Applying geodemographics

Joint CGG OACUG Seminar at the MRS
Northborough Street
3rd November 2008

Martin Callingham
Visiting Professor
Birkbeck College, University of London
martincallingham @yahoo.com

Associations with behaviour

Intuitively, we do think that people who live in different types of places as being different and doing different things.

Logically this should be the case as demographics are associated with behaviour

And geodemographics ***locate*** mixtures of demographics

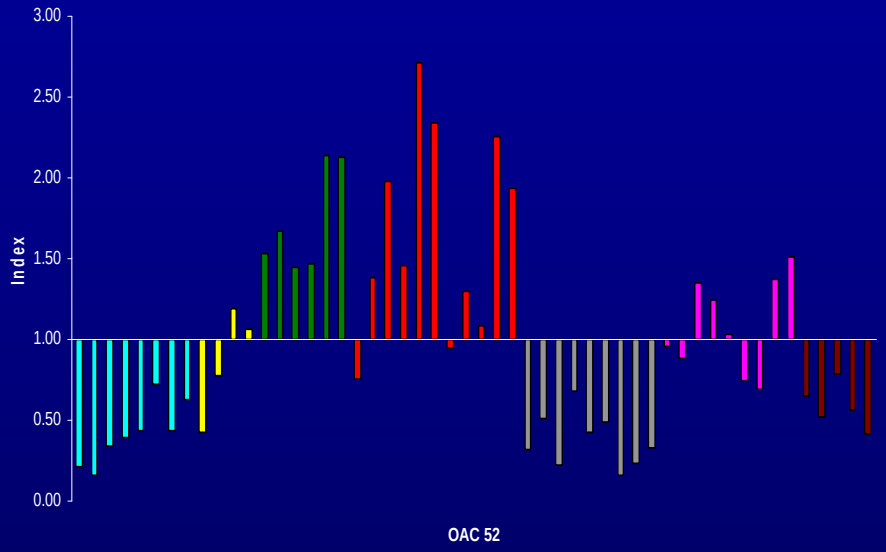
A break through

In 1979, three workers at BRMB (J Bermingham, C McDonald and K Baker), using TGI, found that there were strong associations between purchasing behaviour and geodemographic classifications

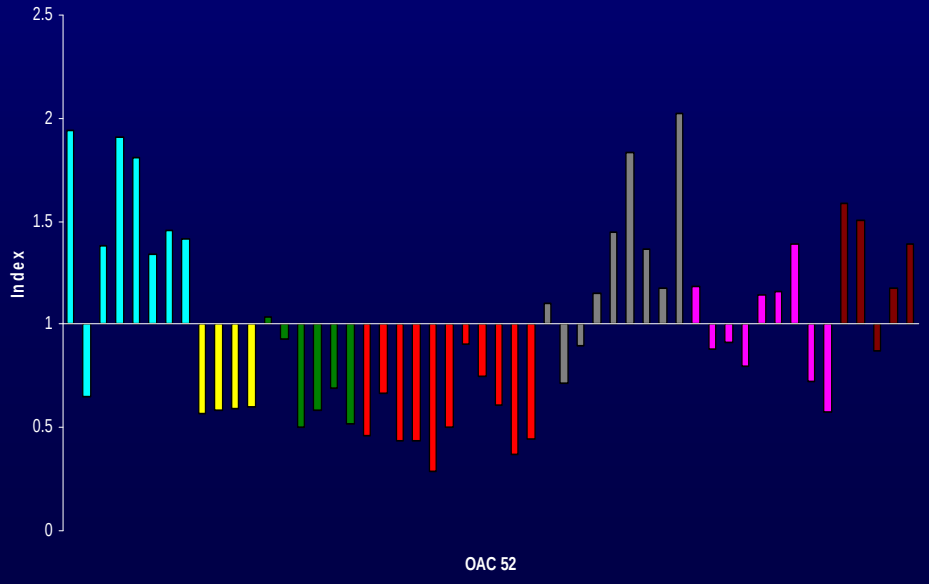
This set the scene for the world wide growth in proprietary geodemographic systems in which Richard Webber had a substantial hand.

Profiling users – newspaper readership

Daily Telegraph



Daily Sport



Blue Collar Workers
City Living
Countryside
Prosperous Suburbs
Constrained by Circumstances
Typical Traits
Multicultural

Various geodemographic systems

Mosaic from Experian

Acorn from CACI

People and Places from Beacon Dodsworth

Output Area Classification from ONS (OAC)

Geodemographics systems vary

All use similar methods to classify areas of clustering areal profiles, but :

- Vary in the hierarchy of the classification
- Vary in areal units (postcode or output area)
- Seem to vary in the amount of non- census data used to
  - *make* the classification
  - to *describe* the classification
- Vary in price and level of support given

But these variations are nothing like as important as knowing what use geodemographics can be put to. They will all do the job.

OAC's hierarchy of names

Super	Groups	Sub-groups
1: Blue Collar Communities	1a: Terraced Blue Collar	3
	1b: Younger Blue Collar	2
	1c: Older Blue Collar	3
2: City Living	2a: Transient Communities	2
	2b: Settled in the City	2
3: Countryside	3a: Village Life	2
	3b: Agricultural	2
	3c: Accessible Countryside	2
4: Prospering Suburbs	4a: Prospering Younger Families	2
	4b: Prospering Older Families	4
	4c: Prospering Semis	3
	4d: Thriving Suburbs	2
5: Constrained by Circumstances	5a: Senior Communities	2
	5b: Older Workers	4
	5c: Public Housing	3
6: Typical Traits	6a: Settled Households	2
	6b: Least Divergent	3
	6c: Young Families in Terraced Homes	2
	6d: Aspiring Households	2
7: Multicultural	7a: Asian Communities	3
	7b: Afro-Caribbean Communities	2

What can they be used for

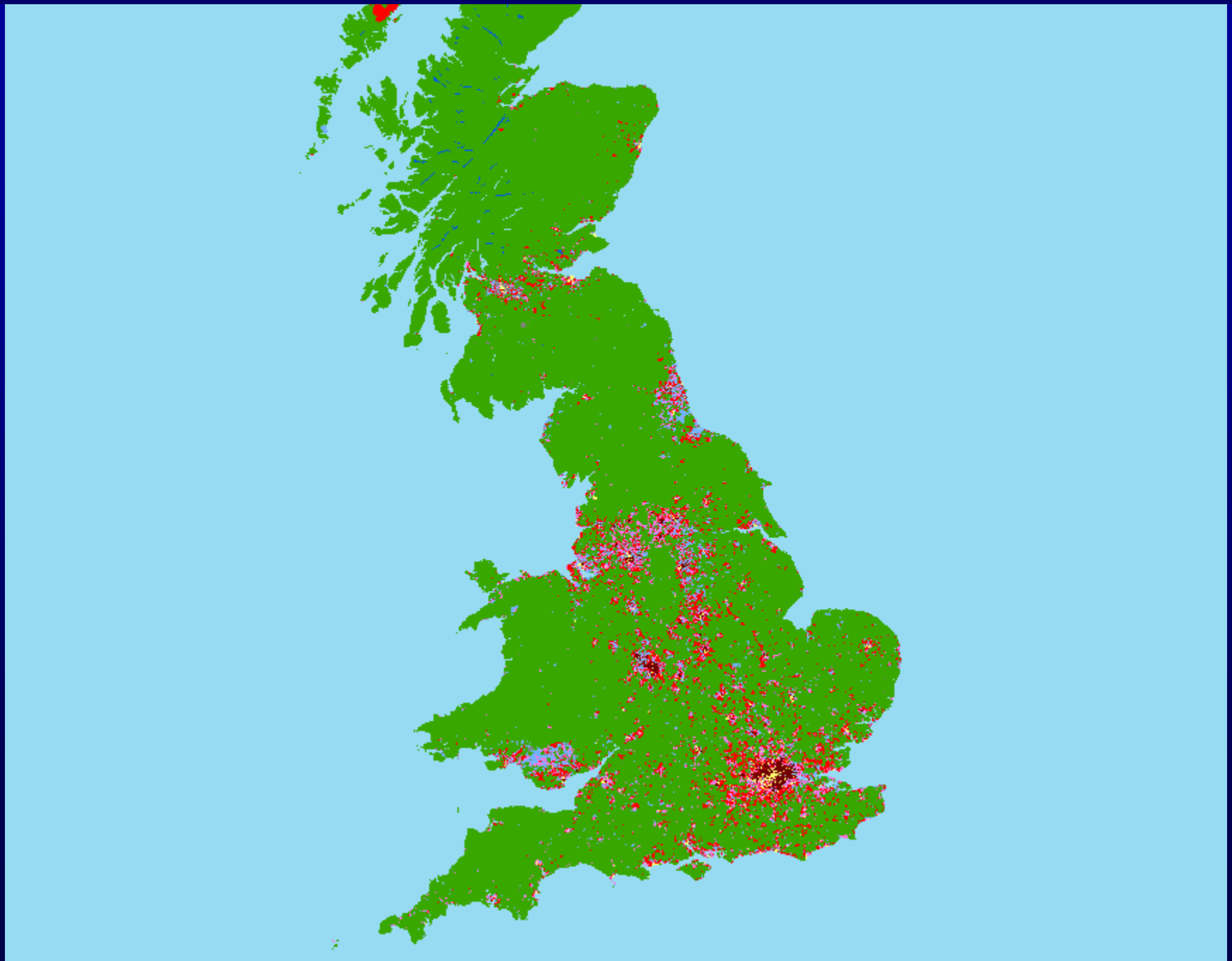
Straight forward mapping of an ***area*** which helps interpret it

Profiling ***users*** of a service which helps understand the nature of them

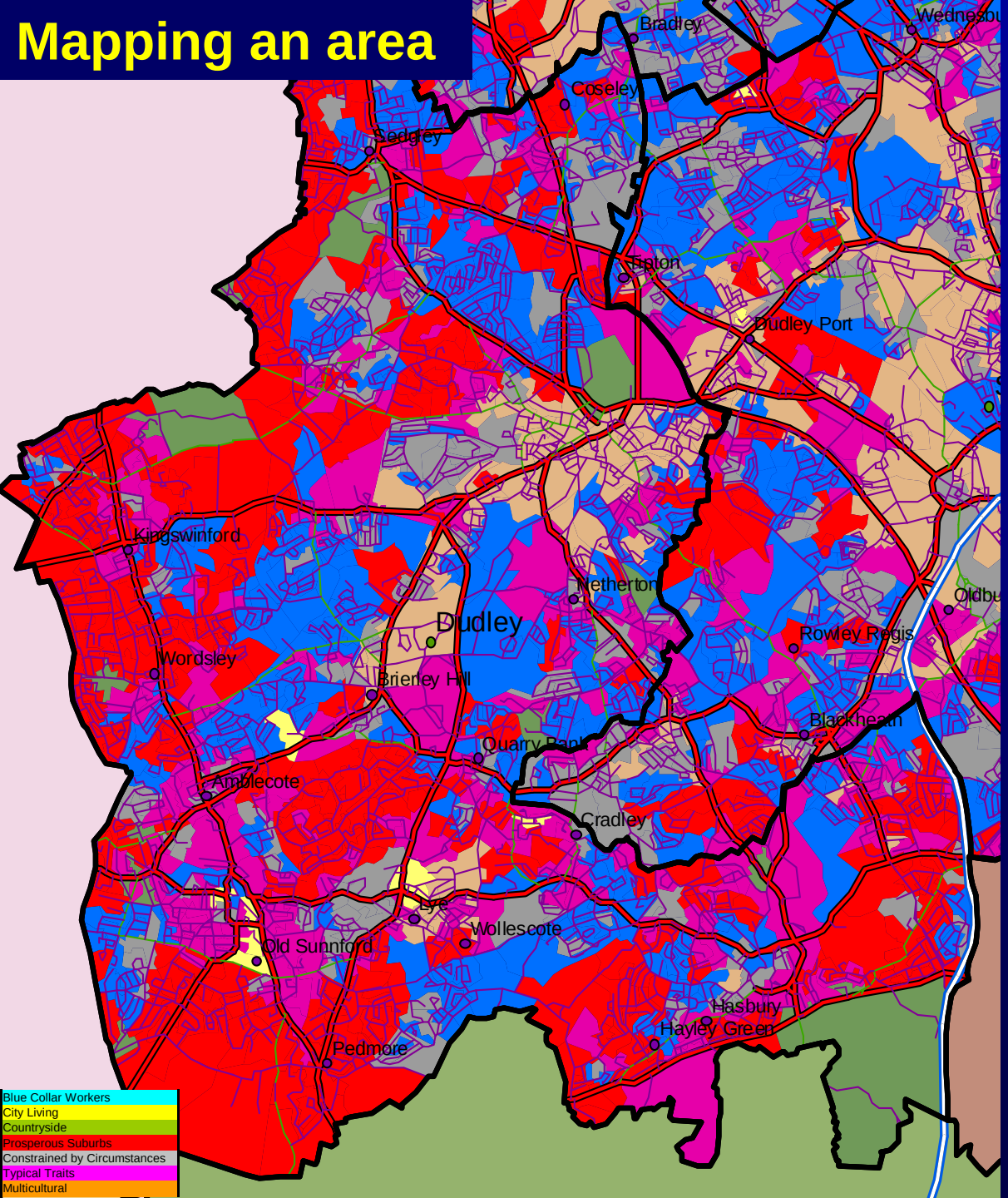
Modelling behaviour or attitude into small areas

