Open Geodemographics

The Creation of the Output Area Classification

Dr Dan Vickers
Department of Geography, University of Sheffield

d.vickers@sheffield.ac.uk

http://www.sasi.group.shef.ac.uk/

Market Research Society Conference

Monday 27th November 2006

The University Of Sheffield.

Social and Spatial Inequalities Research Group

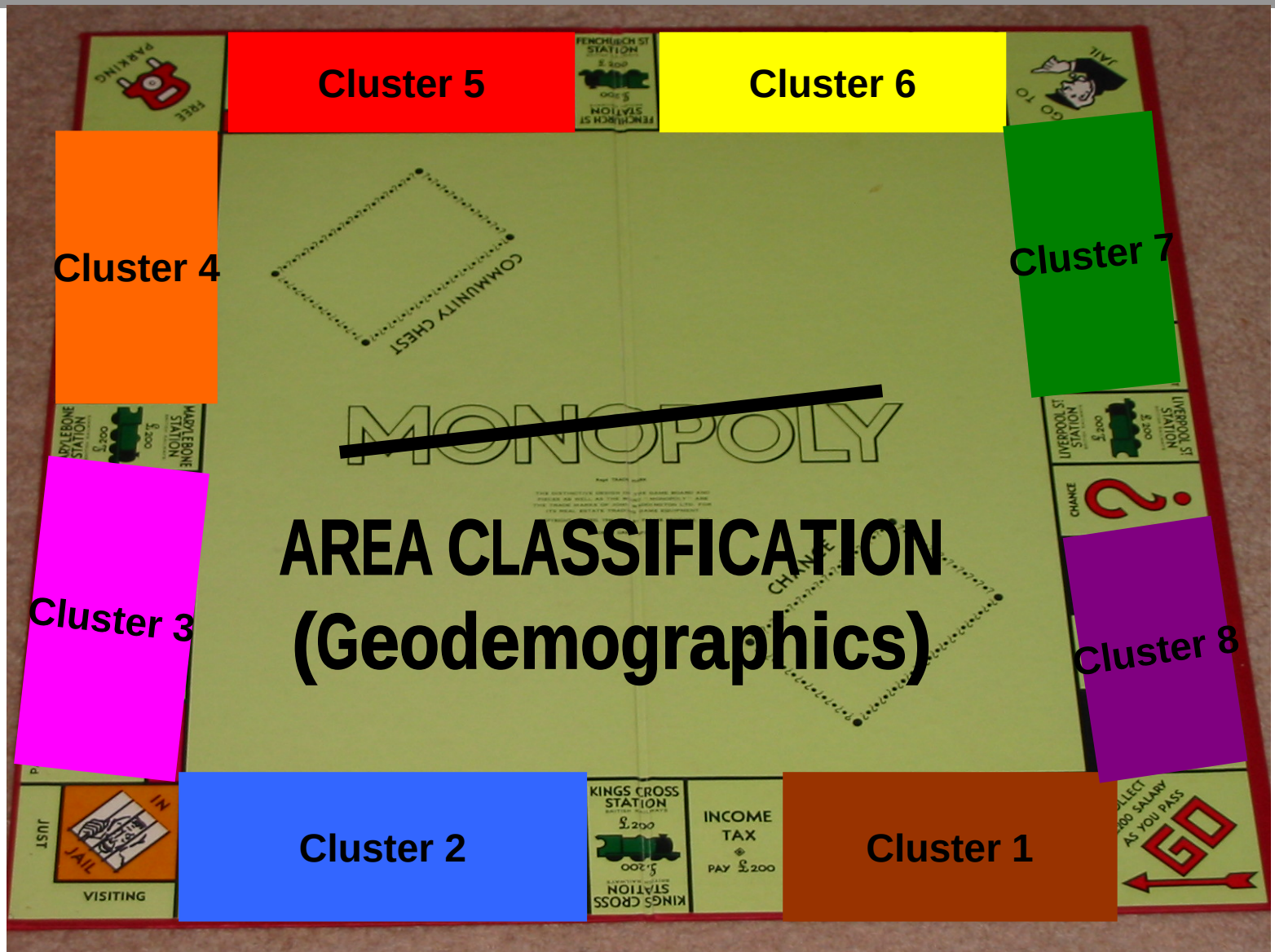
Open Geodemographics

Simplfying the world

"All the real knowledge which we possess depends on methods by which we distinguish the similar from the dissimilar. The greater number of natural distinctions this method comprehends the clearer becomes our idea of things. The more numerous the objects which employ our attention the more difficult it becomes to form such a method and the more necessary." (Linnæus 1737)

Open Geodemographics

What is Geodemographics?



Open Geodemographics

What is Geodemographics?

