Applied Spatial Analysis and Policy (2021) 14:51–79
https://doi.org/10.1007/s12061-020-09343-6

Open Geodemographics: Classification of Small Areas, Ireland 2016

Burcin Yazgi Walsh[1] · Chris Brunsdon[1] · Martin Charlton[1]

Received: 22 May 2019 / Accepted: 12 May 2020 /
Published online: 7 June 2020

Abstract

Geodemographics is a tool to summarize the characteristics of spatial units based on socio-economic data. It has been used over several decades to classify the characteristics of areas based on the similarities in such data, generally working by identifying groups or *clusters* of similar areas. It has seen use in the academic and private sectors but mostly became popular by 1980s for market research purposes – this rise in use drew attention to many issues that needed to be addressed in the literature. The purpose of this paper is to provide a geodemographic classification for the smallest scale administrative units in Ireland (Small Areas) based on the latest 2016 census population data. A further aim is to support reproducible research by following the methodology of an earlier study for the 2011 Census and to bring up the opportunity to compare the cluster results between the 2011 and 2016 census data. The comparison output provides useful insights for policy discussions in the Irish context. One of the other focal points of this study is to bring some clarity regarding some criticisms of geodemographics from an academic perspective. The paper also argues for the use of open data, open source software and supports open analysis.

Keywords Geodemographics · Ireland · Open data · Open research

Introduction

Geodemographic classification is a tool that groups spatial units based on some dataset, typically comprised of socio-economic variables. The approach goes back a long time. It was first proposed to investigate social area analysis and factorial ecology. The works done by human ecologists can be counted as the early versions of geodemographic

✉ Burcin Yazgi Walsh
burcin.yazgiwalsh@mu.ie

[1] National Centre for Geocomputation, Maynooth University Social Sciences Institute, Maynooth University, Maynooth, Ireland

studies (Singleton and Spielman 2014; Webber and Burrows 2018). Booth's poverty maps for London prepared in the late nineteenth century can be regarded as arguably the first example of geodemographic studies. In these maps - which were focusing on the spatial social structure of the city - the streets of London were classified into seven different groups and each classification was given a description. A few decades later, Chicago School of Sociology developed models for urban structure (Park et al. 1925; Hoyt 1939). These charts were representing the social class through the spatial hierarchical distribution. Besides, these were the underlying techniques for the development of social area analysis as well as the factorial ecology (Singleton and Spielman 2014; Longley 2005; Webber and Burrows 2018). Booth's attempt of combining the information about people and where they live through the reports of school board visitors (nineteenth century), Chicago school's approach to urban structure (1920s), social area analysis (1930s) and factorial ecology studies (1940s) could be the underlying works for geodemographic approach. Early work in the USA at city level was undertaken by Shevky and Williams (1949) and Shevky and Bell (1955). Tryon (1955) applies cluster analysis to identify social areas within San Francisco. In the UK Herbert (1967) applies Shevky and Bell's approach to identify social areas in Newcastle-under-Lyme.

However, geodemographic classifications go beyond all of these techniques. There was another wave for the development of geodemographic studies around the 1960s regarding the technological changes, available data, and improvements in computational power. Even though they played an important role in enabling geodemographics to gain some popularity, it could be argued that widespread recognition of this topic came only in the 1970s. Notable in the UK is the work of Richard Webber, then at the Centre for Environmental Studies using 1971 Census Data – then available to researchers in machine readable form (Webber 1975; Webber 1977; Webber and Craig 1978). Webber's 1977 national classification of wards and parishes drew criticism from Openshaw et al. (1980). Openshaw later developed his own national classification at census enumeration district level (Charlton et al. 1985). Webber's work was taken up by the CACI consultancy and developed into a commercial product: ACORN (A Classification of Residential Neighbourhoods) (Webber and Burrows 2018). Experian used Webber to develop the competing MOSAIC classification. Parallel work in the USA by John Robbin led to the development of the Potential Rating Index by Zip Market (PRIZM) by Claritas in 1974.

We may also note the emergence of the k-means algorithm, and perhaps as important, computer code. Macqueen (1967) provides an early description. Hartigan (1975) describes the k-means algorithm in some detail, and follows this with code in the FORTRAN 66 language. Hartigan (1975) comments on the code "The ideal user is an experienced Fortran programmer who is willing to adapt the programs to his own needs". Hartigan and Wong (1979) provide, as Algorithm AS136, code in ISO FORTRAN, and make a suggestion on the selection of the initial k clusters. The existence of such code greatly reduces the effort to implement k-means for a clustering exercise with a large dataset.

