

# Creating a Geodemographic Classification

Dr Daniel Vickers

Department of Geography, University of Sheffield

[D.Vickers@sheffield.ac.uk](mailto:D.Vickers@sheffield.ac.uk)

[www.shef.ac.uk/sasi](http://www.shef.ac.uk/sasi)

[www.areaclassification.org.uk](http://www.areaclassification.org.uk)

# Creating a Geodemographic Classification

Why classify areas?



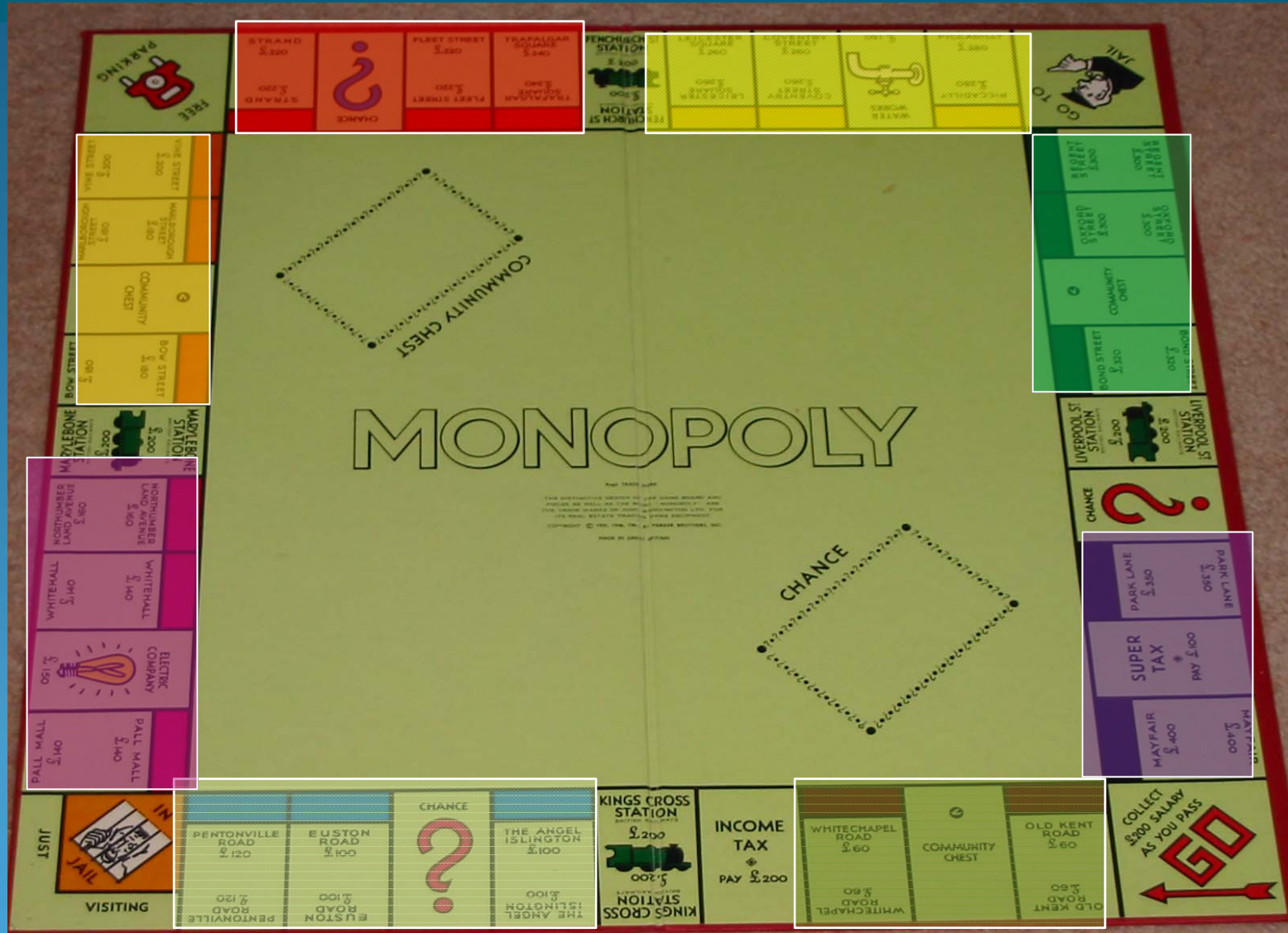
*“All the real knowledge which we possess depends on methods by which we distinguish the similar from the dissimilar. The greater number of natural distinctions this method comprehends the clearer becomes our idea of things. The more numerous the objects which employ our attention the more difficult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

(Linnæus 1737)



# Creating a Geodemographic Classification

Why classify areas?



# Creating a Geodemographic Classification

## The Myths of Geodemographics

- The more data the better?
- The smaller the areas the more meaningful?
- These are distinct groups?
- A true depiction of reality?

# Creating a Geodemographic Classification

## The Objective of Creating OAC



- To provide a publicly accessible classification, free at the point of use for researchers, policy makers and the public, avoiding the expense of commercial classifications.
- To provide a transparent and replicable methodology with the input data as well as the classification output.
- To stimulate value added research either using the classification or the raw data.

# Creating a Geodemographic Classification

## The Seven Steps of Cluster Analysis

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.

# Creating a Geodemographic Classification

## 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

# Creating a Geodemographic Classification

## 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



# Creating a Geodemographic Classification

## 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.

# Creating a Geodemographic Classification

## Clustering the Data

### K-means clustering

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 therefore 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).