Geodemographics *"The analysis of people by where they live"* Sleight 1993

Open *"allowing access, passage, or view; not closed, fastened or restricted"* Oxford English Dictionary

Why Open?

- Geodemographics has been developed as a marketing tool.
- Made behind closed doors due to commercial confidentiality.
- Little released about how the systems are put together.
- Some myths circulated about what makes a good system.

Open Geodemographics

The Seven Steps of Cluster Anlysis

1. Clustering elements (objects to cluster, also known as “operational taxonomic units”)
2. Clustering variables (attributes of objects to be used)
3. Variable standardisation
4. Measure of association (proximity measure)
5. Clustering method
6. Number of clusters
7. Interpretation, testing and replication

(adapted from Milligan 1996)

Milligan, G. W. (1996) Clustering validation: Results and implications for applied analyses. in Arabie, P., Hubert, L. J. and De Soete, G. Eds., *Clustering and Classification*. Singapore: World Scientific.

Open Geodemographics

What Goes into the Classification

- Output Areas are the smallest area for general census output.
- 223,060 in the UK
- England & Wales
  - Number of OAs: 174,434
  - Min size: 40 households, 100 people
  - Mean size: 124 households, 297 people
- Scotland
  - Number of OAs: 42,604
  - Min size: 20 households, 50 people
  - Mean size: 52 households, 119 people

Northern Ireland

- Number of OAs: 5,022
- Min size: 40 households, 100 people
- Mean size: 125 households, 336 people

Open Geodemographics

What Goes into the Classification

41 Census Variables covering:

- **Demographic attributes**
  - Including - age, ethnicity, country of birth and population density
- **Household composition**
  - Including - living arrangements, family type and family size.
- **Housing characteristics**
  - Including - tenure , type & size, and quality/overcrowding
- **Socio-economic traits**
  - Including - education, socio-economic class, car ownership & commuting and health & care.
- **Employment attributes**
  - Including - level of economic activity and employment class type.

How many data inputs are involved?

223,060 Output Areas, 41 Variables = 9,145,460 data points

Open Geodemographics

Standardising the Variables

- Log Transformation
  - Reduce the effect of extreme values

- Range Standardisation (0-1)
  - Problems will occur if there are differing scales or magnitudes among the variables. In general, variables with larger values and greater variation will have more impact on the final similarity measure. It is necessary therefore to make each variable equally represented in the distance measure by standardising the data.

Open Geodemographics

Clustering the Data

K-means is an iterative relocation algorithm based on an error sum of squares measure. **The basic operation of the algorithm is to move a case from one cluster to another to see if the move would improve the sum of squared deviations within each cluster** (Aldenderfer and Blashfield, 1984). **The case will then be assigned/re-allocated to the cluster to which it brings the greatest improvement.** The next iteration occurs when all the cases have been processed. **A stable classification is reached when no moves occur during a complete iteration of the data.** After clustering is complete, it is then possible to examine the means of each cluster for each dimension (variable) in order to assess the distinctiveness of the clusters (Everitt *et al*., 2001).

Open Geodemographics

Issues of Cluster Number Selection

When choosing the number of clusters to have in the classification there were three main issues which need to be considered.

- Issue 1: **Analysis of average distance from cluster centre** for each cluster number option. The ideal solution would be the number of clusters which gives smallest average distance from the cluster centre across all clusters.

