

The Added Value of Geospatial Data in Crime Analysis

Introduction

Crime is not distributed randomly. There are distinctive patterns in where crimes happen and when they take place. Look at any hotspot map over a 24-hour period and you will see that the hotspots are not fixed but rather move around a city or a neighbourhood according to the time of day; some being there during the day but vanishing by the evening, others springing up at night but disappearing by breakfast time (e.g. violence and anti social behaviour in the town centre with many pubs).

This is because crime patterns are influenced by a number of factors that we can see all around us. These would include, for example, the layout of the streets (e.g. whether narrow or wide, through roads or cul-de-sacs) the terrain (e.g. the presence of lakes, rivers, bridges, railway lines), the presence of particular buildings (e.g. shopping parades, pubs, schools) or amenities (bus stops, car parks, taxi ranks) and differences in the type of housing and its design (e.g. terraced housing with back alleys, semis with gardens, larger detached houses).

The people who are present in different places will also affect crime patterns. Not all residential areas, for example, will be the same. Some will have a relatively large number of students and single people coming and going during the day and when they're not there leaving behind iPods laptops mobile phones and perhaps other highly desirable electrical goods just waiting to be stolen! People who commit crimes knows this and if all of the houses in a street have the same design an offender only needs to break into one house to know the layout of all the others when thinking about doing another burglary. Other neighbourhoods may have young families or pensioners who seldom leave the property unattended. How people are doing is also important. If they are well off there may be able to afford burglar alarms better security. Poorer communities may be less well protected but also may have a larger number of offenders living within them.

And it's not just residential areas that shape crime patterns. People can be in places because they are working in them or because they are visiting them and these include people up to no good i.e. those pursuing a criminal career.

In short, understanding crime patterns cannot be achieved just by looking at recorded crime data held by the police. You need information about all of the other influences on crime patterns mentioned above. The good news is that much of the data about the geographical context of the crime or what we call the 'crime environment' is already available in the form of geospatial datasets. When we use these as well as crime data we can begin to see what's happening on the ground and the links between crime patterns and what shapes them.

The [GeoCrimeData](#) project has generated datasets about places produced from what we call the *primary geospatial components of 'place'*. These are referred to as polygons, lines and points - which translates into:

- Information about places with a geographical boundary (i.e. 'polygons'). For example, crime rates for areas and information about what type of area it is;
- Information about streets (i.e. 'lines'), such as where these streets are and how accessible they are to other streets;

- Information about individual houses (i.e. points), what they are like (detached, semis, terraced) and whether or not they have been burgled.

The following examples, all relating to the City of Leeds, show you what we can learn about places and burglaries happening within them when we bring together these different geospatial datasets. We can begin by just looking at what places are like without considering crime (i.e. 'context without crime').

The Output Area Classification

A geo-demographic classification is an important geospatial data set that tells us what different neighbourhoods are like. Table 1 describes the different neighbourhood types or Super Groups that make up the [Output Area Classification](#) (OAC) which is freely available for everyone to use. You'll see from this is that there are seven different super groups and they differ from each other both in their housing, levels of affluence and population characteristics. OAC is one of many that have been produced but you can only get access to these by buying them.

Table 1 Pen pictures of OAC Super Groups

1 Blue Collar Communities	Housing in these areas is more likely to be terraced rather than flats and residents mainly rent from the public sector. There is a high proportion of 5-14 year-olds. Residents tend to have fewer higher educational qualifications than the national average. A high proportion work in manufacturing, retail or construction.	
2 City Living	Residents in these urban areas are more likely to live alone. They are more likely to hold higher educational qualifications and are often first generation immigrants to the UK. Housing is often made up of flats and detached homes are rare and residents typically rent their homes from the private sector	
3 Countryside	Residents in these rural areas are likely to work from home and to be employed in agriculture or fishing. They often live in detached houses; in households with more than one car. Areas are less densely populated than other parts of the country.	
4 Prospering Suburbs	Residents in these prosperous areas often live in detached houses and less frequently in flats or terraced housing. Fewer residents rent their homes and homes are more likely to have central heating. Households often have access to more than one car.	
5 Constrained by Circumstances	Residents in these less well off areas typically live in flats and rent from the public sector. They are less likely to have higher qualifications. They rarely live in detached houses or in households with more than one car.	
6 Typical Traits	These are areas of terraced housing, where residents are unlikely to rent from the public sector. There are a range of ethnic backgrounds and types of households. Residents work in a range of industries.	