# Creating a Geodemographic Classification

## 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 centres 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.

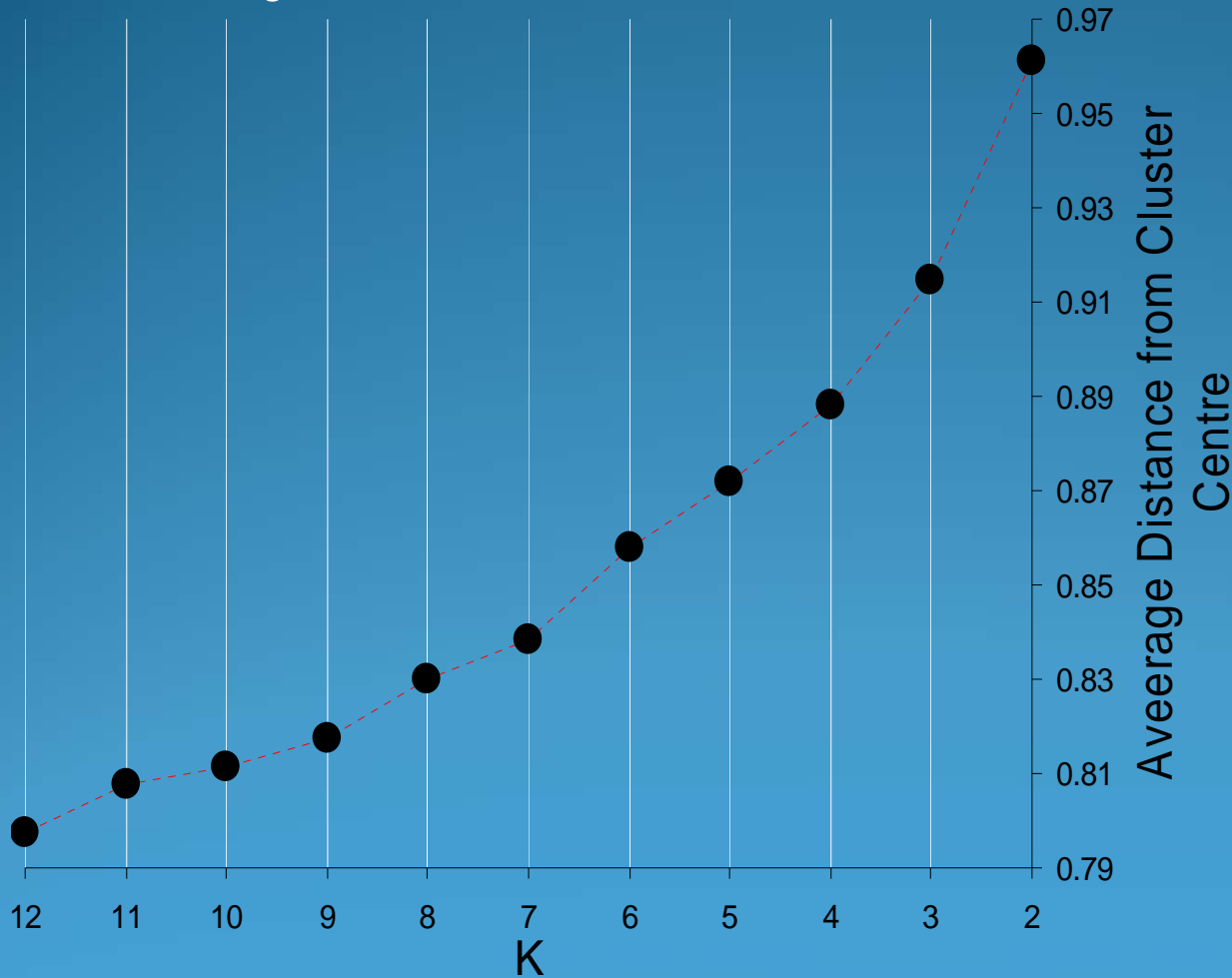
➤ **Issue 3:** The number of clusters produced should be as close to the perceived ideal as possible. This means that the number of clusters needs to be of a size that is useful for further analysis.

# Creating a Geodemographic Classification

## Issues of Cluster Number Selection

A three tier hierarchy 7, 21 & 52 clusters

Average Distance from Cluster Centre



➤ 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.

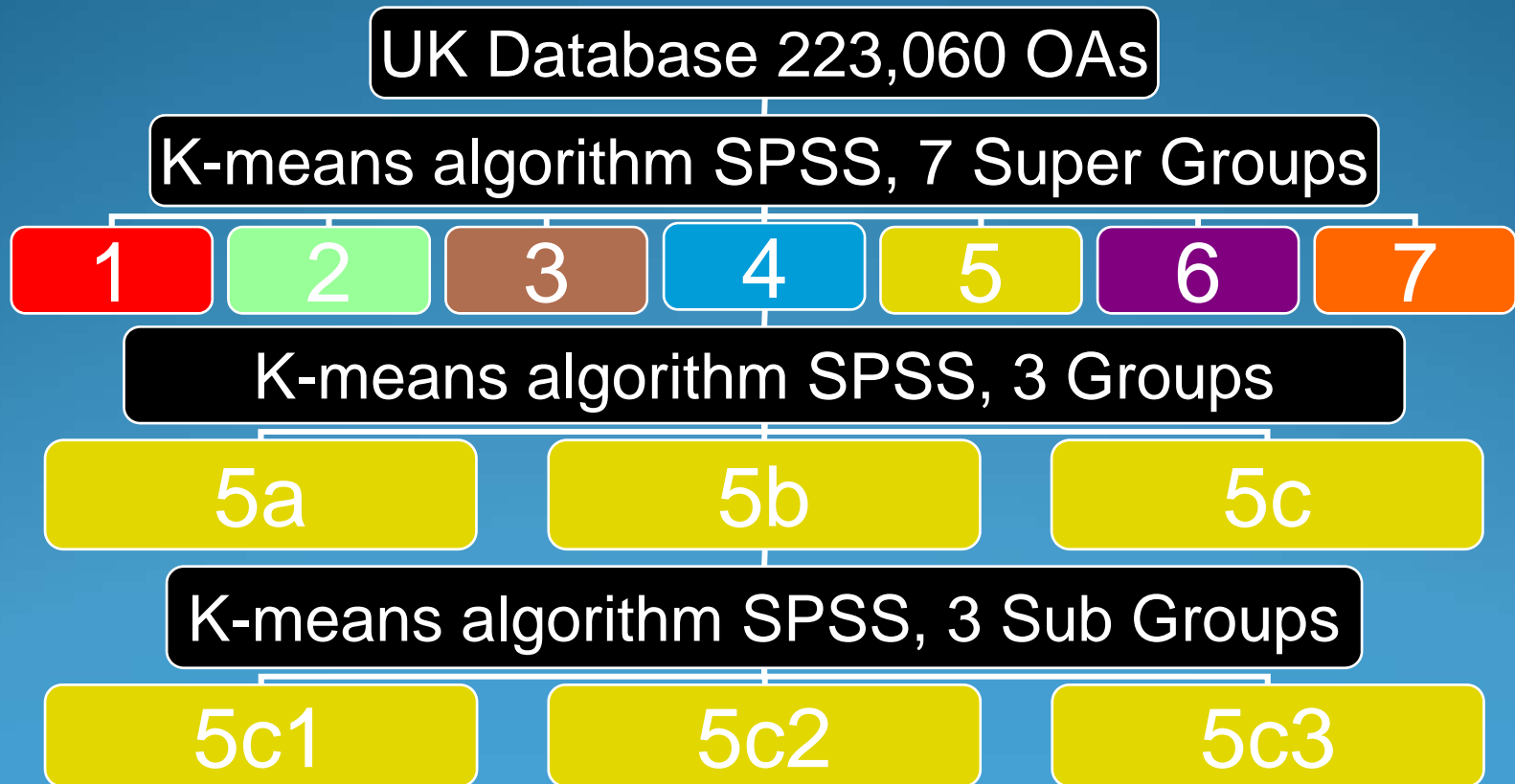
➤ Third Level target 50, 52 selected based on size of each cluster. Split into either 2 or 3 groups



