UCL

TECHNICAL ANNEX

HIGH CRIME: HIGH DISORDER NEIGHBOURHOODS

SPATIAL ANALYSIS AND GEODEMOGRAPHICS

David Ashby and Richard Webber,
Centre for Advanced Spatial Analysis

It was agreed that UCL would provide a technical briefing note to accompany the main Consultancy Report. This document meets that requirement.

1 Data Sources

1.0.1 All data sets have been returned to the Audit Commission in geocoded form. These data sets are provided in MapInfo format and as Excel spreadsheets / Access databases as appropriate.

1.0.2 All data with sufficient georeferenced information (derived or directly supplied) have been tagged with a Mosaic UK geodemographic code and submitted to the Commission.

1.0.3 Additional files have been supplied to the Commission including a full profile library of BCS variables by neighbourhood group and type, and the geodemographic composition profiles of each study area.

1.0.4 Whilst the absolute population distributions observed in the ten wards were reported in the initial sections of the report, all latter statistical analyses were conducted using population distributions which may contain minor revisions to promote more robust findings. All Mosaic UK types containing less than 100 persons in any one ward were redistributed to the most similar type. This redistribution of small numbers of persons to similar types is detailed below. A strict hierarchical process of conditions determined any redistribution of populations; if less than 100 persons were identified in any one type, the subject population would be reattributed to the following type:

1. Linked to the most-similar type defined by the nearest-neighbour of the minimum spanning tree (Figure 1)
2. Linked to the average for all types within the same Mosaic Group
3. Omitted from analyses

1.0.5 These population redistributions are logged in the xls file: Population_Modified.xls and are summarised in Table 1.

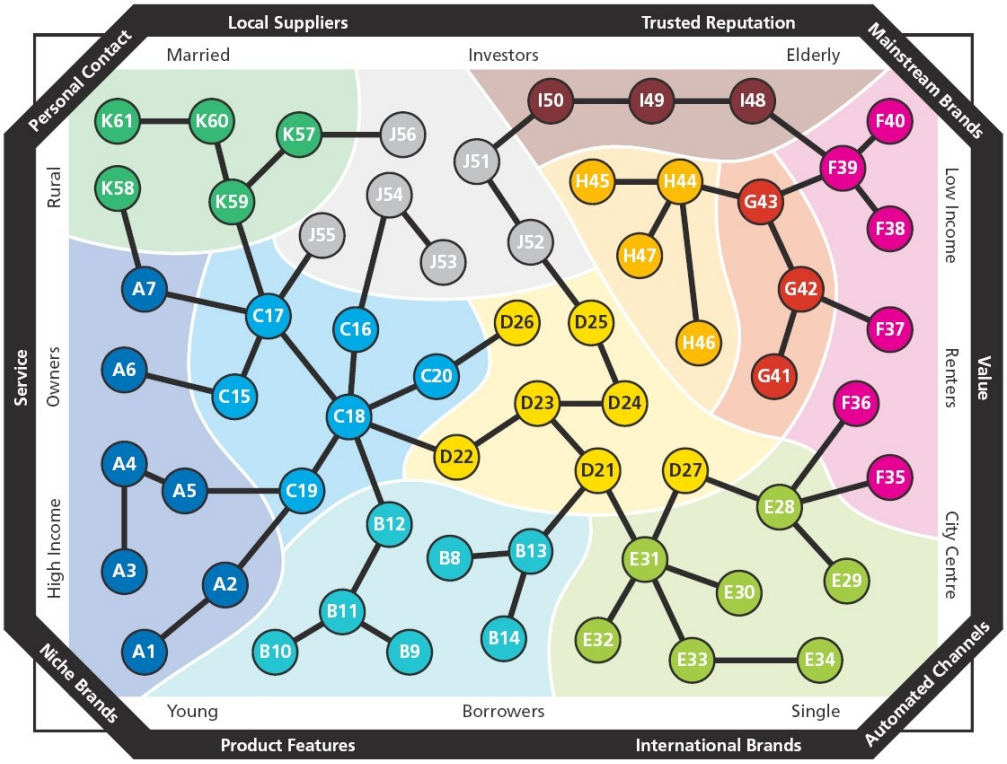


Figure 1: Mosaic UK minimum spanning tree.
(Source: Experian, 2003)

		Population redistributed	From	To	Condition
BRADFORD	**Eccleshill**	97	J55	J54	2
		76	F37	G42	1
	Tong	5	A03	A07	2
		63	C15	C17	1
		80	H46	H47	2
		3	J55	J54	2
		78	K58	A07	1
KERRIER	**Illogan South**	36	B08	B13	1
		11	B11	B12	1
		92	C16	J54	1
		84	D21	B13	1
		97	D25	D24	1
		61	F37	-	3
		27	H45	H44	1
		22	H46	H44	1
	Redruth North	13	B08	B13	1
		48	C16	J54	1
		77	F37	G42	1
		53	F39	G43	1
		19	H47	H44	1
		9	K58	K59	2
LIVERPOOL	**Anfield**	11	B08	B12	2
		44	B13	B12	2
		57	H47	H44	1
	Warbreck	14	B08	B11	2
		22	B13	B11	2
		75	C16	C18	1
		81	D25	D24	1
		26	F40	F39	1
RHONNDA	**Pen-y-waun**	73	G41	G42	1
		97	H45	H44	1
		24	J55	-	3
	Talbot Green	36	A04	A05	1
		36	D23	D22	1
THANNET	**Cliftonville West**	11	B08	B12	2
		82	F37	F35	2
		73	I50	I49	1
		61	J54	J55	2
	Newington	84	G43	G42	1
		46	H45	H44	1
		22	I50	-	3
		72	J54	J55	2

Table 1: Reallocation of small populations by neighbourhood type.