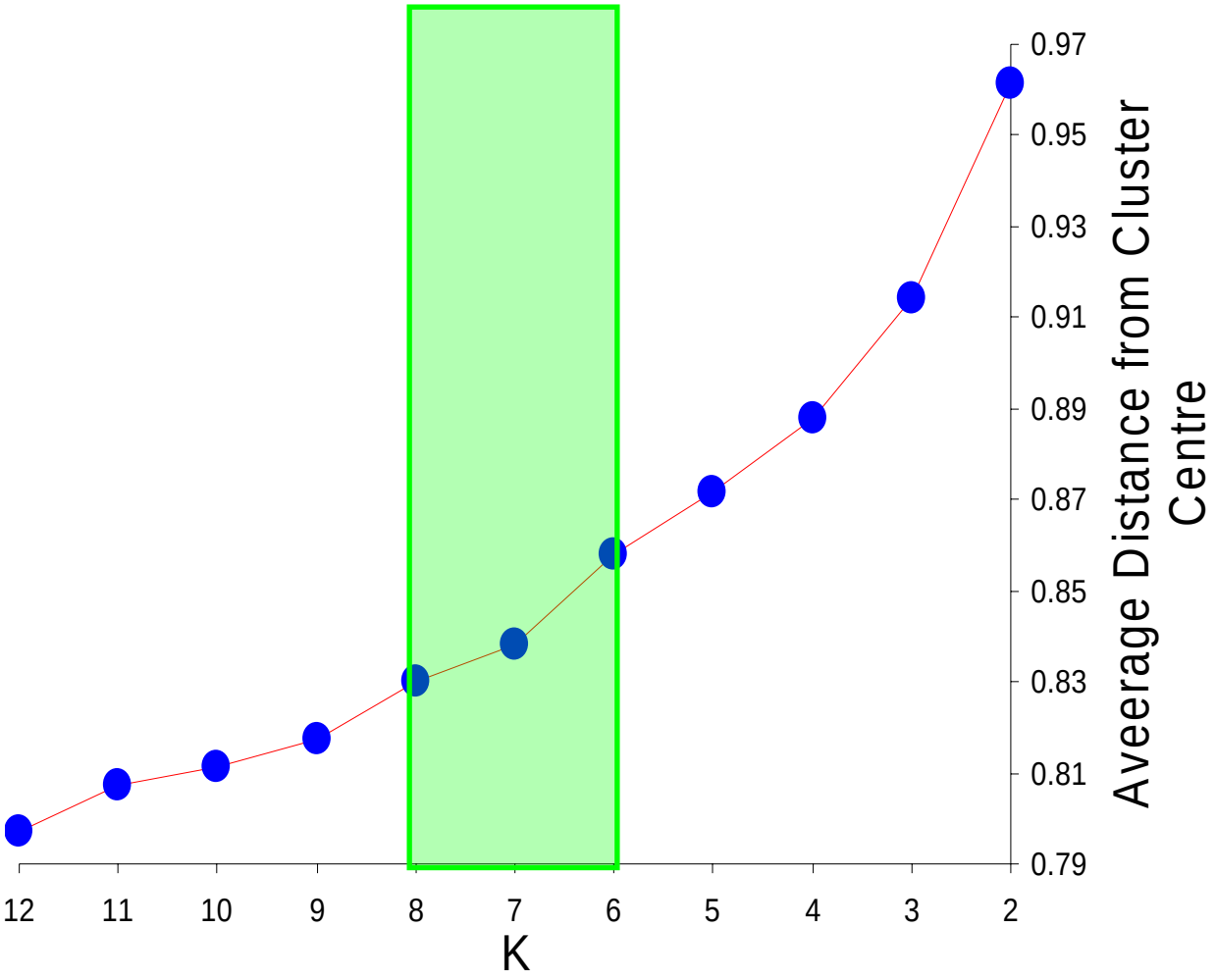
- Issue 2: **Analysis of cluster size homogeneity** for each cluster number option. It would be useful, where possible, to have clusters of as similar size as possible in terms of the number of members within each.

- Issue 3: The number of clusters produced should be **as close to the perceived ideal as possible**. This means that the number of clusters needs to be of a size that is useful for further analysis.

