup 5 was characterized by its location in low density and rural areas which are heavily constrained by infrastructure provision and performance. Despite limited access in these areas, this group displays a higher than average propensity for shopping online, likely due to a limited provision of local retail and increased journey times to access key services. Figure 5.19 shows how this group differs from the national average across a range of engagement measures. Interpretation of the cluster means resulted in the cluster name "Group 5b: Constrained by Infrastructure" and the following Pen Portrait:

"The Constrained by Infrastructure Group is characterised by locations in low-density rural areas where there is poor provision and performance of local Internet infrastructure, both fixed line and mobile. This limits engagement with some online applications. Fixed line broadband performance falls significantly below the national average and is the lowest within the Supergroup as distances to local telephone exchanges are much higher. Distances to the nearest mobile base station for cellular and data coverage are also higher than the national average, and as such further constrains performance and usability. Perhaps as a result, the use of mobile broadband through devices such as smartphones or dongles is below average. Despite poor infrastructure, general interest in the Internet is in line with the national average and members of this Group display above average rates of purchasing online, comparing prices, online banking and paying bills, most likely as this saves travelling to a local retail centre to access these services. Internet enabled device ownership is again lower than the national average with the exception of e-readers, likely due to the prevailing age structure of this Group, which is middle aged and elderly. Those who are not retired are generally highly qualified and work in managerial, professional or technical occupations. Internet non-use is above average but reflects the prevailing age profile of the Group. The Constrained by Infrastructure Group accounts for 11% of all Lower Super Output Areas nationally."

5.4.16 Supergroup 5: Group 5c: Low Density but High Connectivity

The last of the three Groups formed from the re-clustering of Supergroup 5 was characterized by rural and low-density locations that are well connected in terms of infrastructure provision and performance. In addition to high connectivity for such disparate areas, local populations show a higher than average propensity for general Internet use and Internet enabled device ownership. Figure 5.20 shows how this group differs from the national average across a range of engagement measures. Interpretation of the cluster means resulted in the cluster name "Group 5c: Low Density but High Connectivity" and the following Pen Portrait:

"The Low Density but High Connectivity Group is found in areas that are sparsely populated, typically rural and semi-rural areas, or areas with urban parkland. Despite disparate populations, this Group is generally well connected and displays the strongest infrastructure and performance characteristics within the Supergroup, generally falling in line with the national average. Internet use in general is higher across all applications than the Supergroup average, and this Group shows a higher than average propensity for ordering food and groceries online. These characteristics are representative of the prevailing demographic of well-educated workers (often with degrees or higher degrees) who work in high-grade professional occupations. Similarly, Internet enabled device ownership is above the national average, perhaps because local infrastructure is able to support this. Age structure is mixed, although members of this Group are most likely to be aged 45 to 59 with young or teenage children. General interest in the Internet is above the national average and is

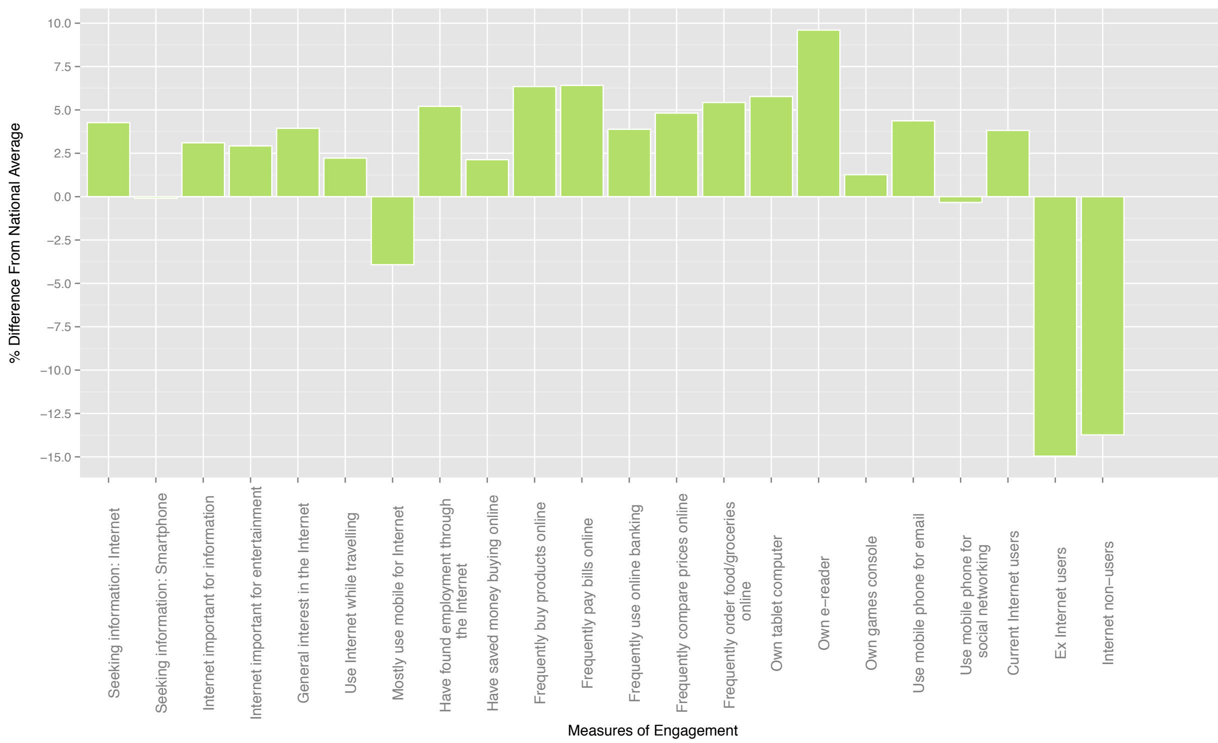


Figure 5.20: Group 5c: Low Density but High Connectivity: Group Characteristics

the highest within the Supergroup. As would be expected, rates of Internet non-use are below the national average. The Low Density but High Connectivity Group accounts for 9.4% of all Lower Super Output Areas nationally."

5.5 Final IUC Hierarchy

Table 5.10 presents the finalised hierarchy of the IUC.

Table 5.10: Final Classification hierarchy

Supergroup	Group
1: E-unengaged	1a: Too Old to Engage
	1b: E-marginals: Not a Necessity
	1c: E-marginals: Opt Out
2: E-professionals and Students	2a: Next Generation Users
	2b: Totally Connected
	2c: Students Online
3: Typical Trends	3a: Uncommitted and Casual Users
	3b: Young and Mobile
4: E-rural and Fringe	4a: E-fringe
	4b: Constrained by Infrastructure
	4c: Low Density but High Connectivity

An interactive map of the classification is available on the companion website:
`http://maps.cdrc.ac.uk/#/geodemographics/iuc14/`

and can also be downloaded here:
`https://data.cdrc.ac.uk/dataset/cdrc-2014-iuc-geodata-pack-england`

5.6 Conclusions

This chapter has presented the process through which a classification of Internet use and engagement has been developed, drawing on a range of complex univariate outputs explored through previous chapters to create a composite nested typology.

A total of 76 variables pertaining to LSOA level observations were assessed for use in this classification as they were deemed to capture important characteristics relating to Internet use and engagement. Of these 76 variables, 50 per cent were deemed to have some level of skew, with around 28 per cent highly skewed. Most of the highly skewed variables reflected real world observations that are naturally unevenly distributed, such as the proportion of elderly residents or population density measurements. However, given that variable skew can influence cluster formation, variables were transformed using log10 and Box-Cox transformation methods to reduce overall skew. Through the application of these two methods, it was found that Box-Cox transformations significantly outperformed log10 in reducing overall variable skew. Despite exploring these techniques, naturally distributed data was used to build a final classification. A comparison of cluster outputs from naturally distributed and transformed data suggested that applying normalisation techniques to reduce variable skew and suppress outliers resulted in less distinctive clusters.

The classification was built using a top down clustering method with a K-means algorithm, using Clustergrams to aid in the selection of an optimum number of clusters at the Supergroup and Group levels. After clustering, Supergroups and Groups were assessed in terms of their aggregate characteristics, mapped and summarised with Pen Portraits. For the group level of the classification, a series of variables were graphed against their national averages for each cluster, demonstrating how each group's characteristics varied from that of the national average. The primary output of this chapter is the Internet User Classification (IUC), a geodemographic classification of Internet use and engagement covering England at LSOA level.

Potential uses of the IUC are broad, and fields of use may include; data profiling, online survey stratification, targeted marketing, location planning, customer insight, and public policy formation and delivery. In a public sector setting, such profiling could be necessary in the planning and implementation of the 2021 national census, which will be predominantly completed online. Such a delivery mechanism is heavily dependent on the online engagement and access characteristics of small area populations. As such, the IUC may assist in highlight areas where local populations may not be equipped (in terms of both access to, and general engagement with, Internet technologies) to complete a census questionnaire online. In the private sector, the IUC could potentially be used in profiling of existing customer databases to identify trends, assisting in the development of targeted marketing strategies. This may be valuable for businesses that operate online, or are interested in the aggregate Internet engagement characteristics of their customer base.

Chapter 6

Geodemographic Application

The purpose of this chapter is to validate the Internet User Classification through presentation of a detailed use case. The use case presented here represents an Economic and Social Research Council (ESRC) funded study, conducted in 2015 and focusing on the e-Resilience of retail centres in England, assessing the extent to which retail centres have spatially differentiated vulnerability to the impacts of online consumption. The IUC formed the basis of this study and the following chapter demonstrates the applicability of the classification in a research context, where the Internet use and engagement characteristics of small area populations are of interest.

The following chapter is adapted from a co-authored paper, a copy of which is included in the appendix, and has been published as:

Singleton, A. D., Dolega, D., Riddlesden, D. and Longley, P. A. (2016). Measuring the Spatial Vulnerability of Retail Centres to Online Consumption through a Framework of e-Resilience. *Geoforum*, 69(1): 5-18

6.1 Introduction and Background

UK Government initiatives aimed at revitalisation of British high streets (Portas, 2011; Digital High Street Advisory Board, 2015) highlight the importance of digital technology in redefining traditional retail spaces. Evidence suggests that growth in online consumption impacts upon the health of retail centres in complex ways, and can be viewed as a source of long-term change to their structure, often referred to in the literature as a 'slow burn' (Pendall et al., 2010). Adjustments to the market share of traditional town centre retailing have been mainly considered with respect to their supply side effects: for example, the extent to which online shopping has substituted, modified or complemented traditional town centres (Doherty and Ellis-Chadwick, 2010). However, there has been less focus on how the structure of traditional high streets are or might be impacted by consumer propensity for online shopping, how such effects could be modelled, or what might be an appropriate adaptive response by stakeholders in retail. Despite evidence to suggest that factors impacting decisions about whether or not to shop online are linked to demographic and socio-economic characteristics of populations (Longley and Singleton, 2009a), there is limited knowledge about the geography of online sales (Forman et al., 2008).

The research explored these challenges through developing the concept of e-resilience, which provided both a theoretical and methodological framework that defined the vulnerability of retail centres to the effects of rapidly growing Internet sales, balancing characteristics of both supply and demand. It was argued that the concept of e-resilience added value to existing research in the following ways:

- It provides insight into wider debates on the performance of UK town centres in the rapidly transforming retail landscape, in particular by assessing their resilience and adaptability to the growth of online sales
- It provides insight into demand through examination of local catchment demographics, and thus rebalances current debates on the resilience of high streets, which hitherto have predominantly focused on supply effects, measured through outcomes such as retailer failures or vacancy rates
- It delivers valuable outputs including: an operational measure of e-resilience, which is implemented to define those retail centres in England that are the most or least e-resilient
- It presents a national geodemographic of Internet user behaviour and it is anticipated that such outputs will be of interest to a wide range of stakeholders in retail policy and provision

The general concept of resilience has been established for some time to describe how various types of system respond to unexpected shocks. There are three widely recognised concepts of resilience adopted between different scientific traditions (Simmie and Martin, 2010): (a) the engineering resilience interpretation found in physical sciences; (b) the ecological resilience interpretation found in biological sciences; and (c) the adaptive resilience interpretation found in complex systems theory. The first two interpretations refer to the notion of equilibrium, which suggest that a resilient economic system would adapt successfully to disturbance and either resume, or even improve its long-run equilibrium growth path. Conversely, a non-resilient system would fail to transform itself successfully and instead become 'locked' into a long-run outdated trajectory or decline (Simmie and Martin, 2010; Dawley et al., 2010). The third interpretation, identified by Martin (2011) as 'adaptive resilience', stresses the anticipatory or reactive reorganisation of the form and function of a system to minimise the impact of a destabilising shock, and focuses on resilience as a dynamic and evolutionary process. Complex system theory is characterised by non-linear dynamics and self-reinforcing interactions among a system of components (Martin and Sunley, 2007), and highlights self-organisation, with adaptive growth relative to changes in the external environment (e.g. the impact of online sales on traditional retailers).

Increasing numbers of social scientists have also begun to use resilience as a mechanism to help explain the impact and response to disruptions and more gradual processes of change in a range of socio-economic systems (Christopherson et al., 2010; Pendall et al., 2010; Hassink, 2010; Simmie and Martin, 2010; Martin, 2011). For example, resilience was first considered within the context of the UK high street by Wrigley and Dolega (2011), who investigated the dynamics of performance and adjustment to the shock of the global economic crisis. In this work they rejected the notion that town centres and high streets could return to their pre-shock configurations, and developed the concept of 'adaptive resilience' whereby the resilience of UK town centres was viewed as a dynamic and evolutionary process. More specifically, they argued that the response of UK town centres to economic and competitive shocks can be seen as a function of the mix and interdependencies of existing business, the dynamics of centres, diversity, attractiveness, accessibility, national planning policies and the socio-demographic characteristics of local catchments. Such characteristics and actions are responsible for building town centre adaptive capacity. Often an economic or competitive shock creates new opportunities for development and innovation (Pendall et al., 2010; Raco and Street, 2012) which, in turn, leads to the emergence of more adaptable town centres characterised by enhanced resilience and ability to more effectively withstand future disturbances. The resilience framework strengthens some basic arguments derived from evolutionary economics such as the advantages of diversity, seeing regional economies as path-dependent systems (Hassink, 2010), or the potential of novelty and selection in economic systems as they adjust to evolving circumstances

(Simmie and Martin, 2010). Furthermore, Dolega and Celinska-Janowicz (2015) argue that future resilience of town centres is crucially dependent upon recognising and acting upon the challenges arising from current trends. A good example of such pre-emptive action in the UK was the establishment of the Digital High Street Advisory Board in 2014 to provide an independent assessment of strategies to revitalise high streets in the context of a digital future.

Equally important to retail centre resilience is an understanding of the geodemographic characteristics of local catchments (Birkin et al., 2002), as consumer choices and behaviours are fundamental drivers of demand, and therefore are closely related to evolution of the retail landscape. Importantly, the behaviours and attitudes of consumers vary spatially, yet are directly linked to the geography of demand for retail facilities. Understanding the geography of consumer behaviour (such as varied propensity for online shopping) at a small area level is crucial to understanding the vitality and viability of both retail centres and the retailers themselves. Indeed, the resilience of retail centres is intertwined with the underlying dynamics of their catchments as variations in consumer confidence (Wrigley and Dolega, 2011) and basic digital skills (Digital High Street Advisory Board, 2015) are likely to shape demand for retail spaces in the digitally transformed retail landscape. The current debate on economic health of UK town centres seems to acknowledge the key role consumers have in that transformation, and a direct response of retail spaces to consumers' needs is being perceived as key to their success (DCLC, 2013).

6.2 A Framework to Understand and Measure e-resilience

The Internet enhances opportunities for price comparison, enables 24/7 convenience, provides a selection of products not limited by physical space, and enables distribution with a wider geographical reach (Williams, 2009). As such, it is perhaps unsurprising that online sales have been growing exponentially; essentially tripling over the past eight years, and are forecasted to reach 15.2% of all UK retail sales by the end of 2015 (Centre for Retail Research, 2015). Furthermore, in recent years, there has also been a shift toward using mobile devices for online shopping, such as tablets and smartphones, which are now estimated to account for the majority of growth in UK online retail sales (Capgemini, 2015). In consequence, the rapid expansion of online shopping has been increasingly viewed as a major cause of change to the structure of traditional UK high streets (Digital High Street Advisory Board, 2015; Wrigley and Lambiri, 2014). Weltevreden (2007) investigated the implications of e-commerce on traditional physical shopping space, and established the extent to which online retailing could be associated with processes of substitution, complementarity, and modification of traditional retail channels. Substitution occurs when e-commerce replaces physical

shopping; however, complementarity and modification pertain to a blending of e-commerce with traditional retailing. These latter two omni-channel processes are however, less well understood (Wrigley and Lambiri, 2014; Weltevreden, 2007).
For instance, in the UK a number of national retailers such as Borders, Zavvi, Jessops and Game have either entirely withdrawn or substantially limited their physical retail offerings within the past few years, while some other major retailers such as John Lewis, Next, Boots or Argos have successfully embraced new technologies through opening 'click and collect' points, or by developing mobile applications (Turner and Gardner, 2014).

The basic concept of e-resilience defines the vulnerability of retail centres to the effects of growing Internet sales, and estimates the likelihood that their existing infrastructure, functions and ownership will govern the extent to which they can adapt to or accommodate these changes. Essentially, e-resilience can be expressed as a balance between the propensity of localised populations to engage with online retailing and the physical retail provision and mix that might increase or constrain these effects, as not all retail categories would be equally impacted. However, estimating the interaction between potential consumers and retail destinations is increasingly complex. For example, there is emerging evidence that choice is related to both experiential factors (Wrigley and Lambiri, 2014; Shobeiri et al., 2015) and a provision of a broader range of services, technologies and activities within shopping localities (Hart and Laing, 2014; Digital High Street Advisory Board, 2015). Although some of these factors may be difficult to quantify for a national extent, empirical evidence suggests that presence of anchor stores and various service providers (typically those difficult to digitise) such as leisure, are associated with lower online substitution rates (Weltevreden, 2007). In other words, customers who have relatively easy access to the most attractive stores that are enhanced by adjacent leisure facilities tend to spend more time in town centres, and are normally expected to spend more in them (Hart and Laing, 2014; Wrigley and Lambiri, 2014). Furthermore, it has also been well documented that the impact of online shopping is not uniform across retail types (Zenter, 2008; Ryan and Been, 2013; Parker and Weber, 2013). Typically, retailers who merchandise goods that can be easily digitised such as music, videos, computer games or books are amongst the most vulnerable (Zenter et al., 2013). Use of the Internet for these retail types is estimated at 44% (ONS, 2014), which makes them susceptible to competition from online retailers.

Constructing a measure of e-resilience for a retail centre requires an array of knowledge about the characteristics and mix of retail offerings, alongside demographic and probable Internet engagement traits of likely consumers. An empirical model must therefore ensure that influences of both supply and demand can be estimated, and consideration is required about how such measures interact.

Such issues were explored through this research, however, Figure 6.1 summarises the range of influences on retail e-resilience as including: connectivity, behaviour, demographics and the retail / service offer.

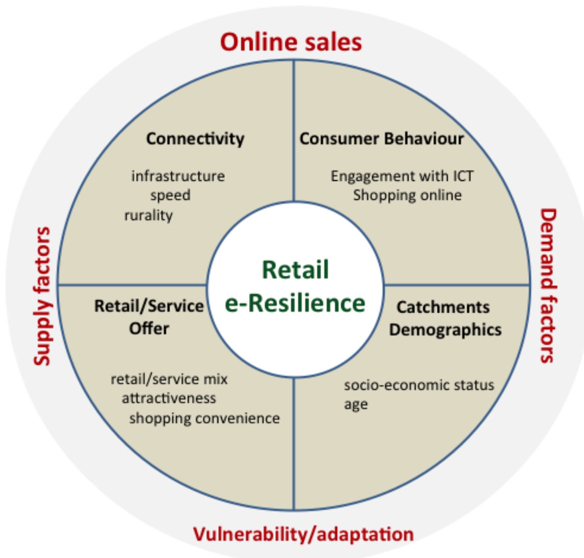


Figure 6.1: A Conceptual Framework of e-resilience

6.3 Demand Factors and Internet Engagement Behaviour

Demand and Internet engagement behaviours have been shown to map onto a range of influencing factors pertaining to the characteristics of people and the places in which they live (Longley and Singleton, 2009a; van Deursen and van Dijk, 2010). Across a range of contexts, influences have been identified as including: demographics, such as age (Rice and Katz, 2003; Warf, 2013), socio-economic status (Silver, 2014), ethnicity (Wilson et al., 2003) or gender (Prieger and Hu, 2008); context, including rurality (usually measured by population density or road connectivity: (Warren, 2007)), and education (Helsper and Enyon, 2010); and finally, Internet connectivity, such as infrastructure availability (Maillé and Tuffin, 2010; Grubesic and Murray, 2002) and speed of connection (Riddlesden and Singleton, 2014). Connectivity relates to the underlying infrastructure that is available within an area to facilitate users getting online. A behavioural component captures the decision whether or not to use the Internet for a given activity, over any number of other modes of access. Influencing such decisions are both demographic effects, mainly age and socio economic status and local retail supply including 'softer factors' such as convenience and accessibility. Such factors have been examined throughout this thesis and have been shown to reveal differentiated levels of both access and engagement across socio-economic indicators and geographic scales (see Chapters 3, 4 and 5). As such there exists an extensive body of current research, from which composite measures of Internet engagement behaviour can be extracted and validated.

Representing the multidimensional and interacting geography of such influences is complex, but has nevertheless previously been illustrated as tractable through geodemographic classification (Longley et al., 2008). The advantage of such methods vis-à-vis univariate measures or scaled composite indicators (e.g. such as a measure of deprivation) are that geodemographics enable the summary of a wider range of influences on Internet user behaviours, and furthermore, enable greater opportunity for differentiation where influencing factors are not necessarily co-linear. For example, differentiating between areas of low engagement that are constrained by infrastructure, versus those constrained by demographics and attitudes.
Geodemographics ascribe categories that aim to summarise the salient characteristics of small areas through comparison of attributes related to resident populations, associated behaviours and features of the built environment (Harris et al., 2005). Such classifications have been applied in a variety of international settings over numerous substantive contexts (Singleton and Spielman, 2013); and is a technique commonly used in retail analysis for consumer segmentation (Birkin et al., 2002). Although general-purpose geodemographic classifications have demonstrated utility for exploration of Internet usage behaviours (Bunyan and Collins, 2013), as illustrated by Longley et al. (2008), there are sound theoretical and empirical justifications for developing purpose-specific classifications within this context. Logic follows that area differentiation is most effectively achieved through a geodemographic where the input attributes are tailored to those outcomes that the classification is being designed to measure, providing enhanced performance and a stronger theoretical rationale for those input attributes selected (Longley and Singleton, 2009b). As such, it is necessary to capture a composite of influences on demand, which are assembled within the e-resilience framework through creation of a purpose specific geodemographic, referred to as the Internet User Classification (IUC). Chapters 4 and 5 of this thesis present a comprehensive overview of how the IUC was created and the subsequent sections of this chapter aim to demonstrate the application of such a classification in a research context.

6.4 Retail Centre Vulnerability and Supply

Measuring the vulnerability of competing retail destinations to consumers of differential Internet engagement characteristics requires an understanding of the location and geographic extent of retail centres, combined with some assessment of their composition and size. A widely-accepted measure of retail area extent in the UK was developed through work funded by the Office of the Deputy Prime Minister in the late 1990 and early 2000s. This technique was later employed by the Department for Communities and Local Government (DCLG) to derive a set of retail centre zones that were used to form a database of information that featured in the State of the Cities Report[1]. These boundaries provide a systematic estimate of where the main concentrations of shops are found between different locations, although do not give information about the composition of competing retail opportunities. However, a nationally expansive record of the location, occupancy and facia of UK retail stores are generated by the Local Data Company[2], a commercial organisation that employs a large survey team to collect these data on a rolling basis. A national extract for February 2014 was made available for this research, with each record comprising the location of a retail premise with latitude and longitude coordinates, and details of the current occupier.

These data were used to calculate a series of measures for physical retail destinations that are detailed in the literature to enhance physical store attractiveness, or, represent retail category vulnerability, where there would be risk of the main product offerings switching from physical to online channels. A composite of these measures forms a 'supply vulnerability index' that is later integrated into the e-resilience score. Input measures to this index included the weighted percentage of anchor stores[3] (Feinberg et al., 2000; Damian et al., 2011) and leisure outlets (Reimers and Clulow, 2009); which are countered by the prevalence of 'digitalisation retail'. The latter measure captured the following categories: newsagents, booksellers, audio-visual rental, computer games, home entertainment, records, tapes and CDs and video libraries, as specified by the Oxford Institute of Retail Management (2013). As such, higher proportions of 'digitalisation retail' are associated with enhanced vulnerability of retail centres, whereas higher proportions of anchor store and leisure units indicated greater resilience. A 'supply vulnerability index' was then generated for each retail centre by creating a composite z score for each variable, and computing an average for each centre. The final score was scaled between 1 and 100.

[1] http://goo.gl/mtX1aB

[2] http://www.localdatacompany.com/

[3] Anchor stores were defined as the 20 most attractive/largest stores as presented by Wrigley and Dolega (2011)

6.5 Reconciling Supply and Demand

Estimating the exposure of retail centres to populations who are active Internet users as defined through the IUC required a method of modelling consumer flows to probable retail destinations. There is a long history and well developed literature on the ways in which such supply and demand for retail centres can be reconciled through catchment area estimation (Woods and Reynolds, 2012; Birkin et al., 2002, 2010). These techniques range in sophistication from calculating the geographic extent that people might be willing to travel to a retail centre in a given time (Grewal et al., 2012), through to more complex mathematical models that are calibrated on the basis of how attractive different retail offerings are to consumers living in different places (Newing et al., 2015). This latter group of models typically makes assumptions that larger towns with more compelling retail and leisure offerings are more attractive, but these effects decay with distance.

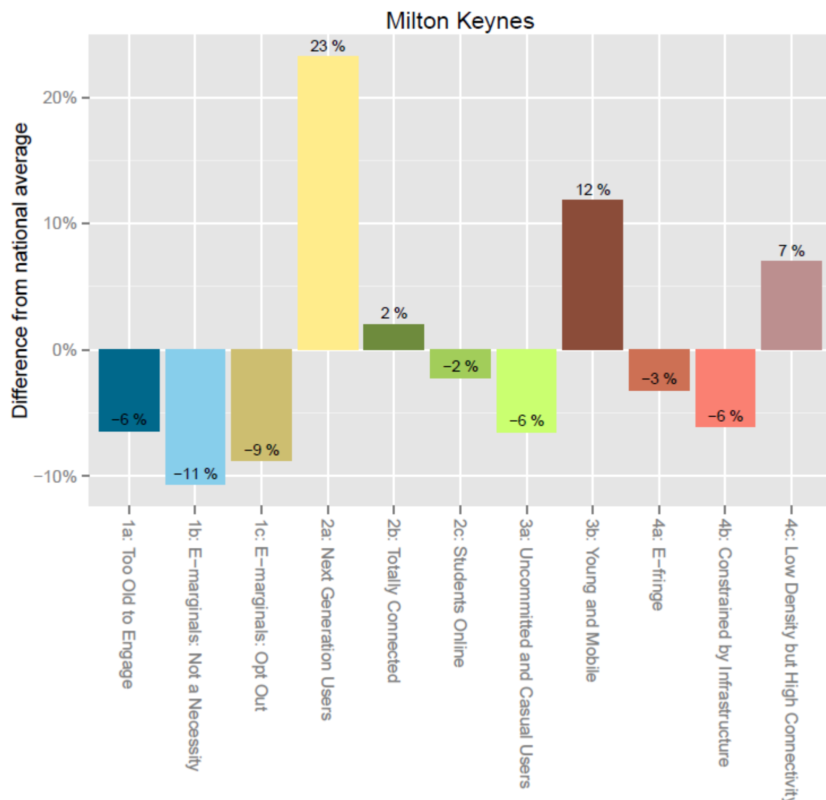


Figure 6.2: IUC Catchment Profile for Central Milton Keynes Retail Centre

Full details of the methodology and software to calibrate bespoke models lies outside the scope of this paper, and can be found in Dolega et al. (2016). However, in brief, catchments were estimated using a product constrained Huff model (Huff, 1964), with inputs including measures of town centre

composition and vacancy rates. Once catchment areas had been estimated, exposure through the intersection of the IUC groups (see Chapter 5, Section 5.5) was examined. An example of a catchment profile for the Milton Keynes retail centre, a city north of London, is shown in Figure 6.2. This considers the proportion of the population within each of the IUC Groups, relative to the England average. The percentage differences are shown for each IUC Group. In this example, it can be seen that the three of the eleven groups are over represented within this retail catchment, and similar profiles were calculated for all retail centres.
As discussed in Chapter 5, the IUC captures a range of influences on Internet user behaviour, however for the purposes of this analysis, those IUC groups with the highest and lowest propensity for online shopping were identified using the OXIS (see Figure 6.3).

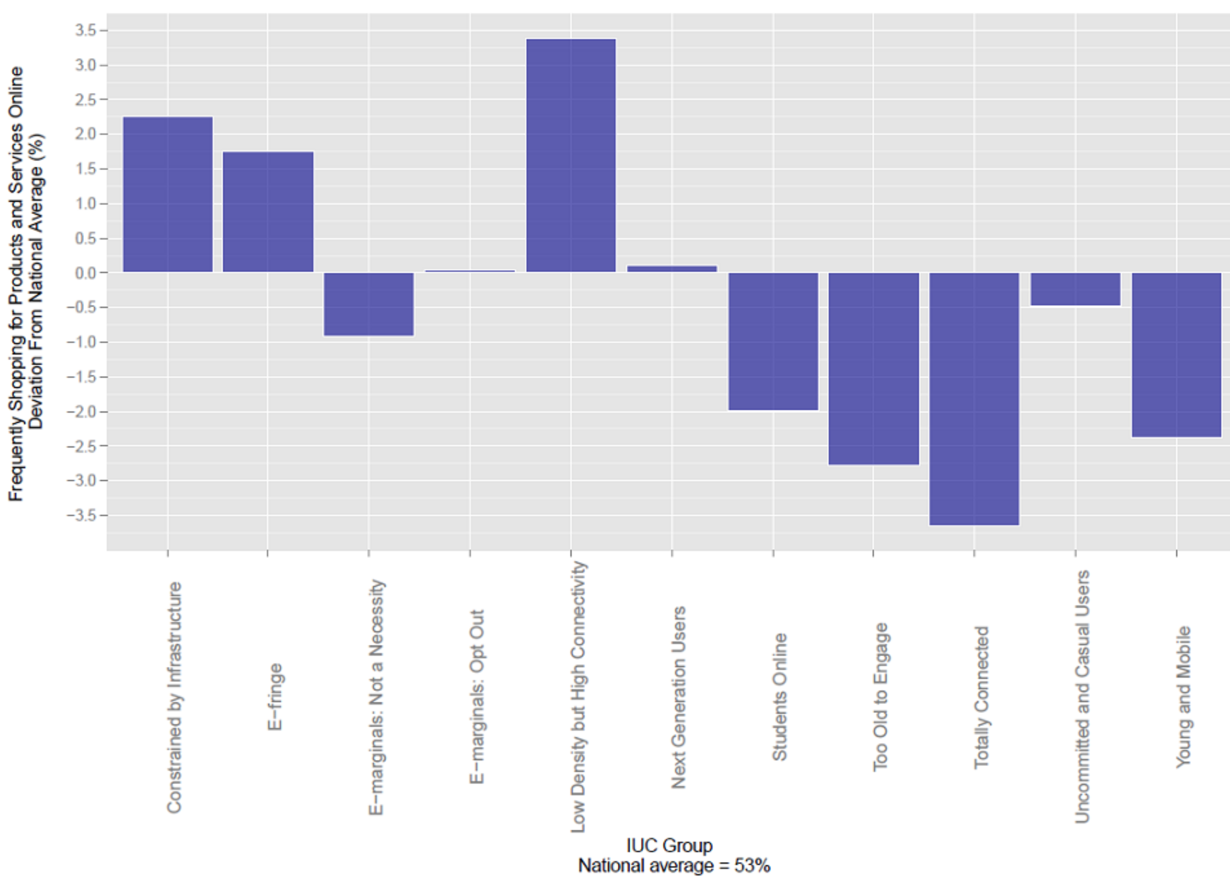


Figure 6.3: Propensity to Shop Online by IUC Groups (OXIS Question - QC30b 'Frequently buy products online')

Nationally, rates of online shopping equate to 53%, however there are differences between IUC Groups. For example, 4c (low density but high connectivity), 4b (constrained by infrastructure), 4a (e-fringe) and 2a (next generation users) are most likely to engage in online shopping; whereas: 3a (uncommitted and casual users), 1b (e-marginals: not a necessity) and 3b (young and mobile) have lower than average propensities. As such, the proportion of people within the overrepresented groups (4a, 4b and 4c) were calculated for each retail centre catchment, and again scaled into the range 1 and 100, forming an Index of high exposure. The Index of high exposure indicates a rather remarkable spatial pattern, and one which has been examined in previous chapters of this thesis, whereby rural and fringe populations are identified as displaying a higher propensity to engage with online retailing (see Chapter 4, Section 4.6)

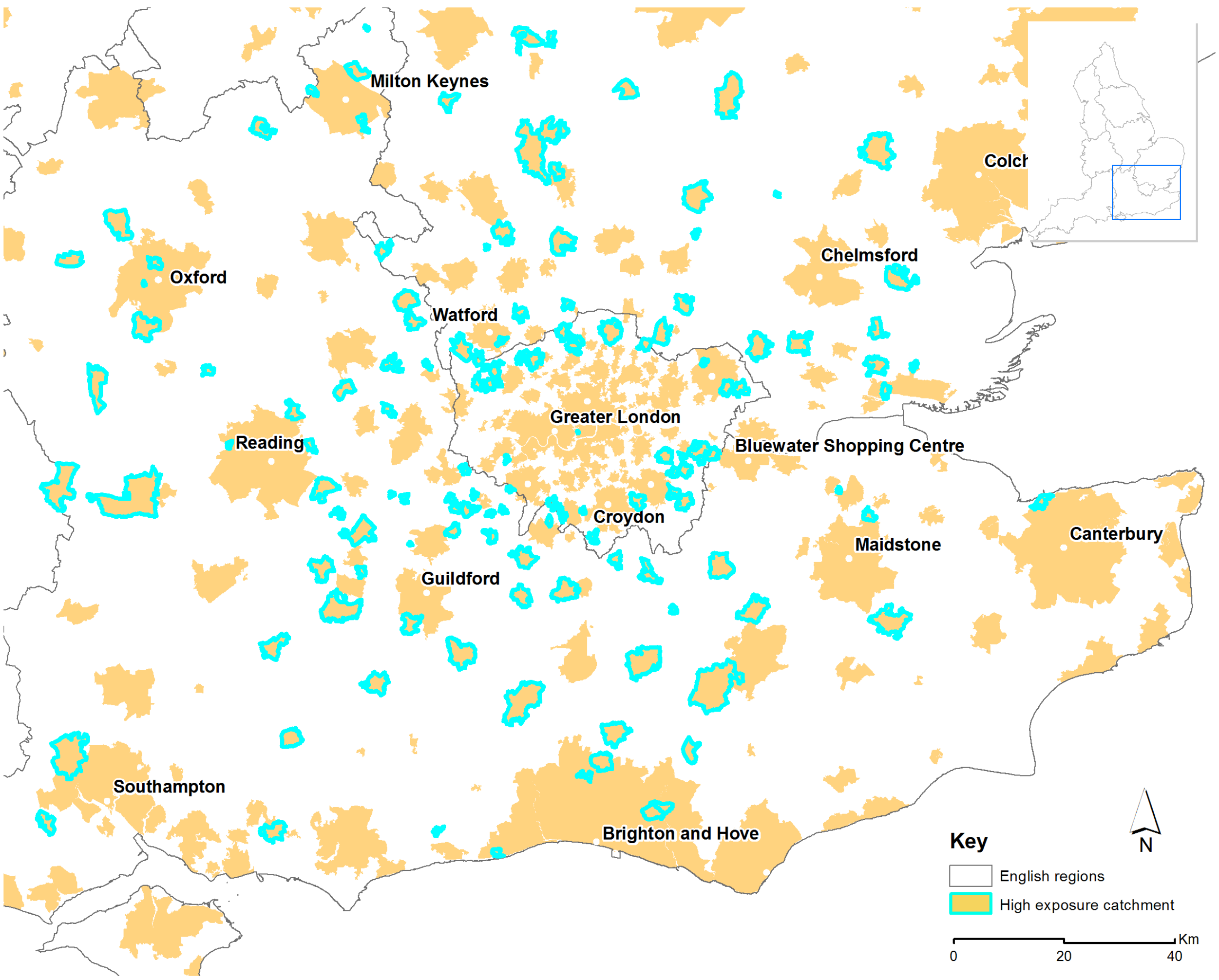


Figure 6.4: Highly Exposed Retail Centres in SE England

Figure 6.4 maps those catchments with high exposure to online retail, defined here as possessing an index over the mean. The pattern emerging from this analysis is that predominantly secondary and tertiary retail centres (Dennis et al., 2002) located in more rural areas, including the satellite centres of more urbanised areas, reveal the greatest exposure to the impacts of online sales. This trend is reiterated for other parts of the country, although the majority of the highly exposed retail centres can be found within the South East. Moreover, based on those attractiveness scores that fed into the catchment model, it is worth noting that none of the highly exposed centres were drawn from the larger, most attractive centres, unlike the fortunes of many of the surrounding smaller towns and local shopping centres. The index of high exposure and the supply vulnerability index were then combined to ascribe a measure of e-resilience to each individual retail centre. The indices were summed, and then the final score scaled into the range 1 to 100. The simple flow diagram in Figure 6.5 shows the input datasets, different stages and outputs used to calculate the e-resilience scores.

