

Metropolitan Police Service – Neighbourhood groupings

Grouping methodology

May 2014

Introduction

The aim of this summary is to show the steps involved in creating the Metropolitan Police Service (MPS) neighbourhood most similar groups. The process places each of the 107 neighbourhoods¹ into a group with those other neighbourhoods to which it is most similar in terms of chosen indicators. This enables similar areas to be classified according to their particular combination of characteristics.

The methodology closely follows that designed by University College London (UCL), in their work on the Office for National Statistics (ONS) Output Area and London Output Area Classifications (OAc/LOAc). Further information can be found [here](#).

Choice of variables

Initial variables were chosen from 2 key sources – the [London Datastore Ward Atlas](#) (128 variables) and the [2011 Census](#). A specific set of representative datasets had already been selected from the 2011 Census by UCL in conjunction with ONS as part of their OAc project, narrowing the variable selection from 167 to 60. After consultation with GLA census experts, these datasets were deemed suitable for this piece of work also. This process is documented in their [methodology](#).

As outlined in the [Geographical Alignment document](#) all neighbourhoods were aligned with existing Ordnance Survey electoral ward boundaries, so that the ward data from the sources above could be simply aggregated up to a neighbourhood level.

A total of 79 datasets were then removed from the Ward Atlas source that were either duplicates², crime-related data (not to be used in profiling the neighbourhoods), or not relevant for grouping work. A list of all datasets and the decisions made at each stage of the process here and below is shown in Appendix A, giving an initial total of 109 datasets (60+49).

Format, shape and scale considerations

There are a number of changes that needed to be made to the datasets to prepare them for any use in a grouping methodology. Any number or combination of these changes might help towards the objective of

¹ *Westminster – ORB* neighbourhood is merged into *Westminster – West End* for the purposes of grouping. See the [Geographical Alignment document](#) for more information.

² Census 2011 datasets took precedence over Ward Atlas datasets to maintain alignment with ONS/UCL methodology.

better realising the variation within the data, and therefore must all be considered as potential options for input into the final grouping process.

Firstly consideration needed to be given to the suitable formats for the raw data. Three main options were considered:

- 1) the **percentage** that each neighbourhood value (the numerator) accounts for within a relevant larger dataset for that neighbourhood (the denominator)

$$\text{"Percentage"} = \frac{\text{Numerator}}{\text{Denominator}} \times 100$$

eg. the percentage of 10-18 years olds in a neighbourhood within a count of all ages in that neighbourhood;

NB. A few variables, such as those relating to area and population density could not be converted into percentages and were left unchanged.

- 2) an **index score** where the percentage calculated above is made comparable across all neighbourhoods by dividing by the percentage of the denominator total for which the neighbourhood denominator accounts.

$$\text{"Index Score"} = \frac{\text{"Percentage"}}{\left[\frac{\text{Denominator}}{\text{Denominator Sum}} \times 100 \right]} \times 100$$

eg. the percentage of 10-18 years olds in a neighbourhood within a count of all ages in that neighbourhood, divided by the percentage of all ages that that neighbourhood has compared to the London total.

- 3) a calculation of **difference from the mean** where a mean value is calculated through the proportion of the numerator and denominator at London level multiplied by the neighbourhood denominator, and then subtracted from the neighbourhood numerator value.

$$\text{"Mean Difference"} = \text{Numerator} - \left[\left[\frac{\text{Numerator Sum}}{\text{Denominator Sum}} \right] \times \text{Denominator} \right]$$

eg. the difference between the number of 10-18 year olds in a neighbourhood and an average that takes into account the London proportion of 10-18 year olds in the wider population and the overall population in that neighbourhood.

Next, issues of 'shape' were considered. Grouping methods work best with data that is normally distributed, that is where data tends to be evenly distributed around a central value with no bias left or right, often called a 'bell-shaped curve' e.g. peoples heights. Skewed distributions are however experienced regularly when considering the diverse geographical areas of London where outliers and non-normal distributions are to be expected; in particular in the high tourist / low residence areas of, for example, Westminster - West End. Three different 'normalisation' techniques were therefore applied separately to the data to try to lessen

the impact of any outliers and thus 'normalise' the distribution – these being a BoxCox transformation, the application of a Natural Logarithm, and the use of the Inverse-Hyperbolic Sine (IHS). NB. Each piece of data had a value of one added to it prior to normalisation to remove any errors resulting from the normalisation of zero values.

Finally consideration was given to the 'scale' of the datasets. All grouping techniques are based on the similarity or dissimilarity of the cases (neighbourhoods) to be grouped. This is measured by considering 'distances' between all the variables in the dataset for each neighbourhood, and clearly problems will occur if there are differing scales among the variables eg. *population density* varying between 10 and 165 persons per hectare, and *% full-time employed* varying between 0 and 100%. It was therefore necessary to ensure each variable was equally represented when measuring the 'distances' by standardising the data using 3 different standardisation techniques – Zscore, Range and Inter-Decile.

Explanations of the normalisation and standardisation techniques can be found in the UCL publication.

This resulted in the creation of 27 different datasets (3x3x3) of the 109 variables representing the different combinations of format, shape and scale outlined above. Using the open-source [R](#) software, in conjunction with programming code developed by UCL (and used in the ONS OAC), the datasets were all run through the grouping methodology (see section below) to consider their impact on the overall data variation.

