A classification for English primary schools using open data

Stephen Clark, Nik Lomax, Mark Birkin[1]

[1] University of Leeds, Leeds, United Kingdom

Abstract. England has statutory regulations in place that ensure state funded schools deliver broadly the same curriculum. However, there still exists a wide range of contexts in which this education takes place, including: the management of schools; how schools choose to spend their budgets; individual policies in regards to staffing; behaviour and attendance; and perhaps most importantly, the composition of the pupil population. Given these factors, one outcome of interest is the attainment profile of schools, and it is important that this performance is judged in context, for the benefits of pupils, parents and schools. To this end, this study develops a classification using contemporary data for English primary schools. The open data used captures aspects of the gender, ethnic, language, staffing and affluence makeup of each school. The nature of these derived groupings is described and made available as a mapping resource. These groupings allow the identification of "families of schools" to act as a resource for fostering better collaboration between schools and more nuanced benchmarking.

1 Introduction

The learning that takes place in a child's early years is often cited as one of the most critical phases in their education (Bruce 2012, Nores, Barnett 2010, Sammons 2011). Therefore, parents understandably want to ensure that their child receives a good education, particularly at the start of their education experience. In England, parents are able to rank their choices of schools, not being limited to the closest (Burgess et al. 2006, Harris, Johnston 2008). However, selecting a school does not necessarily mean that their child will be allocated a place there, especially if the school is oversubscribed. However, in the primary phase of English education, covering ages up to 11, parents are often able to send their children to local schools (Burgess et al. 2011).

Not all primary schools are the same. They are shaped by the composition of their pupil intake (e.g. gender, ethnicity or deprivation) (Harris 2010) and the ethos of the school (Day et al. 2016). These characteristics can have an important impact on the performance of the pupils and the school. Thus, many authorities and parents are keen to benchmark schools, in particular in regards to their academic performance. The question then arises as to which schools to benchmark against. Commonly, the options are benchmarked against a pool of schools within the same administrative area, or all schools nationally. However, given the heterogeneity of schools, this comparison can be unfair or meaningless.

Therefore, this study aims to capture this diversity in the characteristics of mainstream primary schools in England and establish a grouping of such schools. This in turn allows

for benchmarking against schools in the same group, or those in the same group but in close geographical proximity. This categorisation allows for a fairer assessment of schools' performance against their natural peers, which is critical if we are to ensure that the funding system does not favour schools purely on headline comparisons, penalising the ones that are performing better than headline attainment and progress statistics may suggest.

The mapping of these groupings of schools is available via this interactive map resource https://qgiscloud.com/tra6sdc/Map_QGISCloud/ with an accompanying guide in the Appendix. The map shows the neighbourhood contexts of schools in terms of the percentage of the non-White British or Irish living in the area and the rank of the degree of income deprivation affecting children (where 1 is the most deprived).

2 Capturing school heterogeneity

The issue of school effectiveness has received much academic attention. Whilst issues around leadership and teaching should not be neglected (Sammons et al. 2011, 2014), a common finding is that the socio-demographic and socio-economic compositions of a school's pupil population can have a big influence on its effectiveness and academic performance (Ainscow et al. 2016, Dustmann et al. 2010, Strand 2010, 2014).

2.1 Geodemographics

The method used here to develop a typology of English primary schools is a classification based on the characteristics of the schools. Such approaches are widely deployed in the field of geodemographics (Singleton, Spielman 2014), which attempts to classify neighbourhoods based on the characteristics of the people who live in the area (Gale et al. 2016) or work there (Cockings et al. 2015). However, such techniques are not limited to geographic areas; they can also be applied to other typologies such as individuals (Burns et al. 2017) or organisations (Phillip, Iyer 1975).

2.2 Groupings of schools

An early article by Bennett (1975) provides an introduction to an approach for categorisation, outlining many of the concepts needed to ensure a meaningful outcome. Dorabawila et al. (2002) classified schools in the Galle district of Sri Lanka into six groups, using information on school facilities and pupil performance. The authors commend the utility of their classification since it enables a fair distribution and targeting of funds to schools. A classification of French middle schools by Thaurel-Richard, Thomas (2006) used information on family socio-economic status, foreign national pupils, progress, attainment and the nature of the school to derive five classes of schools: urban privileged; under-privileged urban; small; under-privileged socially mixed and privileged socially mixed. Johnston et al. (2005) defined a grouping of English Secondary schools based on their ethnic composition and determined the membership into five groups as a function of: (1) the percentage of the White pupil population; and (2) the dominance of a non-White group. This approach placed schools into a grid based on the mono- or multi-ethnic nature of their pupil population. In perhaps the closest study to the work here, Gibbs et al. (2011) used ethnicity and deprivation data from London primary schools to identify 14 classes, which were then allocated into four groups defined by their relative position on a deprivation (well-off vs in-need) and ethnicity (White vs non-White) scale.