# Creating a Geodemographic Classification

## Clustering: Using k-means to Create a Hierarchy

- First level run as standard k-means
- Second level, first level is split into separate files and each file is clustered separately
- Third level, second level is split into separate files and each file is clustered separately



# Creating a Geodemographic Classification

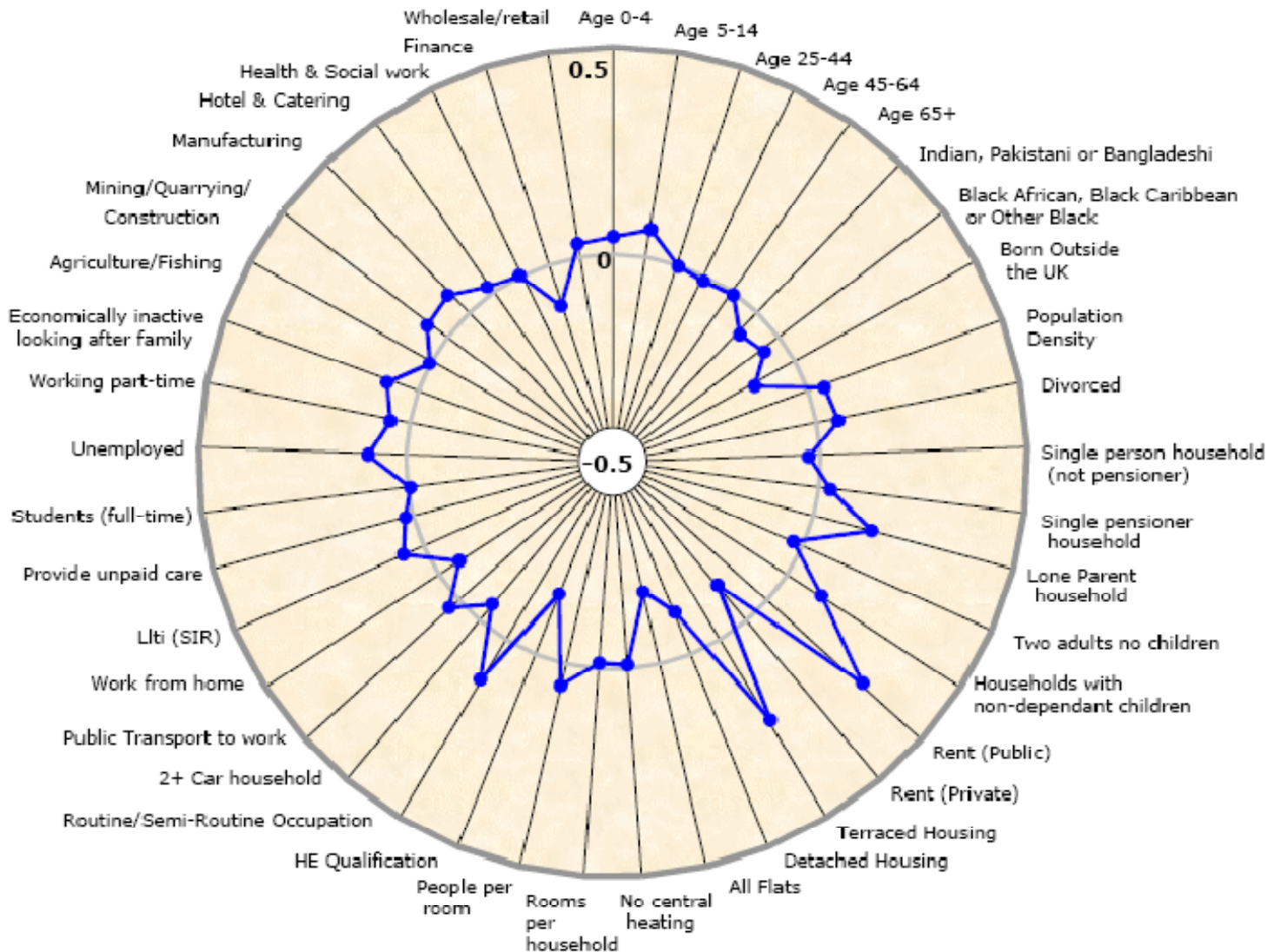
## Cluster Names

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

# Creating a Geodemographic Classification

## 1: Blue Collar Communities

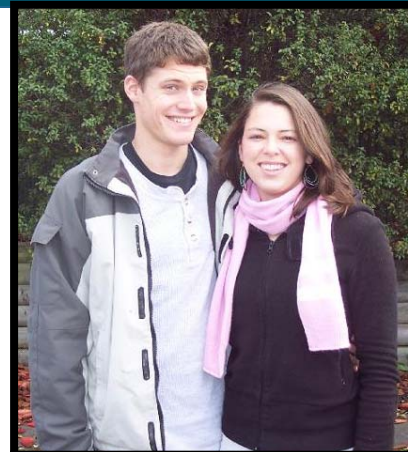
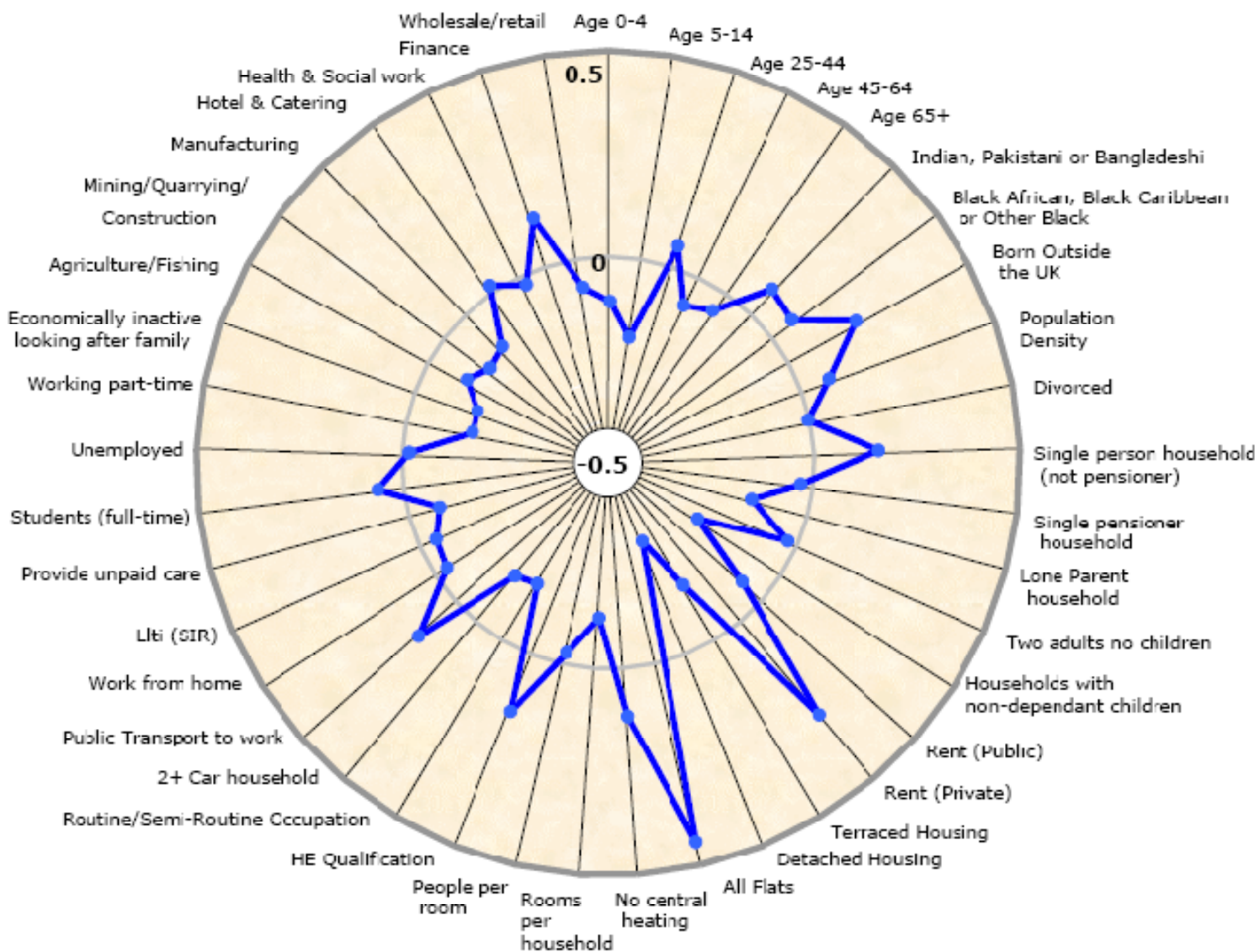
## Cluster Profiling