Of the combinations, it was clear that the permutation of Percentage data format, Box-Cox normalisation and Range standardisation gave the most 'normal' results – this being a level of skew closest to zero whilst still recognising the presence of outliers that are key to representing London's variation and diversity.

Correlations

Academic research suggests that whilst highly correlated variables (e.g. height and weight) can make a high proportion of the component data redundant, they can also be highly predictive and descriptive for grouping classifications such as these. Continuing to use the UCL R code, but only on the Percentage-BoxCox-Range dataset, a Pearson correlation technique was run to identify any pairs of variables which correlated highly with one another. With the Census variables having already been considered for correlation by ONS, consideration was only given to either pairs of non-census datasets or for pairs of non-census and census datasets. Considering the results of running the code above, any pairs of variables with correlation coefficients greater than 0.6 and -0.6 were analysed, resulting in the removal of ten variables from the subsequent grouping analysis (see Appendix A). The final 99 variables can be seen in Appendix B.

Weightings

For this work it was decided that all variables should be weighted equally as 1. There were several reasons for this: the purpose of this work was to create a classification that where possible was not subjective, and had a transparent rationale – an aim clearly inhibited by assigning greater weight to one variable over another. It was very difficult to define whether any particular weighting had a 'positive effect' on the classification, as the classification itself has no specific thematic aim; and by being more selective in how the variables were chosen initially, and not weighting, any classification process was made far simpler.

Group counts

The grouping methodology selected required an initial direction for the number of groups required. Academic research suggests that this be based on 1) what the data suggests (a number of groups that minimises the 'within-group' variation whilst maximising the 'between-group' variation), 2) a perceived ideal scenario (the MPS suggested a number of groups less than 15), and 3) a scenario that makes sense visually to those with knowledge of the characteristics of London. This methodology varies at this stage significantly

from that of ONS/UCL as only a simple single-tier, non-hierarchical structure is required, as opposed to the three-tier classification used in the OAc. A comparison of the two methodologies is shown in flow-charts in Appendix C.

Grouping methodology

Continuing to use the UCL/ONS methodology, the 'k-means' grouping methodology was chosen. K-means is an iterative relocation algorithm whose basic premise is to move a neighbourhood from one grouping to another to see if the move would improve (lower) the variation within the group. The neighbourhood is then assigned or reallocated to the group to which it brings the greatest improvement. When all the neighbourhoods have been assigned, the next iteration starts which then repeats the process. These iterations continue until a stable classification is reached where no more allocations/moves can occur during a complete iteration of the data. Once this point is reached, it is then possible to analyse the distances or 'means' of each group for each variable to assess the distinctiveness of the groupings.

In order to obtain the most stable set of classifications the R code was used to carry out an academically-suggested 10,000 iterations of the algorithm on group counts of 2-15 (using the 99 variables as the Percentage-BoxCox-Range dataset).

The statistical outputs suggested that a group count of 7 gave an optimum 'within-group variation' however on viewing how this split London's neighbourhoods visually it was felt that a further level of granularity was required. It was only after further group discussion that it was felt that a higher number of groupings were needed, which naturally also meant a preferable lower 'within-group variation'. This resulted in the allocation of the 107 neighbourhoods into 12 groupings as outlined in tabular and map form in Appendix D.

Outputs

Once the groupings had been finalised the secondary aim of being able to compare these 'most similar groups' with their relative crime and confidence data was considered. Using tools from within the mapping software ArcGIS and the Javascript programming code, an online interactive London map was created which allowed users to click on their desired neighbourhood and see the other neighbourhoods in its grouping also highlighted. The user was then able to select from a range of contextual confidence and crime datasets to see thematically how these varied across the neighbourhoods in the group. The raw data for each neighbourhood was also displayed.

Profiles were also produced for each grouping which provided a short text and visual summary of the group, focusing on the geography and the key contributing variables. To identify the key variables, the average value for each variable for each group was calculated and compared to the London average. Therefore for each group it was possible to calculate which variables were close to London mean (+/- 5% difference), significantly greater than the London mean (+30%) or significantly less than the London mean (-30%). These results were plotted on a radial plot.

Error checking

Error checking in this process was key, as a single error could have a significant effect on the final classification, especially when dealing with around 10,000 data points. The method for retrieving Census data was run multiple times to ensure no errors, and the input commands independently checked by another experienced Census software user within the GLA Intelligence Unit. Non-Census data was checked and rerun independently by the authors of the GLA Ward Atlas to ensure the same results were gained.

One way of limiting errors was to minimise human interaction with the source data. The benefit of the R code was that as long as the data was formatted in the correct way initially, there was little necessity for

human involvement. The only human input was for the removal of the correlated datasets, and the direction for choice of format-shape-scale permutation and group count as outlined above.

