

How a People Classification Can Add Value to Census Data

Simon Perry

Presentation outline

- Why the census is useful and what's better this time
- Disclosure protection and spatial analysis
- What the census doesn't tell you
- How people classifications can tell you more than raw census data
- Comparing one classification with another

Census data

- Consistent
- Comprehensive
- Compulsory
- FREE

What is better this time?

- One number census
- Well designed Output Areas
- Better links to postcode geography
- Better data on where people work
- Social Grade / deprivation indicators
- Students recorded at term time addresses
- New questions on floor level, religion, health, care, time since last employment, supervisor status, professional qualifications, size of workforce

Disclosure Protection

- Census act requires that no data on individuals be produced for 100 years
- To prevent this three techniques are used:
- Small cell adjustment
 - a small count appearing in a table cell is adjusted
- Record swapping
 - a sample of records is swapped with similar records in other geographical areas
- Thresholds
 - data not released for areas of less than 100 residents and 40 households

Minimising Error Caused by Disclosure Protection

- Minimise the number of cells used for a calculation
- Perform calculations with cells from the same table
- Clustering variables calculated as rates from one table
- Chosen rates show a wide range of different values between Output Areas

Minimising the number of cells used for a calculation

- Important for spatial analysis
- 20 minute drive time could use:
- Sum of 500 Output Area values

OR

- Sum of 16 whole Ward values + 100 Output Area values from split wards
- Option 2) gives more accurate results because fewer small cells are used

What the census doesn't tell us

- Snap-shot in time
- No data about individuals
- Not 100% sample
- Disclosure protection
- No direct data about income or wealth
- Rurality not well measured
- Nothing about product/brand/media usage or attitudes

How can these problems be overcome?

- **By using census data in combination with other datasets to produce classifications**

PEOPLE



&

PLACES

External Consultants helping to develop P²

- Working with Prof Peter Batey and Dr Peter Brown
- Both based at Department of Civic Design, Liverpool University
- Set-up first Teaching Company in the geography and planning fields
- They developed Super Profiles with Littlewoods
- Over 20 years of commercial spatial targeting system experience
- Both worked with 1966 census data

A History of Super Profiles

- 1985 Developed by Openshaw, Batey and Brown
- 1986 Launched commercially and subsequently bought by CDMS (Part of Littlewoods)
- 1994 Updated using 91 Census data
- 1998 Sold by Littlewoods to Claritas UK
- 2003 Claritas UK sold to Axciom
- 2004 Will be updated for postcode changes but not redeveloped using 2001 census data

How P² People & Places is being developed

- Similar hierarchical structure to Super Profiles
- Separate clustering for:
 - Northern Ireland
 - Scotland
 - London ISBA region
 - Rest of England and Wales

Categories of Census Variables in P²

- Age
- Sex
- Social class and social grade
- Ethnic group and language
- Employment/Qualifications
- Housing/Household composition
- Car ownership
- Health
- Religion

What are the broad conclusions of analysis?

- Following cluster and principle component analysis we can identify the 3 strongest dimensions of variation in consumer behaviour
- They are:
- Income
- Lifestage
- Rurality

Strengthening P² in these dimensions

- Neighbourhood Statistics data at Ward level
- Store location data
- PAF information
- TGI Data

Neighbourhood Statistics

- **Child Benefit, Family Credit and Income Support claimants**
- **Working Families tax credit**
- **Attendance Allowance, Disability Living Allowance**
- **House prices and council tax bands**
- **Indices of Deprivation**
- **Countryside Agencies ward level definition of rural areas**

Target Group Index Data and P²

- Survey from BMRB with a sample size of about 30,000 and covers:
- Product usage
- Brand usage
- Media consumption
- Attitudinal questions
- Demographic questions including income

Keeping P² up to date

- ONS produces mid-year population estimates
- Neighbourhood statistics data is updated
- Store location data is updated
- PAF is updated quarterly
- TGI is updated quarterly

How can I tell which classification is best?

