National Statistics 2001 Area Classifications

John Charlton, ONS

see   http://neighbourhood.statistics.gov.uk

areaclassifications@ons.gov.uk

What are the Area Classifications

- Summarise 2001 Census data on 2003 boundaries

- Group together similar areas according to key characteristics of individuals and households

- Produced at Local Authority (LA), Health Area (HA), Ward and Output Area (OA) level for the UK

- Ongoing product, first produced in 1971, (now UK, previously GB)

Examples where is it used

- ONS: Mortality (SMRs), Fertility Rates, ILO Unemployment Estimates

- Local Authorities - to compare "similar" LAs

- DoH: NHS Performance Indicators including Health Outcome measures

- Academic research: provides "type of area of residence"

- SARS: used for modelling multilevel area effects

Aim

- To describe the 2001 Area Classifications

  - Data used
  - Selection of variables
  - Standardisation methods
  - Clustering techniques
  - LA and HA methodology
  - Ward level methodology
  - OA level methodology

Dataset

- 70 variables, from Key Statistics, aggregated up to percentages for LA/HA/Ward

  – Dimensions of individual/household characteristics
    - Demographic
    - Household composition
    - Housing
    - Socio-economic
    - Employment
    - Industry

- Suggestions provided by advisory board

Selection of variables

- Construct a correlation matrix using Pearson's correlation coefficient

- Aim to remove one of a pair of strongly correlated (>0.85)/(<-0.85) variables by:
    - Looking at distributions
    - Using judgements based on experience, logic and dimensions
   - e.g. Students and 15-24 age group

Cluster analysis

- Classification produced using cluster analysis

- Clustering based on the distances between cases to be clustered - constructed a between-area distance matrix based on all the variables in the data set

- Problems will occur if there are differing scales or magnitudes among the variables - variables with larger dispersion have more impact on similarity measure

- Standardising the data will represent each variable equally in the distance measure
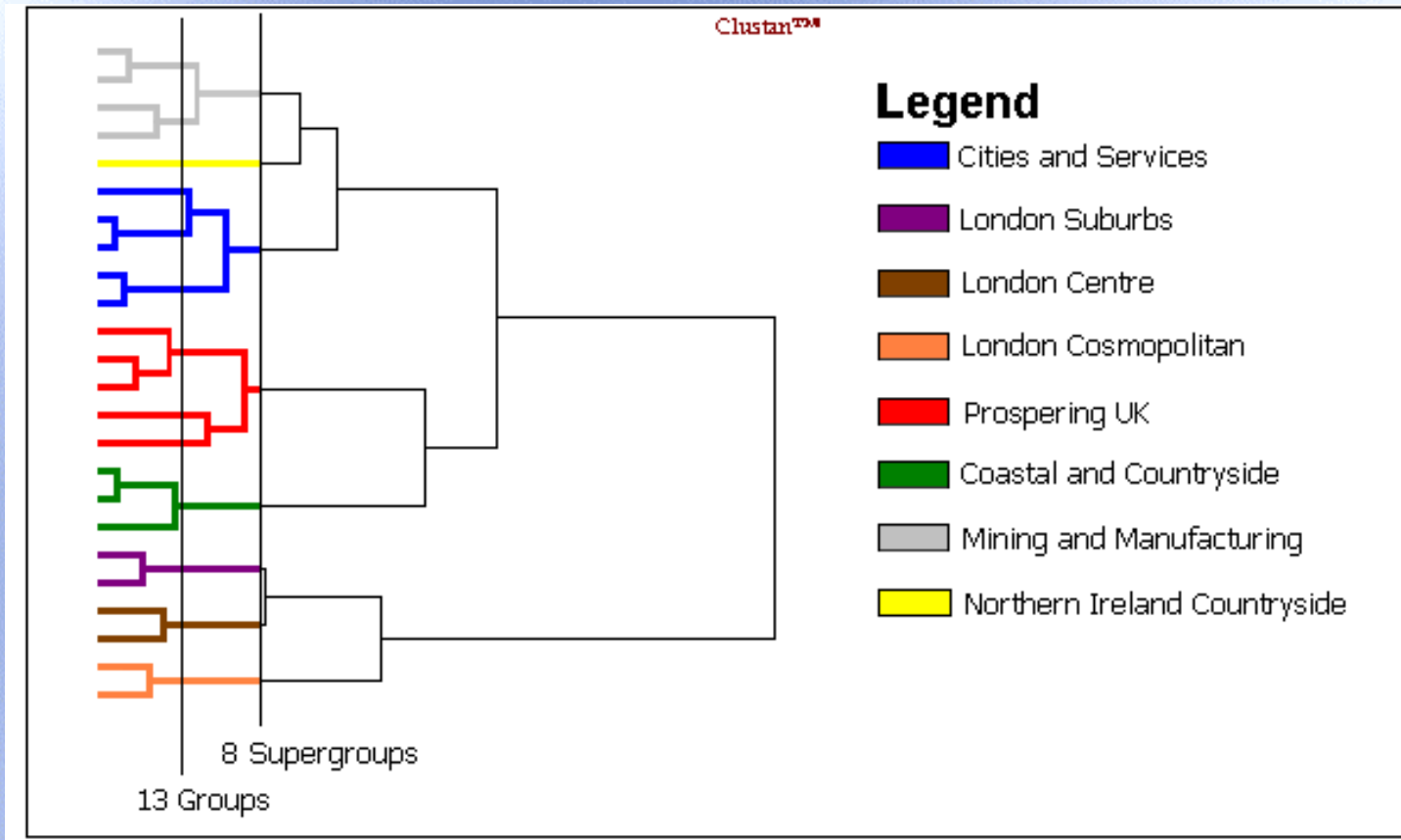
Ward's method

- Starts with all cases as separate clusters.
- Next forms one cluster containing 2 cases, then fuses another pair and so on, at each stage minimising within-cluster sum of squares, also known as 'Error Sum of Squares' (ESS).
- Essentially, for each case, distance to cluster means is calculated.
- ESS is calculated for all possible solutions to find solution with the least ESS, then process is repeated until there is just one cluster containing all cases.
- The cut-off points are chosen based on the agglomeration schedule.