Limitations

- The methodology used for this work is recommended for use with small geographical areas such as Output and Super Output Area to ensure that variation within an area is minimised. Using a larger geography of ward aggregations, as has been requested for this work, is likely to result in groups that cover too much variation within each group ie. many different views within a group. Although the same methodology is used for ward and Local Authority level, these situations can lead to relationships at one geographic level (e.g. group) that are then not seen (even inversed) at a different geographic level (e.g. subgroup), known as the Modifiable Areal Unit Problem. This methodology uses the smallest level of geography palatable to the requesting party, in their full knowledge of the issues outlined above.
- With the MPS neighbourhoods built using aggregations of wards, the risk exists of future ward-boundary changes. The Local Government Boundary Commission for England (LGBCE) reviews arrangements based on changes in population (the electorate), and boundary changes can occur every year, usually on the first Thursday in May when local government elections take place. Boundaries may also be affected by parish boundary changes, which can occur throughout the year. If significant ward boundary changes occur that require the MPS to reallocate wards within neighbourhoods, then naturally consideration will be given (in conjunction with the MPS) to recalculating the set of most similar groupings.
- Key to the processes used above is alignment with the methodologies of UCL/ONS. Whilst a number of other potential methodologies were explored initially, the UCL/ONS option provided a current, nationally accepted methodology which had synergies with this project. A number of the other methodologies would likely have also involved a large amount of subjective decision making within the methodological steps, which although could have brought perceived benefits, an objective alignment to an accepted and current methodology was deemed more satisfactory.

References

- Adnan, M (2011), "Towards real-time geodemographic information systems: design, analysis and evaluation" <http://discovery.ucl.ac.uk/1335608/1/1335608.pdf>
- Bailey, S., Charlton, J., Dollamore, G. and Fitzpatrick, J. (1999a), "Which authorities are alike?" *Population Trends*, 98, 29-41.
- Bailey, S., Charlton, J., Dollamore, G. and Fitzpatrick, J. (1999b), "The ONS classification of local and health authorities of Great Britain: revised for authorities in 1999", *Studies in Medical and Population Subjects* No. 63
- Bailey, S., Charlton, J., Dollamore, G., and Fitzpatrick, J. (2000), "Families, groups and clusters of local and health authorities of Great Britain: Revised for authorities in 1999" *Population Trends*, 99, 37-52.
- Charlton M.E., Openshaw S., Wymer C. (1985) "Some new classifications of census enumeration districts in Britain: a poor man's ACORN" (*Journal of Economic and Social Measurement*, volume 13)

Ding, C & He, X (2004), "K-means Clustering via Principal Component Analysis"
<http://ranger.uta.edu/~chqding/papers/KmeansPCA1.pdf>

Everitt, B. S., Landau, S. and Leese, M. (2001), *Cluster Analysis 4th Ed. London, Arnold.*

Gordon, A. D. (1999), *Classification 2nd Ed., London, Chapman and Hall.*

Harris, R., Sleight, P. and Webber, R. (2005), "Geodemographics, GIS and Neighbourhood Targeting", *London, Wiley.*

Makarenkov, V. and Legendre, P. (2001), "Optimal variable weighting for ultrametric and additive trees and k-means partitioning: Methods and software", *Journal of Classification*, 18, 245-271.

Norusis, M - *IBM SPSS Statistics Guides, Chapter 16 Cluster Analysis*
http://www.norusis.com/pdf/SPC_v13.pdf

Singleton, A. D., Longley, P. A. (2009). "Creating Open Source Geodemographics - Refining a National Classification of Census Output Areas for Applications in Higher Education". *Papers in Regional Science*, 88(3), 643-666.

Vickers, Rees and Birkin (2005) "Creating the National Classifications of Census Output Areas: Data, Methods and Results"
<http://www.geog.leeds.ac.uk/fileadmin/documents/research/csap/05-02.pdf>

Vickers, D (2006), "Multi-Level Integrated Classifications based on the 2001 Census"
<http://etheses.whiterose.ac.uk/15/>

Voas, D. and Williamson, P. (2001a), "The diversity of diversity: a critique of geodemographic Classification", *Area*, 33(1), 63-76.

Appendix A – Removed Datasets

Supplementary data can be found on the datasets at <http://data.london.gov.uk/datastore/package/ward-profiles-and-atlas> and <http://data.london.gov.uk/census>.

Stage of removal	Theme	Dataset	Reason for removal	Secondary Source	Primary source
Initial Collation	Crime & Disorder	All Ambulance Incidents	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	All weapon injuries	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	Animal Attack Incidents	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	Assault Incidents attended by Ambulance	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	Number of ambulance call outs for alcohol related illness	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	Burglary rate	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	Criminal Damage rate	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	Drugs rate	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	Fraud or Forgery rate	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	Other Notifiable Offences rate	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	Robbery rate	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	Sexual offences rate	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	Theft and Handling rate	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	Total crime rate	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	Violence against the person rate	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	Deliberate Fires	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Crime & Disorder	Deliberate Fires per 1,000 population	Crime-related	London Datastore Ward Atlas	SafeStats Data
	Socio-Economic Character	Number of SOAs in ward	Does not provide enough relevant information for profiling	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
	Environment	% Other Land Uses	Does not provide enough relevant information for profiling	London Datastore Ward Atlas	Department of Communities and Local Government on behalf of ONS' Neighbourhood Statistics
	Demographic Structure	Country of Birth - 2011 Census	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Demographic Structure	Ethnic Group 18 groups - 2011 Census	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011

