

# Multi-Level Integrated Classifications

## Based on the 2001 Census

Daniel William Vickers

The University of Leeds

School of Geography

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

January 2006

The Candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

## **Acknowledgements**

First and foremost I must thank my supervisors Phil Rees and Mark Birkin. Phil especially has provided support, advice and guidance far beyond that required of a PhD supervisor. I will be forever grateful and indebted to him for his constant encouragement and the unwavering belief that he has shown in me. I would also like to thank Graham Clarke and John Stillwell, for their valuable nuggets of advice in their role as members of my research support group.

I would like to acknowledge the funding that was provided for the project by the Economic and Social Research Council (ESRC) and CASE partner, the Office for National Statistics (CASE Award PTA-033-2002-00067). Without such generous financial support the project would not have been possible.

Thanks must go to all three census agencies ONS, GROS and NISRA for providing 2001 Census data free of charge through the 'click-use' licensing agreement. Also to the dissemination units of ESRC/JISC Census Programme, whose websites I made extensive use.

Staff at the ONS have been invaluable in the completion of the classification and were instrumental in its publication as a 'National Statistic'. John Charlton my initial CASE supervisor, is worthy of a special mention. John was the driving force of the project within the ONS and provided much needed support in the early stages of the project. Sadly John passed away before the completion of the classification and never got to see its publication. Other ONS staff who helped in the development of the classification include Andrew Botterill, Simon Compton, Chris Denham, Greg Phillpotts, Samantha Poole, Nargis Rahman, Nick Ormiston-Smith, Dev Virdee, Alison Whitworth and Susan Williams. I would also like to thank anyone else within ONS who was involved in the publication of the classification, but of whom I am unaware.

Many thanks must go to Keith Dugmore (Demographic Decisions Ltd.), and Martin Callingham (Birkbeck College, University of London). At the start of the project I could not have hoped to find two such enthusiastic and vocal supporters of my work.

Thanks to Pete Shepherd for all the umpteen discussions about classification and the sharing of ideas and references. Talking to someone who was equally obsessed with cluster analysis as me was a great help. Thanks to Ian Shuttleworth (Queens University Belfast) for casting his experienced eye over my great simplification of demographic issues in Northern Ireland.

---

Thank you to everybody who took part in the consultation exercise, the small amount of time and effort that you gave me was of immeasurable value.

I would like to thank all my friends, family and colleagues with whom I have shared the last three years of my life and to anyone else who offered, help, support, advice or just showed an interest.

---

---

## Abstract

The purpose of this thesis is to describe and explain the processes and decisions that were involved in the creation of the National Area Classification of 2001 Census Output Areas (OAs). The thesis describes the creation of the classification: selection of the variables, assembly of the classification database, the methods of standardisation and the clustering procedures, together with some discussion of alternative methodologies that were considered for use. The processes used for creating the clusters, their naming and description are outlined. The classification is mapped and visualised in a number of different ways.

In order to enable a classification of OAs to be possible the document starts with a review of the history of area classification and issues surrounding its future development. The methodological and theoretical issues in the creation of a classification system are also discussed. In order to test out the practicalities of creating a classification system, a classification of UK local authorities was created prior to the construction of the OA classification.

The thesis describes the quality assurance procedures that the OA classification was put through. This included an innovative consultation exercise. This ensured that the classification was of enough quality and without error, enabling it to be published as a 'National Statistic'. Examples of use of the classification are presented, outlining the value and relevance of the classification to social research. The OA classification is connected to other scales of classification to form a multi-scale classification system enabling the socio-demographic pattern of the UK to be examined at multiple scales.