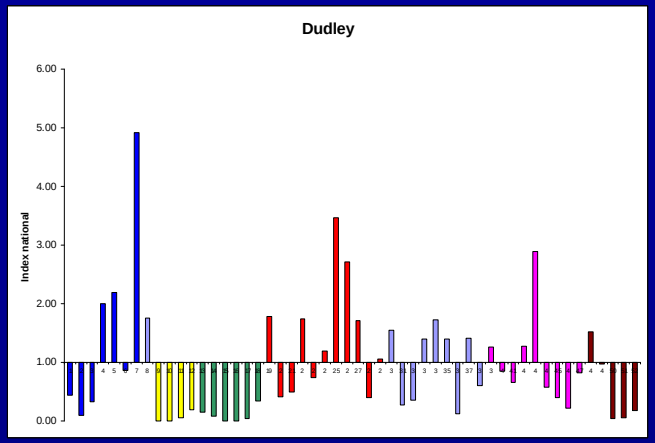
Fusing different data sets

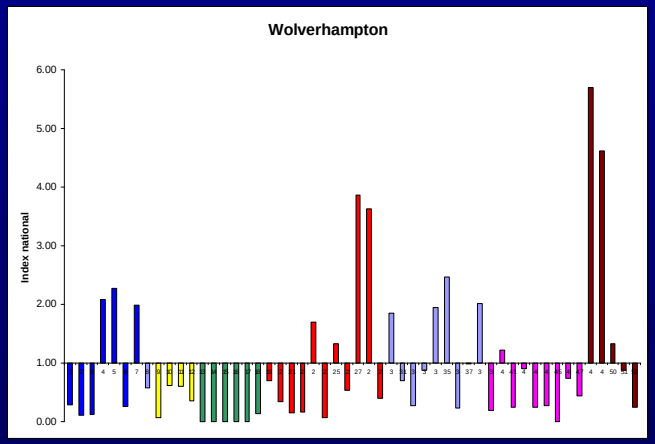
Map based on output areas OAC Super Groups

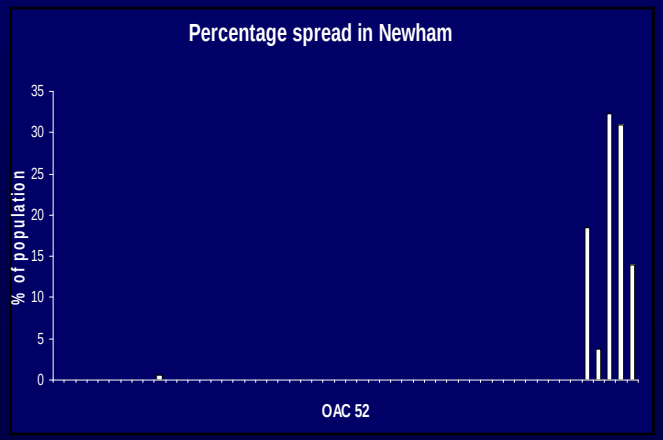


Mapping an area









How is it done - Measuring propensities

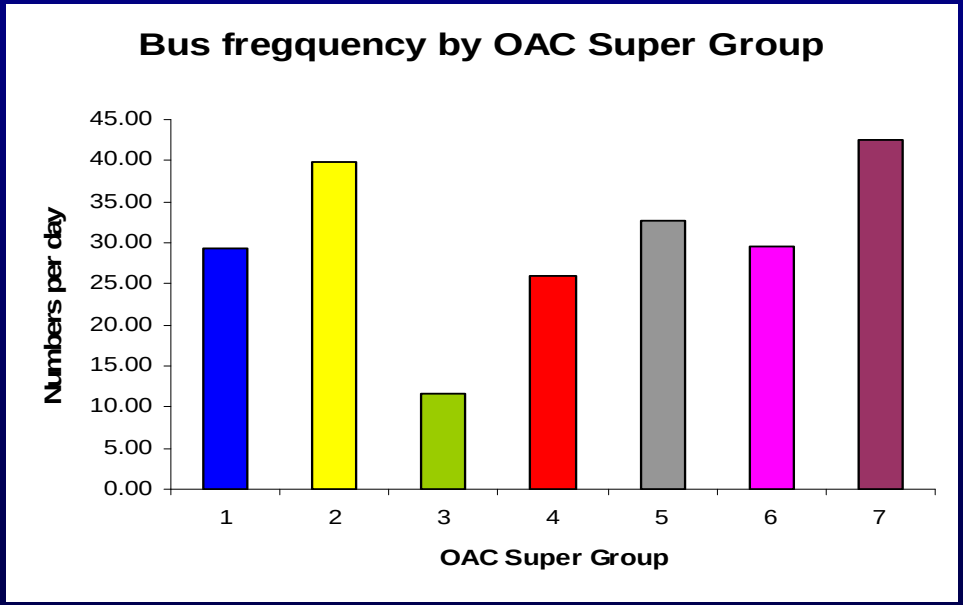
Absolute method

- Conduct a market research survey
- Code respondents from their postcode with a geodemographic type
- Cross tabulate the survey by geodemographics and the question of interest
- Quantify the behaviour of people who live in different Geodems caegories eg how often they eat out a week

Absolute methods

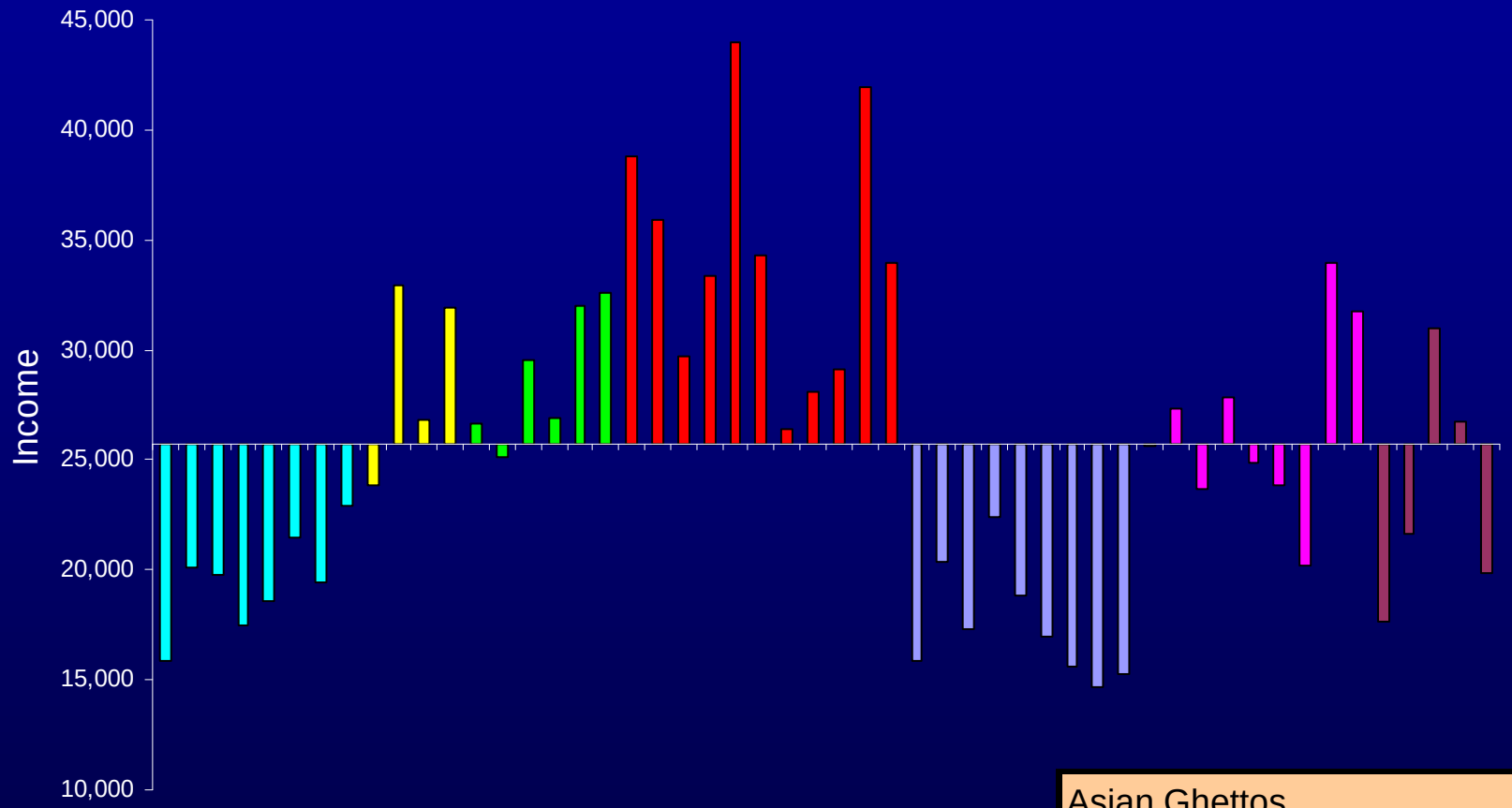
			OAC Super Group						
Frequency of bus service	**Weight**	**Total**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
		%	%	%	%	%	%	%	%
Less than once a day	0.5	0.79	0.00	0.00	5.80	0.21	0.00	0.06	0.00
At least once a day	1	6.28	2.78	0.60	30.68	4.78	1.58	3.14	0.00
At least 1 an hour	12	19.29	18.08	4.42	39.82	23.86	13.46	18.43	1.67
At least 1 every half hour	24	37.69	45.09	25.90	20.65	46.26	40.06	42.66	20.33
At least 1 every quarter hour	48	35.95	34.06	69.08	3.05	24.90	44.90	35.70	77.99
Number of buses		28.68	**29.37**	**39.91**	**11.53**	**25.96**	**32.80**	**29.62**	**42.52**

Sample is 8122



Student Bedsits
Young Cosmopolitan Professionals
Genteel Flatland
Professional Town Families