Open Geodemographics

Issues of Cluster Number Selection

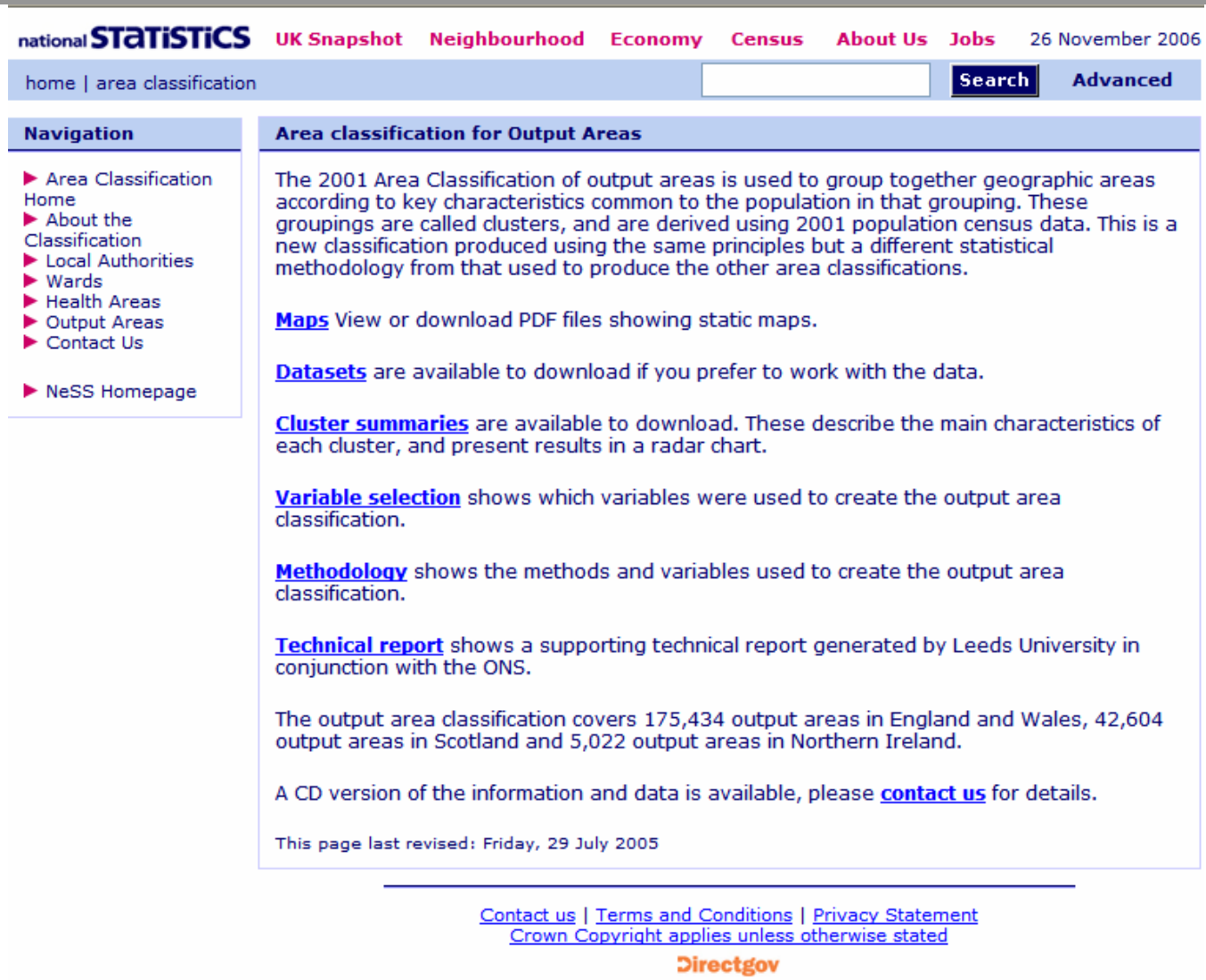
A three tier hierarchy 7, 21 & 52 clusters



- First Level target 6, 7 selected based on analysis of, average distance from cluster centre and size of each cluster.
- Second Level target 20, 21 selected based on analysis of, average distance from cluster centre and size of each cluster.
- Third Level target 50, 52 selected based on size of each cluster. Split into either 2 or 3 groups

Open Geodemographics

Outputs



national STATISTICS UK Snapshot Neighbourhood Economy Census About Us Jobs 26 November 2006

home | area classification Search Advanced

Navigation

- Area Classification Home
- About the Classification
- Local Authorities
- Wards
- Health Areas
- Output Areas
- Contact Us

- NeSS Homepage

Area classification for Output Areas

The 2001 Area Classification of output areas is used to group together geographic areas according to key characteristics common to the population in that grouping. These groupings are called clusters, and are derived using 2001 population census data. This is a new classification produced using the same principles but a different statistical methodology from that used to produce the other area classifications.

Maps View or download PDF files showing static maps.

Datasets are available to download if you prefer to work with the data.

Cluster summaries are available to download. These describe the main characteristics of each cluster, and present results in a radar chart.

Variable selection shows which variables were used to create the output area classification.

Methodology shows the methods and variables used to create the output area classification.

Technical report shows a supporting technical report generated by Leeds University in conjunction with the ONS.

The output area classification covers 175,434 output areas in England and Wales, 42,604 output areas in Scotland and 5,022 output areas in Northern Ireland.

A CD version of the information and data is available, please **contact us** for details.