The project had to overcome numerous methodological issues due to the size of dataset that was used. The project used a new methodology to create the first free-to-use small scale classification of the UK. The classification was published as a 'National Statistic' on the 29<sup>th</sup> July 2005 and is freely available. The classification can be downloaded from the ONS website at [http://www.statistics.gov.uk/about/methodology\\_by\\_theme/area\\_classification/oa/default.asp](http://www.statistics.gov.uk/about/methodology_by_theme/area_classification/oa/default.asp) or via the School of Geography, University of Leeds website at: <http://www.geog.leeds.ac.uk/people/d.vickers/OAclassinfo.html>. Alternatively it can be ordered on CD by contacting [info@statistics.gov.uk](mailto:info@statistics.gov.uk). The publication of the classification as a 'National Statistic' has created a resource that can be used by private, public and academic researchers.

---

## Table of Contents

Acknowledgements.....	i
Abstract.....	iii
Table of Contents.....	iv
List of Figures.....	viii
List of Tables.....	xi
Abbreviations.....	xiii
<b>Chapter 1: Introduction, Project Outline, Aims and Objectives.....</b>	<b>1</b>
1.1: The 2001 Census, the Importance of Place and Area Classification.....	1
1.2: Aims and Objectives.....	4
1.2.1: Methods.....	6
1.2.2: Outputs.....	7
1.3: Thesis Structure.....	8
<b>Chapter 2: Introducing Clustering, Area Classification and Geodemographics.....</b>	<b>11</b>
2.1: Introduction.....	11
2.2: Why Classify?.....	11
2.2.1: Clusters All Around You.....	12
2.3: What are Area Classifications and Geodemographics?.....	15
2.4: The Development of Area Classification.....	17
2.4.1: Modest Beginnings.....	17
2.4.2: Contemporary Geodemographics and the Rise of the Commercial Segmentation System (post 1981 Census).....	21
2.5: Some Current Uses of Area Classifications.....	24
2.6: The Geographical Building Blocks of the Census.....	25
2.6.1: Geographical Hierarchies.....	26
2.6.2: Administrative Geography of the United Kingdom.....	27
2.7: Criticisms and Limitations: Lies, Damn Lies and Geodemographics.....	30
2.7.1: Ecological Fallacy.....	30
2.7.2: The Modifiable Areal Unit Problem.....	30
2.7.3: The Labelling of Clusters.....	33
2.7.4: Out of Date as Soon as it is Made?.....	34
2.7.5: Arbitrariness, Transparency and Lack of Validation.....	35
2.7.6: Lacking in Theory and Statistical Grounding.....	36
2.7.7: Ethics, Privacy and Persecution.....	37
2.7.8: Misrepresentation: Over-precision and Over-stating of Capabilities.....	38
2.8: A Fuzzy Future?.....	40
2.9: Conclusions: Extracting a Research Agenda.....	41
<b>Chapter 3: Making a Classification System: a Guide to Methods and Procedures.....</b>	<b>43</b>
3.1: Introduction.....	43
3.2: Inputs.....	47
3.2.1: Data, the More the Better?.....	47
3.2.2: Census Data.....	47
3.2.3: Issue of Census Data Quality.....	49
3.2.4: Other Data Sources.....	50
3.2.5: Data Quality Issues from Other Data Sources.....	51
3.2.6: The Theory of Selecting Input Variables.....	52
3.2.7: Correlation Analysis.....	55
3.2.8: Data Reduction.....	56
3.3: Processes.....	58
3.3.1: Methods of Standardisation.....	58
3.3.2: Z-score Standardisation.....	58
3.3.3: Range Standardisation.....	59

