Creating an Output Area Classification of Cultural and Ethnic Heritage to Assist the Planning of Ethnic Origin Foods in Supermarkets in England and Wales

Guy Lansley[*1], Yiran Wei [†1] and Tim Rains[‡2]

[1]Department of Geography, UCL
[2]J Sainsbury's plc

04 January, 2015

This paper presents a Cultural, Ethnic and Linguistic Output Area Classification for England and Wales built from clustering census variables which pertain to cultural identity. The study provides a quick insight into the broad patterns in ethnic segregation based on the residential geography recorded from the 2011 Census and is therefore a useful tool for supermarket planners seeking to identify areas where to target particular ethnic origin foods. To confirm this association, the classification has also been compared with the total sales of a selection of ethnic origin foods using supermarket customer loyalty data.

KEYWORDS: ethnicity, k-means clustering, census, food consumption

1. Introduction

Many minority ethnic and cultural groups in Britain have distinctive food consumption habits which emanate from their cultural origins (Uskul and Platt, 2014). With the ethnic minority population of the UK growing (Simpson, 2013), understanding a basic segmentation of ethnic compositions across England and Wales is useful to supermarket planners aiming to make their stores more appealing to local ethnic groups.

Ethnic groups can be considered as distinctive groups of individuals who share a common identity through kinship, religion, language, location, nationality and physical similarities from ancestry (Bulmer, 1996). However, each one of these domains can singularly be a defining characteristic of an individual's cultural identity. Therefore, the definition and classification of ethnicities have attracted on-going debate due to its multidimensional, subjective and complex nature (Mateos *et al*, 2009).

The 2011 Census for England and Wales identified that the population was becoming more ethnically diverse, largely due to immigration and higher fertility rates amongst most ethnic minority groups compared with the national average (Simpson, 2013). Typically many minority groups residentially cluster within urban areas due to a range of structural social and economic forces (Finney, 2013). While minority ethnic groups have been found to be dispersing more recently (Stillwell and Hussain, 2010), most metropolitan neighbourhoods still bare a more diverse ethnic composition than the rest of the country.

Using data at the output area level from the 2011 Census, this paper aims to identify the major spatial traits in ethnic identity across the residential geography of England and Wales by producing a Cultural, Ethnic and Linguistic Output Area Classification (CELOAC).

[*] G.Lansley@ucl.ac.uk
[†] Yiran.Wei.13@ucl.ac.uk
[‡] Tim.Rains@sainsburys.co.uk

2. The 2011 Census for England and Wales

Output areas are the smallest geographical unit available from the 2011 Census data and have an average population of 309 (ONS, 2014). At this geography, over 400 variables relevant to the classification are available from seven Quick Statistics census tables. These cover key dimensions of cultural identity as labeled in Table 1.

Table 1 Census tables used in the classification

Census Table	Name
QS203EW	Country of birth (detailed)
QS204EW	Main language (detailed)
QS205EW	Proficiency in English
QS208EW	Religion
QS211EW	Ethnic group (detailed)
QS802EW	Age of arrival in the UK
QS803EW	Length of residence in the UK

Prior to the analysis an initial variable reduction was conducted to filter out variables which were not appropriate for a k-means classification. Variables with total populations below 10,000 were aggregated into broader groups based on their global regions of origin or removed altogether if they were considered too distinctive to merge.

3. Methods

The methodological approach for this study draws heavily on existing literature surrounding conventional geodemographic classifications (Harris et al, 2005), and most notably, the open source Output Area Classifications which are also produced exclusively from census data (Vicker and Rees, 2007). In a similar approach the data was standardised, tested for suitability and then clustered using the k-means algorithm, as displayed in figure 1.