# Creating a Geodemographic Classification

## 2: City Living

## Cluster Profiling

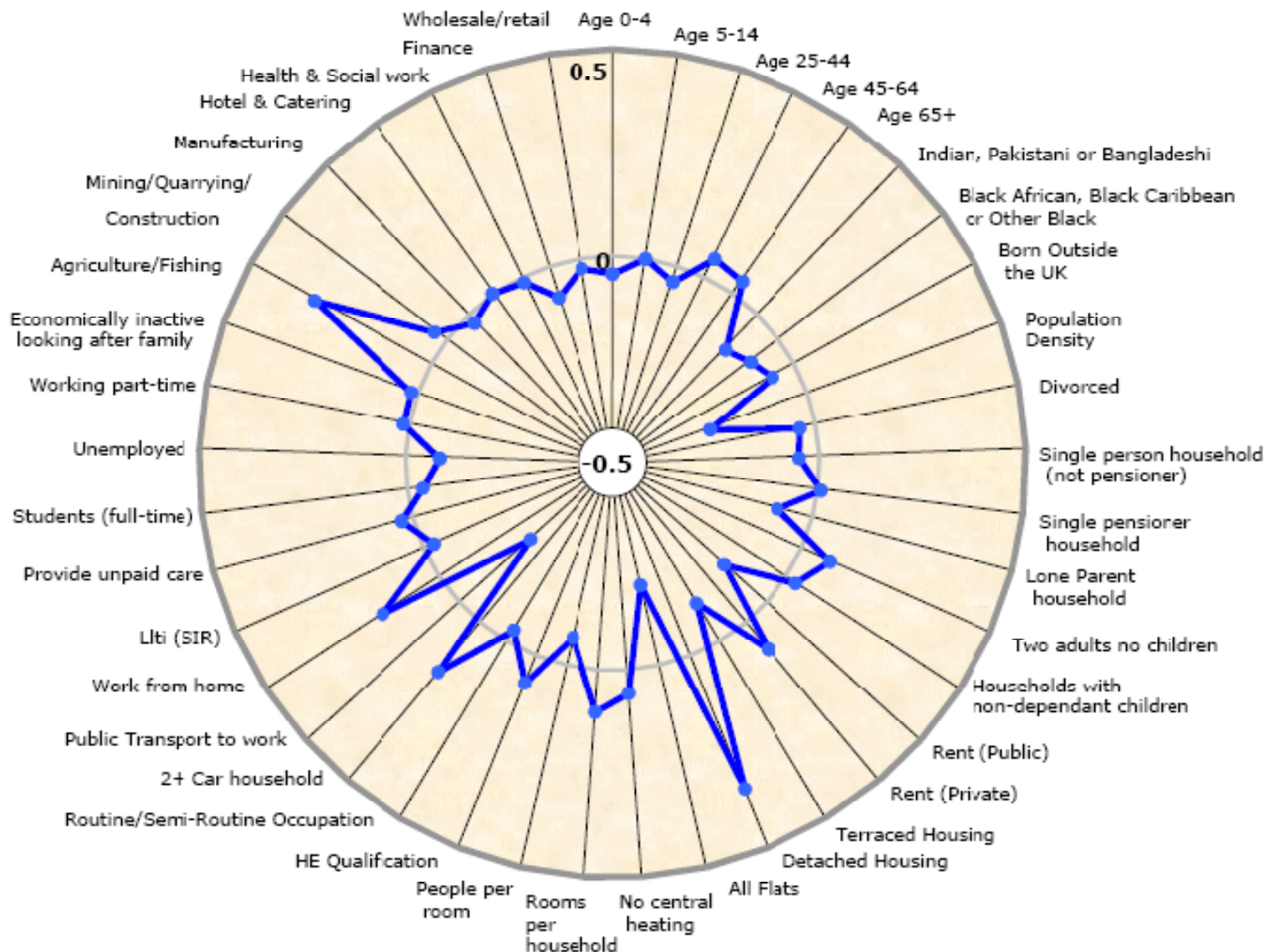




# Creating a Geodemographic Classification

## 3: Countryside

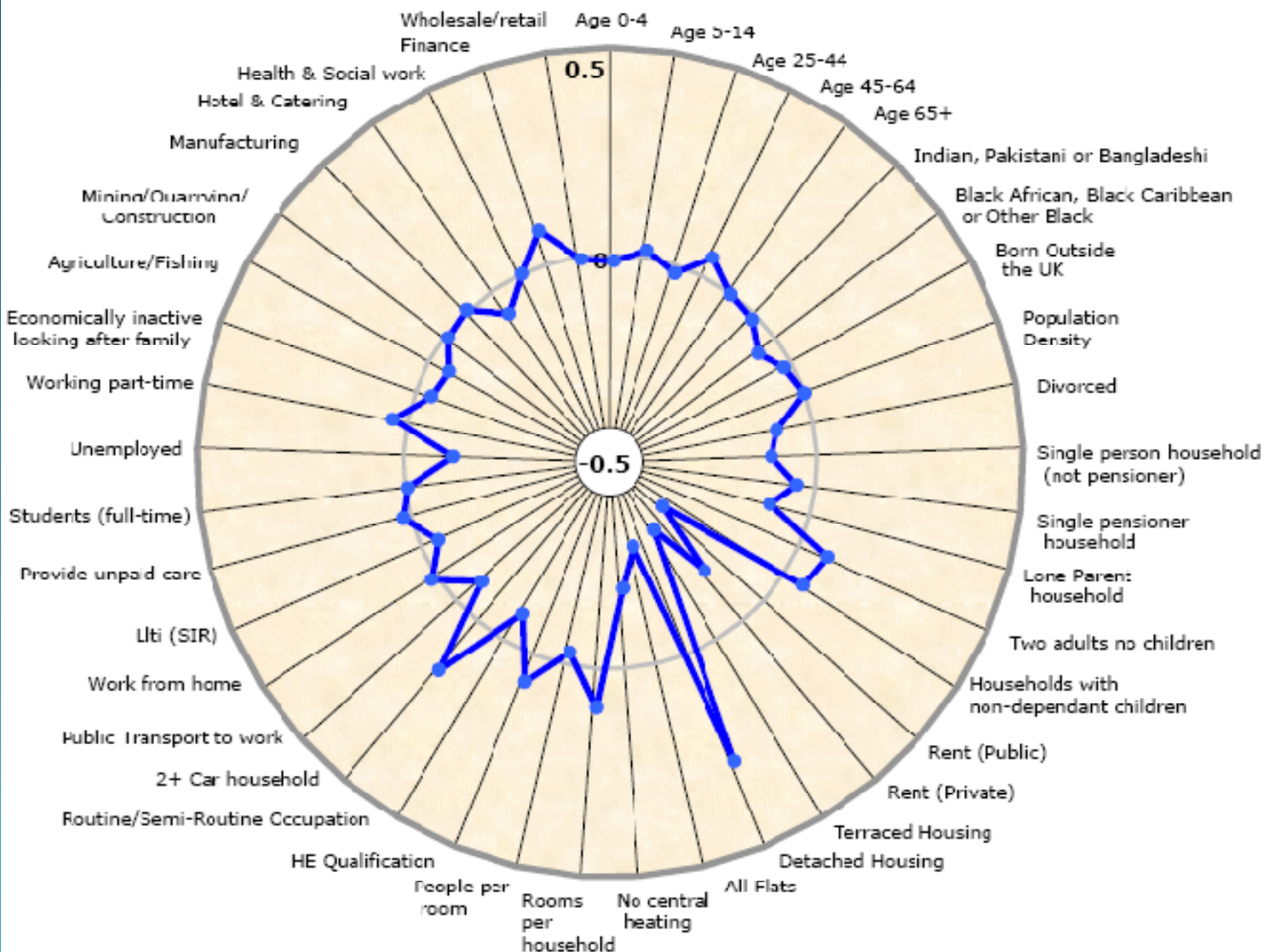