3.3.4: Inter-decile Range Standardisation.....	59
3.3.5: Weighting of Variables.....	60
3.3.6: Methods of Clustering.....	62
3.3.7: Ward's Hierarchical Clustering Algorithm.....	65
3.3.8: K-means Classification .....	66
3.3.9: Distance Measures.....	67
3.4: Outputs.....	68
3.4.1: Selecting the Cluster Numbers and Classification Structure.....	68
3.4.2: Naming the Clusters.....	69
3.4.3: Pen Portraits, Maps, Photos and Visual Descriptions.....	70
3.5: Conclusions.....	71
<b>Chapter 4: A Classification of the UK Local Authorities.....</b>	<b>72</b>
4.1: Introduction.....	72
4.2: The Areas to be Clustered.....	73
4.3: What Went Before? A Comparative Evaluation of Variables Used in Previous Systems.....	74
4.4: Inputs.....	79
4.4.1: Potential Variables to be used in the Classification.....	79
4.4.2: Variable Selection Procedure.....	82
4.4.3: Final Variable List.....	90
4.5: Processes.....	91
4.5.1: Variable Standardisation.....	91
4.5.2: Clustering Method.....	91
4.6: Outputs.....	93
4.6.1: Structure and Naming.....	95
4.6.2: Pen Portraits.....	95
4.6.3: Additional Information.....	98
4.6.4: Mapping out the Clusters.....	100
4.7: A Comparison with the 2001 ONS Classification of Local Authorities.....	105
4.7.1: Variable Choices Comparison.....	105
4.7.2: Methodology Comparison.....	105
4.7.3: Results Comparison.....	106
4.7.4: Cluster Numbers Comparison.....	106
4.7.5: Matching the Clusters.....	107
4.8: Conclusions.....	111
<b>Chapter 5: A Classification of Census Output Areas.....</b>	<b>113</b>
5.1: Introduction.....	113
5.2: Introducing Output Areas: the Objects to be Clustered.....	114
5.2.1: Output Area Design.....	114
5.3: Inputs.....	117
5.3.1: Variable Selection.....	117
5.3.2: Initial Set of Variables Considered.....	117
5.3.3: Reducing the Initial Set of Variables.....	119
5.3.4: Highly Correlated Variables.....	119
5.3.5: Variables with Badly Behaved Distributions.....	121
5.3.6: Composite Variables.....	122
5.3.7: Geographic Consistency of Variables.....	123
5.3.8: Vague or Uncertain Variables.....	123
5.3.9: Uninteresting Geographic Distribution of Data.....	124
5.3.10: Consistency of the Variable for the Life Time of the Classification.....	125
5.3.11: Standardisation of Limiting Long Term Illness.....	125
5.3.12: The Process of Variable Selection.....	126
5.3.13: The Final List of 41 Variables Used in the Classification.....	130
5.3.14: Weighing of Variables.....	132
5.3.15: Database Assembly.....	132

5.3.16: A Program Used to Extract the Variables.....	134
5.3.17: Data Checking.....	134
5.4: Processes.....	137
5.4.1: Variable Standardisation.....	137
5.4.2: The Hierarchy that is to be Created.....	138
5.4.3: The Original Methodology.....	140
5.4.4: The Final Methodology: Creating a Hierarchical System Using K-means.....	145
5.4.5: Elucidating Log Transformation.....	146
5.4.6: The Creation of the Classification.....	148
5.5: Outputs.....	152
5.5.1: Naming and Describing the Clusters.....	152
5.5.2: Cluster Names.....	153
5.5.3: Cluster Portraits.....	156
5.5.4: Other Outputs.....	160
5.5.5: Mapping the Classification.....	161
5.5.6: Visualising the Classification in Alternative Ways.....	162
5.6: Conclusions.....	168
<b>Chapter 6: Quality Assuring and Adding Value to the OA Classification.....</b>	<b>169</b>
6.1: Introduction.....	169
6.2: Variability Reduction from Clustering and the Power of Each Variable.....	169
6.2.1: Sensitivity Analysis.....	170
6.2.2: Variability Reduction from the Clustering Process.....	173
6.3: Variability Reduction Within the Cluster Hierarchy.....	178
6.3.1: Within Cluster Distance.....	178
6.3.2: Between Cluster Centre Distance.....	180
6.4: Atypicality.....	184
6.5: Ground Truthing.....	189
6.6: Fuzzyfying the Classification: Distance to All Cluster Centres.....	191
6.7: The Consultation Exercise.....	197
6.7.1: Method of Consultation.....	197
6.7.2: Consultation Results.....	199
6.7.3: Statistical Results.....	200
6.7.4: Feedback on the Classification.....	201
6.7.5: General Comments.....	201
6.7.6: Teaching Resource.....	202
6.7.7: Life Course and Change Over Time.....	202
6.7.8: Fuzzy Classification.....	203
6.7.9: Names of Clusters.....	203
6.7.10: Preconceptions and Idealisation of Home.....	204
6.7.11: More Difficult than First Thought.....	204
6.7.12: OA Boundaries and Disclosure Control.....	206
6.7.13: Map Colours.....	206
6.7.14: Distinguishing the Function of an Area from its Residential Pattern.....	207
6.7.15: Conclusions and Action Points Resulting from the Consultation.....	208
6.8: Conclusions.....	209
<b>Chapter 7: Testing the OA Classification: Accounting for behaviours and patterns ...</b>	<b>211</b>
7.1: Introduction.....	211
7.2: Focus on Leeds: the Geography of a Regional City.....	212
7.3: A Tale of Eight Cities: Profiling England's Core Cities Group.....	216
7.4: The Swingometer: Change in British Electoral Patterns 1997-2005.....	217
7.5: Out in the Country Air: Disaggregating the Urban-Rural Classification.....	222
7.6: How do the OA Classes Score in Terms of Deprivation?.....	226
7.7: Beth yw hwn y Gymraeg? The Geography of the Welsh Language.....	227
7.8: Demography Across the Divide: Religious Tensions in Northern Ireland.....	230
7.9: Accounting for Migration Rates Across the UK.....	236