	Demographic Structure	Ethnic Group 5 groups - 2011 Census	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Demographic Structure	Household Language - 2011 Census	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Demographic Structure	Religion - 2011 Census	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Demographic Structure	Age structure (numbers) - 2013	Duplicated in Census variables	London Datastore Ward Atlas	GLA SHLAA Trend based Population Projection data
	Demographic Structure	Age structure (percentage) - 2013	Duplicated in Census variables	London Datastore Ward Atlas	GLA SHLAA Trend based Population Projection data
	Education	Qualifications and Students - 2011 Census	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Employment & Industry	Adults not in Employment - 2011 Census	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Employment & Industry	Economic Activity - 2011 Census	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Employment & Industry	Lone Parent Not in Employment - 2011 Census	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Employment & Industry	Employees working in Accommodation and food service activities	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Activities of extraterritorial organisations and bodies	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Activities of households as employers etc	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Administrative and support service activities	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Agriculture, forestry and fishing	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Arts, entertainment and recreation	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Construction	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Education	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Electricity, gas, steam and air conditioning supply	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Financial and insurance activities	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Human health and social work activities	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Information and communication	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Manufacturing	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Mining and quarrying	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Other service activities	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Professional, scientific and technical activities	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)

	Employment & Industry	Employees working in Public administration and defence; compulsory social security	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Real estate activities	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Transportation and storage	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Water supply; sewerage, waste management and remediation activities	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Employees working in Wholesale and retail trade; repair of motor vehicles and motorcycles	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Total employees	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Number of Part-time Employees	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Employment & Industry	Number of Full-time employees	Duplicated in Census variables	London Datastore Ward Atlas	Business Register and Employment Survey (BRES)
	Health	Health - 2011 Census	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Households	Household composition - 2011 Census	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Households	All Household spaces - 2011 Census	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Households	Dwellings, Household Spaces and Accommodation Type - 2011 Census	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Households	Tenure of households - 2011 Census	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Transport	Cars per household - Census 2011	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Transport	Number of cars or vans in household % - Census 2011	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Transport	Sum of all cars or vans in the area - Census 2011	Duplicated in Census variables	London Datastore Ward Atlas	Census 2011
	Socio-Economic Character	Average Rank of Deprivation	Not suitable for scaling from Ward to Neighbourhood level	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
	Socio-Economic Character	Rank of average rank (within London)	Not suitable for scaling from Ward to Neighbourhood level	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
	Socio-Economic Character	Average Score of Deprivation	Not suitable for scaling from Ward to Neighbourhood level	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
	Socio-Economic Character	Rank of average score (within London)	Not suitable for scaling from Ward to Neighbourhood level	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
	Demographic Structure	Population Estimates	Similar dataset already exists within Ward Atlas dataset	London Datastore Ward Atlas	Office for National Statistics
	Environment	% area that is greenspace	Similar dataset already exists within Ward Atlas dataset	London Datastore Ward Atlas	Greenspace Information for Greater London and Ordnance Survey

	Environment	% homes deficient in access to Regional Park, Metropolitan Park, District Park, Local, Small or Pocket Park	Similar dataset already exists within Ward Atlas dataset	London Datastore Ward Atlas	Greenspace Information for Greater London and Ordnance Survey
	Households	% of dwellings sold during year	Similar dataset already exists within Ward Atlas dataset	London Datastore Ward Atlas	Land Registry
	Socio-Economic Character	Turnout Mayoral election	Similar dataset already exists within Ward Atlas dataset	London Datastore Ward Atlas	London Boroughs
	Socio-Economic Character	% of LSOAs in worst 10% nationally	Similar dataset already exists within Ward Atlas dataset	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
	Socio-Economic Character	% of LSOAs in worst 5% nationally	Similar dataset already exists within Ward Atlas dataset	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
	Socio-Economic Character	% of LSOAs in worst 50% nationally	Similar dataset already exists within Ward Atlas dataset	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
	Socio-Economic Character	Rank of employment scale (within London)	Similar dataset already exists within Ward Atlas dataset	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
	Socio-Economic Character	Rank of extent (within London)	Similar dataset already exists within Ward Atlas dataset	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
	Socio-Economic Character	Rank of IDACI (within London)	Similar dataset already exists within Ward Atlas dataset	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
	Socio-Economic Character	Rank of IDAOPI (within London)	Similar dataset already exists within Ward Atlas dataset	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
	Socio-Economic Character	Rank of income scale (within London)	Similar dataset already exists within Ward Atlas dataset	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)

Correlation	Environment	Annual Mean of Nitrogen Dioxide (NO2)	Too highly correlated with Annual Mean of Particulate Matter (PM10) (Non-Census)	London Datastore Ward Atlas	London Atmospheric Emissions Inventory
	Environment	Annual Mean of Nitrogen Oxide (NOx)	Too highly correlated with Annual Mean of Particulate Matter (PM10) (Non-Census)	London Datastore Ward Atlas	London Atmospheric Emissions Inventory
	Demographic Structure	Population density (persons per sq km)	Too highly correlated with Density (number of persons per hectare) (Census)	London Datastore Ward Atlas	GLA SHLAA Trend based Population Projection data
	Socio-Economic Character	Income Deprivation affecting Children Index (IDACI)	Too highly correlated with Deprivation Extent (Non-Census)	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
	Socio-Economic Character	Income Deprivation affecting Older People Index (IDAOPI)	Too highly correlated with Deprivation Extent (Non-Census)	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
	Socio-Economic Character	Income Support Claimants	Too highly correlated with JSA Claimants	London Datastore Ward Atlas	Department for Work and Pensions (DWP)
	Socio-Economic Character	Female JSA Claimants	Too highly correlated with Total JSA Claimants (Non-Census)	London Datastore Ward Atlas	Department for Work & Pensions (DWP) via NOMIS

	Socio-Economic Character	Male JSA Claimants	Too highly correlated with Total JSA Claimants (Non-Census)	London Datastore Ward Atlas	Department for Work & Pensions (DWP) via NOMIS
	Socio-Economic Character	JSA Claimants Aged 16-24	Too highly correlated with Total JSA Claimants (Non-Census)	London Datastore Ward Atlas	Department for Work & Pensions (DWP) via NOMIS
	Socio-Economic Character	Employment Scale of Deprivation	Too highly correlated with Unemployed (Census)	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)

Appendix B – Final Datasets

Supplementary data can be found on the datasets at <http://data.london.gov.uk/datastore/package/ward-profiles-and-atlas> and <http://data.london.gov.uk/census>.