An Agglomeration schedule

Increase in distance between most dissimilar local authority within merged clusters
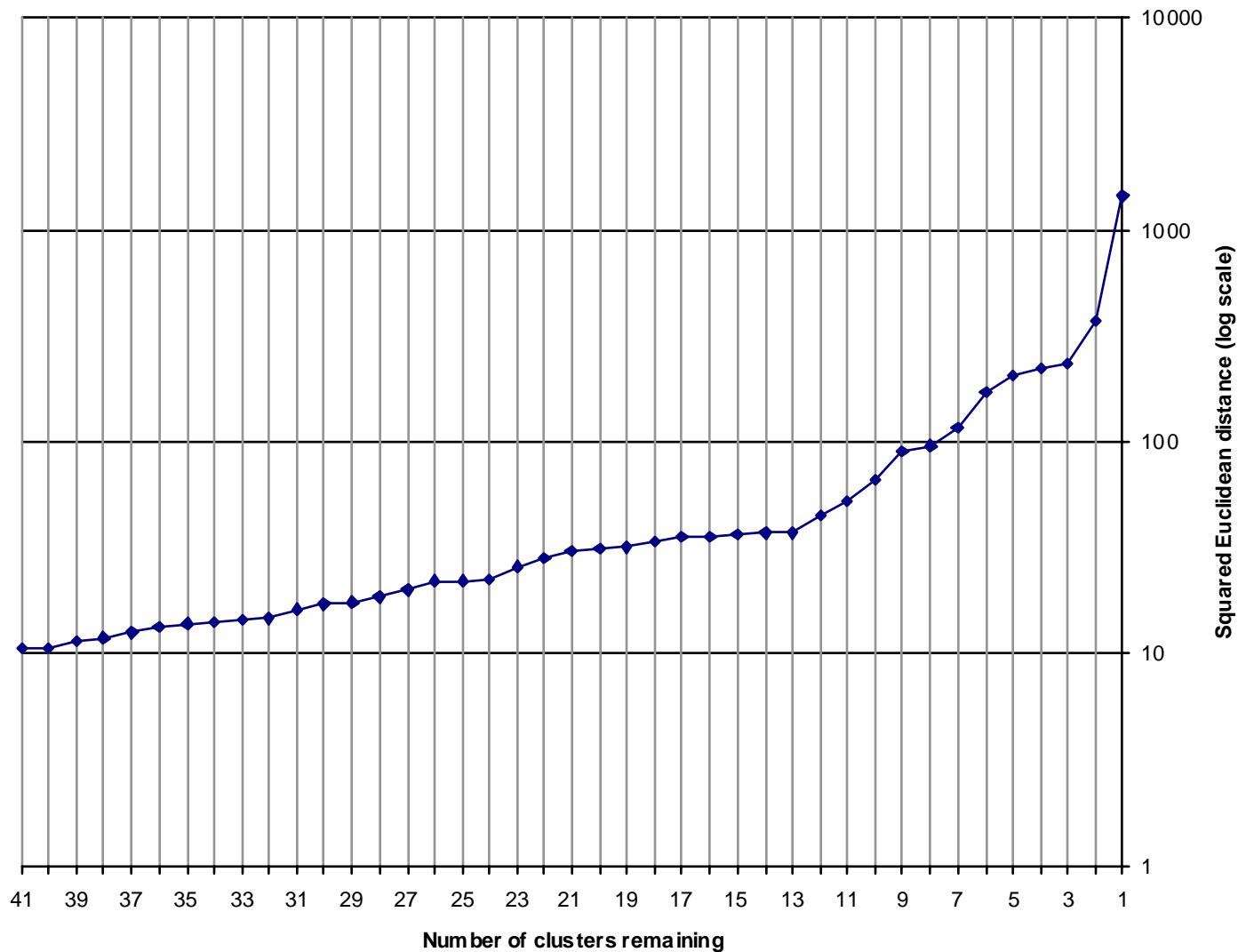
Hierarchical structure (LADs)

K- means

- With Ward's method, cluster centroids will be changing at each step. K-means sed to re-assign cases to cluster with smallest distance between case and cluster centroid over all variables

- Centroids obtained at the lowest level are used to begin the process

- Iterative method, continues until a stable result is achieved (until the difference between the case and other clusters is minimal)

- Higher levels can then be obtained using the hierarchy obtained from Ward's method
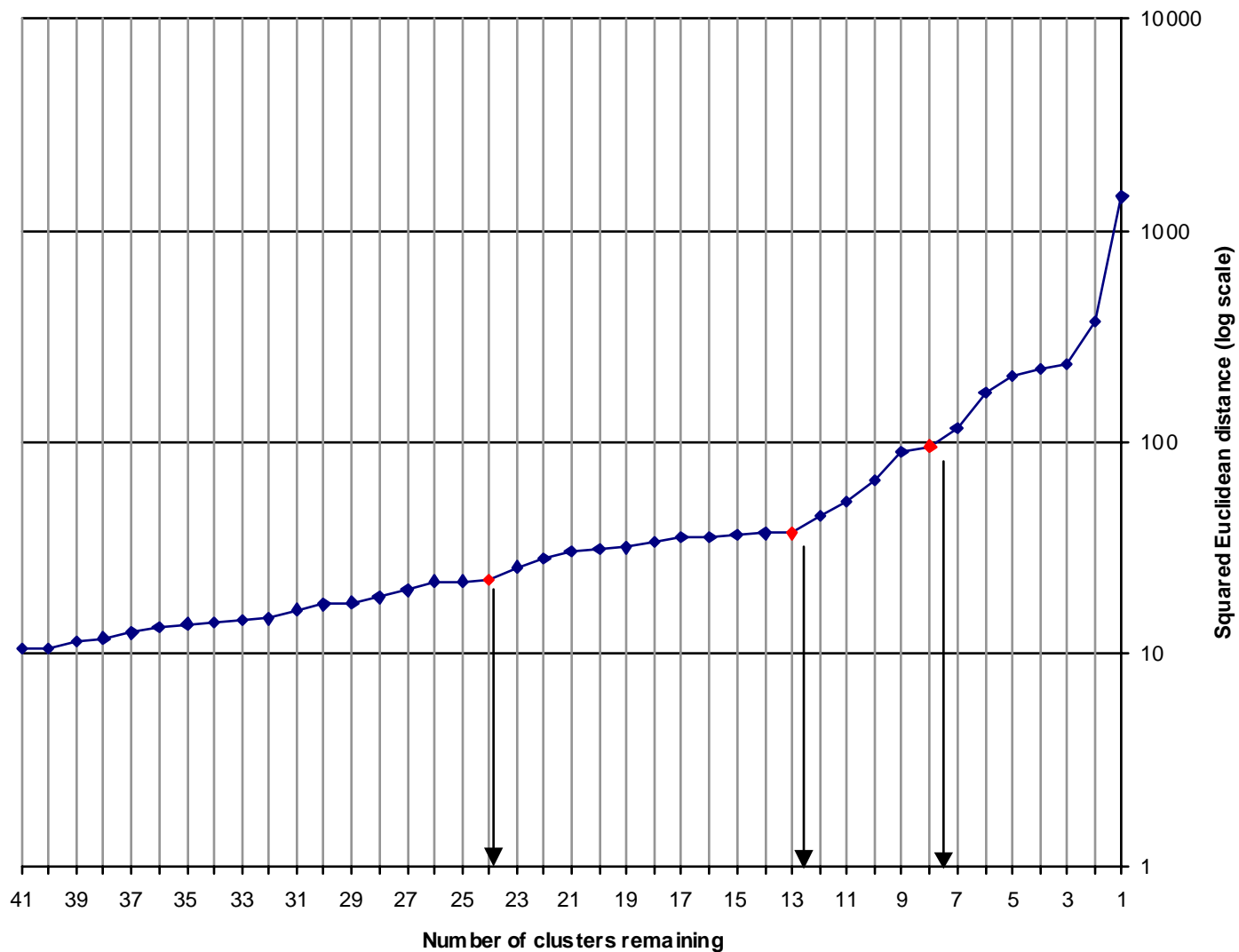
Methodology - Local Authority

- Dataset consisted of 42 variables, 432 cases

- Used Inter-decile range-standardisation

- Constructed a distance matrix to determine how 'different' LAs were from each other (Squared Euclidean Distance)

Methodology - Local Authority

- Ward's method/K-means cluster Analysis used to produce hierarchical classification