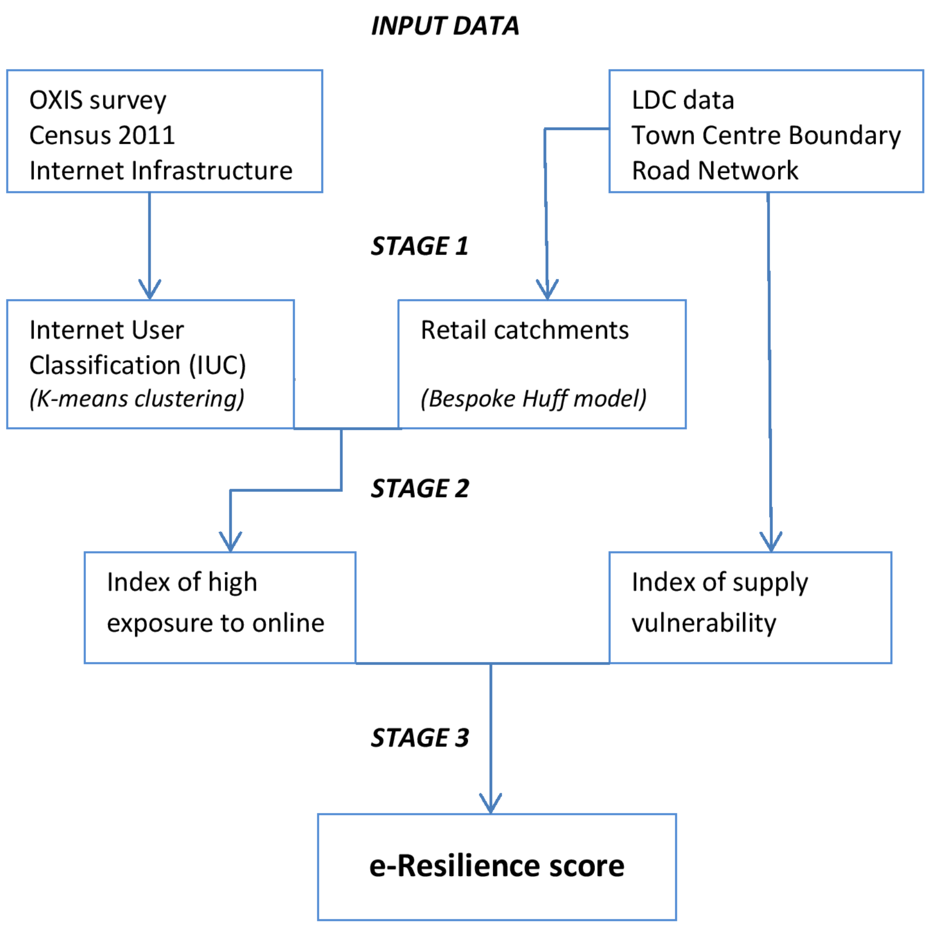


Figure 6.5: Process of Calculating e-resilience Scores

Tables 6.1 and 6.2 show the 20 most and least e-resilient retail centres, with Figure 6.6 mapping the e-resilience levels for the national extent, with scores divided into quartiles. The intersection of these two indicators reveals a remarkable spatial pattern. The most attractive retail centres, in particular the inner areas of the larger urban areas such as Greater London, Birmingham or Manchester demonstrated the highest levels of e-resilience, followed by the small local centres. Conversely, the least e-resilient centres were predominantly located in the suburban and rural areas of South East England, and to a lesser degree around other major conurbations of the country.

Table 6.1: The 20 Most e-resilient Town Centres

Town Centre	Region	e-Resilience Score
Boughton	East Midlands	100.00
Ravenside Retail Park, Bexhill-on-Sea	South East	97.58
Corbridge	North East	93.27
Torport	South West	71.61
Hersham	South East	70.29
Halton, Leeds	Yorkshire and the Humber	69.29
Cinderford	South West	68.51
Marsh Road, Luton	East of England	67.01
South Molton	South West	65.41
Parkgate Retail World	Yorkshire and the Humber	64.37
Carcroft	Yorkshire and the Humber	62.88
Chadderton	North West	60.74
Newburn	North East	60.45
Ventura Road Retail Park, Bitterscote	West Midlands	57.54
Feltham	Greater London	57.16
Teesside Park, Middlesbrough	North East	56.98
Kingston Park	North East	56.92
Sky Blue Way, Coventry	West Midlands	56.82
Crediton	South West	56.36
White Rose Centre	Yorkshire and the Humber	56.28

Typically, these were the secondary and medium sized centres, often referred to as ‘Clone Towns’ (Ryan-Collins et al., 2010). It could be argued that this is largely intertwined with the geography of Internet shopping, where customers in more remote locations, typically faced with poorer retail provision, have displayed a higher propensity for online shopping. Nevertheless, these findings can also be associated with a polarisation effect, implying that large and attractive centres function as hubs for higher order comparison shopping and leisure; whereas the small local centres provide everyday convenience shopping, but the mid-sized centres have a less clear function. Combining such effects with higher exposure to online sales, retailers may be increasingly faced with too much physical space, and therefore may be inclined to downsize their store portfolio in such secondary locations first, as the space offered is often of a wrong size and configuration (BCSC, 2013).

Table 6.2: The 20 Least e-resilient Town Centres

Town Centre	Region	e-Resilience Score
Rochford	East of England	1.00
London Road, Leigh-on-Sea	East of England	15.61
North Seaton Industrial Estate	North East	16.86
Whalley	North West	17.20
Oxted	South East	17.25
Barnt Green	West Midlands	17.39
Eccleshall	West Midlands	17.39
Hurstpierpoint	South East	17.98
Botley Road, Oxford	South East	18.14
Woburn Sands	South East	18.52
Potton	East of England	19.05
Shenfield Station	East of England	19.08
Bradford-on-Avon	South West	19.17
Great Dunmow	East Of England	19.83
Longbenton	North East	20.37
Chalfont St. Peter	South East	20.75
Epworth	Yorkshire and the Humber	20.86
Sawbridgeworth	East of England	21.01
Ampthill	East of England	21.32
Old Bexley	Greater London	21.91

6.6 Discussion and Conclusions

The growth of Internet sales is increasingly viewed as one of the most important forces currently shaping the evolving structure of retail centres (Wrigley and Lambiri, 2014; Hart and Laing, 2014). Although current research does not suggest a death of physical space, the consequences for traditional high streets remain unclear as knowledge about the geography and drivers of Internet shopping are still limited. However, what is evident is that the pace of change in some retail centres has been more rapid than in others, and that multi-channel shopping has generated different requirements, not only in the terms of physical shopping space, but also in the expectations of an increasingly technology-driven consumer (BCSC, 2013; Kacen et al., 2013).

This study introduced the concept of e-resilience as a framework through which the vulnerability of physical retail centres to the impact of online shopping behaviour can be assessed at a spatially disaggregate scale and for a national extent. Importantly, the measurement task required a trade-off between a number of challenges such as the degree of generalisation and the availability of data to inform model specification. For instance, the impact of online sales within a centre may range from damaging to some smaller retailers to complementary in the case of various large multiples. In order to capture these complexities a number of assumptions were made such as the type of retail typically associated with the detrimental impacts and vice-versa. Operationalising this measure of e-resilience required a novel methodology that conflated a range data sources to develop two national indicators of retail centre exposure and vulnerability to online sales. These indices of supply and demand were then coupled through a retail centre catchment model.

The combined e-resilience measure revealed a geography where attractive and large retail centres such as the inner cores of large metropolitan areas were highlighted as more resilient, along with smaller more specialist centres, which perhaps served convenience shopping requirements. The centres identified as most vulnerable included many secondary and medium sized centres, which layers additional risk on top of those issues highlighted elsewhere such as a lack of diversity, or space not appropriately configured to a contemporary retail system (Ryan-Collins et al., 2010).

The findings of this study should be viewed as novel, and can be used to inform policy decisions. The three major implications of this project are as follows. First, it establishes the concept of e-resilience that examines retail centre exposure to the impact of Internet sales, and proposes a new methodology about how such interactions can be measured. Second, a comprehensive classification of retail centres based on their e-resilience levels provides a resource that can be used by a wide range of stakeholders including academics, retailers and town centre managers. For example, such outcomes could be used as assessment tools when evaluating retail centre economic performance. Third, the study adds value to and repositions the focus of current debates on the resilience of traditional high streets, which have predominantly concentrated on supply side measure such as vacancy rates.

The concept of e-resilience contributes considerably to the current understanding of the nature and impact that Internet user behaviour is having on retail centres in the UK. International comparisons are clearly a fertile area for future research - for example some technologically advanced nations, such as South Korea, report lower levels of Internet sales than the U.K. As the penetration of online consumption is still steadily increasing, operational tools such as those offered by this study, in particular the IUC, will have increasing policy relevance.

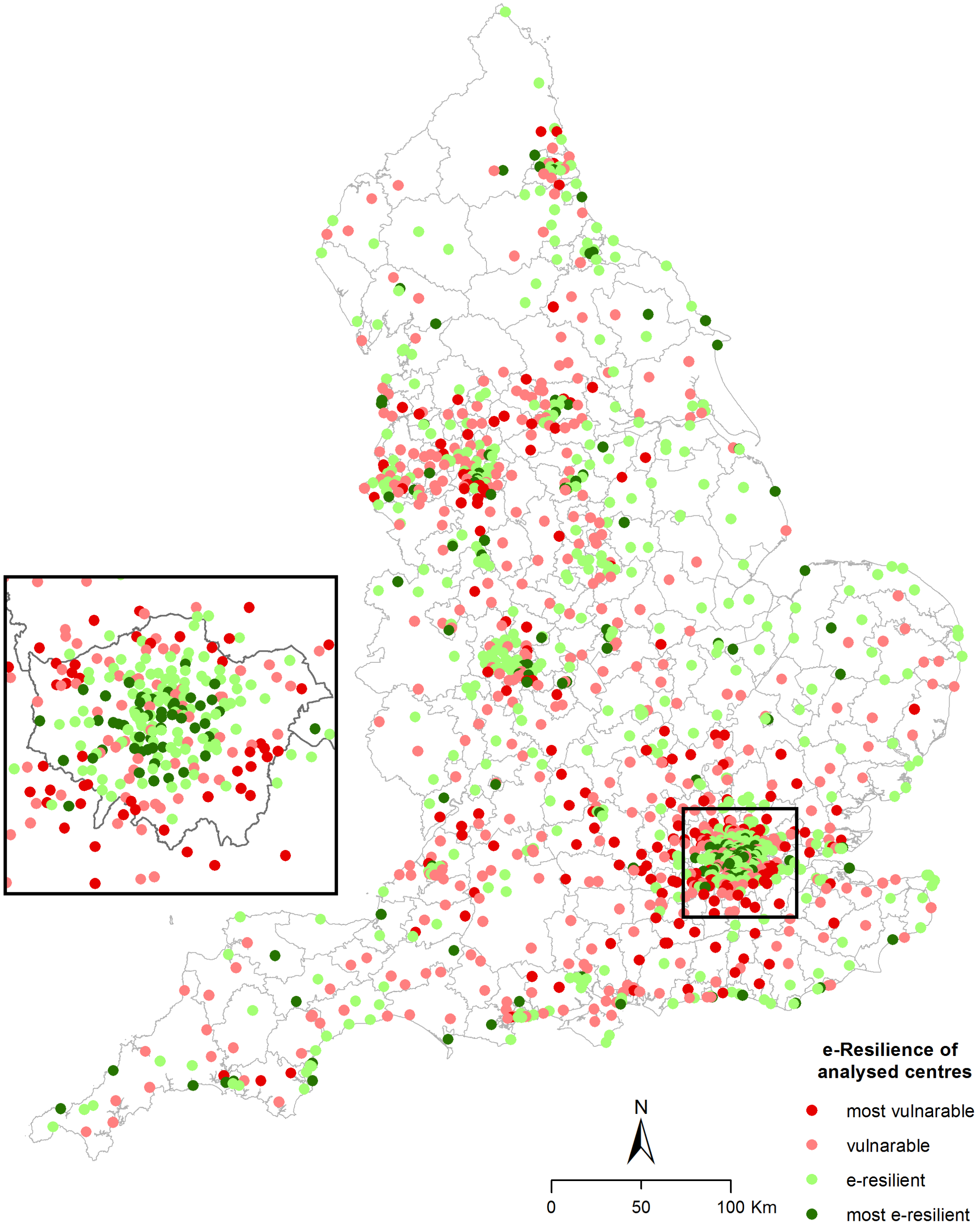


Figure 6.6: The e-resilience of Town Centres in England

Chapter 7

Conclusion

This thesis has aimed to explore the factors that differentiate access to, and engagement with, the Internet, through the use of large quantitative data sources, including novel crowdsourced and open data. While many of the the outputs represent geographically granular analyses that contribute to the wider field of research through the identification of socio-spatially differentiated patterns of Internet access and engagement, this thesis has also contributed a purpose built, domain specific, nested typology of Internet use and engagement. This typology was created at LSOA level for the national extent in the form of the Internet User Classification. This contribution is deemed significant, as it provides a basis for further research across domains where such characteristics of small area populations are of interest. The applicability of the IUC has been demonstrated in Chapter 6 of this thesis. The analysis presented demonstrated how the IUC was applied as a means of profiling resident populations within retail centre catchments to create a composite indicator of retail e-resilience. Without a nested typology of Internet use and engagement, the impact of such a piece of research would likely have been constrained. The following sections revisit each of the quantitative chapters in turn, highlighting the core outputs and their contribution to the field, issues that were faced throughout the work, and those opportunities that may arise for further research.

Chapter 3, the first of the four quantitative studies within this thesis, examined spatial disparities in fixed-line broadband services using crowd sourced speed check data. The use of crowdsourced data should be viewed as novel and it is argued that the paper published from the study represents a significant contribution to the wider field of research, as it is the first instance of crowdsourced data of this type used within the domain. Previously, research into broadband speeds, most of which has been compiled by government funded regulatory bodies such as Ofcom, has relied on relatively small datasets sourced through the deployment of monitoring equipment (see Chapter 3, Section 3.2.1). Whilst the data collection methods vary significantly, aggregate profiling of the crowdsourced data suggested that results are broadly comparable to those reported by Ofcom, thus validating the approach for future research. In addition to the data sources used, the study presented in Chapter 3 has yielded some significant outcomes. In particular, analysis of performance by geographic region has highlighted, at the national scale, exactly which areas are gaining and losing in terms of access to high speed connections. It can be argued that this is a significant contribution to the field given the lack of national performance datasets to support such granular analysis. In this respect, both the aggregate analysis presented as part of the study, as well as the underlying performance datasets, aggregated at varying levels of spatial resolution, represent outputs that can be utilised for further studies.

While there were no immediate issues faced in this study, it is noted that the findings may have limited applicability in the coming years, as the rollout of enhanced infrastructure nationwide will likely result in steep increases in aggregate levels of performance. In addition, this may serve to reduce those performance disparities between rural and urban areas that have been identified, as more geographically disparate locations receive infrastructure upgrades. Whilst this may limit the longevity of some of the findings, it is apparent that future analyses, similar to those presented in this thesis could be implemented to examine these changes, continuing the contribution to the field of research that this thesis has set out to achieve.

Chapter 4 examined Internet demand and engagement, utilising data from the Oxford Internet Survey (OXIS) to create estimates for the national extent across a number of measures at the Output Area level. Again, it is argued that this represents a novel piece of research, as historically, there have been no datasets of this type available for the national extent. It is as a direct result of these data availability constraints that a methodological approach to extrapolate key elements of the OXIS has been presented. Whilst the method of ecological regression is not considered revolutionary for the purpose of small area estimation, the use of decision tree induction to overcome issues arising from small numbers of primary sampling units in the OXIS is argued as novel. Through the use of QUEST modelling it was possible to profile the full OXIS respondent base without any prior aggregation, retaining maximum granularity within the data. Respondent sub-groups could then be created using those limited survey attributes that were derived from the census, and rates

for each question of interest calculated by these groups. The respondent sub-groups could then be recreated at OA level using census data, and rates weighted and fitted to produce area level estimates. The design of such an approach was methodologically complex and required significant exploratory analysis and design. For these reasons it is argued that there is a contribution to the wider field of research. Whilst the methods used are more common, the output represents a collection of univariate measures of engagement with the Internet that were previously unavailable and will likely be made available for future research.
With regards to issues faced, the study presented in Chapter 4 was one of the more problematic. It is apparent that one of the methodological approaches earmarked for extrapolation of OXIS data was deemed redundant due to poor results (see Chapter 4, Section 4.4). However, it is noted that this was predominantly due to the small sample size of the OXIS, and in particular, the small number of geographic areas (both OAs and LSOAs) that are represented in the survey when respondent data are aggregated. Arguably, adopting a second methodological approach alleviated this issue, although the sample size of the OXIS remains small, and as such, it is expected that extrapolation to national coverage may be subject to criticism. A further caveat of the estimation methodology implemented is that QUEST analysis does not produce confidence intervals, and it is noted that confidence intervals for the estimated data would have strengthened subsequent analysis, although in this instance, the method was restrictive. Despite caveats relating to data coverage and method limitations, future studies could revisit such analyses, assessing alternative methods or combining multiple releases of the OXIS in an attempt to yield the most robust estimates possible.

Chapter 5 considered how the complex sets of univariate outputs from previous empirical chapters could be combined using multivariate classification to produce a nested typology of Internet users at the small area level. As a result, a geodemographic classification referred to as the Internet User Classification was created. The main output and perhaps the largest contribution to the wider field of research from this thesis is the IUC, which is open source and publically available. While the applicability of the IUC has been demonstrated through a detailed use case presented in Chapter 6 of this thesis, it is anticipated that there are a wide range of potential uses, including; data profiling, online survey stratification, targeted marketing, location planning, customer insight, and public policy formation and delivery. In terms of issues faced, these were relatively limited, as the classification was predominantly constructed using those outputs of previous analyses within this thesis. It is, however, noted that the classification has coverage of England only, as opposed to the United Kingdom. Whilst this may be viewed as a constraint, it represents a decision made early on in this research. In part this was motivated by coverage of the speed test datasets utilised in Chapter 3, as some areas of Scotland, Wales and Northern Ireland exhibited poorer data coverage

when compared to England. The decision to restrict the geographic extents of the thesis was further strengthened through interrogation of performance measures derived from the speed test data, as these appeared to be sufficiently differentiated to warrant separate studies. Logically, such disparities in performance may influence engagement, and as such, a UK-wide classification would arguably have been less appropriate. Whilst the restriction of the classification to only England is recognised as a caveat, it presents the opportunity for further IUCs to be created for the rest of the UK, thus making further contribution to the field of research.
Currently there are plans to release an updated version of the IUC which will have UK coverage through the Consumer Data Research Centre (CDRC). In part this has been driven by widespread academic and commercial interest in the current classification, which is the first of its kind and has been made publically available to enable the best potential uptake. To date, the current IUC has been used in funding applications for rural broadband investment, where the case for localised infrastructure overhaul has been strengthened by the IUC classes of resident populations. In this sense the classification has help to identify those areas that have resident populations who are engaged with the Internet but suffer from poor performance due to legacy infrastructure, thus enabling stronger return on investment predictions. Further commercial applications of the IUC have seen it used as a profiling tool and predictor (in terms of propensity to interact with digital channels) by a number of UK high street retailers. Such widespread application has opened up discussions with the British Population Survey, with a view to include questions relating to Internet engagement in the survey, such that future iterations of the IUC can utilise this information. The additional data will likely be used to either enrich the current classification or reduce the dependency on closed data sources (such as the OXIS) which have been utilised in the current classification.

Chapter 6 demonstrated the applicability of the IUC through a detailed use case. The use case presented examines the the e-resilience of retail centres in England, assessing the extent to which centres have spatially differentiated vulnerability to the impacts of online consumption. The outcomes of this chapter are twofold, firstly, the study validates the IUC as a scientific contribution, in this instance demonstrating how it can be used to enhance analysis in a related field of research. Secondly, the outputs of the study represent a published research paper that has high policy relevance. It is noted that without the availability of a domain specific geodemographic classification such as the IUC, the outputs of the study would have been more constrained. Instead, the outputs have contributed considerably to the understanding of the nature and impact that Internet user behavior is having on retail centres nationally.

It is anticipated that further public and private sector research will make use of the IUC in both interactive form (see: http://maps.cdrc.ac.uk/#/geodemographics/iuc14/) and through profiling of external data sources using the composite classification. In addition, it is anticipated that the research outcomes from previous chapters within this thesis will be referenced in related studies, given their clear contribution to the field. Understanding how populations engage with the Internet is becoming increasingly valuable to governments, businesses and researchers. It is argued that this thesis represents a significant contribution to the field in consolidating the many factors that interact to form complex geographies of Internet access, performance, use, perceptions and engagement.

Bibliography

Albon, R. (2006). Fixed-to-mobile substitution, complementarity and convergence. *Agenda*, 13(4):309–322.

Alexiou, A. and Singleton, A. D. (2015). Geodemographic Analysis. In *Geocomputation*, pages 137–151. Sage, London.

Allison, C., Miller, A., Oliver, I., Michaelson, R., and Tiropanis, T. (2012). The web in education. *Computer Networks*, 56(18):3811 – 3824.

Angelou, G. N. and Economides, A. A. (2013). Broadband business by utilities infrastructure exploitation: A multistage competition model. *Telecommunications Policy*, 37(1):63 – 79.

Apple Inc. (2012). Apple press info: Apple unveils new iTunes. `http://www.apple.com/pr/library/2012/09/12Apple-Unveils-New-iTunes.html`. Accessed: 2013-07-16.

Apple Inc. (2013). Apple press info: iTunes store sets new record with 25 billion songs sold. `http://www.apple.com/pr/library/2013/02/06iTunes-Store-Sets-New-Record-with-25-Billion-Songs-Sold.html`. Accessed: 2013-07-16.

Asai, Y. and Yano, K. (2001). Construction of geodemographics based on the 1995 Population Census of Japan. In *Papers and Proceedings of the Geographic Information Systems Association*, volume 10, pages 297–284.

Atzori, L., Iera, A., Morabito, G., and Nitti, M. (2012). The social internet of things (SIoT) - when social networks meet the internet of things: Concept, architecture and network characterization. *Computer Networks*, 56(16):3594 – 3608.

Batey, P. and Brown, P. (1995). From human ecology to customer targeting: the evolution of geodemographics. In *GIS for Business and Service Planning*, pages 77–103. GeoInformation International, Cambridge.

Batty, M. (1997). Virtual geography. *Futures*, 29(4-5):337 – 352.

BCSC (2013). Beyond Retail: Redefining the Shape and Purpose of Town Centres. http://policy.bcsc.org.uk/beyondretail/docs/BeyondRetail2013.pdf. Accessed: 2015-05-08.

Beck, N. and Rygl, D. (2015). Categorization of multiple channel retailing in multi-, cross-, and omni-channel retailing for retailers and retailing. *Journal of Retailing and Consumer Services*, 27:170 – 178.

Berners-Lee, T. J. (1992). The world-wide web. *Computer Networks and ISDN Systems*, 25(4-5):454 – 459.

Berthelon, M. and Freund, C. (2008). On the conservation of distance in international trade. *Journal of International Economics*, 75(2):310 – 320.

Bibby, P. and Shepherd, J. (2004). Developing a New Classification of Urban and Rural Areas for Policy Purposes - the Methodology. Technical report, Office for National Statistics.

Biggs, P. and Kelly, T. (2006). Broadband pricing strategies. *info*, 8(6):3–14.

Bimber, B. (2000). Measuring the gender gap on the internet. *Social Science Quarterly*, 81(3).

Birkin, M., Clarke, G., and Clarke, M. (2002). *Retail Geography and Intelligent Network Planning*. Wiley, Chichester NY.

Birkin, M., Clarke, G., and Clarke, M. (2010). Refining and Operationalizing Entropy-Maximizing Models for Business Applications. *Geographical Analysis*, 42(4):422–445.

BIS (2010). Business, innovation and skills committee - fourth report - broadband. http://www.publications.parliament.uk/pa/cm200910/cmselect/cmbis/72/7202.htm. Accessed: 2013-05-21.

Bordea, G., Jothi, N., Rashid, N. A., and Husain, W. (2015). Data mining in healthcare - a review. *Procedia Computer Science*, 72:306 – 313.

Bourdieu, P. and Wacquant, L. (1992). *An Invitation to Reflexive Sociology*. University of Chicago Press, Chicago.

Britain, S. and Liber, O. (1999). A framework for pedagogical evaluation of virtual learning environments. Technical report, JISC, Bristol.

BroadbandGenie (2013). How fast is mobile broadband? a guide to 3G and 4G broadband speed. http://www.broadbandgenie.co.uk/mobilebroadband/help/how-fast-is-mobile-broadband-a-guide-to-3G-broadband-speed. Accessed: 2013-05-08.

Bryant, R. and Bryant, S. (2008). *Investigating Digital Crime*. Wiley.

BT Openreach (2012). Openreach - superfast fibre broadband at home. http://www.superfast-openreach.co.uk/at-home/. Accessed: 2013-02-04.

BT PLC (2011). BT retail launches its fastest-ever broadband. http://www.btplc.com/News/Articles/ShowArticle.cfm?ArticleID=9F9432AE-2A83-4133-A42F-7D9958FD5D96. Accessed: 2013-03-01.

Bunyan, S. and Collins, A. (2013). Digital Exclusion Despite Digital Accessibility: Empirical Evidence from an English City. *Tijdschrift voor economische en sociale geografie*, 104(5):588–603.

Burden, K. and Palmer, C. (2003). Internet crime: Cyber crime - a new breed of criminal? *Computer Law & Security Review*, 19(3):222 – 227.

Burgess, W. (1925). The growth of the city: An introduction to a research project. In *The City*, pages 47–62. University of Chicago Press, Chicago.

Butler, J. C. (1995). An introduction to geoscience education resources on the internet. *Computers & Geosciences*, 21(6):817 – 824.

Cairncross, F. (1997). *The Death of Distance: How the Communications Revolution Will Change Our Lives*. Orion Business Books, London.

Calderwood, E. and Freathy, P. (2014). Consumer mobility in the scottish isles: The impact of internet adoption upon retail travel patterns. *Transportation Research Part A: Policy and Practice*, 59:192 – 203.

Capgemini (2015). Mobile Accounts for 40% of all Online Retail Sales. https://www.uk.capgemini.com/news/uk-news/mobile-accounts-for-40-of-all-online-retail-sales. Accessed: 2015-05-08.

Cardona, M., Schwarz, A., Yurtoglu, B., and Zulehner, C. (2009). Substitution between DSL, cable, and mobile broadband internet services. In *Telecommunication Markets*, Contributions to Economics.

Carlson, J., O'Cass, A., and Ahrholdt, D. (2015). Assessing customers' perceived value of the online channel of multichannel retailers: A two country examination. *Journal of Retailing and Consumer Services*, 27:90 – 102.

Centre for Retail Research (2015). Online Retailing: Britain and Europe 2014. http://www.retailresearch.org/onlineretailing.php. Accessed: 2015-03-21.

Chang, C.-C., Hung, S.-W., Cheng, M.-J., and Wu, C.-Y. (2015). Exploring the intention to continue using social networking sites: The case of facebook. *Technological Forecasting and Social Change*, 95:48 – 56.

Cheong, J. H. and Park, M. C. (2005). Mobile internet acceptance in korea. *Internet Research*, 15(2):125–140.

Cheung, D. W., Kao, B., and Lee, J. (1998). Discovering user access patterns on the world wide web. *Knowledge-Based Systems*, 10(7):463 – 470.

Christopherson, S., Michie, J., and Tyler, P. (2010). Regional Resilience: Theoretical and Empirical Perspectives. *Cambridge Journal of Regions Economy and Society*, 3(1):3–10.

CNET News (2012). Netflix gobbles a third of peak internet traffic in north america. http://news.cnet.com/8301-1023_3-57546405-93/netflix-gobbles-a-third-of-peak-internet-traffic-in-north-america/. Accessed: 2013-05-20.

Cooper, J. (2006). The digital divide: the special case of gender. *Journal of Computer Assisted Learning*, 22(5):320–334.

Coursera (2013). Coursera. https://www.coursera.org/about. Accessed: 2013-05-13.

Damian, D., Curto, J., and Pinto, J. (2011). The Impact of Anchor Stores on the Performance of Shopping Centres: the case of Sonae Sierra. *International Journal of Retail & Distribution Management*, 39(6):456–475.

Dawley, S., Pike, A., and Tomaney, J. (2010). Towards the Resilient Region? *Local Economy*, 25:650–667.

DCLC (2013). The Future of High Streets. Progress since the Portas Review. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/211536/Future_of_High_Street_-_Progress_Since_the_Portas_Review_-revised.pdf. Accessed: 2015-10-28.

DCMS (2009). Digital britain final report. http://webarchive.nationalarchives.gov.uk/20100511084737/http:/interactive.bis.gov.uk/digitalbritain/final-report/. Accessed: 2013-05-20.

DCMS (2010). Britain's superfast broadband future. https://www.gov.uk/government/publications/britains-superfast-broadband-future. Accessed: 2013-05-30.

DCMS (2011). Broadband delivery programme: Delivery model – September 2011. https://www.gov.uk/government/publications/broadband-delivery-programme-delivery-model-september-2011. Accessed: 2013-05-30.

Demirkan, H. and Spohrer, J. (2014). Developing a framework to improve virtual shopping in digital malls with intelligent self-service systems. *Journal of Retailing and Consumer Services*, 21(5):860–868.

Dennis, C., Marsland, D., and Cockett, T. (2002). Central Place Practice: Shopping Centre Attractiveness Measures Hinterland Boundaries and the UK Retail Hierarchy. *Journal of Retailing and Consumer Services*, 9(4):185–199.

Detica (2011). The cost of cyber crime. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60942/THE-COST-OF-CYBER-CRIME-SUMMARY-FINAL.pdf. Accessed: 2013-07-18.

Dholakia, U. M., Kahn, B. E., Reeves, R., Rindfleisch, A., Stewart, D., and Taylor, E. (2010). Consumer behavior in a multichannel, multimedia retailing environment. *Journal of Interactive Marketing*, 24(2):86 – 95. Special Issue on Emerging Perspectives on Marketing in a Multichannel and Multimedia Retailing Enviroment.

Digital High Street Advisory Board (2015). Digital High Street 2020. http://thegreatbritishhighstreet.co.uk/digital-high-street-report-2020. Accessed: 2015-06-15.

Dikaiakos, M. D. (2004). Intermediary infrastructures for the world wide web. *Computer Networks*, 45(4):421 – 447.

Dinev, T., Hart, P., and Mullen, M. R. (2008). Internet privacy concerns and beliefs about government surveillance - an empirical investigation. *The Journal of Strategic Information Systems*, 17(3):214 – 233.

Doherty, N. and Ellis-Chadwick, F. (2010). Internet Retailing: the Past the Present and the Future. *International Journal of Retail & Distribution Management*, 38(11/12):943–965.

Dolega, L. and Celinska-Janowicz, D. (2015). Retail resilience: A theoretical framework for understanding town centre dynamics. *Studia Regionalne i Lokalne - Polish Section of the Regional Studies Association*, 2(60):8–31.

Dolega, L., Pavlis, M., and Singleton, A. (2016). Estimating attractiveness, hierarchy and catchment area extents for a national set of retail centre agglomerations. *Journal of Retailing and Consumer Services*, 28:78 – 90.

Dumais, S. A. and Ward, A. (2010). Cultural capital and first-generation college success. *Poetics*, 38(3):245 – 265.

Dutton, W. H. and Blank, G. (2011). Next generation users: The internet in britain. Technical report, Oxford Internet Institute.

Dutton, W. H. and Blank, G. (2013). Cultures of the Internet: the Internet in Britain. Technical report, Oxford Internet Institute.

Ellison, N. B., Steinfield, C., and Lampe, C. (2007). The benefits of facebook "Friends:" social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168.

Ergen, M. (2009). *Mobile Broadband: Including WiMAX and LTE*. Springer.

European Parliament (2000). Regulation (EC) no 2887/2000. `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2000:336:0004:0004:EN:PDF`. Accessed: 2013-12-03.

Eurostat (2012). Eurostat database. `http://epp.eurostat.ec.europa.eu/portal/page/portal/information_society/data/database`. Accessed: 2013-05-09.

Federal Communications Commission (2011). Connecting america: National broadband plan. `http://www.broadband.gov/download-plan/`. Accessed: 2013-05-13.

Feinberg, R., Stanton, J., Keen, C., Kim, I., Hokama, L., and de Ruyter, K. (2000). Attraction as a Determinant of Mall Choice. In *Proceedings of the 7th International Conference on Retailing and Services Science*, Eindhoven. European Institute of Retailing and Services Science.

Fikkers, D. J. (2013). How to make a €9.2bn broadband investment completely useless. *Regional Insights*, 4(1):15–17.

Fink, C., Mattoo, A., and Neagu, I. C. (2005). Assessing the impact of communication costs on international trade. *Journal of International Economics*, 67(2):428 – 445.

Flamm, K. and Chaudhuri, A. (2007). An analysis of the determinants of broadband access. *Telecommunications Policy*, 31(6-7):321–326.

Fletcher, N. (2007). Challenges for regulating financial fraud in cyberspace. *Journal of Financial Crime*, 14(2):190–207.

Fogel, J. and Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior*, 25(1):153 – 160.

Forman, C., Ghose, A., and Goldfarb, A. (2008). Competition Between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live. *Management Science*, 55(1):47–57.

Freund, C. L. and Weinhold, D. (2004). The effect of the internet on international trade. *Journal of International Economics*, 62(1):171 – 189.

Gale, C. G., Singleton, A. D., Bates, A. G., and Longley, P. A. (2016). Creating the 2011 area classification for output areas (2011 oac). *Journal of Spatial Information Science*. In Press.

Giesler, M. (2006). Consumer gift systems. *Journal of Consumer Research*, 33(2):283–290.

Glowniak, J. (1998). History, structure, and function of the internet. *Seminars in Nuclear Medicine*, 28(2):135 – 144.

Goodchild, M. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69:211–221.

gov.uk (2013). Welcome to GOV.UK. https://www.gov.uk/. Accessed: 2013-05-30.

Graham, M. and Dutton, W. H. (2014). *Society and the Internet: How Networks of Information and Communication are Changing Our Lives*. Oxford University Press, Oxford.

Graham, M., Hale, S., and Stevens, M. (2012). Featured graphic: Digital divide: the geography of internet access. *Environment and Planning A*, 44(5):1009–1010.

Gray, N. J., Klein, J. D., Noyce, P. R., Sesselberg, T. S., and Cantrill, J. A. (2005). Health information-seeking behaviour in adolescence: the place of the internet. *Social Science & Medicine*, 60(7):1467 – 1478.

Greenfield, A. (2006). *Everyware: The dawning age of ubiquitous computing*. New Riders, Boston.

Grewal, D., Kopalle, P., Marmorstein, H., and Roggeveen, A. (2012). Does Travel Time to Stores Matter? The Role of Merchandise Availability. *Journal of Retailing*, 88(3):437–444.

Grow, B. (2005). Hacker hunters: an elite force takes on the dark side of computing. *Business Week*, (30):74–82.

Grubesic, T. and Murray, A. (2002). Constructing the Divide: Spatial Disparities in Broadband Access. *Papers in Regional Science*, 81(2):197–221.

Guardian (2013). Edward snowden identifies himself as source of NSA leaks - as it happened. `http://www.guardian.co.uk/world/2013/jun/09/nsa-secret-surveillance-lawmakers-live?INTCMP=SRCH`. Accessed: 2013-07-18.