This examination into the literature has highlighted that there exists no up to date grouping of primary schools in England. What is available however is a range of databases that attempt to identify the closest "statistical neighbours" for schools based on their characteristics and performance (Education Endowment Foundation 2018, SchoolDash 2018). Such studies allow schools to benchmark against similar schools. However, it is left for schools to decide how extensive this search for statistical neighbours should be and when the comparisons become less valid. The approach proposed in this study allows schools to select statistical neighbours from a defined set of schools that share the same characteristics, recognising that "... geodemographic typologies are structured methods

for making sense of the spatial and socioeconomic patterns [in schools]." (Harris et al. 2007, p. 556).

3 Data and Methods

The data used in this study are obtained from the Department for Education (2018a) and relates to the academic year from September 2018 to August 2019. The data are derived from a number of sources, primarily the annual census of schools and pupils that takes place in the Spring term (Department for Education 2018b).

The composition of the pupils attending the school forms a vital component of the data. These include demographic information about the number of pupils, their gender and ethnic backgrounds. Further information is also available, including the number of pupils eligible for free school meals; number with a statement of Special Educational Needs (SEN); the rate of authorised and unauthorised absences; and the number of pupils whose first language is not English. Information on the staff composition of schools is also available. Finally, there is a measure of deprivation of the schools' catchments.

The data contain 24,952 schools. Not all these schools are appropriate for analysis, with an initial sub-set of 20,472 consisting of those that are designated as primary schools. Of these, 18,683 were open during the whole of the academic year 2018-2019. Some primary schools do not have cohorts covering the required ages, and restricting our sample to schools with starting ages of 2, 3, 4 or 5 and a highest age of 11, gives us 14,091 schools. Further elimination of Special, Independent and unknown school types leaves 13,443 mainstream primary schools for consideration in this study.

Table 1 lists the variables used to define the groups of primary schools. They fall into six sets: the ethnic composition of the schools' pupils; the degree of classroom over-crowding; the staffing structure; the demand for Special Educational Needs SEN provision; the degree of absences; and finally, the deprivation. Most of these variables are expressed as a percentage of the pupil or staff population, whilst two are direct measures: the pupil teacher ratio and a population weighted child-centred deprivation measure taken from Ministry of Housing Communities and Local Government (2019).

For some schools these data items are based on small numbers suppressed in the supplied data tables for confidentiality reasons. The complete case analysis therefore involves 13,363 primary schools. These data have the advantage of being provided by a trusted source, and are all openly available.

The method used to establish the groupings of schools is the widely applied k-means approach (Everitt et al. 2001). This method attempts to form a given number of clusters of schools based on their similarity. This similarity is measured in how close schools are in the "data space", i.e. the variables described above, from a cluster mean. A school is always allocated to the cluster whose mean is closest. However, the process of forming these clusters is iterative, with schools moving between clusters and cluster centres being updated until all schools are stable in their cluster. The quality of the final solution can be measured using a within-group-sum-of-squares, with better solutions having lower values.

This categorisation approach works best when the variables are uncorrelated, not skewed and are measured on a similar scale. The two variables, 'percentage of White British or Irish' and 'percentage with English as first language', have an absolute correlation above 0.75. Using the criteria that the variables that are, on average, highly correlated with the remaining variables should be removed, both the 'percentage White British or Irish' and the 'percentage with English as first language' are discarded as categorisation variables. Similarly, the percentage of pupils eligible for free school meals and the deprivation ranking are also correlated, and the free school meal measure is not included. The percentages of pupils who are boys and who are girls are also highly correlated and the percentage of female pupils is removed. In regards to skewness, Tukey's ladder of powers transformation (Mosteller, Tukey Mosteller, Tukey) is used to correct for a positive skew in these data. For standardisation, a range standardisation is applied.

Similar to the approach used to derive 2001 and 2011 Output Area Classifications from the UK census data, a hierarchical approach to categorisation is adopted (Gale et al.
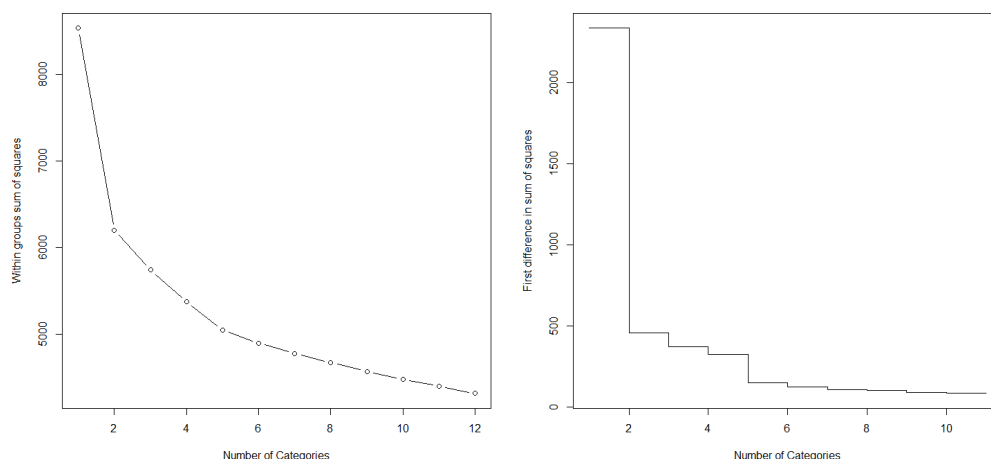
Figure 1: Scree plots of number of groupings and (a) within group sum of squares, (b) first difference in within group sum of squares