- Agglomeration schedule determined the cut-off points (24 subgroups, 13 groups and 8 supergroups)
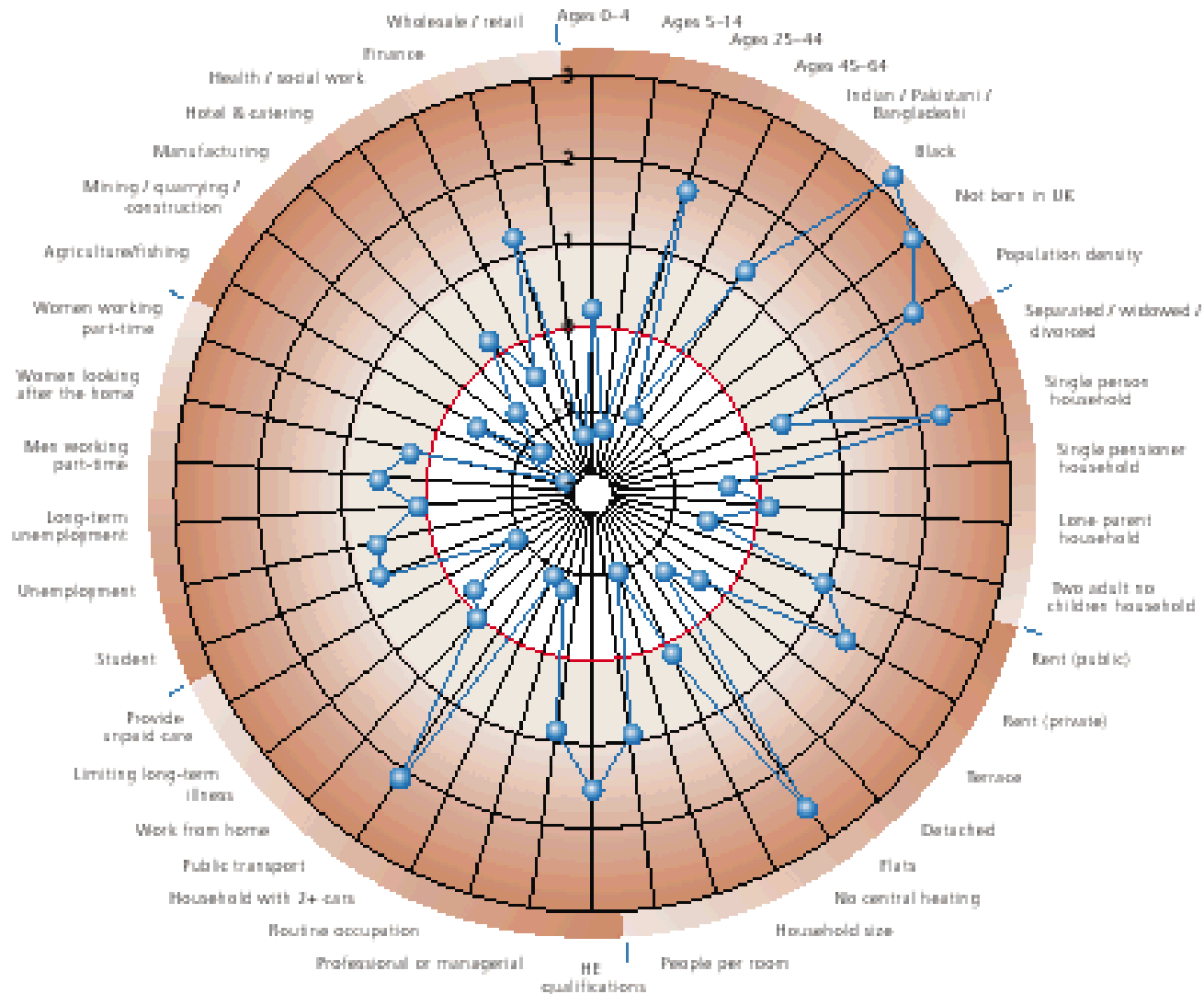
Agglomeration schedule - LADs

Increase in distance between most dissimilar local authority within merged clusters

London centre - radar chart

Range-standardised difference from the UK mean. Positive values show an above average level of the population and negative values show a below average level of the population

Methodology - Health Authority

- DoH asked us to map HAs to LA (24 level)

- HAs were assigned to those subgroups giving the smallest SED from the subgroup centroid

- Same 42 variables which were used as in LA

- HAs standardised to LA data

- Higher levels of the classification were created using the LA hierarchy

2001 Methodology - Ward level

- Dataset consisted of 43 variables

- Wards with a population <1000 people were merged with a neighbouring ward to obtain 10553 standard wards (as per Census statistical wards outputs)

- Data standardised using Range standardisation

2001 Methodology - Ward level

- 1991 ward classification had been based on first classifying a *sample* of wards

- 2001 classification - adopted a different approach (Charlton, Openshaw and Wymer, 1985)

- Generated a random classification of all wards into 1000 clusters using K-means

- The cluster centroids from this random classification were used as starting point to reach the optimum 1000 cluster solution (K-means)

- Ward's method then applied to the 1000 clusters

- Agglomeration schedule determined the cut-off points:
  - 26 subgroups,
  - 17 groups
  - 9 supergroups

- Subgroups obtained from Ward's method were refined using k-means to ensure each ward was assigned to it's correct subgroup
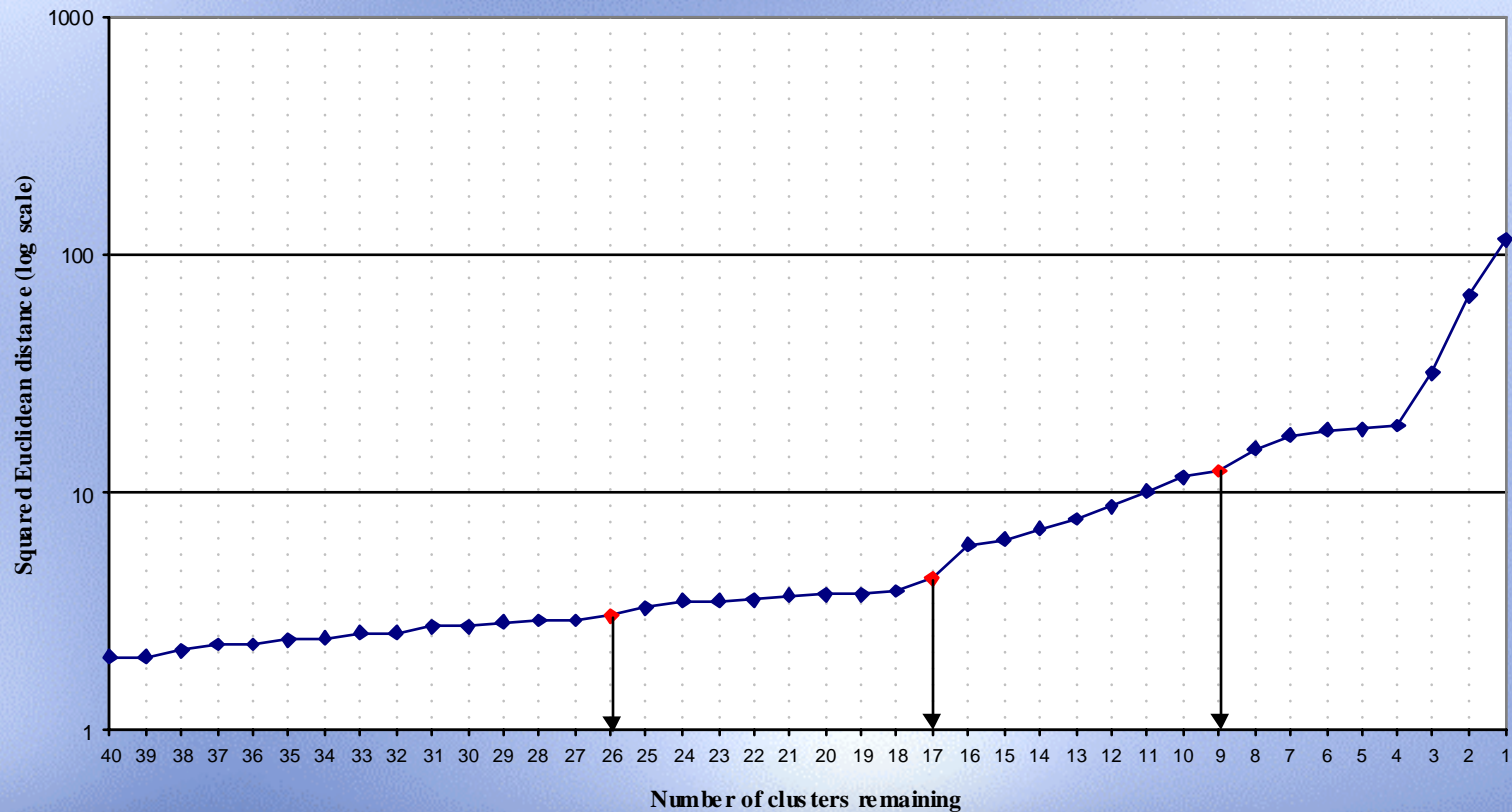
Before reaching final solution, carried out experimental sensitivity analyses:

- varied standardisation method

- altered number of initial clusters

- tried additional variables suggested by Advisory Board

Range standardisation

Increase in distance between most dissimilar statistical wards within merged clusters

- All classifications available for free from the ONS neighbourhood statistics website

- Local Authority level already there, HAs, and wards available March, OAs afterwards

- Quick view of website, and some results for local authorities….

national STATISTICS

22 October 2003

uk at a glance   neighbourhood   £ economic   census 2001

home | neighbourhood statistics

User Name:

Password:

LOGIN

▶ Why register?
▶ Register for site

Navigation

▶ Neighbourhood Statistics
▶ Statistics by subject
▶ Statistics by area
▶ Interactive map
▶ Area Classification

▶ Information about this service
▶ Help
▶ Viewing requirements
▶ Frequently asked questions
▶ Glossary
▶ Contact us

Neighbourhood Statistics

Welcome to Neighbourhood Statistics

Use this site to view, compare or download statistics for local areas on a wide range of subjects including population, crime, health and housing.