1.1 PLASC database

1.1.1 PLASC is an acronym standing for the Pupil Level Annual School Census. This is a database which is collected and maintained by the Department for Education and Skills. The version of the PLASC file used in this analysis relates to information collected in the academic year 2002- 2003.

1.1.2 The PLASC database used in the analysis is derived from a census of English schools only. We have not been able to gather equivalent statistics for pupils or schools in Wales, Scotland or Northern Ireland.

1.1.3 The PLASC database collects information from each school on each child. The files that we have been able to access are those relating to children who have taken keystage 2 and keystage 4 examinations. The reason why we have restricted analysis to these cohorts is because it was our intention to analysis the relationship between performance at these examinations and Mosaic type.

1.1.4 The officials who manage PLASC require schools to include on the records submitted for each pupil their home postcode. This field has made it possible for the officials to append Mosaic codes to the pupil records. We have therefore been able to obtain from DfES files containing the performance at keystage 2 and keystage 4 of each pupil, together with the identity and postcode of the school which they attend and the Mosaic code of their home address.

1.1.5 This has enabled us to accumulate for each school with more than 20 pupils taking GCSE in 2003 the number and percentage of pupils resident in each Mosaic type. We have also been able to calculate for each of the 61 Mosaic types, the average GCSE points score of pupils resident in them.

1.1.6 Using these two files we have been able to measure the performance that each school would have achieved in terms of the average GCSE points of its pupils had each pupil in that school achieve a points score equal to the average of all pupils in England resident in that Mosaic type. This calculation sets a benchmark target for each school which takes into consideration the mix of types of neighbourhood from which it draws its pupils. In some instances the actual average points score of a school will exceed the average points score that was anticipated by this method. In such an instance we would say that the school over-performed in relation to its expected score. Obviously there are an equivalent number of instances where a school's average points score is lower than what one would expect on the

basis of the mix of types of neighbourhood from which it draws its pupils, in which case we would say that the school was underperforming.

1.1.7 When undertaking all these analyses we have sought to exclude from the calculations all children with any form of special education need since schools with high proportions of such children would otherwise tend systematically to underperform. For reasons of statistical reliability we have also removed from the schools file those schools with fewer than 20 pupils taking keystage 4 examinations.

1.1.8 Strangely we have not obtained from the PLASC officials the name of each school – this does not appear to be held on the database. The file we have created therefore contains just the postcode of the school, a unique school reference code, a local education authority identifier, an array with the proportion of pupils by the 61 Mosaic types, the average GCSE points score of the school and the 'expected' GCSE points score based on the mix of pupils by Mosaic type.

1.1.9 This information could easily be updated each year. Likewise equivalent information could be produced for primary schools. A key benefit of the file is that it allows for changes in the social composition of a school when measuring performance.

2 Analysis Methodology

2.0.1 Within the main report the authors made reference to the notion of *profiling* neighbourhoods, administrative regions and datasets. The following two sub-sections detail the simple statistical methods used, and the technical processing tools respectively.

2.1 Statistical profiling

2.1.1 Throughout the main report and much geodemographic analysis, researchers cite 'index scores' and 'propensities' which are used interchangeably. Index scores are derived to assess the relative frequency of an incident (e.g. crime type), attitude (e.g. fear of being attacked), consumer trait (e.g. likelihood to purchase a PDA) or lifestyle characteristic (e.g. internet usage, etc) compared to some base level. In many cases such index scores may merely be the relative frequency of different neighbourhood types in a local area compared to the regional or national average. Similarly, it may provide an indication of the relative propensity of different neighbourhood types to experience burglary dwelling or vehicle crime.

2.1.2 Creating an index score is a simple statistical task, but care must be taken in selecting appropriate numerators, denominators and base population definitions. Essentially, a profile score is a percentage of the average or expected rate. Therefore, a value of 100 indicates the expected rate, a value of 200 twice the propensity, a value of 50 half that which one would expect, and so on. In calculating these proportions it is, however, important to ascertain an appropriate base distribution for comparison. An illustrative example is given below in Table 2.