7 Multicultural	Residents in these areas are often non-white, mainly from Asian or Black British backgrounds. Many are first generation immigrants. Housing is mostly rented from the public or private sectors and is often split into flats. The main means of travelling for residents is by public transport.	
------------------------	---	---

Source: <http://www.localfutures.com/Home/OAC/default.aspx#2153>

Analysis of House Types

Table 2 shows the distribution of different types of houses within each OAC Super Group. The house type information has been produced by software created specifically for our project that examines the shapes of buildings in Leeds as they appear on Ordnance Survey MasterMap and calculates whether they are detached, semi-detached, a terraced house or a new category of house type, a corner property. (For more information about the data and how you can download it, see the [Summary of Data Outputs](#) document).

Table 2 Housing Type by OAC Super Group

OAC Super Group	Percentage of OAC group				
	Detached	Semi	Terraced	Corner	Total
1. Blue-Collar Communities	4.9	59.5	24.5	11.1	100
2. City Living	14.5	33.6	42.5	9.4	100
3. Countryside	47.3	37.6	9.1	6.1	100
4. Prospering Suburbs	40.4	54.4	3.2	2.0	100
5. Constrained by Circumstances	5.9	61.7	21.6	10.9	100
6. Typical Traits	14.5	46.0	29.8	9.7	100
7. Multicultural	4.1	31.8	54.7	9.4	100
Leeds	45,922	131,334	63,335	21,330	261,921

There were large differences in Leeds between the Super Groups in the distribution of house types within them. For example, nearly two thirds of the housing stock in Constrained by Circumstances areas were semis compared with under one third in City Living neighbourhoods.

Just 4.1% of houses in Multicultural areas were detached compared with over 40% in the Prospering Suburbs and in the more rural 'Countryside' communities in Leeds. Across the city as a whole the most predominant form of housing was semidetached properties followed by terraced houses.

Table 3 shows burglary rates for each type of property in each Super Group. Usually burglary rates are calculated by dividing the number of burglaries in general (i.e. across all types of houses) by the total number of households or houses in an area and then multiply by 1,000. The [GeoCrimeData](#) Project has allowed us to improve on this by calculating the burglary rates for each property type. For example, because we know how many burglaries have happened in terraced houses, detached houses and so on and we know the total number of terraced houses, detached houses etc in each Output Area we can calculate the *terraced housing burglary rate* (i.e. the number of burglaries in terraced houses per thousand terraced houses). This gives us information about the risk of crime in particular types of dwelling and how these vary across the different Super Groups.

Table 3 Burglary rates by Housing Type by OAC Super Group

OAC Super Group	Burglary rate				
	Per 1,000 Detached	Per 1,000 Semis	Per 1,000 Terraced	Per 1,000 Corner	Per 1,000 Properties*
1. Blue-Collar Communities	32.2	25.1	14.4	22.4	22.5
2. City Living	93.4	64.5	78.5	83.6	76.4
3. Countryside	21.6	13.2	3.2	14.4	16.4
4. Prospering Suburbs	25.3	23.9	9.6	24.4	24.0
5. Constrained by Circumstances	88.1	31.3	27.1	29.6	33.6
6. Typical Traits	27.6	25.9	16.6	24.6	23.2
7. Multicultural	79.7	45.1	33.7	60.5	41.7
Leeds	33.6	28.7	27.5	33.2	29.7

* This figure excludes flats

There were some very interesting differences between the Super Groups and the likelihood of burglary in particular types of property; something that would not be obvious from studying conventional crime rates. For example, City Living areas had consistently higher burglary rates than any other Super Group for each of the different property types. A detached property in City Living neighbourhoods was over four times more likely to be burgled than a detached property in parts of Leeds classified as Countryside and over three and a half times more likely to be burgled than a detached house in the Prosperous Suburbs.