Please choose one of the options below:

Summary statistics for your area
Please enter your full postcode or a city or town name (England and Wales only)

GO

Find detailed statistics by subject

Find detailed statistics by area name

Interactive map
(England and Wales only)

Latest News

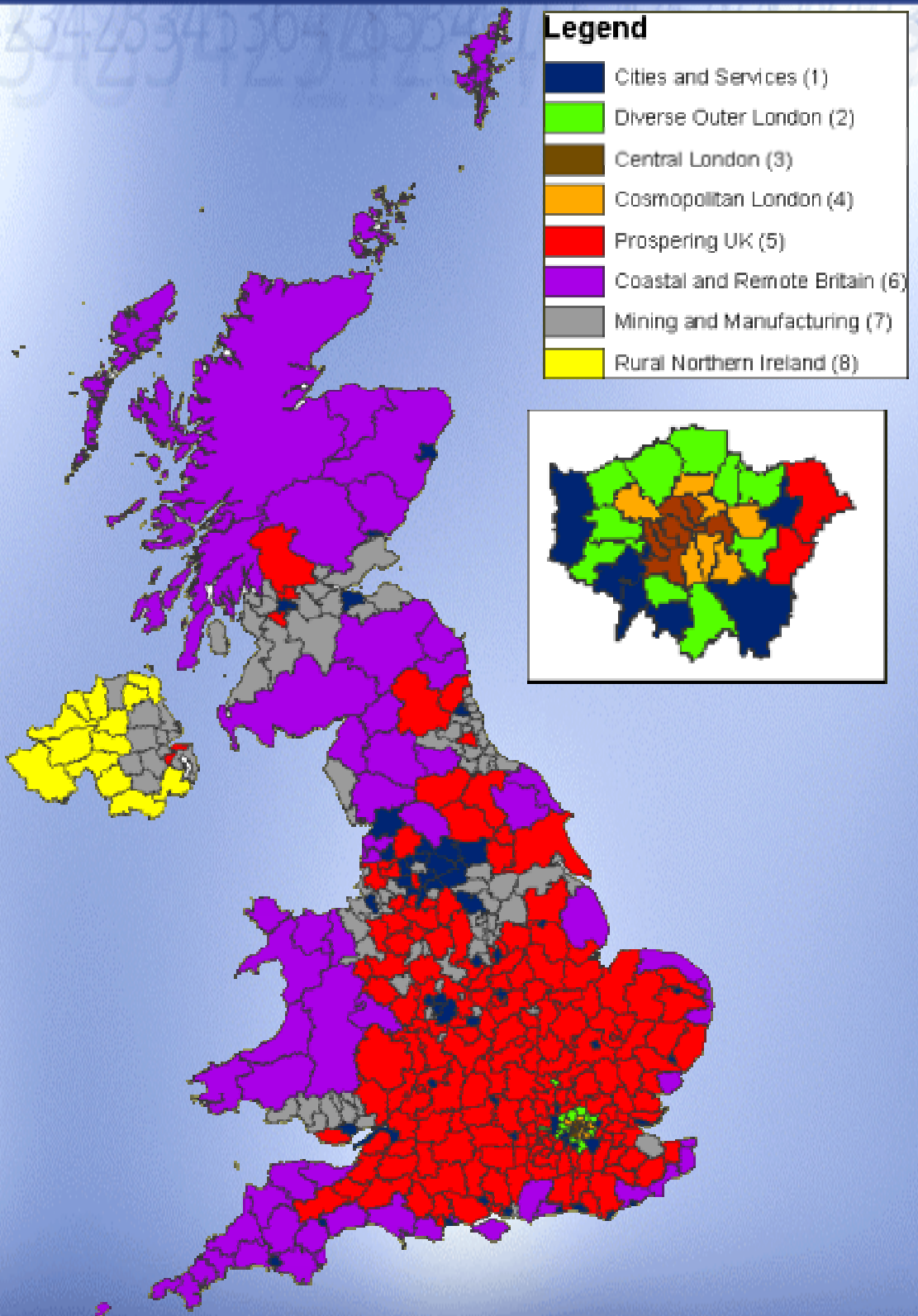▶ Area Classification for local authorities released on the 31st of October.

Related links

▶ Census 2001
▶ Small area population estimates project
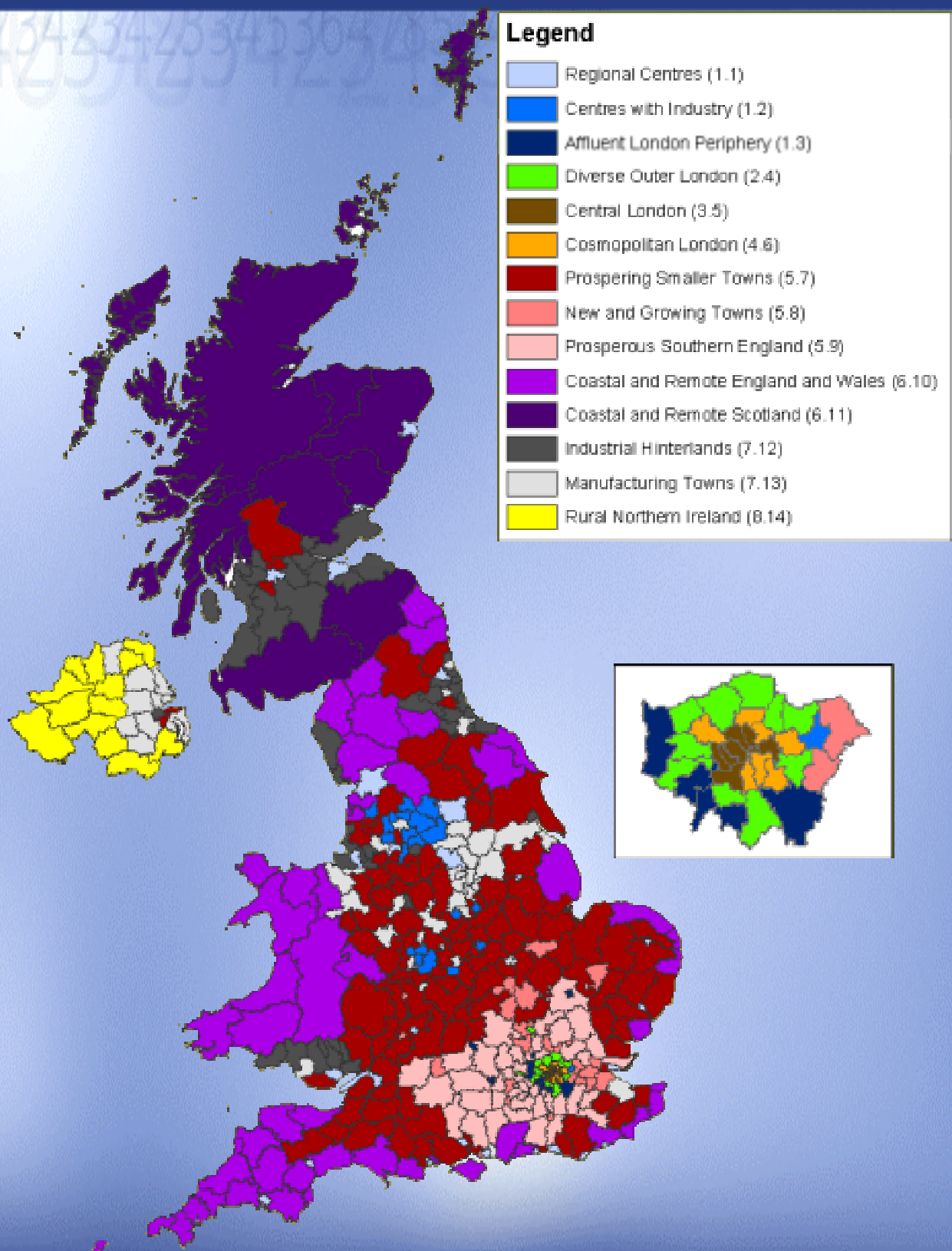▶ Nomis - online labour market statistics