		TARGET		BASE		*PENETRATION*	
	Group Name	**Victims**	**%**	**Population**	**%**	**%**	***Index***
A	**Symbols of Success**	281	1.8	9079	2.5	3.1	***73***
B	**Happy Families**	1,098	7.1	30,288	8.3	3.6	***85***
C	**Suburban Comfort**	1,189	7.7	40,684	11.1	2.9	***69***
D	**Ties of Community**	3,548	22.9	48,548	13.3	7.3	***172***
E	**Urban Intelligence**	1,578	10.2	20,039	5.5	7.9	***186***
F	**Welfare Borderline**	641	4.1	5,039	1.4	12.7	***300***
G	**Municipal Dependency**	549	3.5	7,701	2.1	7.1	***168***
H	**Blue Collar Enterprise**	1,566	10.1	31,559	8.6	5.0	***117***
I	**Twilight Subsistence**	395	2.5	10,619	2.9	3.7	***88***
J	**Grey Perspectives**	2,542	16.4	88,273	24.1	2.9	***68***
K	**Rural Isolation**	2,132	13.7	73,691	20.2	2.9	***68***
		15,519	*100*	*365,520*	*100*	*4.2*	

Table 2: Example geodemographic profiling – creating index scores.

2.1.3 To calculate those index scores of Table 2 one has to compare the proportion of the 'target population' with the proportion of the 'base population' for each neighbourhood group and type. In this example we observe the distribution of victims of all crime (target)

compared to the distribution of the population in the study area (base). The null hypothesis in this case is that 50% of the victims should be found in neighbourhood types accounting for 50% of the base population. In such an instance the index values calculated would be equal to 100. Index scores are calculated by comparing the segmented target population to the segmented base population. Therefore, by definition and index score is akin to a location quotient except 'neighbourhoods types' (defined by geodemographics and social similarity) are used instead of areal units. The index score, or Neighbourhood Type Quotient (NQ), for a given neighbourhood type i is the ratio of the percentage of the total neighbourhood activity in type i to the percentage of the total base in area i. i.e. the ratio of the target to base proportions. If Ai is equal to the level of activity in neighbourhood type i (target) and Bi is the level of the base, then:

$$NQ_i = \frac{A_i / \sum A_i}{B_i / \sum B_i} \times 100$$

2.1.4 Conventionally in the geodemographic industry, index values are multiplied by 100 to give a standard expected rate equal to 100. Index values therefore range from 0 to infinity (∞). Negative values are unachievable. Therefore, to produce the index value column of Table 2 one divides the percentage of victims (target) by the percentage total population (base) and multiplies by 100 for each neighbourhood type. In this study population has been preferred to households or to adults as the base, in most cases. However different bases could have been used or could be used in future studies – burglary dwelling indexes were calculated using households as the base.

2.1.5 Index scores are used to describe and visualise the differential geodemographic composition of local areas (such as wards) compared to both regional (e.g. CDRP area) and national bases. Index values are also used to illustrate relative propensities of victimisation, criminal offences, attitudes and fears, compared to both national and regional bases. It is asserted that geodemographic profiling in this manner is helpful in isolating the differences that exist within the small operational areas (such as wards) as well as the differences between these areas and the UK, or local region. Whilst aggregate census measures are effective in distinguishing the differences between the areas, evidence is presented in this research to illustrate that a geodemographic profiling approach can be more useful when the requirement is to drill down within the ward to identify differences within it.

2.1.6 Whilst the above description provided suggests a relatively simple statistical process to create index scores, which may accompany orthodox descriptive statistics, and pie charts,

bar charts and maps to visualise such distributions, the accurate coding of operational records obtained from police forces is an essential stage in the geodemographic data processing. The following section details this process highlighting theoretical considerations, computer processing methodologies and standard outputs.

