

10. Detailed methodology: a guide to reading the evidence base

Chapter 3 of the evidence base dealt in detail with some important techniques used to visualise mainly spatial data. It also very briefly introduced two contextual data sets, the Indices of Multiple Deprivation (IMD) and Output Area Classification (OAC), as well as stating that proven statistical techniques have been used throughout the report to analyse data. This section serves as an extension to Chapter 3. With the aim of further easing readers into IMD and OAC, it gives a profile of Leicestershire based on both indicators. It also briefly names and defines in simple terms each statistical technique used in the report.

IMD

In Chapter 3 we said that the most logical way of analysing IMD ranks for all 32,482 neighbourhoods in England is by breaking LSOAs into equal parts. The histograms in figure 10a are a useful way of understanding relative levels of deprivation based on these breaks. They split ranks up into 20 percent bandings, from the 20 percent most- to least- deprived in England, and show how many LSOAs in Leicester, Leicestershire and its districts, lie within each.

Probably as expected, relative to the rest of England most LSOAs in Leicestershire experience low levels of deprivation. Forty-six percent, or 182, of the county's LSOAs are amongst the 20 percent least deprived nationally and, moving towards quintiles with moderate and then higher levels of deprivation, the number of neighbourhoods consistently drops. This is almost symmetrically opposite for Leicester City, where 47 percent of its LSOAs are amongst the 20 percent *most* deprived nationally. Despite this, there are in Leicestershire a small number of LSOAs - only 5, or 1 percent of Leicestershire's LSOAs - which do suffer from higher levels of deprivation and which are amongst England's 20 percent most deprived LSOAs. These five LSOAs are located in Charnwood and North West Leicestershire. Although both districts do contain higher levels of deprivation than the rest of Leicestershire, LSOAs in the most deprived quintile still represent only 3 percent of all LSOAs for Charnwood and North West Leicestershire. Looking at the shape of the histograms for both districts, they still experience relatively low levels of deprivation. Charnwood, for instance still has its highest proportion of LSOAs in the least deprived quintile.

Figure 10a: LSOAs in Leicester and Leicestershire by 20 percent deprivation bandings