External links

▶ renewal.net
▶ Neighbourhood Renewal Unit
▶ Social Exclusion Unit
▶ CLIP

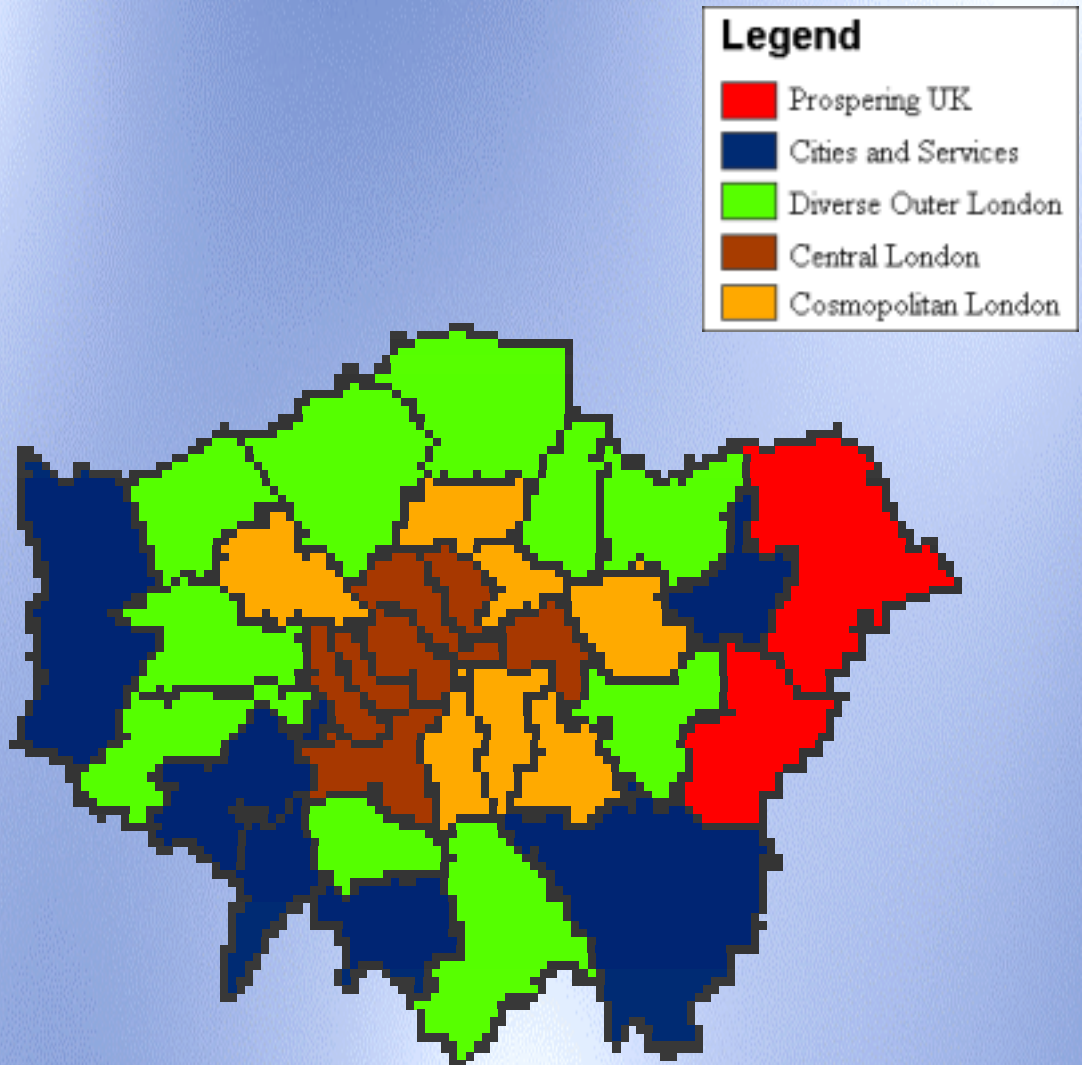# The 8 Families
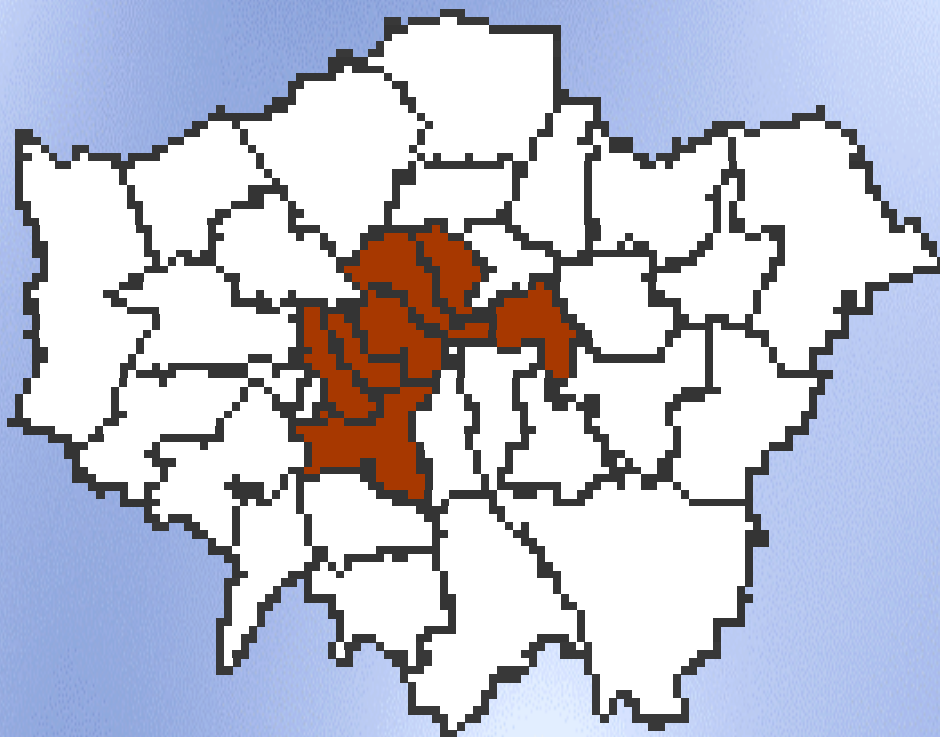

Legend
- Cities and Services (1)
- Diverse Outer London (2)
- Central London (3)
- Cosmopolitan London (4)
- Prospering UK (5)
- Coastal and Remote Britain (6)
- Mining and Manufacturing (7)
- Rural Northern Ireland (8)

The 14 Groups

Legend

- Regional Centres (1.1)
- Centres with Industry (1.2)
- Affluent London Periphery (1.3)
- Diverse Outer London (2.4)
- Central London (3.5)
- Cosmopolitan London (4.6)
- Prospering Smaller Towns (5.7)
- New and Growing Towns (5.8)
- Prosperous Southern England (5.9)
- Coastal and Remote England and Wales (6.10)
- Coastal and Remote Scotland (6.11)
- Industrial Hinterlands (7.12)
- Manufacturing Towns (7.13)
- Rural Northern Ireland (8.14)