Mean Household Income



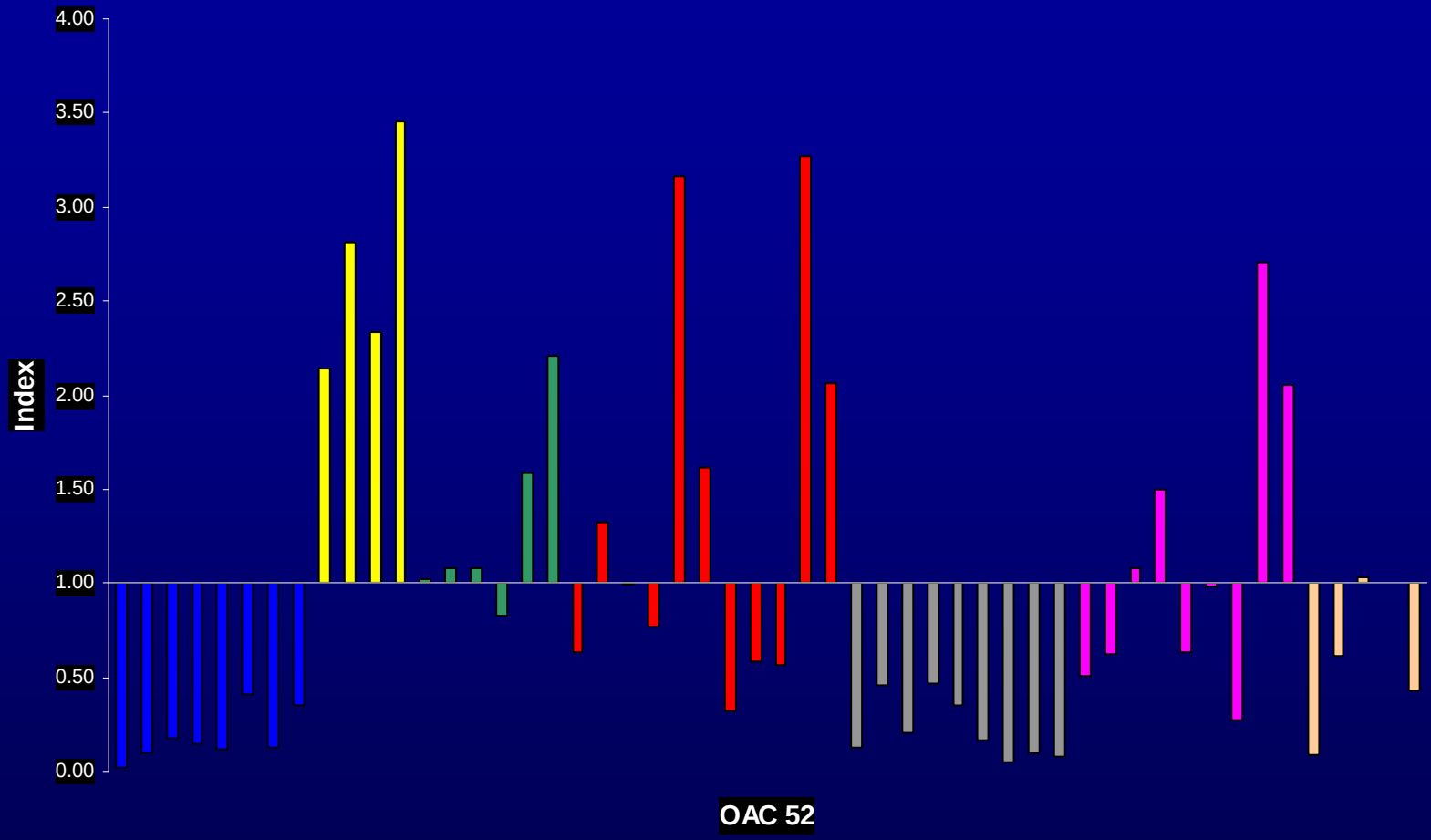
OAC 52

Asian Ghettos
Impoverished Melting Pot
Successful Second Generation
Gentrification
Deprived Tower Blocks

Relative method

- Collect postcodes of customers
- Code them with the geodemographic code
- Calculate the percentage of each code present
- Divide each by the percentage nationally present to get an index
- This is a very easy and powerful way of gaining insight into a list of people who do something.

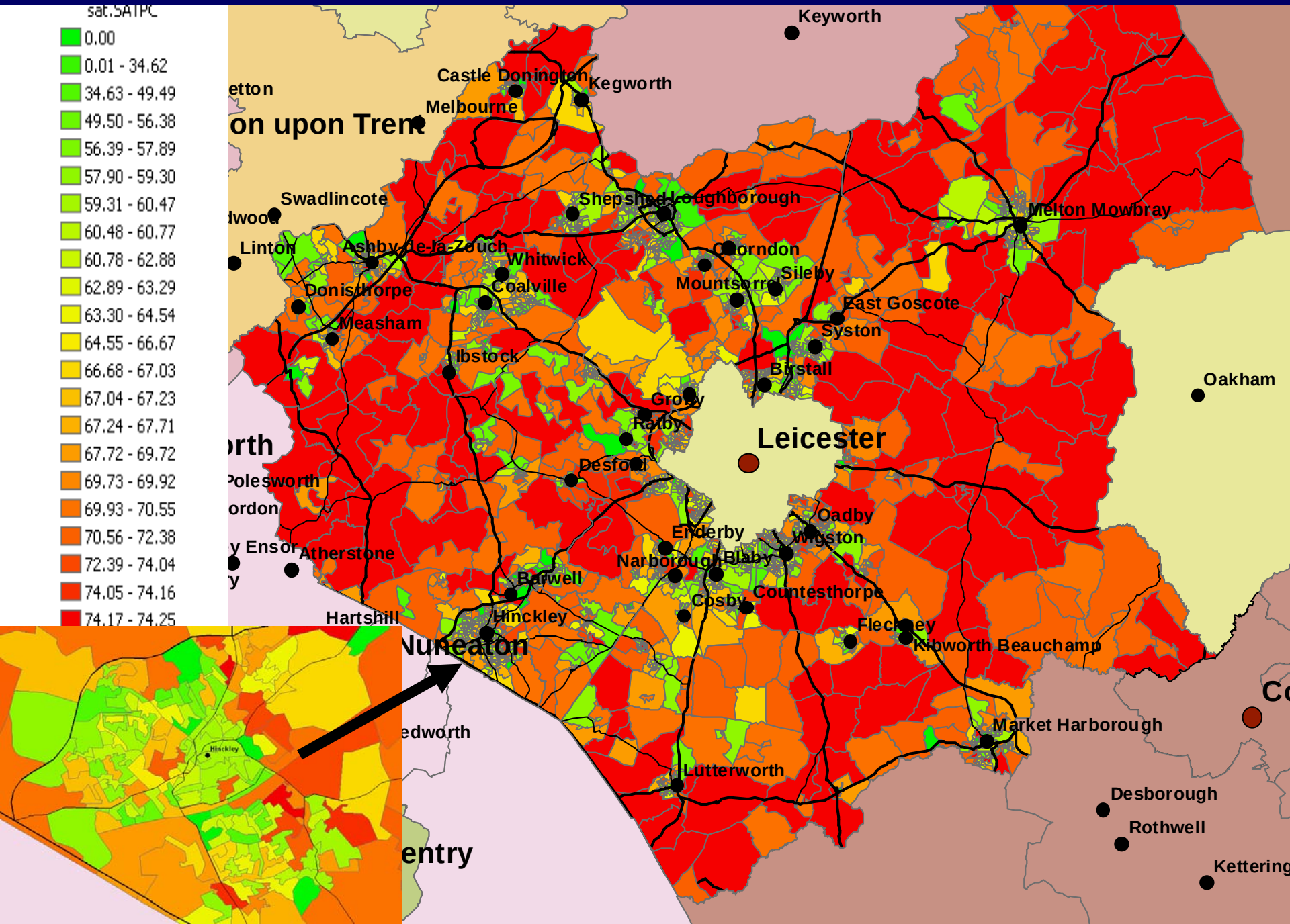
National Index by OAC 52



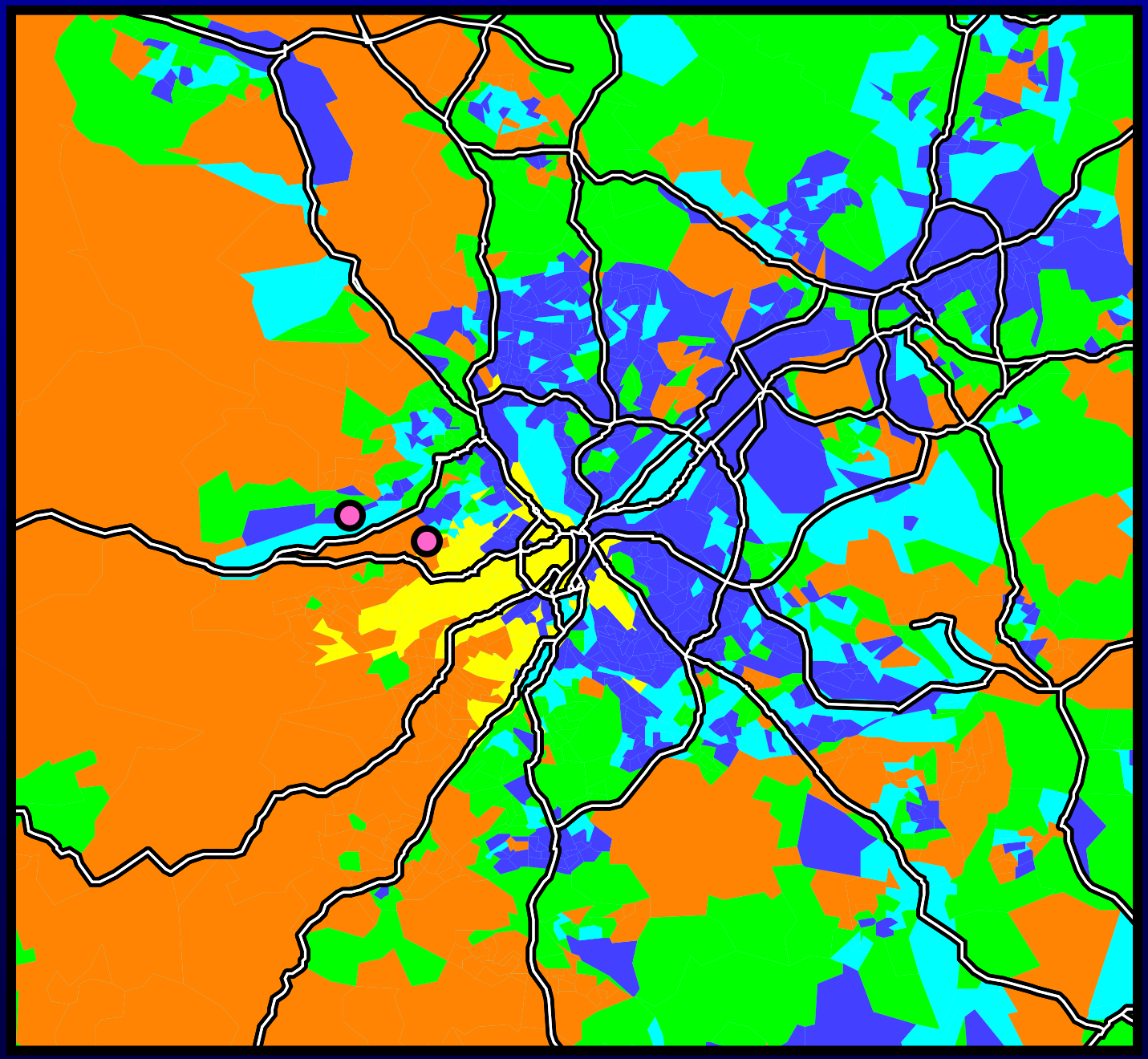
These indices were calculated by firstly getting the proportional spread of the customers across OAC 52 sub groups, ditto for the population of the country and dividing the former by the latter to get an index.

This chart shows substantial variation in index by OAC 52 which is different from the OAC 7 picture we have seen before

Modelled satisfaction with area (% very or fairly satisfied) – mean satisfaction is 64%



Geo-demographics - eating out in Sheffield - Site report



Calculating a site demand report

Collect the output areas within each catchment area by GIS

Ascribe the population to each OA together with its OAC code

Count the number of people of each OAC type in both catchment areas

For each OAC type, we know the frequency of eating out different types of meals per person

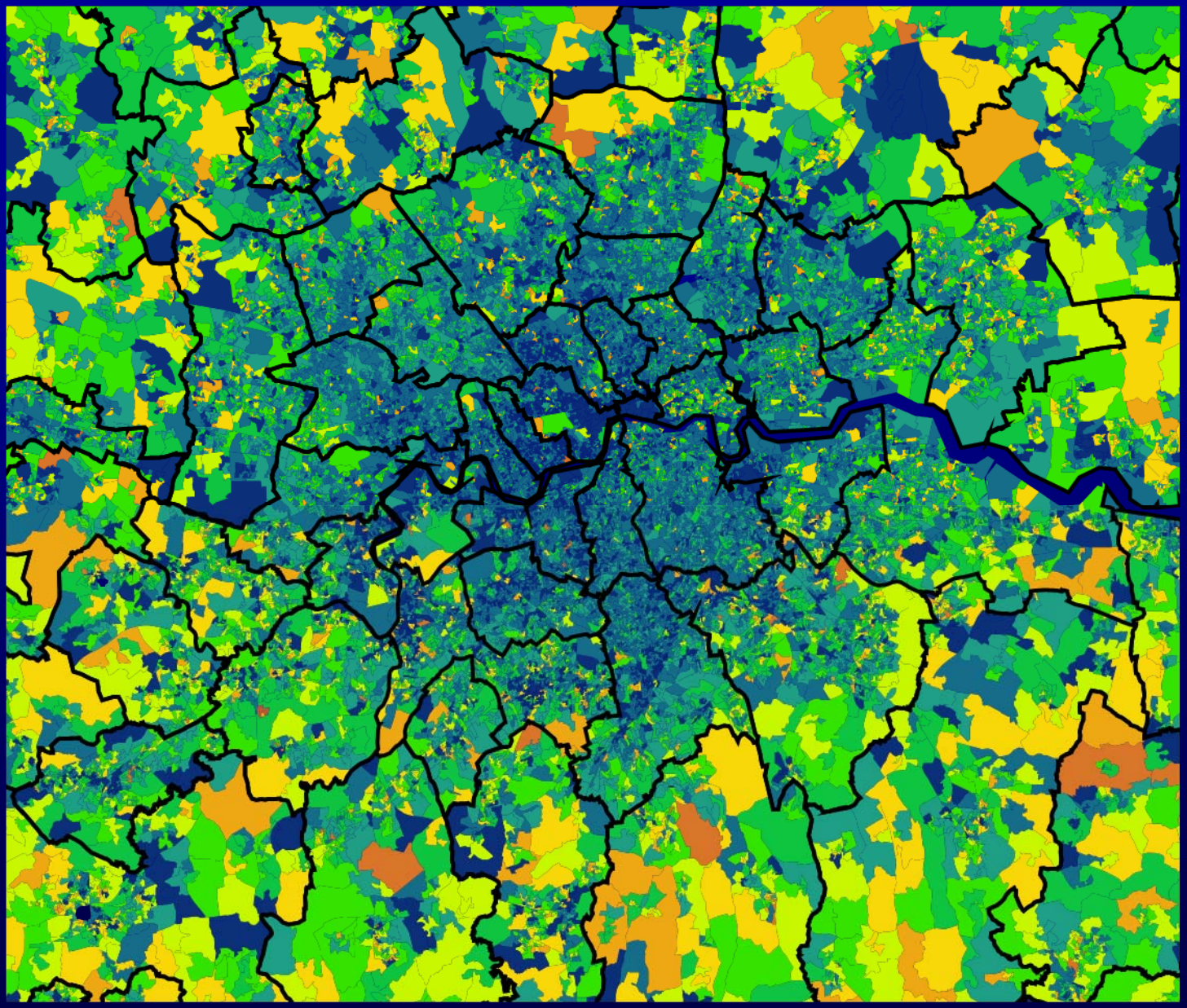
Therefore, can multiply the total people of each OAC type by the number of meals of each type that they consume.