City Living areas have a lot of privately rented property in them and younger single residents (See Table 1) which often means that there are fewer people around during the day to keep an eye on things. This is what criminologists call 'poor guardianship' and neighbourhoods like this are known to have higher levels of crime. These were clearly the types of area in Leeds most vulnerable to burglary.

Countryside consistently had the lowest burglary rate for each type of property in Leeds. This meant that the greatest differences in burglary were between City Living and Countryside.

But some distinctive patterns also emerged. Most property types in Constrained by Circumstances areas had burglary rates that were more or less in line with the average for Leeds except for detached houses which were over two and a half times more likely to be burgled in these areas than in Leeds generally. There were also exceptionally low rates of burglary among terraced properties in Blue-Collar Communities and Prospering Suburbs all of which call for further investigation and, significantly, all of which would have been masked using conventional burglary rates.

Analysis of Road Accessibility

The accessibility of each street segment (that is stretches of road between road junctions) to each other street has also been calculated from street network data (see [here](#) for information about this method or

[here](#) to download street accessibility data). A measure of accessibility can be useful in crime analysis because it has been found to correlate with levels of vehicle or pedestrian traffic - highly accessible roads can expect higher volumes of traffic. Just as it was possible to look at differences in burglary rates by property type across the Super Groups, the GeoCrimeData Project has also made it possible to examine such differences in parts of Leeds that had been grouped together according to how busy the streets are.

Each street segment has been given a busyness score or what is technically called an 'Integration Value'. These integration values for all of the streets occupying each Output Area have then been added up and divided by the number of streets in each OA to give an average integration value. Once these have been calculated for each OA it is fairly straightforward to rank the OAs in decreasing order of their integration values and to identify the top 10% of OAs with the highest values (i.e. those having the busiest streets in Leeds), the next 10% of areas and so on until all OAs have been grouped together. We have defined 10 groups or deciles each containing 10% of OAs ranging from the top 10% (the busiest streets) through to the bottom 10%(OAs with the least busy/ most isolated streets) in Leeds.

This has enabled us to answer the question 'how do burglary rates in the most accessible areas compare with those in the least accessible areas?' and to go on to answer the more precise question 'how do burglary rates within each property type vary by street accessibility in Leeds?'.

Table 4 shows burglary rates by house type for each accessibility decile. The least accessible/most isolated areas are towards the top of the table with accessibility increasing further down the table until Decile 10 is reached which is the 10% of OAs with the most accessible/busiest streets in Leeds.

Table 4 Burglary rates by Housing Type and Accessibility of Streets

Street Accessibility Decile	Burglary rate				
	Per 1,000 Detached	Per 1,000 Semis	Per 1,000 Terraced	Per 1,000 Corner	Per 1,000 Properties*
Decile 1*	21.4	17.6	13.5	20.7	18.2
Decile 2	24.4	20.2	15.0	16.0	19.7
Decile 3	25.6	20.0	25.1	23.5	22.5
Decile 4	27.8	31.2	25.9	28.4	29.1
Decile 5	31.3	28.4	35.2	36.5	31.5
Decile 6	40.2	32.4	33.8	46.5	35.3
Decile 7	38.7	30.6	33.9	41.4	33.6
Decile 8	37.5	34.6	35.2	38.8	35.5
Decile 9	53.7	32.8	22.4	41.9	33.2
Decile 10**	65.6	38.0	26.2	44.2	40.3
Leeds	33.6	28.7	27.5	33.2	29.7

Notes: *10% of OAs with least accessible streets **10% of OAs with most accessible streets

There appears to be a strong link between increasing levels of accessibility and burglary rates within detached houses. This is a positive relationship which means that the busier the streets, the higher the burglary rate. With just two exceptions (Deciles 7 and 8), the higher the decile the higher the burglary rate for detached properties. This suggests that detached properties were much more vulnerable if they

