GLAECONOMICS

Working Paper 30
London's Housing Submarkets
By **Said Hirsh** (Volterra Consulting)

Transport for London

LONDON DEVELOPMENT AGENCY

MAYOR OF LONDON

copyright

Contents

1 Executive summary

London's housing market is inherently complex and diverse, with substantial variations existing among the various attributes that characterise different parts of London. We therefore divide London up into various submarkets for descriptive purposes, rather than attempting to analyse the market for housing in London as a whole. When considering where to buy, purchasers will consider areas with specific attributes and within their price bracket.

This working paper updates the analysis in GLA Economics' 'Working Paper 7: London's Housing Submarkets', published in 2004. All of the variables used in the analysis have been reviewed. More recent data has been used where available, and alternative variables considered for inclusion. Furthermore, the methodology has been reviewed where appropriate. The analysis is carried out at a postcode sector level.

Following a qualitative and quantitative selection process, the following variables were used to group the postcode sectors:

- Percentage of income support claimants
- Number of criminal damage incidents
- Average number of rooms per household
- Transport accessibility (PTAL (Public Transport Accessibility Levels) 2006 scores)

The analysis identifies four distinct housing sub-groups in London. Each had different characteristics, and hedonic stepwise regression modelling was used to identify the key drivers of house prices within the identified groups. The four housing subgroups identified are:

- Central
- High Density Housing
- Outer City
- Spacious Suburbs

Figure 1: The four identified groups
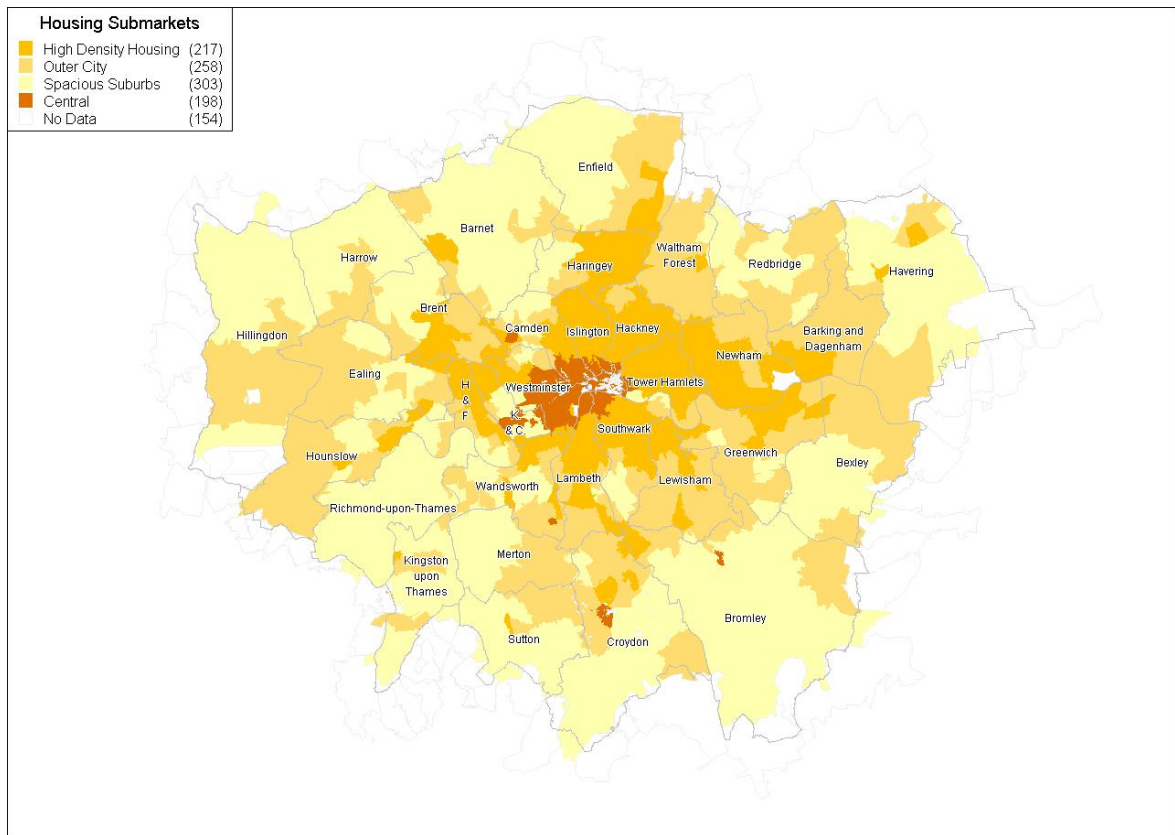
Source: GLA Economics

The Central Group was characterised by some of the highest house prices in London and also had a lower than average incidence of criminal damage. Income level within this group was also particularly high. The Central Group was predominately made up of smaller dwellings.

The High Density Housing Group was heavily populated with high levels of income support claimants and a higher than average level of criminal damage.

The Outer City Group had some of the lowest house prices in London. Geographically, the group covered large areas of Hounslow, Hillingdon, Ealing in the west and Barking and Dagenham and Waltham Forest in the east.

The Spacious Suburbs Group had the second highest house prices of all the groups with a low level of income support claimants. However accessibility scores that determine travel time into central London for these areas was particularly low.

Whilst many of the results are consistent with what is commonly assumed about people's preferences, there are a significant number of unexplained results that indicate further work is required. The casual link assumed between explanatory variables and house prices may not be as strong as expected. This could be due to unexplained variables such as quality of local services and facilities.

2   Introduction

The London housing market is inherently complex and diverse, with substantial variations existing among the various attributes that characterise different parts of London. Area attributes that buyers may consider when searching for a property include accessibility, poverty, demographics, quality of school provision and availability of facilities and services to name just a few. Such attributes mean that a dwelling in one part of London cannot easily be substituted with a dwelling in another. Moreover, different individuals attach different importance to such factors.

This lack of substitutability implies that it may be appropriate to divide London up into various submarkets for descriptive purposes, rather than attempting to analyse the market for housing in London as a whole. When considering where to buy, purchasers will consider areas within their price bracket which may cover more than one submarket.

GLA Economics' 'Working Paper 7: London's Housing Submarkets', published in 2004, used a clustering analysis to define and analyse submarkets in London. The advantage of this type of analysis is that it provides local information without the need to have specific knowledge of the local area – it is a data driven analytical approach. Once groups of postcode sectors have been defined it enables within-group analysis, e.g. regression analysis, which may perform better than analysis at an overall level where there are too many unobserved drivers of house prices to make analysis of the whole of London possible.

This working paper takes the analysis in Working Paper 7 as a starting point. All of the variables used in the analysis have been reviewed. More recent data has been used where available, and alternative variables were considered for inclusion. Furthermore, the methodology has been reviewed where appropriate. This includes relaxing the restriction in the within-group regressions that there should be precisely five variables.

The structure of this working paper is as follows. Section 2 covers the data used and the selection of the variables for the clustering analysis. Section 3 presents the groups, which are described in Section 4. Within-group analysis is initially conducted using stepwise regression. The methodology and regression results are set out in Section 5 and they are discussed in Section 6. Section 7 concludes, and further detail is presented in the appendix.

3   Data and variable selection

3.1   Data

Data was collected for the 1,130 postcode sectors in London[1] where possible. Due to missing data, a number of adjustments to the dataset had to be made. These consisted of:

---

[1] This includes all postcodes sectors that intersect the London boundary

- The exclusion of 153 postcode sectors due to missing House Price Data
- The exclusion of 1 postcode sector due to missing Income Data

The house price data that was used in the analysis was supplied by Hometrack, and is based on Land Registry Data.

A number of variables that could determine house prices were considered for inclusion in the study. Following removal of variables that were unreliable or exhibited a high degree of closeness to other variables (collinearity) a total of 27 variables, in addition to the house price variable were selected for the study. A description of the variables used in the analysis is presented in the appendix.

3.2 Selection of the grouping variables

In order to conduct the cluster analysis, key variables from the dataset were selected. The variables were selected using expert judgement informed by statistical analysis and expertise within the GLA including that from the Housing Policy team.

In assessing variable suitability the following criteria were used:

- Variables should represent the main attributes that are believed to determine house prices
- Variables should not be correlated with each other
- The number of variables selected should be small so that results can be understood

The statistical analyses performed to enhance understanding of the dataset and to inform the selection of grouping variables were:

- Pairwise correlations among all the variables, including log house prices
- Univariate regressions of log house prices against each of the other variables
- A stepwise regression of log house prices against the other variables.

Following this analysis the following variables were accordingly selected:

- Percentage of income support claimants
- Number of criminal damage incidents
- Average number of rooms per household
- Transport accessibility (PTAL 2006 scores)

Maps of these variables are presented in the appendix.

These variables were then used to split London into groups using a probabilistic logic clustering technique[2]. The identified groups are discussed in the following sections.

---

[2] A full technical discussion of the technique is presented in the appendix

4   Determining the groups

The analysis identified four groups of postcode sectors within London. The number of groups was determined by examining the following factors:

- Whether each of the resulting groups were well defined (goodness of fit)
- The stability of the groups in terms of the central values of the variables used to identify the groups following a number of trials
- That the groups made intuitive sense

Following analysis it was found that a maximum of four groups could be used to create well-defined, stable groups.

The four groups identified are:
- Central
- High Density Housing
- Outer City
- Spacious Suburbs

Table 1 sets out the core values of the variables for each of the identified groups.

Table 1: Core values for the four groups

	CLAIMANT RATE	CRIMINAL DAMAGE	AVG RMS / HOUSEHOLD	PTAL 2006
Central	5.54	1.73	3.61	61.48
High Density Housing	8.53	14.11	4.20	19.57
Outer City	5.38	15.56	4.80	11.98
Spacious Suburbs	2.70	9.23	5.53	9.87

The mean and median house prices for the groups are set out in Table 2.

Table 2: Mean and median house prices for the four groups

	Mean House Prices	Median House Prices
Central	£627,164	£532,688
High Density Housing	£363,110	£296,977
Outer City	£299,177	£231,583
Spacious Suburbs	£440,801	£320,526

Figure 2 presents a map of London showing the location of the postcode sectors within the various groups.