## Cluster Profiling



# Creating a Geodemographic Classification

## 4: Prospering Suburbs

## Cluster Profiling

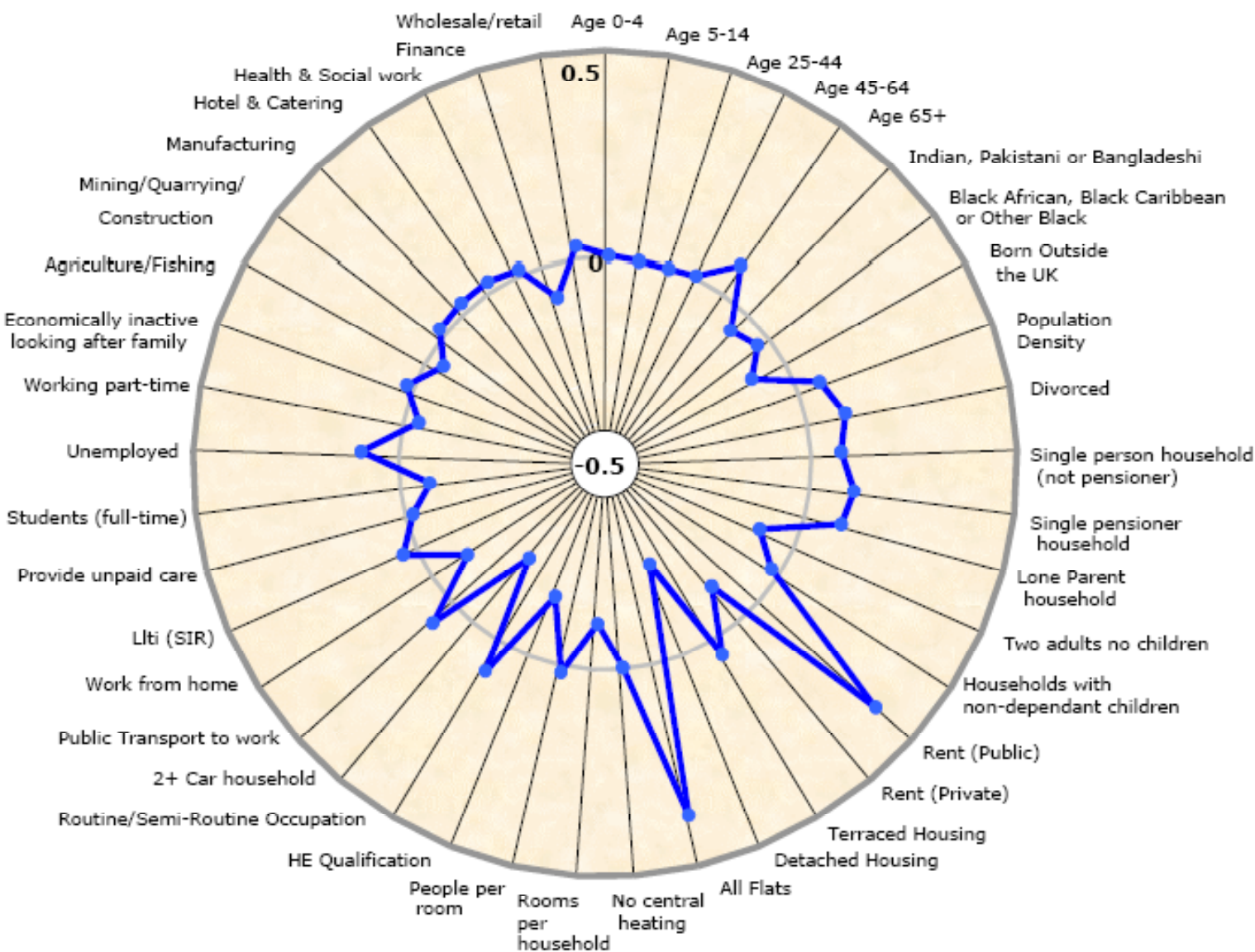




# Creating a Geodemographic Classification

## 5: Constrained by Circumstances

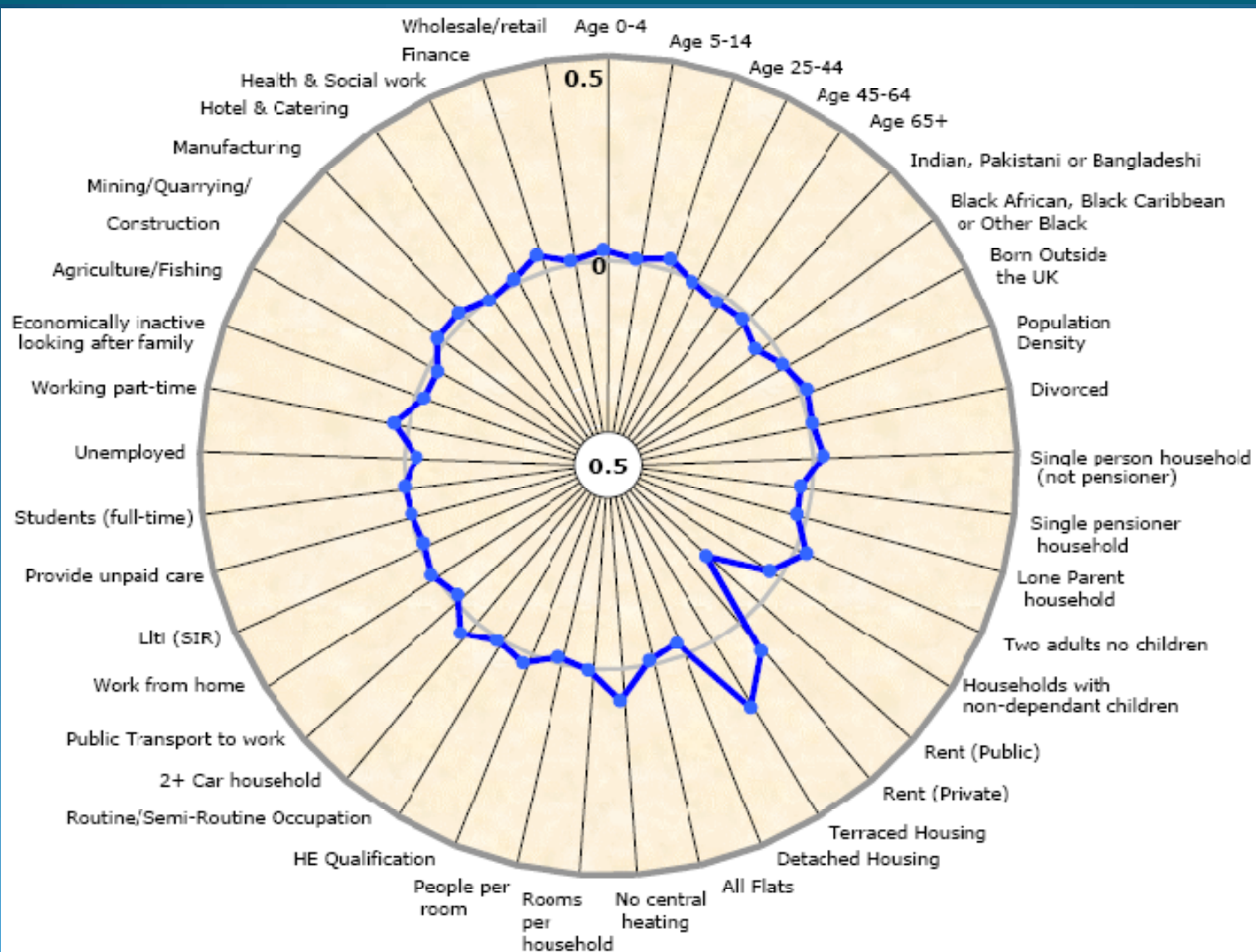