The main type of data used in the classifications was generally derived from censuses of population but over time, mostly via commercial use, other datasets were gradually added to the pool of data used to provide geodemographic classifications. In geographical terms, categorization of areas initially started at city level and moved on to national level. The geodemographics approach has been applied in many different countries across the world (e.g. United States, United Kingdom, Japan, South Africa, Mexico, Spain and Ireland). As well as consumer market related studies (Mitchell and McGoldrick 1994; Goss 1995; Harris et al. 2005) there is also an interest for the improvement of public sector services through geodemographics approach, such as education (Tonks and Farr 1995; Batey et al. 1999; Singleton 2010; Brunsdon et al. 2011a), health (Brown et al. 1991; Openshaw and Blake 1995; Farr et al. 2008; Petersen et al. 2011) and public service delivery (Longley 2005).

Since it seeks to classify the characteristics of areas based on the similarities found in dataset, selecting the variables and the methods for the classifications are important processes and they may vary in different contexts. This was the main focus of the discussions by different scholars after the first attempts of geodemographics studies. Some critiques of the geodemographic approach started to arise in the 1980s. One of them was Openshaw (1983) who criticised the choices made during the classification process, such as the choice of number of the clusters and the method used for clustering. He argued that there is no possible classification method that would suit all purposes. Charlton et al. (1985) supported the criticism and also suggested an alternative approach for a better general-purpose classification may exist. Criticisms continued around these concerns especially related to geodemographics being used increasingly for market research purposes. One of the main discussions - more in the recent decades – focussed on geodemographics being a 'black box' process. The variables, the methods and the clustering techniques used and the criteria behind their selection are all hidden. There are some classification groups obtained, some statements and assumptions made as a result of these processes but the background information that forms these results is unknown to users.

Typically, geodemographic classifications are undertaken by running a clustering algorithm based on derived census variables and possibly other administrative data and survey results in an attempt to reflect the socio-economic characteristics of the entire study area. The inclusion of these other datasets promoted some discussions around the topic of their reliability (Longley 2012). For the clustering algorithm, mostly the k-means approach was applied to form the classifications but there are some alternative approaches that could be substituted depending on the number of the data points, robustness to outliers and the amount of noise in the data (Feng and Flowerdew 1998, 1999; Grekousis and Hatzichristos 2012; Son et al. 2012; Fisher et al. 2014; Fisher and Tate 2015). After selecting variables and applying a clustering algorithm, the last stage of a geodemographic study is to label the groups identified by the algorithm. Often these labels are intended to be distinctive and concise, such as: "money & brains"; "big fish & small pond"; "the underprivileged"; "metro singles" and so on. It can be argued that the idea behind labelling the classifications goes back to Booth's (1902) early representations of London poverty. He was classified the streets of London into seven different categories – see Table 1.

When the size and the content of the area covered by Booth's map and the time it was produced are taken into account, the motivation to label the categories (and hence the associated areas) may be understood, however given today's awareness of the

Table 1 Booth's social survey classifications

Classification	Lowest Class	Very Poor	Poor	Mixed	Fairly Comfortable	Middle Class	Upper-Middle and Upper Classes
Description	Vicious, semi criminal	Casual, chronic want	18 s. to 21 s. a week for a moderate family	Some comfortable others poor	Good ordinary earnings	Well to do	Wealthy

complexity of socio-spatial conditions, we believe we need to exercise caution rather than over-simplify or distort the results of analysis by applying such labelling. Therefore, the groups are deliberately not labelled in this study. All these concerns mentioned above are the main drivers of this study. In order to deal with the hidden 'behind the scenes' criticism, in this study we are bringing an openness perspective for every stage of the study. The data, software and analysis used in this study all follow an open approach. With the provided links and information, the whole process can be reproduced, applied to other studies and checked to understand more or built upon it.

Even though the choice of underpinning methodology can have an important effect on the outcomes of a geodemographic analysis, once this is accepted it is still believed to be a helpful tool. This can be supported by Charlton et al.'s study (Charlton et al. 1985) where they state that "the results provide an exploratory spatial description of small-scale areal census data for an entire country". Another comment on this is by Singleton and Spielman (2014) who stated that a broader usage of the approach, "could support the development of a more robust theory of socio-spatial structure". Since more areas of application are included in geodemographic studies in recent years, experiences of its use in different contexts is growing increasingly. This promises more comprehensive discussions and yet more diverse uses of the technique.