Figure 2: The four identified groups
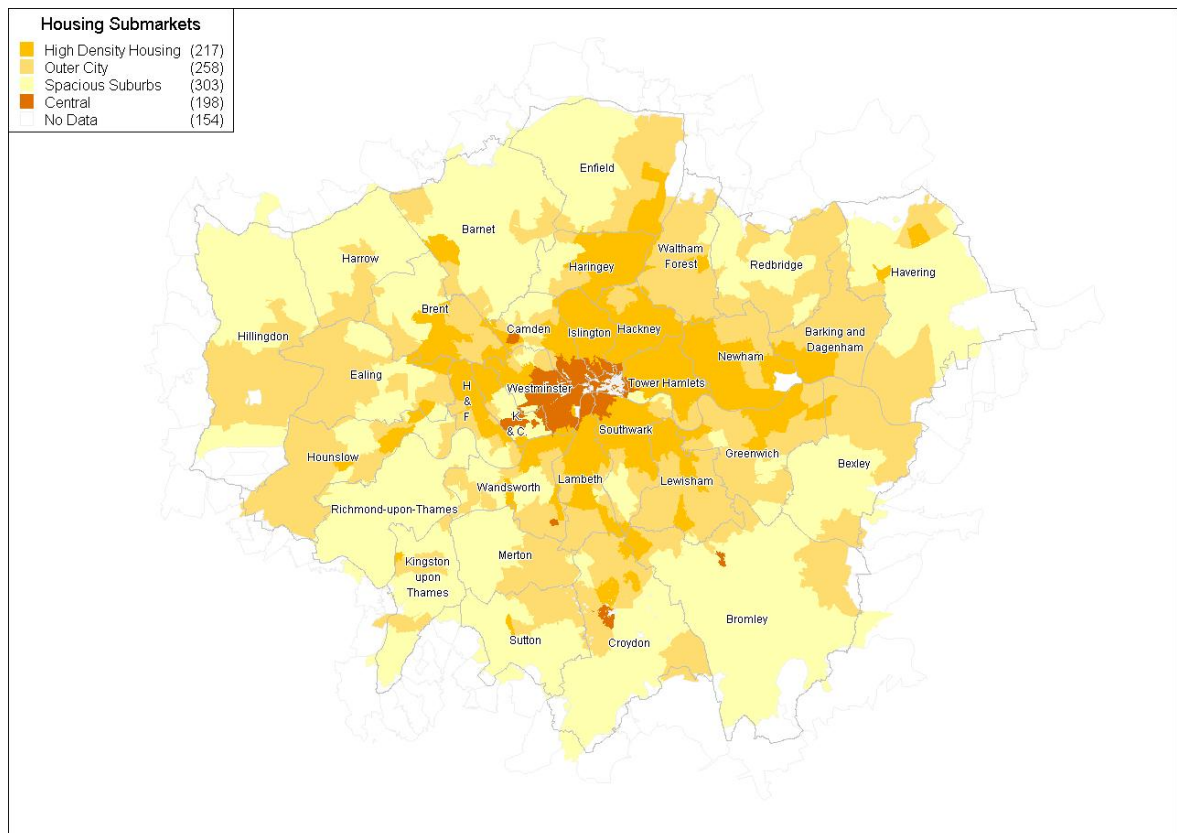
Source: GLA Economics

Summary statistics can be used to further characterise the nature of the groups. For example the percentage of dwellings that are social rented is 44 per cent in the High Density Housing Group, compared with just 10 per cent in the Spacious Suburbs Group.

Summary statistics giving the average value for each variable are presented in the appendix. These are referred to in the context of the description of the four groups in the following section[3].

5   Description of groups

Central Group

The Central Group has the highest house prices of all the groups with a mean of around £627,000. This group has extremely high transport accessibility and a very low incidence of criminal damage. Of all the groups it has the lowest number of rooms per household and it has a middling level of income support.

---

[3] Analysis of how well these groups are defined is provided in the appendix

Other key features of the Central Group include its high average income and high percentage of part time workers. Not only is the group made up predominantly of small dwellings, but it is also the most overcrowded[4].

Whilst the Central Group has a particularly low incidence of criminal damage crimes it is interesting to note that the incidence of theft and handling crimes is comparable to that in the Spacious Suburbs Group. This may be because criminal damage is particularly deterred in the Central Group by the large number of CCTV cameras in the area, whereas theft and handling cannot be so easily deterred.

Geographically, the Central Group covers much of Westminster, along with postcode sectors close to the City of London in boroughs such as Camden, Islington, Hackney, Tower Hamlets, Lambeth and Southwark. It also covers a few postcodes in the borough of Kensington and Chelsea. Notably, the northern half of Westminster and a few postcode sectors towards the south of the borough do not fall within this group.

A few interesting exceptions can be observed. A number of postcode sectors outside Central London fall within this group due to their high accessibility. These include NW6 3 in Camden, CR0 1 in Croydon and BR1 1 in Bromley. Of these, the most accessible is Croydon Town Centre due to National Rail stations and the Croydon tramlink.

One other exception is SW17 6 in Wandsworth. Unlike the others, this postcode sector does not have a high accessibility score, and it falls within this group due to its other attributes such as its low number of incidents of criminal damage.

High Density Housing Group

The High Density Housing Group has the second lowest house prices, with a mean of around £363,000. It is the group with highest average population density at 9,700 people per square kilometre, over 40 per cent higher than that in the Central Group. It has high levels of income support claimants and incidences of criminal damage. The group does, however, benefit from good public transport accessibility.

Average incomes in the High Density Housing Group are the lowest of any of the groups, and the group has the highest percentage of social rented dwellings. A relatively large proportion of the inhabitants of this group are from ethnic minorities.

---

[4] Overcrowding is measured by whether a household lacks one or more rooms according to the Census 'occupancy rating'. For example a value of -1 implies that there is one room too few and that there is overcrowding. It relates the actual number of rooms to the number of rooms 'required' by the members of the household (based on an assessment of the relationship between household members, their ages and gender). For example, a one person household is assumed to require three rooms – two common rooms and a bedroom. In the case of two or more people households, the rating is slightly more complicated as this is dependent on the relationships, ages and gender.

The results for Central London are skewed due to its relatively high proportion of studio flats, which are usually defined as overcrowded by this measure. Preferences for more open plan living may also mean that this variable is less valid nowadays.

Geographically, this group covers large parts of Islington, Hackney, Tower Hamlets, Haringey, Southwark and Lambeth. The group also covers some northern postcodes in Westminster, Kensington and Chelsea and Hammersmith and Fulham along with parts of Brent, Wandsworth, Barking and Dagenham, Greenwich and Enfield.

Other postcode sectors fall within this group on the outskirts of London due to the particular characteristics of the area. For example, postcode sector KT1 1 in Kingston has a high income support claimant rate in line with the rest of this group. Postcode sector NW9 5 in the borough of Barnet has a high proportion of income support claimants and a high incidence of criminal damage.

Outer City Group

The Outer City Group has the lowest house prices of all the groups, with a mean of £299,000. This group has high incidences of criminal damage and relatively low accessibility. The income support claimant rate is similar to that in the Central Group and it has a relatively high number of rooms per household.

This group has a high proportion of dwellings that are terraced houses. Its low accessibility is illustrated not only in the PTAL scores but also in the high distances to a mainline rail or underground station.

Geographically, this group covers large areas in Outer London in boroughs such as Hounslow, Hillingdon, Ealing, Barking and Dagenham, Waltham Forest, Bexley, along with areas in Greenwich, Lewisham, and parts of Enfield, Merton, Sutton and Croydon.

Other postcode sectors fall within this group, for example NW1 4 in Westminster is in this group due to its income support claimant rate and incidences of criminal damage being typical of the group.

Spacious Suburbs Group

The Spacious Suburbs Group has the second highest house prices, in spite of the fact that it has the lowest public transport accessibility of all the groups. The group is characterised by its high proportion of large dwellings both in terms of rooms per household and people per household. It has a very low proportion of income support claimants, and a relatively low incidence of criminal damage.

The group has the highest proportion of owner occupied dwellings, and the proportion of detached and semi-detached dwellings are also the highest of any of the groups. The group also has the highest proportion of couples with dependent children and the highest percentage of children obtaining five A*s-Cs at GCSE or equivalent. A high proportion of inhabitants of postcode sectors within the group are white.

Geographically, this group includes large areas of Outer London, including the majority of the boroughs of Bromley, Croydon, Sutton, Richmond, Kingston, Harrow, Barnet, Enfield and Havering along with parts of Merton, Hillingdon and Redbridge.

Somewhat surprisingly, a large number of postcode sectors in Kensington and Chelsea fall within this group. This is due to the very low income support claimant rates in these particular postcode sectors.

5.1   Descriptive statistics on house prices

The degree of variation in house prices across the groups is shown in Table 3 and Figure 3. This shows that the Central and Spacious Suburbs Groups contain the largest degree of variation with the Spacious Suburbs Group also containing a large number of outlier values as shown in the box plots in Figure 3.

Table 3: Summary statistics of house prices by group

	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
Central	198	£627,164.2	£335,118.1	£158,127.7	£1,900,000.0
High Density Housing	217	£363,110.0	£214,914.3	£156,126.4	£1,625,387.0
Outer City	258	£299,177.0	£192,808.7	£139,463.2	£1,493,099.0
Spacious Suburbs	303	£440,801.3	£332,084.9	£175,490.5	£2,167,398.0

Figure 3: Box plots of house prices by group
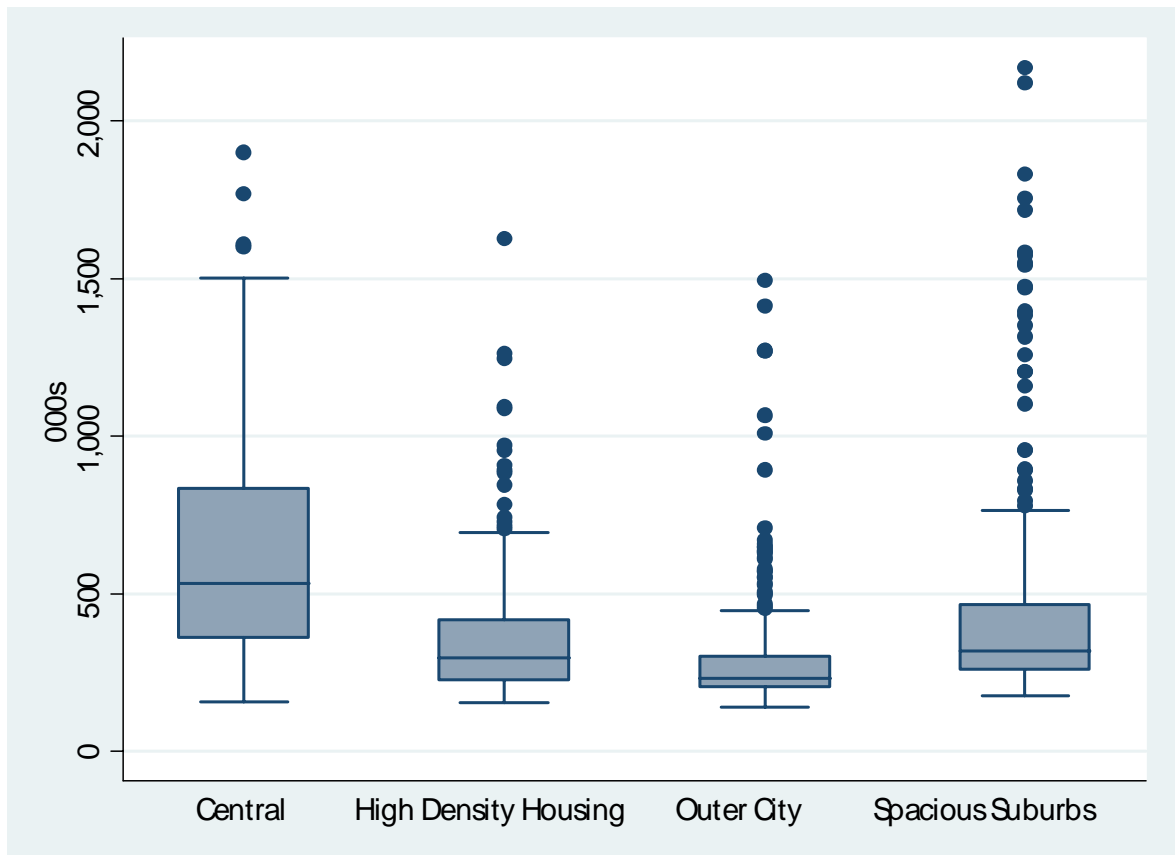
5.2   How well defined are the groups?

The probabilistic logic clustering technique used in this working paper assigns each postcode sector a percentage membership in each of the four groups identified. Each postcode sector is then assigned to the group that it has the highest membership in. It is interesting to consider the strength of membership of the postcode sectors assigned to each of the groups to understand how well these groups are defined[5].