## Cluster Profiling



# Creating a Geodemographic Classification

## 6: Typical Traits

## Cluster Profiling

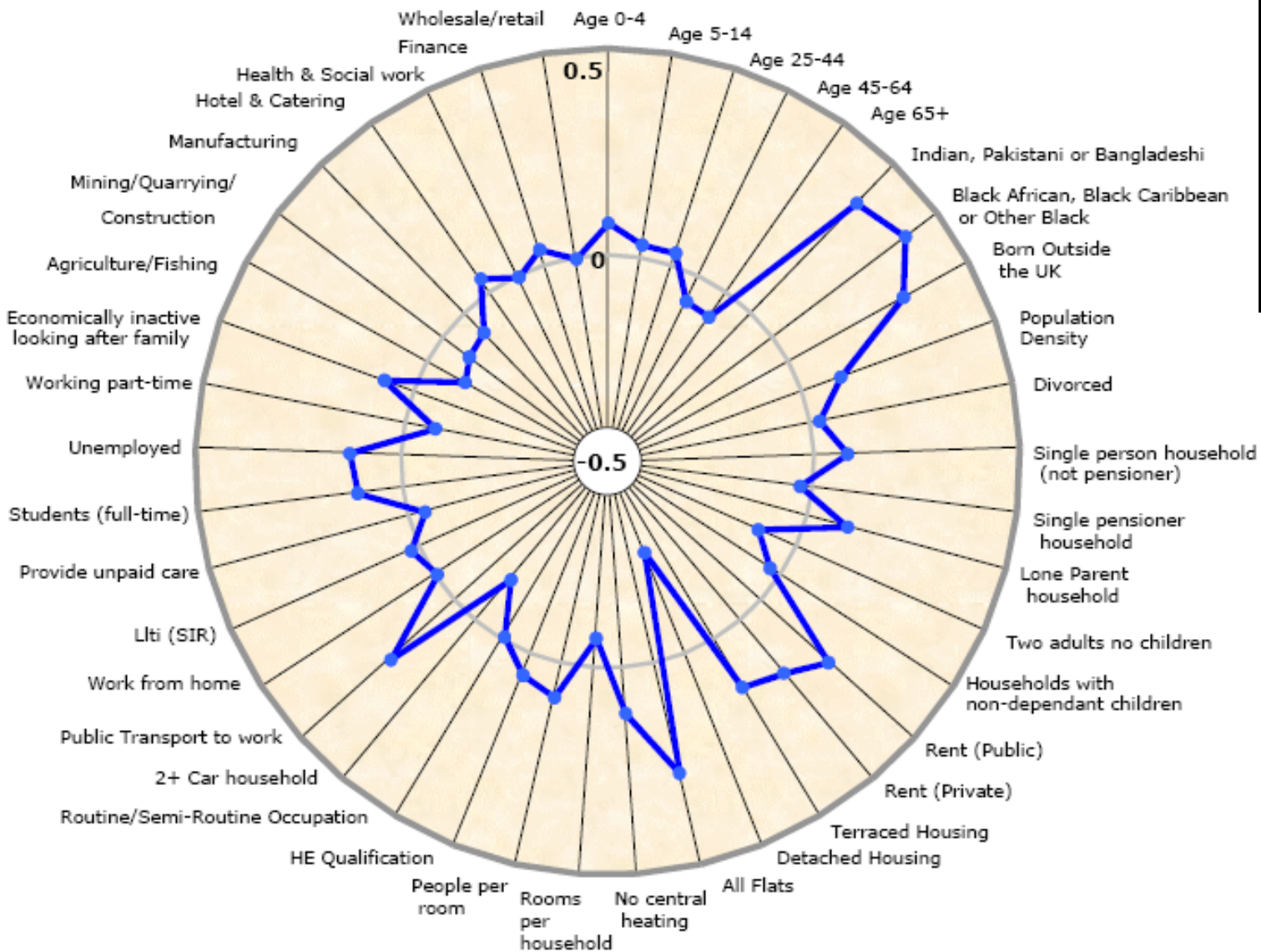




# Creating a Geodemographic Classification

## 7: Multicultural

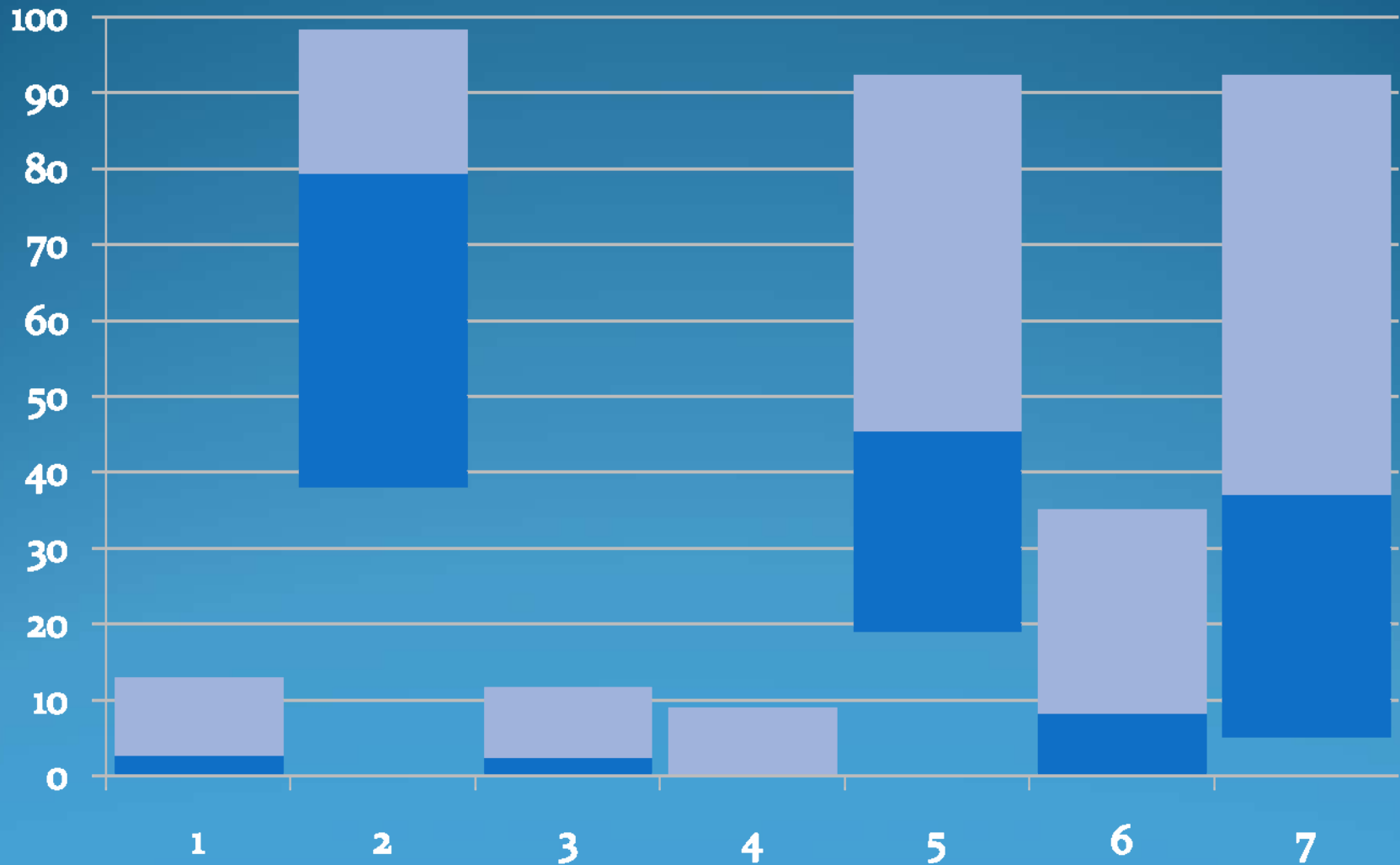