- TEST
- Don't be influenced by the data used to build the classification
- Profile your customer files with all the classifications you want to shortlist
- Choose the classification which gives the best discrimination at the lowest cost

How customer data is profiled

- Take the customer file for a product
- Assign a people classification code to each customer
- Measure which groups are over and under represented

A Profiling report

Profile report of "Marque Customers" by Lifestyle Households										
I	Cc	Description	Sample		Base		IIUEA	100		
			Cour	%	Cour	%				
1	J	Affluent Achievers	1:	2:	232€	9.30	215	[Bar chart showing percentage of total sample]		
2	I	Country Life	2	4:	82€	3.31	137	[Bar chart showing percentage of total sample]		
3	I	Thriving Greys	€	1:	290	11.62	130	[Bar chart showing percentage of total sample]		
4	(Settled Suburbans	7	1:	276	11.07	109	[Bar chart showing percentage of total sample]		
5	I	Nest Builders	€	1:	359	14.38	104	[Bar chart showing percentage of total sample]		
6	I	Urban Venturers	5	€	259	10.39	81	[Bar chart showing percentage of total sample]		
8	(Senior Citizens	3	€	212€	8.49	63	[Bar chart showing percentage of total sample]		
9	I	Producers	5	€	377	15.08	61	[Bar chart showing percentage of total sample]		
10		Hard-pressed Familie	1	2	1704	6.81	40	[Bar chart showing percentage of total sample]		
11	.	Have-nots	2	€	2341	9.36	38	[Bar chart showing percentage of total sample]		

How do you compare classifications?

Leventhal, B (1995)

**“Evaluation of Geodemographic Classifications”
Journal of Targeting, Measurement and Analysis
for Marketing ISSN 0967 3237
Vol 4 Part 2 Pages 173 - 183**

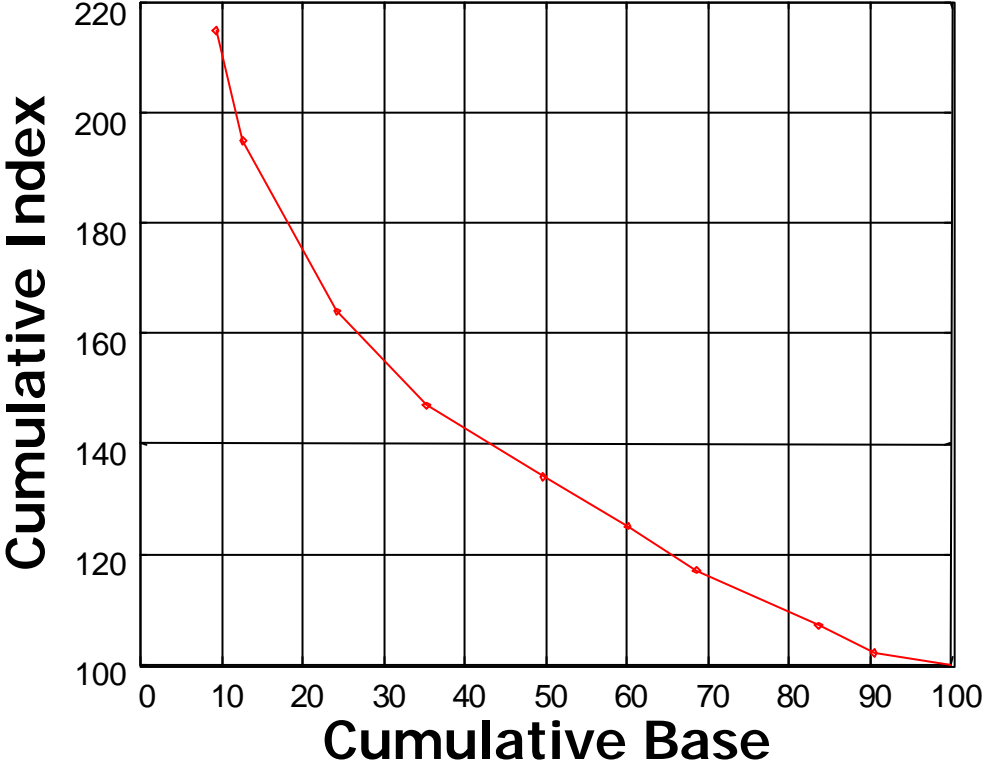
Comparing classifications using gains charts

- Specify common base and obtain profile reports from suppliers
- Order by index
- Calculate cumulative sample, base and index
- Graph cumulative base against cumulative index
- Compare at fixed percentages of the base
e.g. the 10 decile points

Calculate cumulative sample, base and index

Sample	Base	Cumulative Sample	Cumulative Base	Cumulative Index
20.04	9.30	20.04	9.30	215
4.53	3.31	24.57	12.61	195
.....
3.55	9.36	100	100	100

Gains Chart



Summary

- Census 2001 gives us a wealth of detailed data with no economic barriers to access
- But you need to be aware of disclosure protection error
- And you will probably get better results by adding other data
- To compare classifications don't rely on blurb
- TEST