Figure 4 shows the degree of membership for each postcode sector in the group it belongs to. This shows that the groups are generally well defined with 70 per cent of the postcode sectors having a degree of membership of over 50 per cent.

---

[5] The strength of membership is discussed in more detail in the appendix.

Figure 4: Strength of membership
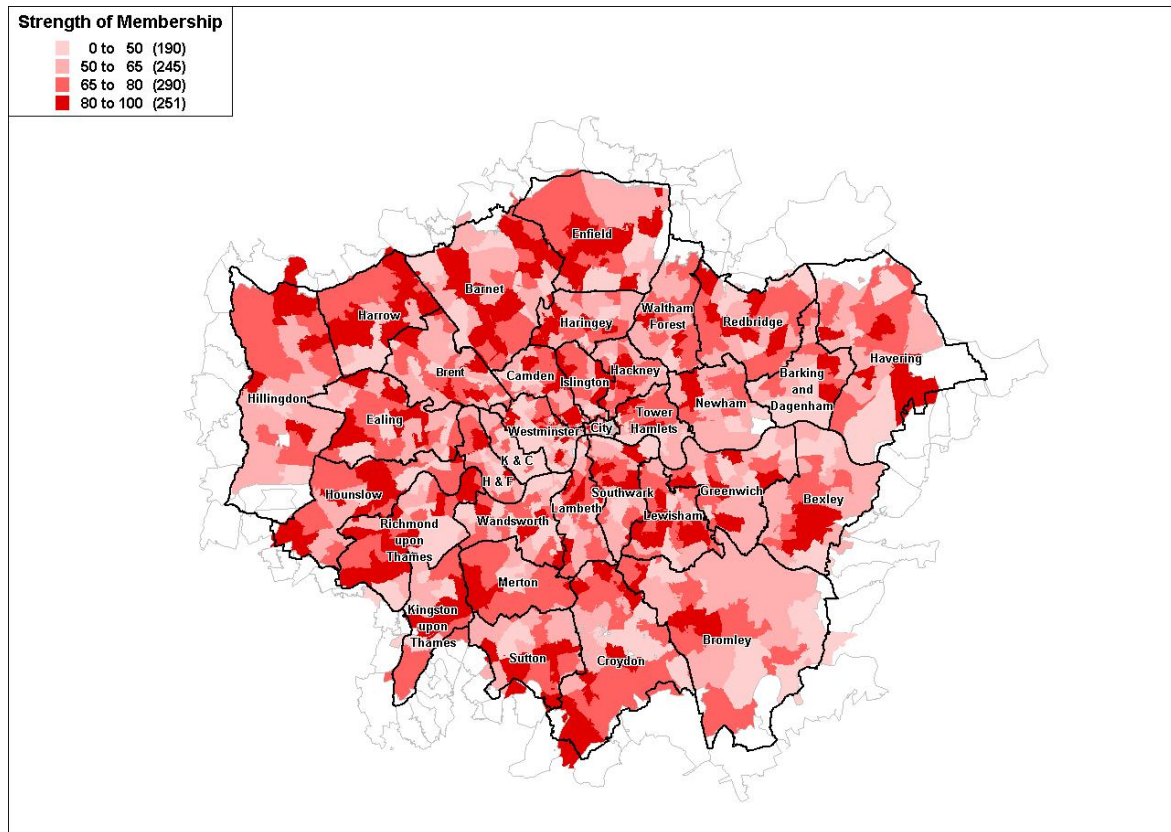
Source: GLA Economics

6   Within-group analysis

6.1   Methodology

Stepwise regression[6] was used to select variables in the hedonic regression models[7] that were analysed. The advantage of this approach is that it makes use of the data to determine which variables most powerfully describe house prices rather than taking an *a priori* view as to which variables should be included.

A log-linear functional form was used for the hedonic regression modelling.

The hedonic log-linear form is given by the following equation:

$$\ln HP_i = \alpha_0 + \sum_{j}^{n} \beta_j X_{ij} + e_i$$

Where HP_i is the ith observation of the average House price
α_0 is the intercept
β_j is the coefficient of the jth explanatory variable

---

[6] The method used was backwards elimination at the 1% significance level.
[7] Hedonic modelling is based on the proposition that the price of a house is based on various attributes such as geographic location, socio-economic characteristics and accessibility to services.

X_{ij} is the ith observation of the jth variables used to explain the variations in house prices
e_i is the error term of the ith observation.

In Working Paper 7 the equations were restricted so that only five explanatory variables could appear in each of the final models. This restriction has been relaxed, so that all statistically significant variables are included.

6.2 Regression results

Tables 4 to 7 show the full regression results within each of the groups. The signs on the coefficients are expected to reflect people's preferences if the technique is working well given that a hedonic pricing model is used. For example, as it has generally been established that people prefer living in low crime areas, other things being equal, negative coefficients would be expected to be observed on the crime variables.

Whilst the majority of the signs observed are consistent with established knowledge of people's preferences, a significant number of the observed signs differ from that which would be expected. A few examples of this phenomenon are discussed in the following section.

Table 4: Regression results for Central Group

Central Group				
Dependent Variable			**N**	198
Log House Prices			**R squared**	0.49
	Coef.	**Std. Err.**	**T**	**P>|t|**
Constant	14.40	0.48	30.33	0.00
Percentage Overcrowded	-0.02	0.00	-5.65	0.00
Theft and Handling Crimes	0.01	0.00	5.02	0.00
Percentage Social Rented	-0.01	0.00	-4.11	0.00
Criminal Damage Crimes	-0.08	0.02	-3.87	0.00
Population Density	0.00	0.00	3.33	0.00
Percentage Economically Active	-0.02	0.01	-3.16	0.00
Percentage Self Employed	0.02	0.01	3.04	0.00
Percentage Couples with dep. child.	0.03	0.01	2.82	0.00

Table 5: Regression results for High Density Housing Group

High Density Housing Group				
Dependent Variable			**N**	217
Log House Prices			**R squared**	0.84
	Coef.	**Std. Err.**	**T**	**P>|t|**
Constant	8.17	0.62	13.28	0.00
Percentage Overcrowded	0.04	0.01	7.25	0.00
Average Rooms per Household	0.79	0.12	6.58	0.00
Percentage Self Employed	0.06	0.01	5.47	0.00
Claimant Rate	0.02	0.00	4.89	0.00
Percentage White	0.01	0.00	4.88	0.00
Percentage Terraced	-0.01	0.00	-4.77	0.00
Percentage Part Time	-0.08	0.02	-4.20	0.00
Population Density	0.00	0.00	4.09	0.00
Average Income	0.00	0.00	3.83	0.00
Average Household Size	-0.63	0.20	-3.13	0.00
Percentage Couples with dep. child.	0.02	0.01	2.61	0.01

Table 6: Regression results for Outer City Group

Outer City Group				
Dependent Variable			**N**	258
Log House Prices			**R squared**	0.88
	Coef.	**Std. Err.**	**T**	**P>|t|**
Constant	9.94	0.52	18.98	0.00
Percentage Self Employed	0.04	0.01	6.00	0.00
Average Income	0.00	0.00	5.42	0.00
Average Rooms per Household	0.35	0.07	5.32	0.00
Percentage Overcrowded	0.02	0.00	5.03	0.00
Percentage White	0.01	0.00	4.97	0.00
Percentage Detached	-0.01	0.00	-4.54	0.00
Percentage Part Time	-0.05	0.01	-4.49	0.00
Percentage Flats and Maisonettes	0.01	0.00	3.97	0.00
Percentage Economically Active	-0.01	0.00	-3.31	0.00
Primary Schools within 2 km	0.01	0.00	3.08	0.00
Percentage Terraced	-0.00	0.00	-2.86	0.00

Table 7: Regression results for Spacious Suburbs Group

Spacious Suburbs Group				
Dependent Variable			**N**	303
Log House Prices			**R squared**	0.92
	Coef.	**Std. Err.**	**T**	**P>|t|**
Constant	12.50	0.45	27.89	0.00
Percentage Self Employed	0.05	0.00	10.65	0.00
Percentage Economically Active	-0.03	0.00	-9.09	0.00
Population Density	0.00	0.00	6.97	0.00
Percentage Private Rented	0.02	0.00	6.68	0.00
Average Rooms per Household	0.27	0.04	6.29	0.00
Percentage Terraced	0.01	0.00	5.91	0.00
Percentage Flats and Maisonettes	0.01	0.00	5.74	0.00
Percentage Semi-detached	0.01	0.00	5.33	0.00
Average Household Size	-0.68	0.14	-5.03	0.00
Claimant Rate	-0.04	0.01	-4.44	0.00
Average Income	0.00	0.00	4.16	0.00
Percentage Part Time	-0.03	0.01	-3.16	0.00
Percentage Couples with dep. child.	0.02	0.01	3.07	0.00
Percentage Social Rented	0.01	0.00	2.74	0.01

7 Discussion of results

7.1 Discussion

This section deals with those results that are unexpected in the context of a hedonic pricing model. A number of possible explanations for these results are put forward. For instance, in the case of the percentage of overcrowded dwellings, it could be strongly argued that, other things being equal, people dislike living in such dwellings. Thus we would expect to observe a negative sign on the overcrowding variable where it appears in the equations. Although this is the case in the Central Group, a positive sign is observed in the High Density Housing Group and the Outer City Group. This phenomenon may be due to house prices being high in areas with high overcrowding as a result of an unobserved variable such as the extent to which an area is desirable, as opposed to overcrowding driving house prices. The positive signs observed on overcrowding indicate that the hedonic regression approach may indicate a poorly specified model. Also important to note that the official definition of overcrowding is more than 1.5 persons living per habitable room which therefore assumes that couples co-habiting in a one bedroom flat would be overcrowded.

Another result that requires explanation is the sign on population density. In principle, population density could have a positive or negative effect on house prices. In the less accessible areas people may prefer to live in areas that are denser if these areas have better accessibility to services. There may also be a 'buzz' or greater diversity of services associated with living in close proximity to a lot of people. Conversely people may dislike high density housing as there may be greater noise and less open space per head of population.

Population density has a positive association with house prices in some of the groups. We do not, however, conclude that this means that people necessarily prefer living in denser areas as this result may be caused by similar reasons as were discussed in the context of overcrowding. For example, in the High Density Housing Group many expensive houses are located in boroughs such as Islington which also have high population densities. When choosing to live in these areas, for example due to the desirable nature of parts of Islington, it may be difficult to live in less dense areas or in dwellings that are not crowded. In addition to this, the private rented market may also have an impact on houses being more expensive in areas where population is particularly dense, if landlords can extract higher rents in these areas.

Another variable that it is worth considering is the proportion of detached dwellings. We would expect house prices to be higher where there are a higher proportion of detached dwellings, as such dwellings tend to sell for more money than other dwellings. The opposite result is observed in the Spacious Suburbs Group, however, with the other types of dwellings (flats, terraced, semi-detached) all having a positive coefficient on house prices (indicating that detached dwellings have a negative effect relative to the other categories). One possible explanation for this is that there are fewer sales of detached houses and this will therefore affect the average house prices in the relevant postcode sector.

The income support claimant rate is positively related to house prices in the High Density Housing Group. This result will arise due to the particular circumstances of Inner London such as the construction of post-war estates in bomb-damaged or former industrial areas and the conversion of many street properties to social housing in the 1970s, which have led to some deprived areas being located close to very rich areas and hence high house prices. Given the high proportion of social rented dwellings in this group it would be wrong to conclude that high levels of income support claimants attract people to an area or that such claimants are able to bid up house prices.