This page last revised: Friday, 29 July 2005

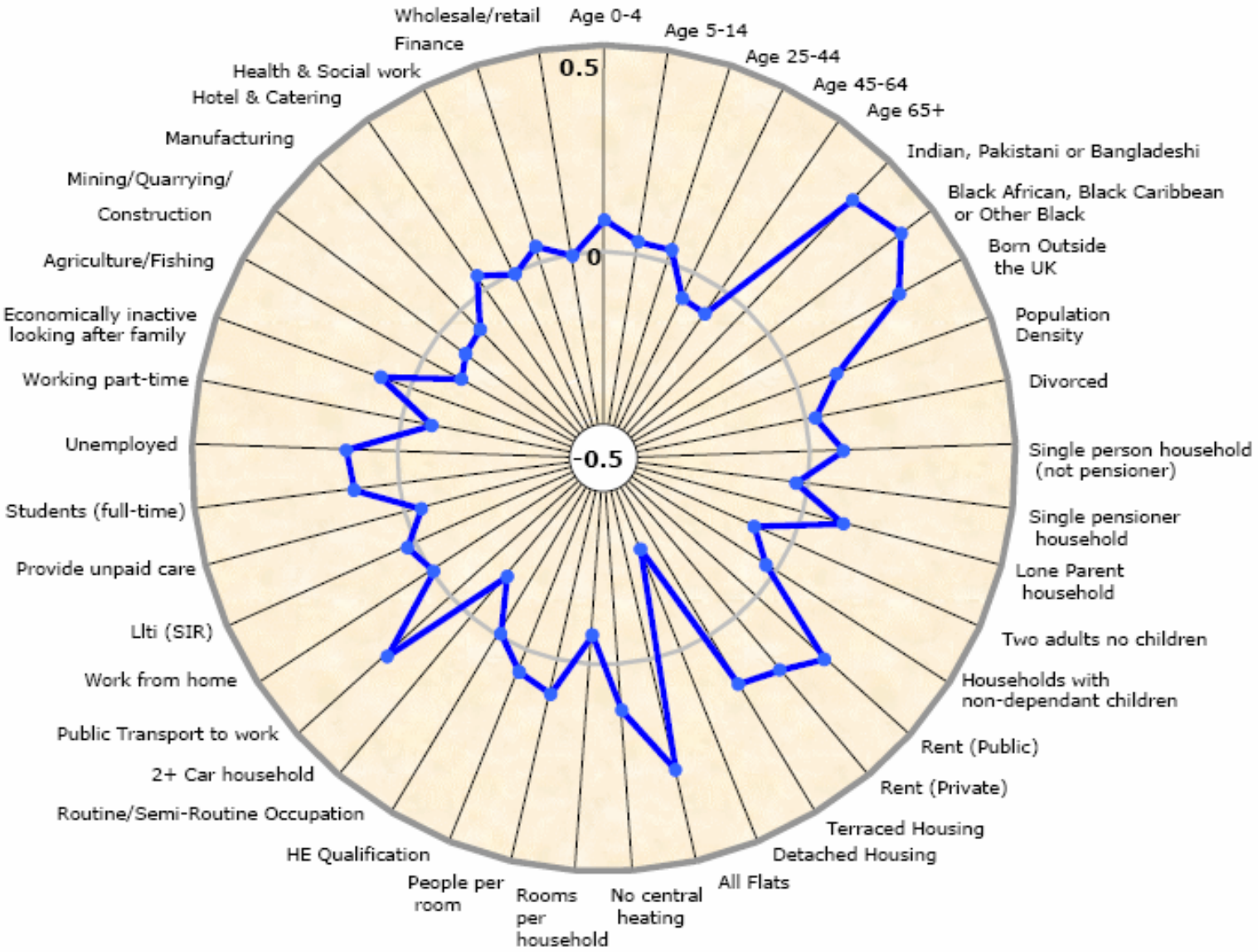
Contact us | Terms and Conditions | Privacy Statement
Crown Copyright applies unless otherwise stated

Directgov

http://www.statistics.gov.uk/about/methodology_by_theme/area_classification/oa/default.asp