2016). Firstly, a number of Groups are formed, and then a further categorisation of the schools within each Group is undertaken, after re-transformation and re-standardisation, to define a series of Sub-groups. To gain an understanding of the nature of each Group and Sub-group, reference is made to the mean centres of each grouping, taken as averages of the variables for all schools in that group. These are presented on the raw scale, and on a scale that standardises each group centre relative to both all schools and schools within the same Group.

4 Results

Choosing the number of groupings using k-means is not an exact science, however, there is a range of methods which can support the decision-making process. The scree plots of the within-group-sum-of-squares for a given value of k and its first difference identify the point at which additional clusters do not materially reduce this measure of fit. In both these plots we are looking for an 'elbow' where the change in trajectory reduces or levels off. For the Group level of categorisation, the within-group-sum-of-squares are calculated using the k-means function in R (R Core Team 2017) with 100 random starting points, and plotted in Figure 1.

   The scree plot of weighted sum of squares (Figure 1a) suggests that there are five groupings at this Group level, and this is confirmed by looking at the first differences (Figure 1b), where they level-off after moving beyond five groupings.

4.1 Category Groups

The raw group centres and the standardised versions of these centres are provided in Table 1 and Table 2 (scree and radial plots of this information are also provided in the supplementary material file SupplimentalScreeRadial.pdf). What is of interest for interpretation purposes are those variables where the group centre is particularly different from either all schools or schools in other groupings.

   In these tables, it is clear that the nature of these groupings is most distinctly defined by the ethnic composition and the deprivation of the schools' pupils or catchments, and can be described as:

A: Multi-ethnic and Affluent : these schools have pupils coming from a range of ethnic backgrounds, but with the White British or Irish group still being dominant at nearly 70%. This is an affluent Group, having one of the largest rankings for

Table 1: Group centres on the raw scale

Variable / Group	A	B	C	D	E
Boys (%)	49.05	49.01	49.10	49.12	49.25
Girls (%) [1]	50.95	50.99	50.90	50.88	50.75
White British or Irish (%) [1]	68.41	91.01	92.09	83.91	30.41
White Other (%)	8.24	2.68	2.36	5.94	12.09
Traveller (%)	0.29	0.33	0.54	0.72	0.76
Mixed (%)	7.74	3.23	2.90	3.94	9.91
Indian (%)	4.15	0.39	0.19	0.48	5.38
Pakistani or Bangladeshi (%)	2.76	0.29	0.10	0.74	15.33
Other Asian (%)	1.86	0.27	0.17	0.60	3.75
Black (%)	3.26	0.40	0.26	1.78	15.45
Other (%)	2.14	0.48	0.31	1.13	5.67
Ethnicity unclassified (%)	1.14	0.91	1.07	0.74	1.26
First language is English (%) [1]	83.43	97.07	97.79	91.02	52.40
First language is not English (%)	16.37	2.83	2.09	8.89	47.33
First language is unclassified (%)	0.20	0.10	0.13	0.09	0.27
Pupils in classes of 31 to 35 with one teacher (%)	11.73	31.29	0.12	6.53	6.73
Pupils in classes of 36 or more with one teacher (%)	0.81	0.82	0.32	0.91	0.86
Pupil-Teacher ratio	21.80	22.19	18.82	20.64	20.28
Teaching staff (%)	46.74	46.22	48.77	42.79	43.71
Teaching Assistant (%)	33.21	33.71	31.11	37.31	35.48
Non-class based (%)	11.30	11.04	11.72	11.13	12.46
Auxiliary (%)	8.79	9.09	8.51	8.83	8.39
SEN pupils (%)	1.54	1.40	1.57	1.68	1.78
Authorised absence (%)	2.87	2.91	3.15	3.21	2.87
Unauthorised absence (%)	0.69	0.65	0.69	1.25	1.26
FSM pupils (%) [1]	13.59	13.15	13.90	36.06	33.79
Catchment IMD	21987	20884	21750	7384	8338

Notes: [1] Percentage of girls; percentage White British or Irish; percentage with English as first language; and percentage eligible for free school meals are not used in the categorisation but are reported here.

affluence. There are a large number of pupils in over-sized classes, suggesting these schools are popular with parents.

B: White British or Irish and Popular : this is the first of three Groups with a dominant White British or Irish ethnic grouping. These schools are very popular with parents, meaning that nearly a third of pupils are in classes with more than 30 pupils. Whilst not as affluent as some Groups, these schools are located in comfortable neighbourhoods.