The purpose of this study is to classify Small Areas in Ireland based on the 2016 Census Data, using a fully open geodemographics approach. The supporting motivations are to compare the new classifications for Ireland with the previous version; to address the outcomes of this comparison for policy actions; to attract attention to general criticisms in geodemographics studies and to discuss how to overcome these issues. With this framework, the study shares some similarities but differs in some respects from other recent open geodemographics studies (Charlton et al. 1985; Singleton and Longley 2009; Adnan et al. 2010; Brunsdon et al. 2016). Each of these motivations will be explained in the relevant sections in the rest of the paper. After providing some background information the classification results will be presented. Then the attention will be directed to the comparison of the two different sets of classifications based on the two latest Census data of Ireland. The paper will be finalised with some discussion on geodemographics and concluding remarks about open research.

Background

In this paper we create a general-purpose classification of the small areas in Ireland based on the Central Statistics Office's 2016 Census of Population. A new system of small areas was developed in 2010 by subdividing the Electoral Districts (EDs). The main criteria for devising the new spatial units were average population changing from 80 to 120 households. The output was a network of 18,488 discrete Small Areas that were used for compiling results from the 2011 census. Some further changes occurred over the intercensal period such as amalgamating, splitting and adjusting borders of some small areas which brought the total number of small areas to 18,641 for 2016 Census. The 2011 Census was the first time small areas were used. While the total number of Small Areas has increased it is still possible to reconcile the spatial units for 2011 and 2016 and therefore to examine changes between the census dates. Having comparable set of spatial units offers a better spatial data analysis approach especially for studies like geodemographics. The spatial unit used for classifications is also an important issue since data availability generally constrains the limit for this criterion. Generally, the spatial units that are available to analyse can vary based on the variables that want to be included in the study. Data availability for different variables and spatial units related to those variables can change for censuses of different countries.

The Central Statistics Office's (CSO) 2016 Population Census, which was obtained via their open portal includes more than 700 count variables. Selection of the appropriate variables for classifications is the vital part of the analysis. Since there is not a rigid structure for deciding variables, following the previous examples provides a useful background. The general approach is to use the key dimensions in order to cover every geographical and demographic aspect with the analysis and to make the analysis manageable to run. All the researches and studies done so far follow a data reduction method. It is also important to have the correct variables that can represent the socio-spatial characteristics of the study area. These and many other variable choices for geodemographic studies helped to form the structure for 2011 Census classification for Ireland completed by Brunsdon et al. (2011b) and Brunsdon et al. (2016). Webber's (1977) suggestion of variables followed by Charlton et al. (1985) were some of these. This paper is following the methodology applied by Brunsdon et al. (2016) for Ireland's Population Census of 2011, 40 derived variables are selected from the dataset. These variable derivations are identical to those used in the 2011 Census-based study and those were selected to mirror, as far as is possible with Irish data, variables used in the ONS Output Area Classification in the UK. Using identical variables in this way provides some scope for comparison studies. The chosen variables fall under the ten different themes of this dataset: Population: Sex, Age and Marital Status; Migration, Ethnicity, Religion and Foreign Languages; Families; Housing; Education; Principal Status; Motor Car Availability, PC Ownership and Internet Access; Commuting; Disability, Carers and General Health and Industries. For the analysis they are organised under the five different categories of: Demographic; Household composition; Housing; Socio-economic employment and Internet access. The full list and the descriptions of the variables used in the study can be found in Appendix 2.

As well as the set of variables chosen, the workflow of the geodemographic classification followed is parallel to the 2011 Ireland classification study. The similarity in the variable and the methodology choice provide the opportunity to compare the classification outputs for different census years and build a stronger policy framework. The other motivation in doing so is to support open science and reproducible research. In Brunsdon et al. (2016) geodemographic study running a principal component analysis (PCA) is the first step, followed by Partitioning Around Medoids (PAM) clustering and then computing the mean variables of clusters. The outputs are represented through different visualisations including heatmaps.

Similar to those workflow steps, in this paper after organising the variables a principal component analysis was run. This is done to apply a linear transform to possibly correlated variables in the dataset into a group of uncorrelated components. These are used to create a distance matrix (based on Euclidean distances). By using uncorrelated components, distances are based on underlying independent characteristics driving the census variables, rather than on the variables themselves, which in their raw form may under - or over - represent certain characteristics. Principal component analysis shows the proportion of each component's contribution to variance, and so it is possible to investigate the cumulative proportional contribution. It was noticeable that first 14 components explain the 80.0% of the variance in total in our study. Note that the use of PCA here is felt necessary, as the clustering methods used rely on 'distances' between observations. One thing that PCA does is to concentrate correlated variables into individual scores. If several 'raw' variables are correlated, then a distance metric based directly on these counts distinctions between the correlated variables multiply – i.e. if variables x and y are correlated, then the distance between them is measured as (×1 – ×2)^2 + (y1 – y2)^2, and effectively counts as a greater contribution to overall distance than if a single variable reflecting the underlying characteristic of x and y were used. Since the selection of the variables used for the classification was not based on working with uncorrelated (or at least not highly correlated) variables, the PCA approach was preferred here.