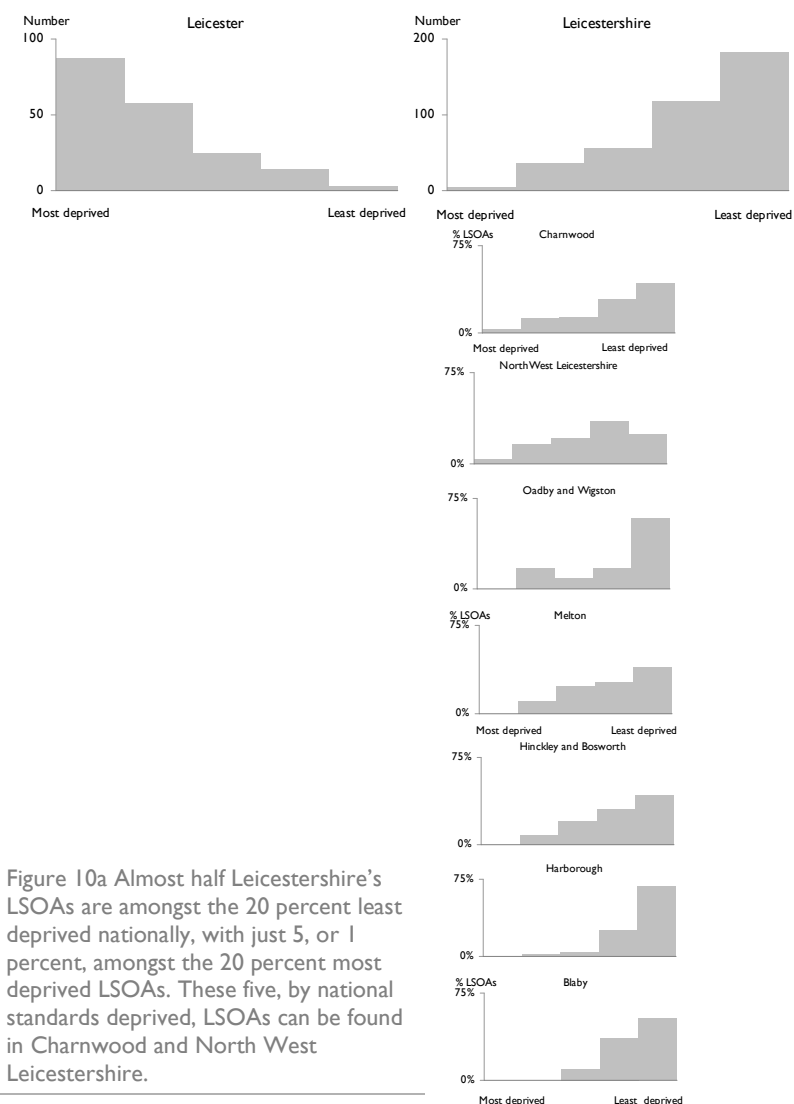


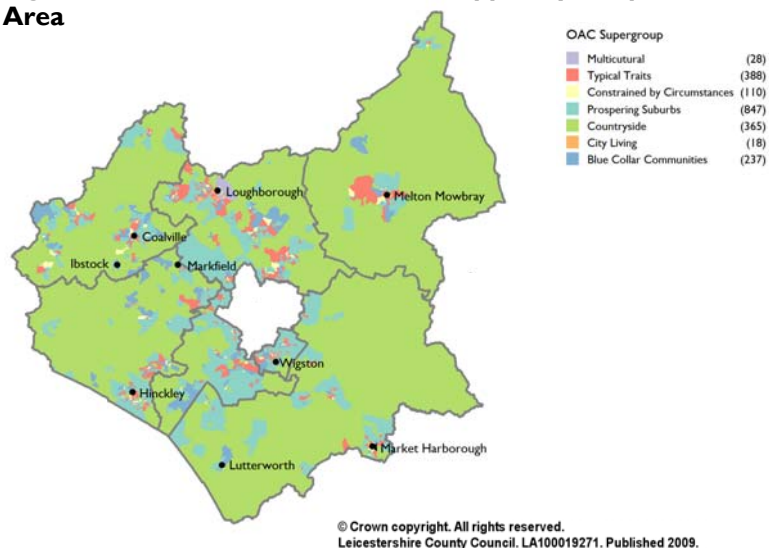
Figure 10a Almost half Leicestershire's LSOAs are amongst the 20 percent least deprived nationally, with just 5, or 1 percent, amongst the 20 percent most deprived LSOAs. These five, by national standards deprived, LSOAs can be found in Charnwood and North West Leicestershire.

Source: IMD, 2007

OAC

As set out in Chapter 3, the 2001 Census Output Area Classification (OAC) is another way of understanding the social geography of the county. It does this by grouping together small geographic areas - Census Output Areas containing roughly 120 households - whose residents share similar or distinct social-economic characteristics. At its most general level, there are seven broad clusters under which the 1,993 OAs in Leicestershire are defined: *prospering suburbs*, *typical traits*, *countryside*, *blue collar*, *constrained by circumstances*, *multicultural communities* and *city living*. Figure 10b shows where Leicestershire and its districts figure in respect to the OAC clusters. Overall, 42 percent of the county's OAs (almost half Leicestershire's population) live in areas labelled as *prospering suburbs*. In Blaby and Oadby and Wigston this cluster makes up 64 and 65 percent respectively of those districts' population. The corollary is that OAs classified as *multicultural communities* and *city living* each constitute just 1 percent of all OAs. This is understandable since Leicestershire is a relatively rural county. That the *constrained by circumstances* classification represents just 6 percent of OAs in the county, and that *prospering suburbs* is the most dominant classification, also makes sense given what we know about levels of deprivation in Leicestershire.