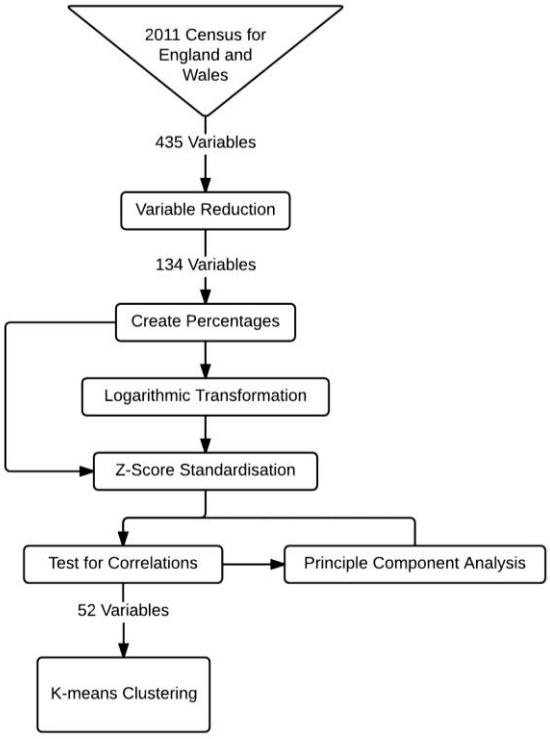

Figure 1 Flow diagram of methodological steps taken

The variables were standardised in order to reduce the effect of outliers across the dataset (Milligan, 1996). Much of the individual variables had positively skewed univariate distributions, largely due to low counts and a well-known tendency for cultural groups to cluster (Finney and Simpson, 2009). Therefore a natural log transformation for these cases was implemented, resulting in data set that is nearer to a normal distribution. In addition, a Z-score standardisation was applied so that each variable was presented on a common scale of standard deviations from mean.

Two steps were then undertaken to gauge the appropriateness of the remaining variables and to remove those which may unnecessarily skew the results. Firstly, a Pearson's correlation test was run between all variables to test for multicollinearity (Vicker and Rees, 2007). Of pairs of variables which correlated highly, either the smallest was removed or they were merged into 'other' groups if both variables were from the same census table and represented similar cultural groups, and did not correlate highly with a variable from a different table. Secondly, a Principal Component Analysis was undertaken to identify variables which may act erratically in the model (Rencher, 1998). Unstable cases were inspected and a handful were merged into broader variables or removed completely. Following these two steps many variables were aggregated and then retested.

In total 52 variables were selected for the classification (table 2). The variable with the smallest population out of the final selection, Russian language, represented over 67,000 persons.

Table 2 52 Variables used in the classification

Variable table	Variables used
Country of birth	China, Ghana, Hong Kong, Kenya, Middle East, Nigeria, Philippines, Romania, Somalia, USA, Other Central and Western Africa countries, Other EU accession countries, Other South and Eastern Africa countries, Other South-East Asia, Other Southern Asia
Main language	French, Russian, Turkish, African Language, East Asian Language, South Asian Language, West or Central Asian Language
Proficiency in English	Cannot speak English
Religion	Buddhist, Christian, Hindu, Jewish, Muslim, Sikh, No religion or not stated
Ethnic group	Afghan, African, Arab, Australian & New Zealander, Baltic States, Bangladeshi & British Bangladeshi, Black British, Caribbean, English/Welsh/Scottish/Northern Irish/British, Greek & Greek Cypriot, Indian & British Indian, Irish, Pakistani or British Pakistani, Polish, Sir Lankan, South East Asian, Other Eastern European, Other Western European
Age of arrival	0 to 4, 45 - 64
Length of residence	10 years or more, Less than 2 years

3.1 k-means clustering

The final 52 variables were clustered to create a composite classification using the k-means algorithm. K-means is an iterative allocation-reallocation method where the number of cluster groups (k) is predefined by the user (Harris et al, 2005). The approach creates distinctive cluster groups by attempting to minimise the sum of the distances from each case to their cluster centre based on the variable distributions. In its simplest form, the algorithm initially randomly seeds the cluster centres in a multidimensional space formed from the variables and every data case is allocated to its nearest centre. The cluster centres are then retested at the centroid of their current data allocation, and the process is repeated until the cluster centres cannot be moved as an optimum solution has been achieved (Harris et al, 2005).

With the intent of creating a classification with a relatively small number of clusters, a series of tests were run to help identify an appropriate number for the CELOAC. An 8 cluster solution was deemed to have the most appropriate average distance to the cluster centre and cluster size distribution.

4. The Cultural, Ethnic and Linguistic Output Area Classification (CELOAC)

The CELOAC consists of 8 culturally distinctive groups. Two groups combined comprise of just over 70% of output areas in England and Wales, both contain higher proportions of the White British ethnic group than the remaining population, with rates of 88.5% and 96.2% respectively. As the focus of this research is on foreign origin ethnic groups the two white British clusters have been merged for the remainder of this paper (group G).