7.10: Reopening the North-South Divide.....	241
7.11: Conclusions.....	244
<b>Chapter 8: A Multi-Scale Integrated Classification System: Investigating Diversity within Area Classifications .....</b>	<b>246</b>
8.1: Introduction.....	246
8.2: Why do we need a Multi-scale Classification System? .....	246
8.3: The Importance of Scale in Assessing Diversity.....	249
8.4: The ONS Local Authority and Ward Classifications.....	250
8.4.1: The ONS Local Authority Classification.....	250
8.4.2: The ONS Statistical Ward Classification.....	252
8.5: Linking the Levels.....	254
8.6: Diversity within the Levels.....	254
8.6.1: Wards within Local Authorities.....	255
8.6.2: OAs within Wards.....	258
8.6.3: OAs within Local Authorities.....	260
8.6.4: Summarising Diversity within the Levels.....	262
8.7: Diversity within the Output Area Classification.....	263
8.7.1: Investigations into the Diversity of Household Types within OAs.....	264
8.8: Conclusions: is there Diversity within Area Classifications? .....	267
<b>Chapter 9: Conclusions: The Way Forward for a Newly Classified Nation.....</b>	<b>270</b>
9.1: Introduction.....	270
9.2: Summary of Research Findings.....	270
9.3: Publication of the Output Area Classification.....	273
9.4: Applications of the classification System.....	278
9.5: Limitations of the Research.....	280
9.6: (Endless) Possibilities for Future Research.....	282
9.6.1: Mapping the Life Course through the OA Classification.....	282
9.6.2: Changing Residential Patterns of the UK 1991 – 2001.....	283
9.6.3: Household Classification.....	285
9.6.4: Super Output Area Classifications.....	287
9.6.5: Expansion Beyond the UK.....	288
9.6.6: Specific Purpose Classifications.....	288
9.6.7: Regional Classifications.....	289
9.7: Concluding Statement.....	289
References.....	290
Appendix A: Full membership list for the Local Authority Classification.....	304
Appendix B: FORTRAN code for Census data extraction program.....	308
Appendix C: Consultation exercise documents.....	313
Appendix D: Percentage of population in each OA Classification cluster.....	320
Appendix E: Further maps of OA classification.....	321
Appendix F: OA Classification on CD.....	330