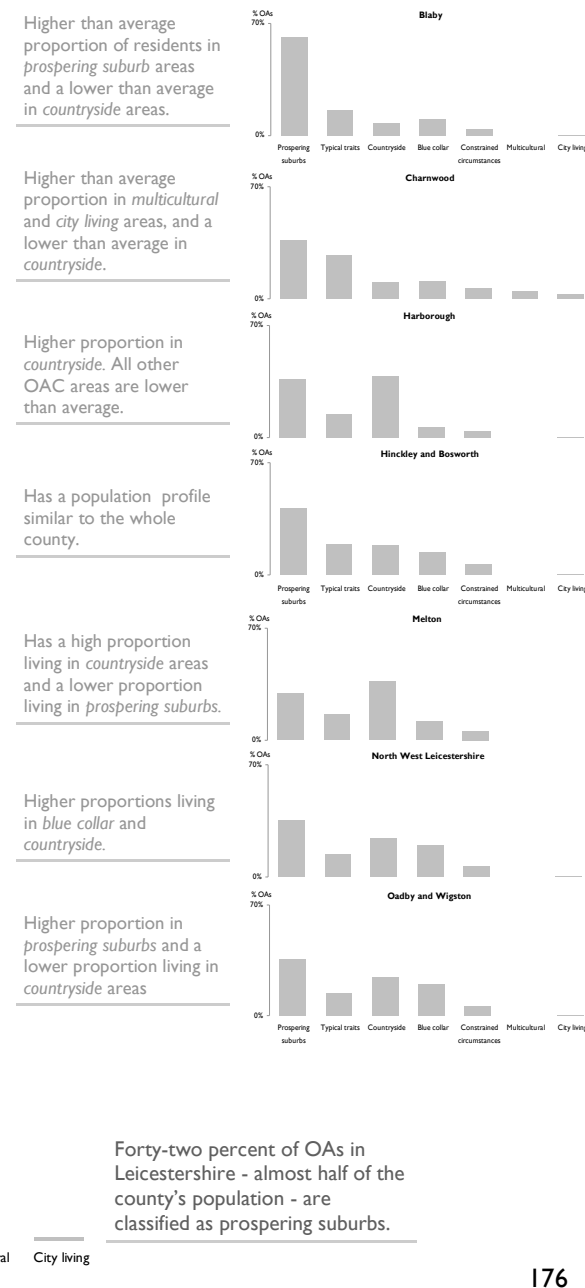
Figure 10c: OAC in Leicestershire mapped by Output Area



© Crown copyright. All rights reserved. Leicestershire County Council. LA100019271. Published 2009.

Source: ONS, 2005

Figure 10b Output Areas in Leicestershire and its districts by OAC



Data analysis

Again Chapter 3 briefly outlined why statistical testing has been used throughout the evidence base. In most instances it has helped to identify people and places likely to be affected by a particular issue or to think a certain way. Using techniques such as Pearson's Chi-square, for example, we have found associations between variables which are statistically significant - that is an event which is extremely unlikely to occur by chance and so which could be of interest. The evidence base also attempts to explain certain occurrences/phenomena by taking into account a combination of factors (or variables). In addition, a method of cluster analysis has been used to group together geographical areas which are similar or distinct, based on characteristics which are of particular interest. The specific methods behind this analysis are defined and explained below.

Cross-tabulations using Pearson's Chi-square test are applied in chapters 5, 7 and 8. Pearson's Chi-square looks at the frequency distribution of variables within a sample. It is assumed that each variable is mutually exclusive of each other - that there is no relationship between variables and that, where there is any variation, this can be explained by chance. Hypothetical values for variables are calculated based on this assumption. These 'expected' values are compared against the original observed ones and a Chi-square test statistic - a measure of the difference between observed and expected values - is produced. A 'hypothetical' distribution of this test statistic is then arrived at assuming there is no relationship between the two original variables. By comparing the observed Chi-square test statistic against its hypothetical distribution, we can say how unlikely it is for the variables to be independent of each other; whether or not differences between the two variables can be explained by chance, or whether there may be some sort of influential relationship between them.

Mosaic Plots are used in Chapter 5, 6, 7, 8 and 9 and are graphical representations of cross-tabulations. In a mosaic plot, assuming a normal distribution, the statistical significance of variables is tested by calculating the standardised residual of each category combination. Whereas chi-square provides an overview of whether or not there is a significant relationship between two variables, by using standardised residuals mosaic plots can tell us which specific category combinations are, or are not, significant.

CHAID is a technique which enables researchers to identify the order and magnitude of statistically significant differences (the output from cross-tabs or mosaic plots). It does this by identifying the most and least important variables (independent variables) which can be used to explain a particular outcome (a dependent variable). The technique is used in Chapters 5, 6, 7 and 8 to identify groups of people who are most likely to think a certain way, or to be affected by a certain issue.