Thus, this number of principal components is used as an input for the next stage of the cluster analysis. Next step then is to use Kaufman and Rousseeuw's (1990) Partitioning Around Medoids (PAM) algorithm with the principal components of the variables to identify the clusters in the data. Following the approach by Brunsdon et al. (2011b), PAM is preferred over k-means since PAM is more robust to noise and outliers, compared to k-means. This is because PAM minimizes the absolute distances between cluster members and cluster centroids, not the squared distances as k-means does. This also can mean that the PAM algorithm is less likely to form very small clusters. Small clusters are often formed when individual observations, or small groups are notably distinct from the remaining observations. K-means clustering attempts to minimise the squared distances from individual observations to cluster centers, and this puts a higher weight on unusual individuals or small groups being further away from the centroids. Thus, in minimising, it tends to assign small groups to their own clusters, which is less desirable here. As a result, the PAM algorithm – which minimises *absolute* distance is used instead.

To decision on how many clusters are in the dataset is another important issue. Both k-means and PAM require that this number that is supplied when the algorithm is run. Choosing the most appropriate number is left to the analyst. In order to have a better understanding, the algorithm is run using a range of numbers of clusters, and the smallest cluster size versus the number of clusters is plotted. As a result of this plot, it was noticeable that after some point the smallest cluster is very small. In particular this occurred when the number of the clusters exceeded 8 (Fig. 1). Based on this

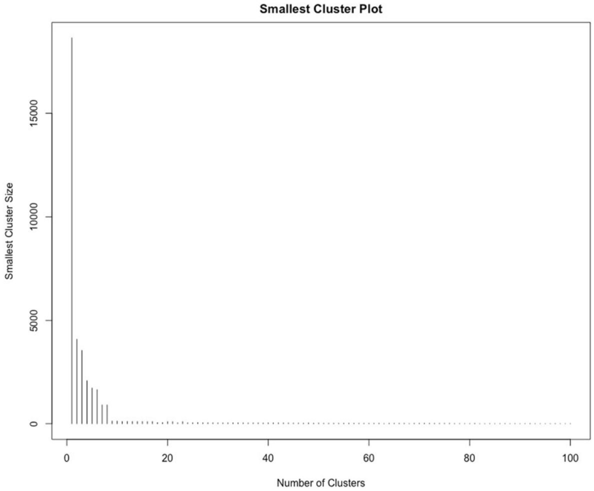


Fig. 1 Number of clusters

information, a choice of 8 clusters was made, giving the final allocation of the spatial units into different clusters. Note that here, even when using PAM – after 8 clusters, very small groups start to appear. This suggests that after this point the algorithm is simply organising the clusters into the broader groups that are required, and then harvesting out the more outlying observations within these clusters.

For the reproducibility purposes as well as to achieve the basic classifications for 2016 Census Population Data for Ireland, the PAM algorithm - which provides a crisp classification - was followed. In a *crisp* setting, the spatial unit can only belong to one group. This can be contrasted to a *fuzzy* approach where spatial units may be viewed as 'partial' members of more than one cluster. Although being generally useful, the authors feel that alternatives to the crisp approach are worth exploring. Thus, in further planned work, authors will be applying fuzzy clustering as an alternative approach to the same dataset in order to further understand the structure in the data and compare the results of the two approaches to cluster creation.

To interpret the clusters and identify each group the appropriate techniques suggested by the previous study were applied. Accordingly, the cluster mean for each variable was calculated and plotted in a heatmap to help to identify the characteristics of a group. The variables are represented in the rows and clusters in the columns (Fig. 2) in a heatmap plot below. Higher values of a variable can be identified by darker shade of blue and low values by brown within each cluster. The row and column arrangements are computed so that similar clusters (in terms of mean values of variables) are near each other, and similar variables (in terms of their levels in clusters) are also near each other. Even though this study does not have any intention to identify the cluster group for labelling it is still useful to see the distribution of the variables for different clusters and identify clusters that are similar to one another. Interpretation of this heatmap will be discussed in detail in the next section.