Gubbi, J., Buyya, R., Marusic, S., and Palaniswami, M. (2013). Internet of things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7):1645 – 1660.

Haklay, M., Singleton, A. D., and Parker, C. (2008). Web Mapping 2.0: The Neogeography of the GeoWeb. *Geography Compass*, 2(6):2011–2039.

Han, J., Kamber, M., and Pei, J. (2012). *Data Mining Concepts and Techniques*. Morgan Kaufmann, San Francisco, CA, third edition.

Hargittai, E. (2002). Second-level digital divide: differences in people's online skills. *First Monday*, 7(4).

Harris (2004). National survey on consumer privacy attitudes. http://epic.org/privacy/survey/. Accessed: 2013-07-19.

Harris, R., Sleight, P., and Webber, R. (2005). *Geodemographics, GIS and Neighbourhood Targeting*. Wiley, West Sussex, England.

Harris and Associates and Westin, A. (1998). Ecommerce and privacy: what net users want. privacy and american business. Technical report, Pricewaterhouse Coopers. LLP.

Hart, C. and Laing, A. (2014). The Consumer Journey through the High Street in the Digital Area. In *Evolving High Streets: Resilience and Reinvention - Perspectives from Social Science*, pages 36–39. University of Southampton, Southampton.

Hassink, R. (2010). Regional Resilience: A Promising Concept to Explain Differences in Regional Economic Adaptability? *Cambridge Journal of Regions Economy and Society*, 3(1):45–58.

Helsper, E. and Enyon, R. (2010). Digital Natives: Where is the Evidence? *British Educational Research Journal*, 36(3):503–520.

Hernández-Murillo, R., Llobet, G., and Fuentes, R. (2010). Strategic online banking adoption. *Journal of Banking & Finance*, 34(7):1650–1663.

HM Treasury (2012). National infrastructure plan: Update 2012. http://www.hm-treasury.gov.uk/d/national_infrastructure_plan_051212.pdf. Accessed: 2013-04-03.

Howard, N. (1969). Least squares classification and principal components analysis: a comparison. In *Quantitative Ecological Analysis in the Social Sciences*, pages 397–412. MIT Press, Cambridge, MA.

Hoy, M. B. (2011). HTML5: a new standard for the web. *Medical Reference Services Quarterly*, 30(1):50–55.

Huff, D. (1964). Defining and Estimating a Trade Area. *Journal of Marketing*, 28(3):34–38.

Hunton, P. (2009). The growing phenomenon of crime and the internet: A cybercrime execution and analysis model. *Computer Law & Security Review*, 25(6):528 – 535.

IEGMP (2000). Mobile Phones and Health. Technical report, Independent Expert Group on Mobile Phones.

Informa (2013). VoIP and IP messaging: Operator strategies to combat the threat from OTT players. https://commerce.informatm.com/reports/main/voip-ip-messaging-revised.html. Accessed: 2013-07-02.

Internet World Stats (2016). World internet users statistics usage and world population stats. http://www.internetworldstats.com/stats.htm. Accessed: 2016-08-17.

ITU (2002). G.993.2: Very high speed digital subscriber line transceivers 2 (VDSL2). http://www.itu.int/rec/T-REC-G.993.2-200602-S/en. Accessed: 2013-03-01.

Jameel, A., Ehsan, N., Mirza, E., and Waseem Bhatti, M. (2012). Factors affecting broadband proliferation in pakistan and initiatives/measures taken by the government of pakistan. *Procedia Technology*, 1:250–257.

James, J. (2007). From origins to implications: key aspects in the debate over the digital divide. *Journal of Information Technology*, 22(3):284–295.

Jupiter Media Metrix (2001). Global napster usage plummets, but new file-sharing alternatives gaining ground. http://web.archive.org/web/20080413104420/http://www.comscore.com/press/release.asp?id=249. Accessed: 2013-07-15.

Kacen, J., Hess, J., and Chiang, W. (2013). Bricks or Clicks? Consumer Attitudes toward Traditional Stores and Online Stores. *Global Economics and Management Review*, 18(1):12–21.

Katz, J. E., Rice, R. E., and Aspden, P. (2001). The internet, 1995-2000. access, civic involvement, and social interaction. *American Behavioral Scientist*, 45(3):405–419.

Kaufmann, A., Lehner, P., and Tödtling, F. (2003). Effects of the internet on the spatial structure of innovation networks. *Information Economics and Policy*, 15(3):402 – 424.

Kenney, M. and Pon, B. (2011). Structuring the smartphone industry: Is the mobile internet OS platform the key? *Journal of Industry Competition and Trade*, 11(3):239–261.

Kim, J. and Park, J. (2005). A consumer shopping channel extension model: attitude shift toward the online store. *Journal of Fashion Marketing and Management: An International Journal*, 9(1):106–121.

Kitchin, R. and Dodge, M. (2011). *Code/Space: Software and Everyday Life*. MIT Press, Cambridge, MA.

Koutroumpis, P. (2009). The economic impact of broadband on growth: A simultaneous approach. *Telecommunications Policy*, 33(9):471 – 485.

Kraut, R. and Kiesler, S. (2003). The social impact of internet use. *Psychological Science Agenda*, 13(5):8–10.

Kühn, P. J. (2004). Location-based services in mobile communication infrastructures. *International Journal of Electronics and Communications*, 58(3):159 – 164.

Kyriakidou, V., Michalakelis, C., and Sphicopoulos, T. (2011). Digital divide gap convergence in europe. *Technology in Society*, 33(3-4):265–270.

Lai, B. and Brewer, G. A. (2006). New york city's broadband problem and the role of municipal government in promoting a private-sector solution. *Technology in Society*, 28(1-2):245–259.

Lane-Fox, M. (2009). Champion for digital inclusion: The economic case for digital inclusion. Technical report, Pricewaterhousecoopers LLP.

LaRose, R., Gregg, J. L., Strover, S., Straubhaar, J., and Carpenter, S. (2007). Closing the rural broadband gap: Promoting adoption of the internet in rural america. *Telecommunications Policy*, 31(6-7):359–373.

LeBaron, J. F., Gibson, J. T., Burke, D. M., and Scollin, P. A. (1998). How educators find education resources on the internet: A discussion of independent search behaviors by graduate education students. *The Internet and Higher Education*, 1(3):191 – 201.

Lenhart, A., Madden, M., and Hitlin, P. (2005). Teens and technology. youth are leading the transition to a fully wired and mobile nation. Technical report, Pew Internet.

Li, K. and Du, T. C. (2012). Building a targeted mobile advertising system for location-based services. *Decision Support Systems*, 54(1):1 – 8.

Liaw, S. S. (2008). Investigating students' perceived satisfaction, behavioral intention, and effectiveness of e-learning: A case study of the blackboard system. *Computers & Education*, 51(2):864 – 873.

Lim, T.-S., Loh, W.-Y., and Shih, Y.-S. (2000). A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning*, 40:203–229.

Lin, N. (2001). *Social Capital: A Theory of Social Structure and Action*. Cambridge University Press, New York.

Lindsay, P. H. and Norman, D. A. (1977). *Human Information Processing: An Introduction to Psychology*. Academic Press.

Liverpool City Council (1969). Social Malaise in Liverpool: Interim Report on Social Problems and their Distribution. Technical report, Liverpool.

Loges, W. E. and Jung, J.-Y. (2001). Exploring the digital divide, internet connectedness and age. *Communication Research*, 28(4):536–562.

Loh, W.-Y. and Shih, Y.-S. (1997). Split selection methods for classification trees. *Statistica Sinica*, 7:815–840.

Longley, P. A. (2003). Towards better understanding of digital differentiation. *Computers, Environment and Urban Systems*, 27(2):103 – 106.

Longley, P. A. and Singleton, A. D. (2009a). Classification Through Consultation: Public Views Of The Geography Of The E-Society. *International Journal of Geographical Information Science*, 23(6):737–763.

Longley, P. A. and Singleton, A. D. (2009b). Linking social deprivation and digital exclusion in england. *Urban Studies*, 46(7):1275–1298.

Longley, P. A., Webber, R., and Li, C. (2008). The UK geography of the e-society: a national classification. *Environment and Planning A*, 40(2):362–382.

Madden, M. and Jones, S. (2008). Networked workers. `http://www.pewinternet.org/~/media/Files/Reports/2008/PIP_Networked_Workers_FINAL.pdf.pdf`. Accessed: 2012-12-18.

Maillé, P. and Tuffin, B. (2010). Price War in Heterogeneous Wireless Networks. *Computer Networks*, 54(13):2281–2292.

Marshall, A. M. and Tompsett, B. C. (2002). Spam 'n' chips - a discussion of internet crime. *Science & Justice*, 42(2):117 – 122.

Martin, R. (2011). Regional Economic Resilience Hysteresis and Recessionary Shocks. *Journal of Economic Geography*, 12(1):1–32.

Martin, R. and Sunley, P. (2007). Complexity Thinking and Evolutionary Economic Geography. *Journal of Economic Geography*, 7(5):573–602.

McMullan, M. (2006). Patients using the internet to obtain health information: How this affects the patient-health professional relationship. *Patient Education and Counseling*, 63(1-2):24 – 28.

Modarres, A. (2011). Beyond the digital divide. *National Civic Review*, 100(3):4–7.

Mowery, D. C. and Simcoe, T. (2002). Is the internet a US invention? an economic and technological history of computer networking. *Research Policy*, 31(8-9):1369 – 1387.

Müller, R., Kijl, B., and Martens, J. (2011). A comparison of inter-organizational business models of mobile app stores: There is more than open vs. closed. *Journal of Theoretical and Applied Electronic Commerce Research*, 6(2):63–76.

Nathan, M. and Rosso, A. (2015). Mapping digital businesses with big data: Some early findings from the UK. *Research Policy*, 44(9):1714 – 1733. The New Data Frontier.

Neslin, S., Grewal, D., Leghorn, R., Shankar, V., Teerling, M., Thomas, J., and Verhoef, P. (2006). Challenges and opportunities in multichannel customer management. *Journal of Service Research*, 9(2):95–112.

Newing, A., Clarke, G., and Clarke, M. (2015). Developing and Applying a Disaggregated Retail Location Model with Extended Retail Demand Estimations. *Geographical Analysis*, 47(3):219–239.

Nie, N. H. (2001). Sociability, interpersonal relations, and the internet : Reconciling conflicting findings. *American Behavioral Scientist*, 45(3):420–435.

Norris, P. (2001). *Digital Divide*. Cambridge University Press, Cambridge.

NTIA (2002). A nation online. Technical report, National Telecommunications and Information Administration.

NTRS (2009). National technology readiness survey 2009 - telecommuting report. http://www.technoreadymarketing.com/NTRS_2009_Telecommuting_Cost.php. Accessed: 2013-05-13.

OECD (2006). Digital broadband content: Public sector information. http://www.oecd-ilibrary.org/science-and-technology/digital-broadband-content_231534841283. Accessed: 2013-05-30.

OECD (2009). Mobile broadband: Pricing and services. http://www.oecd.org/internet/broadbandandtelecom/43280727.pdf. Accessed: 2013-01-08.

Ofcom (2009). UK broadband speeds 2009: Research report. http://stakeholders.ofcom.org.uk/binaries/research/telecoms-research/broadbandspeeds.pdf. Accessed: 2013-07-18.

Ofcom (2010). UK broadband speeds, May 2010: The performance of fixed-line broadband delivered to UK residential consumers. Technical report.

Ofcom (2011). UK fixed-line broadband performance, may 2011 - the performance of fixed-line broadband delivered to UK residential consumers. Technical report.

Ofcom (2012). Infrastructure report - 2012 update. http://stakeholders.ofcom.org.uk/market-data-research/other/telecoms-research/broadband-speeds/infrastructure-report-2012/. Accessed: 2013-05-08.

Ofcom (2013a). The availability of communications services in the UK. http://stakeholders.ofcom.org.uk/market-data-research/market-data/economic-geography/?utm_source=updates&utm_medium=www&utm_campaign=thinkbroadband. Accessed: 2013-05-21.

Ofcom (2013b). The european broadband scorecard. http://stakeholders.ofcom.org.uk/market-data-research/other/telecoms-research/bbresearch/scorecard. Accessed: 2013-05-30.

Ofcom (2014). Infrastructure Report - 2014. http://stakeholders.ofcom.org.uk/market-data-research/market-data/infrastructure/infrastructure-2014/. Accessed: 2015-09-23.

Ofcom (2015). Mobile Phone Base Stations - How Mobile Phone Networks Work. http://www.ofcom.org.uk/static/archive/ra/topics/mpsafety/school-audit/mobilework.htm. Accessed: 2015-10-27.

Oftel (2002). Oftel - access network facilities: Oftel guidelines on condition 83 of BT's licence - september 2000. `http://www.ofcom.org.uk/static/archive/oftel/publications/broadband/llu/llug0900.htm`. Accessed: 2013-03-06.

Ojo, A., Vickers, D., and Ballas, D. (2013). Creating a small scale area classification for understanding the economic social and housing characteristics of small geographical areas in the Philippines. *Regional Science Policy & Practice*, 5:1–24.

O'Neil, C. and Schutt, R. (2013). *Doing Data Science: Straight Talk From the Front Line*. O'Reilly Media, Sebastopol, CA.

ONS (2014). Internet Access - Households and Individuals 2014. `http://www.ons.gov.uk/ons/rel/rdit2/internet-access---households-and-individuals/2014/stb-ia-2014.html`. Accessed: 2015-05-08.

ONS (2014). The national statistics socio-economic classification (NS-SEC rebased on the SOC2010). `http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/soc2010/soc2010-volume-3-ns-sec--rebased-on-soc2010--user-manual/index.html#top`. Accessed: 2014-05-01.

Osbourne, J. (2002). Notes on the Use of Data Transformation. *Practical Assessment, Research & Evaluation*, 8(6).

Ostrower, F. (1998). The arts as cultural capital among elites: Bourdieu's theory reconsidered. *Poetics*, 26(1):43 – 53.

Oxford Institute of Retail Management (2013). Diversity and the Uk's High Streets. `http://oxford-institute.sbsblogs.co.uk/2013/07/19/diversity-and-the-uks-high-streets`. Accessed: 2015-05-08.

Paine, C., Reips, U. D., Stieger, S., Joinson, A., and Buchanan, T. (2007). Internet users' perceptions of 'privacy concerns' and 'privacy actions'. *International Journal of Human-Computer Studies*, 65(6):526 – 536.

Pantano, E. and Viassone, M. (2015). Engaging consumers on new integrated multichannel retail settings: Challenges for retailers. *Journal of Retailing and Consumer Services*, 25:106 – 114.

Papoulis, A. and Pillai, S. (2002). *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, New York, fourth edition.

Parker, B. and Weber, R. (2013). Second-Hand Spaces: Restructuring Retail Geographies in an Era of E-Commerce. *Urban Geography*, 34(8):1096–1118.

Parker, E. B. (2000). Closing the digital divide in rural america. *Telecommunications Policy*, 24(4):281–290.

Pendall, R., Foster, K., and Cowell, M. (2010). Resilience and Regions: Building Understanding of the Metaphor. *Cambridge Journal of Regions Economy and Society*, 3(1):71–84.

Peter, J. and Valkenburg, P. M. (2006). Adolescents' internet use: Testing the "disappearing digital divide" versus the "emerging digital differentiation" approach. *Poetics*, 34(4-5):293–305.

Pew Internet (2001). The internet and education: Pew internet & american life project. `http://www.pewinternet.org/Reports/2001/The-Internet-and-Education/Summary-of-Findings.aspx`. Accessed: 2013-05-09.

Picot, A. and Wernick, C. (2007). The role of government in broadband access. *Telecommunications Policy*, 31(10):660–674.

Pigg, K. E. and Crank, L. D. (2005). Do information communication technologies promote rural economic development? *Community Development*, 36(1):65–76.

Pitt, L., Berthon, P., and Berthon, J.-P. (1999). Changing channels: The impact of the internet on distribution strategy. *Business Horizons*, 42(2):19–28.

Portas, M. (2011). The Portas Review. An independent review into the future of our high streets. `https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6292/2081646.pdf`. Accessed: 2015-06-15.

Portes, A. (2014). Downsides of social capital. *Proceedings of the National Academy of Sciences*, 111(52):18407–18408.

Prieger, J. E. (2013). The broadband digital divide and the economic benefits of mobile broadband for rural areas. *Telecommunications Policy*, 37(6-7):483–502.

Prieger, J. E. and Hu, W. M. (2008). The broadband digital divide and the nexus of race, competition, and quality. *Information Economics and Policy*, 20(2):150–167.

Putnam, R. D. (1995). Bowling alone: America's declining social capital. *Journal of Democracy*, 6(1):65–78.

Quah, D. (2000). Internet cluster emergence. *European Economic Review*, 44(4-6):1032 – 1044.

Raco, M. and Street, E. (2012). Resilience Planning Economic Change and the Politics of Post-recession Development in London and Hong Kong. *Urban Studies*, 49(5):1065–1087.

Rao, J. (2003). *Small Area Estimation*. Wiley series in survey methodology. John Wiley, Hoboken, NJ.

Reimers, V. and Clulow, V. (2009). Retail Centres: It's Time to Make them Convenient. *International Journal of Retail & Distribution Management*, 37(7):541–562.

Rice, R. E. (2006). Influences, usage, and outcomes of internet health information searching: Multivariate results from the pew surveys. *International Journal of Medical Informatics*, 75(1):8 – 28.

Rice, R. E. and Katz, J. E. (2003). Comparing internet and mobile phone usage: digital divides of usage, adoption, and dropouts. *Telecommunications Policy*, 27(8):597–623.

Riddlesden, D. and Singleton, A. D. (2014). Broadband speed equity: A new digital divide? *Applied Geography*, 52(0):25 – 33.

Rokach, L. and Maimon, O. (2010). Classification trees. In *Data Mining and Knowledge Discovery Handbook*, pages 149–174. Springer, New York, second edition.

Ryan, S. and Been, C. (2013). Bricks-and-Mortar: Bookselling and Supermarket Philosophy. *International Journal Of The Book*, 10(2):55–66.

Ryan-Collins, J., Cox, E., Potts, R., and Squires, P. (2010). *Re-imagining the High Street: Escape from Clone Town Britain*. New Economics Foundation, London.

Schoenberger, C. R. (2002). The internet of things. `www.forbes.com/technology/forbes/2002/0318/155.html`. Accessed: 2013-07-15.

Selwyn, N. (2004). Reconsidering political and popular understandings of the digital divide. *New Media & Society*, 6(3):341–362.

Selwyn, N. (2006). Digital division or digital decision? a study of non-users and low-users of computers. *Poetics*, 34(4-5):273–292.

Shah, D. V., Kwak, N., and Holbert, R. L. (2001). "Connecting" and "Disconnecting" with civic life: Patterns of internet use and the production of social capital. *Political Communication*, 18(2):141–162.

Shakshuki, E., Yasar, A., Park, H. J., Kim, P. H., Marsico, M., and Rasheed, N. (2014). Data mining strategies for real-time control in new york city. *Procedia Computer Science*, 32:109 – 116.

Shankar, V., Venkatesh, A., Hofacker, C., and Naik, P. (2010). Mobile marketing in the retailing environment: Current insights and future research avenues. *Journal of Interactive Marketing*, 24(2):111 – 120.

Shevky, E. and Bell, W. (1955). *Social Area Analysis: Theory, Illustrative Application and Computational Procedure*. Stanford University Press, Stanford.

Shobeiri, S., Mazaheri, E., and Laroche, M. (2015). Shopping Online for Goods vs Services: Where do Experiential Features help more? *International Journal of Consumer Studies*, 39(2):172–179.

Silver, M. (2014). Socio-Economic Status over the Lifecourse and Internet use in older adulthood. *Ageing and Society*, 34:1019–1034.

Simmie, J. and Martin, R. (2010). The Economic Resilience of Regions: Towards an Evolutionary Approach. *Cambridge Journal of Regions Economy and Society*, 3(1):27–43.

Singleton, A. D. (2014). London Output Area Classification: London Datastore. `http://data.london.gov.uk/gla-publications/loac/`. Accessed: 2014-01-12.

Singleton, A. D. and Dolega, L. (2015). The e-resilience of UK town centres. In *Evolving High Streets: Resilience & Reinvention*, Perspectives from Social Science, pages 40 – 43. Economic and Social Research Council.

Singleton, A. D. and Spielman, S. E. (2013). The Past, Present and Future of Geodemographic Research in the United States and United Kingdom. *The Professional Geographer*, 66(4):558 – 567.

Sleight, P. (1997). *Targeting Customers: How to Use Geodemographics and Lifestyle Data in Your Business*. World Advertising Research Centre, Henley-on-Thames.

Srinuan, P., Srinuan, C., and Bohlin, E. (2012). Fixed and mobile broadband substitution in sweden. *Telecommunications Policy*, 36(3):237–251.

Steinfield, C., DiMicco, J. M., Ellison, N. B., and Lampe, C. (2009). Bowling online: social networking and social capital within the organization. In *Proceedings of the fourth international conference on Communities and technologies*, pages 245–254, New York, NY, USA.

Tessler, J. (2009). Broadband funding in stimulus plan sparks debate. `http://usatoday30.usatoday.com/tech/news/2009-02-06-broadband-funding_N.htm`. Accessed: 2013-05-13.

thinkbroadband (2013). Coverage of 2 mbps broadband increases to over 95%. `http://www.thinkbroadband.com/news/5848-coverage-of-2-mbps-broadband-increases-to-over-95.html`. Accessed: 2013-05-21.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:234–240.

Tranos, E. (2013). *The Geography of the Internet*. New Horizons in Regional Science. Edward Elgar, Cheltenham.

Trkman, P. and Turk, T. (2009). A conceptual model for the development of broadband and e-government. *Government Information Quarterly*, 26(2):416–424.

Turner, J. and Gardner, T. (2014). Critical Reflections on the Decline of the UK High Street: Exploratory Conceptual Research into the Role of the Service Encounter. In *Handbook of Research on Retailer-Consumer Relationship Development*, pages 127–151. IGI Global, Hershey, PA.

Uzunboylu, H., Kirlidog, M., and Asuk, C. (2012). A fraud detection approach with data mining in health insurance. *Procedia - Social and Behavioral Sciences*, 62:989 – 994.

van Deursen, A. and van Dijk, J. (2010). Internet skills and the digital divide. *New Media & Society*, 13(6):893–911.

van Dijk, J. and Hacker, K. (2003). The digital divide as a complex and dynamic phenomenon. *The Information Society*, 19(4):315–326.

Verhoef, P. C., Kannan, P., and Inman, J. J. (2015). From multi-channel retailing to omni-channel retailing: Introduction to the special issue on multi-channel retailing. *Journal of Retailing*, 91(2):174 – 181.

Verkasalo, H., López-Nicolás, C., Molina-Castillo, F. J., and Bouwman, H. (2010). Analysis of users and non-users of smartphone applications. *Telematics and Informatics*, 27(3):242 – 255.

Vickers, D. and Rees, P. (2007). Creating the UK National Statistics 2001 output area classification. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):379–403.

Wang, S. S. and Hong, J. (2010). Discourse behind the forbidden realm: Internet surveillance and its implications on china's blogosphere. *Telematics and Informatics*, 27(1):67 – 78.

Warf, B. (2013). Contemporary Digital Divides in the United States. *Tijdschrift voor economische en sociale geografie*, 104(1):1–17.

Warren, M. (2007). The Digital Vicious Cycle: Links between Social Disadvantage and Digital Exclusion in Rural Areas. *Telecommunications Policy*, 31(6-7):374–388.

Webber, R. (1977). Technical Paper 23: an Introduction to the National Classification of Wards and Parishes. Technical report, Centre for Environmental Studies, London.

Wei, R. (2008). Motivations for using the mobile phone for mass communications and entertainment. *Telematics and Informatics*, 25(1):36 – 46.

Wellman, B., Quan Haase, A., Witte, J., and Hampton, K. (2001). Does the internet increase, decrease, or supplement social capital? social networks, participation, and community commitment. *American Behavioral Scientist*, 45(3):436–455.

Weltevreden, J. W. J. (2007). Substitution or complementarity? how the internet changes city centre shopping. *Journal of Retailing and Consumer Services*, 14(3):192 – 207.

West, J. and Mace, M. (2010). Browsing as the killer app: Explaining the rapid success of apple's iPhone. *Telecommunications Policy*, 34(5-6):270 – 286.

Williams, D. (2009). The Evolution of E-tailing. *The International Review of Retail Distribution and Consumer Research*, 19(3):219–249.

Willis, I., Gibin, M., Barros, J., and Webber, R. J. (2010). Applying neighbourhood classification systems to natural hazards: a case study of Mt Vesuvius. *Natural Hazards*, 70:1–22.

Wilson, K. R., Wallin, J. S., and Reiser, C. (2003). Social stratification and the digital divide. *Social Science Computer Review*, 21(2):133–143.

Woods, S. and Reynolds, J. (2012). Leveraging Locational Insights within Retail Store Development? Assessing the use of Location Planners' Knowledge in Retail Marketing. *Geoforum*, 43(6):1076–1087.

Wrigley, N. and Dolega, L. (2011). Resilience, fragility, and adaptation: new evidence on the performance of uk high streets during global economic crisis and its policy implications. *Environment and Planning A*, 43(10):2337–2363.

Wrigley, N. and Lambiri, D. (2014). High Street Performance and Evolution: A Brief Guide to the Evidence. Technical report, University of Southampton, Southampton.

Xavier, P. (2003). Should broadband be part of universal service obligations? *info*, 5(1):8–25.

Xiao, F. and Fan, C. (2014). Data mining in building automation system for improving building operational performance. *Energy and Buildings*, 75(0):109 – 118.

Yang, K. and Bergrud, E. (2008). *Civic engagement in a network society*. Research on International civic engagement. Information Age Pub Incorporated.

Yiu, C. and Fink, S. (2012). The superfast and the furious: Priorities for the future of UK broadband policy. `http://www.policyexchange.org.uk/publications/category/item/the-superfast-and-the-furious-priorities-for-the-future-of-uk-broadband-policy`. Accessed: 2013-01-07.

Yu, Z., Haghighat, F., Fung, B., and Yoshino, H. (2010). A decision tree method for building energy demand modeling. *Energy and Buildings*, 42(10):1637–1646.

Zenter, A. (2008). Online Sales, Internet Use, File Sharing and the Decline of Retail Music Specialty Stores. *Information Economics and Policy*, 20(3):288–300.

Zenter, A., Smith, M., and Kaya, C. (2013). How Video Rental Patterns Change as Consumers Move Online. *Management Science*, 59(11):2622–2634.

Appendix A

Applied Geography Paper

Applied Geography 52 (2014) 25–33

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Broadband speed equity: A new digital divide?

Dean Riddlesden*, Alex D. Singleton

Department of Geography and Planning, School of Environmental Sciences, University of Liverpool, United Kingdom

Keywords:
Broadband
Internet
Digital divide
GIS
Crowdsourced data
Geodemographics

A B S T R A C T

The availability and performance of broadband connectivity is becoming an increasingly important issue across much of the developed world as the prevalence of richer media services and growing populations have generated increasing demands on existing networks. The heterogeneous geography of broadband infrastructure and investments results in variable service provision, and as such, there exist large disparities in access and performance within different spatio-temporal locations. This paper presents analysis of 4.7 million crowdsourced Internet speed test results that were compiled between 2010 and 2013 alongside various indicators of socio-spatial structure to map disparities in English broadband speed between and within urban areas. Although average speeds have improved over time, inequity is shown to emerge between different societal groups and locations. Short-term dynamics also reveal that in areas of different density, speeds can fall dramatically during peak hours, thus influencing the availability of services. The apparent disparities in access and performance represent a major issue as Internet use becomes increasingly ubiquitous in our everyday lives, with inequalities evoking social and economic disadvantage at local and national scales. This work resonates with UK government policy that has stimulated considerable investment in improving infrastructure, and presents analysis of an expansive crowd sourced "big data" resource for the first time.

Broadband and the Internet

Within developed countries, access to the Internet in urban areas is predominantly available in the home (Dutton & Blank, 2013) and enabled through wired broadband infrastructure. This availability reduces traditional constraints on communication such as time and cost, enables the consumption of rich media services (e.g. video chat, streaming movies or music) and most importantly, enhances access to a plethora of online information and services that engender benefits for economic development, education, health and wellbeing (Broadbent & Papadopoulos, 2013; Kraut et al., 2002). Particular emphasis has been placed by governments on the link between the provision of broadband infrastructure and economic growth (Freund & Weinhold, 2004; Picot & Wernick, 2007; Xavier, 2003; Yiu & Fink, 2012), leading to promotion and investment in these technologies as part of national infrastructure plans. In the USA, the National Broadband Plan, developed by the Federal Communications Commission (FCC) aims to promote broadband availability through ensuring robust competition and universal service, as well as maximising the benefits of broadband in government influenced sectors (Federal Communications Commission, 2011). Similarly, UK government response has come in the form of the Digital Britain Report, which outlined a nationwide Universal Service Commitment (USC) to be achieved by 2012 (DCMS, 2009). The importance of such investments have been widely cited, with previous studies estimating that broadband infrastructure accounted for 9.53% of the United Kingdom's GDP growth in the period 2002–2007 (Koutroumpis, 2009). Investment in developing countries has also increased in recent years, with heavily subsidised rollout of high-bandwidth infrastructure across much of the global south (Graham & Mann, 2013), although not necessarily fixed line connectivity. For example, some private sector funded projects aim to provide widespread Internet access through the launch of medium orbit satellites (O3B, 2013). Other projects, such as the One Laptop Per Child program (OLPC), have aimed to promote the distribution of inexpensive computing equipment with wireless connectivity to children (Graham, 2011).

Early research into Internet inequalities was concerned with disparities in connectivity between developed and developing countries, and what social impacts these differences would likely engender. Within this context, the term 'Digital Divide' was introduced in the late 1990s (Norris, 2001) to describe differences between the 'haves' and 'have-nots'. Although a valuable concept at the time when the Internet was first developing, as the access divide narrowed (Kyriakidou, Michalakelis, & Sphicopoulos, 2011;

* Corresponding author. Tel.: +44 (0)1517941355.
E-mail address: d.riddlesden@liv.ac.uk (D. Riddlesden).

http://dx.doi.org/10.1016/j.apgeog.2014.04.008

Peter & Valkenburg, 2006), more recent discussion has diversified this binary concept, adopting instead the term 'Digital Differentiation' (Hargittai, 2002; Longley, Webber, & Li, 2008) to signify those new complexities that have emerged in differentials of Internet access. The digital differentiation approach aims to depict more nuanced differences between societal groups in terms of use and engagement patterns, that are evident especially within those countries with more developed Internet infrastructure (Longley, 2003). Such differentiating factors explored by this research have included age (van Dijk & Hacker, 2003; Lenhart, Madden, & Hitlin, 2005; Livingstone & Helsper, 2007; Loges & Jung, 2001), gender (Bimber, 2000; Cooper, 2006), rurality (Pigg & Crank, 2005; Prieger, 2013) and ethnicity (Fairlie, 2004; Prieger & Hu, 2008). Most recently, research has diversified to utilise crowdsourced information generated through location-based social media and Internet use to better understand the complex emerging geographies linked with the built environment (Arribas-Bel, 2014; Sui & Goodchild, 2011). Manifest from digital differentiation are patterns of digital exclusion where individuals will have varying degrees of engagement with the Internet (Bunyan & Collins, 2013). The complexity of such patterns has been discussed elsewhere (Longley, 2003), and have been shown to emerge between the intersections of traditional concepts of material deprivation (Longley & Singleton, 2009). Thus, you may have urban areas with populations that are typically digitally engaged, yet materially deprived, or inversely, urban areas with low digital engagement, but also low levels of material deprivation. Within a UK context, overall levels of digital exclusion have declined steadily in recent years (Dutton & Blank, 2013; Lane-Fox, 2009), albeit, a significant proportion of the population still remain digitally excluded. In 2013, this equated to around 11.4 m people having never used the Internet (18% of the UK population) (Dutton & Blank, 2013). Although these differentials are part of a wider complex of influences (Parayil, 2005), in part, these inequalities will relate to the provision of infrastructure enabling access to the Internet, and as such, underpin a UK government aim to ensure that a minimum threshold of Internet speed is enabled for all of the populations. The Universal Service Commitment outlined in the government's Digital Britain Report set a nationwide minimum connection speed of 2 Mbps to be delivered by 2012. It was stated that this target "would allow virtually everyone to experience the benefits of broadband, including the increasing delivery of public services online" (DCMS, 2009, p. 27).

Geographic disparities in access and performance exist in part due to the physical structure of broadband networks. In particular, performance is affected by the distance (or line length) between a customer's home and the nearest telephone exchange. Such is the effect, that distance to the nearest exchange is often used as a proxy for deliverable speed (Ofcom, 2012). Within rural areas, distance has a vastly limiting effect on service provision due to sparse distribution of both populations and core network infrastructure such as exchanges and backbone networks; as such, large disparities in broadband performance exist between urban and rural areas. Similar limitations of cellular networks also exist as a result of rurality, with mobile broadband coverage being often poor in isolated areas. However, although relevant to the wider field of research, disparities in mobile broadband access fall outside the scope of this paper. Advances in communications technology, such as the use of fibre optics to supply domestic broadband services, aim to increase broadband speed in locations disparate from telephone exchanges. Fibre to the Cabinet (FTTC) connections link street cabinets that supply small neighbourhood areas to a local telephone exchange with a dedicated fibre link: this allows for much faster transmission of data. FTTC connections (where available) can currently deliver speeds of up to 76 Mbps to homes in the UK (BT PLC, 2013). By contrast, Fibre to the Premises (FTTP) connections (often utilised by businesses) supply a direct fibre link to a site, allowing for large volumes of data to be exchanged rapidly. FTTP connections to domestic properties, offering speeds of up to 300 Mbps are currently only available in a small number of areas and as such are very much in their infancy. Within this context, this paper explores how provision and performance of connections to broadband infrastructure within England are both temporally and socio-spatially differentiated; evoking a complex geography of connectivity both between and within urban areas. For the first time, this study utilises a large dataset of crowdsourced Internet speed estimation tests for England, supplied by the company Speedchecker Limited (broadbandspeedchecker.co.uk), and pertaining to 4.7 million test results, with geographic attribution at the level of the unit postcode (zip code). An extract comprising two time periods was provided, covering 1/1/2010 to 31/1/2011 and 1/4/2012 to 31/5/2013.

Measuring broadband connectivity and access

Data were supplied by Speedchecker Limited, who are a provider of a Web-based application (broadbandspeedchecker.co.uk) that enables users to test their Internet connection speed. When users visit the website, a page is loaded with an embedded testing application that when run provides a small file (of known size) that is automatically downloaded and uploaded, thus enabling speeds to be estimated (i.e. size/time). After running a speed test, users are requested to supply a unit postcode and to confirm details of the connecting Internet company/package. The unit postcode details enable speed test results to be geo-located, and the website displays the test outcome within the context of other results proximal to the supplied postcode. All results derived through the website are stored by Speedchecker Limited as part of their terms of use. The geographic resolution of the geo-located speed tests is therefore high, with postcodes relating to on average around 13 households. It is important to note that these tests differ from "official" speed tests in the UK, which are collated on behalf of the industry regulator Ofcom (http://www.ofcom.org.uk) by SamKnows (www.samknows.com), who are an organisation that provide information about broadband performance, providers and usage. The crucial difference between SamKnows data and the data supplied by Speedchecker Limited is the collection method. Rather than relying on users to run speed tests through a Web application, SamKnows supply hardware in the form of a small testing box that sits between participants' existing routers and the rest of their network. Boxes are supplied to a representative sample of Internet users nationwide. In 2010, Ofcom's UK broadband performance report utilised data collected from 1506 testing boxes. The boxes automatically ran speed tests on a user's connection, but only when there was no other network activity. Conversely, data supplied by Speedchecker Limited is much larger, but there are no restrictions to prevent users from performing tests when there is other network activity ongoing (e.g. multiple users online within a property, or a background update being downloaded). As such, the data presented here could be interpreted as those actual speeds people attain when using a service, taking into consideration local constraints related to router configuration, WiFi coverage or coincident household usage. This said, comparison between the regulator estimates and the derived Speedchecker Limited estimates revealed very similar figures. Our data sample suggested a nationwide average download speed of around 4.8 Mbps in 2010, close to that estimated by Ofcom/SamKnows at 5.1 Mbps (Ofcom, 2010).