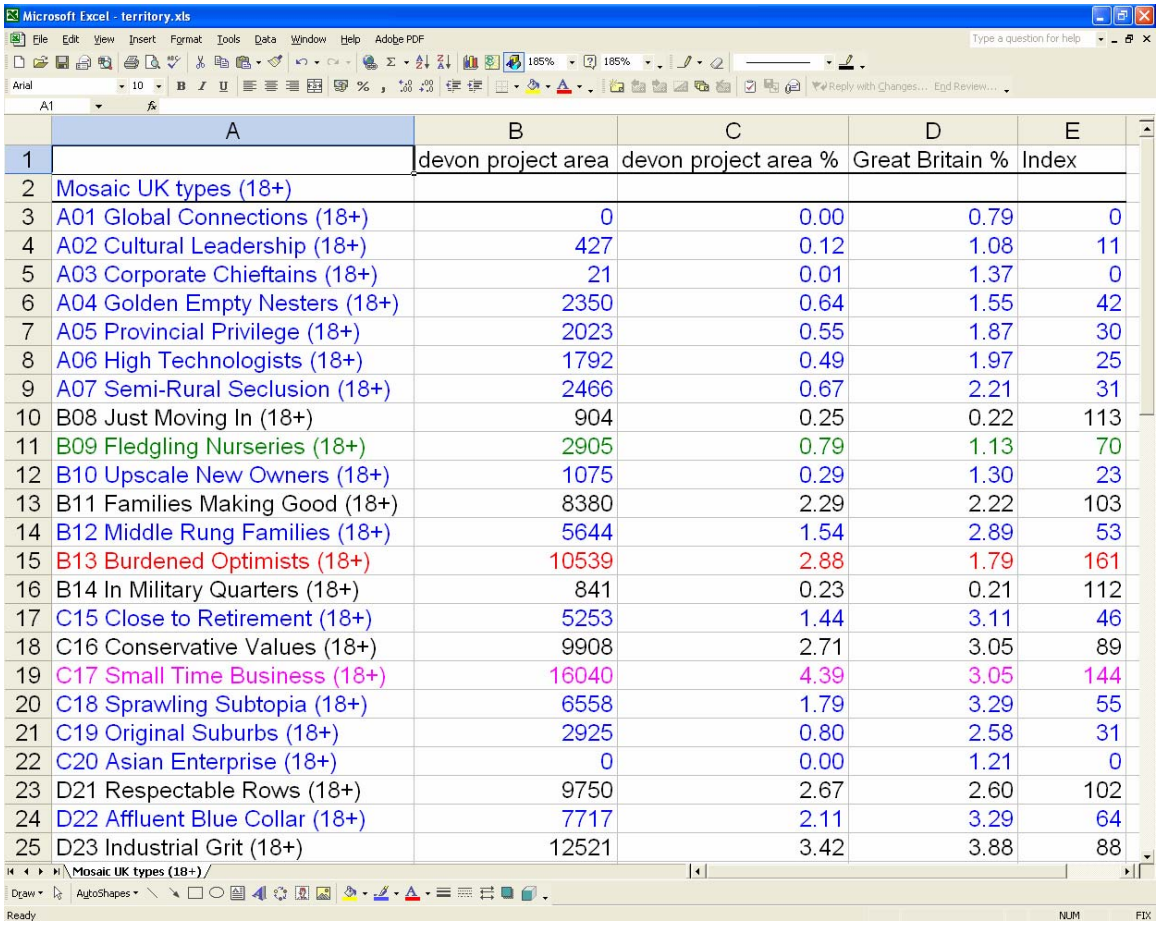
2.2 Processing methodologies

2.2.1 To illustrate the processing methodology which was performed on a wide range of data sets for this research, this section details the general procedure followed throughout. Whilst the methodology outlined below is generalised where possible, specific obstacles, considerations and challenges are often encountered for each unique data set. Therefore, whilst a guiding outline is provided, it has proven challenging in practice to rapidly coach practitioners to conduct such analyses unaccompanied. A steep learning curve and some previous expertise and understanding of the method appear pre-requisites for the course.

Creating Geodemographic Crime Profiles

Input data

2.2.2 Two core data sets are required for the geodemographic analysis of operational data sets. Primarily, a 'territory file' is required. This file is generated using Experian's MicroMarketer software and describes the geodemographic composition of the study region. The output file (.csv, .xls, .dbf, etc) may usefully be observed in spreadsheet format whereby the composition of the study region (e.g. the Liverpool CDRP, or indeed Anfield ward with this) is given by neighbourhood group and type for the required variable. In most cases, it is useful to provide the population distribution by neighbourhood type for the study area, but other options exist, such as number of households, number of adults (18+) and number of adults (15+). An example of the territory output file is given in the screen capture below (see Figure 2).



	A	B	C	D	E
1		devon project area	devon project area %	Great Britain %	Index
2	Mosaic UK types (18+)				
3	A01 Global Connections (18+)	0	0.00	0.79	0
4	A02 Cultural Leadership (18+)	427	0.12	1.08	11
5	A03 Corporate Chieftains (18+)	21	0.01	1.37	0
6	A04 Golden Empty Nesters (18+)	2350	0.64	1.55	42
7	A05 Provincial Privilege (18+)	2023	0.55	1.87	30
8	A06 High Technologists (18+)	1792	0.49	1.97	25
9	A07 Semi-Rural Seclusion (18+)	2466	0.67	2.21	31
10	B08 Just Moving In (18+)	904	0.25	0.22	113
11	B09 Fledgling Nurseries (18+)	2905	0.79	1.13	70
12	B10 Upscale New Owners (18+)	1075	0.29	1.30	23
13	B11 Families Making Good (18+)	8380	2.29	2.22	103
14	B12 Middle Rung Families (18+)	5644	1.54	2.89	53
15	B13 Burdened Optimists (18+)	10539	2.88	1.79	161
16	B14 In Military Quarters (18+)	841	0.23	0.21	112
17	C15 Close to Retirement (18+)	5253	1.44	3.11	46
18	C16 Conservative Values (18+)	9908	2.71	3.05	89
19	C17 Small Time Business (18+)	16040	4.39	3.05	144
20	C18 Sprawling Subtopia (18+)	6558	1.79	3.29	55
21	C19 Original Suburbs (18+)	2925	0.80	2.58	31
22	C20 Asian Enterprise (18+)	0	0.00	1.21	0
23	D21 Respectable Rows (18+)	9750	2.67	2.60	102
24	D22 Affluent Blue Collar (18+)	7717	2.11	3.29	64
25	D23 Industrial Grit (18+)	12521	3.42	3.88	88