## List of Figures

Figure 2.1: The ‘Curry Mile’ by night.....	13
Figure 2.2: Map of the Curry houses on Wilmslow Road.....	13
Figure 2.3: The geography of the 2004 US presidential election results (48 contiguous states).....	14
Figure 2.4: An extract of Camden from Charles Booth’s poverty map of London.....	18
Figure 2.5: The administrative geography of the United Kingdom.....	29
Figure 2.6: Sample dataset to illustrate the effect of aggregation on areal units.....	31
Figure 2.7: An illustration of the effect of data aggregation of socio-economic data.....	31
Figure 2.8: Advert for CACI’s ACORN in the People Newspaper.....	38
Figure 2.9: Visual and verbal portrait of MOSAIC: Group C19 Original Suburbs.....	39
Figure 3.1: The sequence of Principal Component Analysis.....	57
Figure 4.1: The percentage of each domain type used in the reviewed classifications.....	75
Figure 4.2: The percentage of each domain type used in the each of the reviewed classifications.....	76
Figure 4.3: The range of each domain type used in the reviewed classifications.....	76
Figure 4.4: The changing prevalence of each domain type with increasing number of variables.....	78
Figure 4.5: Correlation matrix of age variables.....	84
Figure 4.6: Increase in distance between the most dissimilar LA within merged clusters.....	93
Figure 4.7: Cluster Profile of Family A: Urban UK.....	96
Figure 4.8: Cluster Profile of Family B: Rural UK.....	96
Figure 4.9: Cluster Profile of Family C: Prosperous Britain.....	96
Figure 4.10: Cluster Profile of Family D: Urban London.....	97
Figure 4.11: Cluster Profile of Family E: Northern Irish Heartlands.....	97
Figure 4.12: Map of the Five Families of the Local Authority classification.....	102
Figure 4.13: Map of the 13 Groups of the Local Authority classification.....	103
Figure 4.14: Map of the 26 Classes of the Local Authority classification.....	104
Figure 4.15: Crosstabulation of the Leeds families with the ONS super-groups.....	108
Figure 4.16: Crosstabulation of the Leeds groups with the ONS groups.....	109
Figure 4.17: Crosstabulation of the Leeds classes with the ONS sub-groups.....	110
Figure 5.1: shows OA boundaries for an area of Leeds overlain on OS 1:10,000 mapping....	116
Figure 5.2: The size of observed and expected cluster sizes (when creating 1000 clusters using the k-means procedure).....	141
Figure 5.3: The intricacies of Ward’s Hierarchical Clustering Procedure.....	144
Figure 5.4: The creation of a hierarchical system using the k-means algorithm.....	145
Figure 5.5: An illustration of the inherent problem of clustering cluster centres.....	146
Figure 5.6: the effect of logarithmic transformation on a dataset.....	147
Figure 5.7: Average distance from cluster centre for different values of k.....	149
Figure 5.8: The range in the size of clusters when choosing the number of clusters.....	151
Figure 5.9: Cluster summary of super-group 1: Blue Collar Communities.....	157
Figure 5.10: Cluster summary of super-group 2: City Living.....	157
Figure 5.11: Cluster summary of super-group 3: Countryside.....	158
Figure 5.12: Cluster summary of super-group 4: Prospering Suburbs.....	158