The data used for this study can be considered as 'Volunteered Geographic Information' (VGI) (Goodchild, 2007; Haklay, Singleton,

& Parker, 2008), which although a useful way to generate large amounts of geographic information, require interpretive caution related to data accuracy, coverage and bias. As such, both internal and external validations were implemented to explore sources of potential bias or error. Firstly, speed test results were explored to identify anomalous results, or records that may pertain to non-broadband connections. After researching those packages on offer by suppliers during the period the data was collected, it was deemed that the lower boundary for broadband speed would be set at 512 Kbps (half a Megabit per second) and the upper at 102,400 Kbps (100 Megabits per second). Any results that fell below the minimum threshold, or above the maximum were deemed outliers and removed prior to analysis. In applying these thresholds for speed, we expect the results retained to represent a mix of connection types; predominantly traditional ADSL connections, followed by a significant proportion of more recent FTTC and cable connections and a very limited number of FTTP and satellite connections, thus reflecting the distribution of broadband technologies available in England. Further records were also removed where there was either an invalid or blank postcode, thus preventing geocoding. These tests resulted in the removal of around 790,000 (27.5%) records from our 2010/11 dataset, and around 450,000 (24.1%) from our 2012/13 dataset.

Following this analysis, the remaining results were assigned to their nearest telephone exchange using a nearest neighbour algorithm. The location of telephone exchanges was derived from the website SamKnows and geo-coded by unit postcode. The Euclidean distance between each speed test result and the nearest telephone exchange was then calculated. This analysis showed that the vast majority of speed tests were run from postcodes within 2.5 km of their nearest telephone exchange; and the average distance of a speed test location to the nearest exchange was shown to be around 1.6 km in both datasets. These results are in line with what might be expected given what is known about the technical limitations of digital subscriber line technology, particularly that long distances result in high levels of signal attenuation and reductions in speed. As such, it would be unlikely to find broadband connections at distances over 5 km (Ofcom, 2013).

Further validation may have involved locating users by their IP address at the time of completing the test, and then comparing this result to their supplied postcode. However, the data contained no IP address information, and thus, such a test was not possible. In general, analysis of download speeds by distance from the nearest exchange showed that speed tests conducted in close proximity to the nearest telephone exchange returned the fastest download results, particularly those within 1000 m. Speeds start to deteriorate noticeably at distances of over 2000 m. Fig. 1 presents analysis of average download speeds by distance to the nearest telephone exchange.

However, these analyses also highlighted an apparent anomaly. In both datasets, results appeared between 5000 m and 7000 m from the exchange with increases in mean download speeds; and may represent a percentage of speed tests run through fibre or coaxial-based connections. Such infrastructure allows speeds of up to 76 Mbps on fibre to the cabinet (FTTC) lines (Ofcom, 2012). However, within both datasets only a small percentage (0.5% and 0.6% respectively) of the speed test results were recorded at distances over 5000 m from the closest exchange. Table 1 presents the raw number of connections recorded at these distances in our 2012/13 dataset.

To explore the geographic distribution of test results, the number of speed test results per head of population was calculated for each English local authority district; which is a local administrative geography. For this analysis, both datasets were combined in order to assess whether there was any geographic bias within the data over the entire period of data collection. Fig. 2 illustrates the output of this analysis. It is apparent that there are some districts, particularly those that are within predominantly rural areas such as Suffolk, the Lake District and the Cotswolds, which record a higher propensity for speed testing. This may be due to poorer performance within these areas in general due to infrastructure constraints, and a greater tendency for Internet users to seek methods

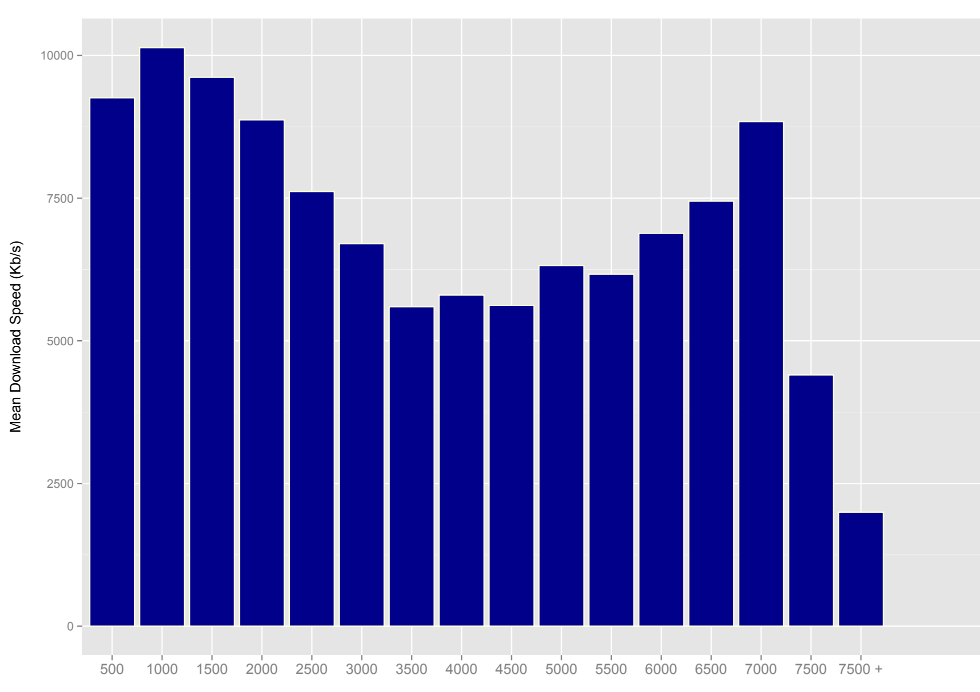


Fig. 1. Mean download speeds by distance to the nearest exchange 2012/13.

Table 1
Number of connections recorded by distance to the nearest exchange 2012/13.

Distance from exchange (m)	Number of speed tests recorded 2012/13
0–500 m	136,697
500–1000 m	279,502
1000–1500 m	292,328
1500–2000 m	247,114
2000–2500 m	181,033
2500–3000 m	119,005
3000–3500 m	73,380
3500–4000 m	41,663
4000–4500 m	22,775
4500–5000 m	11,150
5000–5500 m	4812
5500–6000 m	2458
6000–6500 m	820
6500–7000 m	371
7000–7500 m	150
Over 7500 m	56

of monitoring their connection performance. Conversely, predominantly urban centres such as London, Birmingham, Manchester and Liverpool record fewer speed tests per head of population. This may be due to higher average speeds in these areas as a result of better infrastructure.

To investigate the aggregated socio-spatial structure of test results, the preliminary Office for National Statistics Output Area Classification (OAC) was appended to the test results. This classification was created for England and Wales from 2011 Census Data (www.opendataprofiler.com/). Geodemographic classifications are categorical summary measures of the aggregate social and built environment structure of small areas. These have international use across a wide range of application areas (Singleton & Spielman, 2013) and assist in the analysis and visualization of complex geographies at variable scales (Singleton & Longley, 2009). OAC is appended to Output Area zones, which comprise an area defined by a minimum of 40 households; however, the optimal size is 125 households. OAs contained an average population size of 309 people in 2011. OAC was appended to the records, and the number of speed test results per postcode (the spatial resolution at which our data were geo-tagged) was calculated within each OAC Group (See Fig. 3). For this analysis, we again used the combined datasets, and the number of current postcodes within each OAC group was extracted from the August 2013 Office for National Statistics Postcode Directory.

It is apparent that there are more speed tests performed by residents within certain OAC groups than in others. Group 6b: Established Suburbs records the highest rate at around 4.5 test results per postcode. This is closely followed by groups 6a: Inner Suburbs, 6c: Suburban Aspiration and 4b: Blue Collar Transitions. As we would expect, areas with an over representation of older residents, who are frequently reported to be less engaged with communications technology such as broadband Internet, are under-represented in the dataset. Groups 7c: Elderly in Flats and 8c: Late Retirement record significantly lower rates. The lowest rate is however recorded by group 3A: Urban Deprivation, perhaps where material deprivation has become a more general barrier to Internet use.

Broadband access speeds and equity by indicators of socio-spatial structure

The validated broadband speed test data are utilised in the remaining sections of this paper. Through these analyses we illustrate both the social and spatial patterns of broadband speed inequity, followed by consideration of how these patterns vary over multiple temporal scales. Disparities in broadband speed across England are visualised in Fig. 4 by mapping the average download speeds of our most recent dataset (covering 2012/13) by English local authority district.

There is significant clustering of higher average download speeds around major urban centres such as London, Birmingham, Manchester and Liverpool. As might be expected, predominantly rural regions such as the South West, Norfolk, Yorkshire and Lincolnshire record the majority of the slowest download speeds. On the basis of the averages within the dataset, we estimate that a large proportion of households, particularly those in rural areas, at the time the data were collected, would be falling below the UK government's universal service commitment threshold of 2 Mbps. Of the validated speed test results used for analysis in the 2012/13 dataset, 21.8% fell below the USC threshold.

To examine the relationship between speed and rurality more explicitly, the Office for National Statistics (ONS) Urban/Rural classification was appended to each postcode in the dataset. These definitions were created at a Census Output Area (OA) level of geography for England and Wales, and classified as 'urban' if the majority of the population of an output area lived within settlements with a population of 10,000 or more. In addition, the classification also categorises output areas based on context; such as

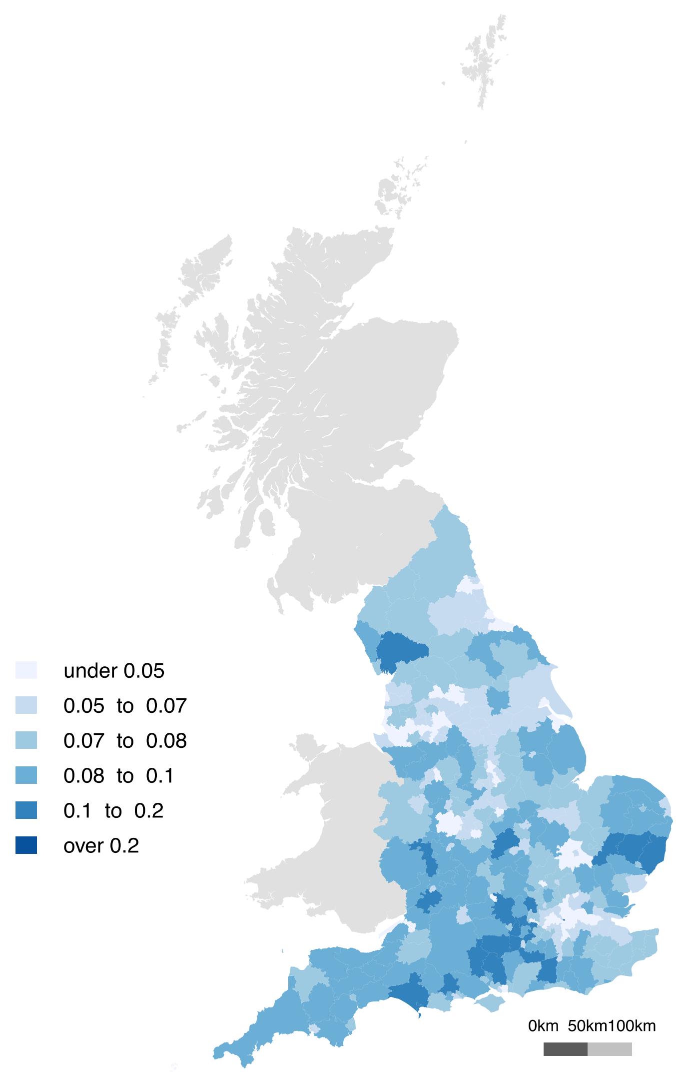


Fig. 2. Speed tests per head of population: English districts.

whether the wider surrounding area of a given output area is sparsely populated or less sparsely populated (Bibby & Shepherd, 2004).

Table 2 shows the average download speed recorded by the Urban/Rural classification alongside average distance to an exchange. The highest average speeds are recorded in those areas classified as 'Urban ≥ 10 K − Less Sparse', essentially areas which are urban and have more densely populated surroundings. Unsurprisingly, the slowest average speeds are recorded in areas classified as 'Hamlets and isolated dwellings' (both 'Sparse' and 'Less Sparse') and villages (again, both 'Sparse' and 'Less sparse'). However, it should be noted that the 'Less Sparse' classes display marginally higher average download speeds than their 'Sparse' counterparts. The results therefore provide further evidence that rurality has a profound effect on the level of service consumers are likely to receive. This is however not a surprising distribution when line length is taken into account, given that longer lines are subject to greater signal attenuation, and as such, likely to deliver slower speeds.

Mean distance to exchange in both the 'Urban ≥ 10 K' and 'Town and Fringe' categories are shorter for the 'Sparse' context than 'Less Sparse'. Initially, this looks anomalous, however, OAs surrounded by less sparsely populated areas would logically be closer to their local exchange, as that exchange would typically be located towards the centre of a populated area in order to supply the best service. In OAs where the surrounding area is classed as 'Less Sparse' (i.e. of higher population density) the logical placement of the exchange may not be as obvious, as the area it must serve is both large and more densely populated throughout. Therefore, it follows that there could be far more OAs at greater distance from their local exchange than those which happen to be closer, thus increasing the average distance.

Although rurality is an important differentiating factor of broadband speed in aggregate, these patterns interact with other social factors. As has been shown elsewhere (Longley & Singleton, 2009), differences in use and engagement with the Internet occur between and at the intersection of patterns of material deprivation. However, to date, equity in broadband speed has not been explored within this context. To examine the relationship between prevailing levels of material deprivation and broadband speeds, the Indices of Multiple Deprivation 2010 (IMD) were converted into ranked deciles and appended to each of the test results. The IMD aims to capture multiple aspects of overall deprivation and is measured at Lower Super Output Area (LSOA) geographic resolution. LSOAs represent small area statistical boundaries that capture between 400 and 1200 households with area populations of between 1000 and 3000 persons. Average speeds were then calculated for each of the IMD decile and are shown in Table 3. These analysis show that average download speeds are higher in those areas ranking as more deprived (particularly so for deciles 1 and 2).

This is due to such areas of high deprivation typically being located within more densely populated urban conurbations, and as such, benefiting from more developed network infrastructure that is necessary to support higher speeds. These results would suggest that prevailing levels of deprivation in England are not necessarily a barrier to broadband access or to higher speed broadband connections. However, one caveat of this analysis is that although speed tests (based on the user supplied postcode) were relatively evenly distributed across all IMD deciles, we possess no data regarding the

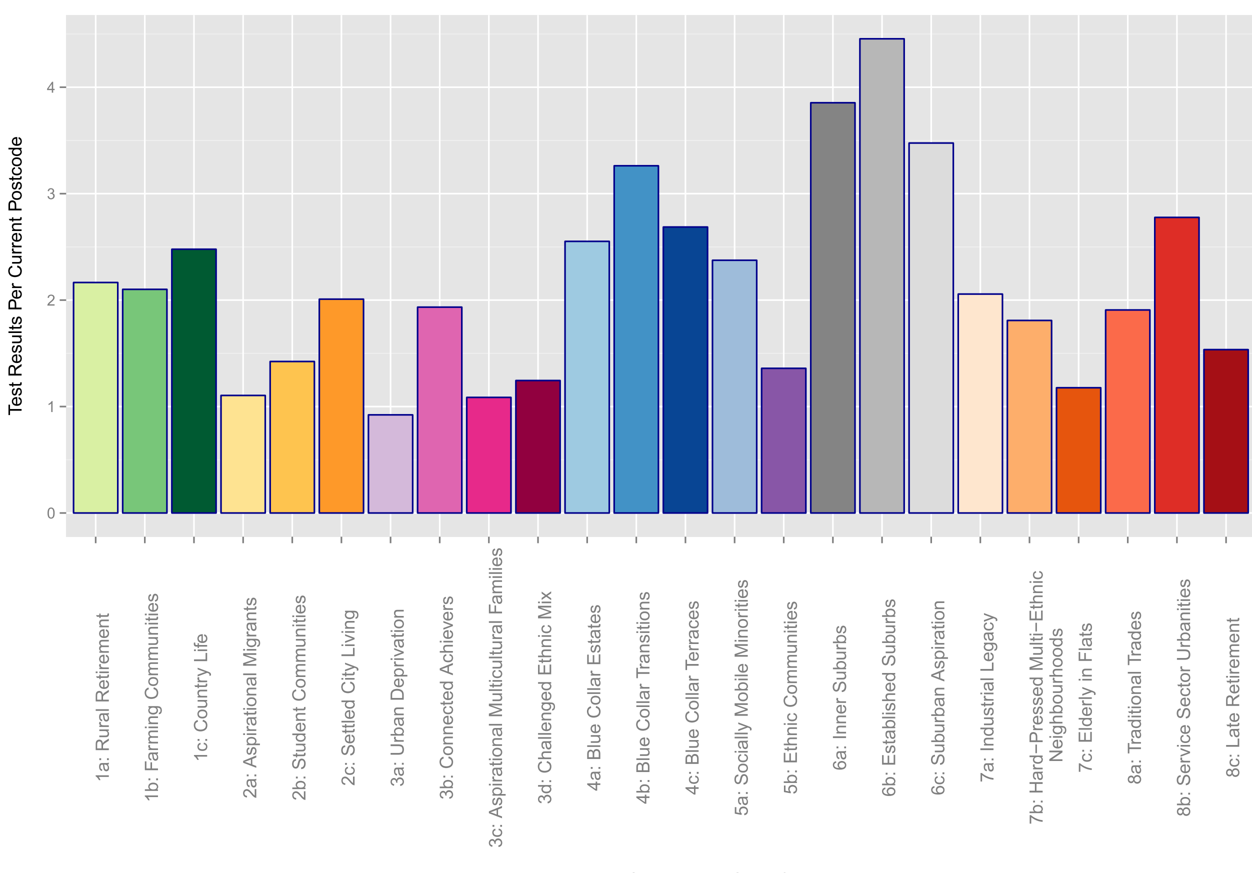


Fig. 3. Test rate (tests per postcode) by 2011 Output Area Classification (OAC).

levels of material deprivation that are experienced by each individual test user, only that at the aggregate level of the LSOA.

The short and long term temporal dynamics of broadband speed

A further attribute for each of the speed test results is a ‘time stamp’ assigned by the Web server that identifies the time and date when each test was performed. This usefully enables expansion of the socio-spatial analysis across varying temporal scales to provide insight into these changing dynamics. Fig. 5 extends the previously presented urban/rural analysis and illustrates how speeds fluctuate throughout the day within different types of conurbation.

Densely populated urban areas classified as ‘Urban ≥ 10 K Less Sparse’ record the highest average download speeds throughout the day, with ‘Town and Fringe’ and sparsely populated urban areas sitting around mid-table. ‘Villages’ and ‘Hamlets and Isolated Dwellings’, both sparse and less sparse record the lowest average download speeds throughout the day. In terms of performance variation, there appear to be larger fluctuations in urban areas that are likely due to higher population densities, and more prevalent “bottlenecking” of data traffic when there are a large number of users connected to the Internet. Bottlenecking is a term used to refer to the slowing of a network in times of high use. Within predominantly rural areas, fluctuations in download speed appear to be smaller, although speeds are generally much slower throughout the day. Despite delivering slower speeds, infrastructure in rural areas appears to be less susceptible to bottlenecking, and is most likely due to their composite smaller populations, and as such, lower demand. In all Urban/Rural categories the effects of peak and off peak hours are apparent, with average speeds spiking in the early hours of the morning when there are fewer users online, allowing for faster download speeds. Although we would expect these increases to be measurable, the large spike at 3am in the ‘Town and Fringe Sparse’ class represents noise in this particular subset of our data. Conversely, the slowest average speeds are recorded between 6pm and 9pm in most areas, which is a peak time for Internet use, and as such, aggregate speeds tend to slow. As well as detailing performance fluctuations, this analysis again highlights the large disparities that exist between urban and rural areas in terms of aggregate broadband speeds.

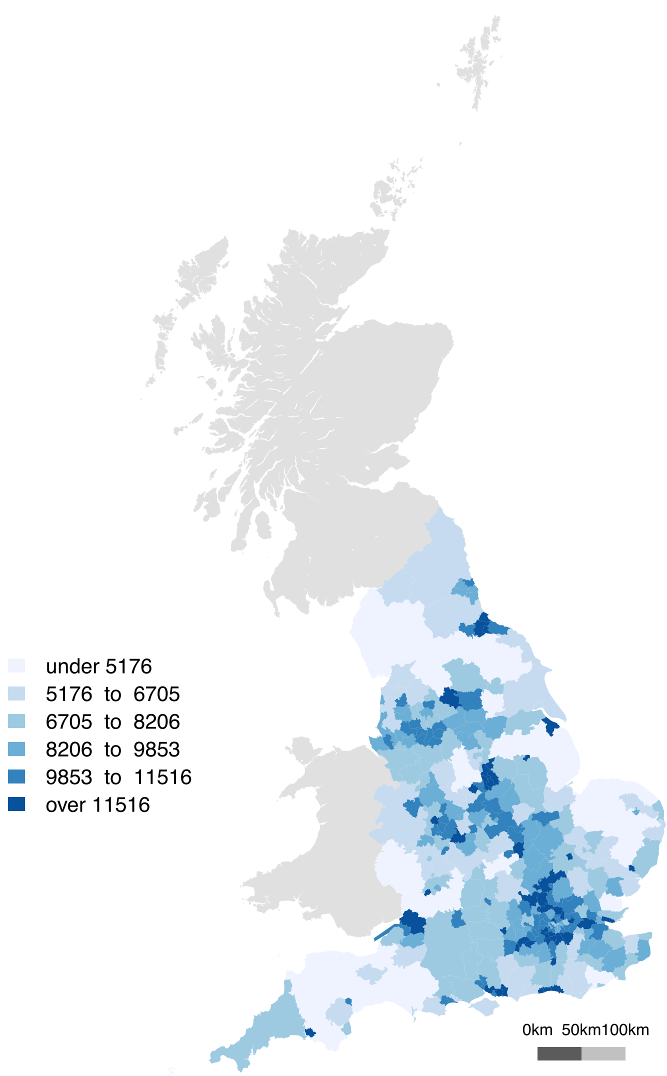


Fig. 4. Mean download speed (Kbps) by English district: 2012/13.

Table 2
Mean download speeds by urban/rural indicator.

Urban/rural indicator	Mean download (Kbps) 2012/13	Mean distance to exchange (m)
Urban ≥ 10 K sparse	6651	971
Town and fringe sparse	6248	736
Village sparse	3380	2272
Hamlet and isolated dwelling sparse	3151	2790
Urban ≥ 10 K less sparse	10,176	1496
Town and fringe less sparse	6952	1575
Village less sparse	4042	2375
Hamlet and isolated dwelling less sparse	3896	2477

However, we can also consider these data at a more granular temporal resolution to highlight how inequalities are changing over time. If the two annual extracts of 2010/11 and 2012/13 are considered in isolation, we find that the national average speed increases from 4.8 Mbps to 8.4 Mbps, with a reduction in the estimated number of people falling bellow the USC 2 Mbps threshold from 30.9% to 21.8%. However, if these results are mapped at the local authority district geography, we see that there is geographic disparity to these improvements (See Fig. 6 and Tables 4 and 5). Unlike the previously presented map of average download speeds (See Fig. 4), change is less obviously clustered around urban conurbations, and instead, high increases in download speed are more widely dispersed. The lowest increases appear predominantly in rural areas, however, some rural areas such as Norfolk and Cornwall have seen relatively high speed increases over the time period. It would be safe to assume that large urban conurbations have not seen large percentage increases in average download speeds as the infrastructure necessary to support high speeds has typically long been in place within these areas. However, the largest increases in average download speeds are predominantly urban, but would not be considered as major urban centres. In most cases, these represent areas that may not previously have benefited from next generation infrastructure that have replaced elements of the cabling carrying broadband with fibre. More densely populated urban centres would logically offer the best return on investment for such enhanced services, and as such, would likely have had infrastructure upgrades prioritised.

As might be expected, the smallest increases in average download speeds were recorded in those districts that are predominantly rural. Districts such as Copeland, Boston and the Isles of Scilly, in particular, are rurally isolated and sparsely populated which constrains the viability of infrastructure upgrades. For example, areas

Table 3
Mean download speeds by IMD decile.

IMD decile	Mean download 2012/13 (Kbps)
Decile 1 (Most deprived)	10,356
Decile 2	9730
Decile 3	8822
Decile 4	8037
Decile 5	7675
Decile 6	7387
Decile 7	8067
Decile 8	8009
Decile 9	8779
Decile 10 (Least deprived)	9326

such as these would probably require government intervention to stimulate the required investment in infrastructure. One such initiative in the UK has been the Rural Community Broadband Fund (RCBF), jointly funded by Defra and Broadband Delivery UK (BDUK), which is aiming to deliver improvements in broadband infrastructure to the most rurally isolated areas.

The relationship between infrastructure investment over these two time periods and broadband speed can be illustrated further by exploring the geography of fibre enabled local exchanges. For a consumer to access a fibre enhanced broadband service, the local telephone exchange serving a property would require this provision to be enabled. In order to investigate the geography of these next generation network upgrades, a comprehensive list of exchanges that are supplying FTTC connections was obtained from BT. The dataset contained the name and a code of each exchange in the UK that was supplying BT's 'Infinity' fibre services. These data were geo-tagged by matching attribute information from our original exchange dataset obtained from SamKnows. The percentage of fibre enabled exchanges within each district was then calculated relative to the total number of exchanges in that district. Fig. 7 shows this analysis on a national scale.

Broadly speaking, the geography of fibre enabled exchanges matches that of broadband speeds, with more FTTC enabled exchanges in predominantly urban areas. London and Manchester in particular have high percentages of FTTC enabled exchanges; however, it is apparent that this is also the case in some rural areas such as Cornwall. Cornwall has been promoting the expansion of fibreoptic broadband through the Superfast Cornwall Program, a £132 m partnership funded by the European Regional Development Fund Convergence Programme, BT and Cornwall Council. This has aimed to upgrade existing broadband infrastructure and enable fibre access to 95% of homes and businesses in Cornwall by 2015. Many other rural areas such as North Yorkshire and Cumbria also have programs in place that make use of BDUK funds to supply superfast broadband to the most hard to reach areas.

Conclusions

This paper has provided an overview of how crowdsourced speed test estimation data can be used to investigate the variable disparities in English broadband infrastructure and access patterns. In doing so, we have presented internal and external validation methods that seek to assess the integrity of the 4.7 million speed test records provided by Speedchecker Limited for the purpose of this study. The validated data were shown to produce broadly similar estimates (here, in terms of nationwide average download speeds) to that of regulator data reported by Ofcom, although, there was some socio-spatial bias within the data which appeared to relate to patterns of overall access participation. In general, analysis of test results by distance to the closest exchange revealed that tests run from close proximity to their nearest exchange returned the highest average download speeds. However, there were distances at which fibre-based connections appeared prevalent, particularly

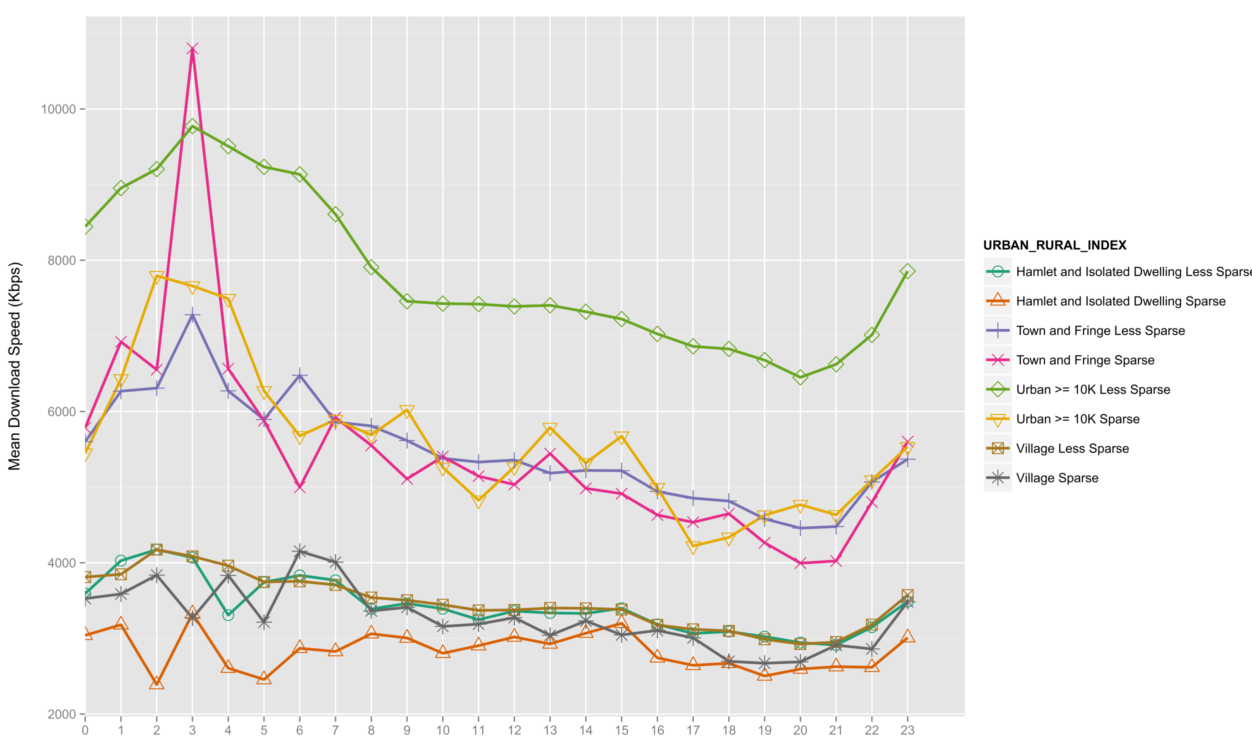


Fig. 5. Hourly fluctuations in mean download speed by urban/rural index.

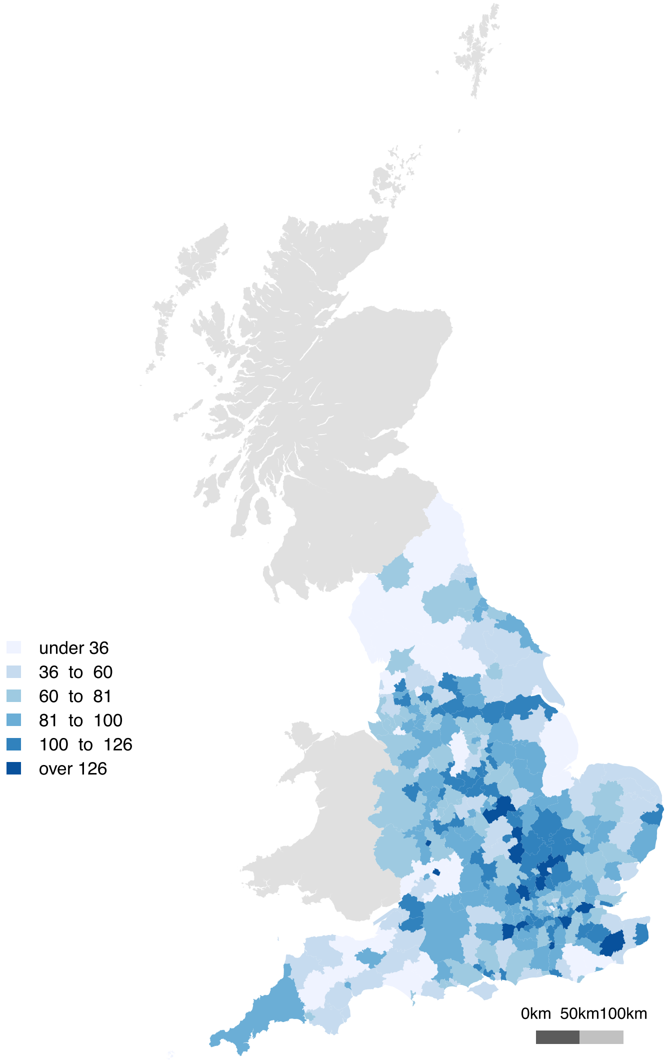


Fig. 6. Percent increase in mean download speed 2010/11 to 2012/13 by English district.

5000–7000 m from a local exchange, and disrupted this trend. More generally, it would appear that predominantly rural areas record a higher propensity for speed testing, possibly due to variable levels of performance as a result of longer lines; motivating a necessity or interest in monitoring speed. Conversely, urban areas tended to record fewer tests per head of population. When profiling by a geodemographic indicator, OAC in this case, further data bias

Table 4
Largest ten increases in mean download speeds by English district 2010/11 to 2012/13.

District	Mean download 2010/11 (Kbps)	Mean download 2012/13 (Kbps)	% Increase
Harborough	3763	10,366	175
Ashford	3725	9363	151
Chiltern	4070	10,193	150
Wellingborough	4635	11,601	150
North Hertfordshire	5302	13,112	147
Worcester	3135	7638	144
Surrey Heath	5265	12,600	139
Bromley	5471	12,915	136
Thurrock	5114	12,013	135
Milton Keynes	3474	8128	134

Table 5
Smallest ten increases in mean download speeds by English district 2010/11 to 2012/13.

District	Mean download 2010/11 (Kbps)	Mean download 2012/13 (Kbps)	% Increase
Copeland	3467	3491	0.67
Allerdale	3634	3726	2.53
Craven	4275	4470	4.56
Mid Devon	3248	3667	12.92
Barrow-in-Furness	3702	4226	14.17
West Somerset	3149	3711	17.83
Richmondshire	3332	3932	18.01
Boston	3279	3870	18.02
Cotswold	3599	4279	18.90
Isles of Scilly	2558	3044	18.99

was identified with some groups such as ‘Established Suburbs’ overrepresented in the data. Equally, some groups such as ‘Elderly in Flats’ were underrepresented, however, these biases are to be expected and broadly support the notion of digital differentiation between those societal groups likely resident within these areas.

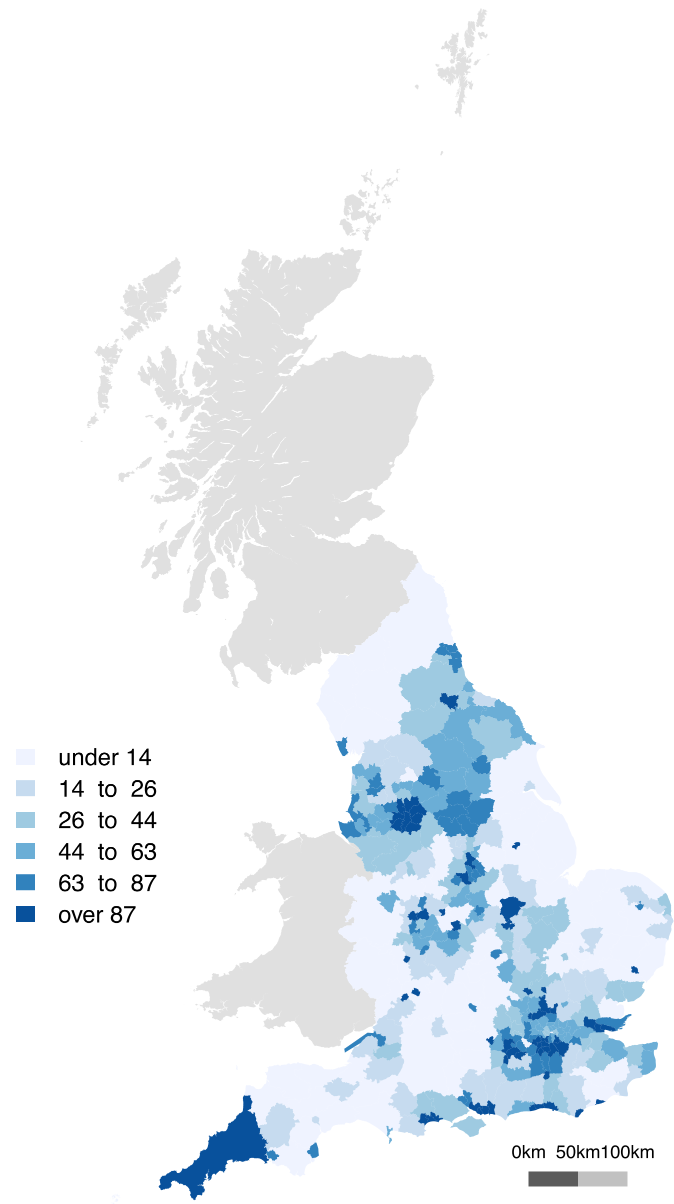


Fig. 7. Percentage of FTTC enabled telephone exchanges by district (October 2013).

Profiling average download speeds at a regional geography revealed that higher speeds are generally clustered around major conurbations. In rural areas, average speeds are shown to be significantly slower. Further investigation revealed that a large proportion of speed test results in our data fell below the USC threshold of 2 Mbps, however the number of these slow connections was shown to be falling across the two annual extracts. Profiling speeds by measures of deprivation showed that average download speeds are higher in those areas ranking as more deprived. It would also appear that prevailing levels of deprivation in England are not necessarily a barrier to broadband access; although, there still remain those constraints of cost and behavioural choice that would result in variable uptake or use.