Finally, the percentage economically active has a negative association with house prices in a couple of the groups. This effect may be linked to the definition of economic inactivity, which covers those individuals are who are not in work and not looking for it, including retired individuals, some students, those on incapacity benefit and those who are sufficiently wealthy not needing to work (or claiming job seeker's allowance whilst looking). This definition means that it is not obvious whether house prices would increase or decrease as economic activity increases.

7.2   Conclusions

Whilst many of the results are consistent with what is commonly assumed about people's preferences, the significant number of unexplained results indicates that further investigation is required. The most likely cause of this issue is that an unobserved variable such as desirability is omitted from the regressions. Furthermore, the causal link between the explanatory variables and house prices may not exist as specified.

This analysis has identified four distinct housing sub-groups in London, based on various housing attributes. Such analysis is appropriate as the lack of substitutability among houses

in different areas of London means that it is inappropriate to analyse the market for housing in London as a whole. This approach is useful as it provides local information without the need to have specific knowledge of a local area, and it allows for further analysis to be conducted of the behaviour of the markets in each of the groups.

The analysis of the groups showed that the identified sub-groups had different characteristics, and hedonic stepwise regression modelling was used to identify the key drivers of house prices within the identified sub-groups. Some of the regression results suggested that the hedonic modelling may not be working well due to certain statistical problems inherent in such models. This highlights that, whilst the within-group regression analysis is useful, we need to explore the variation within groups further to provide a better understanding.

8   Appendix

8.1   Definition of variables used in the study

Table 8: Variables used in the study

No.	Variable	Description	Source
1	House price	Average of all sales 2002-2006	Hometrack
2	Average income	Average Income (weighted average by mean of each pool with £100k+ assumed at £120k)	CACI
3	Income support claimants %	Claimant Rate, percentage (Claimants* per Population)	DWP
4	Crime: criminal damage	Number of criminal damage offences	Metropolitan Police Authority
5	Crime: drugs	Number of drugs offences	Metropolitan Police Authority
6	Crime: theft and handling	Number of theft and handling offences	Metropolitan Police Authority
7	Ethnicity white %	Percentage of people in ethnic groups: White (British, Irish and other)	Census 2001, ONS
8	Economically active %	Percentage of people aged 16-74 employed part-time**, full-time**, self-employed, unemployed or full-time student	Census 2001, ONS
9	Household: detached %	Percentage of household spaces of detached type	Census 2001, ONS
10	Household: semi-detached %	Percentage of household spaces of semi-detached type	Census 2001, ONS
11	Household: terraced %	Percentage of household spaces of terraced type (including end-terrace)	Census 2001, ONS
12	Household: flats %	Percentage of household spaces that are flats, maisonettes or apartments	Census 2001, ONS
13	Household: owner-occupied %	Percentage of households: owner-occupied (owns outright, owns with a mortgage or loan, or shared ownership)	Census 2001, ONS
14	Household: social rented %	Percentage of households: social renting (housing association/registered social landlord or council/local authority)	Census 2001, ONS
15	Household: private rented %	Percentage of households: Private Rented Sector	Census 2001, ONS
16	Average household size	Average number of people per household	Census 2001, ONS

17	Average rooms/household	Average Number of rooms per household^Δ	Census 2001, ONS
18	Overcrowded households %	Percentage of households with an occupancy rating++ of −1 or less	Census 2001, ONS
19	One-person households %	Percentage of households comprising one person (pensioner or other)	Census 2001, ONS
20	Couple with dependent children households %	Percentage of households comprising couples with dependent children (married, cohabiting or others)	Census 2001, ONS
21	Distance to British Rail or London Underground station	Distance to British Rail or London Underground Station in km	Greater London Authority
22	Schools: Distance to Nearest	Distance to Nearest School	The London Pupil Dataset 2006
23	Schools: Number Primary	Number of Primary Schools within 1.5 km	The London Pupil Dataset 2006
24	Schools: Number Secondary	Number of Secondary Schools within 1.5 km	The London Pupil Dataset 2006
25	Accessibility	PTAL 2006 Score	Transport for London, GLA Economics
26	Part-time employment %	Percentage of people aged 16-74 employed part time**	Census 2001, ONS
27	Self-employed %	Percentage of people aged 16-74 who are self-employed	Census 2001, ONS
28	Population Density	Number of people per square km	Census 2001, and Mapinfo Boundaries

Notes:

Δ This variable records the number of rooms in a household space. Bathrooms, toilets, halls or landings, or rooms that can only be used for storage are not counted. All other rooms, for example, kitchens, living rooms, bedrooms, utility rooms and studies are counted. If two rooms have been converted into one, they are counted as one room. Rooms shared between a number of households, for example a shared bathroom or kitchen, are not counted. This figure is then divided by the number of residents in the household. The total number of household rooms are added together and divided by the number of residents to give average figures. Please note that this calculation is not exact as more than 12 rooms and more than eight household residents are rounded down to these figures.

* Income Support is a non-contributory benefit. From October 1996, the Jobseeker's Allowance replaced Income Support for unemployed people. In general Income Support is now only available to people who are not required to be available for work such as pensioners, lone parents, the sick and disabled people. The conditions for entitlement are in the Income Support regulations.

** For the Census, part time is defined as working 30 hours or less a week. Full time is defined as working more than 30 hours a week.

++ The occupancy rating provides a measure of under-occupancy and overcrowding. For example, a value of -1 implies that there is one room too few and that there is over-crowding in the household. The occupancy rating assumes that every household, including one-person households, require a minimum of two common rooms (excluding bathrooms).

An explanation of the postcode sector is shown in Figure 5.

Figure 5: The Postcode Sector

The Postcode Area:
One or two Alpha Characters which identify the main office through which the mail will be processed

The Sector:
A number that defines a district or neighbourhood

SW1A 1AA

The Postcode district:
One or two characters - usually numeric except in central London. These normally relate to the areas covered by Delivery Offices.

The Inward Code:
Two letters which identify the group of up to 30 addresses

Source: Land Registry

8.2   Maps of variables used in the clustering

Maps of the variables used in the clustering analysis are presented in Figures 6 to 9 below.

Figure 6: Income support claimants, percentage of all people

Income Support Claimants
Percentage of all people
15 to 42
8 to 15
5 to 8
3 to 5
0 to 3