London

Legend

- ■ (red) Prospering UK
- ■ (dark blue) Cities and Services
- ■ (green) Diverse Outer London
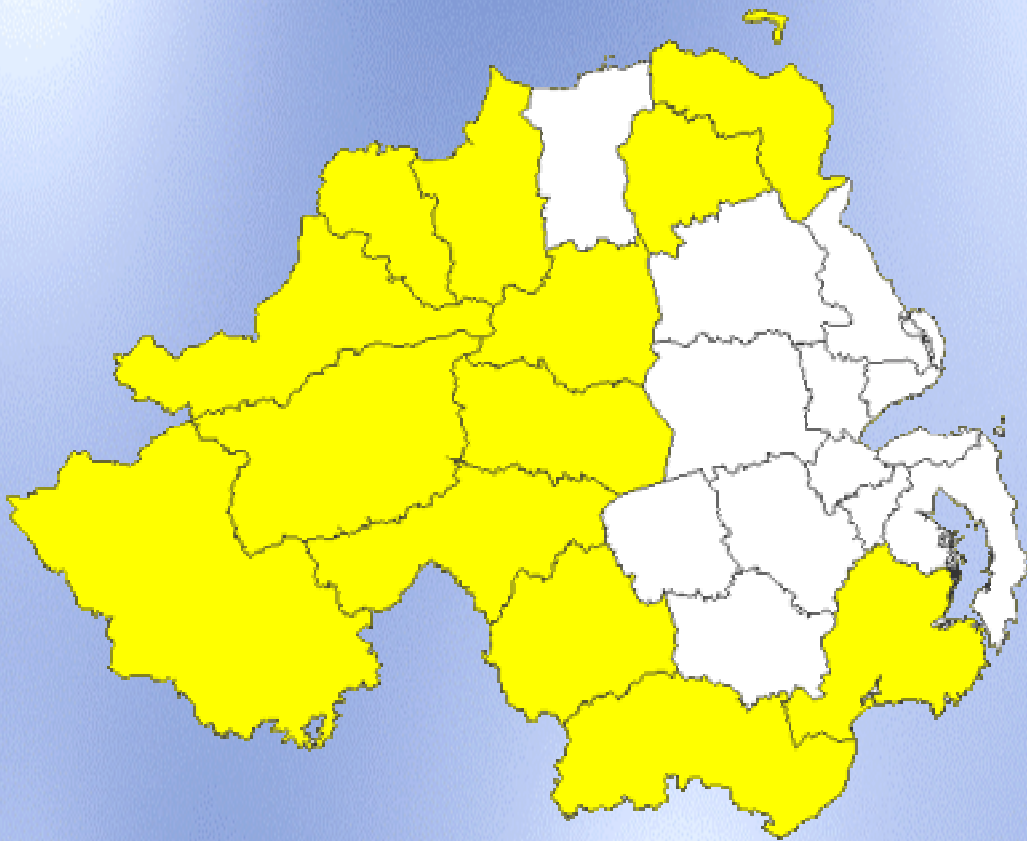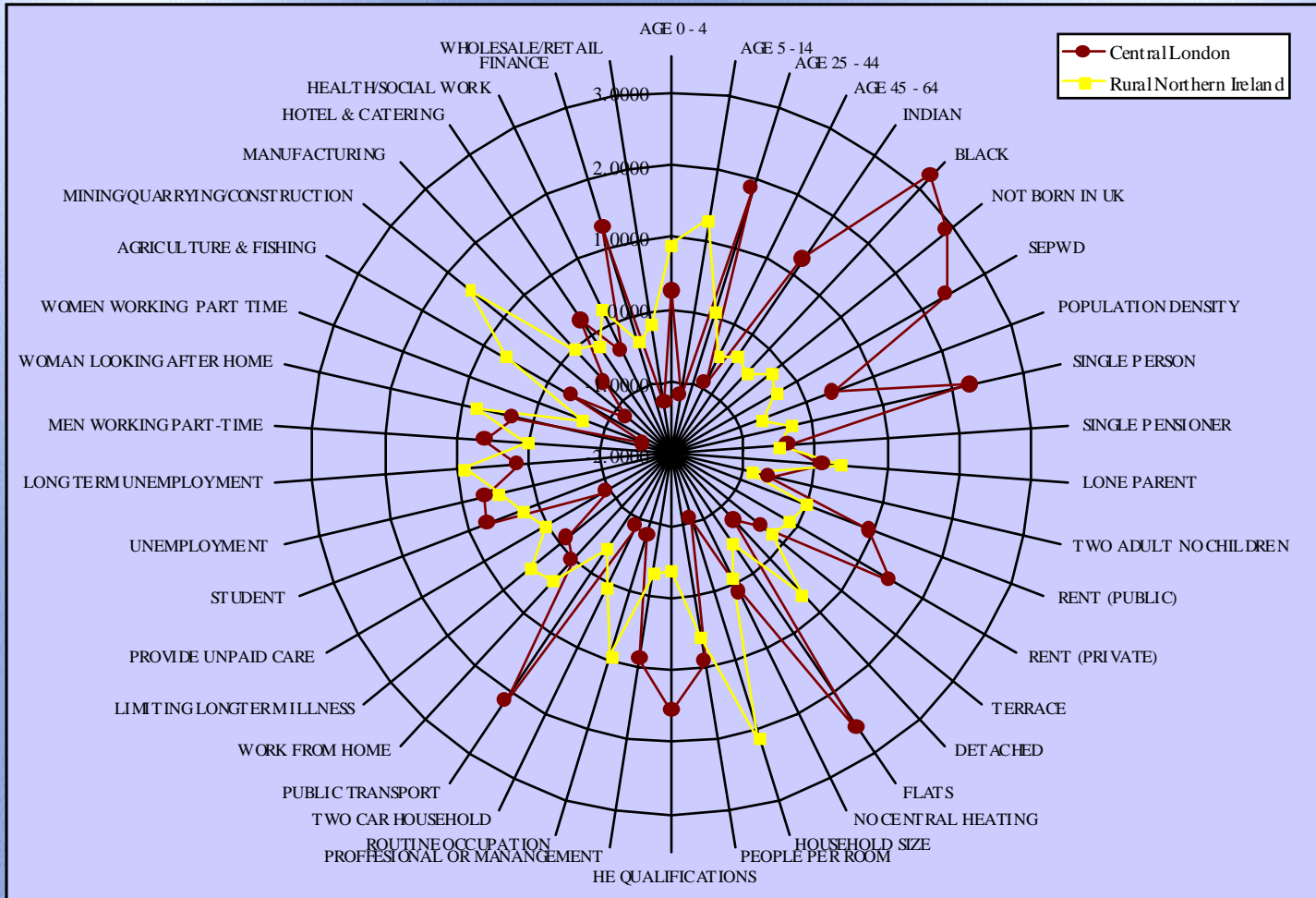- ■ (brown) Central London
- ■ (orange) Cosmopolitan London

Central London

Rural Northern Ireland

Radar Plots

Legend: Central London, Rural Northern Ireland

Axes (clockwise from top): AGE 0 - 4, AGE 5 - 14, AGE 25 - 44, AGE 45 - 64, INDIAN, BLACK, NOT BORN IN UK, SEPWD, POPULATION DENSITY, SINGLE PERSON, SINGLE PENSIONER, LONE PARENT, TWO ADULT NO CHILDREN, RENT (PUBLIC), RENT (PRIVATE), TERRACE, DETACHED, FLATS, NO CENTRAL HEATING, HOUSEHOLD SIZE, PEOPLE PER ROOM, HE QUALIFICATIONS, PROFFESIONAL OR MANANGEMENT, ROUTINE OCCUPATION, TWO CAR HOUSEHOLD, PUBLIC TRANSPORT, WORK FROM HOME, LIMITING LONG TERM ILLNESS, PROVIDE UNPAID CARE, STUDENT, UNEMPLOYMENT, LONG TERM UNEMPLOYMENT, MEN WORKING PART-TIME, WOMAN LOOKING AFTER HOME, WOMEN WORKING PART TIME, AGRICULTURE & FISHING, MINING/QUARRYING/CONSTRUCTION, MANUFACTURING, HOTEL & CATERING, HEALTH/SOCIAL WORK, FINANCE, WHOLESALE/RETAIL