Figure 5.13: Cluster summary of super-group 5: Constrained by Circumstances.....	159
Figure 5.14: Cluster summary of super-group 6: Typical Traits.....	159
Figure 5.15: Cluster summary of super-group 7: Multicultural.....	160
Figure 5.16: Variable by cluster graph using the original data for variable 20 (All Flats)....	161
Figure 5.17: Mapping the OA Classification at the super-group level for the whole UK using centroids.....	163
Figure 5.18a: Mapping the OA classification at super-group level using boundaries for London and surrounding area overlaid on a UK settlement map.....	164
Figure 5.18b: Mapping the OA classification at super-group level using centroids for London and surrounding area overlaid on a UK settlement map.....	165
Figure 5.19: Mapping the OA classification at super-group level using centroids for Leeds and surrounding area overlaid on 1:50,000 Ordnance survey Mapping.....	166
Figure 5.20: Mapping the OA classification at super-group level using centroids for Selby and surrounding area overlaid on SPOT satellite image.....	167
Figure 6.1: Sensitivity analysis results: showing the effect of removing each variable from the classification on the average distance from cluster centre .....	171
Figure 6.2: The frequency distribution of the distances from cluster centre at the three levels of the classification.....	179
Figure 6.3: The difference between cluster centres at the Super-group level.....	181
Figure 6.4: the difference between cluster centres at the Group level.....	182
Figure 6.5: the difference between cluster centres at the Sub-group level.....	183
Figure 6.6: The location of atypical areas in Leeds.....	186
Figure 6.7: Photographs representing the classification at super-group level.....	190
Figure 6.8: Distance from super-group 1 Blue Collar Communities (Top map depicts whole UK, bottom map depicts Leeds Metropolitan District).....	192
Figure 6.9: Distance from super-group 2 City Living (Top map depicts whole UK, bottom map depicts Leeds Metropolitan District).....	192
Figure 6.10: Distance from super-group 3 Countryside (Top map depicts whole UK, bottom map depicts Leeds Metropolitan District).....	193
Figure 6.11: Distance from super-group 4 Prospering Suburbs (Top map depicts whole UK, bottom map depicts Leeds Metropolitan District).....	193
Figure 6.12: Distance from super-group 5 Constrained by Circumstances (Top map depicts whole UK, bottom map depicts Leeds Metropolitan District).....	194
Figure 6.13: Distance from super-group 6 Typical Traits (Top map depicts whole UK, bottom map depicts Leeds Metropolitan District).....	194
Figure 6.14: Distance from super-group 7 Multicultural (Top map depicts whole UK, bottom map depicts Leeds Metropolitan District).....	195
Figure 6.15: A comparison of the <i>ad hoc</i> fuzzy classification with actual membership of the cluster for super-group 5 Constrained by Circumstances, for the Leeds Metropolitan District.....	196
Figure 6.16: Example map used in consultation.....	199
Figure 6.17: The location of the maps looked at by the respondents to the consultation overlain.....	199
Figure 7.1: A Map of Leeds showing the distribution of Super-groups by Community Areas.....	213
Figure 7.2: The location of the Community Areas that contain the highest proportion of each Super-group.....	215
Figure 7.3: The percentage of each Super-Group in the eight Core Cities.....	216

Figure 7.4: The 2001 General Election results for England and Wales assigned to OAs and split by Super-groups.....	219
Figure 7.5: The 2005 General Election results for England and Wales assigned to OAs and split by Super-groups.....	220
Figure 7.6: The 2001-2005 change in the voting pattern in the UK General Election in England and Wales assigned to OAs and split by Super-groups.....	221
Figure 7.7: The percentage of each OA Classification group type which is in each urban-rural classification type.....	224
Figure 7.8: The percentage of each urban-rural classification type that is in each OA Classification group type.....	225
Figure 7.9: The IMD 2004 ranks (SOA) aggregated by the OA Classification at super-group level (Ranks used rather than scores to allow comparability between the different domains).....	226
Figure 7.10: Knowledge of the Welsh language according to the 2001 Census.....	228
Figure 7.11: OA Classification at Super-group level for Wales.....	228
Figure 7.12: The distribution of Welsh speakers by Super-group.....	229
Figure 7.13: Front age of the Belfast Telegraph 19 <sup>th</sup> December 2002.....	231
Figure 7.14: England & Wales religion question.....	232
Figure 7.15: Northern Ireland religion question.....	232
Figure 7.16: The distribution of Catholics in Northern Ireland, 2001 Census OA level.....	233
Figure 7.17: The Distribution of Religious Background in Northern Ireland by Super-Group.....	234
Figure 7.18: The percentage of each super-group type by Catholic deciles.....	235
Figure 7.19: The percentage of each super-group type by Protestant deciles.....	236
Figure 7.20: Profiling UK migration rates against the Super-group level of the classification.....	238
Figure 7.21: Profiling UK migration rates against the Group level of the classification.....	239
Figure 7.22: Profiling UK migration rates against the Sub-group level of the classification..	239
Figure 7.23: The distribution of migration rate deciles by OA classification super-group type.....	240
Figure 7.24: The north and south stratified by the OA Classification.....	242
Figure 7.25: The London and the Archipelago stratified by the OA Classification.....	243
Figure 8.1: What constitutes diversity within areas.....	250
Figure 8.2: A Map of the ONS Local Authority Classification at the super-group level.....	251
Figure 8.3: A Map of the ONS Statistical Ward Classification at the super-group level.....	253
Figure 8.4: Section from the database linking the classifications at different scales.....	254
Figure 9.1: The OA Classification on National Statistics Online.....	274
Figure 9.2: The OA Classification on the University of Leeds website.....	275
Figure 9.3: The download page for the OA Classification on UK Borders.....	276
Figure 9.4: Presentation of award for the OA Classification from the Demographic User Group, titled 'best new information from government'.....	277
Figure 9.5: The distribution of YELL businesses by OA Classification sub-group.....	278
Figure 9.6: Take up of promotional energy efficiency measures OA Classification group...	279