London Borough
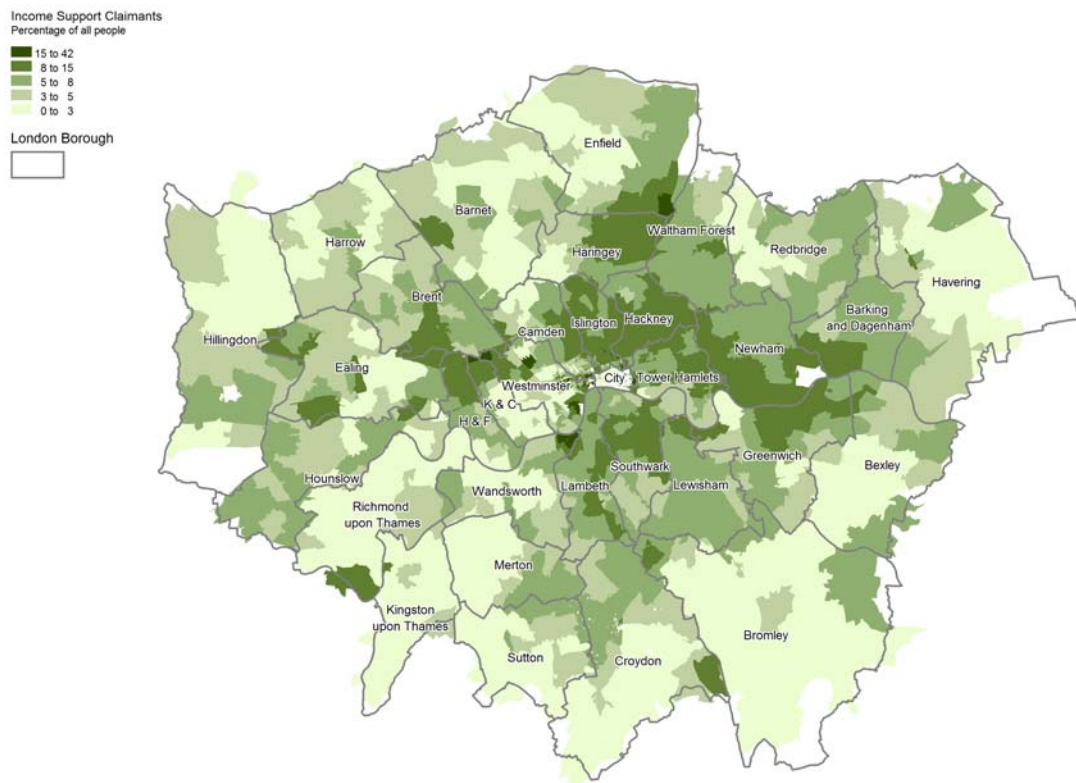
Source: DWP

Figure 7: Incidences of criminal damage

Incidences of criminal damage

- 32 to 38
- 24 to 32
- 16 to 24
- 8 to 16
- 0 to 8

London Borough
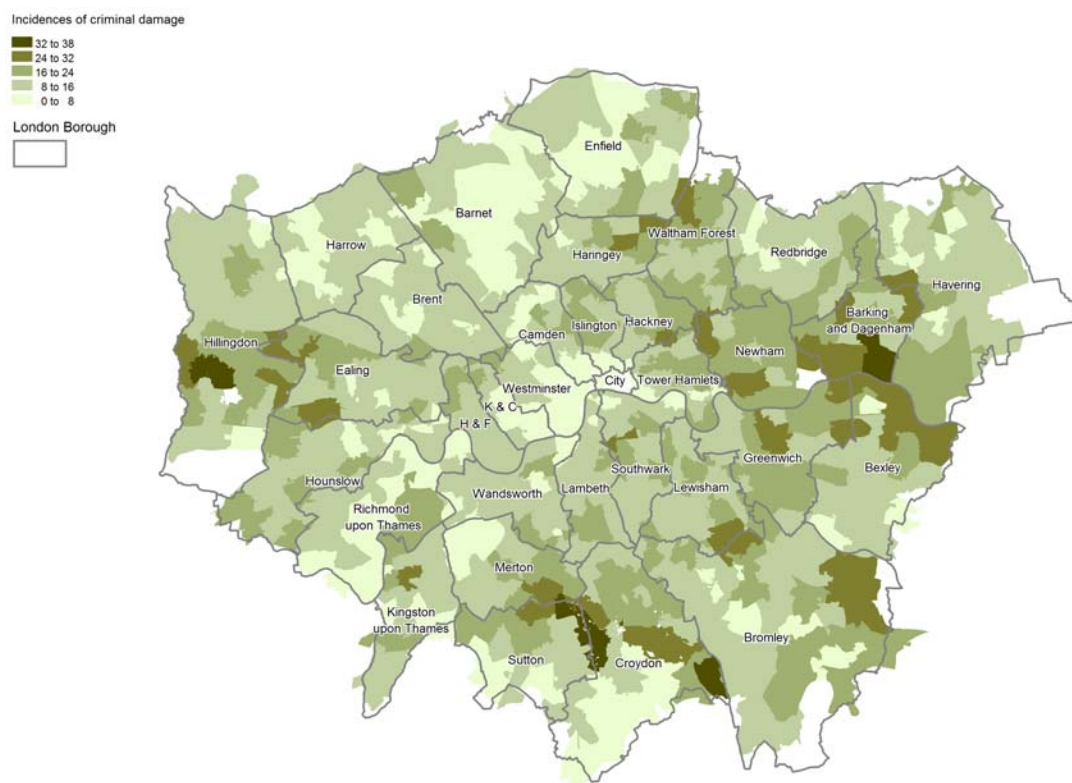
Source: MPA

Figure 8: Average rooms per household
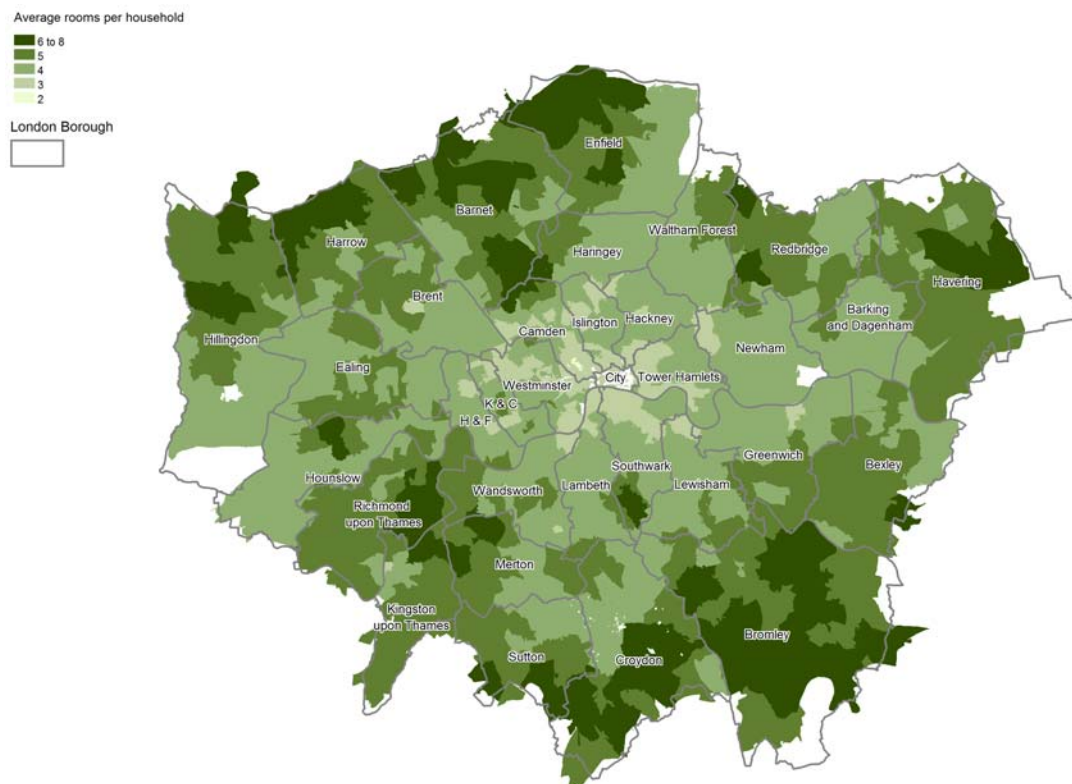
Source: Census, 2001, ONS

Figure 9: PTAL score 2006 (accessibility)
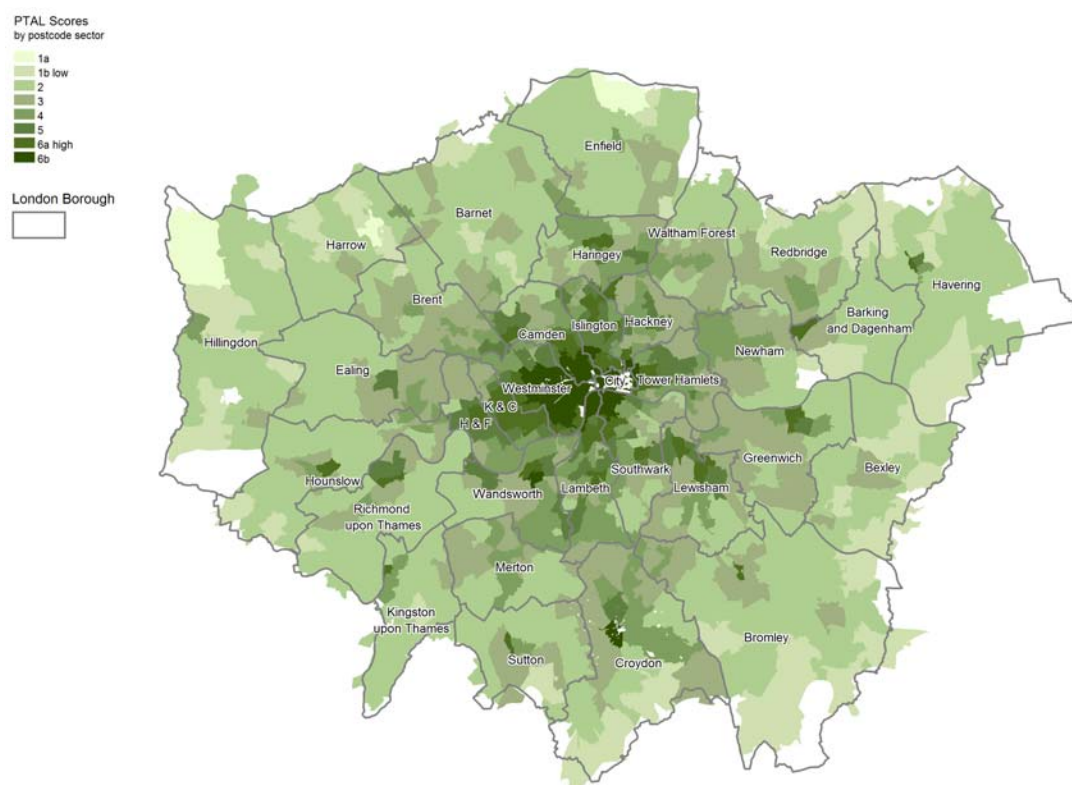
Source: TfL

The description of the level of accessibility for various PTAL scores as defined by Transport for London is set out in Table 9.

Table 9: Description of PTAL scores

PTAL	Range of Index	Description
1a (Low)	0.01 – 2.50	Very poor
1b	2.51 – 5.00	Very poor
2	5.01 – 10.00	Poor
3	10.01 – 15.00	Moderate
4	15.01 – 20.00	Good
5	20.01 – 25.00	Very Good
6a	25.01 – 40.00	Excellent
6b (High)	40.01 +	Excellent

Source: "Measuring Public Transport Accessibility Levels – PTALS – Summary", TFL, 2005

8.3 Postcode sectors by groups

Table 10: Postcode sectors within the Central Group

| Central Group								
BR1 1	CR0 1	E1 7	E1 8	EC1A 4	EC1A 7	EC1A 9	EC1M 4	EC1M 5
EC1M 6	EC1M 7	EC1N 6	EC1N 7	EC1N 8	EC1R 0	EC1R 3	EC1R 4	EC1R 5
EC1V 0	EC1V 1	EC1V 2	EC1V 3	EC1V 4	EC1V 8	EC1V 9	EC1Y 0	EC1Y 2
EC1Y 8	EC2A 1	EC2A 2	EC2A 3	EC2A 4	EC2M 1	EC2M 4	EC2V 5	EC2V 6
EC2V 8	EC2Y 5	EC2Y 8	EC2Y 9	EC3A 5	EC3N 1	EC3N 2	EC3N 4	EC3R 5
EC3R 8	EC3V 1	EC3V 3	EC3V 9	EC4A 1	EC4A 2	EC4A 3	EC4A 4	EC4M 7
EC4N 6	EC4N 7	EC4R 0	EC4R 1	EC4V 3	EC4V 5	EC4V 6	EC4Y 0	EC4Y 1
EC4Y 8	NW1 2	NW1 3	NW1 5	NW1 6	NW6 3	SE1 0	SE1 1	SE1 6
SE1 7	SE1 8	SE1 9	SW17 6	SW1A 1	SW1A 2	SW1E 5	SW1E 6	SW1H 0
SW1H 9	SW1P 1	SW1P 2	SW1P 4	SW1V 1	SW1V 2	SW1V 4	SW1W 0	SW1W 8
SW1W 9	SW1X 0	SW1X 7	SW1X 8	SW1X 9	SW1Y 4	SW1Y 5	SW1Y 6	SW3 1
SW3 3	SW5 0	SW5 9	SW7 3	SW7 4	SW7 5	W1B 1	W1B 5	W1C 1
W1C 2	W1D 3	W1D 4	W1D 5	W1D 6	W1D 7	W1F 0	W1F 7	W1F 8
W1F 9	W1G 0	W1G 6	W1G 7	W1G 8	W1G 9	W1H 1	W1H 2	W1H 4
W1H 5	W1H 7	W1J 0	W1J 5	W1J 6	W1J 7	W1J 8	W1K 1	W1K 2
W1K 3	W1K 4	W1K 5	W1K 6	W1K 7	W1S 1	W1S 4	W1T 1	W1T 2
W1T 3	W1T 4	W1T 5	W1T 6	W1U 1	W1U 2	W1U 3	W1U 4	W1U 5
W1U 6	W1U 7	W1U 8	W1W 5	W1W 6	W1W 7	W1W 8	W2 2	WC1A 1
WC1A 2	WC1B 3	WC1B 4	WC1B 5	WC1E 6	WC1E 7	WC1H 0	WC1H 8	WC1H 9
WC1N 1	WC1N 2	WC1N 3	WC1R 4	WC1V 6	WC1V 7	WC1X 0	WC1X 8	WC1X 9
WC2A 1	WC2A 3	WC2B 4	WC2B 5	WC2B 6	WC2E 7	WC2E 8	WC2E 9	WC2H 0
WC2H 7	WC2H 8	WC2H 9	WC2N 4	WC2N 5	WC2N 6	WC2R 0	WC2R 1	WC2R 3

Table 11: Postcode sectors within the High Density Housing Group

High Density Housing Group								
CR0 2	E1 0	E1 1	E1 2	E1 3	E1 4	E1 5	E1 6	E13 0
E13 8	E13 9	E14 0	E14 2	E14 4	E14 5	E14 6	E14 7	E14 8
E14 9	E15 1	E15 2	E15 3	E15 4	E16 1	E16 2	E16 3	E16 4
E17 3	E1W 3	E2 0	E2 6	E2 7	E2 8	E2 9	E3 2	E3 3
E3 4	E3 5	E5 0	E5 8	E5 9	E6 1	E6 6	E8 1	E8 2
E8 3	E8 4	E9 5	E9 6	E9 7	EC1R 1	EC1V 7	EN3 4	HA8 0
HA9 0	IG11 0	IG11 7	KT1 1	N1 0	N1 1	N1 2	N1 3	N1 4
N1 5	N1 6	N1 7	N1 8	N1 9	N15 3	N15 4	N15 5	N15 6
N16 7	N16 8	N16 9	N17 0	N17 6	N17 7	N17 8	N17 9	N18 2
N18 3	N19 3	N19 4	N19 5	N22 5	N22 6	N22 8	N4 1	N4 2
N4 3	N4 4	N5 1	N5 2	N7 0	N7 6	N7 7	N7 8	N7 9
N8 0	N8 7	N9 0	NW1 0	NW1 1	NW1 7	NW1 8	NW1 9	NW10 0
NW10 2	NW10 4	NW10 6	NW10 7	NW10 8	NW10 9	NW2 3	NW5 2	NW5 3
NW5 4	NW6 2	NW6 4	NW6 5	NW6 6	NW8 8	NW9 5	RM1 1	RM1 3
RM3 8	SE1 2	SE1 3	SE1 4	SE1 5	SE10 9	SE11 4	SE11 5	SE11 6
SE13 6	SE13 7	SE14 5	SE14 6	SE15 1	SE15 2	SE15 3	SE15 4	SE15 5
SE15 6	SE16 2	SE16 3	SE16 4	SE17 1	SE17 2	SE17 3	SE18 4	SE18 5
SE18 6	SE18 7	SE19 1	SE19 2	SE2 9	SE20 8	SE25 5	SE27 9	SE28 0
SE5 0	SE5 7	SE5 9	SE6 3	SE8 3	SE8 4	SE8 5	SM1 1	SW10 0
SW11 2	SW11 3	SW11 4	SW12 9	SW18 4	SW1P 3	SW1V 3	SW2 1	SW2 2
SW2 3	SW2 4	SW2 5	SW4 6	SW4 7	SW6 1	SW8 1	SW8 2	SW8 4
SW8 5	SW9 0	SW9 6	SW9 7	SW9 8	SW9 9	TW3 1	TW8 0	TW8 8
W10 4	W10 5	W10 6	W11 1	W11 2	W11 4	W12 0	W12 7	W12 8
W14 0	W14 9	W2 1	W2 5	W2 6	W3 8	W6 7	W6 8	W9 2
W9 3								