Sum these meals up for each catchment area to get a demand report.

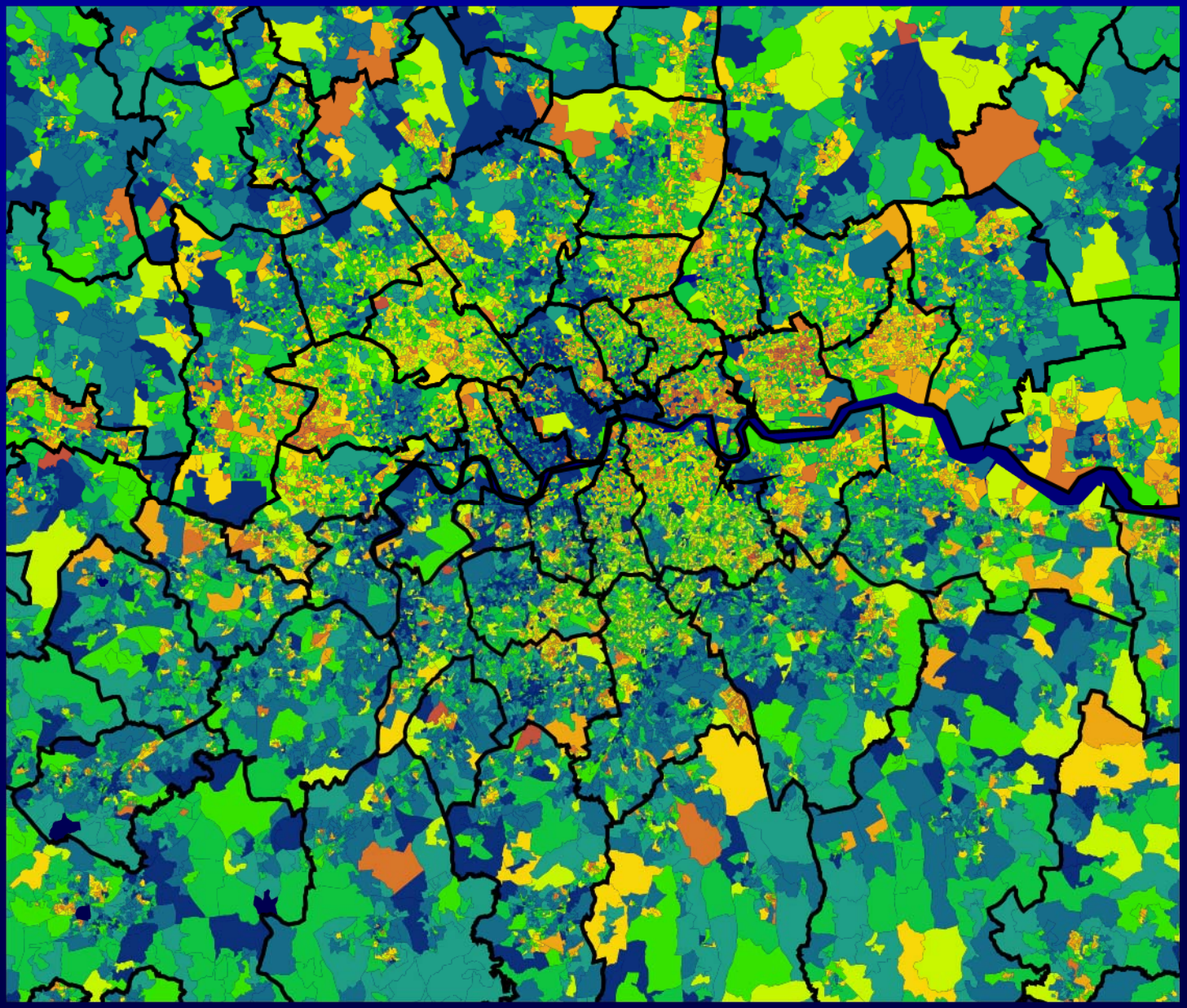
	National
Population	90,468
Total meals (last 7 days)	57,898
	%
Fast Food	***26.80***
Burger House	
Fish and chips	
Other	
Casual Dinning off	***18.60***
Pizza House	
Chinese Take-away	
Indian Take-away	
Other	
Casual Dining On	***26.9***
Pizza - On	
Chinese- On	
Indian-On	
Pubs	
Steak House	
Other Casual On	
Formal	***10.90***
Hotel	
Other Restaurant	
Other	***17.60***
Cafe	
Other	
Meal pers head	**0.64**

	National	Site 1	Site 2
Population	**90,468**	**34,092**	**38,010**
Total meals (last 7 days)	**57,898**	**19,303**	**29,402**
	%		
Fast Food	***26.80***	***6,049***	***6,936***
Burger House			
Fish and chips			
Other			
Casual Dinning off	***18.60***	***4,133***	***5,140***
Pizza House			
Chinese Take-away			
Indian Take-away			
Other			
Casual Dining On	***26.9***	***4,203***	***8,212***
Pizza - On			
Chinese- On			
Indian-On			
Pubs			
Steak House			
Other Casual On			
Formal	***10.90***	***1,610***	***4,072***
Hotel			
Other Restaurant			
Other	***17.60***	***3,388***	***5,311***
Cafe			
Other			
Meal pers head	**0.64**	**0.57**	**0.77**

Modelled stroke Incidence

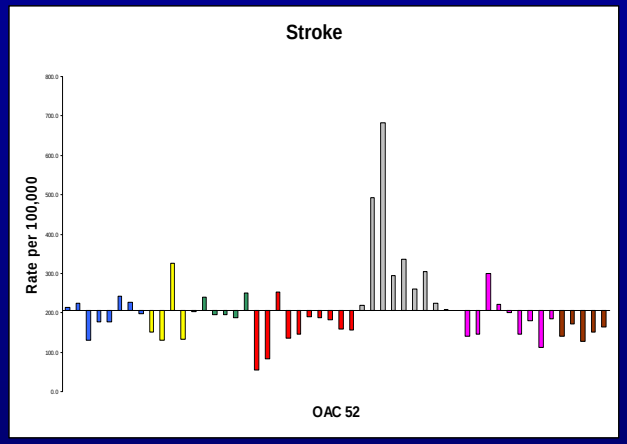


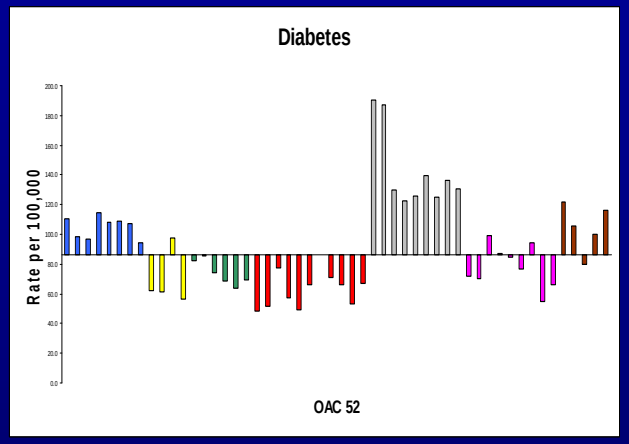
Modelled diabetes incidence

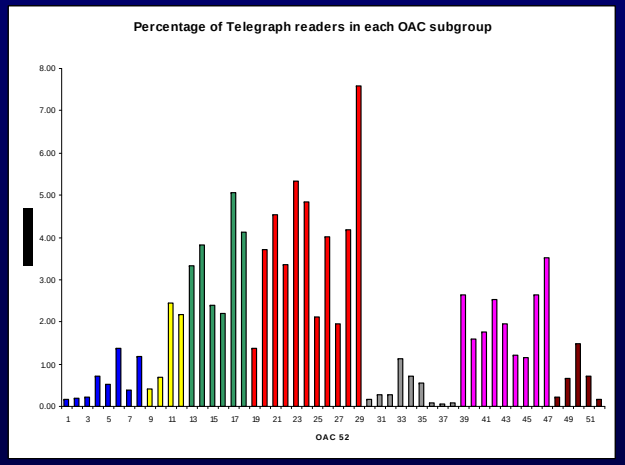


Fusing data sets

If two data sets have Geodem profiles, then it is possible to fuse information from one to the other.







Telegraph readers have:

	% of national rate
Stroke	96%
Diabetes	82%
Asthma	76%

Modelled disease rates by Local Authority

	National	S Tyneside	Wokingham	S Tyneside	Wokingham
Strokes	200	215	161	1.08	0.81
Diabetes	90	104	65	1.16	0.72
Asthma	110	136	80	1.24	0.73

Expansion from the classic use

Use of higher levels of cluster numbers

Modelling from large areas to smaller areas

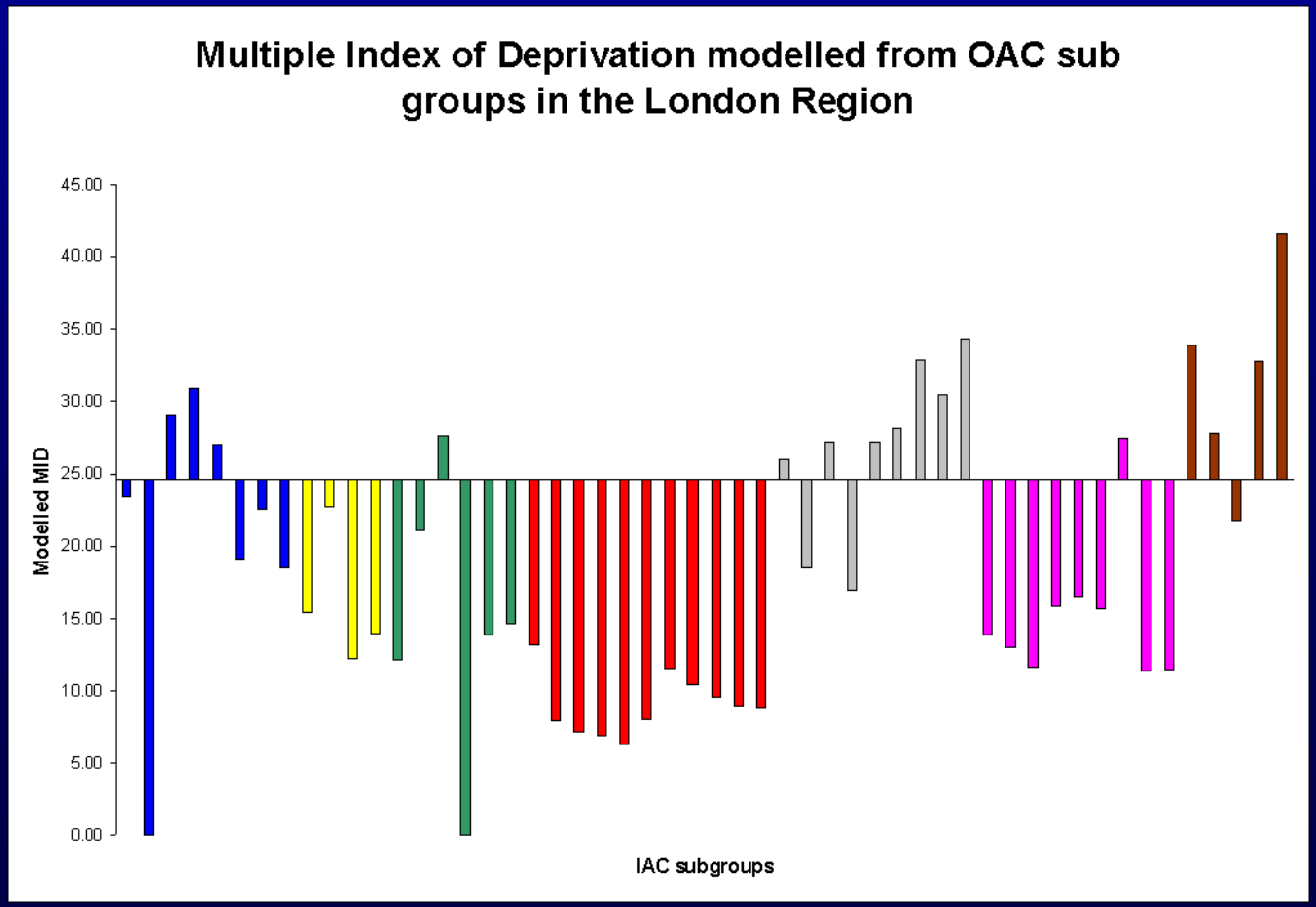
Estimating residuals for undisclosed health or crime incidence.