Open Geodemographics

Cluster Profiling



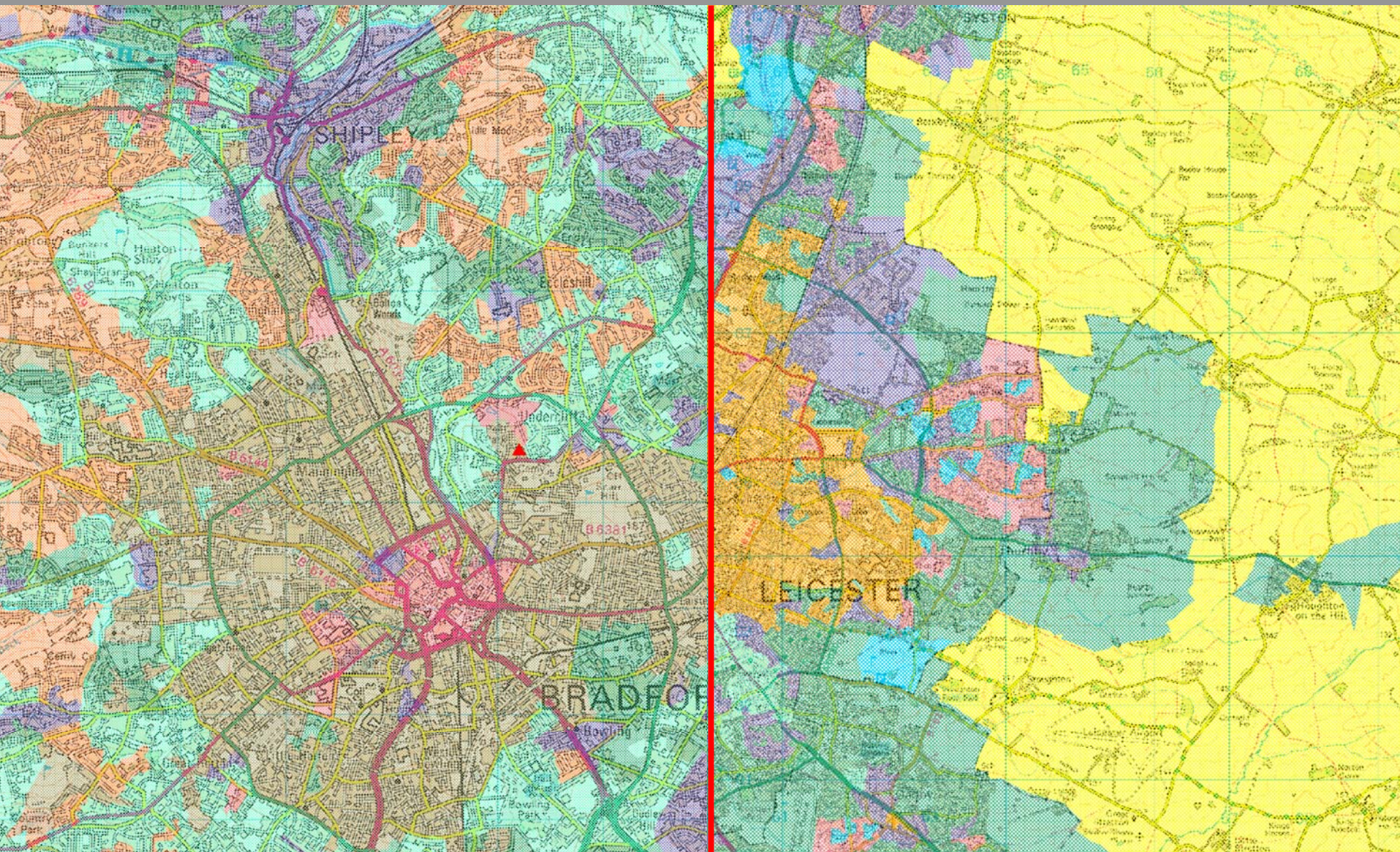
7: Multicultural

Open Geodemographics

	Cluster Names
1: BLUE COLLAR COMMUNITIES	1a: Terraced Blue Collar
	1b: Younger Blue Collar
	1c: Older Blue Collar
2: CITY LIVING	2a: Transient Communities
	2b: Settled in the City
3: COUNTRYSIDE	3a: Village Life
	3b: Agricultural
	3c: Accessible Countryside
4: PROSPERING SUBURBS	4a: Prospering Younger Families
	4b: Prospering Older Families
	4c: Prospering Semis
	4d: Thriving Suburbs
5: CONSTRAINED BY CIRCUMSTANCES	5a: Senior Communities
	5b: Older Workers
	5c: Public Housing
6: TYPICAL TRAITS	6a: Settled Households
	6b: Least Divergent
	6c: Young Families in Terraced Homes
	6d: Aspiring Households
7: MULTICULTURAL	7a: Asian Communities
	7b: Afro-Caribbean Communities