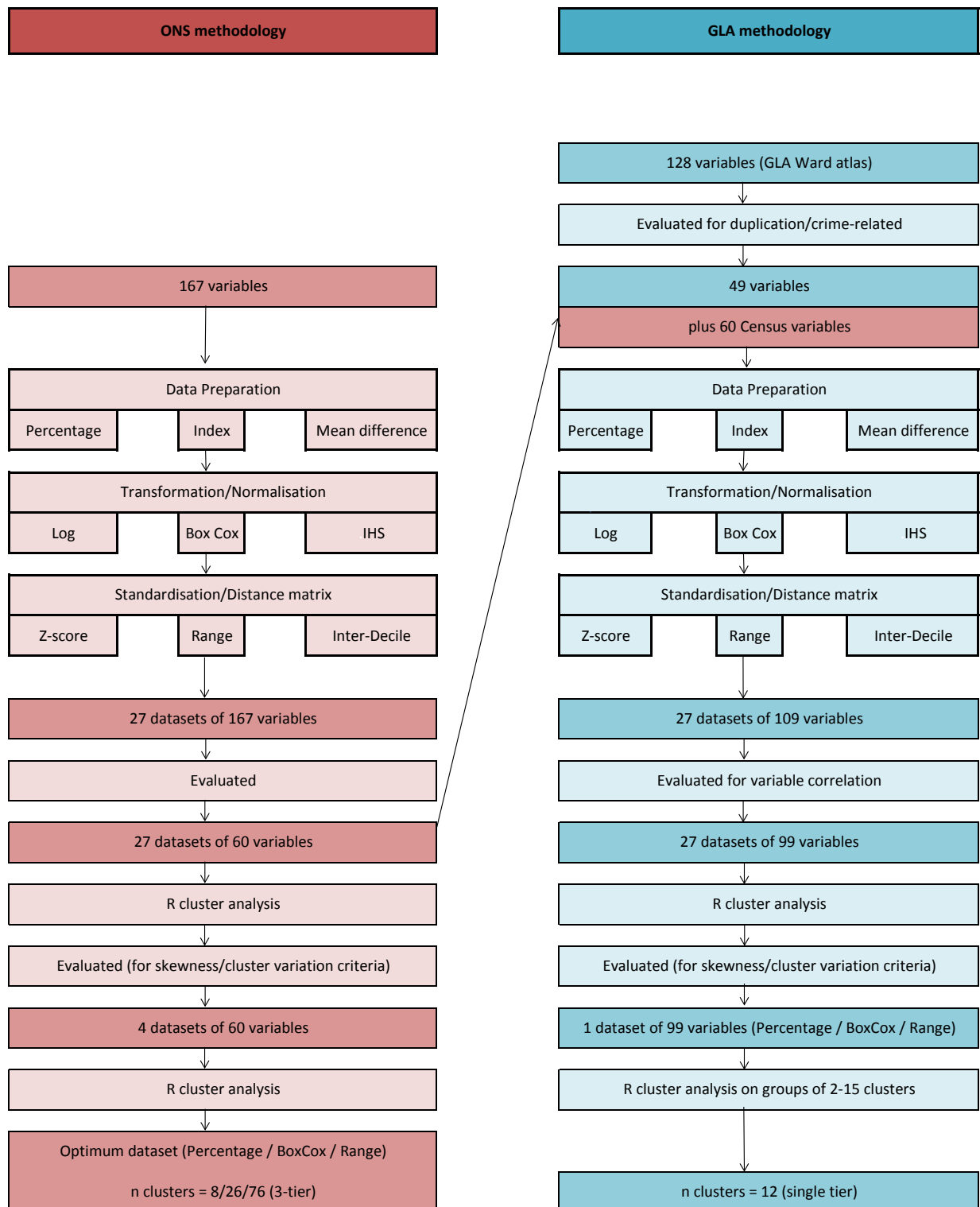
Theme	Dataset	Secondary Source	Primary source
Demographic Structure	Births	London Datastore Ward Atlas	Office for National Statistics
Demographic Structure	Deaths	London Datastore Ward Atlas	Office for National Statistics
Demographic Structure	Average age	London Datastore Ward Atlas	GLA SHLAA Trend based Population Projection data
Demographic Structure	GLA Projections	London Datastore Ward Atlas	GLA SHLAA Trend based Population Projection data
Demographic Structure	White	Census	Census 2011
Demographic Structure	Mixed/multiple ethnic group	Census	Census 2011
Demographic Structure	Asian/Asian British: Indian	Census	Census 2011
Demographic Structure	Asian/Asian British: Pakistani	Census	Census 2011
Demographic Structure	Asian/Asian British: Bangladeshi	Census	Census 2011
Demographic Structure	Asian/Asian British: Chinese and Other	Census	Census 2011
Demographic Structure	Black/African/Caribbean/Black British	Census	Census 2011
Demographic Structure	Arab or other ethnic groups	Census	Census 2011
Demographic Structure	Single	Census	Census 2011
Demographic Structure	Married or in a registered same-sex civil partnership	Census	Census 2011
Demographic Structure	Divorced or Separated	Census	Census 2011
Demographic Structure	Age 0 to 4	Census	Census 2011
Demographic Structure	Age 5 to 14	Census	Census 2011