Modelling from large to smaller areal units

This is the modelled MID for each OAC subgroup(OAC 52) in the London Region.

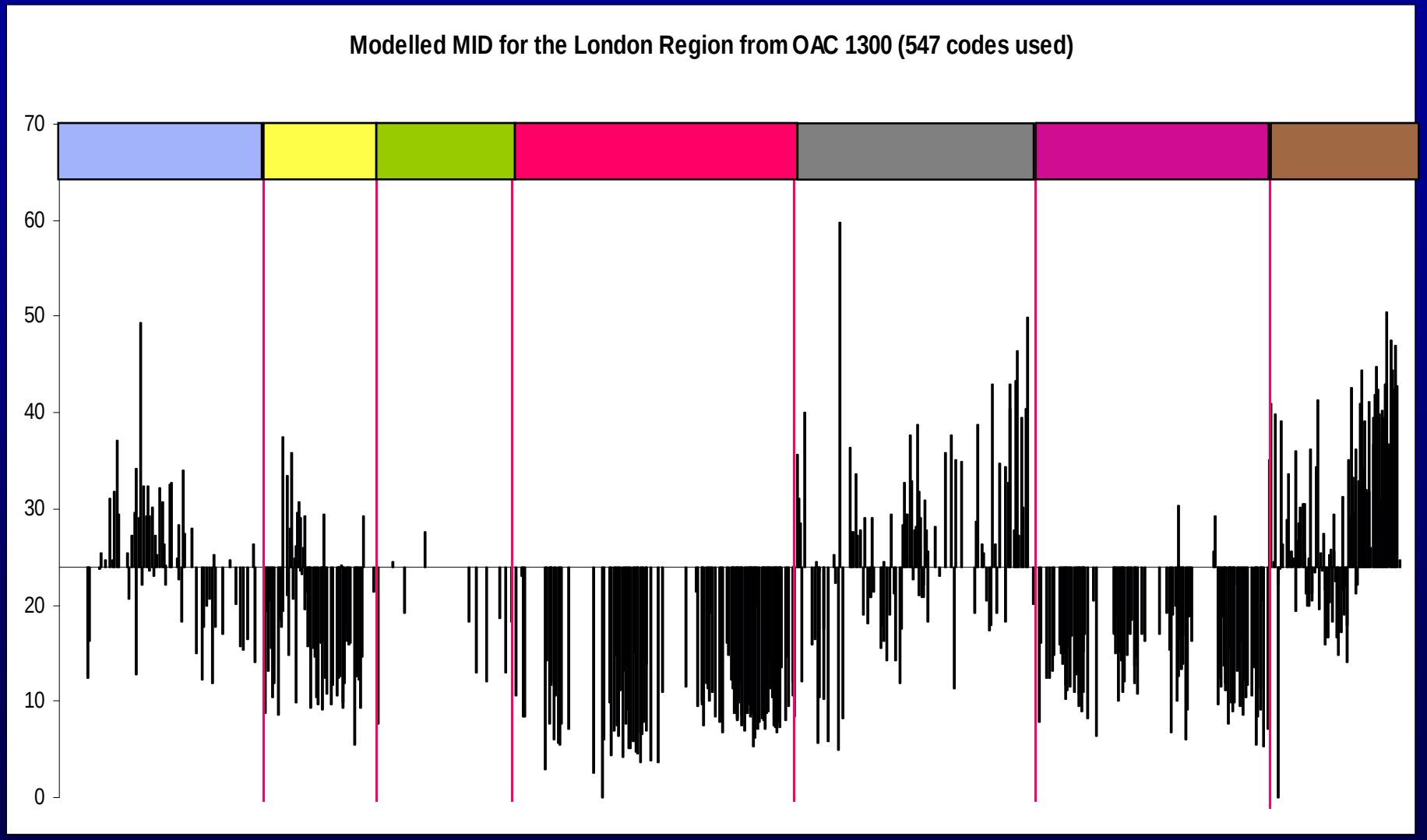
Note that some OAC subgroups are only barely present.

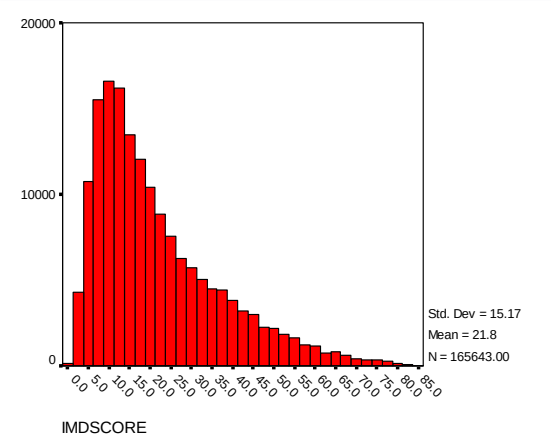
Two are not present at all and are shown as zero MID.

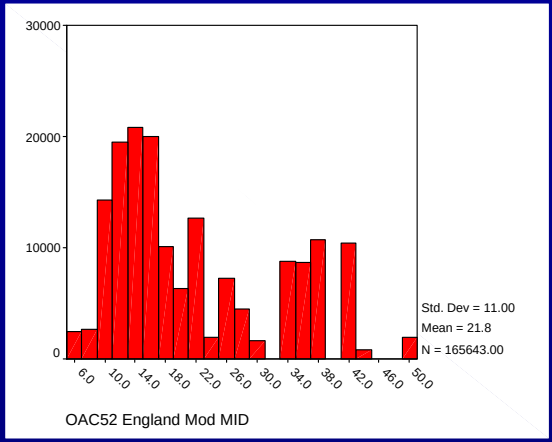


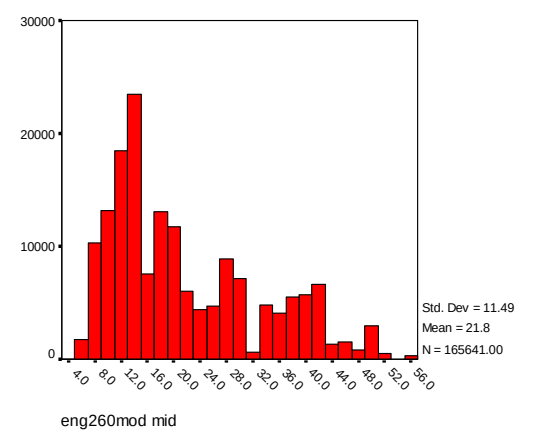
Sub group	Modelled MID	Subgroup Counts
1a1	0.00	0
1a2	23.46	30
1a3	29.10	57
1b1	30.98	288
1b2	27.13	156
1c1	19.22	19
1c2	22.62	25
1c3	18.58	31
2a1	15.40	141
2a2	22.71	2452
2b1	12.33	485
2b2	13.94	2095
3a1	12.17	3
3a2	21.16	8
3b1	27.63	1
3b2	0.00	0
3c1	13.84	4
3c2	14.69	5
4a1	13.25	3
4a2	7.93	36
4b1	7.19	1
4b2	6.84	23
4b3	6.36	238
4b4	8.05	10
4c1	11.56	1
4c2	10.46	88
4c3	9.56	364
4d1	9.00	913
4d2	8.89	105
5a1	26.07	38
5a2	18.54	14
5b1	27.17	52
5b2	16.94	54
5b3	27.19	293
5b4	28.08	16
5c1	32.96	16
5c2	30.38	50
5c3	34.38	59
6a1	13.90	22
6a2	12.99	769
6b1	11.71	3
6b2	15.82	161
6b3	16.54	5
6c1	15.73	279
6c2	27.46	2
6d1	11.36	923
6d2	11.49	266
7a1	33.84	203
7a2	27.86	1249
7a3	21.77	4092
7b1	32.80	4788
7b2	41.63	3200
Group Tota	24.62	24136

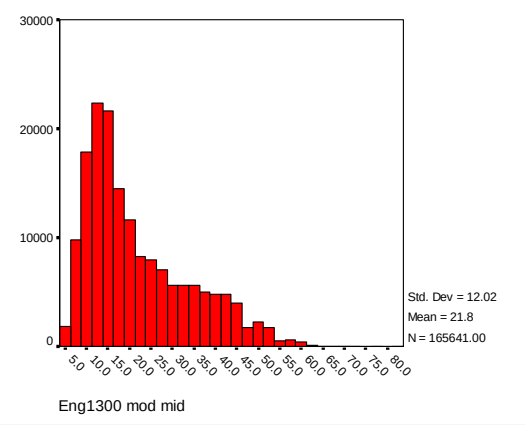
Modelling MID to OAC 1300



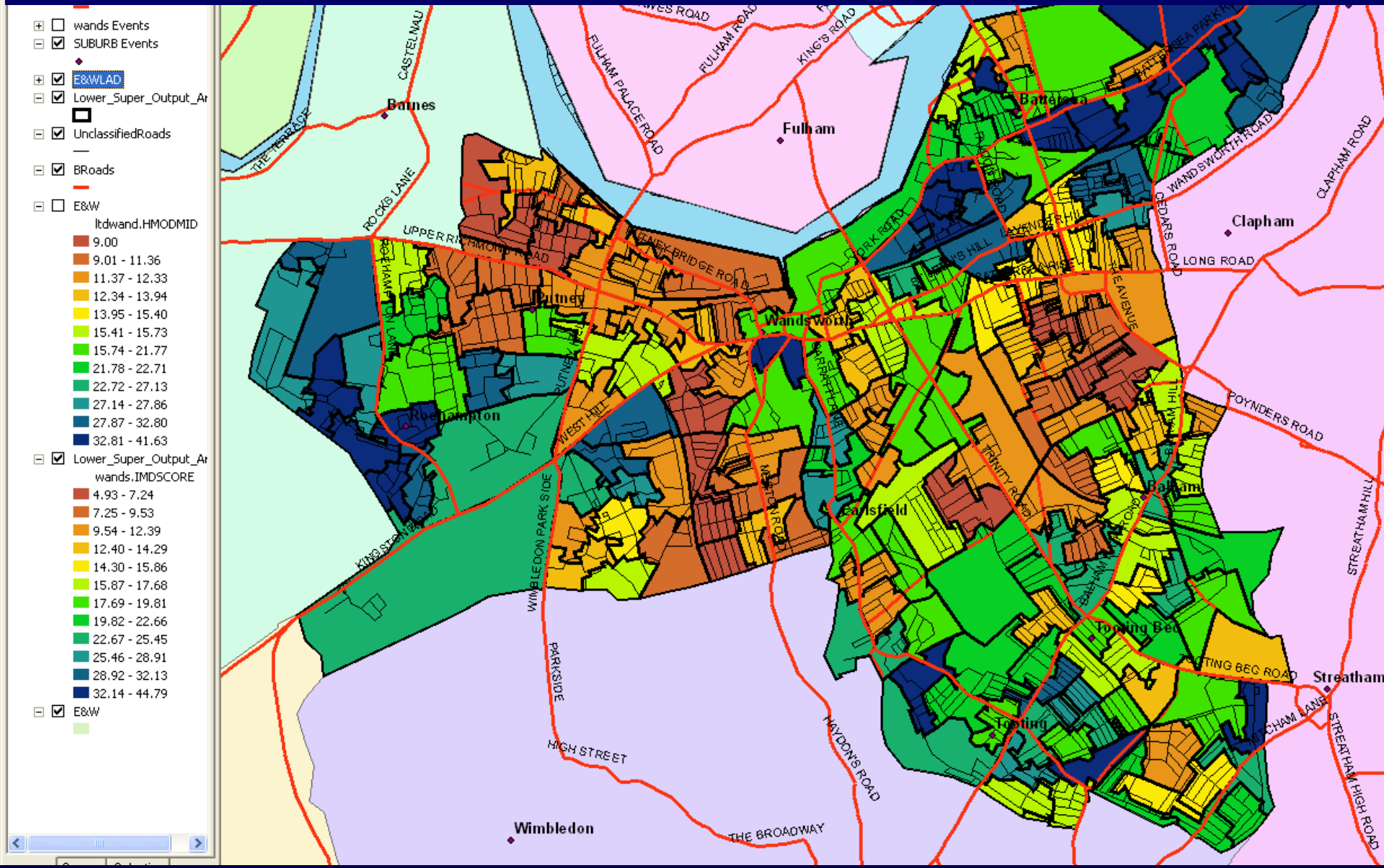

20000
10000
0
0,0 5,0 10,0 15,0 20,0 25,0 30,0 35,0 40,0 45,0 50,0 55,0 60,0 65,0 70,0 75,0 80,0 85,0
Std. Dev = 15.17
Mean = 21.8
N = 165643.00
IMDSCORE