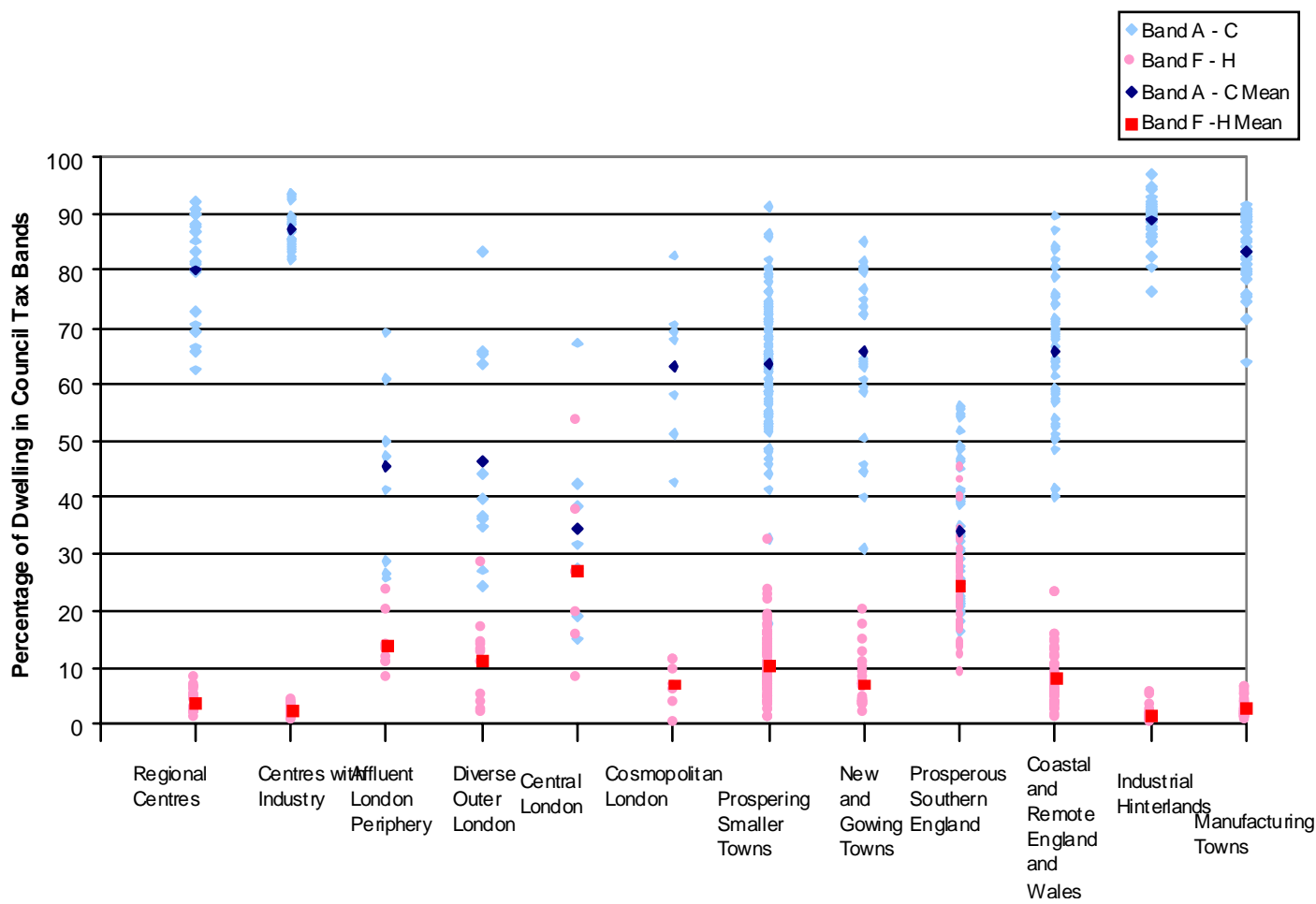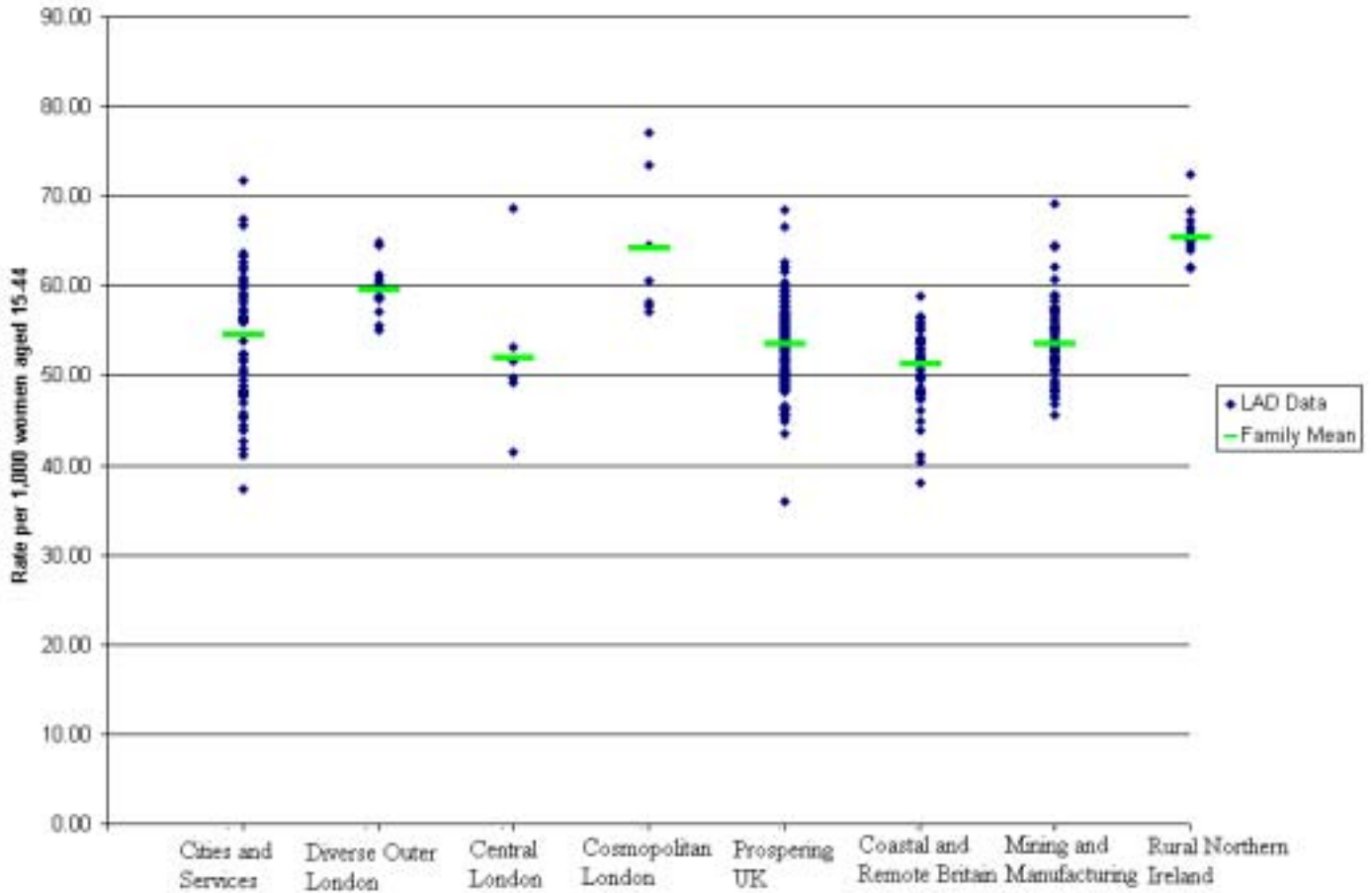Radial scale: 1.0000, 0.0000, 1.0000, 0.0000, 1.0000, 2.0000, 3.0000