C: White British or Irish and Affluent : This is another Group dominated by pupils of a White British or Irish ethnicity, but in contrast to the previous Group, there is no evidence of oversubscription from larger class sizes. These schools are also located in affluent neighbourhoods.

D: White and Deprived : in this Group the White British or Irish in combination with the White groups of other ethnic backgrounds form a large proportion, at nearly 90%. This Group is further differentiated by the level of deprivation, which is high, both from the perspective of the percentage of pupils that are eligible for free school meals and also the deprivation of the schools' neighbourhoods.

Table 2: Group centres on the standardised scale

Variable / Group	A	B	C	D	E
Boys (%)	0.9987	0.9980	0.9998	1.0002	1.0029
Girls (%) [2]	1.0013	1.0020	1.0002	0.9998	0.9972
White British or Irish (%) [2]	0.9488	1.2622	1.2772	1.1638	0.4217
White Other (%)	1.2981	0.4225	0.3713	0.9357	1.9041
Traveller (%)	0.5337	0.6117	1.0053	1.3453	1.4154
Mixed (%)	1.3776	0.5754	0.5160	0.7021	1.7631
Indian (%)	1.9093	0.1782	0.0895	0.2225	2.4738
Pakistani or Bangladeshi (%)	0.6580	0.0703	0.0246	0.1763	3.6547
Other Asian (%)	1.3435	0.1964	0.1239	0.4349	2.7012
Black (%)	0.7155	0.0880	0.0566	0.3908	3.3901
Other (%)	1.0490	0.2350	0.1520	0.5548	2.7774
Ethnicity unclassified (%)	1.1065	0.8778	1.0364	0.7134	1.2179
First language is English (%) [2]	0.9991	1.1624	1.1710	1.0899	0.6274
First language is not English (%)	1.0022	0.1733	0.1279	0.5444	2.8984
First language is unclassified (%)	1.2465	0.5991	0.7801	0.5641	1.7086
Pupils in classes of 31 to 35 with one teacher (%)	1.0731	2.8621	0.0109	0.5969	0.6158
Pupils in classes of 36 or more with one teacher (%)	1.1021	1.1141	0.4315	1.2304	1.1731
Pupil-Teacher ratio	1.0544	1.0733	0.9100	0.9982	0.9810
Teaching staff (%)	1.0236	1.0124	1.0681	0.9372	0.9573
Teaching Assistant (%)	0.9733	0.9878	0.9116	1.0933	1.0396
Non-class based (%)	0.9770	0.9538	1.0126	0.9616	1.0768
Auxiliary (%)	1.0104	1.0443	0.9773	1.0143	0.9638
SEN pupils (%)	0.9643	0.8733	0.9797	1.0485	1.1164
Authorised absence (%)	0.9581	0.9711	1.0487	1.0691	0.9557
Unauthorised absence (%)	0.7524	0.7169	0.7503	1.3646	1.3782
FSM pupils (%) [2]	0.6095	0.5897	0.6235	1.6170	1.5153
Catchment IMD	1.3801	1.3108	1.3652	0.4635	0.5234

Notes: [2] Percentage of girls; percentage of White British or Irish; percentage with English as first language; and percentage eligible for free school meals are not used in the categorisation but are reported here.

E: Multi-ethnic and Deprived : this final Group is the most multi-ethnic, with all ethnicities being present in large numbers, and the White British or Irish ethnic group comprising less than a third of pupils at the schools. There is also substantial deprivation associated with these schools, but less so than in Group D.

4.2 Category Sub-groups

The Groups presented above provide useful summary measures, but there is also some variation at a Sub-group level. The Sub-groups are constructed by applying k-means to just those schools in each group, following re-transformation and re-standardisation and using the methods outlined in Section 3. The scree plots, first difference in scree plots and tables of group centres are provided in the supplementary materials. Table 3 provides the number of schools in each sub-group along with information on how the Sub-groups differ within their Groups.

Table 3: Names for the Sub-groups

Code	Description	# schools
A	**Multi-ethnic Affluent**	**2402**
A.1	Comfortable	425
A.2	White Other	385
A.3	Affluent	448
A.4	Oversubscribed	442
A.5	Unclassified	364
A.6	Traveller	338
B	**White British/Irish & Over subscribed**	**2591**
B.1	White Other	414
B.2	Deprived	511
B.3	Affluent	624
B.4	Very oversubscribed	123
B.5	High absences	332
B.6	Comfortable	408
B.7	Unclassified	179
C	**White British/Irish**	**2873**
C.1	Girls	287
C.2	Unclassified	158
C.3	Traveller	267
C.4	Oversubscribed	227
C.5	Very White British/Irish	626
C.6	Low teachers	241
C.7	White Other	433
C.8	Deprived	634
D	**White British/Irish & Deprived**	**2461**
D.1	Comfortable	847
D.2	Unclassified	205
D.3	Very deprived	732
D.4	Oversubscribed	677
E	**Multi-ethnic Deprived**	**3036**
E.1	Black	558
E.2	Deprived	507
E.3	Indian	328
E.4	Oversubscribed	596
E.5	Very oversubscribed	207
E.6	Oversubscribed	492
E.7	Pakistan/Bangladesh	348