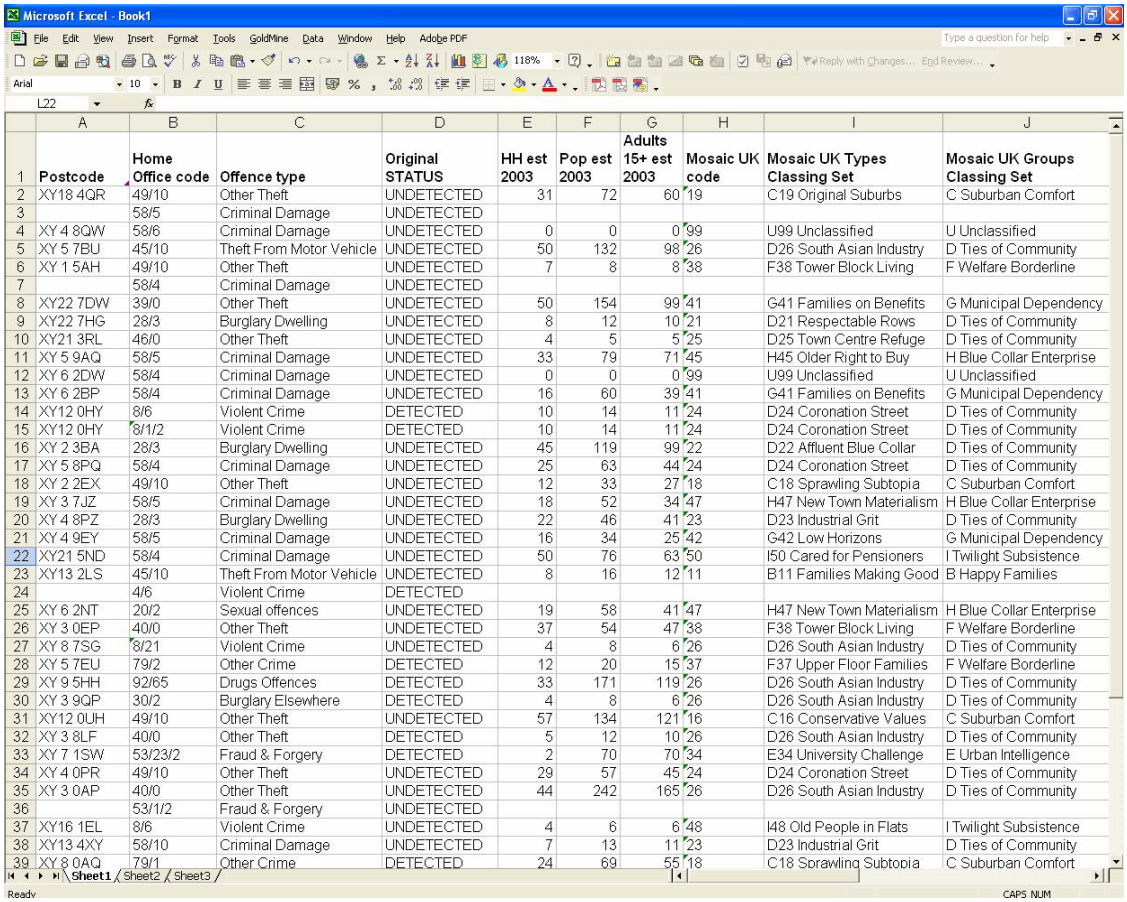
Figure 2: Screen capture of an example 'territory file' MicroMarketer output – viewed in a spreadsheet.

2.2.3 The second data set required is the input data from the operational database. This must contain those data to be profiled and the appropriate georeference for each record – e.g. location of victim, offender and/or crime. In some cases it has been necessary to link the operational data and georeference information as an additional stage in the process. However, essentially the analyst requires both those data to be processed and a spatially reference, to which a geodemographic code can be attributed. Full postcodes are required for the classification of data by Mosaic, so if only a grid reference location is known it may be required to conduct a GIS point-in-polygon query using the operational records and a polygon postcode layer. Postcodes can thus be derived for all records with a spatial reference, subject to the accuracy of the original geocoding.

Geodemographic coding

2.2.4 The geocoded data set should be parsed through Experian's MicroMarketer software to attribute a geodemographic code to each postcode of the file. Ideally, this should be conducted for both victim and offender postcodes where available. This is inevitably dependent upon the quality of the data provided and their associated georeferencing. MicroMarketer identifies postcodes in any format and appropriately appends a number of

columns to the original dataset – these include the Mosaic classifier (A1- K61), a postcode population estimate, and any other required georeferenced information such as a ward qualifier, LAD code or NHS region. Figure 3 below provides an example of key fields output from an annonymised data set, when victim postcodes are coded by Mosaic.