CHAID stands for CHi-square Automatic Interaction Detector. It divides up cases, in our example *Place Survey* respondents, into sub-groups based on the best predictor/explainer of a dependent variable. An example might be the extent to which car ownership (independent variable) determines/explains ease of access (dependent variable). The analysis then splits each of these into smaller sub-groups based on other explanatory variables (so maybe, age and social-economic status). This continues until no more statistically significant predictors can be found. CHAID is, then, a way of finding meaningful sub-groups of a population, from a potentially complex range of variables, which explain a certain outcome.

Logistic regression is a form of multivariate analysis which can be used to predict the presence or absence of an outcome based on a number of predictor (independent) variables. It is used where the outcome (dependent) variable is dichotomous - where there can be only two possible outcomes (in this instance satisfied or dissatisfied). Chapter 8 draws on logistic regression analysis in order to estimate an odds ratio for the variables which are most likely to predict whether or not residents are satisfied with their local transport system.

K-means cluster analysis is used in Chapter 4. Rather than considering differences within a data set and testing whether there might be a relationship between values, cluster analysis is an explanatory technique which is concerned with finding out whether or not values are similar and can fall into natural groups. First, each case, or unit of interest, is partitioned into a cluster, which contain similar characteristics or values. The cases are then reallocated numerous times to improve the quality of the clusters. A cluster can be defined as reliable if the average distances between cases in each cluster is small. Cluster analysis enables us to not only identify cases whose characteristics are very different to each other, but also to statistically group together cases which are similar.

11. Glossary

Cartograms are used to represent spatially, data on people. Whilst conventional Ordnance Survey maps are ideal for showing the physical geography of a local area, they usually give emphasis to areas where few people live. Overlaying information on people onto these maps is therefore problematic. One solution is a cartogram. The Leicestershire cartogram consists of 396 hexagons, each representing an LSOA. Although with cartograms there is a risk of losing some geographical orientation, each hexagon is ordered spatially and it is therefore possible to identify spatial patterns in whatever data are encoded in them.

Gross Value Added (GVA) is a measure of the total value of goods and services that a geographical area produces. At a local level it is calculated by summing all incomes generated in the production process.

Census Output Areas (OAs) are small neighbourhoods of roughly equal size - each contains approximately 120 households. There are 1,993 OAs in Leicestershire.

Decile and quintile A decile is a means of ordering continuous data and then splitting its records into 10 equally sized groups. A quintile works in the same way but splits records into 5 equal groups.

The **Indices of Multiple Deprivation (IMD)** data set is one of the most comprehensive measures of neighbourhood level deprivation in England. It uses a number of Census 2001 indicators aggregated to lower Super Output Area level to give a composite deprivation score. IMD breaks down into seven 'deprivation domains'. These are: Employment, Education, Skills and Training, Income, Crime, Health, Barriers to Housing and Services and Living

Environment. IMD scores are best analysed relative to each other. Throughout this report we therefore consider how Leicestershire's 396 LSOAs rank compared to the 32,483 in England.

Lower-layer Super Output Areas (LSOA). Lower Super Output Areas are a collection of Census Output Areas which are generally smaller than wards and have about 1,500 people living in them. There are 32,482 LSOAs in England, and 396 in Leicestershire.

Output Area (OA) There are 175,500 Output Areas in England and Wales and 1,993 in Leicestershire. Each contains approximately 120 households. Output Areas are the smallest geographical areas to which Census 2001 data is aggregated.

Output Area Classification. The Census Output Area Classification (OAC) is a way of understanding the social geography of an area. It looks at 50 Census indicators to group together small geographic areas - Census Output Areas - whose residents share similar or distinct social-economic characteristics.

A **sub-region**, or city-region, represents the space over which local economies and their key markets operate. Although sub-regions do not adhere neatly to administrative boundaries, where the term is used in this report it refers to all activity concentrated within the boundary of Leicester and Leicestershire.

Electoral **Wards** are residential areas which vary in size but contain on average around 5,500 people. They are built up from Lower Super Output Areas.