30000
20000
10000
0
6,0 10,0 14,0 18,0 22,0 26,0 30,0 34,0 38,0 42,0 46,0 50,0
Std. Dev = 11.00
Mean = 21.8
N = 165643.00
OAC52 England Mod MID


30000
20000
10000
0
4,0 8,0 12,0 16,0 20,0 24,0 28,0 32,0 36,0 40,0 44,0 48,0 52,0 56,0
Std. Dev = 11.49
Mean = 21.8
N = 165641.00
eng260mod mid

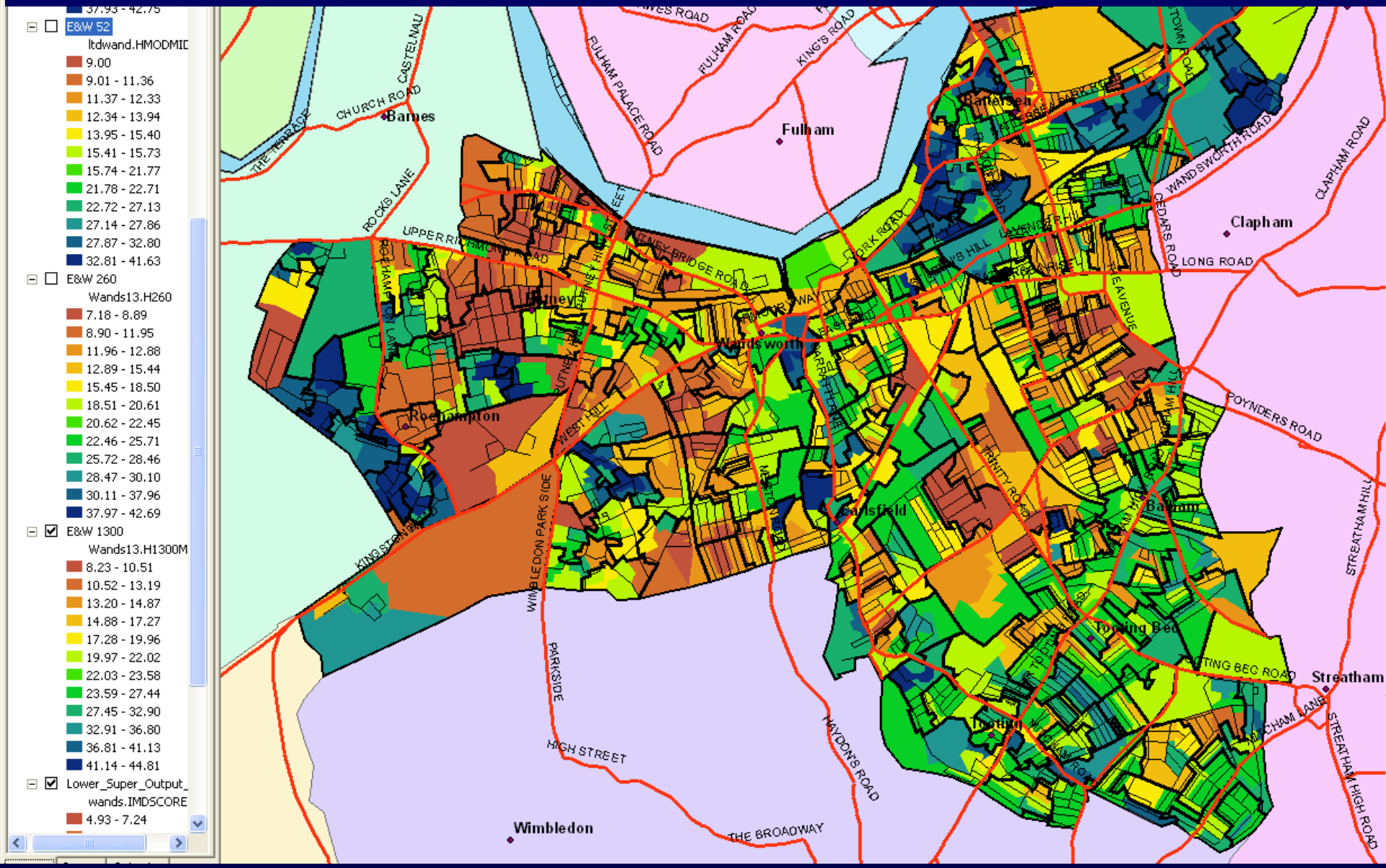

30000
20000
10000
0
5,0 10,0 15,0 20,0 25,0 30,0 35,0 40,0 45,0 50,0 55,0 60,0 65,0 70,0 75,0 80,0
Std. Dev = 12.02
Mean = 21.8
N = 165641.00
Eng1300 mod mid

MID at Super Output areas for Wandsworth



This is the Multiple Index of Deprivation overall score shown at super output area – the bluer the more deprived

MID at OAC 1300 for Wandsworth



Fuzzy geodemographic modelling

Normal geodemographics modelling is done by weighting a propensity by the population of an output area separately by each geodemographic type.

Geodemographic	A	B	C	D	E
Propensity	10	5	2	8	1

OAs	Category	Pop	Propensity	Market
1	A	250	10	2500
2	A	231	10	2310
3	E	213	1	213
4	C	198	2	396
5	A	260	10	2600

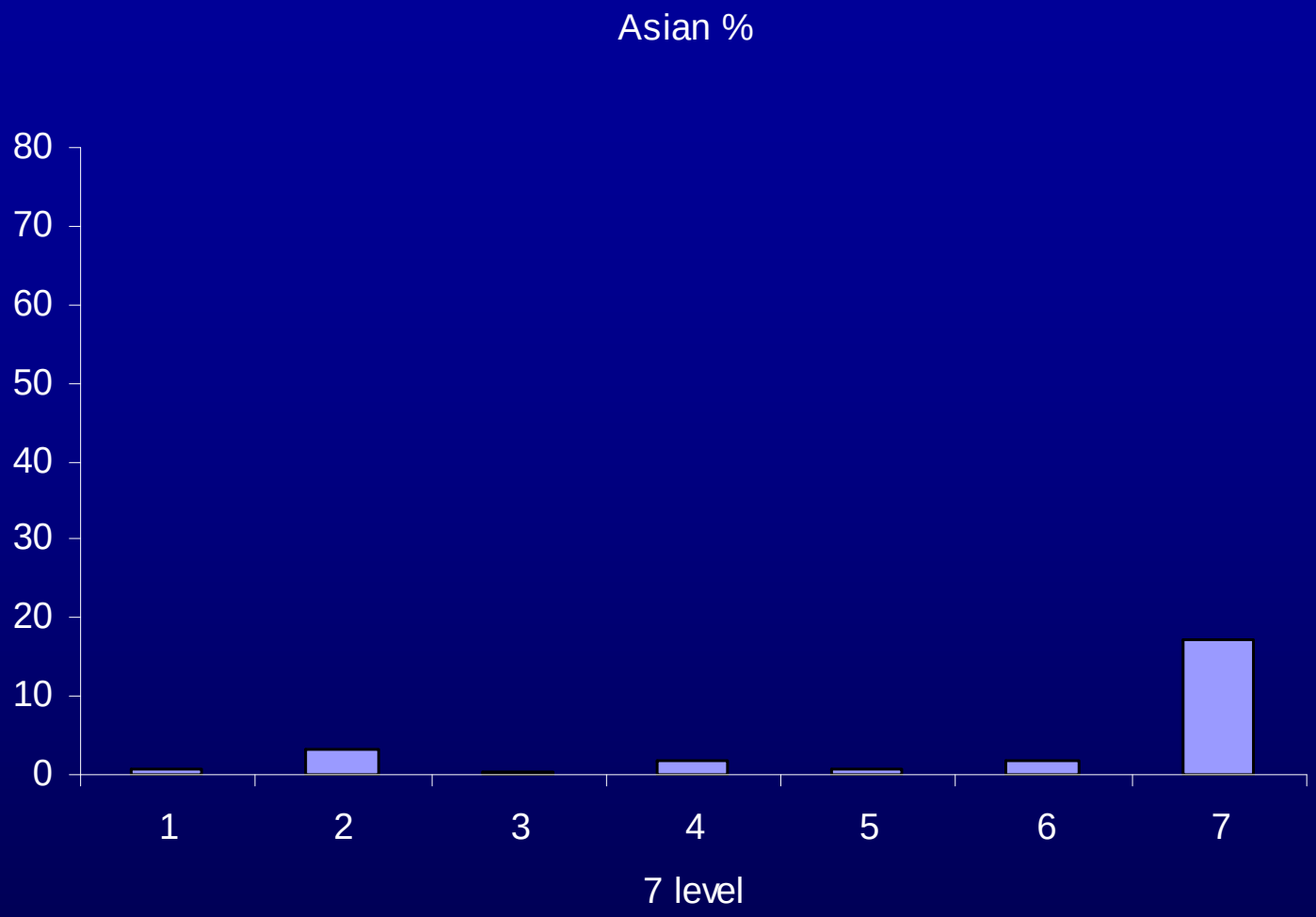
This can be improved by calculating the fuzzy geodemographics set for each output area and then

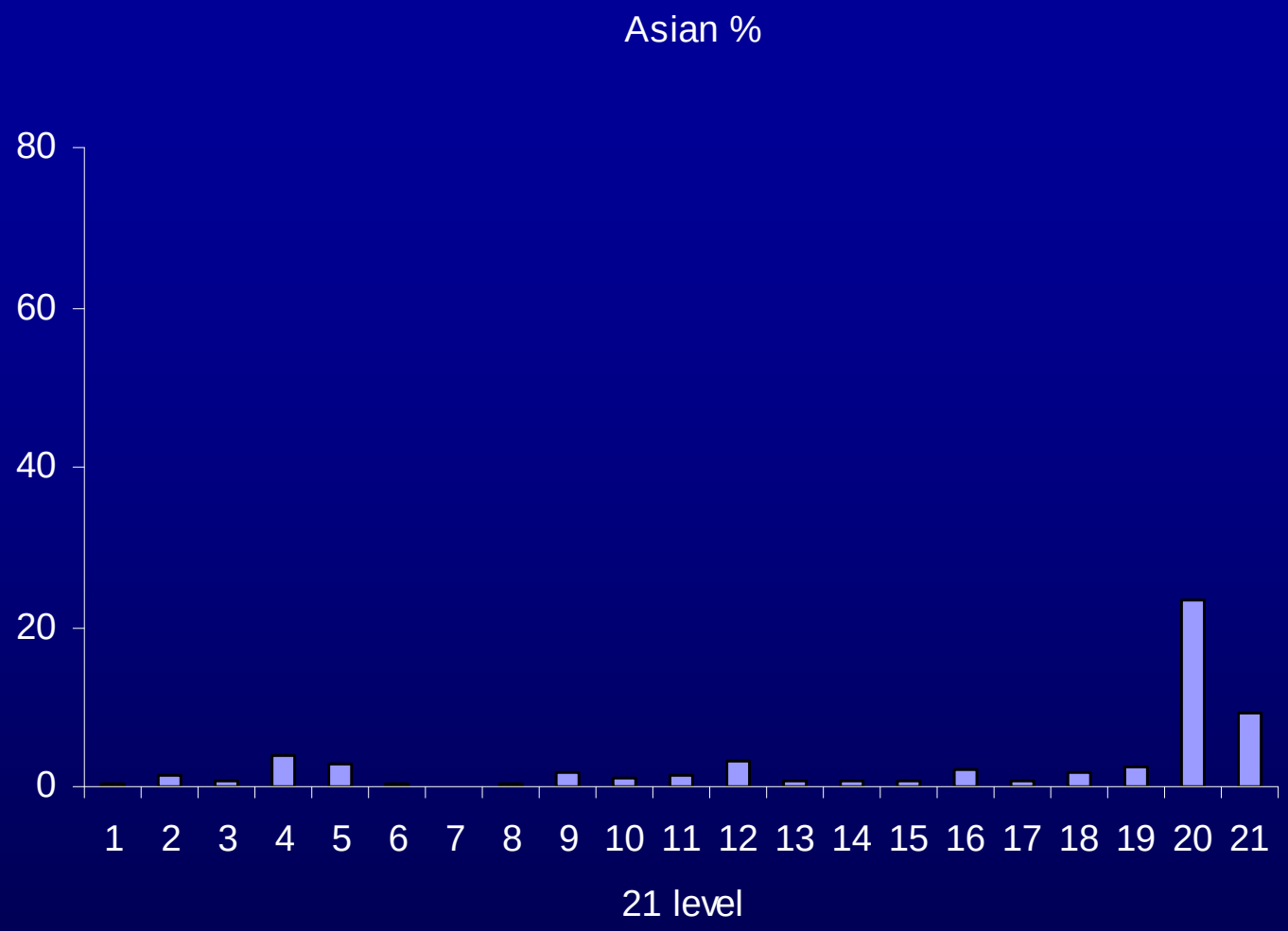
OAs	Category	Pop	A 10	B 5	C 2	D 8	E 1	Wgted Cat	Wgted mkt
1	A	250	72	12	6	7	3	8.51	2128
2	A	231	51	45	1	2	1	7.54	1742
3	E	213	5	5	5	5	80	2.05	437
4	C	198	10	31	45	1	13	3.66	725
5	A	260	37	27	27	1	8	5.75	1495