	Postcode	Home Office code	Offence type	Original STATUS	HH est 2003	Pop est 2003	Adults 15+ est 2003	Mosaic UK code	Mosaic UK Types Classing Set	Mosaic UK Groups Classing Set
2	XY18 4QR	49/10	Other Theft	UNDETECTED	31	72	60	19	C19 Original Suburbs	C Suburban Comfort
3		58/5	Criminal Damage	UNDETECTED						
4	XY 4 8QW	58/6	Criminal Damage	UNDETECTED	0	0	0	99	U99 Unclassified	U Unclassified
5	XY 5 7BU	45/10	Theft From Motor Vehicle	UNDETECTED	50	132	98	26	D26 South Asian Industry	D Ties of Community
6	XY 1 5AH	49/10	Other Theft	UNDETECTED	7	8	8	38	F38 Tower Block Living	F Welfare Borderline
7		58/4	Criminal Damage	UNDETECTED						
8	XY22 7DW	39/0	Other Theft	UNDETECTED	50	154	99	41	G41 Families on Benefits	G Municipal Dependency
9	XY22 7HG	28/3	Burglary Dwelling	UNDETECTED	8	12	10	21	D21 Respectable Rows	D Ties of Community
10	XY21 3RL	46/0	Other Theft	UNDETECTED	4	5	5	25	D25 Town Centre Refuge	D Ties of Community
11	XY 5 9AQ	58/5	Criminal Damage	UNDETECTED	33	79	71	45	H45 Older Right to Buy	H Blue Collar Enterprise
12	XY 6 2DW	58/4	Criminal Damage	UNDETECTED	0	0	0	99	U99 Unclassified	U Unclassified
13	XY 6 2BP	58/4	Criminal Damage	UNDETECTED	16	60	39	41	G41 Families on Benefits	G Municipal Dependency
14	XY12 0HY	8/6	Violent Crime	DETECTED	10	14	11	24	D24 Coronation Street	D Ties of Community
15	XY12 0HY	8/1/2	Violent Crime	DETECTED	10	14	11	24	D24 Coronation Street	D Ties of Community
16	XY 2 3BA	28/3	Burglary Dwelling	UNDETECTED	45	119	99	22	D22 Affluent Blue Collar	D Ties of Community
17	XY 5 8PQ	58/4	Criminal Damage	UNDETECTED	25	63	44	24	D24 Coronation Street	D Ties of Community
18	XY 2 2EX	49/10	Other Theft	UNDETECTED	12	33	27	18	C18 Sprawling Subtopia	C Suburban Comfort
19	XY 3 7JZ	58/5	Criminal Damage	UNDETECTED	18	52	34	47	H47 New Town Materialism	H Blue Collar Enterprise
20	XY 4 8PZ	28/3	Burglary Dwelling	UNDETECTED	22	46	41	23	D23 Industrial Grit	D Ties of Community
21	XY 4 9EY	58/5	Criminal Damage	UNDETECTED	16	34	25	42	G42 Low Horizons	G Municipal Dependency
22	XY21 5ND	58/4	Criminal Damage	UNDETECTED	50	76	63	50	I50 Cared for Pensioners	I Twilight Subsistence
23	XY13 2LS	45/10	Theft From Motor Vehicle	UNDETECTED	8	16	12	11	B11 Families Making Good	B Happy Families
24		4/6	Violent Crime	DETECTED						
25	XY 6 2NT	20/2	Sexual offences	UNDETECTED	19	58	41	47	H47 New Town Materialism	H Blue Collar Enterprise
26	XY 3 0EP	40/0	Other Theft	UNDETECTED	37	54	47	38	F38 Tower Block Living	F Welfare Borderline
27	XY 8 7SG	8/21	Violent Crime	UNDETECTED	4	8	6	26	D26 South Asian Industry	D Ties of Community
28	XY 5 7EU	79/2	Other Crime	DETECTED	12	20	15	37	F37 Upper Floor Families	F Welfare Borderline
29	XY 9 5HH	92/65	Drugs Offences	DETECTED	33	171	119	26	D26 South Asian Industry	D Ties of Community
30	XY 3 9QP	30/2	Burglary Elsewhere	DETECTED	4	8	6	26	D26 South Asian Industry	D Ties of Community
31	XY12 0UH	49/10	Other Theft	UNDETECTED	57	134	121	16	C16 Conservative Values	C Suburban Comfort
32	XY 3 8LF	40/0	Other Theft	DETECTED	5	12	10	26	D26 South Asian Industry	D Ties of Community
33	XY 7 1SW	53/23/2	Fraud & Forgery	DETECTED	2	70	70	34	E34 University Challenge	E Urban Intelligence
34	XY 4 0PR	49/10	Other Theft	UNDETECTED	29	57	45	24	D24 Coronation Street	D Ties of Community
35	XY 3 0AP	40/0	Other Theft	UNDETECTED	44	242	165	26	D26 South Asian Industry	D Ties of Community
36		53/1/2	Fraud & Forgery	UNDETECTED						
37	XY16 1EL	8/6	Violent Crime	UNDETECTED	4	6	6	48	I48 Old People in Flats	I Twilight Subsistence
38	XY13 4XY	58/10	Criminal Damage	UNDETECTED	7	13	11	23	D23 Industrial Grit	D Ties of Community
39	XY 8 0AQ	79/1	Other Crime	DETECTED	24	69	55	18	C18 Sprawling Subtopia	C Suburban Comfort

Figure 3: Screen capture of an example operational data set output from MicroMarketer.

Pre-processing

2.2.5 Once a full operational data set with postcodes and geodemographic qualifiers has been established one is almost ready to begin to calculating those relevant index values. However, exploratory analyses with a series of data sets of this kind helped the author identify a number of key factors which need to be considered and often excluded from the data set. These are detailed further in Section 3: Recommended Standards below.