5   Geographic distribution

The regional distribution for each group is shown in Figure 2. This map reveals some significant spatial variations. There are few multi-ethnic schools in the North East and South West of England, whilst there are a sizeable number of such schools in the West Midlands and the South East. London stands out as particularly different to all the other regions. The two multi-ethnic Groups (A: Multi-ethnic Affluent and E: Multi-ethnic Deprived) dominate London schools, and the remaining mono-ethnic white groupings are very uncommon in London. A closer look at the distribution by London Boroughs in Figure 3 also reveals differences within London. Multi-ethnic schools dominate in the inner Boroughs of Tower Hamlets and Newham, whilst they are far less common in some outer Boroughs (Bromley and Barnet) where some white groupings are represented.

To furher illustrate the utility of this categorisation, an example map in Figure 4 is provided for the city of Derby. Each primary school is displayed by its Group and labeled with its Sub-group. The background maps show the Index of Multiple Deprivation (IMD) (the higher the rank, the higher the deprivation) and the percentage of the White British or Irish for the neigbourhood from the 2011 Census. This map shows that the
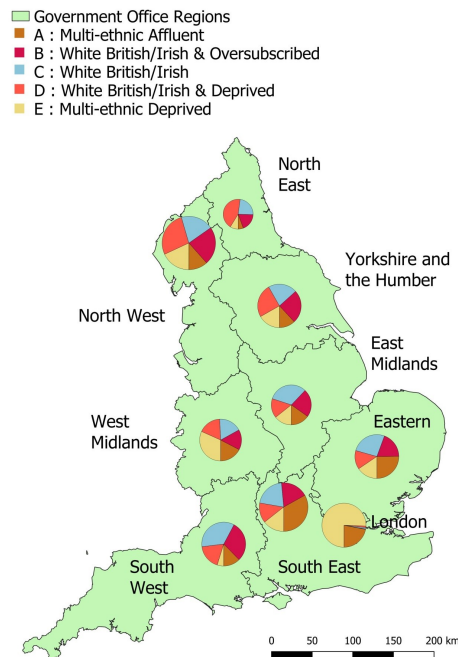
Figure 2: Geographic distribution of groupings by English region

more prosperous schools, in Groups A and B are located in areas with low deprivation ranks and that the multi-ethnic schools, Groups A and E are located in areas with lower percentages of the White British or Irish populations.

6   Discussion

This study has demonstrated that ethnicity is an important characteristic that differentiates primary schools in England. Within England, there are large areas within towns and cities with concentrations of particular ethnic groups, e.g. White British in rural towns, South Asians in ex-industrial northern towns, and Black populations in London. Given that the pupil catchments of primary schools are concentrated around their location, it is inevitable that such schools will have an intake of the dominant ethnic groups in their vicinity (this is especially the case in London, where Gibbs et al. (2011) note that "... *the vast majority of schools reflect the ethnic mix of their immediate neighbourhood.*". page 37-38). The affluence or deprivation of the school's pupil population is also similar, for concentrations of these measures in the locality of the school will dominate in the character of that school. These two aspects of schools are largely constrained by geography (Ainscow et al. 2016), however, the school can influence some of the other characteristics. A school may choose to employ more classroom teachers in preference over teaching assistants or non-classroom based staff – at a cost. This will then influence the composition of its staff and also whether pupils will be taught in large classes of 30 or more . However, some schools could struggle to manage an ideal staffing structure with overcrowding resulting from either the school being popular and having to take more pupils than its capacity allows or from the school being unable to attract enough teachers to provide a full complement of staff.

Another area over which a school has some control is the absence profile of pupils (Taylor 2012). Education welfare officers can be employed by schools to work with families whose children are failing to attend regularly. Schools can also fine parents for days that they take their children out of school. In reality, the percentage of authorised absences is similar between all the Groups and Sub-groups, but the percentage of un-authorised absences is larger in the groupings that are defined as challenged.