The remaining groups account for neighbourhoods where the proportion of ethnic minorities is above the national average, and these are largely in metropolitan regions. At the broadest level, the remaining groups identify four main characteristics, neighbourhoods with higher proportion of Asian ethnicities (groups A and B), neighbourhoods with higher proportions of black ethnicities (group C), and neighbourhoods with higher proportions of White, Middle Eastern and East Asian ethnicities (groups D and E). There is also a large group which represents a cosmopolitan mixture of ethnic groups (group F), the White British ethnic group is also well integrated here. The average proportion of key ethnic groups from the 2011 Census for each CELOAC group are presented in table 3.

Table 3 The average percentage of ethnic groups by each CELOAC group

Ethnic Group	A	B	C	D	E	F	G
White British	43.29	22.50	33.14	53.73	42.93	72.15	93.01
White Irish	1.21	1.70	1.80	2.82	1.71	1.47	0.64
Other White	4.69	8.62	12.40	20.63	14.24	7.96	2.06
Mixed & multiple	3.43	3.81	6.64	4.88	4.53	3.49	1.30
Indian	9.84	25.14	3.01	2.53	5.55	3.06	0.73
Pakistani	20.23	9.36	2.85	0.78	3.46	1.42	0.34
Bangladeshi	5.87	2.83	3.10	1.10	1.96	0.66	0.14
Chinese	0.66	0.90	1.35	2.09	5.89	1.30	0.30
Other Asian	3.11	10.24	3.92	2.88	5.66	2.49	0.43
Black ethnicities	5.49	10.81	27.80	5.30	7.72	4.60	0.70
Arab	0.84	1.59	1.14	1.35	3.67	0.49	0.10
Other	1.25	2.38	2.69	1.85	2.60	0.77	0.17

The CELOAC groups have been labeled and mapped in figures 2 and 3. The labels only consider the key ethnicities which are overrepresented in each group and are only intended to aid interpretation for this study.