Demographic Structure	Age 25 to 44	Census	Census 2011
Demographic Structure	Age 45 to 64	Census	Census 2011
Demographic Structure	Age 65 to 89	Census	Census 2011
Demographic Structure	Age 90 and over	Census	Census 2011
Demographic Structure	Density (number of persons per hectare)	Census	Census 2011
Demographic Structure	Lives in a communal establishment	Census	Census 2011
Demographic Structure	Main language is not English and cannot speak English well or at all	Census	Census 2011
Demographic Structure	United Kingdom and Ireland	Census	Census 2011
Demographic Structure	Other EU: Member countries in March 2001	Census	Census 2011
Demographic Structure	Other EU: Accession countries April 2001 to March 2011	Census	Census 2011
Education	Highest level of qualification: Level 1, Level 2 or Apprenticeship	Census	Census 2011
Education	Highest level of qualification: Level 3 qualifications	Census	Census 2011
Education	Highest level of qualification: Level 4 qualifications and above	Census	Census 2011
Education	Schoolchildren and full-time students: Age 16 and over	Census	Census 2011
Environment	% homes by number of ways deficient to access to public open space (0-4)	London Datastore Ward Atlas	Greenspace Information for Greater London and Ordnance Survey
Environment	% homes with deficiency in access to nature	London Datastore Ward Atlas	Greenspace Information for Greater London and Ordnance Survey
Environment	Annual Mean of Particulate Matter (PM10)	London Datastore Ward Atlas	London Atmospheric Emissions Inventory
Environment	% Domestic Buildings	London Datastore Ward Atlas	Department of Communities and Local Government on behalf of ONS' Neighbourhood Statistics
Environment	% Domestic Gardens	London Datastore Ward Atlas	Department of Communities and Local Government on behalf of ONS' Neighbourhood Statistics
Environment	% Greenspace	London Datastore Ward Atlas	Department of Communities and Local Government on behalf of ONS' Neighbourhood Statistics
Environment	% Non Domestic Buildings	London Datastore Ward Atlas	Department of Communities and Local Government on behalf of ONS' Neighbourhood Statistics
Environment	% Path	London Datastore Ward Atlas	Department of Communities and Local Government on behalf of ONS' Neighbourhood Statistics
Environment	% Rail	London Datastore Ward Atlas	Department of Communities and Local Government on behalf of ONS' Neighbourhood Statistics
Environment	% Road	London Datastore Ward Atlas	Department of Communities and Local Government on behalf of ONS' Neighbourhood Statistics
Environment	% Water	London Datastore Ward Atlas	Department of Communities and Local Government on behalf of ONS' Neighbourhood Statistics
Environment	Area of Admin Geography (Hectares)	London Datastore Ward Atlas	Department of Communities and Local Government on behalf of ONS' Neighbourhood Statistics
Health and Care	Female life expectancy	London Datastore Ward Atlas	London Health Programmes (LHP) using ONS mortality data and GLA population projections

Health and Care	Male life expectancy	London Datastore Ward Atlas	London Health Programmes (LHP) using ONS mortality data and GLA population projections
Health and Care	Day-to-day activities limited a lot or a little Standardised Illness Ratio	Census	Census 2011
Health and Care	Provides unpaid care	Census	Census 2011
Households	Whole house or bungalow: Semi-detached	Census	Census 2011
Households	No children household	Census	Census 2011
Households	Non-dependent children household	Census	Census 2011
Households	Full-time student household	Census	Census 2011
Households	Whole house or bungalow: Detached	Census	Census 2011
Households	% dwellings in council tax bands A or B	London Datastore Ward Atlas	Neighbourhood Statistics (ONS)
Households	% dwellings in council tax bands C, D or E	London Datastore Ward Atlas	Neighbourhood Statistics (ONS)
Households	% dwellings in council tax bands F, G or H	London Datastore Ward Atlas	Neighbourhood Statistics (ONS)
Households	Number of dwellings	London Datastore Ward Atlas	Neighbourhood Statistics (ONS)
Households	Median House Price	London Datastore Ward Atlas	Land Registry
Households	Number of properties sold	London Datastore Ward Atlas	Land Registry
Households	Private rented	Census	Census 2011
Households	Whole house or bungalow: Terrace and end-terrace	Census	Census 2011
Households	Flats	Census	Census 2011
Households	Owned or Shared Ownership	Census	Census 2011
Households	Social rented	Census	Census 2011
Households	One fewer or less rooms than required	Census	Census 2011
Industry & Employment	Part-time	Census	Census 2011
Industry & Employment	Full-time	Census	Census 2011
Industry & Employment	GCSE capped point scores	London Datastore Ward Atlas	Department for Education (on Neighbourhood Statistics)
Industry & Employment	Authorised Absence in All Schools (%)	London Datastore Ward Atlas	Department for Education (on Neighbourhood Statistics)
Industry & Employment	Overall Absence in All Schools (%)	London Datastore Ward Atlas	Department for Education (on Neighbourhood Statistics)
Industry & Employment	Unauthorised Absence in All Schools (%)	London Datastore Ward Atlas	Department for Education (on Neighbourhood Statistics)
Industry & Employment	Agriculture, forestry and fishing	Census	Census 2011
Industry & Employment	Mining, quarrying and construction	Census	Census 2011