---

**List of Tables**

Table 1.1: Thesis Outline .....	8
Table 2.1: List of British/UK geodemographic systems currently available .....	23
Table 2.2: An example of ecological fallacy using 2001 Census data .....	30
Table 2.3: The effect of data aggregation on the sample dataset (as shown in figure 2.7)...	32
Table 4.1: the variation in size of the UK's LAs in terms of population and area.....	73
Table 4.2: Sources and dates of the ten reviewed classifications .....	74
Table 4.3: Variables which are used in at least three of the ten reviewed classifications ...	77
Table 4.4: The comprehensive list of 129 variables considered for use in the LA Classification.....	80
Table 4.5: First 10 Rows of component I.....	83
Table 4.6: First 10 Rows of component II.....	83
Table 4.7: First 10 Rows of component III.....	83
Table 4.8: Variables with the highest and lowest standard deviation across all LAs.....	85
Table 4.9: The list of Inclusion, Rejection or Merger of variables.....	86
Table 4.10: The final list of 56 variables to be used in the classification.....	90
Table 4.11: The Structure of Families, Groups and Classes.....	94
Table 4.12: Classes with that have the highest positive and negative values for each variable .....	99
Table 4.13: Distance to five most similar local authorities, for selected local authorities.....	100
Table 4.14: The distance in scale between the families, groups and classes in the two classification systems.....	107
Table 5.1: The average size of OAs in the constituent countries of the UK.....	115
Table 5.2: The initial set of variables considered for inclusion in the classification.....	118
Table 5.3: Highly Correlated Variables from Original Variable List (ordered by redundancy).....	120
Table 5.4: Percentage of ethnic group in London, W Midlands, G Manchester & West Yorkshire.....	124
Table 5.5: Full list of 41 variables selected for input to the classification.....	131
Table 5.6: Full variable definitions and sources (specified in terms of Key Statistics Table and column number).....	133
Table 5.7: Example of the Data checking results.....	135
Table 5.8: Example of the Data checking results (for the North East GOR).....	136
Table 5.9: The number of clusters in the most commonly used classification systems.....	139
Table 5.10: Number of OAs in each cluster based on the original methodology: attempt 1	142
Table 5.11: Number of OAs in each cluster based on the original methodology (ranks)....	143