It is evident that densely populated urban areas record the highest average download speeds throughout the day, but suffer the most in terms of "bottlenecking" of data traffic during peak hours; conversely, small conurbations, and those which are rurally isolated, record far lower speeds, but, are impacted less at peak hours which is likely due to smaller populations. Splitting our data into annual extracts and mapping increases in average download speed at a district level revealed that large increases in speed were not as obviously clustered around urban conurbations, instead, high increases in download speed are more widely dispersed. Some rural areas were shown to be investing heavily in fibre broadband provision to ensure universal access to superfast connections.

The policy implications of this work are diverse, but initially show a requirement for further investment in broadband infrastructure to eliminate poor service below the minimum threshold set by UK Government, and to ensure more equitable delivery of next generation services. Analysis of the dynamics of broadband speeds throughout the day has highlighted implications for use, particularly so for those residents in areas that suffer from large performance fluctuations at peak hours. Upgrading traditional copper networks to carry broadband over fibre will help to eliminate such unstable connections and ensure robust connectivity both geographically and temporally, with longer-term benefits for the UK in an international economic context.

References

Arribas-Bel, D. (2014). Accidental, open and everywhere: emerging data sources for the understanding of cities. *Applied Geography, 49*, 45–53.

Bibby, P., & Shepherd, J. (2004). *Developing a new classification of urban and rural areas for policy purposes – The methodology*. Technical report. Office for National Statistics.

Bimber, B. (2000). Measuring the gender gap on the internet. *Social Science Quarterly, 81*(3), 868–876.

Broadbent, R., & Papadopoulos, T. (2013). Impact and benefits of digital inclusion for social housing residents. *Community Development, 44*(1), 55–67.

BT PLC. (2013). *BT products and services: BT infinity*. http://www.productsandservices.bt.com/products/broadband/infinity Accessed 14.11.13.

Bunyan, S., & Collins, A. (2013). Digital exclusion despite digital accessibility: empirical evidence from an English City. *Tijdschrift voor Economische en Sociale Geografie, 104*(5), 588–603.

Cooper, J. (2006). The digital divide: the special case of gender. *Journal of Computer Assisted Learning, 22*(5), 320–334.

DCMS. (2009). *Digital Britain final report*. http:/interactive.bis.gov.uk/digitalbritain/final-report/ Accessed 20.05.13.

Dutton, W., & Blank, G. (2013). *Cultures of the internet: The internet in Britain*. Technical report. Oxford Internet Institute.

Fairlie, R. (2004). Race and the digital divide. *Contributions to Economic Analysis and Policy, 3*(1), 155–194.

Federal Communications Commission. (2011). *Connecting America: National broadband plan*. http://www.broadband.gov/download-plan/ Accessed 13.05.13.

Freund, C. L., & Weinhold, D. (2004). The effect of the internet on international trade. *Journal of International Economics, 62*(1), 171–189.

Goodchild, M. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal, 69*, 211–221.

Graham, M. (2011). Time machines and virtual portals: the spatialities of the digital divide. *Progress in Development Studies, 11*(3), 211–227.

Graham, M., & Mann, L. (2013). Imagining a silicon savannah? Technological and conceptual connectivity in Kenya's BPO and software development sectors. *The Electronic Journal of Information Systems in Developing Countries, 56*, 1–19.

Haklay, M., Singleton, A. D., & Parker, C. (2008). Web mapping 2.0: the neogeography of the GeoWeb. *Geography Compass, 2*(6), 2011–2039.

Hargittai, E. (2002). Second-level digital divide: differences in people's online skills. *First Monday, 7*(4). http://firstmonday.org/issues/issue7_4/hargittai/index.html Accessed 17.05.13.

Koutroumpis, P. (2009). The economic impact of broadband on growth: a simultaneous approach. *Telecommunications Policy, 33*(9), 471–485.

Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., & Crawford, A. (2002). Internet paradox revisited. *Journal of Social Issues, 58*(1), 49–74.

Kyriakidou, V., Michalakelis, C., & Sphicopoulos, T. (2011). Digital divide gap convergence in Europe. *Technology in Society, 33*(3–4), 265–270.

Lane-Fox, M. (2009). *Champion for digital inclusion: The economic case for digital inclusion*. Technical report. Pricewaterhousecoopers LLP.

Lenhart, A., Madden, M., & Hitlin, P. (2005). *Teens and technology: Youth are leading the transition to a fully wired and mobile nation*. Technical report. Pew Internet.

Livingstone, S., & Helsper, E. (2007). Gradations in digital inclusion: children, young people and the digital divide. *New Media and Society, 9*(4), 671–696.

Loges, W. E., & Jung, J.-Y. (2001). Exploring the digital divide, internet connectedness and age. *Communication Research, 28*(4), 536–562.

Longley, P. A. (2003). Towards better understanding of digital differentiation, computers. *Environment and Urban Systems, 27*(2), 103–106.

Longley, P. A., & Singleton, A. D. (2009). Linking social deprivation and digital exclusion in England. *Urban Studies, 46*(7), 1275–1298.

Longley, P. A., Webber, R., & Li, C. (2008). The UK Geography of the e-society: a national classification. *Environment and Planning A, 40*(2), 362–382.

Norris, P. (2001). *Digital divide*. Cambridge: Cambridge University Press.

O3B. (2013). *Our technology*. http://www.o3bnetworks.com/o3b-advantage/our-technology Accessed 04.06.13.

Ofcom. (2010). *UK broadband speeds, May 2010: The performance of fixed-line broadband delivered to UK residential consumers* (Technical report).

Ofcom. (2012). *Infrastructure report: 2012 Update*. http://stakeholders.ofcom.org.uk/market-data-research/other/telecoms-research/broadband-speeds/infrastructure-report-2012/ Accessed 08.05.13.

Ofcom. (2013). *UK fixed-line broadband performance, May 2013: The performance of fixed-line broadband provided to UK residential consumers* (Technical report).

Parayil, G. (2005). The digital divide and increasing returns: contradictions of informational capitalism. *Information Society, 21*(1), 41–51.

Peter, J., & Valkenburg, P. M. (2006). Adolescents' internet use: testing the "disappearing digital divide" versus the "emerging digital differentiation" approach. *Poetics, 34*(4–5), 293–305.

Picot, A., & Wernick, C. (2007). The role of government in broadband access. *Telecommunications Policy, 31*(10), 660–674.

Pigg, K. E., & Crank, L. D. (2005). Do information communication technologies promote rural economic development? *Community Development, 36*(1), 65–76.

Prieger, J. E. (2013). The broadband digital divide and the economic benefits of mobile broadband for rural areas. *Telecommunications Policy, 37*(6–7), 483–502.

Prieger, J. E., & Hu, W. M. (2008). The broadband digital divide and the nexus of race, competition, and quality. *Information Economics and Policy, 20*(2), 150–167.

Singleton, A. D., & Longley, P. A. (2009). Geodemographics, visualisation, and social networks in applied geography. *Applied Geography, 29*(3), 289–298.

Singleton, A. D., & Spielman, S. E. (2013). *The past, present and future of geodemographic research in the United States and United Kingdom*. The Professional Geographer.

Sui, D., & Goodchild, M. (2011). The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographical Information Science, 25*(11), 1737–1748.

van Dijk, J., & Hacker, K. (2003). The digital divide as a complex and dynamic phenomenon. *The Information Society, 19*(4), 315–326.

Xavier, P. (2003). Should broadband be part of universal service obligations? *Info, 5*(1), 8–25.

Yiu, C., & Fink, S. (2012). *The superfast and the furious: Priorities for the future of UK broad-band policy*. http://www.policyexchange.org.uk/publications/category/item/the-superfast-and-the-furious-priorities-for-the-future-of-uk-broadband-policy Accessed 07.01.13.

Appendix B

Geoforum Paper

Geoforum 69 (2016) 5–18

Contents lists available at ScienceDirect

Geoforum

journal homepage: www.elsevier.com/locate/geoforum

Measuring the spatial vulnerability of retail centres to online consumption through a framework of e-resilience

A.D. Singleton [a,*], L. Dolega [a], D. Riddlesden [a], P.A. Longley [b]

[a] Department of Geography and Planning, University of Liverpool, Roxby Building, Liverpool L69 7ZT, UK
[b] Department of Geography, University College London, Pearson Building, Gower Street, WC1E 6BT, UK

ARTICLE INFO

Article history:
Received 5 August 2015
Received in revised form 14 November 2015
Accepted 18 November 2015
Available online 11 December 2015

ABSTRACT

This paper presents e-resilience as a framework for assessing the extent to which retail centres have spatially differentiated vulnerability to the impacts of online consumption. This extends the conceptual model of resilience as applied to retail, and is operationalised through a novel methodology that develops two indices that balance both supply and demand influences. We describe the creation of a composite e-resilience indicator, and then calculate it for retail centres across England. Our findings suggest a geographic polarising effect, with least vulnerable centres identified as large and more attractive or as smaller local destinations with a focus on convenience shopping. Mid-sized centres were typically shown to be the most exposed, and are argued as having a less clearly defined function in contemporary retail. Such findings have wide policy relevance to stakeholders of retail interested in the future configuration of sustainable and resilient provision.

1. Introduction and background

UK Government initiatives aimed at revitalisation of British high streets (Portas, 2011; Digital High Street Advisory Board, 2015) highlight the importance of digital technology in redefining traditional retail spaces. Evidence suggests that growth in online consumption impacts upon the health of retail centres in complex ways (Weltevreden, 2007), and can be viewed as a source of long-term change to their structure, often referred to in the literature as a ‘slow burn’ (Pendall et al., 2010). Adjustments to the market share of traditional town centre retailing have been mainly considered with respect to their supply side effects: for example, the extent to which online shopping has substituted, modified or complemented traditional town centres (Weltevreden, 2007; Doherty and Ellis-Chadwick, 2010). However, there has been less focus on how the structure of traditional high streets are or might be impacted by consumer propensity for online shopping, how such effects could be modelled, or what might be an appropriate adaptive response by stakeholders in retail. Despite evidence to suggest that factors impacting decisions about whether or not to shop online are linked to demographic and socio-economic characteristics of populations (Longley and Singleton, 2009), we nevertheless posses limited knowledge about the geography of online sales (Forman et al., 2008).

This paper explores these challenges through developing the concept of *e-resilience*, which provides both a theoretical and methodological framework that defines the vulnerability of retail centres to the effects of rapidly growing Internet sales, balancing characteristics of both supply and demand. We argue that the concept of e-resilience adds value to existing research in the following ways:

(i) It provides insight into wider debates on the performance of UK town centres in the rapidly transforming retail landscape, in particular by assessing their resilience and adaptability to the growth of online sales.
(ii) It provides insight into demand through examination of local catchment demographics, and thus rebalances current debates on the resilience of high streets, which hitherto have predominantly focused on supply effects, measured through outcomes such as retailer failures or vacancy rates.
(iii) It delivers valuable outputs including: an operational measure of e-resilience, which is implemented to define those retail centres in England that are the most or least e-resilient.
(iv) It presents a national geodemographic of Internet user behaviour. It is anticipated that such outputs will be of interest to a wide range of stakeholders in retail policy and provision.

The general concept of resilience has been established for some time to describe how various types of system respond to

* Corresponding author.
E-mail address: alex.singleton@liverpool.ac.uk (A.D. Singleton).

http://dx.doi.org/10.1016/j.geoforum.2015.11.013
0016-7185/

unexpected shocks. There are three widely recognised concepts of resilience adopted between different scientific traditions (Simmie and Martin, 2010): (a) the engineering resilience interpretation found in physical sciences; (b) the ecological resilience interpretation found in biological sciences; and (c) the adaptive resilience interpretation found in complex systems theory. The first two interpretations refer to the notion of equilibrium, which suggest that a resilient economic system would adapt successfully to disturbance and either resume, or even improve its long-run equilibrium growth path. Conversely, a non-resilient system would fail to transform itself successfully and instead become 'locked' into a long-run outdated trajectory or decline (Simmie and Martin, 2010; Dawley et al., 2010). The third interpretation, identified by Martin (2011) as 'adaptive resilience', stresses the anticipatory or reactive reorganisation of the form and function of a system to minimise the impact of a destabilising shock, and focuses on resilience as a dynamic and evolutionary process. Complex system theory is characterised by non-linear dynamics and self-reinforcing interactions among a system of components (Martin and Sunley, 2007), and highlights self-organisation, with adaptive growth relative to changes in the external environment (e.g. the impact of online sales on traditional retailers).

Increasing numbers of social scientists have also begun to use resilience as a mechanism to help explain the impact and response to disruptions and more gradual processes of change in a range of socio-economic systems (Christopherson et al., 2010; Pendall et al., 2010; Hassink, 2010; Simmie and Martin, 2010; Martin, 2011). For example, resilience was first considered within the context of the UK high street by Wrigley and Dolega (2011), who investigated the dynamics of performance and adjustment to the shock of the global economic crisis. In this work they rejected the notion that town centres and high streets could return to their pre-shock configurations, and developed the concept of "adaptive resilience" whereby the resilience of UK town centres was viewed as a dynamic and evolutionary process. More specifically, they argued that the response of UK town centres to economic and competitive shocks can be seen as a function of the mix and interdependencies of existing business, the dynamics of centres, diversity, attractiveness, accessibility, national planning policies and the socio-demographic characteristics of local catchments. Such characteristics and actions are responsible for building town centre adaptive capacity. Often an economic or competitive shock creates new opportunities for development and innovation (Pendall et al., 2010; Raco and Street, 2012) which, in turn, leads to the emergence of more adaptable town centres characterised by enhanced resilience and ability to more effectively withstand future disturbances. The resilience framework strengthens some basic arguments derived from evolutionary economics such as the advantages of diversity, seeing regional economies as path-dependent systems (Hassink, 2010), or the potential of novelty and selection in economic systems as they adjust to evolving circumstances (Simmie and Martin, 2010). Furthermore, Dolega and Celinska-Janowicz (2015) argue that future resilience of town centres is crucially dependent upon recognising and acting upon the challenges arising from current trends. A good example of such pre-emptive action in the UK was the establishment of the Digital High Street Advisory Board in 2014 to provide an independent assessment of strategies to revitalise high streets in the context of a digital future.

Equally important to retail centre resilience is an understanding of the geodemographic characteristics of local catchments (Birkin et al., 2002), as consumer choices and behaviours are fundamental drivers of demand, and therefore are closely related to evolution of the retail landscape. Importantly, the behaviours and attitudes of consumers vary spatially, yet are directly linked to the geography of demand for retail facilities. Understanding the geography of consumer behaviour (such as varied propensity for online shopping) at a small area level is crucial to understanding the vitality and viability of both retail centres and the retailers themselves. Indeed, the resilience of retail centres is intertwined with the underlying dynamics of their catchments as variations in consumer confidence (Wrigley and Dolega, 2011) and basic digital skills (Digital High Street Advisory Board, 2015) are likely to shape demand for retail spaces in the digitally transformed retail landscape. The current debate on economic health of UK town centres seems to acknowledge the key role consumers have in that transformation, and a direct response of retail spaces to consumers' needs is being perceived as key to their success (DCLG, 2013).

2. A framework to understand and measure e-resilience

The Internet enhances opportunities for price comparison, enables 24/7 convenience, provides a selection of products not limited by physical space, and enables distribution with a wider geographical reach (Williams, 2009). As such, it is perhaps unsurprising that online sales have been growing exponentially; essentially tripling over the past eight years, and are forecasted to reach 15.2% of all UK retail sales by the end of 2015 (Centre for Retail Research, 2015). Furthermore, in recent years, there has also been a shift towards using mobile devices for online shopping, such as tablets and smartphones, which are now estimated to account for the majority of growth in UK online retail sales (Capgemini, 2015). In consequence, the rapid expansion of online shopping has been increasingly viewed as a major cause of change to the structure of traditional UK high streets (Digital High Street Advisory Board, 2015; Wrigley and Lambiri, 2014). Weltevreden (2007) investigated the implications of e-commerce on traditional physical shopping space, and established the extent to which online retailing could be associated with processes of substitution, complementarity, and modification of traditional retail channels. Substitution occurs when e-commerce replaces physical shopping; however, complementarity and modification pertain to a blending of e-commerce with traditional retailing. These latter two omni-channel processes are however, less well understood (Wrigley and Lambiri, 2015; Weltevreden, 2014). For instance, in the UK a number of national retailers such as Borders, Zavvi, Jessops and Game have either entirely withdrawn or substantially limited their physical retail offerings within the past few years, while some other major retailers such as John Lewis, Next, Boots or Argos have successfully embraced new technologies through opening 'click and collect' points, or by developing mobile applications (Turner and Gardner, 2014).

The basic concept of e-resilience defines the vulnerability of retail centres to the effects of growing Internet sales, and estimates the likelihood that their existing infrastructure, functions and ownership will govern the extent to which they can adapt to or accommodate these changes. Essentially, e-resilience can be expressed as a balance between the propensity of localised populations to engage with online retailing and the physical retail provision and mix that might increase or constrain these effects, as not all retail categories would be equally impacted. However, estimating the interaction between potential consumers and retail destinations is increasingly complex. For example, there is emerging evidence that choice is related to both experiential factors (Wrigley and Lambiri, 2014; Shobeiri et al., 2015) and a provision of a broader range of services, technologies and activities within shopping localities (Hart and Laing, 2014; Digital High Street Advisory Board, 2015). Although some of these factors may be difficult to quantify for a national extent, empirical evidence suggests that presence of anchor stores and various service providers (typically those difficult to digitise) such as leisure, are associated with lower

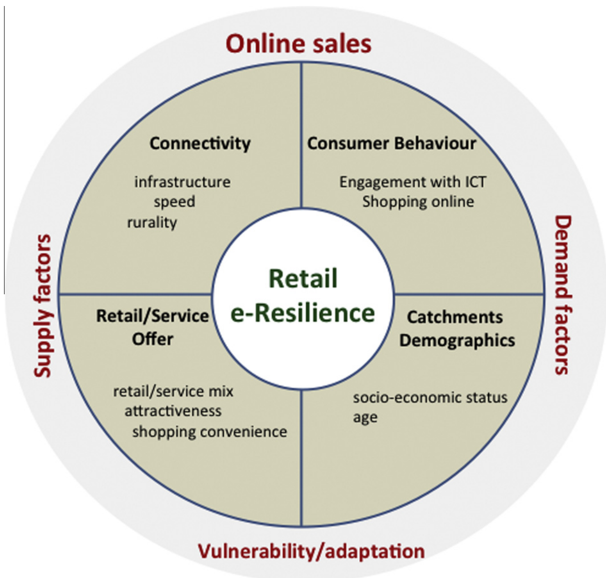


Fig. 1. A conceptual framework of e-resilience.

online substitution rates (Weltevreden, 2014). In other words, customers who have relatively easy access to the most attractive stores that are enhanced by adjacent leisure facilities tend to visit town centres for longer periods of time, and are normally expected to spend more within them (Hart and Laing, 2014; Wrigley and Lambiri, 2014). Furthermore, it has also been well documented that the impact of online shopping is not uniform across retail types (Zentner, 2008; Ryan and Been, 2013; Parker and Weber, 2013). Typically, retailers who merchandise goods that can be easily digitised such as music, videos, computer games or books are amongst the most vulnerable (Zentner et al., 2013). Use of the Internet for these retail types is estimated at 44% (ONS, 2014), which makes them susceptible to competition from online retailers.

Constructing a measure of e-resilience for a retail centre requires an array of knowledge about the characteristics and mix of retail offerings, alongside demographic and probable Internet engagement characteristics of likely consumers. An empirical model must therefore ensure that influences of both supply and demand can be estimated, and consideration is required about how these measures interact. Such issues are explored in the remaining sections of this paper, however, Fig. 1 summarises the range of influences on retail e-resilience as including: connectivity, behaviour, demographics and the retail/service offer.

3. Demand factors and Internet engagement behaviour

Demand and Internet engagement behaviours have been shown to map onto a range of influencing factors pertaining to the characteristics of people and the places in which they live (Longley and Singleton, 2009; Van Deursen and Van Dijk, 2011). Across a range of contexts, influences have been identified as including: demographics, such as age (Rice and Katz, 2003; Warf, 2013), socio-economic status (Silver, 2014), ethnicity (Wilson et al., 2003) or gender (Prieger and Hu, 2008); context, including rurality (usually measured by population density or road connectivity: Warren, 2007), education (Helsper and Eynon, 2010); and finally, Internet connectivity, including the underlying infrastructure that is available within an area to facilitate users getting online (Maillé and Tuffin, 2010; Grubesic and Murray, 2002), and the speed of connection that these access modes enable (Riddlesden and Singleton, 2014). A behavioural component captures the decision whether or not to use the Internet for a given activity, over any number of other modes of access. Influencing such decisions are both demographic effects, mainly age and socio economic status, and local retail supply including ‘softer factors’ such as convenience and accessibility.

Representing the multidimensional and interacting geography of such influences is complex, but has nevertheless previously been illustrated as tractable through geodemographic classification (Longley et al., 2008). The advantage of such methods vis-à-vis univariate measures or scaled composite indicators (e.g. such as a measure of “deprivation”) are that geodemographics enable the summary of a wider range of influences on Internet user behaviours, and furthermore, enable greater opportunity for differentiation where influencing factors are not necessarily co-linear. For example, differentiating between areas of low engagement that are constrained by infrastructure, versus those constrained by demographics and attitudes.

Geodemographics ascribe categories that aim to summarise the salient characteristics of small areas through comparison of attributes related to resident populations, associated behaviours and features of the built environment (Harris et al., 2005). Such classifications have been applied in a variety of international settings over numerous substantive contexts (Singleton and Spielman, 2014); and is a technique commonly used in retail analysis for consumer segmentation (Birkin et al., 2002). Although general-purpose geodemographic classifications have demonstrated utility for exploration of Internet usage behaviours (Bunyan and Collins, 2013), as illustrated by Longley et al. (2008), there are sound reasons for developing purpose-specific classifications within this context. Logic follows that area differentiation is most effectively achieved through a geodemographic where the inputs are tailored to those outcomes that the classification is being designed to measure, providing enhanced performance and a stronger theoretical rationale for those attributes selected (Singleton and Longley, 2009).

As such, it is necessary to capture a composite of influences on demand, which are assembled within our e-resilience framework through creation of a purpose specific geodemographic, referred to going forward as the Internet User Classification (IUC). Guided by those past empirical studies that were highlighted in the literature presented earlier in this section, attributes of influence were organised into a typology of inputs comprising domains and more specific sub-domains. These are summarised in Table 1 and mapped onto input measures in the Appendix. These data were assembled at a 2011 Lower Layer Super Output Area (LSOA) scale for England, which comprises 32,844 zones of between 1,000 and 3,000 people, and 400 and 1,200 households. Most data were available for all of Great Britain, albeit that those datasets available for England were more robust, and so the decision was taken to exclude Scotland and Wales from the analysis.

An important source of data forming input to the IUC was the Oxford Internet Survey (OXIS), which was launched by the Oxford Internet Institute (OII)[1] in 2003, with subsequent surveys conducted biannually. Each survey implements a multi-stage national probability sample design for around 2000 people in Great Britain, enabling projection of estimates to Great Britain as a whole and comparison over time. Full details of the survey and methods can be found on the OXIS website,[2] and for this research we used responses from the latest 2013 study. A Small Area Estimation (SAE) technique was applied to each question and generated a predicted response rate at the LSOA level. This process was multi-staged, first implementing decision tree modelling for each OXIS survey

[1] http://oxis.oii.ox.ac.uk/.
[2] http://oxis.oii.ox.ac.uk/research/methodology/.

Table 1
Domains and subdomains of influence upon Internet behaviour.

Domain	Sub domain
Demographic	Age Density
Education	HE Qualifications
Employment	Occupation Access
Engagement	Attitude Business Civic Engagement Commerce Finance Finance Mobile Mobile Retail
Infrastructure	Connectivity Other Wireless/Mobile

Table 2
Structure and labels of the Internet User Classification (IUC).

Supergroup	Group
1: E-unengaged	1a: Too Old to Engage 1b: E-marginals: Not a Necessity 1c: E-marginals: Opt Out
2: E-professionals and students	2a: Next Generation users 2b: Totally Connected 2c: Students Online
3: Typical trends	3a: Uncommitted and Casual Users 3b: Young and Mobile
4: E-rural and fringe	4a: E-fringe 4b: Constrained by Infrastructure 4c: Low Density but High Connectivity

response, with covariates derived from non-response survey attributes and by geocoding the responder postcode. Predictor variables selected were based on those factors known to influence Internet use and behaviour that were identified earlier in this section, and include: age, social grade and population density quintile (as a proxy for rurality). A range of other attributes were evaluated, however these had limited influence, and as such were removed to improve model parsimony. The models output a series of rates which were then fitted to Output Areas zones (OA: minimum 40 households and 100 people) by examining the distribution of these population sub-groups within each OA nationally. In a sense, the OA rate for each question was a weighted average derived from all the population sub-groups present within it. We performed external validation of the OA estimates by profiling with a geodemographic (ONS Output Area Classification[3]) to ensure that the propensity for certain responses (e.g. the use of smartphones) were in line with responses given the general demographic profile of the clusters. Secondly, we also examined how our nation-wide and regional estimates differed from those derived through the original OXIS sample. In all instances our estimates were consistent with those from the OXIS sample, with no statistically significant differences in their distributions. Input into the classification required aggregation of the rates from OA to LSOA.

In addition to OXIS, attributes related to both fixed line and mobile Internet enabling infrastructures were assembled. Data comprising 4.7-million unit postcode level crowd sourced Internet speed test results were made available from broadbandspeedchecker.co.uk,[4] enabling average access speeds to be compiled for each LSOA. Detailed consideration of the spatio-temporal characteristics of these data can be found elsewhere in Riddlesden and Singleton (2014). Additionally, given that cellular signal strength becomes constrained as the distance from a cell tower increases (Godara, 2001), a proxy for access speed was created for each LSOA by calculating the population weighted centroid distance to the nearest phone mast using the Ofcom 'Sitefinder' database.[5] Finally, a range of socio-demographic indicators from the 2011 census was collated, including: levels of education, employment sector, prevalence of full-time students, age structure and population density. All variables considered as inputs to the classification were evaluated for their discrimination potential, and where possible, were limited to those without strong correlation; as such effects can overly influence the form of a final classification (Harris et al., 2005). Inputs were normalised using a Box–Cox transformation (Box and Cox, 1964), and were then range standardised onto the same measurement scale (Wallace and Denham, 1996). These are common types of transformation and standardisation implemented in the creation of geodemographic classifications (Vickers and Rees, 2007), and are deemed necessary to reduce the influence of skew, and additionally to ensure that all variables are ascribed equal weighing. After input measures are assembled, a geodemographic classification is created using a clustering algorithm, which is a class of computational method that aims to assign each record (in this case an LSOA) into a cluster based on similarity across the full range of input attributes.

This classification was created using the *K*-means clustering algorithm[6] with a "top-down" implementation; running the classification procedure for multiple iterations in order to attain an optimised result for each cluster frequency selected (Spielman and Singleton, 2015). This process first created an initial 'coarse' tier referred to as 'Supergroups' before re-clustering data within these assignments to form a second nested 'Group' level. Numerous different cluster frequencies were tested, with varying interpretability of the cluster characteristics and classification performance assessed. This evaluation included mapping and empirical testing of cluster fit through within sum of squares statistics. The final classification formed a nested hierarchy of 4 Supergroups and 11 Groups. Cluster mean values were then calculated for each of the input attributes and used to create 'Pen Portraits' (see Appendix) by considering variability in these scores between clusters. These textual descriptions provide an overview of the salient characteristics of each cluster, and are also summarised with Supergroup and Group names (see Table 2). An interactive map of the classification is available on the companion website: http://maps.cdrc.ac.uk/#/geodemographics/iuc14/, and can also be downloaded here: https://data.cdrc.ac.uk/dataset/cdrc-2014-iuc-geodata-pack-england.

4. Retail centre vulnerability and supply

Measuring the vulnerability of competing retail destinations to consumers of differential Internet engagement characteristics requires an understanding of the location and geographic extent of retail centres, combined with some assessment of their composition and size. A widely-accepted measure of retail area extent in the UK was developed through work funded by the Office of the Deputy Prime Minister in the late 1990 and early 2000s. This technique was later employed by the Department for Communities and

[3] http://www.opengeodemographics.com/.
[4] http://broadbandspeedchecker.co.uk.
[5] http://sitefinder.ofcom.org.uk/.

[6] *K*-means is stochastic and sensitive to those conditions used to initiate the algorithm; as such, and as is common practice in the creation of geodemographics, the algorithm was run 10,000 times, with an optimal run isolated through comparison of a within sum of squares statistic.

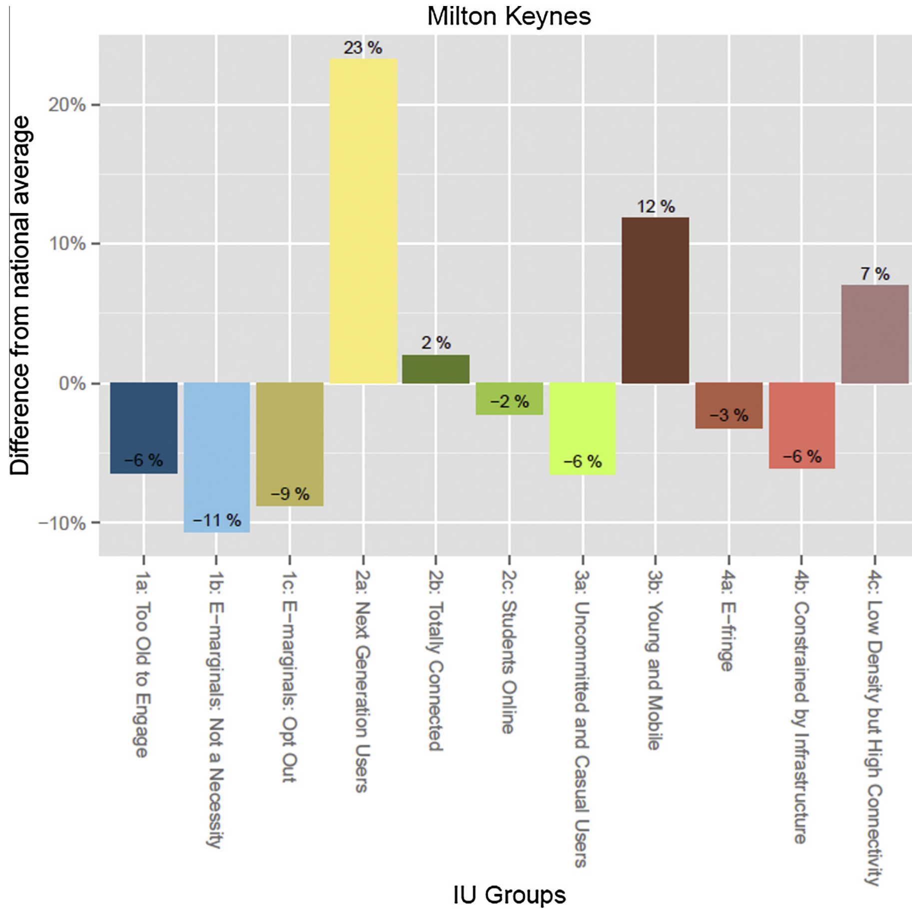


Fig. 2. IUC catchment profile for central Milton Keynes retail centre.

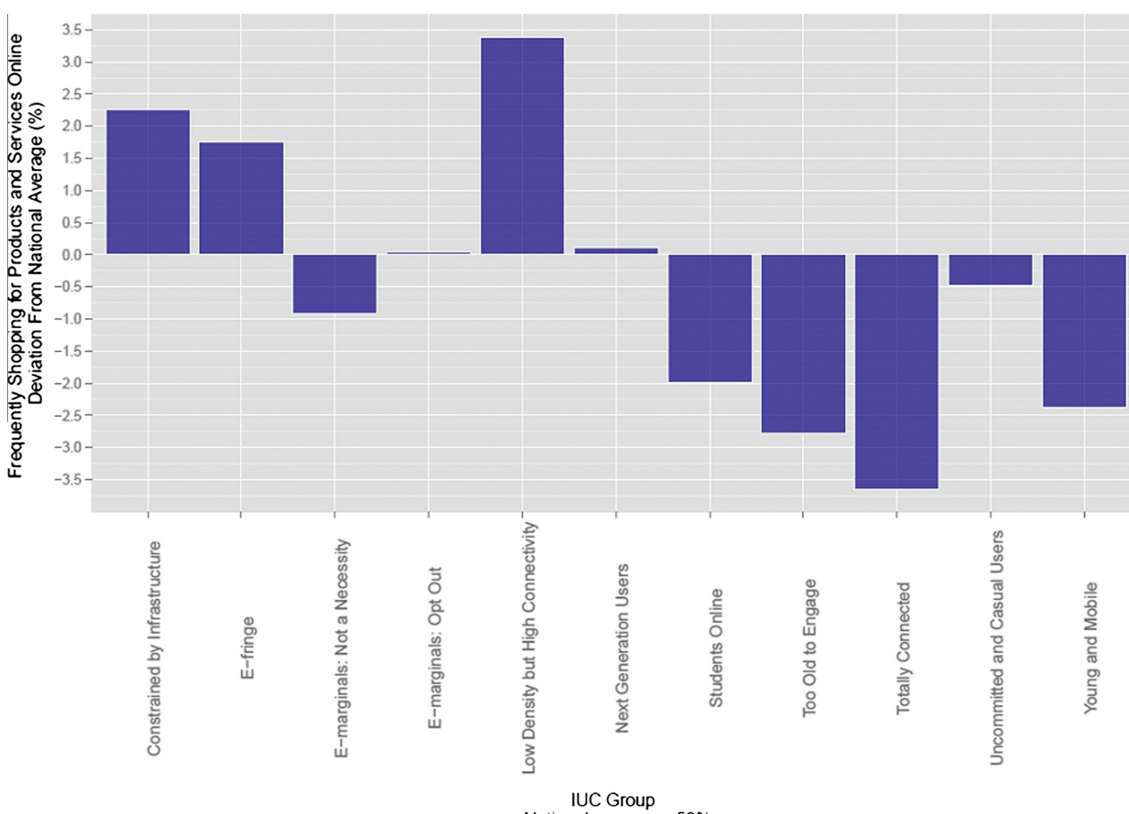


Fig. 3. Propensity to shop online by IUC Groups (OXIS Question – QC30b ‘Frequently buy products online’).

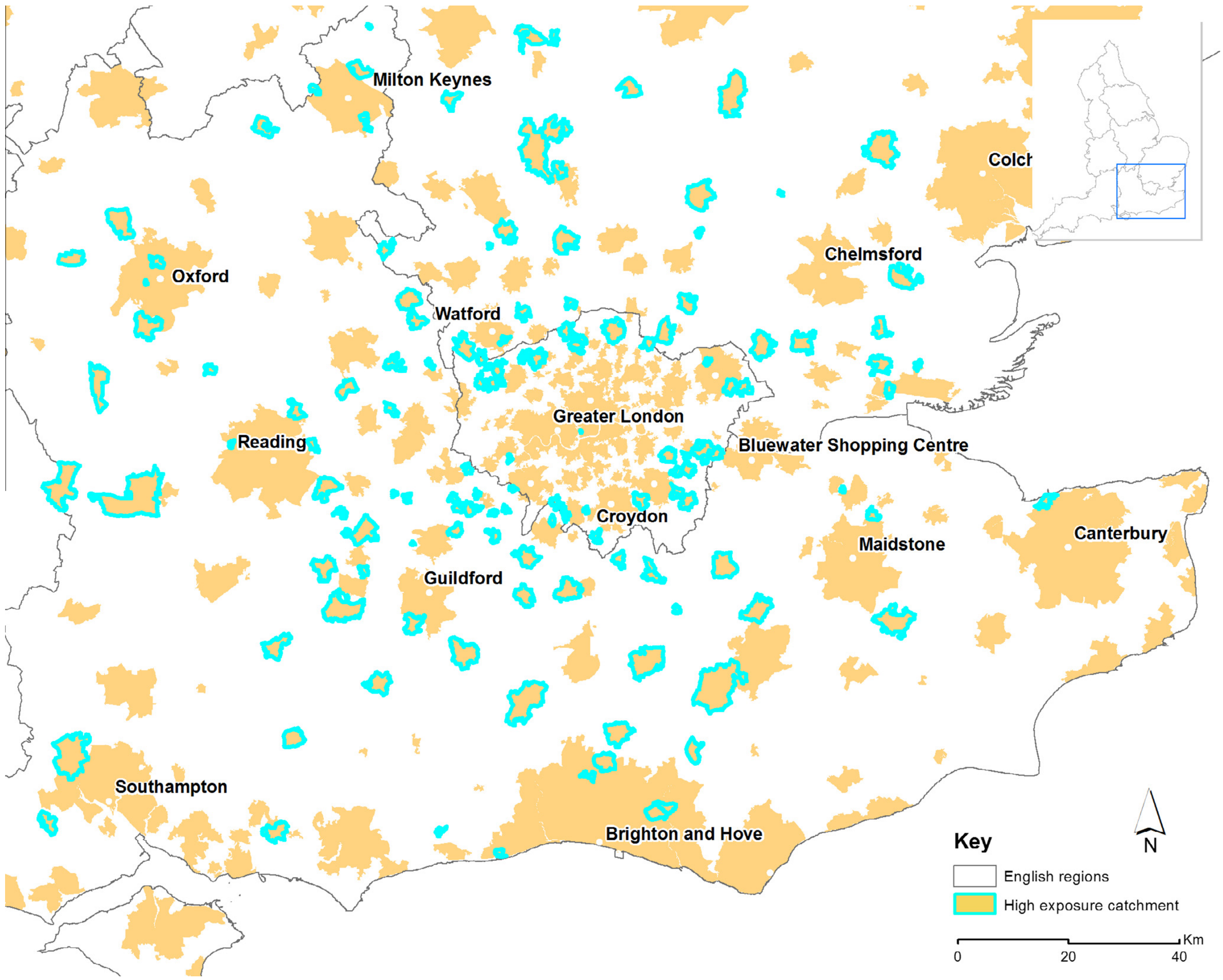


Fig. 4. Highly exposed retail centres in SE England.