Industry & Employment	Manufacturing	Census	Census 2011
Industry & Employment	Energy, water and air conditioning supply	Census	Census 2011
Industry & Employment	Wholesale and retail trade; repair of motor vehicles and motor cycles	Census	Census 2011
Industry & Employment	Transport and storage	Census	Census 2011
Industry & Employment	Accommodation and food service activities	Census	Census 2011
Industry & Employment	Information and communication and professional, scientific and technical activities	Census	Census 2011
Industry & Employment	Financial, insurance and real estate activities	Census	Census 2011
Industry & Employment	Administrative and support service activities	Census	Census 2011
Industry & Employment	Public administration and defence; compulsory social security	Census	Census 2011
Industry & Employment	Education	Census	Census 2011
Industry & Employment	Human health and social work activities	Census	Census 2011
Industry & Employment	Unemployed	Census	Census 2011
Socio-Economic Character	Incapacity Benefit Claimants	London Datastore Ward Atlas	Department for Work and Pensions (DWP)
Socio-Economic Character	Total JSA Claimants	London Datastore Ward Atlas	Department for Work & Pensions (DWP) via NOMIS
Socio-Economic Character	Children living in Out-of-work Benefit Claimant Households	London Datastore Ward Atlas	Department for Work and Pensions (DWP)
Socio-Economic Character	Turnout Borough election	London Datastore Ward Atlas	London Boroughs
Socio-Economic Character	% of LSOAs in worst 20% nationally	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
Socio-Economic Character	Extent of Deprivation	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
Socio-Economic Character	Income Scale	London Datastore Ward Atlas	Department of Communities and Local Government (with supplementary GLA calculations)
Socio-Economic Character	NINo Registrations	London Datastore Ward Atlas	Department for Work and Pensions (DWP)
Transport	Average PTAL score	London Datastore Ward Atlas	Transport for London (TfL), further calculations by GLA
Transport	Underground Footfall	London Datastore Ward Atlas	Transport for London (TfL)
Transport	Overground Footfall	London Datastore Ward Atlas	Office of Rail Regulation
Transport	Public Transport	Census	Census 2011
Transport	Private Transport	Census	Census 2011
Transport	On foot, Bicycle or Other	Census	Census 2011
Transport	2 or more cars or vans in household	Census	Census 2011

Appendix C: Comparison of ONS and GLA clustering methodologies



Appendix D – Final Neighbourhood Groupings

Group	Neighbourhood
-------	---------------

Group 1	Camden - North
	Hammersmith and Fulham - Fulham
	Greenwich - Greenwich
	Hounslow - East
	Merton - Wimbledon
	Wandsworth - Tooting
	Wandsworth - Battersea
	Wandsworth - Putney
	Ealing - Acton
	Ealing - Ealing
	Haringey - West

Group 2	Kensington and Chelsea - Kensington
	Kensington and Chelsea - Chelsea
	Westminster - Central
	Westminster - South

Group 3	Westminster - West End
----------------	------------------------

Group 4	Redbridge - West
	Bromley - North-West
	Barnet - Barnet
	Richmond upon Thames - Richmond
	Richmond upon Thames - Teddington
	Richmond upon Thames - Twickenham
	Kingston upon Thames - North
	Kingston upon Thames - South
	Merton - Morden
	Sutton - West

Group 5	Lambeth - South
	Southwark - South-West
	Southwark - South-East
	Lewisham - North
	Lewisham - South
	Lewisham - Central
	Brent - Kilburn
	Croydon - North-East
	Croydon - Central
	Croydon - North-West

Group	Neighbourhood
-------	---------------

Group 8	Havering - Central
	Havering - South
	Bromley - South-West
	Bromley - South-East
	Bromley - North-East
	Bexley - Central
	Bexley - South
	Hillingdon - North
	Croydon - South-West
	Sutton - East

Group 9	Waltham Forest - North
	Havering - North
	Barking and Dagenham - Whalebone
	Barking and Dagenham - Dagenham
	Greenwich - Eltham
	Bexley - North
	Enfield - Enfield & North
	Croydon - South-East
	Sutton - North

Group 10	Hackney - Stoke Newington
	Hackney - Homerton
	Hackney - Shoreditch
	Tower Hamlets - Stepney & Wapping
	Tower Hamlets - Poplar Isle of Dogs
	Tower Hamlets - Bricklane & Globe
	Tower Hamlets - Bow and Mile End
	Lambeth - Central
	Lambeth - North
	Southwark - North-East
	Southwark - North-West

Group 11	Kensington and Chelsea - Notting Hill
	Westminster - North
	Camden - South
	Camden - Central
	Hammersmith and Fulham - Shepherds Bush
	Islington - North
	Islington - East
	Islington - West