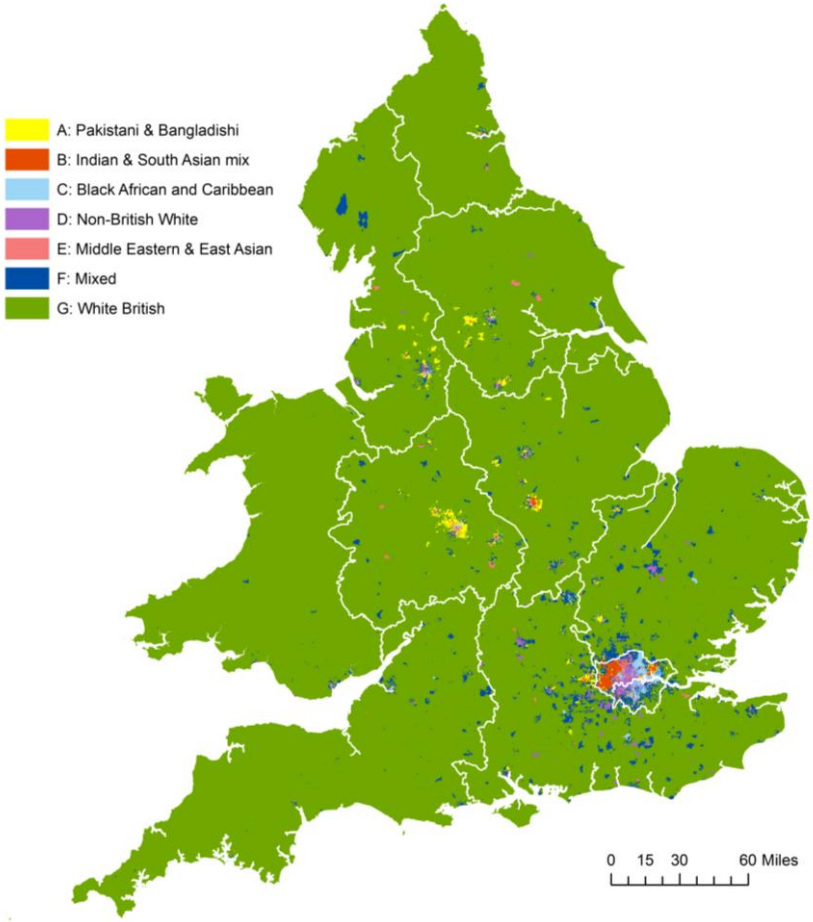

Figure 2 A map of CELOAC in England and Wales

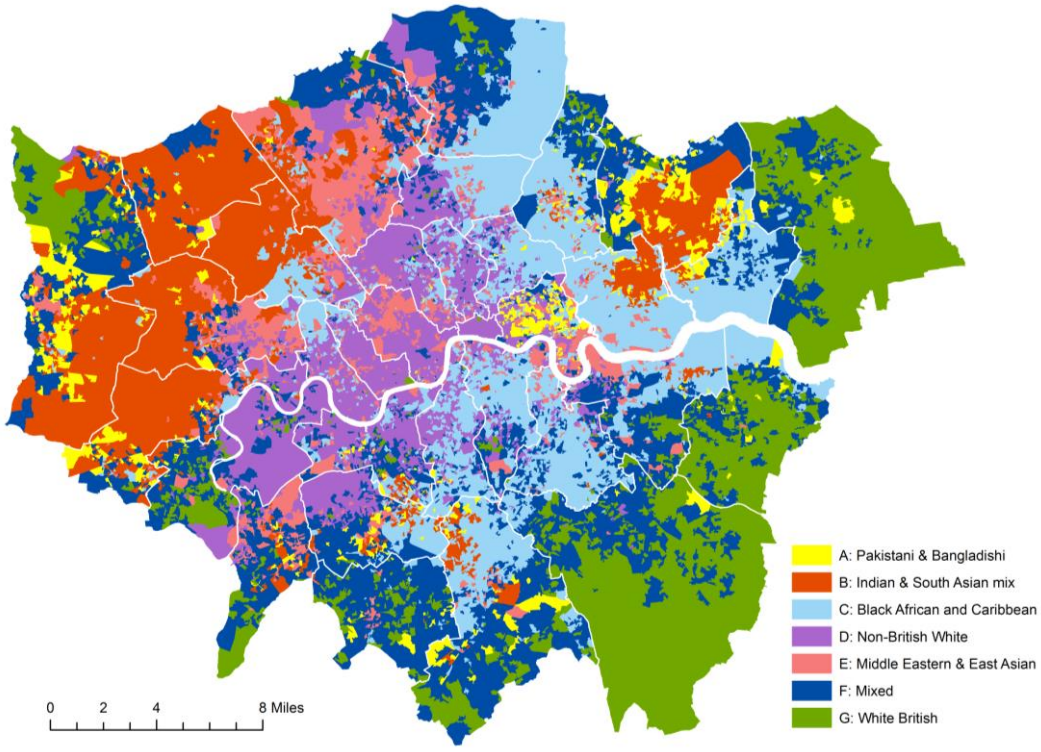

Figure 3 A map of CELOAC in London

4.1. Cultural heritage and ethnic origin food consumption

Ethnic and cultural identity can greatly influence consumption habits, and food especially (Kershen, 2002; Hamlett et al, 2008). The supermarket chain Sainsbury's provided the number of sales for six pre-selected grocery products by OA as recorded from their customer loyalty database, the data represented the total sales within a 52 week period commencing in May 2011. Each of the foods were chosen due to their distinctive cultural heritage with minority groups. As the total grocery expenditure per OA was not available, the data was standardised by the Census population. This data has been cross-tabulated by the CELOAC groups and the results are shown as location quotients whereby 100 represent an average penetration (table 4).

Table 4 Sales of six ethnic origin foods by CELOAC group

Group	Black Eye Beans	Chickpeas	Chinese Leaf	Ghee	Halal	Ogorki
A	215.85	80.99	81.49	250.52	163.20	122.03
B	472.69	111.15	137.82	601.02	711.46	312.35
C	305.82	120.15	131.56	277.24	598.45	286.26
D	218.65	230.27	229.67	216.07	413.44	356.12
E	202.49	122.70	260.88	229.66	402.39	286.00
F	151.65	136.29	151.40	151.36	109.55	184.47
G:	50.40	87.58	78.99	44.77	19.14	49.37

The results identified substantial variations in the consumption of ethnic origin foods across the CELOAC groups. Generally, while each of the products may sell particularly well in one of the 'minority' clusters, they will often sell better in other minority clusters too, reflecting the cosmopolitan composition of their populations. The results are especially compelling considering migrant groups may be less likely to patronise Sainsbury's stores than the White British population.