## Cluster Profiling



# Creating a Geodemographic Classification

## Cluster Profiling

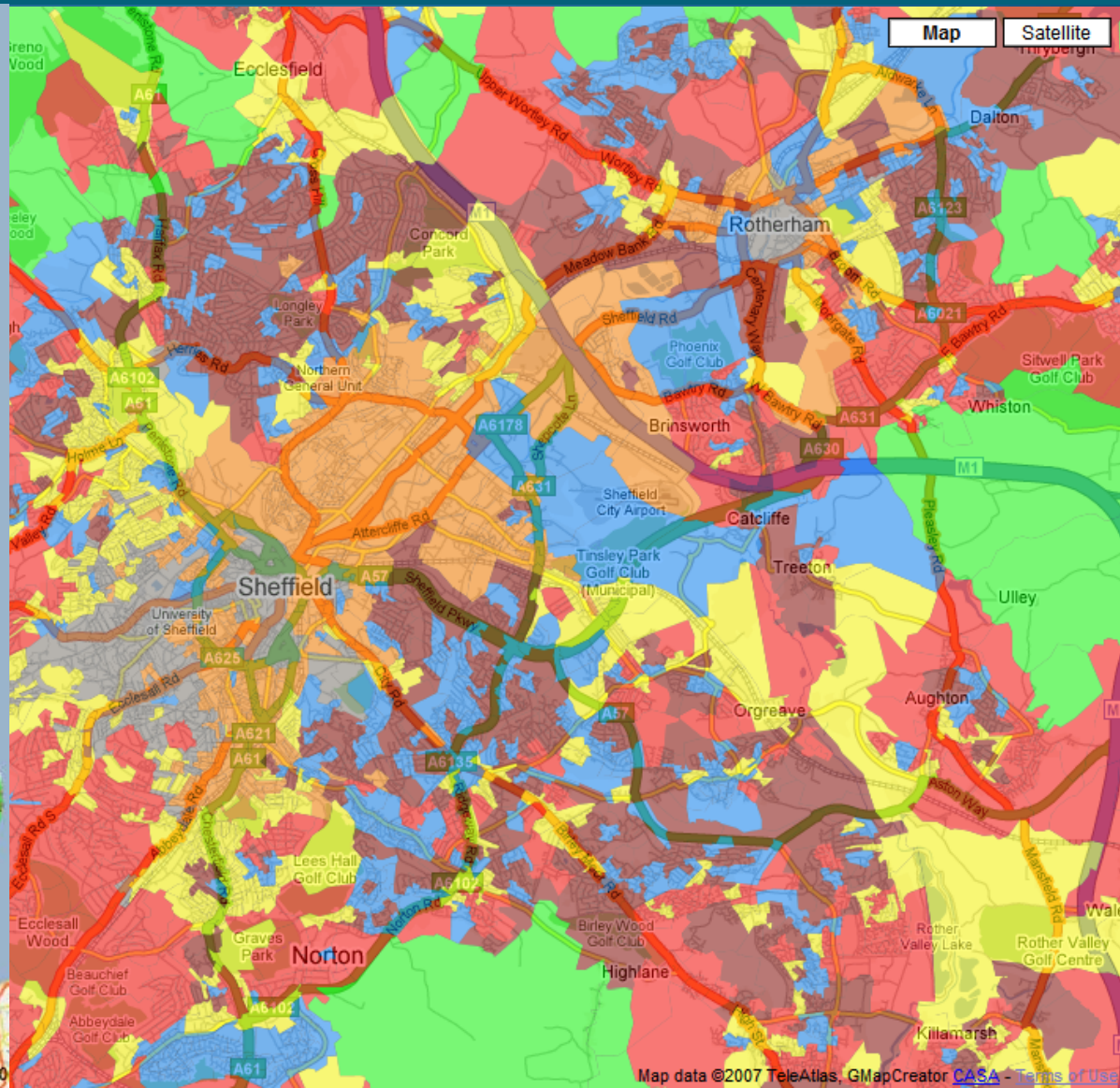
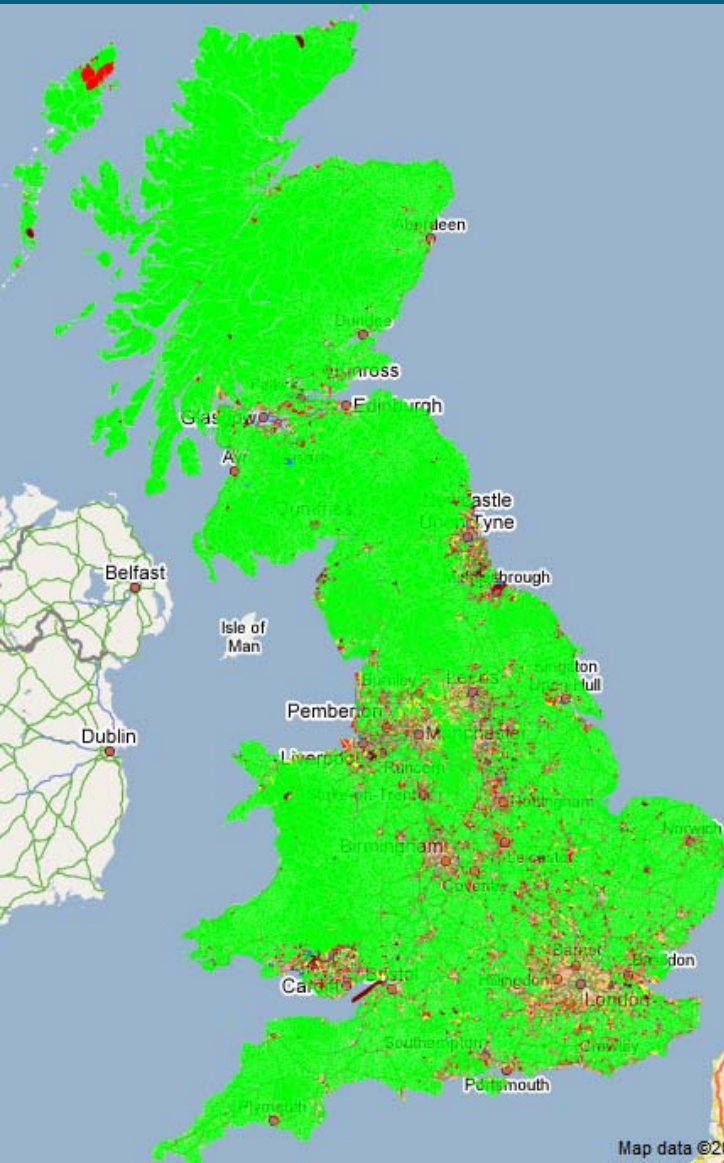
Alternative variable profiles by super-group: Flats (All)





# Creating a Geodemographic Classification

## Mapping the Classification





# Creating a Geodemographic Classification

Further Information

For more information visit:  
[www.areaclassification.org.uk](http://www.areaclassification.org.uk)