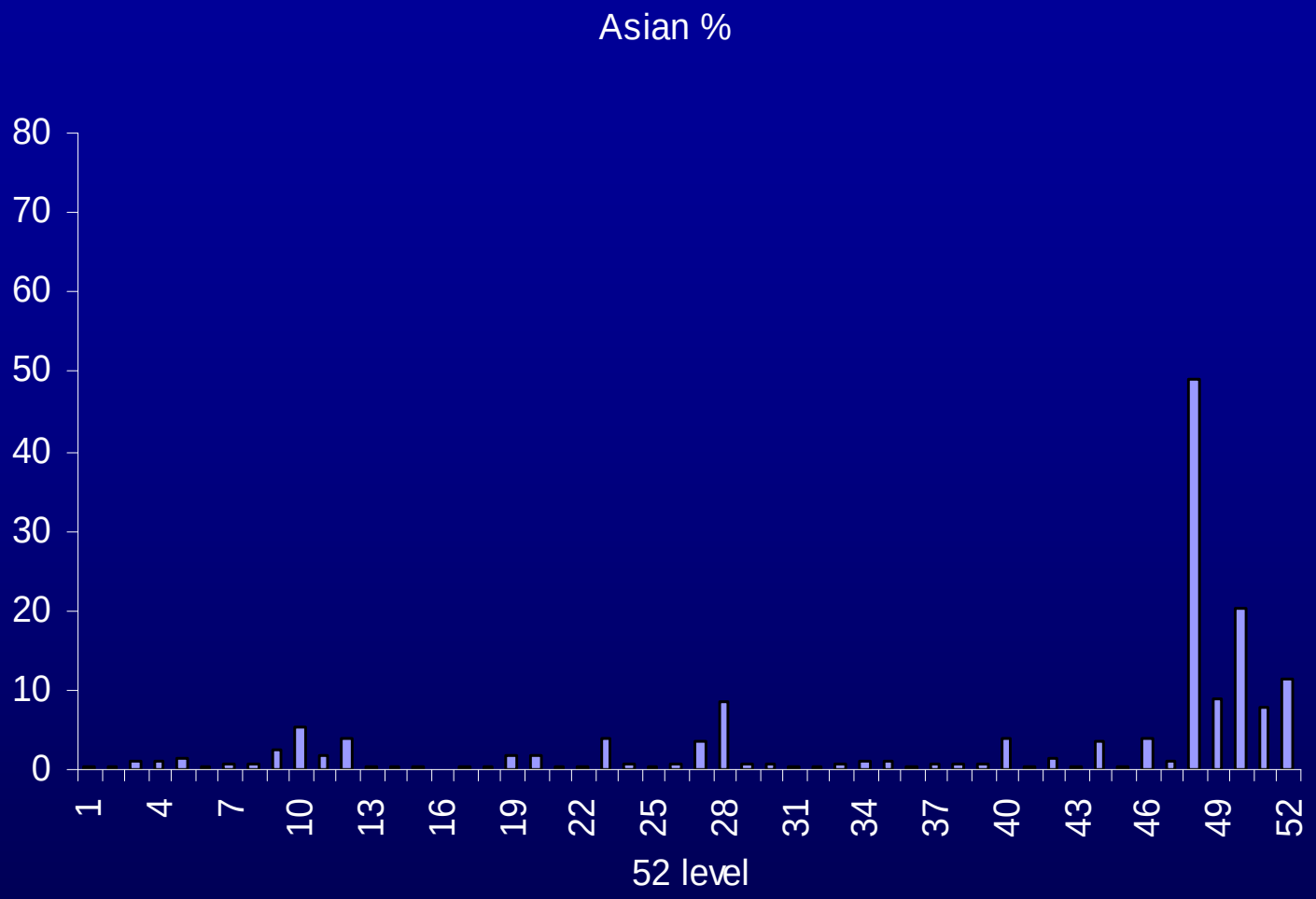
Extending OAC – increasing its discrimination

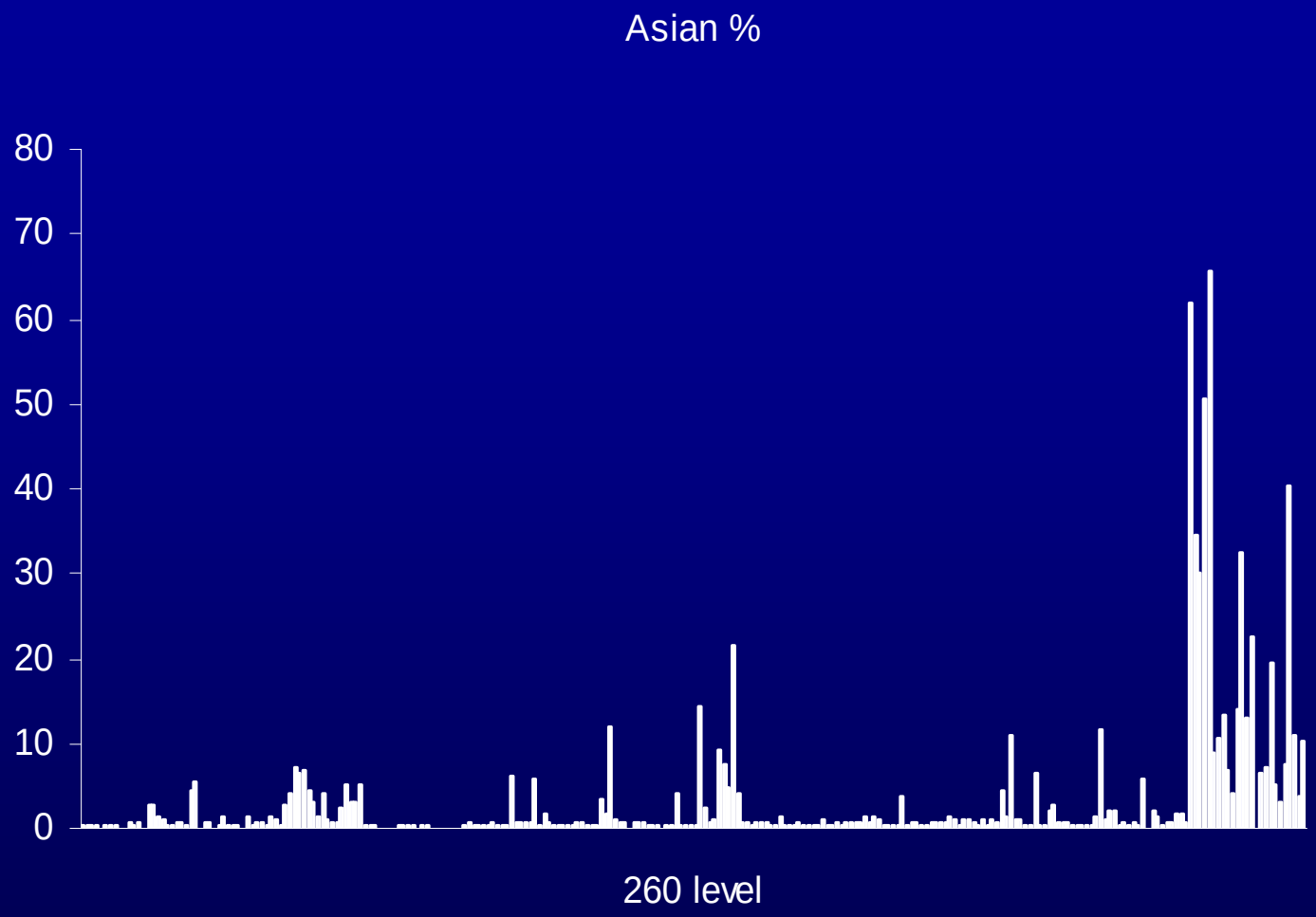
OAC was further sub-divided into 260, 1300 and 10,400 clusters

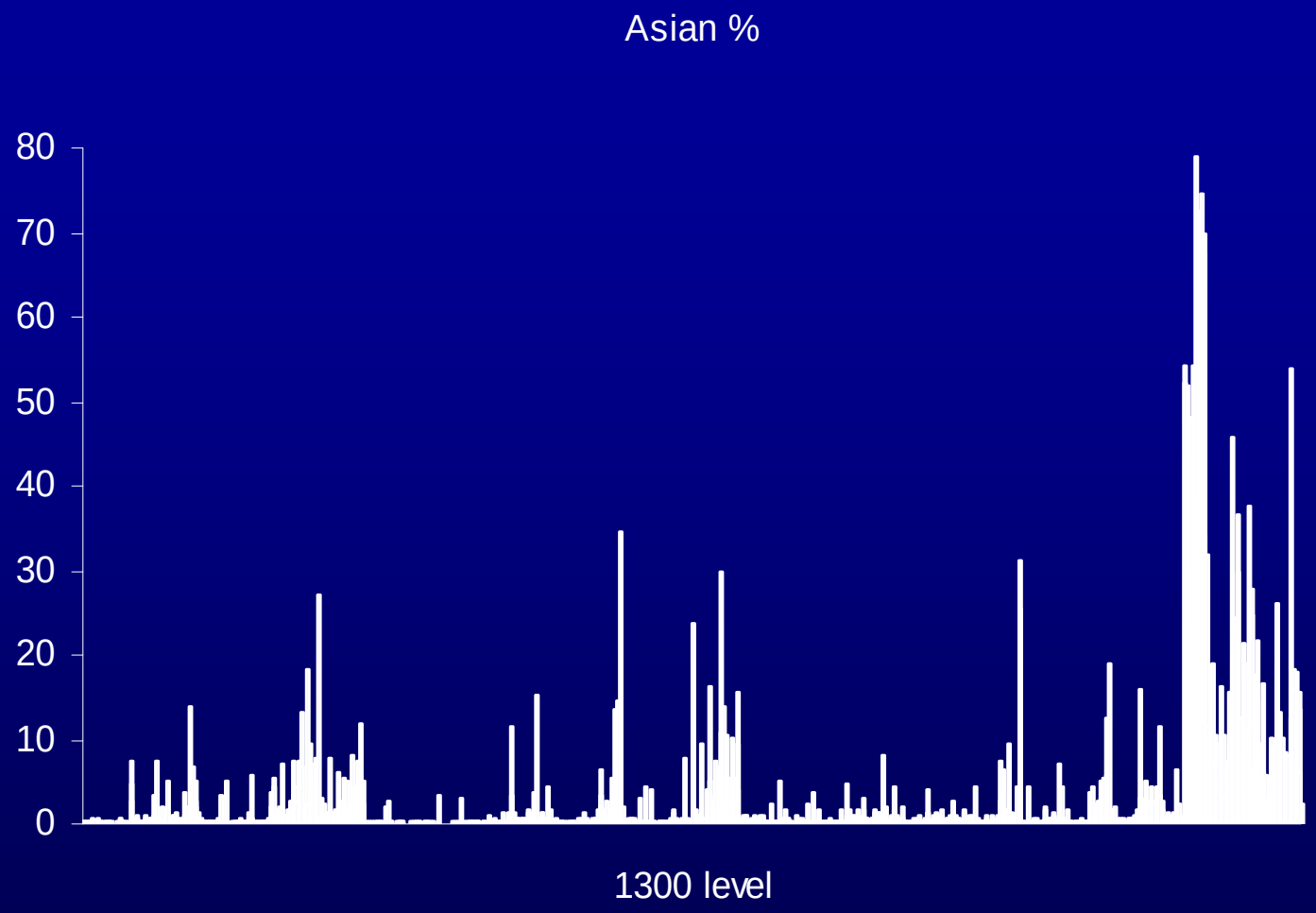
The following charts shows the improved discrimination in terms of the percentage of Asians (3.64% nationally)