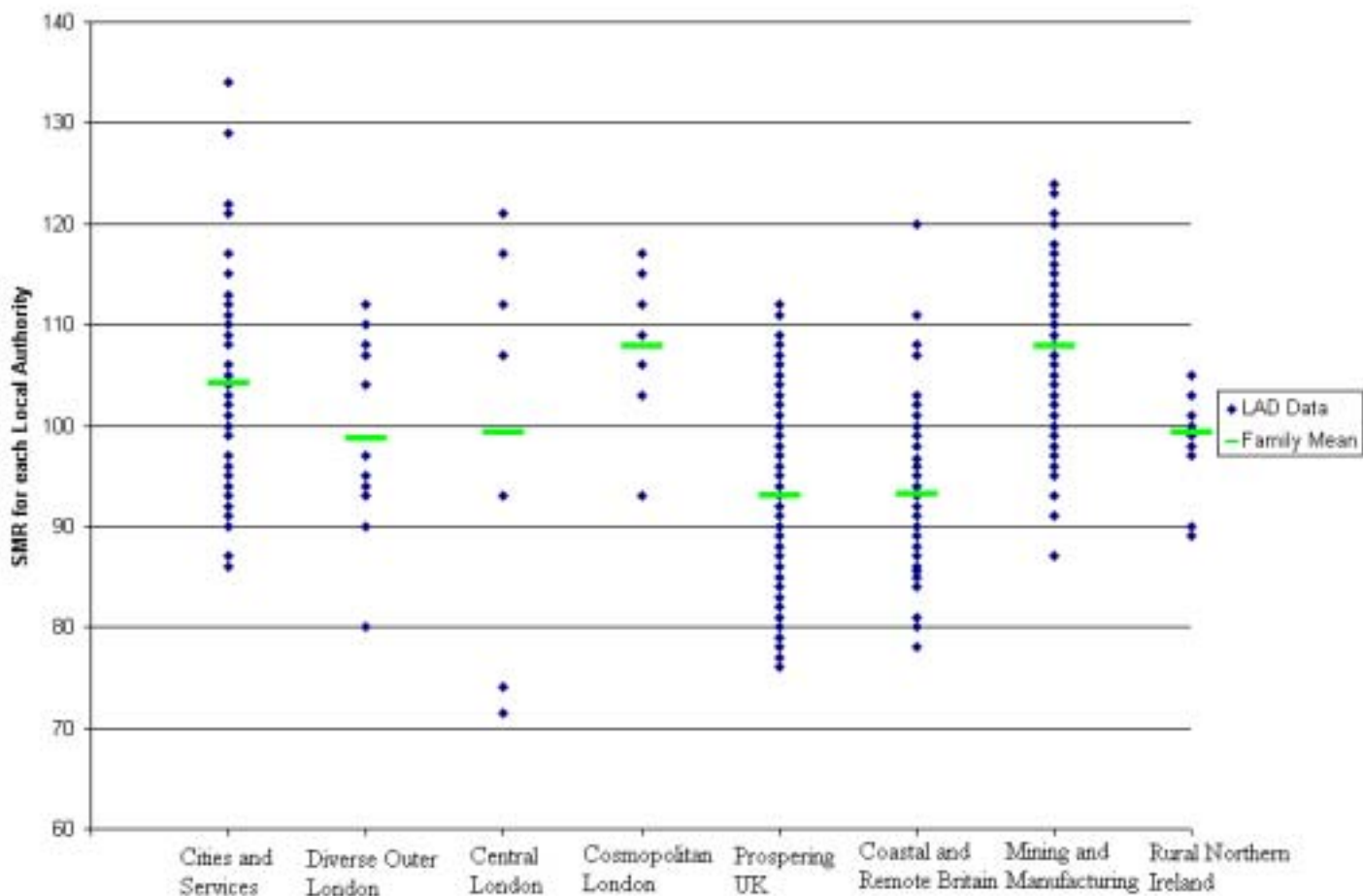# Percentage of VAT registrations, by industry
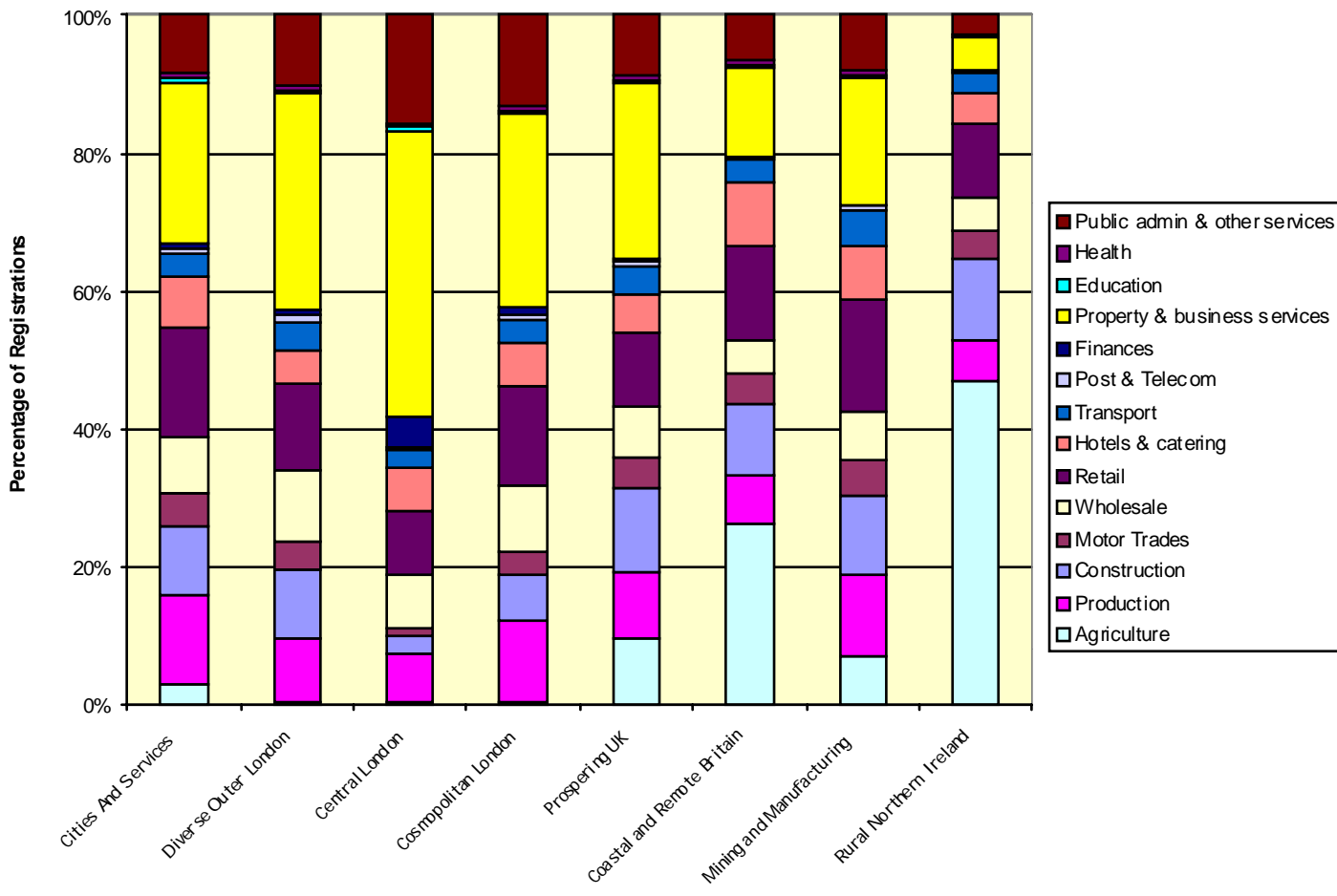
Percentage dwellings in council tax bands - England

General Fertility Rate by Family

Standardised Mortality Ratio by Family

Standardised Mortality Ratio by Group