Group 6	Redbridge - South
	Hounslow - Central
	Hounslow - North
	Hounslow - West
	Merton - Mitcham
	Ealing - Greenford/Northolt
	Ealing - Southall
	Hillingdon - West Drayton
	Hillingdon - Hayes

Group 7	Hackney - Hackney North-East
	Waltham Forest - South
	Waltham Forest - Central
	Newham - East
	Newham - South
	Newham - Central
	Newham - West
	Barking and Dagenham - Barking
	Brent - Harlsden
	Greenwich - Plumstead
	Enfield - Edmonton & South
	Haringey - North
	Haringey - East

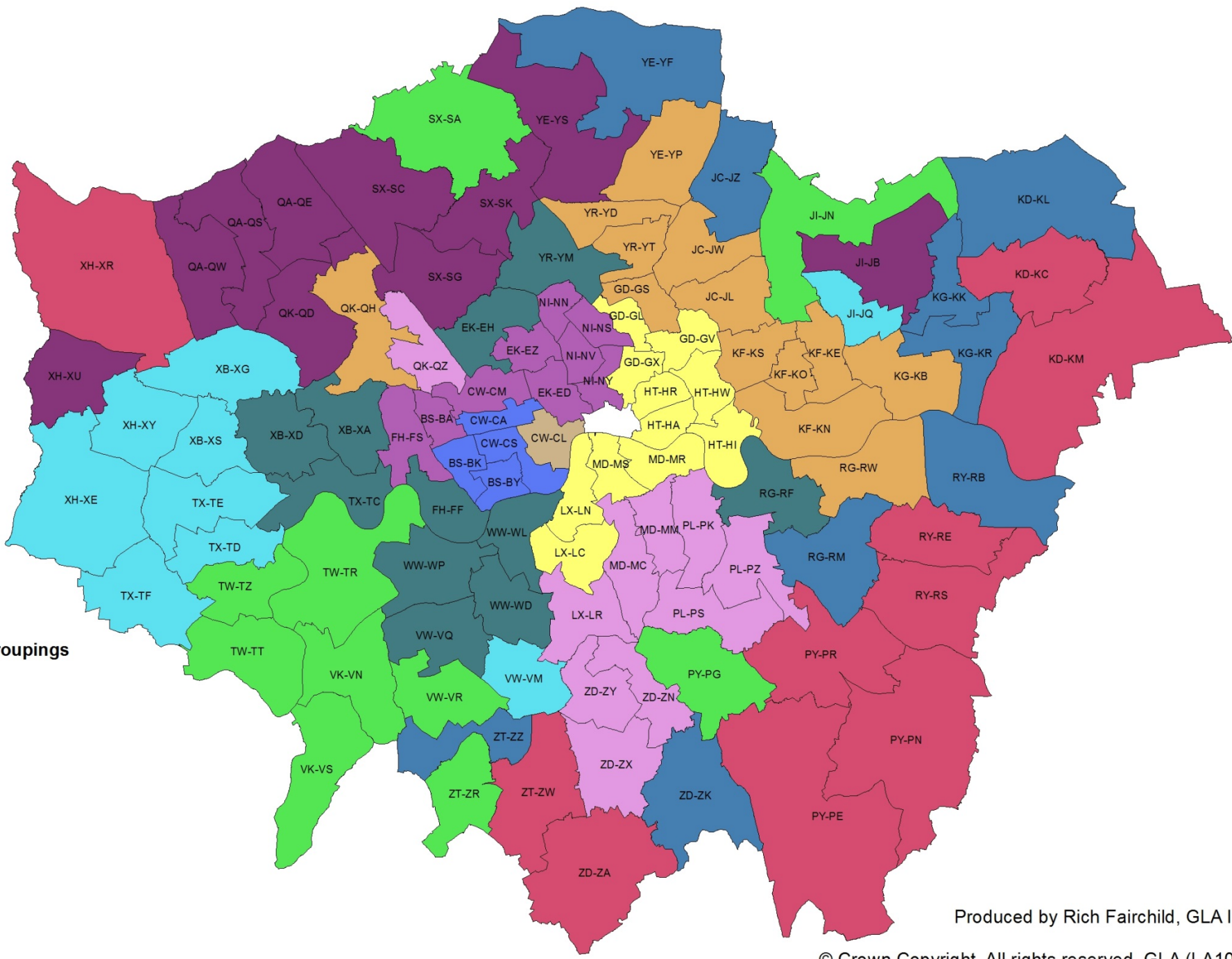
	Islington - South
--	-------------------

Group 12	Redbridge - Central
	Harrow - East
	Harrow - Central
	Harrow - West
	Brent - Wembley
	Barnet - Colindale
	Barnet - Golders Green
	Barnet - Whetstone
	Hillingdon - Uxbridge
	Enfield - Southgate & West

Map of Metropolitan Police Service neighbourhood most similar groups after GLA grouping analysis

Legend
Neighbourhood Groupings

- Group 1
- Group 2
- Group 3
- Group 4
- Group 5
- Group 6
- Group 7
- Group 8
- Group 9
- Group 10
- Group 11
- Group 12



GLA INTELLIGENCE

For more information please contact GLA Intelligence

Richard Fairchild,
Greater London Authority,
City Hall,
The Queen's Walk,
More London,
London SE1 2AA

Tel: 020 7983 4723
e-mail: richard.fairchild@london.gov.uk

Copyright © Greater London Authority, 2014

MAYOR OF LONDON