5. Conclusions

This study has presented an open-source output area classification of ethnic, cultural and linguistic characteristics for England and Wales and identified distinctive ethnic clusters. Whilst most rural and many suburban areas are homogenously White British in composition. The inner cities of larger, more globally connected urban areas are composed of a more heterogeneous mix of cultural groups. As cultural groups cluster and segregate themselves from dissimilar communities, spatial mosaics of culturally distinctive neighbourhoods occur in major metropolitan areas, and London especially. This distinctive residential geography exerts an associated spatial variation in the consumption of ethnic origin foods, as identified by the supermarket data.

Whilst there are several disadvantages from devising a discrete categorisation of neighbourhoods, the approach presented provides an insightful snapshot into the contemporary cultural and ethnic geography of England and Wales.

6. Biography

Guy Lansley is a Research Associate at the Consumer Data Research Centre, UCL, an ESRC Data Investment. His previous research at UCL has included exploring the temporal geo-demographics derived from social media data, and identifying socio-spatial patterns in car model ownership in conjunction with the Department for Transport. His current work entails exploring population data derived from large consumer datasets.

Yiran Wei is a recent alumnus of UCL. She studied the Geospatial Analysis MSc at the Department of Geography, UCL and specialised in geodemographics and ethnic clustering. She previously studied the

Environment and Development MA at King's College London and Environmental Science BSc at the University of Greenwich.

Tim Rains is a senior GIS analyst at J Sainsbury's plc. He also holds a GIS MSc degree from Birkbeck College and a Geography BA from the University of Plymouth.

References

Bulmer, M. (1996) 'The ethnic group question in the 1991 Census of population' in Coleman, D. and Salt J. (des.) Ethnicity in the 1991 Census. vol.1 Demographic characteristics of the ethnic minority populations London: HMSO

Finney, N. (2013) "How ethnic mix changes and what this means for integration." In van Ham, M., Manley, D., Bailey, N., Simpson, L., Maclennan, D (Eds) Understanding Dynamic Neighbourhoods, New York: Springer

Finney, N. & Simpson, L. (2009) *'Sleepwalking to segregation'? Challenging Myths about Race and Migration*, London: Policy Press

Hamlett, J., Bailey A., Alexander, A. and Shaw G. (2008) Ethnicity and Consumption South Asian food shopping patterns in Britain, 1947 – 1975. Journal of Consumer Culture 8(1) 91-116.

Harris R, Sleight P, Webber R. (2005) Geodemographics: neighbourhood targeting and GIS. Chichester, UK: John Wiley and Sons.

Kershen, A. J. ed. (2002) *Food in the Migrant Experience*. Aldershot: Ashgate.

Mateos P, Singleton A, Longley P (2009) Uncertainty in the analysis of ethnicity classifications: issues of extent and aggregation of ethnic groups. Journal of Ethnic and Migration Studies 35(9), 1437–1460

Milligan, G. W. (1996), Clustering validation: Results and implications for applied analyses, in Arabie, P., Hubert, L. J. and De Soete, G. Eds., Clustering and Classification, Singapore, World Scientific.

ONS (2012) 2011 Census, Population and Household Estimates for Small Areas in England and Wales, 23/11/ 2012. [Online] http://www.ons.gov.uk/ons/dcp171778_288463.pdf (Accessed 19th November 2014)

Rencher, A. C. (1998) *Multivariate statistical inference and applications*. New York: Wiley.

Simpson, L. (2013) What makes ethnic group populations grow? Age structures and immigration. Dynamics of Diversity: Evidence from the 2011 Census. Centre on Dynamics of Ethnicity [Online] http://www.ethnicity.ac.uk/medialibrary/briefings/dynamicsofdiversity/what-makes-ethnic-group-populations-grow-age-structures-and-immigration.pdf (Accessed 11th November 2014)

Stillwell, J and Hussain, S. (2010) Internal migration of ethnic groups in Britain, Chapter 5 in Stillwell, J., Finney, N. and Van Ham, M. (eds.) Understanding Population Trends and Processes Volume 3: Ethnicity and Integration, Springer, Dordrecht

Uskul, A. K. and Platt, L (2014) A note on maintenance of ethnic origin diet and healthy eating in Understanding Society - Institute for Social and Economic Research (ISER). Working Paper. [Online] https://www.iser.essex.ac.uk/publications/working-papers/iser/2014-03 (Accessed 9th November 2014)

Vickers D and Rees P. (2007). Creating the UK National Statistics 2001 Output Area Classification. Journal of the Royal Statistical Society: Series A (Statistics in Society) 170(2): 379-403