Table 12: Postcode sectors within the Outer City Group

Outer City Group								
BR1 4	BR1 5	BR3 1	BR5 2	BR5 3	BR5 4	CR0 0	CR0 3	CR0 4
CR0 6	CR0 9	CR2 6	CR4 1	CR4 2	CR4 3	CR4 4	CR7 7	CR7 8
DA1 4	DA14 5	DA17 5	DA17 6	DA18 4	DA8 1	DA8 2	DA8 3	E10 5
E10 6	E10 7	E11 1	E11 3	E11 4	E12 5	E12 6	E14 3	E17 4
E17 5	E17 6	E17 7	E17 8	E17 9	E4 6	E4 8	E4 9	E6 2
E6 3	E6 5	E7 0	E7 8	E7 9	EN1 1	EN1 3	EN1 4	EN3 5
EN3 6	EN3 7	HA0 1	HA0 2	HA0 4	HA1 1	HA1 2	HA1 3	HA2 0
HA2 8	HA3 5	HA3 7	HA4 6	HA8 8	IG1 1	IG1 2	IG1 3	IG1 4
IG11 8	IG11 9	IG3 8	IG6 3	IG7 4	IG8 8	KT1 2	KT1 3	KT3 3
KT9 1	N11 1	N11 2	N11 3	N12 0	N12 8	N13 6	N16 0	N16 5
N16 6	N18 1	N6 5	N8 8	N8 9	N9 7	N9 8	N9 9	NW1 4
NW10 1	NW10 3	NW10 5	NW2 1	NW2 4	NW2 5	NW2 6	NW2 7	NW3 2
NW3 3	NW3 6	NW5 1	NW6 1	NW6 7	NW8 0	NW9 6	NW9 7	RM10 7
RM10 8	RM10 9	RM13 8	RM13 9	RM3 7	RM3 9	RM6 4	RM6 5	RM6 6
RM7 0	RM8 1	RM8 2	RM8 3	RM9 4	RM9 5	RM9 6	SE10 0	SE10 8
SE12 0	SE12 8	SE12 9	SE13 5	SE16 6	SE16 7	SE18 1	SE18 2	SE18 3
SE19 3	SE2 0	SE20 7	SE21 8	SE22 0	SE22 9	SE23 1	SE23 2	SE23 3
SE24 0	SE25 4	SE25 6	SE26 4	SE26 5	SE26 6	SE27 0	SE28 8	SE3 0
SE3 8	SE3 9	SE4 1	SE4 2	SE5 8	SE6 1	SE6 2	SE6 4	SE7 7
SE7 8	SE9 2	SE9 4	SE9 5	SE9 6	SM1 4	SM4 5	SM4 6	SM5 1
SM5 2	SM6 7	SW11 5	SW12 0	SW15 3	SW15 5	SW16 1	SW16 2	SW16 3
SW16 5	SW16 6	SW17 0	SW17 7	SW17 8	SW17 9	SW18 1	SW19 2	SW19 6
SW4 0	SW4 8	SW4 9	SW6 2	SW6 5	SW6 7	SW8 3	TW13 4	TW13 5
TW13 6	TW13 7	TW14 0	TW14 8	TW14 9	TW3 3	TW3 4	TW4 5	TW4 6
TW4 7	TW5 0	TW5 9	TW7 6	TW7 7	TW8 9	UB1 1	UB1 2	UB1 3
UB10 0	UB2 4	UB2 5	UB3 1	UB3 2	UB3 3	UB3 4	UB4 0	UB4 8
UB4 9	UB5 4	UB5 5	UB5 6	UB6 8	UB6 9	UB7 7	UB7 8	UB7 9
UB8 2	UB8 3	W12 9	W13 0	W14 8	W3 6	W3 7	W3 9	W4 2
W5 2	W6 0	W6 9	W7 1	W7 3	W9 1			

Table 13: Postcode sectors within the Spacious Suburbs Group

Spacious Suburbs Group								
BR1 2	BR1 3	BR2 0	BR2 6	BR2 7	BR2 8	BR2 9	BR3 3	BR3 4
BR3 5	BR3 6	BR4 0	BR4 9	BR5 1	BR6 0	BR6 6	BR6 7	BR6 8
BR6 9	BR7 5	BR7 6	CR0 5	CR0 7	CR0 8	CR2 0	CR2 7	CR2 8
CR2 9	CR5 1	CR5 2	CR7 6	CR8 1	CR8 2	CR8 3	CR8 4	CR8 5
DA14 4	DA14 6	DA15 7	DA15 8	DA15 9	DA16 1	DA16 2	DA16 3	DA5 1
DA5 2	DA5 3	DA6 7	DA6 8	DA7 4	DA7 5	DA7 6	E11 2	E18 1
E18 2	E1W 1	E1W 2	EN1 2	EN2 0	EN2 6	EN2 7	EN2 8	EN2 9
EN4 0	EN4 8	EN4 9	EN5 1	EN5 2	EN5 3	EN5 5	HA0 3	HA1 4
HA2 6	HA2 7	HA2 9	HA3 0	HA3 6	HA3 8	HA3 9	HA4 0	HA4 7
HA4 8	HA4 9	HA5 1	HA5 2	HA5 3	HA5 4	HA5 5	HA6 1	HA6 2
HA7 1	HA7 2	HA7 3	HA7 4	HA8 5	HA8 6	HA8 7	HA8 9	HA9 6
HA9 7	HA9 8	HA9 9	IG2 6	IG2 7	IG3 9	IG4 5	IG5 0	IG6 1
IG6 2	IG8 0	IG8 7	IG8 9	KT1 4	KT2 5	KT2 6	KT2 7	KT3 4
KT3 5	KT3 6	KT4 7	KT4 8	KT5 8	KT5 9	KT6 4	KT6 6	KT6 7
KT8 9	KT9 2	N10 1	N10 2	N10 3	N12 7	N12 9	N13 4	N13 5
N14 4	N14 5	N14 6	N14 7	N2 0	N2 8	N2 9	N20 0	N20 8
N20 9	N21 1	N21 2	N21 3	N22 7	N3 1	N3 2	N3 3	N6 4
N6 6	NW11 0	NW11 6	NW11 7	NW11 8	NW11 9	NW2 2	NW3 1	NW3 4
NW3 5	NW3 7	NW4 1	NW4 2	NW4 3	NW4 4	NW7 1	NW7 2	NW7 3
NW7 4	NW8 6	NW8 7	NW8 9	NW9 0	NW9 8	NW9 9	RM1 2	RM1 4
RM11 1	RM11 2	RM11 3	RM12 4	RM12 5	RM12 6	RM13 7	RM14 1	RM14 2
RM2 5	RM2 6	RM3 0	RM5 2	RM5 3	RM7 7	RM7 8	RM7 9	SE16 5
SE21 7	SE22 8	SE24 9	SE3 7	SE9 1	SE9 3	SM1 2	SM1 3	SM2 5
SM2 6	SM2 7	SM3 8	SM3 9	SM4 4	SM5 3	SM5 4	SM6 0	SM6 8
SM6 9	SW10 9	SW11 1	SW11 6	SW12 8	SW13 0	SW13 8	SW13 9	SW14 7
SW14 8	SW15 1	SW15 2	SW15 4	SW15 6	SW16 4	SW18 2	SW18 3	SW18 5
SW19 1	SW19 3	SW19 4	SW19 5	SW19 7	SW19 8	SW20 0	SW20 8	SW20 9
SW3 2	SW3 4	SW3 5	SW3 6	SW6 3	SW6 4	SW6 6	SW7 1	SW7 2
TN16 3	TW1 1	TW1 2	TW1 3	TW1 4	TW10 5	TW10 6	TW10 7	TW11 0
TW11 8	TW11 9	TW12 1	TW12 2	TW12 3	TW2 5	TW2 6	TW2 7	TW3 2
TW7 4	TW7 5	TW9 1	TW9 2	TW9 3	TW9 4	UB10 8	UB10 9	UB3 5
UB6 0	UB6 7	UB7 0	UB8 1	UB9 6	W11 3	W13 8	W13 9	W2 3
W2 4	W3 0	W4 1	W4 3	W4 4	W4 5	W5 1	W5 3	W5 4
W5 5	W7 2	W8 4	W8 5	W8 6	W8 7			

Table 14: Postcode sectors excluded due to lack of available data

No Data								
BR8 7	BR8 8	CR3 0	CR3 5	CR5 3	CR6 9	CR9 0	CR9 1	CR9 2
CR9 3	CR9 4	CR9 5	CR9 6	CR9 7	CR9 8	DA1 3	DA1 5	DA2 7
E4 7	E6 7	EC1A 1	EC1A 2	EC1M 3	EC1N 2	EC1Y 1	EC1Y 4	EC2M 2
EC2M 3	EC2M 5	EC2M 6	EC2M 7	EC2N 1	EC2N 2	EC2N 3	EC2N 4	EC2P 2
EC2R 5	EC2R 6	EC2R 7	EC2R 8	EC2V 7	EC3A 1	EC3A 2	EC3A 3	EC3A 4
EC3A 6	EC3A 7	EC3A 8	EC3M 1	EC3M 2	EC3M 3	EC3M 4	EC3M 5	EC3M 6
EC3M 7	EC3M 8	EC3N 3	EC3P 3	EC3R 6	EC3R 7	EC3V 0	EC3V 4	EC4M 6
EC4M 8	EC4M 9	EC4N 1	EC4N 4	EC4N 5	EC4N 8	EC4P 4	EC4R 2	EC4R 3
EC4R 9	EC4V 2	EC4V 4	EC4Y 7	EC4Y 9	EN5 4	EN6 2	EN6 4	EN6 5
EN7 5	EN8 7	EN8 8	EN9 1	EN9 3	HA6 3	IG10 4	IG7 5	IG7 6
IG9 5	IG9 6	KT10 0	KT17 2	KT17 3	KT18 7	KT19 0	KT19 7	KT19 8
KT19 9	KT21 1	KT21 2	KT22 0	KT6 5	KT7 0	KT8 1	RM14 3	RM15 4
RM15 5	RM15 6	RM19 1	RM4 1	SL0 0	SL0 9	SL3 0	SM7 1	SM7 3
SW1A 0	TN14 7	TN16 2	TW15 1	TW15 2	TW16 5	TW16 7	TW19 6	TW19 7
TW6 1	TW6 2	TW6 3	UB11 1	UB9 4	UB9 5	W1A 3	W1B 2	W1B 3
W1B 4	W1D 1	W1D 2	W1H 6	W1J 9	W1S 2	W1S 3	W1T 7	WC1R 5
WC2A 2	WC2R 2	WD19 6	WD23 1	WD3 1	WD3 8	WD3 9	WD6 1	WD6 2
WD6 3								

8.4 Summary statistics by group

Table 15 sets out the average value of each variable by identified group. Lowest values per row are given in red and highest values in purple.