Local Government (DCLG) to derive a set of retail centre zones that were used to form a database of information that featured in the State of the Cities Report (http://goo.gl/mtX1aB). These boundaries provide a systematic estimate of where the main concentrations of shops are found between different locations, although do not give information about the composition of competing retail opportunities. However, a nationally expansive record of the location, occupancy and facia of UK retail stores are generated by the Local Data Company (LDC: http://www.localdatacompany.com/), a commercial organisation that employs a large survey team to collect these data on a rolling basis. A national extract for February 2014 was made available for this research, with each record comprising the location of a retail premise with latitude and longitude coordinates, and details of the current occupier.

These data were used to calculate a series of measures which were informed by the literature to either enhance physical store attractiveness, or, to represent retail category vulnerability, where there would be risk of the main product offerings switching from physical to online channels. A composite of these measures forms a "supply vulnerability index" that is later integrated into the e-resilience score. Input measures to this index included the weighted percentage of anchor stores[7] (Feinberg et al., 2000; Damian et al., 2011) and leisure outlets (Reimers and Clulow, 2009); which are countered by the prevalence of 'digitalisation retail'. The latter measure captured the following categories: newsagents, booksellers, audio-visual rental, computer games, home entertainment, records, tapes & CDs and video libraries, as specified by the Oxford Institute of Retail Management (2013). As such, higher proportions of 'digitalisation retail' are associated with enhanced vulnerability of retail centres, whereas higher proportions of anchor store and leisure units indicated greater resilience.[8] A *supply vulnerability index* was then generated for each retail centre by creating a composite *z* score for each variable, and computing an average for each centre. The final score was scaled between 1 and 100.

5. Reconciling supply and demand

Estimating the exposure of retail centres to populations who are active Internet users as defined through the IUC required a method of modelling consumer flows to probable retail destinations. There is a long history and well developed literature on the ways in which such supply and demand for retail centres can be reconciled

[7] Anchor stores were defined as the 20 most attractive/largest stores as presented by Wrigley and Dolega (2011).

[8] Due to data availability, the percentage of stores within each retail centre rather than share of floor area was used. However such measure may be prone to a degree of bias as typically anchor stores are of larger size, therefore, this was addressed by increasing their weighting.

through catchment area estimation (Wood and Reynolds, 2012; Birkin et al., 2002, 2010). These techniques range in sophistication from calculating the geographic extent that people might be willing to travel to a retail centre in a given time (Grewal et al., 2012), through to more complex mathematical models that are calibrated on the basis of how attractive different retail offerings are to consumers living in different places (Newing et al., 2015). This latter group of models typically makes assumptions that larger towns with more compelling retail and leisure offerings are more attractive, but these effects decay with distance. Full details of the methodology and software to calibrate bespoke models lies outside the scope of this paper, and can be found in Dolega et al. (2016). However, in brief, catchments were estimated using a product constrained Huff model (Huff, 1964), with inputs including town centre composition and vacancy, and was implemented using distance decay functions calibrated using road network distance, and retail centre morphology proxied by ease of access and centre's position within retail hierarchy.

Once catchment areas had been estimated, we examined exposure through the intersection of the IUC groups presented earlier (Table 2). An example of a catchment profile for the Milton Keynes retail centre, which is within a city north of London, is shown in Fig. 2. This considers the proportion of the population within each of the IUC Groups, relative to the England average. In this example, it can be seen that the three of the eleven groups are over represented within this retail catchment, and similar profiles were calculated for all retail centres.

As discussed earlier, the IUC captures a range of influences on Internet user behaviour, however for the purposes of this analysis, those IUC groups with the highest and lowest propensity for online shopping were identified using the OXIS (see Fig. 3). Nationally, rates of online shopping equate to 53%, however there are differences between IUC Groups. For example, 4c (low density but high connectivity), 4b (constrained by infrastructure), 4a (e-fringe) and 2a (next generation users) are most likely to engage in online shopping; whereas: 3a (uncommitted and casual users), 1b (e-marginals: not a necessity) and 3b (young and mobile) have lower than average propensities. As such, the proportion of people within the overrepresented groups (4a, 4b and 4c) were calculated

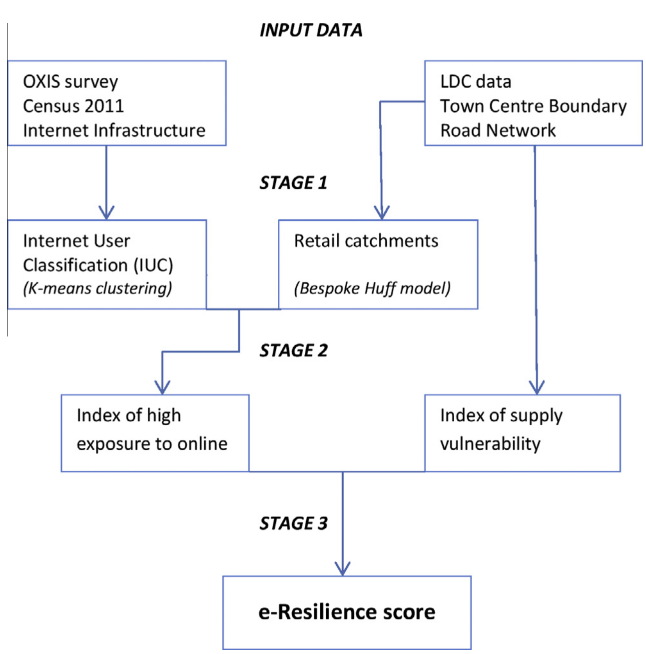


Fig. 5. Flow diagram showing how the e-resilience scores were calculated.

Table 2a
The 20 most e-resilient town centres.

Town centre	Region	e-resilience score
Boughton	East Midlands	100.00
Ravenside Retail Park, Bexhill-on-Sea	South East	97.58
Corbridge	North East	93.27
Torport	South West	71.61
Hersham	South East	70.29
Halton, Leeds	Yorkshire and the Humber	69.29
Cinderford	South West	68.51
Marsh Road, Luton	East of England	67.01
South Molton	South West	65.41
Parkgate Retail World	Yorkshire and the Humber	64.37
Carcroft	Yorkshire and the Humber	62.88
Chadderton	North West	60.74
Newburn	North East	60.45
Ventura Road Retail Park, Bitterscote	West Midlands	57.54
Feltham	Greater London	57.16
Teesside Park, Middlesbrough	North East	56.98
Kingston Park	North East	56.92
Sky Blue Way, Coventry	West Midlands	56.82
Crediton	South West	56.36
White Rose Centre	Yorkshire and the Humber	56.28

Table 2b
The 20 least e-resilient town centres.

Town centre	Region	e-resilience score
Rochford	East of England	1.00
London Road, Leigh-on-Sea	East of England	15.61
North Seaton Industrial Estate	North East	16.86
Whalley	North West	17.20
Oxted	South East	17.25
Barnt Green	West Midlands	17.39
Eccleshall	West Midlands	17.39
Hurstpierpoint	South East	17.98
Botley Road, Oxford	South East	18.14
Woburn Sands	South East	18.52
Potton	East of England	19.05
Shenfield Station	East of England	19.08
Bradford-on-Avon	South West	19.17
Great Dunmow	East of England	19.83
Longbenton	North East	20.37
Chalfont St. Peter	South East	20.75
Epworth	Yorkshire and the Humber	20.86
Sawbridgeworth	East of England	21.01
Ampthill	East of England	21.32
Old Bexley	Greater London	21.91

for each retail centre catchment, and again scaled into the range 1 and 100, forming an *index of high exposure*.

The index of high exposure indicates a rather remarkable spatial pattern. Fig. 4 maps those catchments with high exposure to online retail, defined here as possessing an index over the mean. The pattern emerging from this analysis is that predominantly secondary and tertiary retail centres (Dennis et al., 2002) located in more rural areas, including the satellite centres of more urbanised areas, reveal the greatest exposure to the impacts of online sales. This trend is reiterated for other parts of the country, although the majority of the highly exposed retail centres can be found within the South East. Moreover, based on those attractiveness scores that fed into the catchment model, it is worth noting that none of the

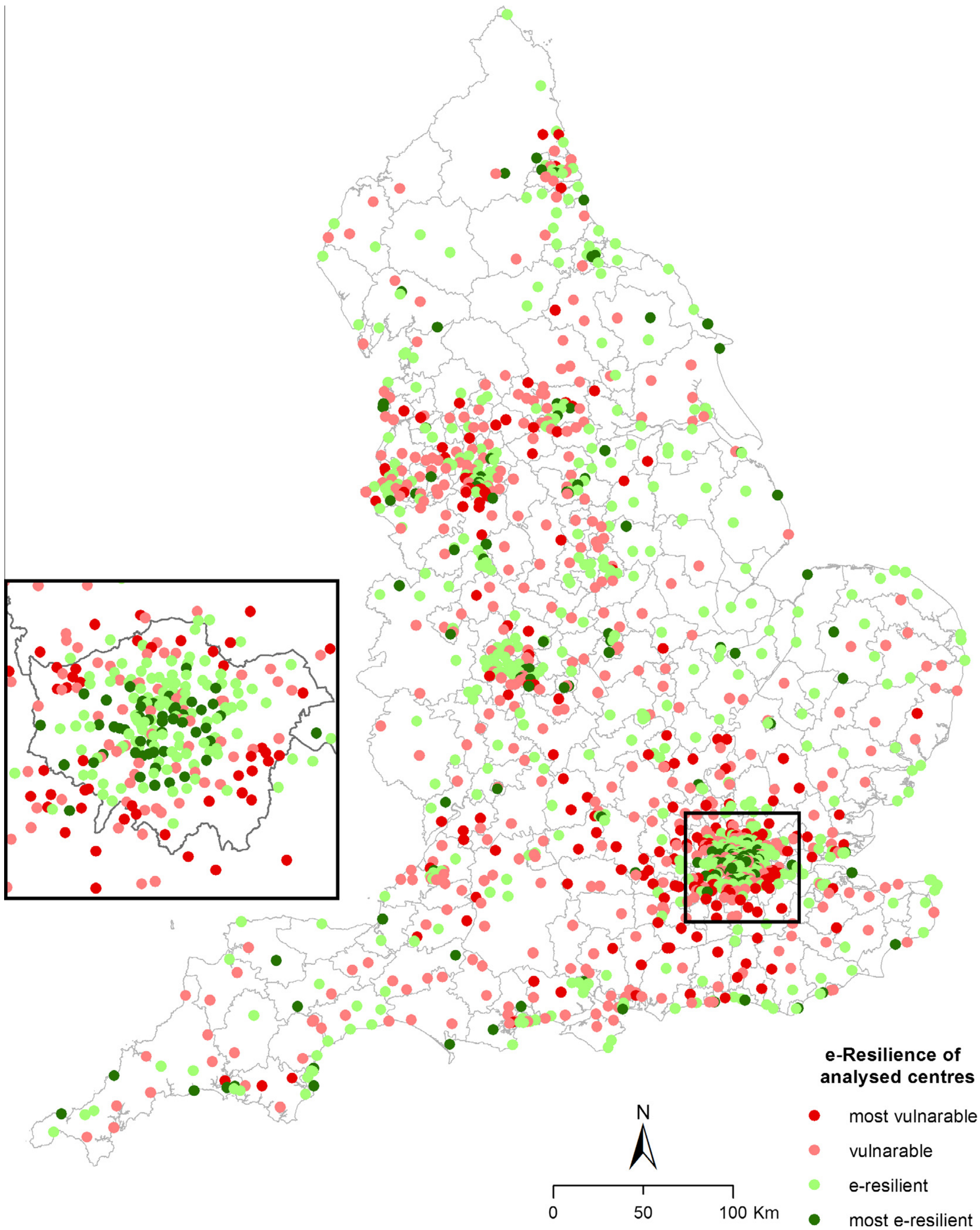


Fig. 6. The e-resilience of town centres in England.

highly exposed centres were drawn from the larger, most attractive centres, unlike the fortunes of many of the surrounding smaller towns and local shopping centres.

The index of high exposure and the supply vulnerability index were then combined to ascribe a measure of e-resilience to each individual retail centre. The indices were summed, and then the final score scaled into the range 1 and 100. The simple flow diagram in Fig. 5 shows the input datasets, different stages and outputs used to calculate the e-resilience scores.

Tables 2a and 2b show the 20 most and least e-resilient retail centres, with Fig. 6 mapping the e-resilience levels for the national extent, with scores divided into quartiles. The intersection of these two indicators reveals a remarkable spatial pattern. The most attractive retail centres, in particular the inner areas of the larger urban areas such as Greater London, Birmingham or Manchester demonstrated the highest levels of e-resilience, followed by the small local centres. Conversely, the least e-resilient centres were predominantly located in the suburban and rural areas of South East England, and to a lesser degree around other major conurbations of the country. Typically, these were the secondary and medium sized centres, often referred to as 'Clone Towns' (Ryan-Collins et al., 2010). It could be argued that this is largely intertwined with the geography of Internet shopping, where customers in more remote locations, typically faced with poorer retail provision, have displayed a higher propensity for online shopping. Nevertheless, these findings can also be associated with a polarisation effect, implying that large and attractive centres function as hubs for higher order comparison shopping and leisure; whereas the small local centres provide everyday convenience shopping, but the mid-sized centres have a less clear function. Combining such effects with higher exposure to online sales, retailers may be increasingly faced with too much physical space, and therefore may be inclined to downsize their store portfolio in such secondary locations first, as the space offered is often of a wrong size and configuration (BCSC, 2013).

6. Discussion and conclusions

The growth of Internet sales is increasingly viewed as one of the most important forces currently shaping the evolving structure of retail centres (Wrigley and Lambiri, 2014; Hart and Laing, 2014). Although current research does not suggest a death of physical space, the consequences for traditional high streets remain unclear as knowledge about the geography and drivers of Internet shopping are still limited. However, what is evident is that the pace of change in some retail centres has been more rapid than in others, and that multi-channel shopping has generated different requirements, not only in the terms of physical shopping space, but also in the expectations of an increasingly technology-driven consumer (BCSC, 2013; Kacen et al., 2013).

This study has introduced the concept of *e-resilience* as a framework through which the vulnerability of physical retail centres to the impact of online shopping behaviour can be assessed at a spatially disaggregate scale and for a national extent. Importantly, the measurement task required a trade-off between a number of challenges such as the degree of generalisation and the availability of data to inform model specification. For instance, the impact of online sales within a centre may range from damaging to some smaller retailers through to complementary in the case of various large multiples. In order to capture these complexities a number of assumptions were made such as the type of retail typically associated with the detrimental impacts and vice versa. Operationalising this measure of e-resilience required a novel methodology that conflated a range of data sources to develop two national indicators of retail centre exposure and vulnerability to online sales. These indices of supply and demand were then coupled through a retail centre catchment model.

The combined e-resilience measure revealed a geography where attractive and large retail centres such as the inner cores of large metropolitan areas were highlighted as more resilient, along with smaller more specialist centres, which perhaps served convenience shopping requirements. The centres identified as most vulnerable included many secondary and medium sized centres, which layers additional risk on top of those issues highlighted elsewhere such as a lack of diversity, or space not appropriately configured to a contemporary retail system (Ryan-Collins et al., 2010).

The findings of this study should be viewed as novel, and can be used to inform policy decisions. The three major implications of this project are as follows. First, it establishes the concept of e-resilience that examines retail centre exposure to the impact of Internet sales, and proposes a new methodology about how such interactions can be measured. Second, a comprehensive classification of retail centres based on their e-resilience levels provides a resource that can be used by a wide range of stakeholders including academics, retailers and town centre managers. For example, such outcomes could be used as assessment tools when evaluating retail centre economic performance. Third, the study adds value to and repositions the focus of current debates on the resilience of traditional high streets, which have predominantly concentrated on supply side measure such as vacancy rates.

Although there is no doubt that the concept of e-resilience, and the deliverable measures have both intuitive validity and practical application, this study is not free from limitations. The first relates to the availability of data that can be used to measure the e-resilience of retail centres. For instance, it may be difficult to capture comprehensively the vulnerability of retail centres using merely quantitative indicators; and in particular, those softer experiential factors may not be well reflected. The second is the extent to which supply and demand factors influence the e-resilience scores. The measure of e-resilience was calculated by simply adding the indexes of high exposure and supply vulnerability together. This presupposes that each index (that captures demand and supply respectively) is equally weighted and hence supply and demand has equal importance in terms of measuring e-resilience. These arbitrary equal weights for demand and supply might be validated against data on changing retail centre fortunes as these become available. A potential route to such validation may be sourced through the pooling of consumer data related to de facto online and offline consumption patterns. Finally, by examining e-resilience at a centre level, we have imposed a degree of generalisation in terms of composite retailer function, configuration and ownership. Further research is also required to explore how individual retailers respond to variable levels of exposure to consumers with differing Internet consumption characteristics. Such measures might be refined by sourcing retail floorspace estimates as a substitute for number of retail units and the arbitrary weighting of anchor stores. Opening hours and the range and availability of leisure facilities are also of clear importance in establishing the competitiveness of retail centres in comparison with online offers.

The concept of e-resilience contributes considerably to our current understanding of the nature and impact that Internet user behaviour is having on retail centres within the UK. International comparisons are clearly a fertile area for future research – for example some technologically advanced nations, such as South Korea, report lower levels of Internet sales than the U.K. As the penetration of online consumption is still steadily increasing, operational tools such as those offered by this study will have increasing policy relevance.

Acknowledgments

This research was funded by the Economic and Social Research Council, grant numbers ES/L011840/1; ES/L003546/1. We gratefully acknowledge the helpful and constructive comments of three anonymous reviewers.

Appendix A. IUC input measures

Measure	Sub domain	Domain	Source
Persons aged 0–4	Age	Demographic	Census
Persons aged 5–7	Age	Demographic	Census
Persons aged 8–9	Age	Demographic	Census
Persons aged 10–15	Age	Demographic	Census

(*continued on next page*)

Appendix A (*continued*)

Measure	Sub domain	Domain	Source
Persons aged 16–17	Age	Demographic	Census
Persons aged 18–19	Age	Demographic	Census
Persons aged 20–24	Age	Demographic	Census
Persons aged 25–29	Age	Demographic	Census
Persons aged 30–44	Age	Demographic	Census
Persons aged 45–59	Age	Demographic	Census
Persons aged 60–64	Age	Demographic	Census
Persons aged 65–74	Age	Demographic	Census
Persons aged 75–84	Age	Demographic	Census
Persons aged 85–89	Age	Demographic	Census
Persons aged 90 plus	Age	Demographic	Census
Population density persons per hectare	Density	Demographic	Census
Persons with no qualifications	Qualifications	Education	Census
Persons with level one qualifications	Qualifications	Education	Census
Persons with level two qualifications	Qualifications	Education	Census
Persons who are apprentices	Qualifications	Education	Census
Persons with level three qualifications	Qualifications	Education	Census
Persons with level four qualifications	Qualifications	Education	Census
Persons with other qualifications	Qualifications	Education	Census
Full time students	HE	Education	Census
Managers, directors and senior officials	Occupation	Employment	Census
Professional occupations	Occupation	Employment	Census
Associate professional and technical occupations	Occupation	Employment	Census
Administrative and secretarial occupations	Occupation	Employment	Census
Skilled trades occupations	Occupation	Employment	Census
Caring, leisure and other service occupations	Occupation	Employment	Census
Sales and customer service occupations	Occupation	Employment	Census
Process, plant and machine operatives	Occupation	Employment	Census
Elementary occupations	Occupation	Employment	Census
Seeking info holiday/journey – Internet	Commerce	Engagement	OXIS
Seeking info holiday/journey – Smartphone	Commerce	Engagement	OXIS
Frequently compare prices online	Commerce	Engagement	OXIS
Frequently order food or groceries online	Commerce	Engagement	OXIS
Frequently sell things online	Commerce	Engagement	OXIS
Seeking info topic/professional project – Internet	Business	Engagement	OXIS
Seeking info topic/professional project – Smartphone	Business	Engagement	OXIS
Have found a job through the Internet	Business	Engagement	OXIS
Frequently pay bills online	Finance	Engagement	OXIS
Frequently use online banking	Finance	Engagement	OXIS
Use mobile phone for email	Mobile	Engagement	OXIS
Use mobile for posting videos and photos online	Mobile	Engagement	OXIS
Use mobile phone for navigation	Mobile	Engagement	OXIS
Use mobile phone for social networking	Mobile	Engagement	OXIS
Use mobile phone for apps	Mobile	Engagement	OXIS
Use mobile phone for browsing the Internet	Mobile	Engagement	OXIS
Internet important for information	Attitude	Engagement	OXIS
Internet important for entertainment	Attitude	Engagement	OXIS
Interested in the Internet	Attitude	Engagement	OXIS
Use Internet while travelling – mobile/dongle	Mobile	Engagement	OXIS
Mostly use mobile phone for Internet	Mobile	Engagement	OXIS
Have saved money buying online	Finance	Engagement	OXIS
Frequently buy products online	Retail	Engagement	OXIS
Current Internet users	Access	Engagement	OXIS
Ex Internet users	Access	Engagement	OXIS
Internet non users	Access	Engagement	OXIS
Seeking info MP – Internet	Civic Engagement	Engagement	OXIS
Seeking info MP – Smartphone	Civic Engagement	Engagement	OXIS
Seeking info council tax – Internet	Civic Engagement	Engagement	OXIS
Seeking info council tax – Smartphone	Civic Engagement	Engagement	OXIS
Households that have Internet access at present	Connectivity	Infrastructure	OXIS
Households that don't have Internet access but have had in past	Connectivity	Infrastructure	OXIS

Appendix A (*continued*)

Measure	Sub domain	Domain	Source
Households that have never had Internet access	Connectivity	Infrastructure	OXIS
Households that have had Internet access for ten years or more	Connectivity	Infrastructure	OXIS
Households with wireless access in home through Wi-Fi	Wireless/Mobile	Infrastructure	OXIS
Households with a tablet computer	Wireless/Mobile	Infrastructure	OXIS
Households with e reader	Other	Infrastructure	OXIS
Households with games console	Other	Infrastructure	OXIS
Households with a smart TV	Other	Infrastructure	OXIS
Mobile phone ownership	Wireless/Mobile	Infrastructure	OXIS
Local download speed	Connectivity	Infrastructure	Broadbandspeedchecker
Distance to closest mobile base station	Connectivity	Infrastructure	Ofcom

Appendix B. IUC Group Pen Portraits

Group 1a: Too Old to Engage

The Too Old to Engage Group is characterised by large elderly populations who show little or no engagement with the Internet across all applications. The proportion of residents aged 75 plus is higher than any Group in the IUC. As a result, Internet enabled device ownership is lower than the Supergroup average, and the lowest of any Group in the IUC. Abstinence from Internet use is higher than the Supergroup average and far above the national average. Enclaves of this Group are found in coastal and lower density rural areas that serve as retirement destinations. Infrastructure provision and performance is typically slightly below the national average. The Too Old to Engage Group accounts for 4% of all Lower Super Output Areas nationally.

Group 1b: E-marginals: Not a Necessity

Members of the E-marginals: Not a Necessity Group typically have low engagement with Internet applications, lower than average qualifications and higher than average rates of employment in blue collar occupations that are not heavily reliant on digital skills. Of those that do access the Internet, many do so using a smartphone. Residents of this Group tend to be found within urban areas characterised by high levels of material deprivation, although infrastructure provision and performance are in line with the national average. The E-marginals: Not a Necessity Group accounts for 10.4% of all Lower Super Output Areas nationally.

Group 1c: E-marginals: Opt Out

The E-marginals: Opt Out Group are characterised by low levels of engagement with the Internet for applications such as seeking information and entertainment, preferring instead more traditional media such as newspapers and television, in part reflecting the elderly demographic of this Group. Typically residents of this Group are aged 60 plus, with significantly higher than average incidence of those aged 65–84. Geographically, this Group tends to be found in affluent rural and fringe areas that are more sparsely populated and where infrastructure provision and performance is below the national average. Access to the Internet through mobile devices is below the national average. Those who do choose to use the Internet tend to use it for price comparison and occasional online shopping. Levels of qualifications are generally above the national average, and those members who are not retired will typically be employed in senior managerial, professional or skilled trade occupations. Abstinence is significantly higher than the national average, but the lowest within the Supergroup. The E-marginals: Opt Out Group accounts for 10.4% of all Lower Super Output Areas nationally.

Group 2a: Next Generation Users

The Next Generation Users Group is characterised by high levels of engagement across all applications of the Internet. Members of this Group are heavy smartphone users and typically access the Internet on the move and for applications such as email, social networking and navigation. However, they favour fixed line connections for most other tasks such as general browsing and seeking information. Device ownership is higher than the national average, and members of this Group are likely to own several Internet enabled devices, such as tablet computers, e-readers and smart TVs. Levels of qualification are high within this Group, with higher than average rates of degree and higher degree level qualifications. The age structure is young to middle aged, with members of this Group most likely aged between 25 and 44, and in some cases with young children. Employment tends to be in managerial, professional and technical occupations. General interest in the Internet is above the national average. Members of this Group are found in affluent, higher density suburban and city fringe areas where infrastructure provision and performance is above the national average. Next Generation Users are the second most heavily engaged Group within the IUC, behind Group 2b: Totally Connected and account for 10.2% of all Lower Super Output Areas nationally.

Group 2b: Totally Connected

The Totally Connected Group is characterised by the highest levels of engagement within the IUC and score higher than the Supergroup and national averages for most measures of engagement. This Group displays a clear preference to use the Internet by default for almost all applications. Members of this Group access the Internet through multiple devices, whilst on the move and in the home to ensure seamless connectivity. As such, device ownership is significantly higher than the national and Supergroup averages and members of this Group own a wide range of Internet enabled hardware. Levels of qualification are significantly higher than the national average. Professional occupations are most prevalent, with the age structure of residents being young to middle aged, sometimes with young children. Geographically, this Group tends to be found in affluent city centre and city fringe areas that are densely populated and where infrastructure provision and performance is above the national average. Members of this Group show below average rates of online shopping, perhaps given good local retail choice. However, rates of online shopping for food and groceries are significantly above the national and Supergroup averages as this enables wider choice and convenience in highly populated areas. Totally Connected are the most heavily engaged Group within the IUC and account for 4.8% of all Lower Super Output Areas nationally.

Group 2c: Students Online

Students Online represents a small but very distinct Group that is comprised almost entirely of student areas. The Group is characterised by very high levels of Internet usage, particularly through mobile devices. Smartphones are the device of choice for electronic communication and are used for a wide range of applications including email, social networking, third party applications, web browsing and sharing photos and videos. Members of this Group are typically aged between 18 and 24 and are registered as full time students. Interest in the Internet for information and entertainment is above the national average, and a higher than average proportion of the local population is likely to have found, or to be seeking, employment through the Internet. Employment across all sectors is below the national average with the exception of sales and customer service roles, in which some students choose to work, most likely on a part-time basis to support their studies. Geographically, this Group is often found in the major urban conurbations, usually within city centres and university campus areas where there are highly concentrated student populations. Infrastructure provision and connection performance is above the national average in these areas. The Students Online Group accounts for 1.7% of all Lower Super Output Areas nationally.

Group 3a: Uncommitted and Casual Users

The Uncommitted and Casual Users Group are characterised by mixed levels of engagement with the Internet. Access to the Internet through smartphones is marginally above the national average and access through fixed-line connections falls marginally below. Members of this Group show below average rates for purchasing online but above average rates for price comparison and selling online. Age structure is generally young to middle aged, with higher than average proportions of young and teenage children. Qualifications tend to be of a lower level and members of this Group are most likely to work in service, sales and elementary occupations. Overall, abstinence from Internet use is marginally higher than the national average and general interest in the Internet falls below the national average. This Group also contains higher than average numbers of lapsed Internet users. Geographically, this Group tends to be found in major urban and city fringe areas that suffer higher levels of material deprivation, but where infrastructure provision and performance is above the national average. The Uncommitted and Casual Users Group accounts for 15.5% of all Lower Super Output Areas nationally.

Group 3b: Young and Mobile

The Young and Mobile Group is predominantly young and has a tendency to access the Internet using mobile devices rather than fixed line connections. This Group is found in major urban conurbations where population density is above average and infrastructure provision is sufficient to support heavy mobile broadband usage. These areas are typically inner city or city fringe and experience mixed levels of material deprivation. As a Group there are higher than average proportions of young and teenage children and adults aged 25–44. Conversely, the proportion of adults aged over 45 falls below the national and Supergroup averages. All levels of qualification are below the national average and those who work are likely to be employed in elementary, sales or service occupations. Interest in the Internet for entertainment and information is above the national average, most likely reflecting the prevailing age structure. This Group displays a lower than average tendency to purchase online, and would be expected to shop locally in most cases. The Young and Mobile Group accounts for 11.5% of all Lower Super Output Areas nationally.

Group 4a: E-fringe

The E-fringe Group is distinguished by its location around the fringes of urban areas that are typically low density or semi-rural. Age structure is middle aged to elderly and there are fewer than average numbers of young adults aged 18–29, a group who are likely to have moved to more major urban conurbations. General interest in the Internet within this Group is slightly below the national average and the lowest within the Supergroup, rates of current Internet users are also below average and numbers of Internet non-users are above the national average. Members of this Group generally have mixed levels of qualifications and are most likely to work in administrative and secretarial or skilled trade occupations. The most common uses of the Internet within this Group are paying bills and banking online, comparing prices and buying products, which score above the national average. Below average rates are recorded for seeking information and entertainment purposes, consistent with the age profile of this Group. Equally, ownership of Internet enabled devices is below average, with the exception of e-readers, which are popular amongst this Group. Infrastructure provision and performance is marginally below the national average but would be unlikely to limit access. The E-fringe Group accounts for 11.1% of all Lower Super Output Areas nationally.

Group 4b: Constrained by Infrastructure

The Constrained by Infrastructure Group is characterised by locations in low-density rural areas where there is poor provision and performance of local Internet infrastructure, both fixed line and mobile. This limits engagement with some online applications. Fixed line broadband performance falls significantly below the national average and is the lowest within the Supergroup as distances to local telephone exchanges are much higher. Distances to the nearest mobile base station for cellular and data coverage are also higher than the national average, and as such further constrains performance and usability. Perhaps as a result, the use of mobile broadband through devices such as smartphones or dongles is below average. Despite poor infrastructure, general interest in the Internet is in line with the national average and members of this Group display above average rates of purchasing online, comparing prices, online banking and paying bills, most likely as this saves travelling to a local retail centre to access these services. Internet enabled device ownership is again lower than the national average with the exception of e-readers, likely due to the prevailing age structure of this Group, which is middle aged and elderly. Those who are not retired are generally highly qualified and work in managerial, professional or technical occupations. Internet non-use is above average but reflects the prevailing age profile of the Group. The Constrained by Infrastructure Group accounts for 11% of all Lower Super Output Areas nationally.

Group 4c: Low Density but High Connectivity

The Low Density but High Connectivity Group is found in areas that are sparsely populated, typically rural and semi-rural areas, or areas with urban parkland. Despite disparate populations, this Group is generally well connected and displays the strongest infrastructure and performance characteristics within the Supergroup, generally falling in line with the national average. Internet use is higher across all applications than the Supergroup average, and this Group shows a higher than average propensity for ordering food and groceries online. These characteristics are representative of the prevailing demographic of well-educated workers (often with degrees or higher degrees) who work in high-grade professional occupations. Similarly, Internet enabled device ownership is above the national average, perhaps because local infrastructure is able to support this. Age structure is mixed, although members of this Group are most likely to be aged 45–59 with young or teenage children. General interest in the Internet is above the national average and is the highest within the Supergroup. As would be expected, rates of Internet non-use are below the national average. The Low Density but High Connectivity Group accounts for 9.4% of all Lower Super Output Areas nationally.

References

BCSC, 2013. Beyond Retail: Redefining the Shape and Purpose of Town Centres. BCSC, London. <http://policy.bcsc.org.uk/beyondretail/docs/BeyondRetail2013.pdf> (accessed 05.08.15).

Birkin, M., Clarke, G., Clarke, M., 2002. Retail Geography and Intelligent Network Planning. Wiley, Chichester, NY.

Birkin, M., Clarke, G., Clarke, M., 2010. Refining and operationalizing entropy-maximizing models for business applications. Geogr. Anal. 42 (4), 422–445.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. R. Stat. Soc.: Ser. B (Methodol.) 26 (2), 211–252.

Bunyan, S., Collins, A., 2013. Digital exclusion despite digital accessibility: empirical evidence from an English city: digital exclusion despite digital accessibility. Tijdschr. Econ. Soc. Geogr. 104 (5), 588–603.

Capgemini, 2015. Mobile Accounts for 40% of All Online Retail Sales. <https://www.uk.capgemini.com/news/uk-news/mobile-accounts-for-40-of-all-online-retail-sales> (accessed 05.08.15).

Centre for Retail Research, 2015. Online Retailing: Britain and Europe 2014. <http://www.retailresearch.org/onlineretailing.php> (accessed 21.03.15).

Christopherson, S., Michie, J., Tyler, P., 2010. Regional resilience: theoretical and empirical perspectives. Cambridge J. Reg. Econ. Soc. 3 (1), 3–10.

Dennis, C., Marsland, D., Cockett, T., 2002. Central place practice: shopping centre attractiveness measures hinterland boundaries and the UK retail hierarchy. J. Retail. Consum. Serv. 9 (4), 185–199.

Damian, D.S., Curto, J.D., Pinto, J.C., 2011. The impact of anchor stores on the performance of shopping centres: the case of Sonae Sierra. Int. J. Retail Distrib. Manage. 39 (6), 456–475.

Dawley, S., Pike, A., Tomaney, J., 2010. Towards the resilient region? Local Econ. 25, 650–667.

DCLG, 2013. The Future of High Streets. Progress since the Portas Review. <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/211536/Future_of_High_Street_-_Progress_Since_the_Portas_Review_-revised.pdf> (accessed 28.10.15).

Digital High Street Advisory Board, 2015. Digital High Street 2020. <http://thegreatbritishhighstreet.co.uk/digital-high-street-report-2020> (accessed 21.03.15).

Doherty, N.F., Ellis-Chadwick, F., 2010. Internet retailing: the past the present and the future. Int. J. Retail Distrib. Manage. 38 (11/12), 943–965.

Dolega, L., Celinska-Janowicz, D., 2015. Retail resilience: a theoretical framework for understanding town centre dynamics. Stud. Region. Lokalne – Polish Sect. Reg. Stud. Assoc. 2 (60), 8–31.

Dolega, L., Pavlis, M., Singleton, A., 2016. Estimating attractiveness, hierarchy and catchment area extents for a national set of retail centre agglomerations. J. Retail. Consum. Serv. 28, 78–90.

Forman, C., Ghose, A., Goldfarb, A., 2008. Competition between local and electronic markets: how the benefit of buying online depends on where you live. Manage. Sci. 55 (1), 47–57.

Feinberg, R., Stanton, J., Keen, C., Kim, I.-S., Hokama, L., de Ruyter, K., 2000. Attraction as a determinant of mall choice. In: Proceedings of the 7th International Conference on Retailing and Services Science, European Institute of Retailing and Services Science, Eindhoven.