2.2.6 When analysing incidents by Mosaic it is easy to skew or distort the results of an analysis by including victims and/or offenders who live outside the area. To assess incident patterns it is important to be able to compare the distribution of incidents by Mosaic with the distribution of a relevant *base* population. This base may be the population or total number of households within the study area by neighbourhood type. However, the Mosaic profile of out of area victims and out of area offenders is likely to be very different to the profile of within area victims and offenders. Therefore, such data should not be included when calculating

local index values. However, it is recommended that whilst these data can not be included in initial geodemographic profiling, one should not simply delete all out of area records. A standard has been provided by the authors for previous geodemographic analysis whereby if either the victim or the offender (or both) are determined to reside outside the study area, those associated Mosaic codes should be removed. This should only be executed for whichever party lives outside the area, and not necessarily both parties. It may also be worthwhile retaining such records in a separate file so that one can analyse the type of victim or offender entering the area. However, profiles can not be created from the local denominator bases (population, adults, households, victims).

2.2.7 In removing any geodemographic codes for out-of-area offenders/victims all data are retained, but only those victims and offenders resident in the study area will contribute to the study area index scores. This prevents skew and bias in any profiles. For example, if a commuter living in an A01 neighbourhood in London commutes to Exeter and becomes a victim of vehicle crime in a city centre car-park, it is unfair to attribute such victimisation to those very few A01 neighbourhoods in the local area. However, it may be that a career-criminal habitually perpetrates such vehicle crime on commuter vehicles, and this offender does live in the local Exeter region. Using the offender's resident postcode and geodemographic code in the creation of local profile scores is required in this case.

2.2.8 Additional considerations are also required in the pre-processing stage. It is important to also exclude so called 'victim-less' crime (or perhaps more accurately, crime against victims who are not private residents) from geodemographic analysis in much the same way that the BCS does. For example, it is important to exclude the Mosaic codes of the victims of 'theft from shop'. Shops are disproportionately concentrated in certain Mosaic types and so the inclusion of shop theft distorts the profile since shops have no enumerated population and therefore do not feature in the base. Again, it is sensible to retain these records and their postcodes, and to Mosaic code the offenders (if known) but not the 'victims'. Similarly, some offences, such as possession of drugs, are 'victimless' offences. Thus any analysis of victims will not pick up offences of these types.

2.2.9 Essentially, all records to be geodemographically analysed should represent a victim or perpetrator of crime resident in the study area which is to be used as the 'base' population.

Creating profiles

2.2.10 Once all of those stages of pre-processing above have been conducted the analyst can proceed with the calculation of index values. A number of software packages and macros are available for automated processing, but it is important to fully understand the extent of

each data set and to appreciate why such pre-processing and exclusions are necessary. Furthermore, this entire process can all be completed in a generic spreadsheet or database package such as Microsoft Excel or Access; tools which are both widely available and familiar to crime analysts.

2.2.11 To create an index score one simply takes the entire dataset with postcode and Mosaic codes appended and creates *N* pivot tables, segmented by both neighbourhood group and crime category, etc. Pivot tables may be required to ascertain;

1. Total number of victims by neighbourhood type
2. Total number of offenders by neighbourhood type
3. Crime type by neighbourhood type
4. Seasonality of crimes by neighbourhood type
5. Detection rate by neighbourhood type

2.2.12 Using pivot tables one is able to ascertain the counts for each of the above phenomena by class category and by neighbourhood type (e.g. all victims by crime type {e.g. burglary dwelling, criminal damage, violence against person, etc} by neighbourhood type {e.g. A01, A02, …K61}). It is worthwhile conducting this process for both the 61-type and 11-group level, especially in those cases where insufficient data volumes are available within each cell to enable statistically robust calculation of index values at the 61-type level.

2.2.13 Once those counts have been calculated for the segmented categories of the database one is ready to compare these 'target' data to some 'base' dataset. The base distribution of population or households was calculated earlier in the territory data set. Using standard spreadsheet functions it is therefore possible to calculate the proportion of a target crime variable (e.g. burglary dwelling) in all neighbourhood types and compare this to the relevant base population/household distribution, to create an index score. In some instances it is important to take the distribution of all crimes/victims as the base rather than the underlying population – for example, in evaluating differential detection rates one should use all crimes as the base instead, as it is given that a crime occurred. Using population in this case would actually reflect the variance in crime rates across neighbourhood types, rather than any inequalities in the police performance / detection rates across all neighbourhoods whenever a crime is committed.

2.2.14 This section has detailed the general process conducted in geodemographically coding and analysing data for crime and policing purposes. The process does require some understanding of the limitations and pre-processing stages required and indeed some efficient method of attributing a geodemographic code to each record. Using Experian software this process of attributing a Mosaic identifier is relatively simple, but does require not insignificant investment in the purchase of the software. Similarly, whilst the creation of

index values can be a relatively simple process utilising pivot table functions, considerable effort is required in the first instance in creating appropriate templates and formulae for batch processing of large volumes of data. The extraction of out of area offenders, and the derivation of postcodes from grid reference co-ordinates is also a relatively straightforward task but does require a clear understanding of postal and OS geographies, a working knowledge of GIS overlay functionality and access to all relevant data and software licences. Whilst such data, software, expertise and resource are not insignificant, it is recognised that economies of scale are required and that as such, these processing tasks should ideally be conducted as the Police Force level at Force Headquarters rather than local divisions thereof.