Table 15: Average value[8] by group

	Central	High Density Housing	Outer City	Spacious Suburbs
Claimant Rate	6.13	**8.92**	5.38	**2.56**
Criminal Damage Crimes	**1.91**	14.20	**16.26**	9.37
Average Household Size	**1.75**	2.28	**2.45**	2.44
PTALs 2006	**58.54**	19.10	10.84	**10.19**
Average Income (£)	**41,784**	**35,220**	35,754	41,405
Percentage 5 A*-C GCSE or equivalent	53.99	**42.25**	48.87	**65.20**
Drugs Crimes	**1.37**	**5.87**	3.49	1.45

---

[8] In respect of the variables used in the probabilistic logic clustering, the average values will differ from the core values obtained by the probabilistic logic clustering technique. This is because the core values are determined by an optimisation process that locates the four clusters that best describe the data, whereas the average values take the mean of the values for the postcodes sectors that fall within a particular cluster.

Theft and Handling Crimes	**21.85**	**44.54**	36.49	23.50
Percentage White	79.65	**64.16**	71.15	**82.38**
Percentage Economically Active	**70.79**	**64.75**	68.02	69.12
Percentage Detached	**0.81**	2.19	5.86	**17.72**
Percentage Semi-detached	**1.16**	6.79	22.98	**30.98**
Percentage Terraced	**5.06**	22.48	**32.62**	19.96
Percentage Flats and Maisonettes	**93.04**	68.31	38.22	**31.12**
Percentage Owner Occupied	**30.25**	36.07	61.34	**73.81**
Percentage Social Rented	23.17	**43.76**	22.30	**10.09**
Percentage Private Rented	**46.58**	20.16	16.36	**16.10**
Average Rooms per Household	**3.68**	4.16	4.81	**5.51**
Percentage Overcrowded	**28.56**	25.49	15.39	**9.36**
Percentage One Person Households	**54.46**	39.23	31.49	**29.59**
Percentage Couples with Dependent Children	**8.18**	17.85	23.09	**24.84**
Distance to BR or LU Station in km	**0.30**	0.53	**0.91**	0.90
Distance to Schools	**0.80**	**0.64**	0.69	0.79
Primary Schools within 2 km	7.88	**9.18**	6.21	**5.97**
Secondary Schools within 2 km	**0.10**	0.58	**0.92**	0.81
Percentage Part Time	**4.71**	7.19	8.75	**9.34**
Percentage Self Employed	**12.87**	**7.97**	8.52	12.06
Population Density (People per sq. km.)	6,873	**9,701**	6,216	**4,754**

8.5   Probabilistic logic clustering mathematical details

We start with a data set X consisting of n observations, where each observation is a vector in d dimensions. The aim of probabilistic logic clustering is to divide the data into c clusters, where c can be between 2 and n. The divisions should be such that within the clusters the data have similar characteristics and the average difference between cluster characteristics is maximised.

$$X = \{x_1, x_2, \ldots, x_n\} \qquad x_j \in R^d \tag{1}$$

The attributability of observation x_j to cluster k is u_{kj}. With classical clustering u_{kj} can only take the value 0 or 1, but with probabilistic logic clustering it can take any value between 0 and 1.

$$u_{kj} \in \{0, 1\} \qquad \text{Classical clustering} \tag{2}$$

$$u_{kj} \in [0, 1] \qquad \text{Probabilistic logic clustering} \tag{3}$$

However for each type of clustering we still have the condition:

$$\sum_j u_{kj} > 0 \qquad \sum_k u_{kj} = 1 \tag{4 \& 5}$$

The objective function, whose size is to be minimised for an optimal solution is:

$$J_m(U, v) = \sum_{j=1}^{n} \sum_{k=1}^{c} (u_{kj})^m \|x_j - v_k\|^2 \qquad 1 \leq m < \infty \tag{6}$$

In this equation we have U, the matrix of memberships and v_k, the centre of cluster k. The variable m determines the type of clustering that is done. When $m = 1$ and $u_{kj} \in \{0, 1\}$ the minimalization of (6) is what is known as ordinary k-means. When m takes a value greater than 1 and $u_{kj} \in [0, 1]$ we have probabilistic logic clustering. In this case the values of u_{kj} and v_k that minimise (6) are:

$$\hat{u}_{kj} = \left(\sum_{g=1}^{c} \left(\frac{\|x_j - \hat{v}_k\|}{\|x_j - \hat{v}_g\|} \right)^{\frac{2}{(m-1)}} \right)^{-1} \qquad \forall j, k \tag{7}$$

$$\hat{v}_k = \frac{\sum_{j=1}^{n} (\hat{u}_{kj})^m x_j}{\sum_{j=1}^{n} (\hat{u}_{kj})^m} \qquad \forall k \tag{8}$$

As the centres of the clusters are not known before the clustering process, the memberships can not be calculated directly, and an iterative process has to be used. The optimal u_{kj} can be found by repeating the following process:

(i) m and cluster number c are assumed, and a norm in equation (6) is defined appropriately (for our purposes, the standard Euclidean norm). In addition, an initial value $U^{(0)} \in M_{fc}$ is set for U (where M_{fc} is the space satisfying the above conditions (3), (4) and (5)). The value can be taken at random.

(ii)   The cluster centre $v_k^{(0)}$ is calculated using $U^{(0)}$ and equation (8)

(iii) $U^{(1)}$ is calculated using $v_k^{(0)}$ and equation (7)

(iv)   Defining an appropriate norm and threshold value ε, the preceding steps are repeated until

$$\|U^{(p)} - U^{(p-1)}\| \leq \epsilon$$

When the inequality in step (iv) is satisfied, we are left with the c optimal cluster centres, $v_k^{(p)}$, whose memberships $U^{(p)}$ are given by equation (7).

8.6   Analysis of membership

As mentioned earlier, the probabilistic logic clustering technique used in this working paper assigns each postcode sector a percentage membership in each of the four groups identified. Each postcode sector is then assigned to the group that it has the highest membership in.
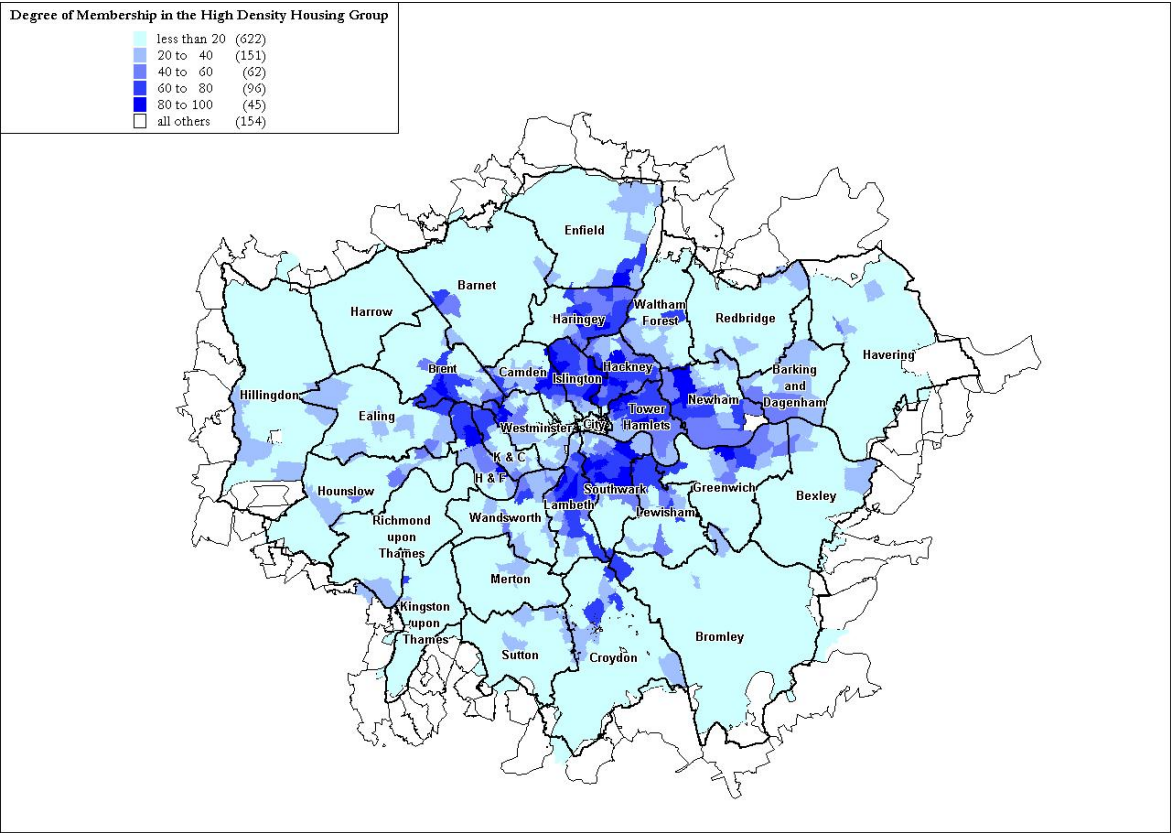
Figures 10 to 13 show the degree of membership in each of the groups for those postcode sectors that are assigned to each particular group.
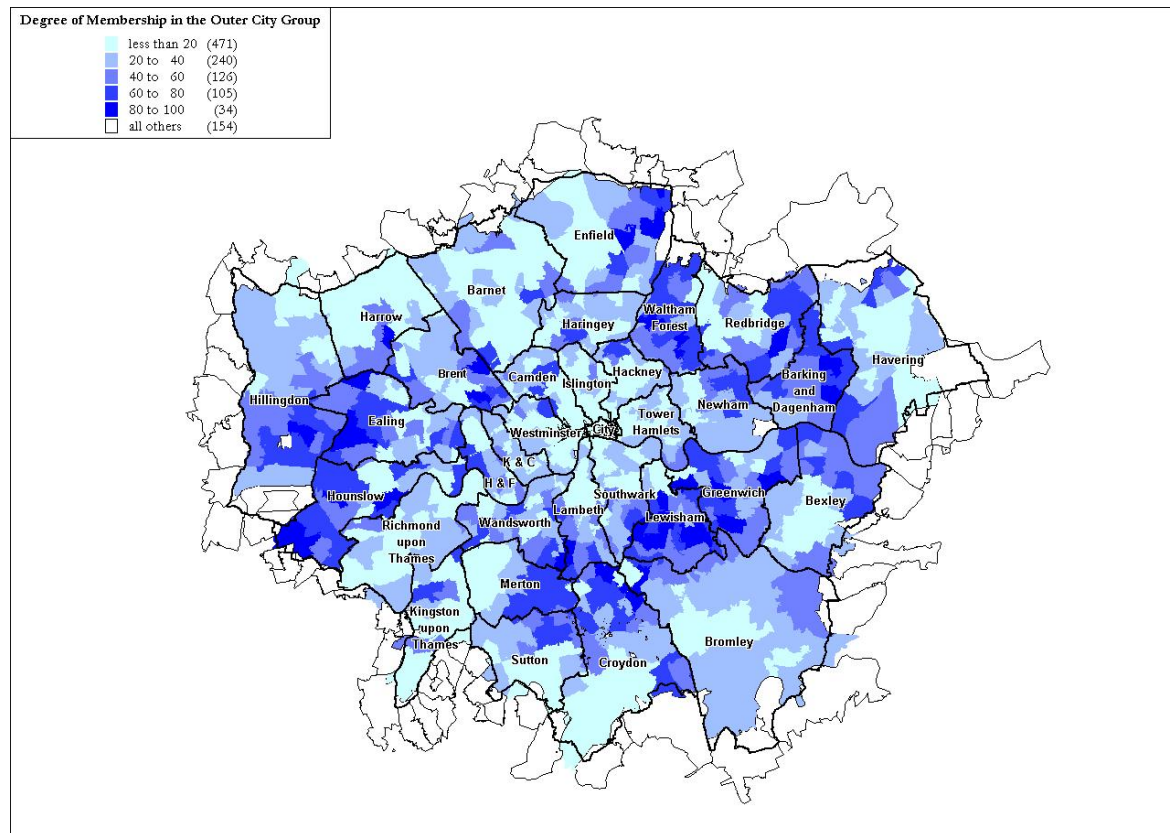
Figure 10: Degree of membership in the Central Group
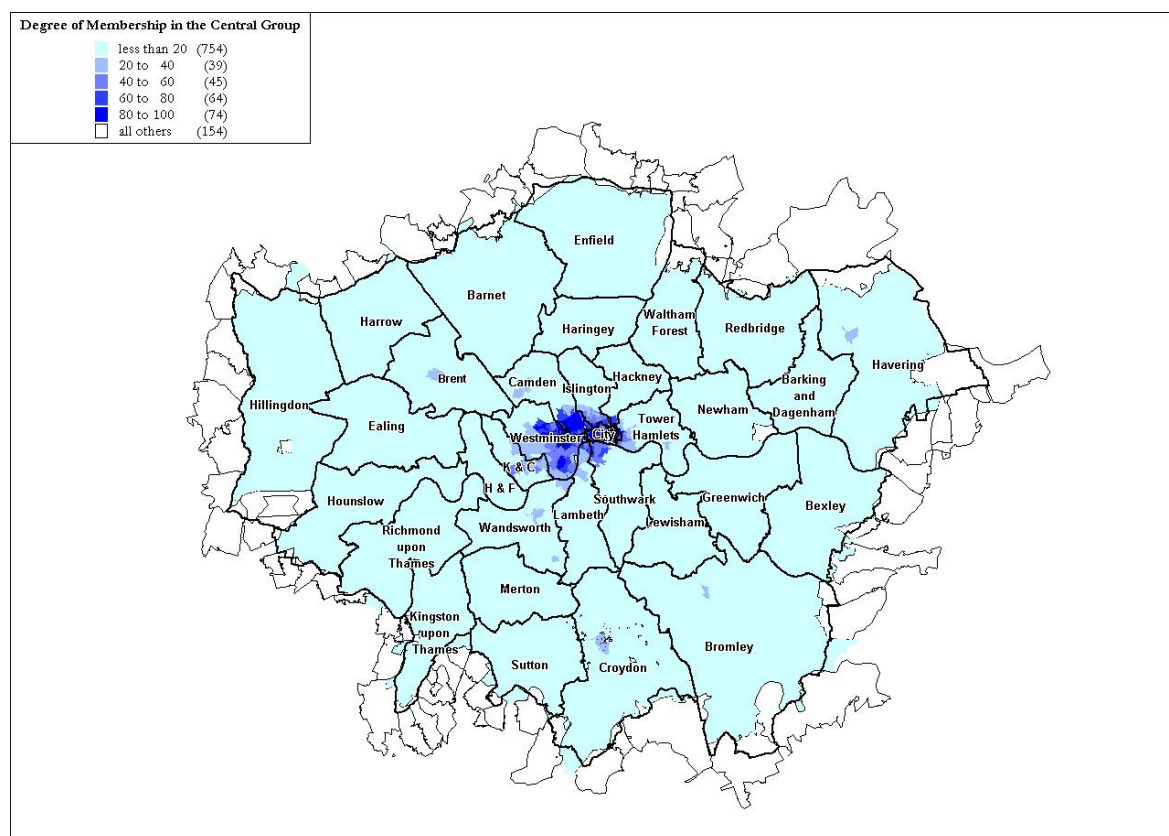
Source: GLA Economics