Grewal, D., Kopalle, P., Marmorstein, H., Roggeveen, A.L., 2012. Does travel time to stores matter? The role of merchandise availability. J. Retail. 88 (3), 437–444.

Grubesic, T.H., Murray, A.T., 2002. Constructing the divide: spatial disparities in broadband access. Pap. Reg. Sci. 81 (2), 197–221.

Godara, L.C., 2001. Handbook of Antennas in Wireless Communications. CRC Press, Boca Raton, FL.

Harris, R., Sleight, P., Webber, R., 2005. Geodemographics GIS and Neighbourhood Targeting. Wiley, Chichester, UK.

Hart, C., Laing, A., 2014. The consumer journey through the high street in the digital area. In: Wrigley, N., Brookes, E. (Eds.), Evolving High Streets: Resilience and Reinvention – Perspectives from Social Science. University of Southampton, Southampton, UK, pp. 36–39, <http://thegreatbritishhighstreet.co.uk/pdf/GBHS-HighStreetReport.pdf> (accessed 05.08.15).

Hassink, R., 2010. Regional resilience: a promising concept to explain differences in regional economic adaptability? Cambridge J. Reg. Econ. Soc. 3 (1), 45–58.

Helsper, E.J., Eynon, R., 2010. Digital natives: where is the evidence? Brit. Educ. Res. J. 36 (3), 503–520.

Huff, D.L., 1964. Defining and estimating a trade area. J. Market. 28 (3), 34–38.

Kacen, J.J., Hess, J.D., Kevin Chiang, W.-Y., 2013. Bricks or clicks? Consumer attitudes toward traditional stores and online stores. Glob. Econ. Manage. Rev. 18 (1), 12–21.

Longley, P.A., Singleton, A.D., 2009. Classification through consultation: public views of the geography of the E-society. Int. J. Geogr. Inform. Sci. 23 (6), 737–763.

Longley, P.A., Webber, R.J., Chao, L., 2008. The UK geography of the e-society: a national classification. Environ. Plan. A 40 (2), 362–382.

Maillé, P., Tuffin, B., 2010. Price war in heterogeneous wireless networks. Comput. Netw. 54 (13), 2281–2292.

Martin, R.L., 2011. Regional economic resilience hysteresis and recessionary shocks. J. Econ. Geogr. 12 (1), 1–32.

Martin, R.L., Sunley, P.J., 2007. Complexity thinking and evolutionary economic geography. J. Econ. Geogr. 7 (5), 573–602.

Newing, A., Clarke, G.P., Clarke, M., 2015. Developing and applying a disaggregated retail location model with extended retail demand estimations. Geogr. Anal. 47 (3), 219–239.

ONS, 2014. Internet Access – Households and Individuals 2014. Office for National Statistics, London. <http://www.ons.gov.uk/ons/rel/rdit2/internet-access—households-and-individuals/2014/stb-ia-2014.html> (accessed 05.08.15).

Oxford Institute of Retail Management, 2013. Diversity and the UK's High Streets. <http://oxford-institute.sbsblogs.co.uk/2013/07/19/diversity-and-the-uks-high-streets/> (accessed 06.08.15).

Parker, B., Weber, R., 2013. Second-hand spaces: restructuring retail geographies in an era of E-commerce. Urban Geogr. 34 (8), 1096–1118.

Pendall, R., Foster, K.A., Cowell, M., 2010. Resilience and regions: building understanding of the metaphor. Cambridge J. Reg. Econ. Soc. 3 (1), 71–84.

Portas, M., 2011. The Portas Review. An Independent Review into the Future of Our High Streets. <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6292/2081646.pdf> (accessed 15.06.15).

Prieger, J.E., Hu, W.-M., 2008. The broadband digital divide and the nexus of race competition and quality. Inf. Econ. Policy 20 (2), 150–167.

Raco, M., Street, E., 2012. Resilience planning economic change and the politics of post-recession development in London and Hong Kong. Urban Stud. 49 (5), 1065–1087.

Reimers, V., Clulow, V., 2009. Retail centres: it's time to make them convenient. Int. J. Retail Distrib. Manage. 37 (7), 541–562.

Rice, R.E., Katz, J.E., 2003. Comparing Internet and mobile phone usage: digital divides of usage adoption and dropouts. Telecommun. Policy 27 (8–9), 597–623.

Riddlesden, D., Singleton, A.D., 2014. Broadband speed equity: a new digital divide? Appl. Geogr. 52, 25–33.

Ryan, S., Been, C., 2013. Bricks-and-mortar: bookselling and supermarket philosophy. Int. J. Book 10 (2), 55–66.

Ryan-Collins, J., Cox, E., Potts, R., Squires, P., 2010. Re-imagining the High Street Escape from Clone Town Britain. New Economics Foundation, London.

Shobeiri, S., Mazaheri, E., Laroche, M., 2015. Shopping online for goods vs services: where do experiential features help more? Int. J. Consum. Stud. 39 (2), 172–179.

Silver, M.P., 2014. Socio-economic status over the lifecourse and Internet use in older adulthood. Age. Soc. 34, 1019–1034.

Simmie, J., Martin, R.L., 2010. The economic resilience of regions: towards an evolutionary approach. Cambridge J. Reg. Econ. Soc. 3 (1), 27–43.

Singleton, A.D., Longley, P.A., 2009. Creating open source geodemographics – refining a national classification of census output areas for applications in higher education. Pap. Reg. Sci. 88 (3), 643–666.

Singleton, A.D., Spielman, S.E., 2014. The past present and future of geodemographic research in the United States and United Kingdom. Prof. Geogr. 66 (4), 558–567.

Spielman, S.E., Singleton, A.D., 2015. Studying neighborhoods using uncertain data from the American community survey: a contextual approach. Ann. Assoc. Am. Geogr. 105 (5), 1003–1025.

Turner, J.J., Gardner, T., 2014. Critical reflections on the decline of the UK high street: exploratory conceptual research into the role of the service encounter. In: Musso, F., Drucia, E. (Eds.), Handbook of Research on Retailer–Consumer Relationship Development. IGI Global, Hershey, PA, pp. 127–151.

van Deursen, A., van Dijk, J., 2011. Internet skills and the digital divide. New Media Soc. 13 (6), 893–911.

Vickers, D., Rees, P., 2007. Creating the UK national statistics 2001 output area classification. J. R. Stat. Soc. Ser. A: Stat. Soc. 170 (2), 379–403.

Wallace, M., Denham, C., 1996. The ONS Classification of Local and Health Authorities of Great Britain. Studies on Medical and Population Subjects ONS Number 59.

Warf, B., 2013. Contemporary digital divides in the United States. Tijdschr. Econ. Soc. Geogr. 104 (1), 1–17.

Warren, M., 2007. The digital vicious cycle: links between social disadvantage and digital exclusion in rural areas. Telecommun. Policy 31 (6–7), 374–388.

Weltevreden, J., 2007. Substitution or complementarity? How the Internet changes city centre shopping. J. Retail. Consum. Serv. 14 (3), 192–207.

Weltevreden, J., 2014. The digital challenge for the high street: insights from Europe. In: Wrigley, N., Brookes, E. (Eds.), Evolving High Streets: Resilience and

Reinvention – Perspectives from Social Science. University of Southampton, pp. 32–35.
Williams, D.E., 2009. The evolution of E-tailing. Int. Rev. Retail Distrib. Consum. Res. 19 (3), 219–249.
Wilson, K.R., Wallin, J.S., Reiser, C., 2003. Social stratification and the digital divide social science. Comput. Rev. 21 (2), 133–143.
Wood, S., Reynolds, J., 2012. Leveraging locational insights within retail store development? Assessing the use of location planners' knowledge in retail marketing. Geoforum 43 (6), 1076–1087.
Wrigley, N., Dolega, L., 2011. Resilience, fragility and adaptation: new evidence on the performance of UK high streets during global economic crisis and its policy implications. Environ. Plan. A 43 (10), 2337–2363.
Wrigley, N., Lambiri, D., 2014. High Street Performance and Evolution: A Brief Guide to the Evidence. University of Southampton, Southampton, <http://thegreatbritishhighstreet.co.uk/pdf/GBHS-HighStreetReport.pdf> (accessed 05.08.15).
Wrigley, N., Lambiri, D., 2015. British High Streets: from Crisis to Recovery? A comprehensive review of the evidence. RIBEN. Available from: <http://www.riben.org.uk/Cluster_publications_%26_media/BRITISH%20HIGH%20STREETS_MARCH2015.pdf>.
Zentner, A., 2008. Online sales, Internet use, file sharing and the decline of retail music specialty stores. Inf. Econ. Policy 20 (3), 288–300.
Zentner, A., Smith, M., Kaya, C., 2013. How video rental patterns change as consumers move online. Manage. Sci. 59 (11), 2622–2634.

Appendix C

IUC User Guide

Internet User Classification (IUC)
User Guide

Dean Riddlesden
University of Liverpool
email: d.riddlesden@liverpool.ac.uk

October 2014

Contents

1 Introduction **3**

2 Pen Portraits: Supergroup Level **5**
2.1 Supergroup 1: E-unengaged 5
2.2 Supergroup 2: E-professionals and Students 6
2.3 Supergroup 3: Typical Trends 7
2.4 Supergroup 4: E-rural and Fringe 8

3 Pen Portraits: Group Level **9**
3.1 Group 1a: Too Old to Engage 9
3.2 Group 1b: E-marginals: Not a Necessity 10
3.3 Group 1c: E-marginals: Opt Out 11
3.4 Group 2a: Next Generation Users 12
3.5 Group 2b: Totally Connected 13
3.6 Group 2c: Students Online 14
3.7 Group 3a: Uncommitted and Casual Users 15
3.8 Group 3b: Young and Mobile 16
3.9 Group 4a: E-fringe 17
3.10 Group 4b: Constrained by Infrastructure 18
3.11 Group 4c: Low Density but High Connectivity 19

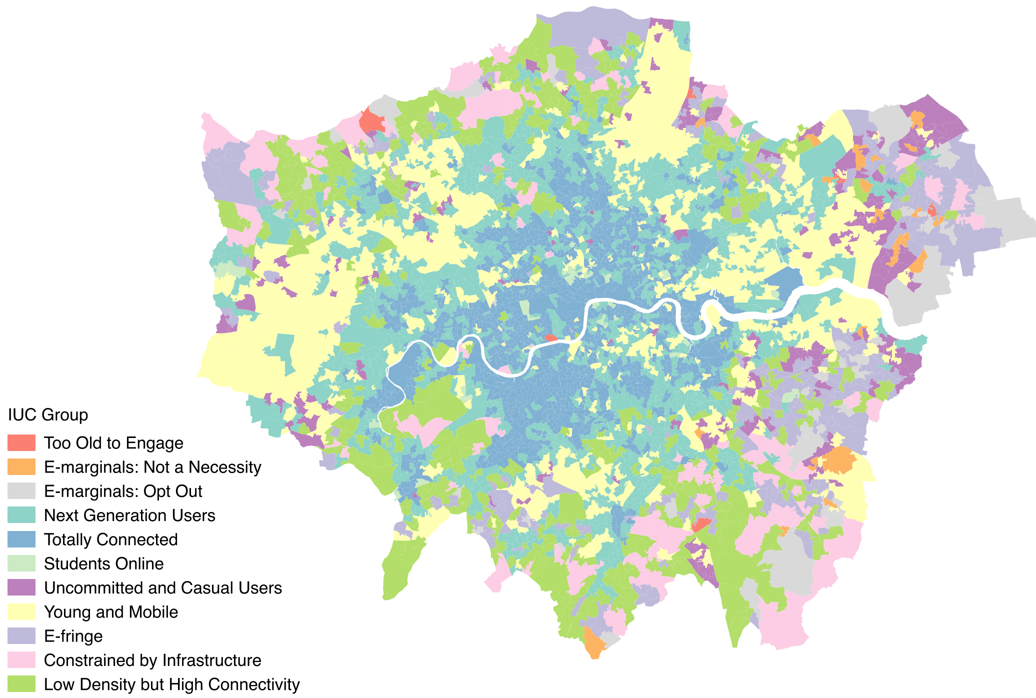
4 Queries and Further Information **20**

1 Introduction

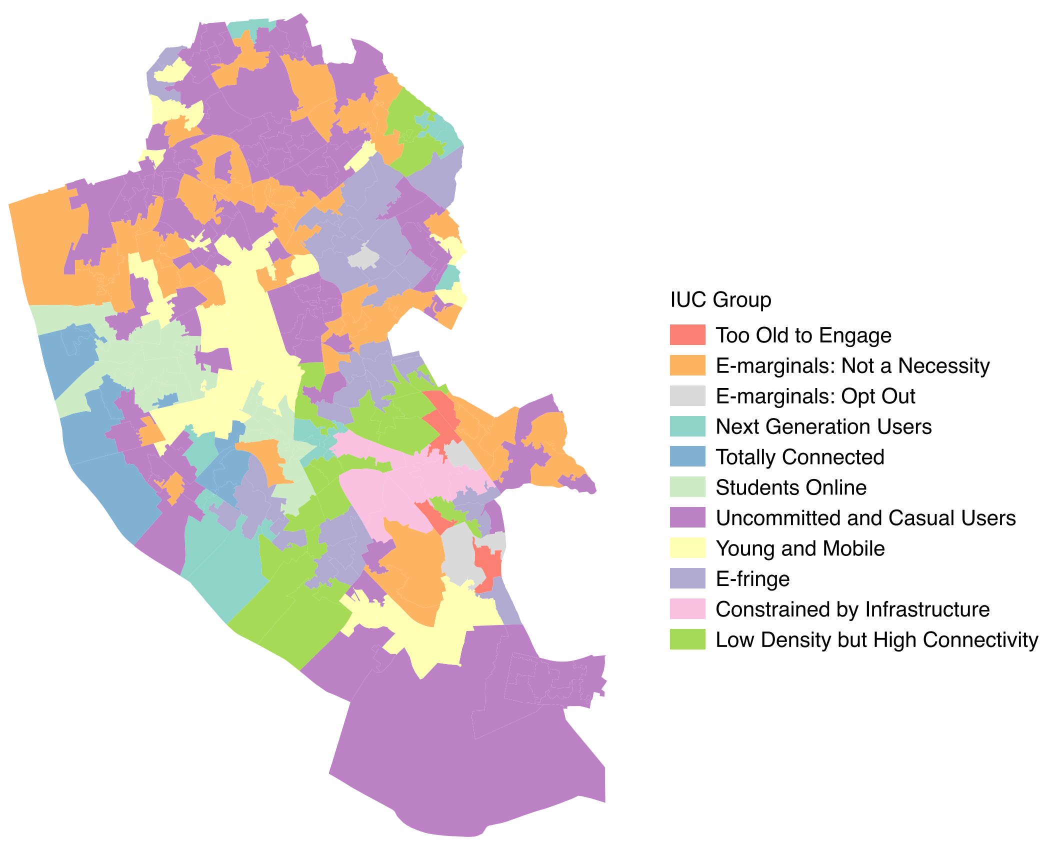
This bespoke geodemographic classification maps the geography of digital consumers within England by combining over seventy measures selected from survey and lifestyle data, alongside census and infrastructure performance statistics. This project forms an output of an ESRC funded PhD and research project, both undertaken at the University of Liverpool. This project aims to assist the profiling of customer databases, targeted marketing applications, policy delivery and strategic planning, where Internet use and engagement characteristics of a target population are of high importance. The Internet User Classification (IUC) has a two-tier structure comprising four Supergroups and eleven Groups that are used to classify all Lower Layer Super Output areas in England.

The table below details IUC Supergroup and Group assignments.

Supergroup	Group
1: E-unengaged	1a: Too Old to Engage
	1b: E-marginals: Not a Necessity
	1c: E-marginals: Opt Out
2: E-professionals and Students	2a: Next Generation Users
	2b: Totally Connected
	2c: Students Online
3: Typical Trends	3a: Uncommitted and Casual Users
	3b: Young and Mobile
4: E-rural and Fringe	4a: E-fringe
	4b: Constrained by Infrastructure
	4c: Low Density but High Connectivity



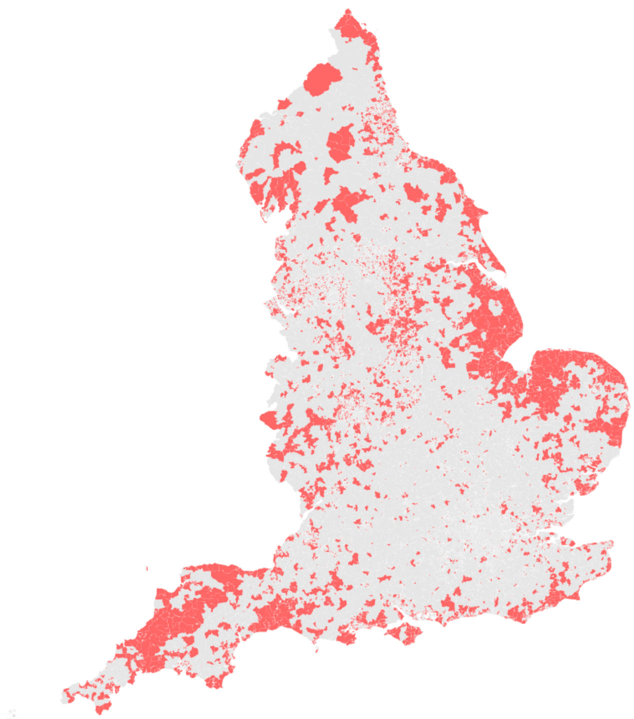
London by IUC Group



Liverpool by IUC Group

2 Pen Portraits: Supergroup Level

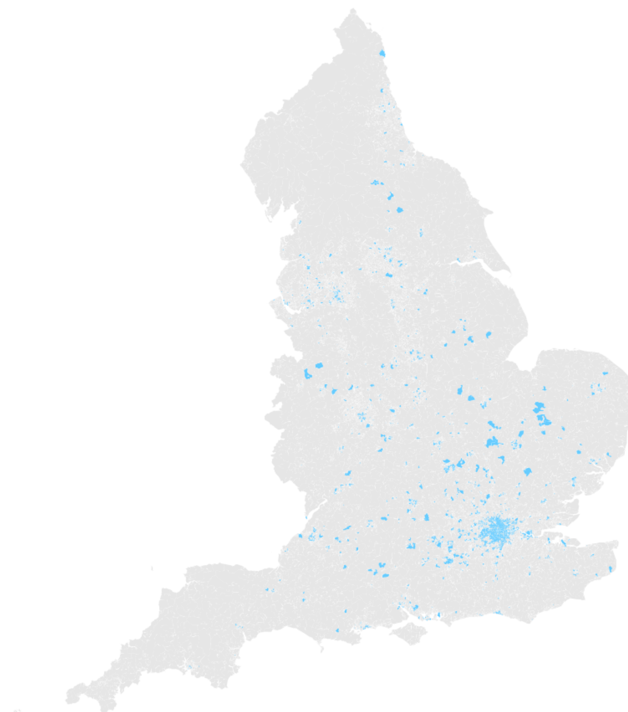
2.1 Supergroup 1: E-unengaged



Supergroup 1: E-unengaged National Distribution

The E-unengaged Supergroup display apparent low levels of engagement with Internet applications across all measures including; seeking information online, purchasing online, device ownership, general interest and mobile access. The age structure of the E-unengaged Supergroup is significantly skewed towards the elderly, with members most likely to be 60 plus. This Supergroup also has the highest proportion of residents aged 75 plus, of any Supergroup in the IUC. Members of this Supergroup generally favour traditional means of communication such as telephone and newspapers over their online equivalents. As such, device ownership including smartphones, tablets, e-readers, smart TVs and games consoles is far lower than the national average and the lowest of all Supergroups in the IUC. Infrastructure provision is in line with the national average, although usage falls far below, with Internet non-use significantly higher than the national average. Members of this Supergroup are most likely to be retired or to work in skilled trades or service occupations. Rates of higher-level (level three and above) qualifications are below the national average, most likely due to elderly populations. Geographically, this Supergroup tends to cluster around rural and coastal areas that attract elderly populations, although it is not uncommon for this Supergroup to also appear in urban areas, typically long-established suburbs as opposed to city centre areas. The E-unengaged Supergroup accounts for 24.8% of all Lower Super Output Areas nationally.

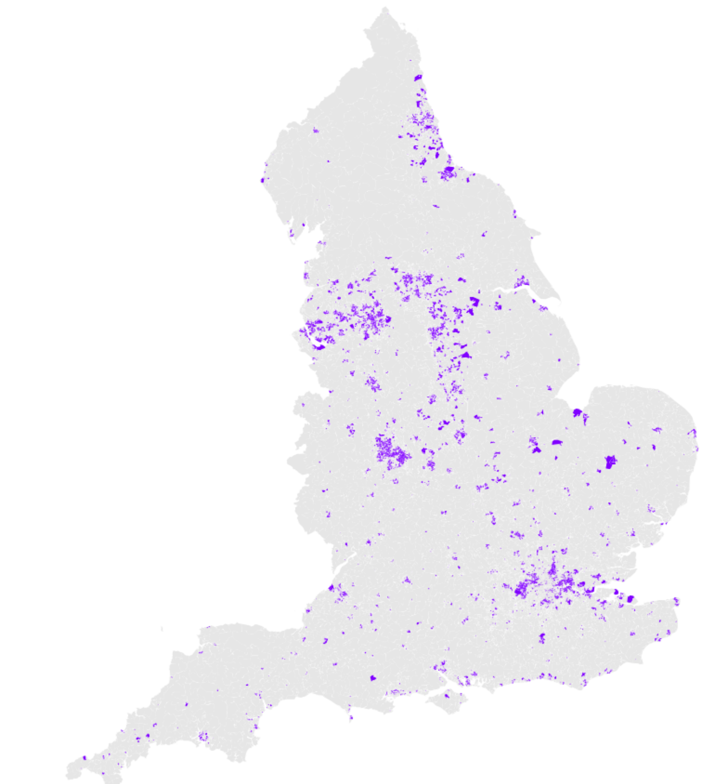
2.2 Supergroup 2: E-professionals and Students



Supergroup 2: E-professionals and Students National Distribution

The E-professionals and Students Supergroup display very high levels of engagement with Internet applications across all measures. Members of this Supergroup typically access the Internet using multiple devices, favouring access across mobile or fixed line Internet connections to ensure 'always online' connectivity. Seeking information is 'online by default' for this Supergroup, as are most everyday tasks such as banking, account and bill payments and food and grocery shopping. General interest in the Internet for information and entertainment is significantly higher than the national average. Mobile device ownership in this Supergroup is also higher than the national average, and the highest of all Supergroups in the IUC. Device ownership often extends to more recent types of device, such as tablet computers, smart TVs, e-readers and networked games consoles. Mobile phone ownership is high, with most users using smartphones to support email, social networking, navigation, mobile Internet access and third party apps. As would be expected, time spent online is higher than the national average within this Supergroup and the highest of any Supergroup in the IUC. In socio-economic and demographic terms, users are typically aged between 25 and 44 and highly qualified, in most cases to degree or higher-degree level and are likely to have found employment through the Internet. Equally, large numbers of full time students fall within this Supergroup and are typically aged between 18 and 24. Students within this Supergroup are characterised by very high levels of Internet usage, particularly through mobile devices such as smartphones and tablet computers. Geographically, this Supergroup tends to be found clustered within densely populated urban centres that have good Internet infrastructure and above average broadband performance. The E-professionals and Students Supergroup accounts for 16.7% of all Lower Super Output Areas nationally.

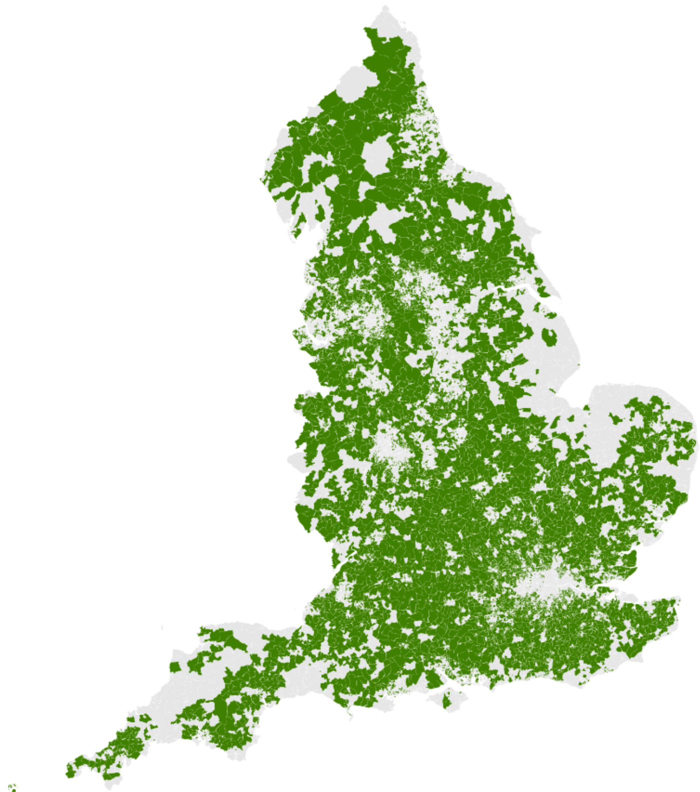
2.3 Supergroup 3: Typical Trends



Supergroup 3: Typical Trends National Distribution

The Typical Trends Supergroup displays levels of engagement that are the closest to the national average. Individuals are characterised by average engagement in terms of seeking information, device ownership and general interest in the Internet. Use of commercial applications such as online shopping, online banking and online bill payments are slightly below the national average and the use of mobile devices for Internet access is above the national average, in part because the younger individuals within this Supergroup favour mobile use. The Supergroup contains more individuals aged 10 to 17 of any Supergroup in the IUC. Members of this Supergroup who are of a typical working age are not highly qualified, and generally work in elementary or service occupations. Geographically, this Supergroup is clustered within and around urban areas, which in many cases also have higher than average levels of material deprivation. These areas are, however, well connected in terms of Internet infrastructure, and have above average broadband performance. The Typical Trends Supergroup accounts for 27% of all Lower Super Output Areas nationally.

2.4 Supergroup 4: E-rural and Fringe

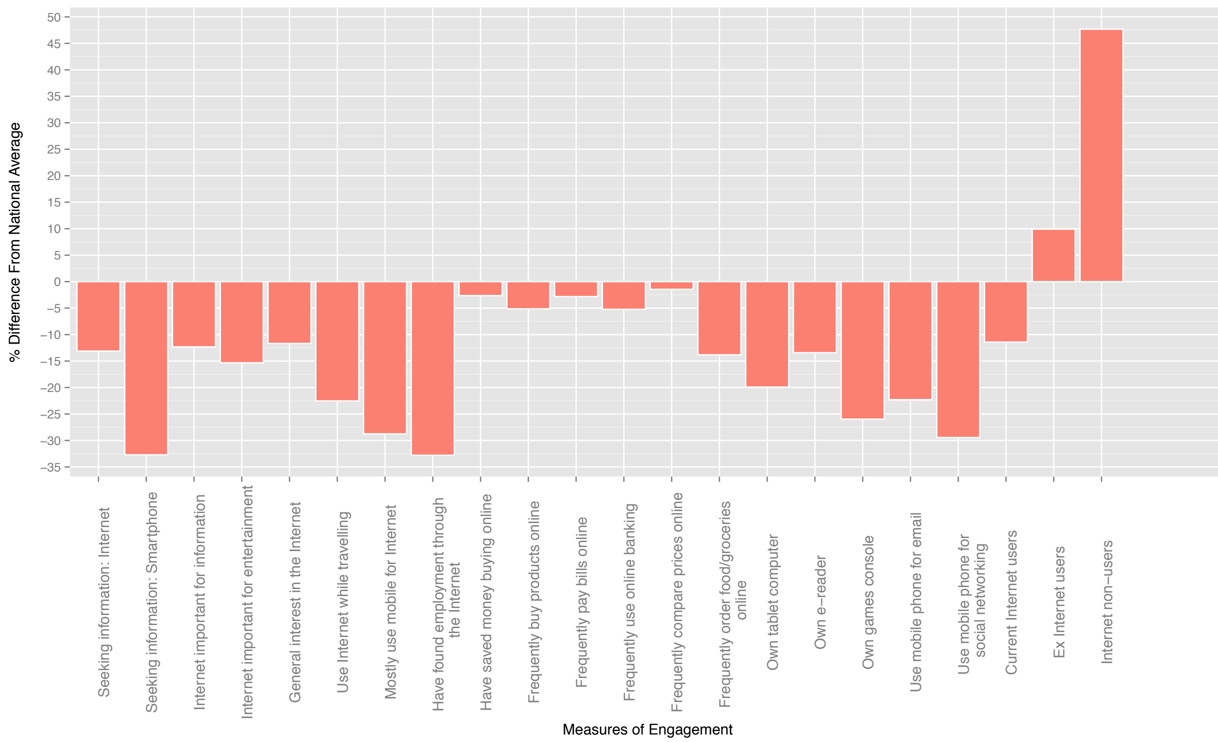


Supergroup 4: E-rural and Fringe National Distribution

Use of the Internet by members of the E-rural and Fringe Supergroup is constrained by poor infrastructure provision, typically because of their predominantly rural locations. Although engagement with Internet applications is only around the national average, this Supergroup display higher than average use of online shopping for products and groceries, online banking, online bill and account payments and price comparisons. This may in part arise because of the limited provision of these services locally. Fixed line broadband connections are used more than mobile broadband. Although mobile phone ownership is in line with the national average, the use of smartphones for data-dependent applications is significantly lower, given poorer infrastructure in these predominantly rural areas. The age structure within this Supergroup is middle aged to elderly with residents most likely to be aged between 45 and 75. Members of this Supergroup are generally well qualified and likely to work in managerial, professional or technical occupations. Device ownership is close to the national average, although devices such as e-readers are favoured over games consoles, consistent with the age structure of these areas. Performance of local broadband connections is below the national average, and the most constrained of all Supergroups in the IUC. The E-rural and Fringe Supergroup accounts for 31.5% of all Lower Super Output Areas nationally.

3 Pen Portraits: Group Level

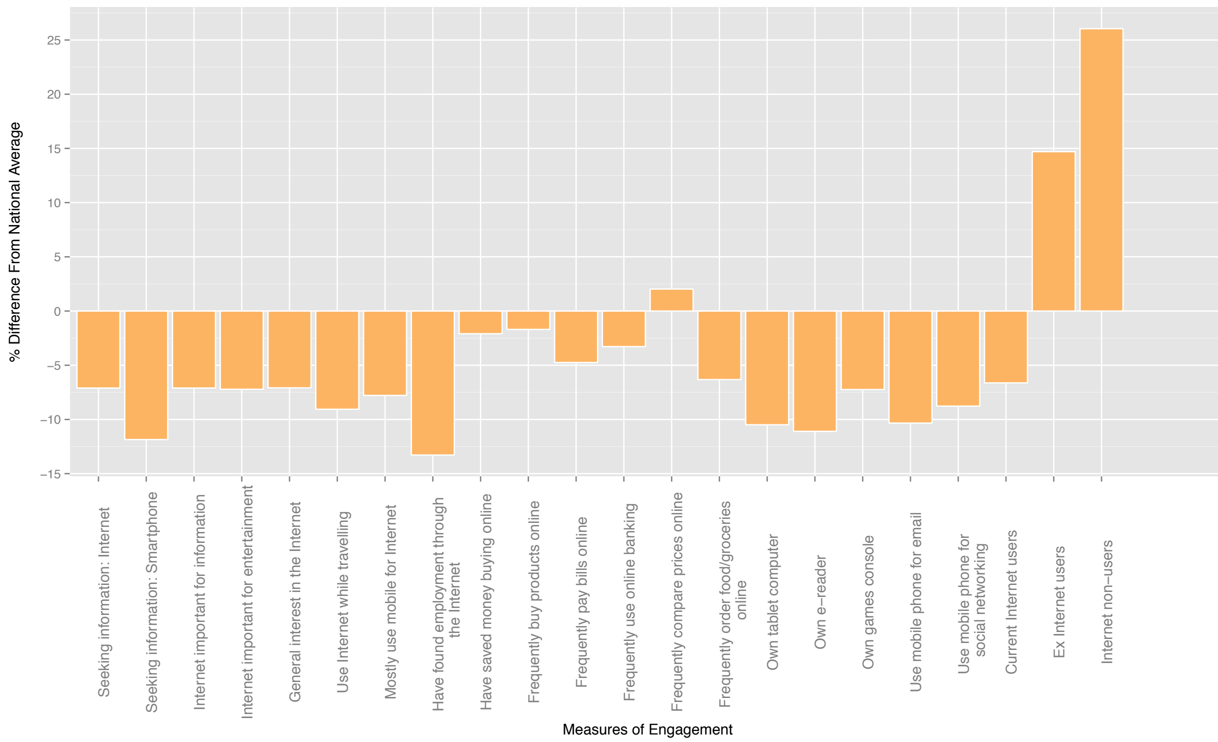
3.1 Group 1a: Too Old to Engage



Group 1a: Too Old to Engage: Group Characteristics

The Too Old to Engage Group is characterised by large elderly populations who show little or no engagement with the Internet across all applications. The proportion of residents aged 75 plus is higher than any Group in the IUC. As a result, Internet enabled device ownership is lower than the Supergroup average, and the lowest of any Group in the IUC. Abstinence from Internet use is higher than the Supergroup average and far above the national average. Enclaves of this Group are found in coastal and lower density rural areas that serve as retirement destinations. Infrastructure provision and performance is typically slightly below the national average. The Too Old to Engage Group accounts for 4% of all Lower Super Output Areas nationally.

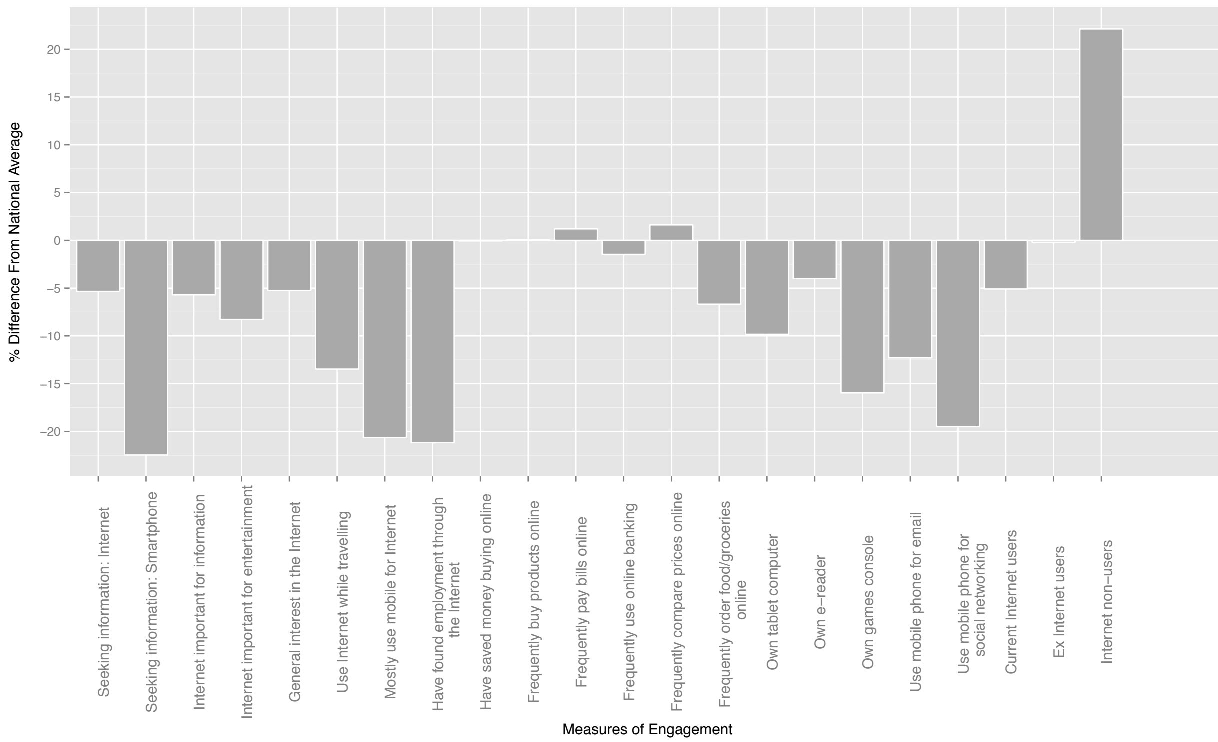
3.2 Group 1b: E-marginals: Not a Necessity



Group 1b: E-marginals: Not a Necessity: Group Characteristics

Members of the E-marginals: Not a Necessity Group typically have low engagement with Internet applications, lower than average qualifications and higher than average rates of employment in blue collar occupations that are not heavily reliant on digital skills. Of those that do access the Internet, many do so using a smartphone. Residents of this Group tend to be found within urban areas characterised by high levels of material deprivation, although infrastructure provision and performance are in line with the national average. The E-marginals: Not a Necessity Group accounts for 10.4% of all Lower Super Output Areas nationally.