3 Recommended Standards

Frequency of extract

3.0.1 It is suggested that crime incident data should be extracted either once a year or once every six months. The speed of change in patterns by Mosaic are such that more frequent analysis is unlikely to show real change. In other words the ranking of neighbourhoods types is unlikely to change as frequently as the ranking of local areas.

Date of extract

3.0.2 It is suggested that extracts are undertaken for incidents within a financial rather than a calendar year and that half yearly extracts take incident data for the first six months or second six months of the financial year.

Length of period for extract

3.0.3 The longer the period over which incident data is extracted the more robust will be the size of the sample. For this reason it is worth considering taking a full year's incident data even if you refresh the date each six months. Taking a full year's data will also avoid the possible bias in the pattern of data between winter and summer.

Definition of qualifying incidents

3.0.4 Most incident data files will give the time of an incident as being between time 'x' and time 'y'. Obviously there are occasions where, within an extract period, there will be incidents, which have been notified before the start of the extract period window and others, which will remain open even after the window. It is suggested that before analysis all incidents with an opening time before the start of the time window are deleted. These incidents are likely to be skewed towards certain incident types.

Offender, victim and location postcodes

3.0.5 In general it is worth extracting as much information within reason as possible. In particular it is useful where possible to extract the postcode of the offender as well as the postcode of the victim. If possible try to extract an indicator of where the offence takes place, whether using a grid reference or a postcode. This information can be useful for measuring the distance of the location from the victim and/or the offender's home address.

Obviously the postcode of the location, if known, should not be analysed by Mosaic if they are not the home location of the victim or the offender (see above).

Fields for analysis

3.0.6 Fields which have proved valuable for analysis and which can be extracted from a crime information system include:

- How the police were made aware of the incident
- The time periods between which the incident occurred
- The offence category
- The clear up code

From this information it is possible to generate additional fields such as:

- Time of day within certain higher-level bands (e.g. early evening)
- Seasons of the year
- Postcode of victim same as postcode of offender
- Postcode of victim same as postcode of incident location
- Distance between location of incident and home location of victim or offender

Incident types

3.0.7 There are wide disparities in the frequencies of different offence codes. Many offence categories have very low frequencies. It may be worth ranking offence codes by frequency and disregarding all offence categories with fewer than 100 records during the analysis period. An alternative when presenting results is to sort your excel spreadsheet by column, so that offence codes are presented in frequency rather than in alphabetic order. Offence codes with very infrequent occurrences can be concealed from public view using the 'hide' function in Excel.

Coding by Mosaic

3.0.8 It is worth coding and summarising by both the 61 types and the 11 groups. The groups are useful when analysing offence codes with low frequencies.

Mosaic code 99

3.0.9 The Mosaic code '99' indicates non-residential postcodes, for instance schools, shops etc. When you are analysing the percentage of incidents by Mosaic type you would be well advised to use victims in Mosaic codes 1-61 as your denominator rather than in all Mosaic codes including 99. This is because when you compare your distribution with that of the

base area there will be no occurrences of people in Mosaic type 99 in your study area. However you might want to flag records where the victim is in Mosaic code 99. These are likely to be incidents against commercial organisations or the local authority rather than crimes against other individuals.

Valid postcodes

3.0.10 Postcodes can be quite difficult to match. Experian software can identify all current and historic postcodes and append Mosaic codes to both types. This is a useful facility if for example there has been postcode reorganisation within parts of your study area. Likewise Experian software will reformat postcodes to a consistent standard and translate zeroes to 'o' where appropriate. Without using this software you may in danger of not being able to code 'old' postcodes or indeed very recent ones. The Mosaic postcode list is updated annually so it will fail to code very recently added postcodes. There may be some further loss of data when matching postcodes in any GIS / data source with postcode attributes.

Selection of base distribution

3.0.11 In order to analyse incident rates you will need to obtain a distribution for your study area of population, households and adults 15+ by Mosaic type. Experian can provide this information. In due course the authors aim to provide documented guidance on which types of incidents should be analysed using which of these bases. Obviously the Mosaic percentage distribution within the study area will differ slightly between these three.

Missing Mosaic categories

3.0.12 Whilst most Mosaic codes are represented in most police force areas there are occasions when a type is missing from an area. In these cases it is worth keeping the count in the analysis, even it is zero. This makes it easier then to compare results between different police forces.

3.0.13 However, an alternative for mapping (probably most appropriate option for importing into a GIS) is to replace these cells with the value 100 – i.e. the average for the region. In constructing hot cold maps it is easy to use the value 100 as the point of inflection between red and blue (probably coloured white) thus indicating average or 'no data'. This will also work for charts and graphical representations where bars go above and below the x-axis at the value 100. Such cells will not be visible in the charts but will not cause the software package any trouble over missing values. This method should not be used for statistical analyses, rather only for visualisation purposes.

Dr. David Ashby
Tel: 020 7679 9802
Fax: 020 7679 9801
Email: d.ashby@ucl.ac.uk
Web: www.spatial-literacy.org

A report presented to the Audit Commission July 2005