Asian %
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
7 level

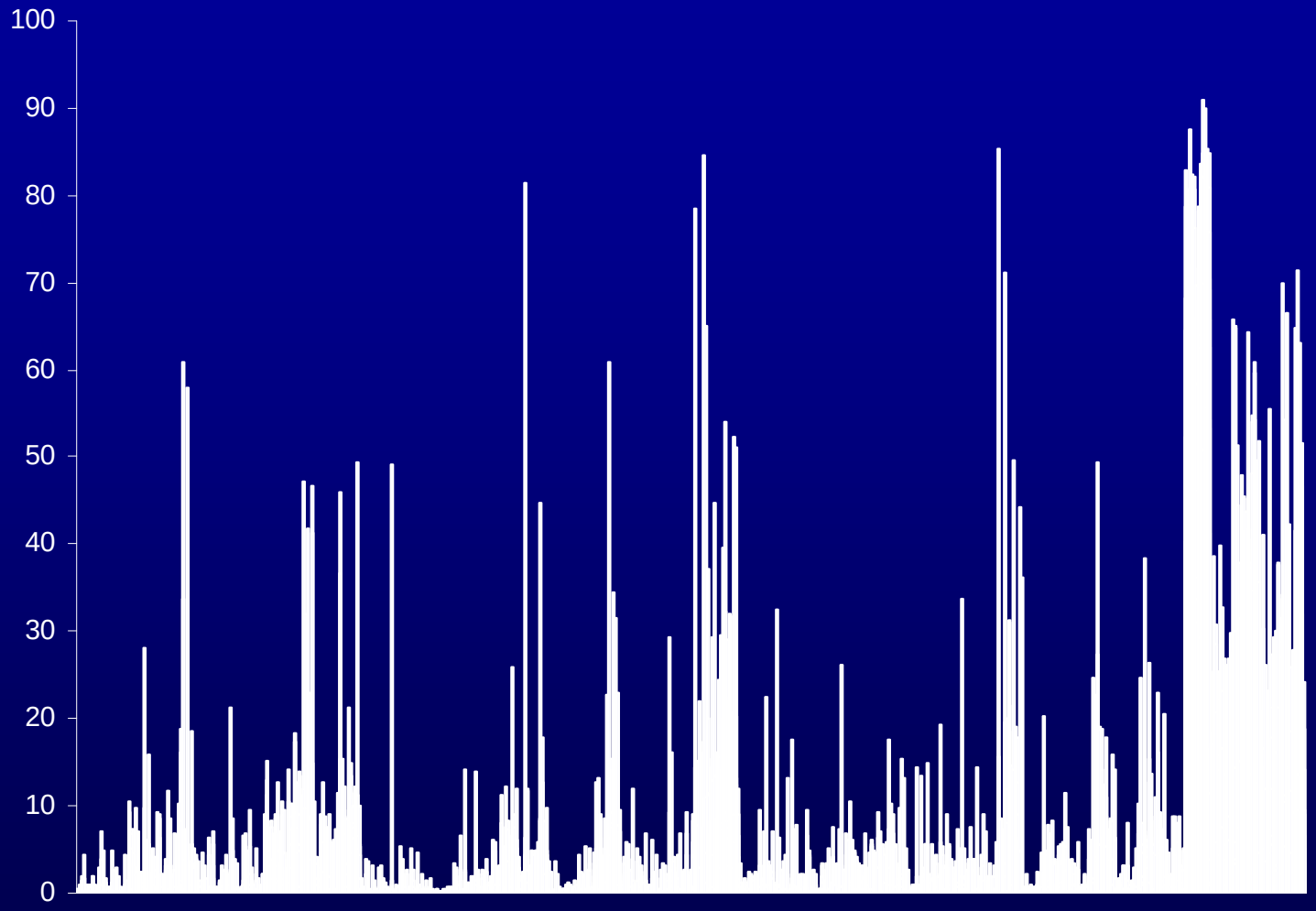

Asian %
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
21 level


Asian %
80
70
60
50
40
30
20
10
0
1
4
7
10
13
16
19
22
25
28
31
34
37
40
43
46
49
52
52 level


Asian %
80
70
60
50
40
30
20
10
0
260 level


Asian %
80
70
60
50
40
30
20
10
0
1300 level

Asian %



OAC 10400

Net flows between OAC supergroups

UNIVERSITY OF LEEDS

Origins	Destinations							
	Blue Collar Communities	City Living	Countryside	Prospering Suburbs	Constrained by Circumstances	Typical Traits	Multicultural	*Net outflow*
Blue Collar Communities	0	8863	6672	27850	-92	12256	-15631	*39918*
City Living	-8863	0	-24837	-52629	937	4257	14044	*-67091*
Countryside	-6672	24837	0	-20648	1297	-5272	-4086	*-10544*
Prospering Suburbs	-27850	52629	20648	0	-3384	-37071	-9316	*-4344*
Constrained by Circumstances	92	-937	-1297	3384	0	-540	-11496	*-10794*
Typical Traits	-12256	-4257	5272	37071	540	0	-28885	*-2515*
Multicultural	15631	-14044	4086	9316	11496	28885	0	*55370*

Source: 2001 Census – Special Migration Statistics Level 3

Learning to do this yourself

The manipulation of geodemographics is actually quite simple

You need:

- A series of appropriate files - in OAC's case down loadable free
- The ability to join files through ACCESS (simple really)
- The ability to use Excel to a moderate level
- Your own data that is postcoded (survey or people data)
- The motivation to explore this fascinating area.
- To produce maps, use the Geodata or CrimeStat –both free
- The most important GIS files are also free

The OAC User Group website has a section on getting started which will be a helpful point to start

Future developments of OAC

OAC User Group provides support and help

Classifications of all surveys conducted by ONS

OAC User Group - SUF group at the Royal Statistical Society

(www. Areaclassification.org.uk)

About OACUG

The OAC User Group (OACUG) was formed in late 2006 with the mission of promoting the use of the National Statistics Output Area Classification (OAC) and area classifications. The User Group is affiliated to the RSS' Statistics User Forum.

The Group is focussed around the well established methods and output of area classification, but OAC is distinctive as it is in the public domain - a basis for 'open geodemographics' - encouraging interchanges between sectors on issues ranging from the underlying statistical methodology to interactive graphics for effective presentation. This open approach is possible as a result of relatively recent changes in the policy on access to government information.

The aims of the Group are to help users apply OAC, to provide opportunities to share experience and build expertise, and to help advances to be made through new methods and applications, operating through open meetings, a self help network, and a dedicated website, enabling it to represent the interests of the OAC user community.

Join the mailing list

Any documents relating to the administration of the OAC User Group can be found [here]

For information about OAC see our getting started section: [link]

Last modified: May 27, 2007

BROWSE AND SEARCH

Monthly Archives

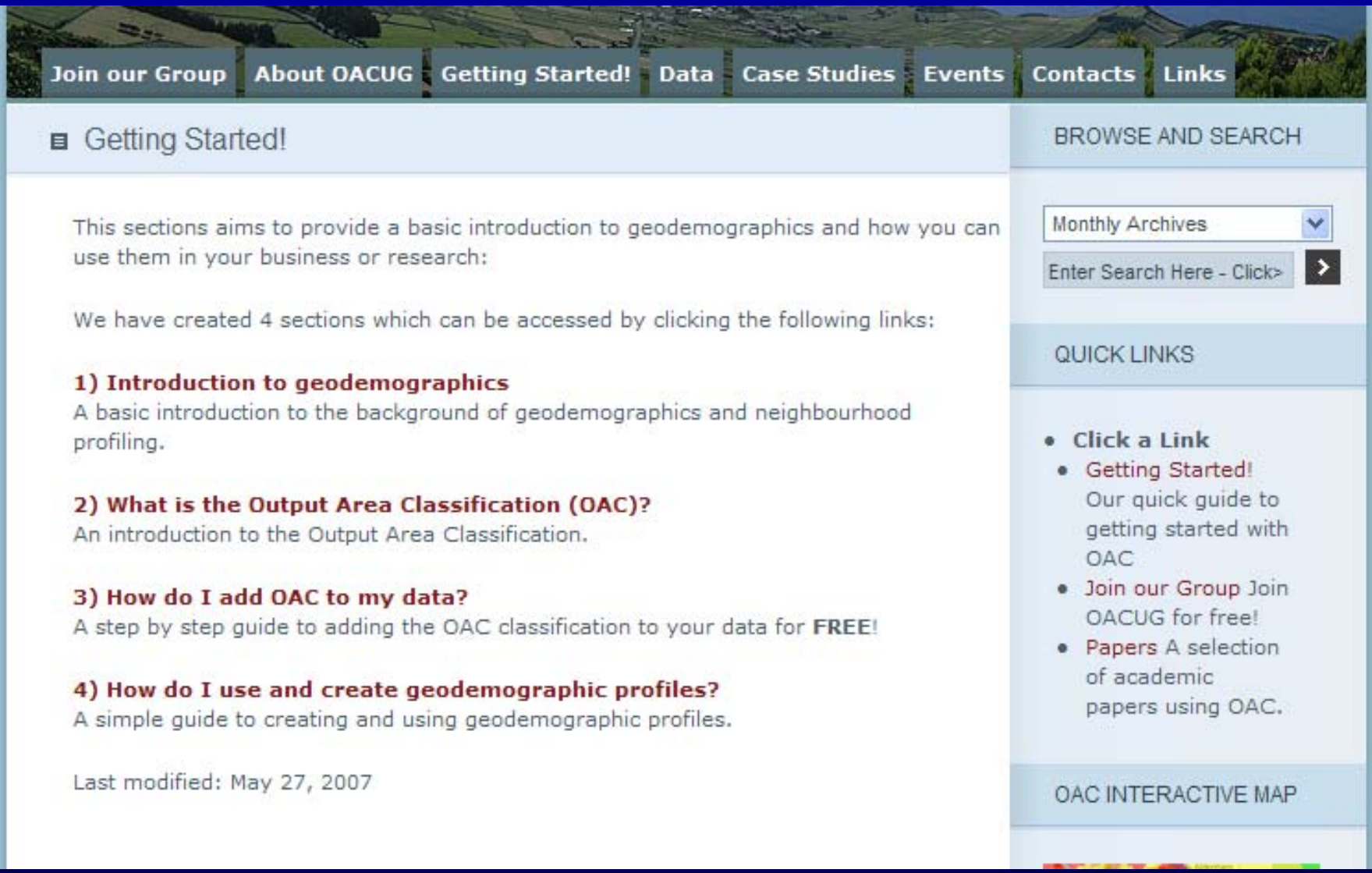
Enter Search Here - Click>

QUICK LINKS

- **Click a Link**
  - Getting Started! Our quick guide to getting started with OAC
  - Join our Group Join OACUG for free!
  - Papers A selection of academic papers using OAC.

OAC INTERACTIVE MAP

Getting started page



Training

OAC USER Group runs a hands-on training course on the use of OAC (equally applicable to other geodemographic systems)

It may consider doing a separate course on elementary visualisation using free GIS and the free boundary mapping files available from ONS.

These two course will allow anybody to do all the stuff I have shown today.

Conclusions

Don’t worry about which system to use

Worry about what you are going to use it for:

- **Areal mapping** for better visualisation
- **Areal profiling** to understand the mix of people better
- **People profiling** to understand who is using a service (or not)
- **Areal modelling** to spread the findings of a survey to places where no interviews were done
- **Data fusion** to take values from one data set to another provided both are profiled

This understanding is about :

Fuelling the creative juices to make better decisions though a more profound understanding of the people

Enabling better marketing

Ultimately about more effective resource allocation

END