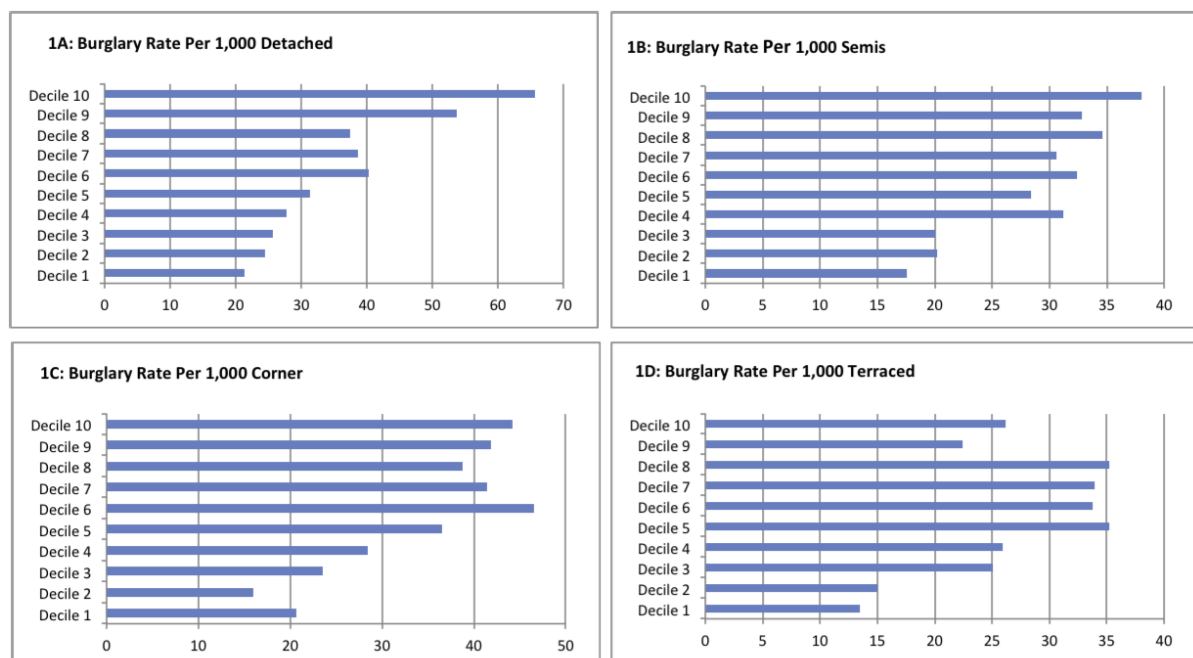
were in more accessible streets. For example, detached houses were over three times more likely to be burgled in the most accessible 10% of OAs compared with the least accessible 10%; a major difference indeed! The relationship between accessibility and burglary detached houses can clearly be seen in Figure 1A.

Semi-detached and corner properties also tended to be at greater risk the higher the levels of accessibility, although, for these two house types the relationship was not as strong as that for detached properties. The links between increasing accessibility and higher burglary for house types can clearly be seen in Figures 1B and 1C. Semi-detached houses and corner properties were just over twice as likely to be burgled in the most accessible 10% of OAs compared with the least accessible areas (Decile 1).

Interestingly, burglary levels of terraced properties were not quite as susceptible to increasing levels of accessibility compared with other house types. Burglary rates in the most accessible decile (10) were just under twice those in the least accessible decile (1). Burglary rates in Deciles 9 and 10 were actually lower than those in deciles 5 through 8. The burglary rates for terraced houses in Decile 4 was just below that in Decile 10. These fluctuations are clearly shown in the bar graph in Figure 1D. What this indicates is that accessibility might be less of a factor in the burglary of terrace properties than it is for detached and semi-detached houses.

Once again, this finding would not have emerged without bringing together different geospatial datasets!

Figure 1 Burglary rates and street accessibility decile graphs



Discussion

There are, of course, many other questions that could be explored by cross-referencing the geospatial datasets, discussed above, in different ways. We have not had time to do this for now, but we are aware of what could be done. For example, we could find out a lot more about different places and what they are like even before we look at crime data (i.e. learning about context without crime). Another topic to explore might be to look at the geo-demographic composition of high and low accessibility areas. A relevant question here would be: Which OAC Super Groups do output areas with the most accessible streets fall into?