The relationship between this national categorisation of primary schools and the
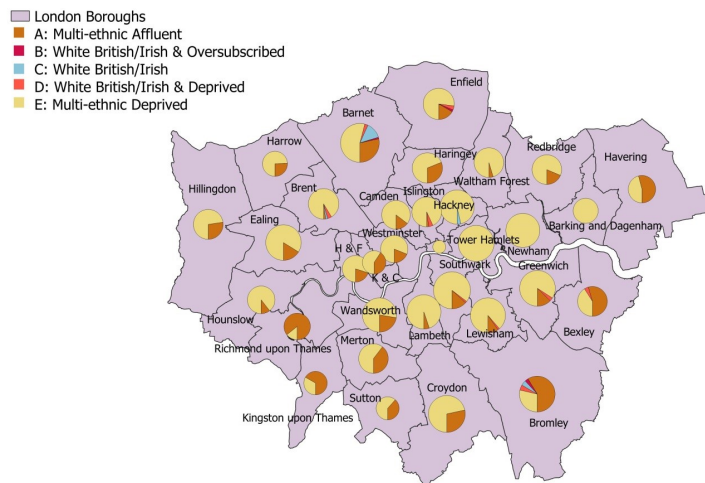
Figure 3: Geographic distribution of groupings by London Borough (H & F : Hammersmith and Fulham and K & C : Kensington and Chelsea)
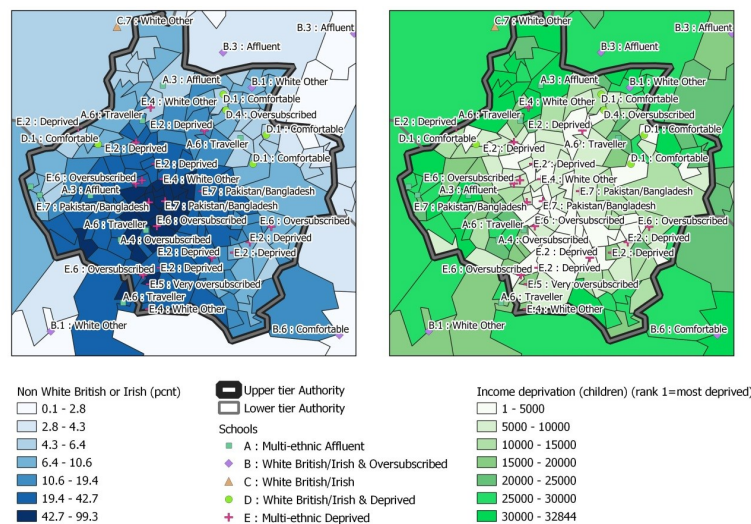

Figure 4: Categorisation of schools in Derby

London categorisation in Gibbs et al. (2011) is similar to that between the national OAC (Gale et al. 2016) and the London specific OAC (Greater London Authority 2017), the motivation for the latter being a desire to provide a categorisation that reflects the unique nature of London.

In this study a number of assumptions have been made. Amongst these is the use of the k-means classification technique in preference to other available methods. K-means is undoubtedly the most widely applied classification technique with applications in a wide number of domains. Another limitation has been our reliance on open data. This is however seen as a desirable feature for a number of reasons. Firstly, it allows the data to be shared with no restrictions, so that the work can be reproduced or extended. Secondly, it ensures that no information about any individual is disclosive. The decision not to include the performance of a school as a categorisation measure was also made. This assumption allows such measures to be treated as an output measure from the school, and there would undoubtedly be interest in how this independent measure varies amongst schools that otherwise look similar.

The circumstances of a school are not fixed over time. Transformations can take place in schools, for example, the ethnic or socio-economic compositions of a school

population can change, being driven by how a school's catchment evolves as a result of population migration or as other neighbouring schools expand or contract their intake. A new leadership team or governance arrangement may also take the school in a different direction. In light of this, it will be useful to re-visit these groupings over time. Luckily, the information used in this study, derived mainly from the annual school census, will keep being generated, allowing for this resource to be maintained.

References

Ainscow M, Dyson A, Hopwood L, Thomson S (2016) Primary schools responding to diversity: Barriers and possibilities. Cambridge primary review trust, York

Bennett N (1975) Cluster analysis in education research: A non-statistical introduction. *Research Intelligence* 1: 64–70

Bruce T (2012) Early childhood education: Hachette UK. Technical report

Burgess S, Briggs A, McConnell B, Slater H (2006) School choice in England: Background facts. DP, 6, 159

Burgess S, Greaves E, Vignoles A, Wilson D (2011) Parental choice of primary school in England: What types of school do different types of family really have available to them? *Policy Studies* 32: 531–547

Burns L, See L, Heppenstall A, Birkin M (2017) Developing an individual-level geodemographic classification. *Applied Spatial Analysis and Policy* 11: 417–437. CrossRef.

Cockings S, Martin D, Harfoot A (2015) A classification of workplace zones for England and Wales (COWZ-EW). Southampton: University of Southampton

Day C, Gu Q, Sammons P (2016) The impact of leadership on student outcomes. *Educational Administration Quarterly* 52: 221–258. CrossRef.

Department for Education (2018a) Get information about schools. Retrieved from https://www.get-information-schools.service.gov.uk/

Department for Education (2018b) School census 2016 to 2017: Guide for schools and local authorities. Retrieved from https://www.gov.uk/government/publications/school-census-2016-to-2017-guide-for-schools-and-las

Dorabawila S, Samita S, Thattil R (2002) Classification of state schools in Sri Lanka: A multivariate approach. *Ceylon Journal of Science: Physical Sciences* 9: 102–109

Dustmann C, Machin S, Schönberg U (2010) Ethnicity and educational achievement in compulsory schooling. *The Economic Journal* 120: F272–F297. CrossRef.