Open Geodemographics

Mapping the Classification



Open Geodemographics

Mapping the Classification

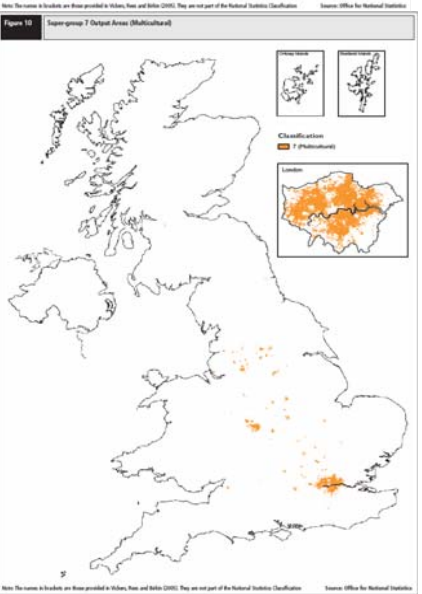
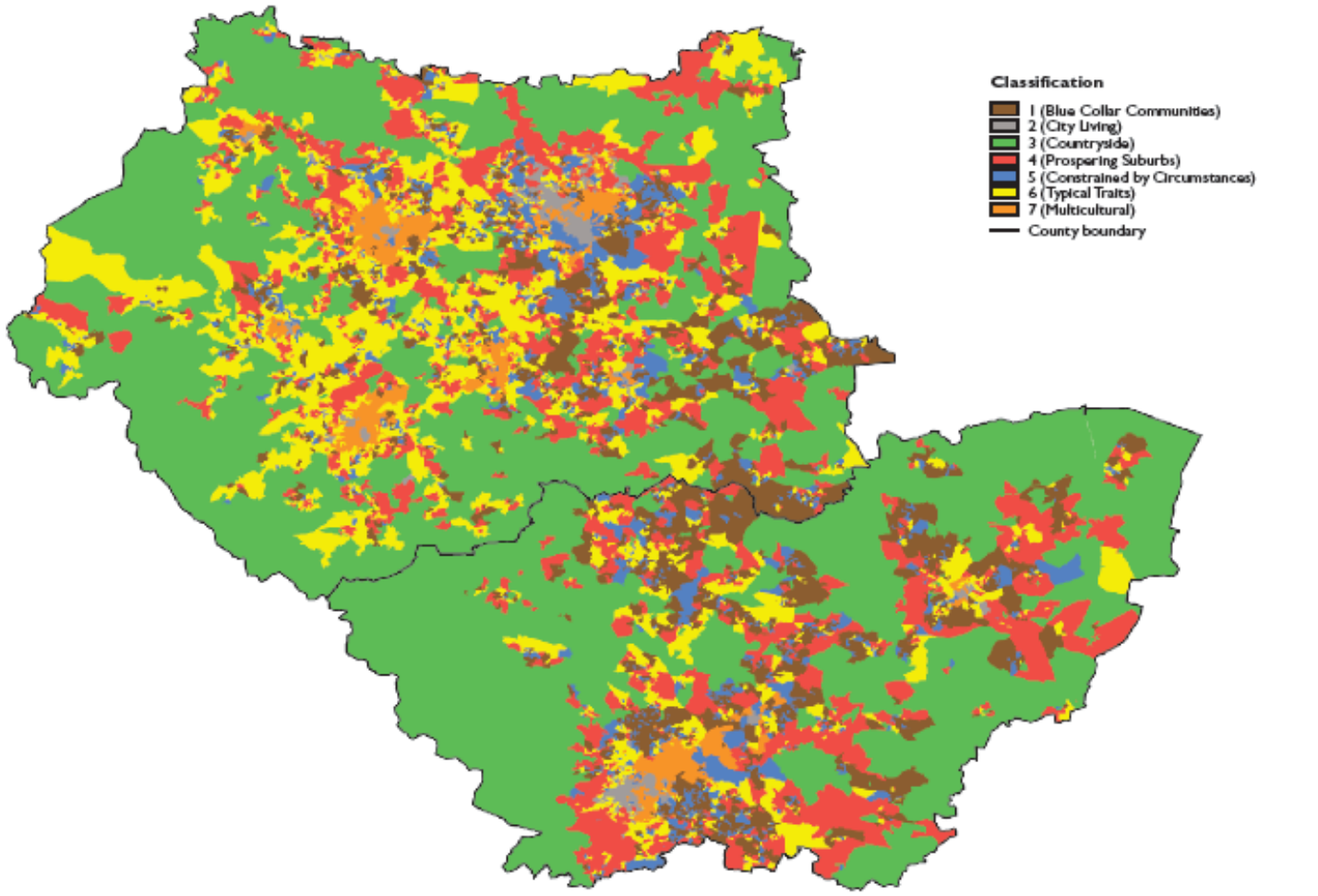