---

---

Table 5.12: Number of OAs in each cluster based on the original methodology (logs).....	143
Table 5.13: The effect of logging data on the distribution of the data.....	148
Table 5.14: Example of method of calculating which solution is most homogenous in terms number of members in each cluster.....	150
Table 5.15: Number of OAs in each cluster based on the final methodology.....	152
Table 5.16: The Cluster Names.....	155
Table 6.1: Reduction in variability for each variable in the classification (ordered by largest reduction at sub-group level).....	175
Table 6.2: Reduction in variability for each variable in previous classification systems (ordered by largest reduction for GB Profiles).....	177
Table 6.3: The average reduction in variability for comparable variables.....	178
Table 6.4: The reduction of within cluster distance (figures refer to average Squared Euclidian distance of OAs from the cluster centre).....	180
Table 6.5: Number of Outliers (defined as Squared Euclidian distance of 1.5 or more away from cluster centre) by Government Office Region or Country.....	187
Table 6.6: Number of Outliers (defined as Squared Euclidian distance of 1.5 or more away from cluster centre) by Super-group.....	188
Table 6.7: Cross-tabulation of Outliers (defined as Squared Euclidian distance of 1.5 or more away from cluster centre) by Super-group and Government Office Region/Country.....	189
Table 6.8: The differences between the final set of names and those used in the consultation.....	199
Table 6.9: Percentage of respondents who identified the super groups correctly.....	200
Table 6.10: The super groups in which respondents live.....	201
Table 7.1: The diversity of OA types within selected community areas.....	214
Table 7.2: The Community Areas that contain the highest proportion of each super-group	215
Table 8.1: The Ward Classification Super-group types that make up the LA Classification Families (percent of wards by type for LA types).....	256
Table 8.2: The OA Classification Super-group types that make up the Ward Classification Super-groups (percent of OAs by type for ward types).....	258
Table 8.3: The OA Classification Super-group types that make up the LA Classification Families (percent of OAs by type for LA types).....	260
Table 8.4: The household types that make up the OA Classification Super-groups (percent of household types by OA type).....	264

---

---

**Abbreviations**

ACORN	A Classification of Residential Neighbourhoods
BBC	British Broadcasting Corporation
BMRB	British Market Research Bureau
BCS	British Crime Survey
CACI	Consolidated Analysis Centers Inc .
CAMS	Controlled Access Microdata Sample
CASE	Collaborative Awards in Science and Engineering
CASWEB	Census Area Statistics Website
CCG	Centre for Computational Geography
CHCC	Collection of Historical and Contemporary Census Data and Related Materials
CIA	Central Intelligence Agency
Con	Conservative Party
DEFRA	Department of the Environment, Food and Rural Affairs
DfES	Department for Education and Skills
DUG	Demographic User Group
DVLA	Driver and Vehicle Licensing Agency
E&W	England and Wales
ED	Enumeration District
EDINA	Edinburgh Data and Information Access
ESRC	Economic and Social Research Council
ESRI	Environmental Systems Research Institute
EuroStat	European Office for Statistics
EU	European Union
GB	Great Britain
GIGO	garage in, garbage out
GIS	Geographical Information System
GOR	Government Office Region
GROS	General Register Office Scotland
HBoS	Halifax Bank of Scotland
HE	Higher Education
IMD	Indices of Multiple Deprivation
KS	Key Statistics
LA	Local Authority
Lab	Labour Party
LD	Liberal Democrats
LLTI	Limiting Long Term Illness

---

---

log	logarithm
LSE	London School of Economics & Political Science
MAUP	Modifiable Areal Unit Problem
MIMAS	Manchester Information and Associated Services (Manchester University)
NHS	National Health Service
NI	Northern Ireland
NISRA	Northern Ireland Statistics Research Agency
NUTS	Nomenclature of Units for Territorial Statistics
OA	Output Area
ODPM	Office of the Deputy Prime Minister
ONS	Office for National Statistics
OPCS	Office of Population Censuses and Surveys
OS	Ordnance Survey
PAM	Partitioning Around Medoids
PC	Personal Computer
PCA	Principal Components Analysis
PDF	Postdoctoral Fellowship
RAC	Royal Automobile Club
RSG	Research Support Group
S4C	Sianel Pedwar Cymru
SARs	Sample of Annonomised Records
SDF	Social Democratic Federation
SDRC	Social Disadvantage Research Centre
SIR	Standardised Illness Ratio
SOA	Super Output Area
SPSS	Statistical Package for the Social Sciences
TGI	Target Group Index
UK	United Kingdom of Great Britain and Northern Ireland
UN	United Nations
US	United States of America
v	Variable

---