Education Endowment Foundation (2018) Families of schools database. Retrieved from https://educationendowmentfoundation.org.uk/tools/families-of-schools-database/

Everitt B, Landau S, Leese M (2001) *Cluster analysis* (4th ed.). Arnold, London

Gale C, Singleton A, Bates A, Longley P (2016) Creating the 2011 area classification for output areas (2011 OAC). *Journal of Spatial Information Science* (12): 1–27. CrossRef.

Gibbs A, Stillwell J, See L (2011) Understanding ethnic sorting and school performance through a classification of primary schools in London. *Education Research Journal* 1: 31–45

Greater London Authority (2017) London output area classification. Retrieved from https://data.london.gov.uk/dataset/london-area-classification

Harris R (2010) Segregation by choice?: The debate so far. Centre for market and public organisation, University of Bristol

Harris R, Johnston R (2008) Primary schools, markets and choice: Studying polarization and the core catchment areas of schools. *Applied Spatial Analysis and Policy* 1: 59–84

Harris R, Johnston R, Burgess S (2007) Neighborhoods, ethnicity and school choice: Developing a statistical framework for geodemographic analysis. *Population Research and Policy Review* 26: 553–579. CrossRef.

Johnston R, Wilson D, Burgess S (2005) England's multiethnic educational system? a classification of secondary schools. *Environment and Planning A* 37: 45–62

Ministry of Housing Communities and Local Government (2019) The English indices of deprivation 2019. Technical report. Retrieved from https://www.gov.uk/government/-publications/english-indices-of-deprivation-2019-research-report

Mosteller F, Tukey J *Data analysis and regression: A second course in statistics.* Addison-Wesley, Reading, Mass.

Nores M, Barnett W (2010) Benefits of early childhood interventions across the world: (under) investing in the very young. *Economics of education review* 29: 271–282. CrossRef.

Phillip P, Iyer R (1975) Classification of community hospitals. *Health Services Research* 10: 349–368. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1241783

R Core Team (2017) R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria

Sammons P (2011) Do the benefits of pre-school last? Investigating pupil outcomes to the end of key stage 2 (aged 11). In: Sylva K, Melhuish E, Sammons P, Siraj-Blatchford I, Taggart B (eds), *Early Childhood Matters: Evidence from the Effective Pre-school and Primary Education project.* Routledge, London and New York, 114–148

Sammons P, Davis S, Day C, Gu Q (2014) Using mixed methods to investigate school improvement and the role of leadership: An example of a longitudinal study in England. *Journal of Educational Administration* 52: 565–589

Sammons P, Gu Q, Day C, Ko J (2011) Exploring the impact of school leadership on pupil outcomes: Results from a study of academically improved and effective schools in England. *International Journal of Educational Management* 25: 83–101

SchoolDash (2018) Understand education. Retrieved from https://www.schooldash.com/-about.html#dashboards

Singleton A, Spielman S (2014) The past, present and future of geodemographic research in the United States and United Kingdom. *Professional Geographer* 66: 558–567. CrossRef.

Strand S (2010) Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement. *School Effectiveness and School Improvement* 21: 289–314

Strand S (2014) School effects and ethnic, gender and socio-economic gaps in educational achievement at age 11. *Oxford Review of Education* 40: 223–245

Taylor C (2012) Improving attendance at school. Retrieved from https://www.gov.uk/-government/publications/improving-attendance-at-school

Thaurel-Richard M, Thomas F (2006) Typologies des colleges publics. *Donnees sociales - La societe francaise*: 147–156. Retrieved from https://www.insee.fr/fr/statistiques/-fichier/1372011/donsoc06p.pdf

A   Appendix A: VIEWING OUR DATASET IN QGIS CLOUD

This short appendix introduces the functionality of QGIS Cloud in order to provide a guide to viewing our dataset.

A.1   FUNCTION BUTTONS

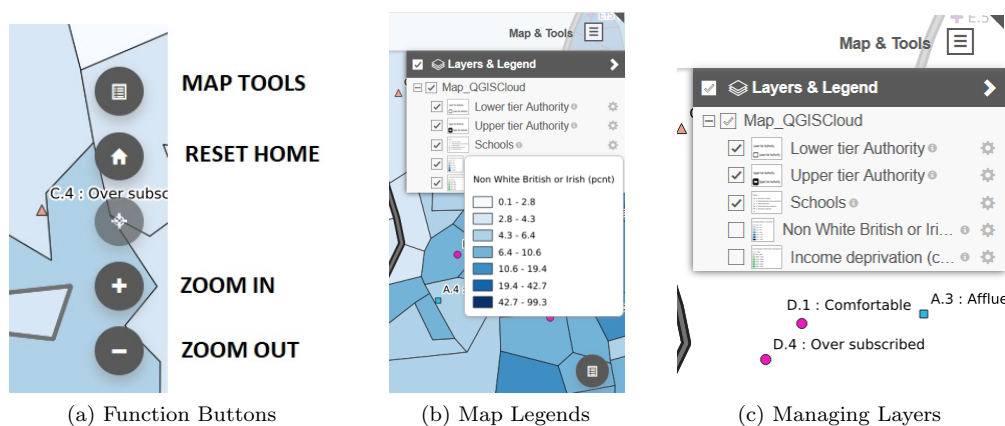To the lower right hand side of the map there are five function buttons in a column (Figure A.1a).