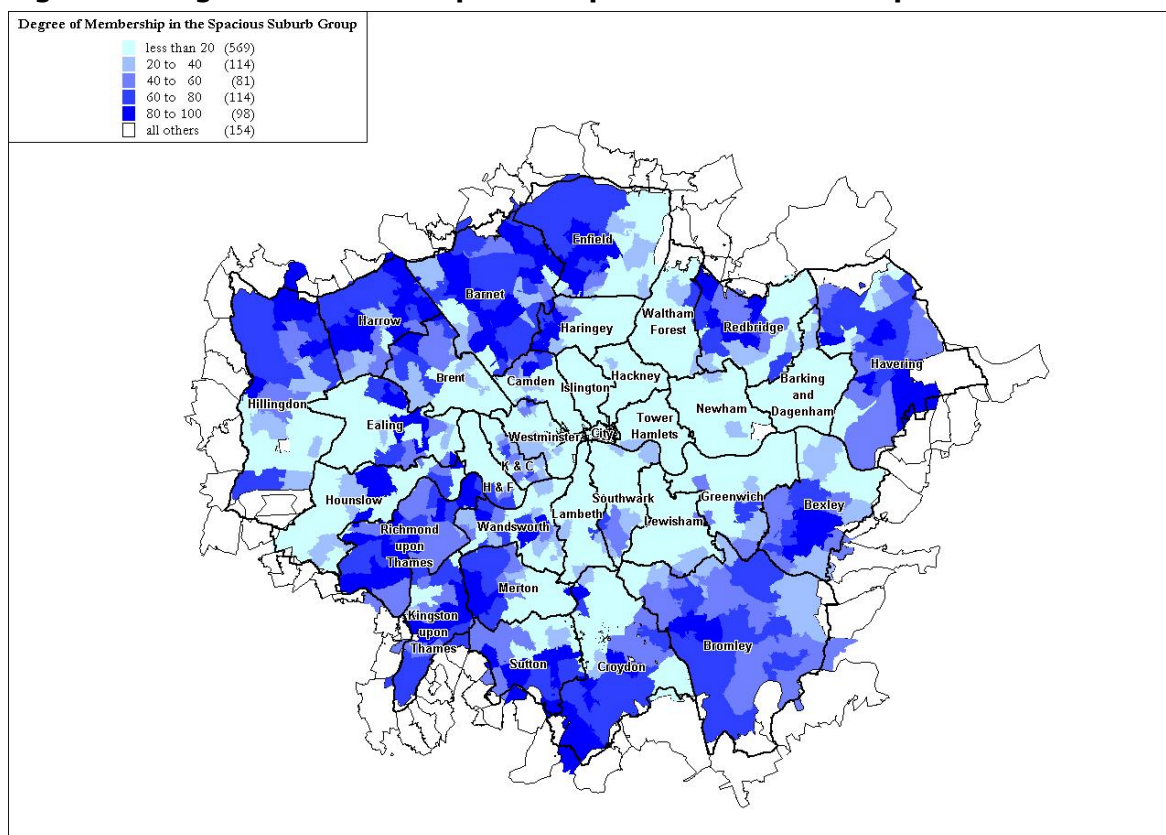## Figure 11: Degree of membership in the High Density Housing Group

Source: GLA Economics

Figure 12: Degree of membership in the Outer City Group

Source: GLA Economics

Figure 13: Degree of membership in the Spacious Suburbs Group

Source: GLA Economics

Figures 10-13 show that the groups are generally well defined with the majority of the postcode sectors in each group having a degree of membership of over 50 per cent. This visual analysis is useful because it highlights those postcode sectors that are particularly well described by their particular group, and those that are somewhat less well described. For example, in the Central Group, the postcode sectors in Camden, Kensington and Chelsea, Wandsworth and Croydon are less well described by the group than postcode sectors in the heart of London, as is to be expected.

The relationship between the different groups is also considered, i.e. the correlation of the membership of each group with the other groups. The results are shown in Table 16. This shows that the most negatively correlated groups are High Density Housing and Spacious Suburbs.

Table 16: Matrix of correlation coefficients for the degrees of membership

	Central	Spacious Suburbs	Outer City	High Density Housing
Central	1.00			
Spacious Suburbs	-0.38	1.00		
Outer City	-0.46	-0.26	1.00	
High Density Housing	-0.24	-0.51	-0.12	1.00

Other formats and languages

For a large print, Braille, disc, sign language video or audio-tape version of this document, please contact us at the address below:

Public Liaison Unit
Greater London Authority
City Hall, The Queen's Walk
London SE1 2AA

Telephone 020 7983 4100
Minicom 020 7983 4458
www.london.gov.uk

You will need to supply your name, your postal address and state the format and title of the publication you require. If you would like a copy of this document in your language, please phone the number or contact us at the address above.

Chinese

如果需要您母語版本的此文件，
請致電以下號碼或與下列地址聯絡

Vietnamese

Nếu bạn muốn có văn bản tài liệu này bằng ngôn ngữ của mình, hãy liên hệ theo số điện thoại hoặc địa chỉ dưới đây.

Greek

Αν θέλετε να αποκτήσετε αντίγραφο του παρόντος εγγράφου στη δική σας γλώσσα, παρακαλείστε να επικοινωνήσετε τηλεφωνικά στον αριθμό αυτό ή ταχυδρομικά στην παρακάτω διεύθυνση.

Turkish

Bu belgenin kendi dilinizde hazırlanmış bir nüshasını edinmek için, lütfen aşağıdaki telefon numarasını arayınız veya adrese başvurunuz.

Punjabi

ਜੇ ਤੁਹਾਨੂੰ ਇਸ ਦਸਤਾਵੇਜ਼ ਦੀ ਕਾਪੀ ਤੁਹਾਡੀ ਆਪਣੀ ਭਾਸ਼ਾ ਵਿਚ ਚਾਹੀਦੀ ਹੈ, ਤਾਂ ਹੇਠ ਲਿਖੇ ਨੰਬਰ 'ਤੇ ਫ਼ੋਨ ਕਰੋ ਜਾਂ ਹੇਠ ਲਿਖੇ ਪਤੇ 'ਤੇ ਰਾਬਤਾ ਕਰੋ:

Hindi

यदि आप इस दस्तावेज की प्रति अपनी भाषा में चाहते हैं, तो कृपया निम्नलिखित नंबर पर फोन करें अथवा नीचे दिये गये पते पर संपर्क करें

Bengali

আপনি যদি আপনার ভাষায় এই দলিলের প্রতিলিপি (কপি) চান, তা হলে নীচের ফোন্ নম্বরে বা ঠিকানায় অনুগ্রহ করে যোগাযোগ করুন।

Urdu

اگر آپ اِس دستاویز کی نقل اپنی زبان میں چاہتے ہیں، تو براہ کرم نیچے دئے گئے نمبر پر فون کریں یا دیئے گئے پتے پر رابطه کریں

Arabic

إذا أردت نسخة من هذه الوثيقة بلغتك، يرجى الاتصال برقم الهاتف أو مراسلة العنوان أدناه

Gujarati

જો તમને આ દસ્તાવેજની નકલ તમારી ભાષામાં જોઇતી હોય તો, કૃપા કરી આપેલ નંબર ઉપર ફોન કરો અથવા નીચેના સરનામે સંપર્ક સાધો.