Figure 11 The Full Output Area Classification for South and West Yorkshire

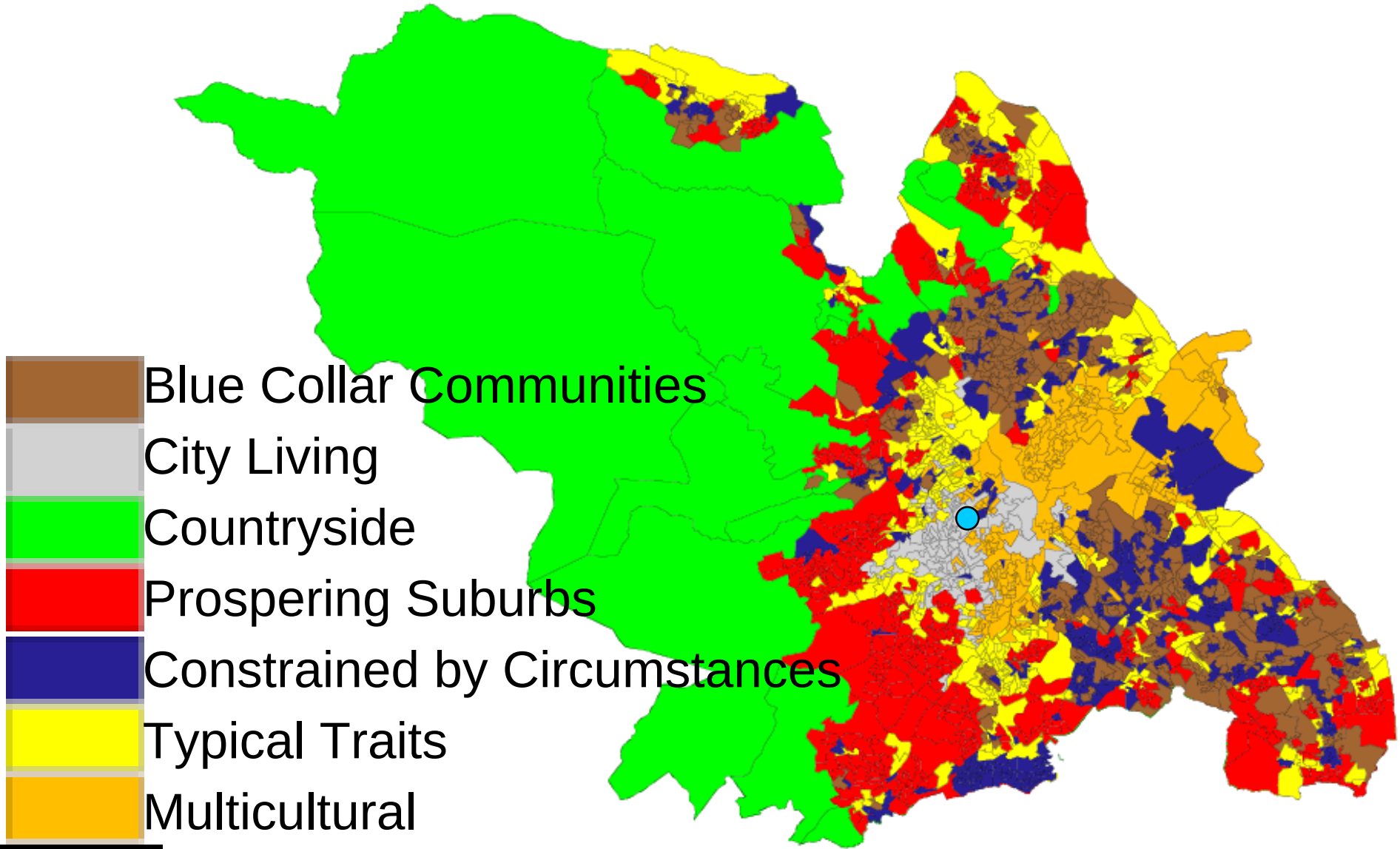


Note: The names in brackets are those provided in Vickers, Rees and Birkin (2005). They are not part of the National Statistics Classification

Source: Office for National Statistics

Open Geodemographics

What I tell the Students



Open Geodemographics

What I tell the Students

	Low		Council Tax Band				High
Cluster	A	B	C	D	E	F	G	H
Blue Collar Communities	84.8	10.1	2.9	1.5	0.6	0.1	0.0	0.0
City Living	46.9	23.4	12.3	7.7	5.3	2.5	1.6	0.2
Countryside	6.4	8.3	22.0	18.8	20.5	15.8	7.8	0.4
Prospering Suburbs	8.9	22.0	31.2	17.7	10.7	5.6	3.6	0.1
Constrained by Circumstances	89.8	5.4	2.5	1.3	0.7	0.2	0.1	0.0
Typical Traits	50.2	25.8	13.3	5.4	3.5	1.2	0.6	0.0
Multicultural	86.0	8.3	4.4	0.8	0.3	0.3	0.1	0.0

Open Geodemographics

Further Information

Web Data Resources:

- ONS Website: http://www.statistics.gov.uk/about/methodology_by_theme/area_classification/oa/default.asp
- Sheffield, SASI Website: http://www.sasi.group.shef.ac.uk/area_classification/index.html

Papers and Reports:

- Vickers, Rees and Birkin (2005) Creating the National Classification of Output Areas: http://www.geog.leeds.ac.uk/wpapers/05-2.pdf
- Vickers and Rees (2006), Introducing the Area Classification of Output Areas, Population Trends (125): http://www.statistics.gov.uk/downloads/theme_population/PT125_main_part2.pdf
- Vickers (2006), Multi-level Integrated Classifications Based on the 2001 Census, PhD Thesis, University of Leeds. http://www.geog.leeds.ac.uk/people/d.vickers/thesis.html