(a) Function Buttons   (b) Map Legends   (c) Managing Layers

Figure A.1: Function buttons and Map tools

Clicking on the top button opens up the MAP TOOLS (more on these later). The next button down returns the map to the HOME view, this is the view displayed when the map was first loaded. The bottom two buttons zoom in and out of the map. It is also possible to pan around the map using the left mouse button; click on the map and, whilst keeping the mouse button pressed down, move the mouse to change the area of the map in view.

A.2   MAP TOOLS

The MAP TOOLS allow the user to see the legends that correspond to the displays on the map and to switch layers on and off.

A.2.1   Opening MAP TOOLS

To see the MAP TOOLS click on the MAP TOOLS button.

A.2.2   Map legends

To see the legend associated with each layer hover the mouse above the miniature image of the legend. This will temporarily display a larger version of the legend. As an example, Figure A.1b shows the legend for the percent White British or Irish layer of our resource. To remove the legend, move the mouse off the miniature image of the legend.

A.2.3   Managing layers

To the left of the name of the layer in the MAP TOOLS display there is a tick box. If this box contains a tick then the layer is displayed on the map. If the box is empty then the layer is not displayed. To toggle the display of the layer, click in this box. If the layer is on (shown with a tick) and this box is clicked then the layer will be turned off. If the layer if off (shown with an empty box) and this box is clicked then the layer will be turned on. Figure A.1c shows the map with the layers for the Percentage White British or Irish and Income Deprivation Rank turned off.

A.2.4   Closing MAP TOOLS

To close the MAP TOOLS click on the MAP TOOLS button.

A.3   IDENTIFYING FEATURES

To identify a feature in the map, click on (or very close to) the feature or within the area of the map of interest. To the left of the map a display will show information on the objects selected in the five layers.
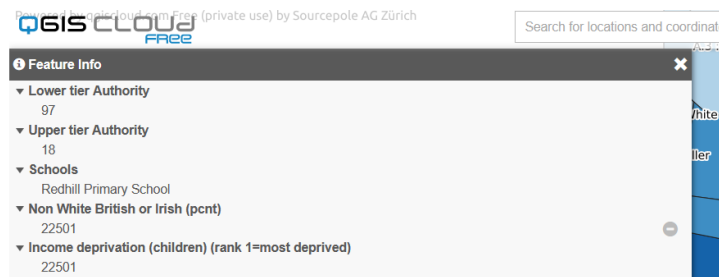
To see more detail on an object, click on the object text. For example, clicking on the text "Redhill Primary School" provides the classification details on Redhill School.
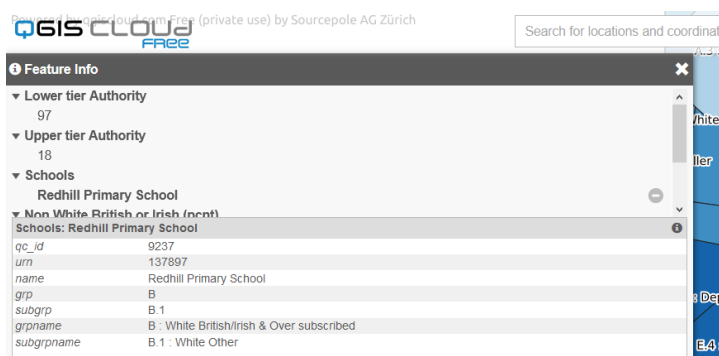
Clicking on the index number 22501 for the Non White British or Irish (pcnt) shows that the percentage of non White British or Irish population in the lower level super output area that contains Redhill School is 4.35%.
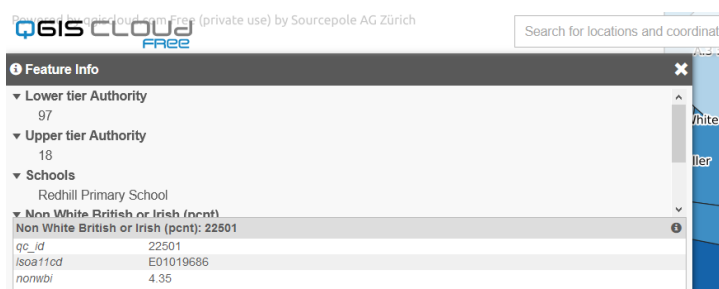
A.3.1   Close the Information Screen

To close the information screen, click on the white cross in the top right hand corner.