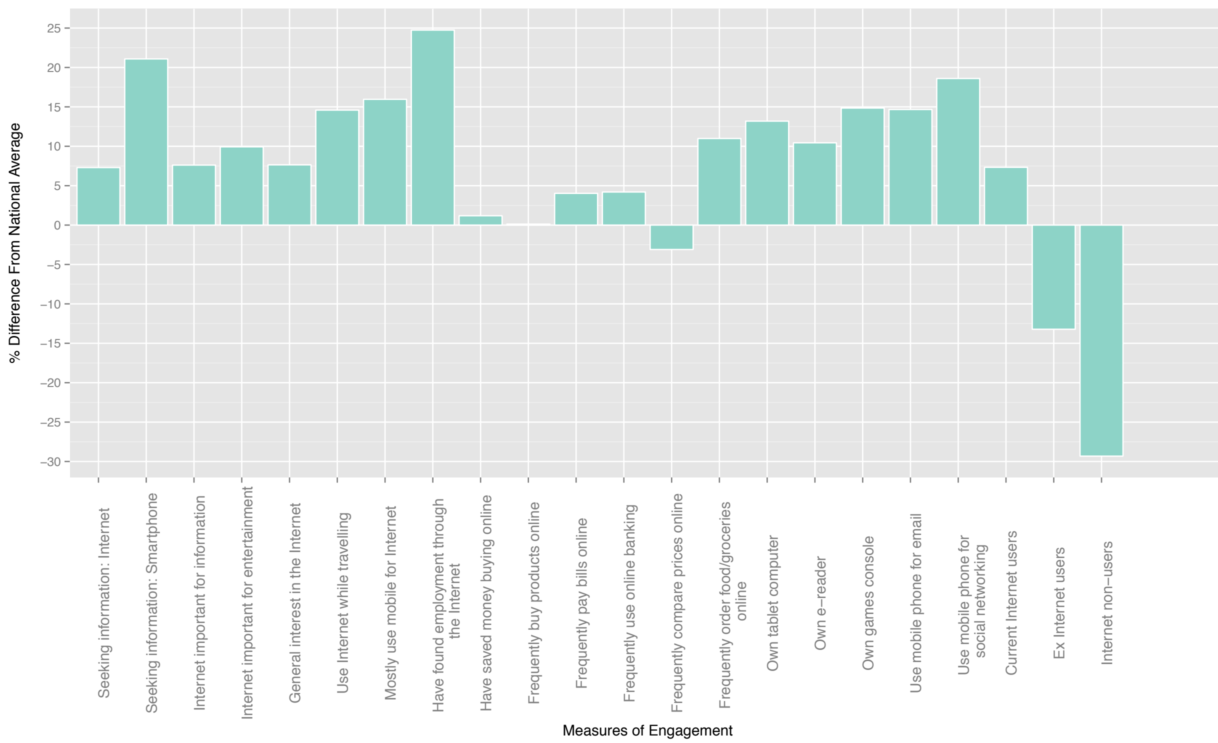
3.3 Group 1c: E-marginals: Opt Out



Group 1c: E-marginals: Opt Out: Group Characteristics

The E-marginals: Opt Out Group are characterised by low levels of engagement with the Internet for applications such as seeking information and entertainment, preferring instead more traditional media such as newspapers and television, in part reflecting the elderly demographic of this Group. Typically this Group is aged 60 plus, with significantly higher than average incidence of those aged 65 to 84. Geographically, this Group tends to be found in affluent rural and fringe areas that are more sparsely populated and where infrastructure provision and performance is below the national average. Access to the Internet through mobile devices is below the national average. Those who do choose to use the Internet tend to use it for price comparison and occasional online shopping. Levels of qualifications are generally above the national average, and those members who are not retired will typically be employed in senior managerial, professional or skilled trade occupations. Abstinence is significantly higher than the national average, but the lowest within the Supergroup. The E-marginals: Opt Out Group accounts for 10.4% of all Lower Super Output Areas nationally.

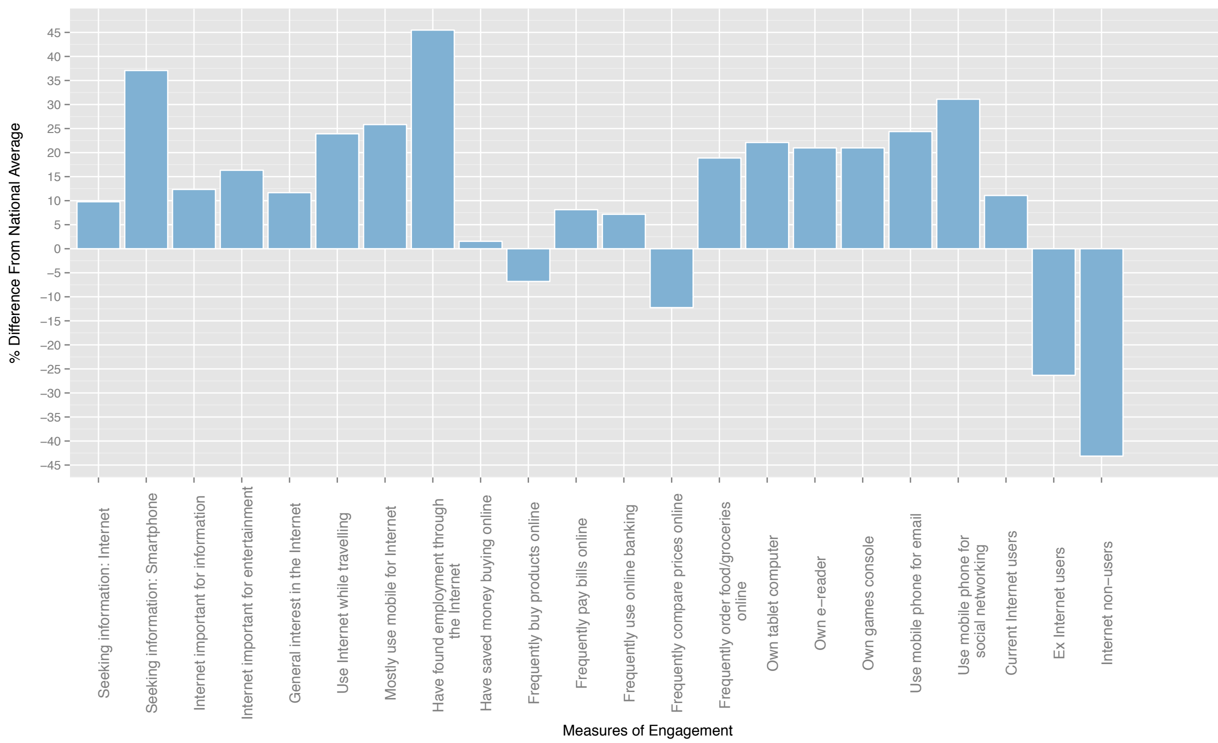
3.4 Group 2a: Next Generation Users



Group 2a: Next Generation Users: Group Characteristics

The Next Generation Users Group is characterised by high levels of engagement across all applications of the Internet. Members of this Group are heavy smartphone users and typically access the Internet on the move and for applications such as email, social networking and navigation. However, they favour fixed line connections for most other tasks such as general browsing and seeking information. Device ownership is higher than the national average, and members of this Group are likely to own several Internet enabled devices, such as tablet computers, e-readers and smart TVs. Levels of qualification are high within this Group, with higher than average rates of degree and higher degree level qualifications. The age structure is young to middle aged, with members of this Group most likely aged between 25 and 44, and in some cases with young children. Employment tends to be in managerial, professional and technical occupations. General interest in the Internet is above the national average. Members of this Group are found in affluent, higher density suburban and city fringe areas where infrastructure provision and performance is above the national average. Next Generation Users are the second most heavily engaged Group within the IUC, behind Group 2b: Totally Connected and account for 10.2% of all Lower Super Output Areas nationally.

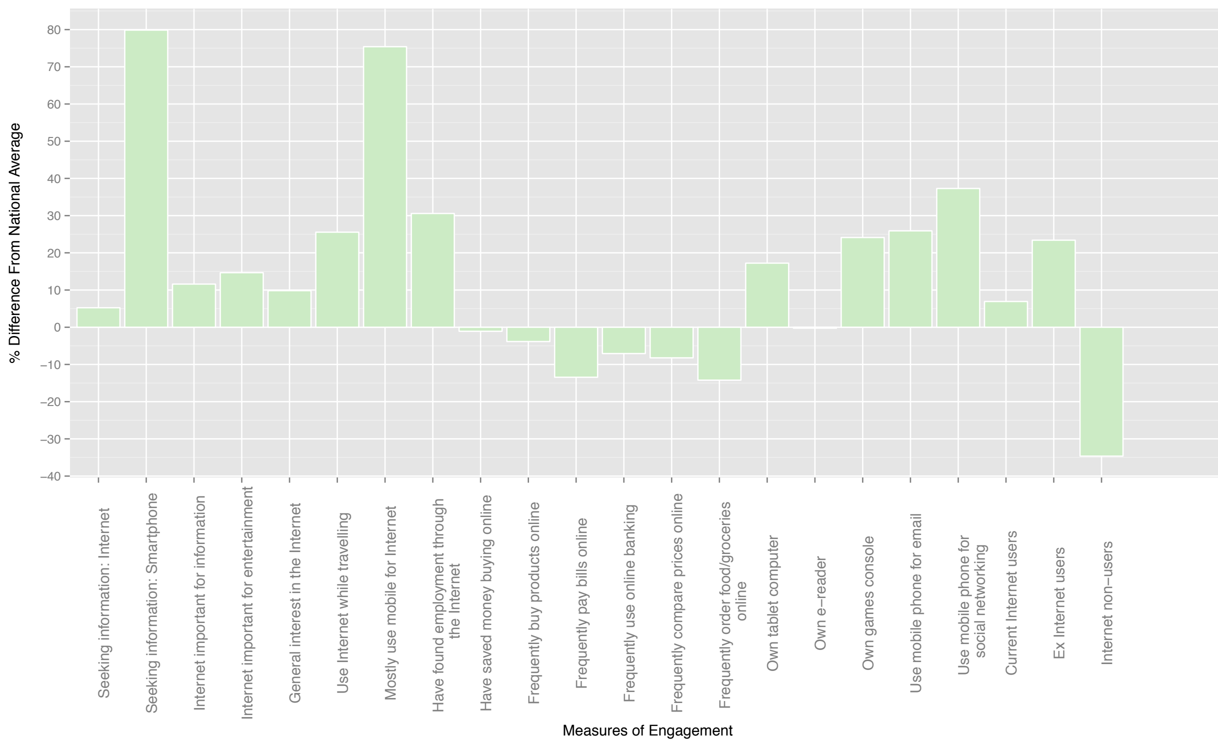
3.5 Group 2b: Totally Connected



Group 2b: Totally Connected: Group Characteristics

The Totally Connected Group is characterised by the highest levels of engagement within the IUC and score higher than the Supergroup and national averages for most measures of engagement. This Group displays a clear preference to use the Internet by default for almost all applications. Members of this Group access the Internet through multiple devices, whilst on the move and in the home to ensure seamless connectivity. As such, device ownership is significantly higher than the national and Supergroup averages and members of this Group own a wide range of Internet enabled hardware. Levels of qualification are significantly higher than the national average. Professional occupations are most prevalent, with the age structure of residents being young to middle aged, sometimes with young children. Geographically, this Group tends to be found in affluent city centre and city fringe areas that are densely populated and where infrastructure provision and performance is above the national average. Members of this Group show below average rates of online shopping, perhaps given good local retail choice. However, rates of online shopping for food and groceries are significantly above the national and Supergroup averages as this enables wider choice and convenience in highly populated areas. Totally Connected are the most heavily engaged Group within the IUC and account for 4.8% of all Lower Super Output Areas nationally.

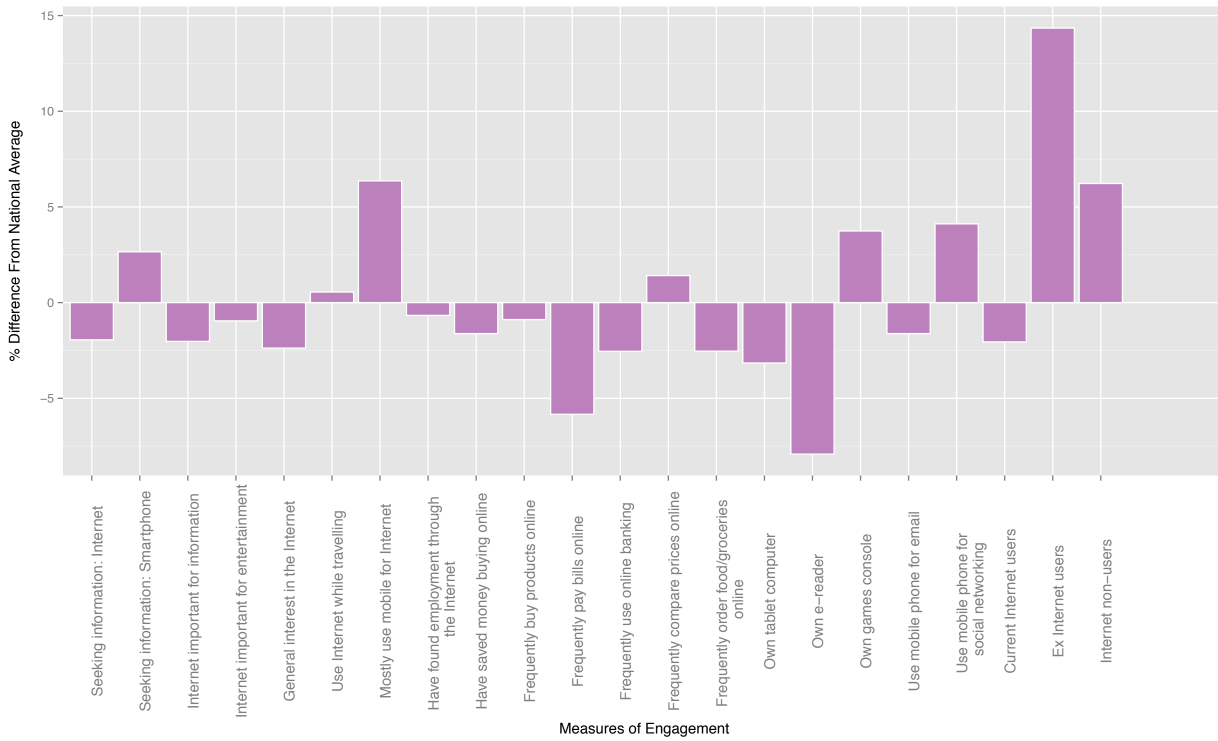
3.6 Group 2c: Students Online



Group 2c: Students Online: Group Characteristics

Students Online represents a small but very distinct Group that is comprised almost entirely of student areas. The Group is characterised by very high levels of Internet usage, particularly through mobile devices such as smartphones and tablet computers. Smartphones are the device of choice for electronic communication and are used for a wide range of applications including email, social networking, third party applications, web browsing and sharing photos and videos. Members of this Group are typically aged between 18 and 24 and are registered as full time students. Interest in the Internet for information and entertainment is above the national average, and a higher than average proportion of the local population is likely to have found, or to be seeking, employment through the Internet. With very high proportions of students in this Group, most members are likely to possess Level Three qualifications or above. Employment across all sectors is below the national average with the exception of sales and customer service roles, in which some students choose to work, most likely on a part-time basis to support their studies. Geographically, this Group is often found in the major urban conurbations, usually within city centres and university campus areas where there are highly concentrated student populations. Infrastructure provision and connection performance is above the national average in these areas. The Students Online Group accounts for 1.7% of all Lower Super Output Areas nationally.

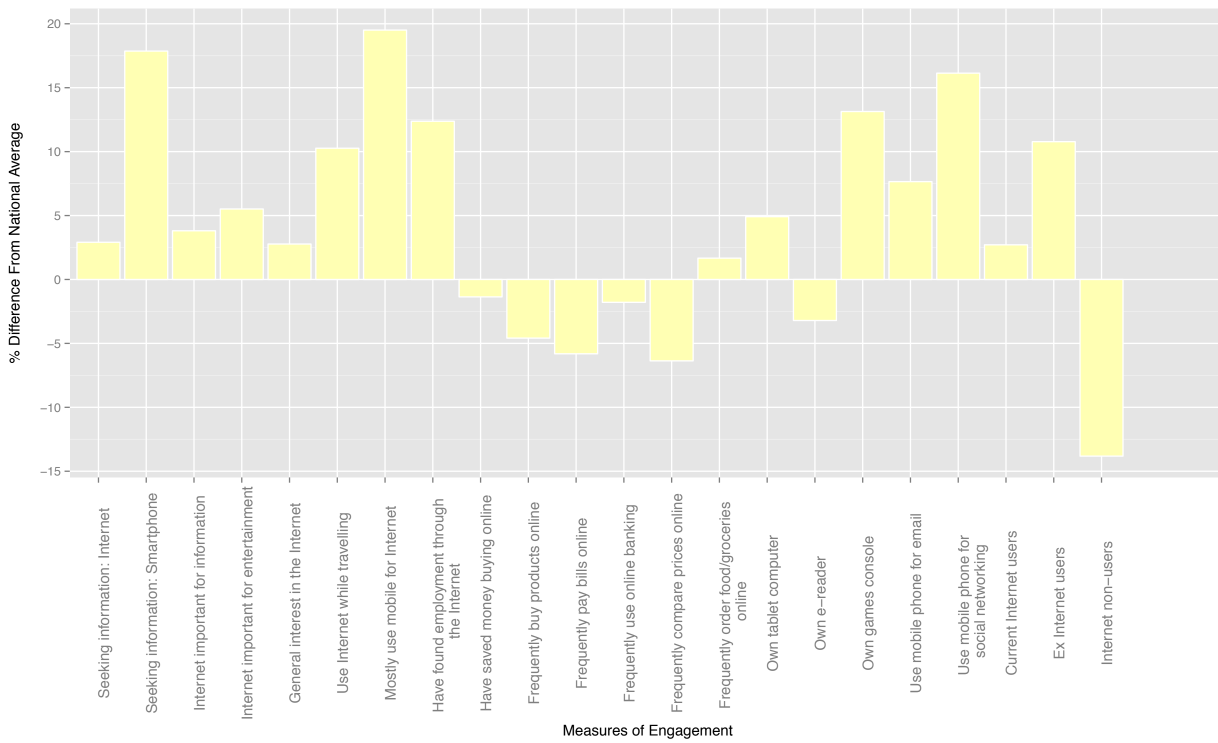
3.7 Group 3a: Uncommitted and Casual Users



Group 3a: Uncommitted and Casual Users: Group Characteristics

The Uncommitted and Casual Users Group are characterised by mixed levels of engagement with the Internet. Access to the Internet through smartphones is marginally above the national average and access through fixed-line connections falls marginally below. Members of this Group show below average rates for purchasing online but above average rates for price comparison and selling online. Age structure is generally young to middle aged, with higher than average proportions of young and teenage children. Qualifications tend to be of a lower level and members of this Group are most likely to work in service, sales and elementary occupations. Overall, abstinence from Internet use is marginally higher than the national average and general interest in the Internet falls shy of the national average. This Group also contains higher than average numbers of lapsed Internet users. Geographically, this Group tends to be found in major urban and city fringe areas that suffer higher levels of material deprivation, but where infrastructure provision and performance is above the national average. The Uncommitted and Casual Users Group accounts for 15.5% of all Lower Super Output Areas nationally.

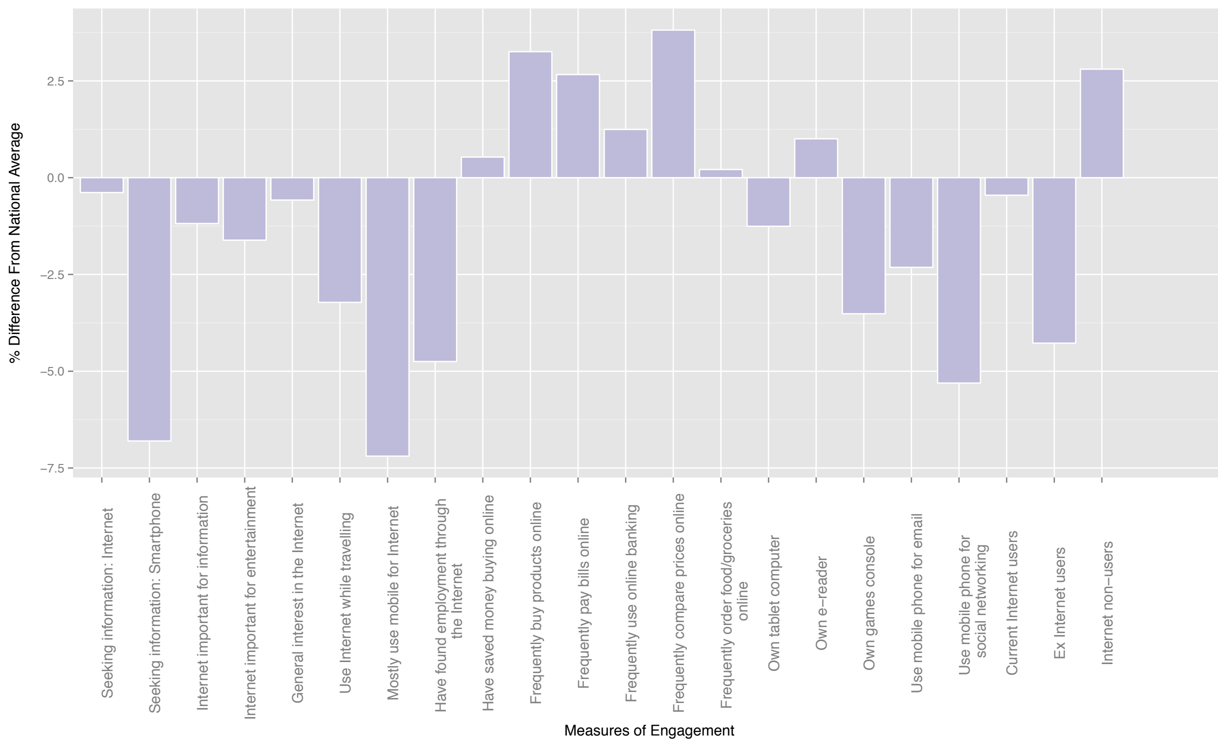
3.8 Group 3b: Young and Mobile



Group 3b: Young and Mobile: Group Characteristics

The Young and Mobile Group is predominantly young and has a tendency to access the Internet using mobile devices rather than fixed line connections. This Group is found in major urban conurbations where population density is above average and infrastructure provision is sufficient to support heavy mobile broadband usage. These areas are typically inner city or city fringe and experience mixed levels of material deprivation. As a Group there are higher than average proportions of young and teenage children and adults aged 25 to 44. Conversely, the proportion of adults aged over 45 falls below the national and Supergroup averages. All levels of qualification are below the national average and those who work are likely to be employed in elementary, sales or service occupations. Interest in the Internet for entertainment and information is above the national average, most likely reflecting the prevailing age structure. This Group displays a lower than average tendency to purchase online, and would be expected to shop locally in most cases. The Young and Mobile Group accounts for 11.5% of all Lower Super Output Areas nationally.

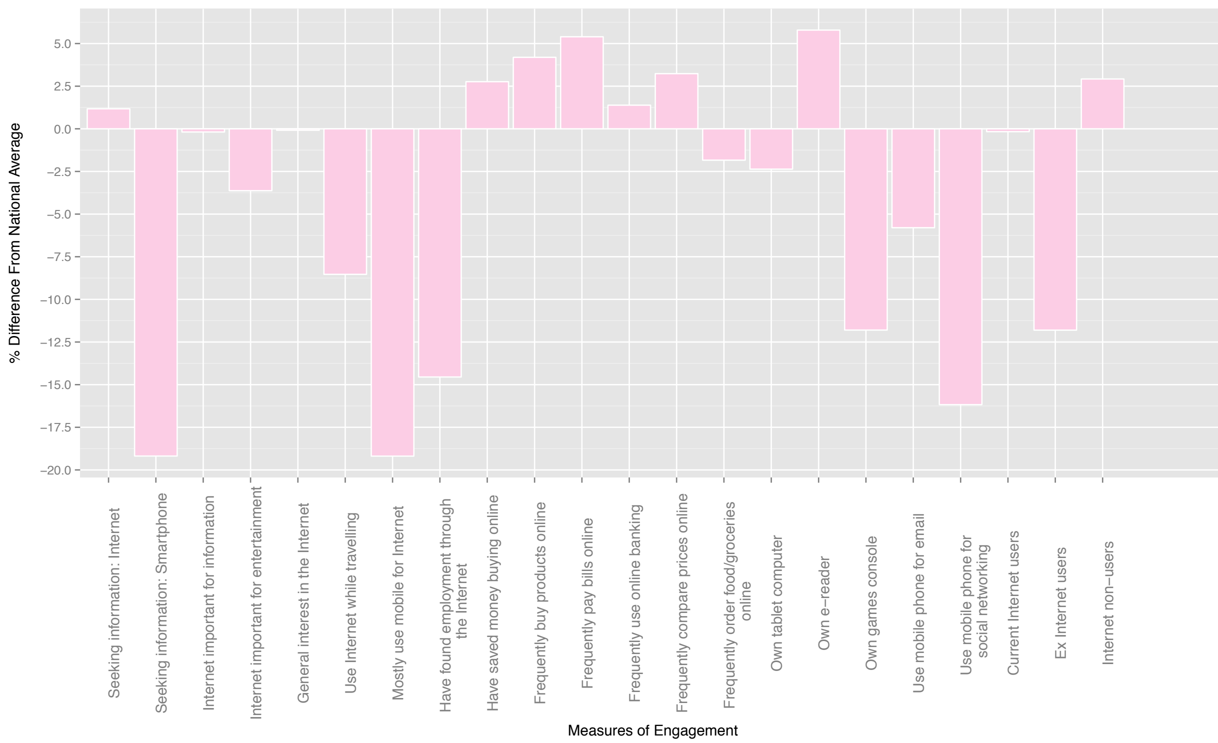
3.9 Group 4a: E-fringe



Group 4a: E-fringe: Group Characteristics

The E-fringe Group is distinguished by its location around the fringes of urban areas that are typically low density or semi-rural. Age structure is middle aged to elderly and there are fewer than average numbers of young adults aged 18-29, a group who are likely to have moved to more major urban conurbations. General interest in the Internet within this Group is slightly below the national average and the lowest within the Supergroup, rates of current Internet users are also below average and numbers of Internet non-users are above the national average. Members of this Group generally have mixed levels of qualifications, rates of level 1 to level 3 qualifications are all above the national average and incidence of persons with no qualifications is below average. Members are most likely to work in administrative and secretarial or skilled trade occupations. The most common uses of the Internet within this Group are paying bills and banking online, comparing prices and buying products, which score above the national average. Below average rates are recorded for seeking information and entertainment purposes, consistent with the age profile of this Group. Equally, ownership of Internet enabled devices is below average, with the exception of e-readers, which are popular amongst this Group. Infrastructure provision and performance is marginally below the national average but would be unlikely to limit access. The E-fringe Group accounts for 11.1% of all Lower Super Output Areas nationally.

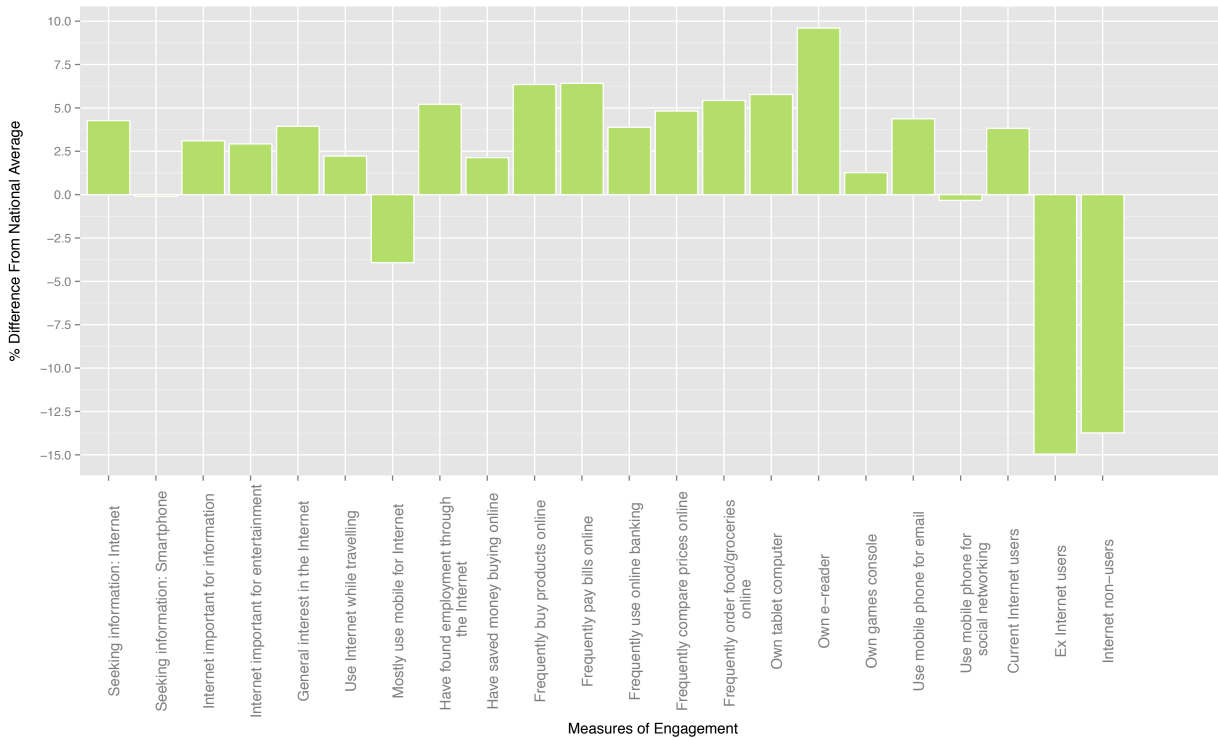
3.10 Group 4b: Constrained by Infrastructure



Group 4b: Constrained by Infrastructure: Group Characteristics

The Constrained by Infrastructure Group is characterised by locations in low-density rural areas where there is poor provision and performance of local Internet infrastructure, both fixed line and mobile. This limits engagement with some online applications. Fixed line broadband performance falls significantly below the national average and is the lowest within the Supergroup as distances to local telephone exchanges are much higher. Distances to the nearest mobile base station for cellular and data coverage are also higher than the national average, and as such further constrains performance and usability. Perhaps as a result, the use of mobile broadband through devices such as smartphones or dongles is below average. Despite poor infrastructure, general interest in the Internet is in line with the national average and members of this Group display above average rates of purchasing online, comparing prices, online banking and paying bills, most likely as this saves travelling to a local retail centre to access these services. Internet enabled device ownership is again lower than the national average with the exception of e-readers, likely due to the prevailing age structure of this Group, which is middle aged and elderly. Those who are not retired are generally highly qualified and work in managerial, professional or technical occupations. Internet non-use is above average but reflects the prevailing age profile of the Group. The Constrained by Infrastructure Group accounts for 11% of all Lower Super Output Areas nationally.

3.11 Group 4c: Low Density but High Connectivity



Group 4c: Low Density but High Connectivity : Group Characteristics

The Low Density but High Connectivity Group is found in areas that are sparsely populated, typically rural and semi-rural areas, or areas with urban parkland. Despite disparate populations, this Group is generally well connected and displays the strongest infrastructure and performance characteristics within the Supergroup, generally falling in line with the national average. Internet use in general is higher across all applications than the Supergroup average, and this Group shows a higher than average propensity for ordering food and groceries online. These characteristics are representative of the prevailing demographic of well-educated workers (often with degrees or higher degrees) who work in high-grade professional occupations. Similarly, Internet enabled device ownership is above the national average, perhaps because local infrastructure is able to support this. Age structure is mixed, although members of this Group are most likely to be aged 45 to 59 with young or teenage children. General interest in the Internet is above the national average and is the highest within the Supergroup. As would be expected, rates of Internet non-use are below the national average. The Low Density but High Connectivity Group accounts for 9.4% of all Lower Super Output Areas nationally.

4 Queries and Further Information

Queries regarding the data and technical methodology that underpin the IUC can be sent to Dean Riddlesden (PhD, University of Liverpool) at:

Email: d.riddlesden@liv.ac.uk
Tel: +44 (0)151 794 1355
Tweet: @deanriddlesden

Alternatively:

Email: alex.singleton@liv.ac.uk
Tel: +44 (0)151 794 3108
Tweet: @alexsingleton

For legal reasons and in accordance with agreements with our data partners, datasets used in the formation of the IUC will not be distributed to third parties. Distribution of the IUC itself is free and open. Please cite the IUC appropriately where it has been used in any published material.

Appendix D

Example Geodemographic Input Correlations

	QC30f	QC30h	QC30i	QH1a	QH1b	QH1c	QH3	QH5	QH7f	QH7g
QC1b	-0.38	0.69	0.94	0.84	-0.07	-0.89	-0.05	0.88	0.93	0.52
QC2a	-0.41	0.32	0.79	0.57	0.28	-0.65	-0.37	0.65	0.71	0.16
QC22a	-0.42	0.77	0.92	0.83	-0.16	-0.86	0.08	0.87	0.93	0.60
QC22b	0.36	0.41	0.12	0.51	-0.65	-0.45	0.61	0.33	0.37	0.63
QC30b	0.92	0.12	-0.10	0.08	-0.14	-0.07	0.37	0.11	0.04	0.29
QC30d	0.21	0.60	0.02	0.45	-0.86	-0.35	0.92	0.30	0.34	0.81
QC30e	0.09	0.87	0.44	0.72	-0.77	-0.65	0.74	0.68	0.65	0.90
QC30f	1.00	-0.14	-0.31	-0.27	0.11	0.27	0.12	-0.16	-0.29	-0.06
QC30h	-0.14	1.00	0.78	0.82	-0.53	-0.80	0.43	0.86	0.81	0.80
QC30i	-0.31	0.78	1.00	0.79	-0.04	-0.84	-0.13	0.94	0.86	0.46
QH1a	-0.27	0.82	0.79	1.00	-0.54	-0.99	0.41	0.89	0.96	0.84
QH1b	0.11	-0.53	-0.04	-0.54	1.00	0.43	-0.85	-0.27	-0.36	-0.76
QH1c	0.27	-0.80	-0.84	-0.99	0.43	1.00	-0.31	-0.91	-0.97	-0.79
QH3	0.12	0.43	-0.13	0.41	-0.85	-0.31	1.00	0.12	0.30	0.81
QH5	-0.16	0.86	0.94	0.89	-0.27	-0.91	0.12	1.00	0.91	0.66
QH7f	-0.29	0.81	0.86	0.96	-0.36	-0.97	0.30	0.91	1.00	0.78
QH7g	-0.06	0.80	0.46	0.84	-0.76	-0.79	0.81	0.66	0.78	1.00
QH7h	-0.35	0.64	0.97	0.75	0.08	-0.82	-0.23	0.88	0.85	0.37
QH7i	-0.38	0.71	0.76	0.82	-0.40	-0.82	0.17	0.81	0.79	0.60
QH11	-0.20	0.75	0.88	0.94	-0.27	-0.96	0.16	0.94	0.95	0.68
QH12b	-0.34	0.74	0.91	0.90	-0.20	-0.94	0.12	0.90	0.98	0.65
QH12e	-0.39	0.66	0.96	0.75	0.04	-0.81	-0.18	0.86	0.86	0.40
QH12h	-0.28	0.87	0.93	0.89	-0.30	-0.92	0.18	0.94	0.95	0.69
QH12i	-0.40	0.60	0.94	0.72	0.09	-0.79	-0.22	0.83	0.84	0.35
QH12j	-0.35	0.74	0.96	0.85	-0.11	-0.90	0.00	0.91	0.94	0.56
QH12k	-0.36	0.72	0.92	0.88	-0.15	-0.92	0.07	0.89	0.97	0.61
QH13_Current	-0.23	0.82	0.83	0.98	-0.44	-0.99	0.38	0.91	0.98	0.83
QH13_Ex	-0.09	-0.69	-0.17	-0.61	0.87	0.52	-0.93	-0.42	-0.52	-0.91
QH13_Non	0.26	-0.79	-0.86	-0.97	0.36	0.99	-0.29	-0.92	-0.99	-0.77