Another research task might be to find more out about the accessibility of streets in each OAC Super Group. This sounds similar to the previous question but actually it is quite different. It would involve selecting the output areas in each OAC group and then calculating their accessibility score (integration value). The results could then be put into a simple table. It would then be possible to look at geo-demographic clusters in a new way i.e. in terms of the accessibility and connectivity of the streets within them.

Further information will be provided if we looked at the accessibility of streets that have different types of housing within them. For example, we could find out what percentage of output areas with the greatest proportion of terraced properties fall into the highest accessibility decile? Alternatively, we could identify what proportion of properties in output areas with the most accessible streets are terraced, semi detached, detached?

Needless to say that a lot more that could be explored bringing crime into the picture (i.e. crime with context). An analysis which we have not done, would be to explore how far there is a difference in accessibility between high burglary and low burglary output areas ? This would involve grouping output areas into deciles on the basis of their burglary rate (e.g. output areas falling into the top 10% on burglary rates through to bottom 10%) and identifying street accessibility within each burglary decile.

Yet another analysis would be to look at the property type profile of high and low burglary communities. For example, what is the mix of housing in the top burglary decile and how does that compare with the house type profile in the lowest burglary decile?

Our online survey and [Workshop](#) in September 2011 highlighted a number of other more detailed analyses that could be carried out using geospatial datasets. Linking variations in burglary by time of day to geospatial data provides one example of what could be done. For example, how far daytime and nighttime burglaries vary by street accessibility could be examined and the extent to which there is any difference in the timing of burglaries by OAC Super Group could be explored.

The single biggest risk factor of burglary is having experienced a previous burglary. When houses have been burgled twice or more within a year, this is called repeat victimisation. There is room to explore the relationship between repeat victimisation and house type and then to extend this further by looking at how far the number of repeat burglaries in particular types of housing (e.g. detached, semis and terraced houses, etc) is greater (or indeed far lower) in particular types of neighbourhood using OAC Super Groups. For example, are detached houses more likely be repeatedly burgled in City Living or Multicultural Areas than in Constrained by Circumstances Neighbourhoods ?

It would also be possible to look all three things together, for example, comparing repeat burglaries of

detached properties in City Living areas that are highly accessible compared with City Living areas that has more isolated with less busy streets. We are sure that you will have lots of other ideas and it would be great if you could share them with us!

However, the crime risks in a neighbourhood might be influenced, not only, by what the neighbourhood itself is like in terms of people, housing and access, but also by the types of areas that are next to it. Using geospatial data it is possible to look at what we call the 'spatial configuration of neighbourhoods' that is, what is next to what (e.g. a rich area surrounded by lots of poorer areas) and how this varies across the city. We could then compare burglary rates for different combinations of neighbourhood (e.g. rich surrounded by poor, student areas next to housing for the elderly, blue-collar neighbourhoods bordering multicultural areas and so on).

There is little doubt that looking at the relationship between OAC Super Groups and burglary rates within particular types of houses is very interesting. But there may also be some important variations in burglary rates within the Super Groups themselves. This can easily be established by examining burglary rates for the OAs belonging to each Super Group.

If there are great variations in levels of crime *within* a single OAC group, it is necessary to explore possible reasons for such variations. For example, although output areas belonging to a particular Super Group are likely to have similar housing and population characteristics (that is why they are in the same Super Group in the first place!), any marked differences in burglary between them might be because they vary in other ways. For example, higher crime might occur among areas *within* an OAC Super Group only where neighbourhoods border a particular combination of other Super Groups. Alternatively, burglary might be higher among the output areas in a particular Super Group because these have a high level of street accessibility.

Other suggestions included using geospatial data to identify the effect on burglary of differences in the design and layout of residential housing developments. Some interesting ideas were put forward, particularly, how burglary might be influenced by the presence of footpaths and how well these were connected to roads, the influence of cul-de-sacs and the visibility of streets from individual houses. Although we have been able to calculate some of these metrics in our household-level data (which is available for [download](#)) we have not begun to analyse them yet.