All stages including the visualisations as well as all prior analysis steps of this study are achieved using the open source R language and different packages/extensions of

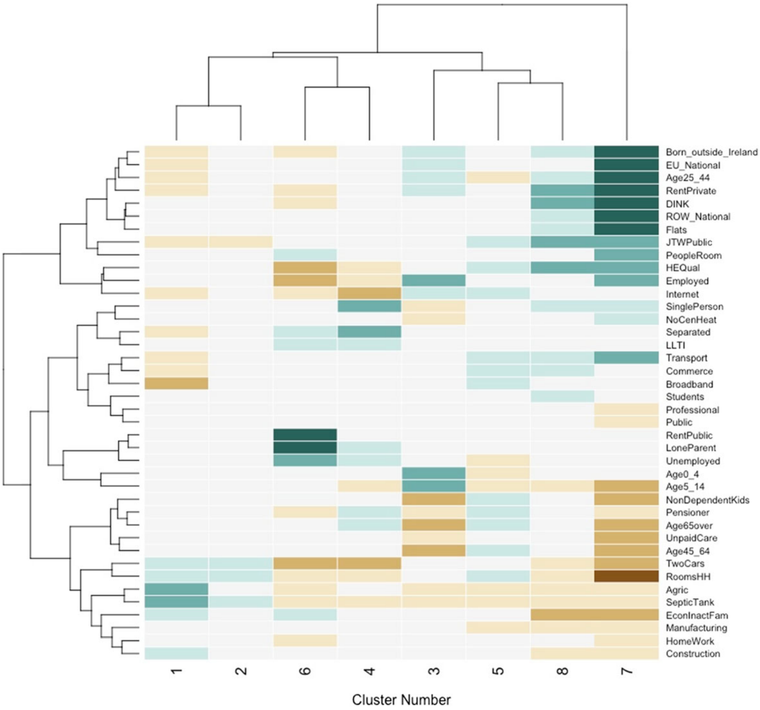


Fig. 2 Heatmap of PAM algorithm cluster outputs

this programing language. During the analysis procedure for the classifications the specific R packages of ‘plyr’ and ‘cluster’ are used. For supporting the test stages, the graph production ‘ggplot’ package is applied. For some of the outputs like heatmaps the ‘RColorBrewer’ package is preferred. For several static map alternatives, the package ‘tmap’ is adapted. For interactive visual outputs ‘ggplot’, ‘leaflet’ and ‘shiny’ are integrated. Through all the stages, documentation is kept in R Markdown which helped to publish a quick, online version of the outputs which gives the opportunity for this research to be reproducible which can be found at https://www.rpubs.com/burcinwalsh/343141.

Results of Classification

Each variable’s cluster mean is computed and presented in a heatmap form above (Fig. 2). Dark green and blue represent the higher average values of that variable in the cluster and dark brown on the contrary represents the lowest values. The arrangements of rows and columns make the interpretations easier since similar characteristics will be close to each other in the plot. It is shown that overall there are 8 different groups and the general characteristic of these clusters transform from urban to rural moving from left to right. Variables like septic tank and agricultural employment are both playing an important role for this interpretation. Each cluster has their own strong features that would differentiate them from the other clusters. However, only one of the clusters (Cluster 2) is not much different than Cluster 1.

It is worth mentioning that one of the groups stands out from the others. Cluster 7 has the highest values of people between age of 25–44, people renting privately, living in flats with an international background and families with double incomes and no children. Cluster 1 on the other hand, has the highest scores in septic tank usage and agricultural activities. Cluster 6 shows a different pattern with its higher unemployment rates, lone parents, separated families and renting public houses conditions. Cluster 3 shows a high proportion of children between the age of 0–14. Cluster 8 has some similarities with Cluster 7 but differs from it with high student population. Cluster 5 differs from the rest of the clusters with its population of over 45 years old and having mostly non-dependent kids. For Cluster 4 the dominant characteristic is people living on their own.

If we examine the general properties of Cluster 1, we can notice that broadband/ internet connections are at a low level. People born outside of Ireland or Europe are also less in this group compared to other groups in Ireland. The other variables that have high value in this group except than septic tank and agricultural activity are number of cars owned and number of the rooms per household, economically inactive families and construction as an economic activity. As mentioned earlier Cluster 2 has some similar characteristic to Cluster 1, such as high values in septic tank, car ownership rates and number of the rooms per household as well as low values in journeys to work with public transport. The main difference is this group doesn't have a strong evidence in agricultural activity. If we follow the order from rural to urban next cluster to examine is Cluster 6. Besides the earlier mentioned characteristics, the high values are around economic inactivity, low health conditions, separated families and number of people per room. Some of the variables with low values for this group are, higher education qualifications, employment, internet connection, car ownership. Cluster 4 as a group reflects mostly singlehood life. This includes separated people, lone parents and single people as well as people over 65 years of age. The number of pensioners is high, and the health conditions are low. Within this group higher education qualifications are at a low level as well as the internet connection rates. The next group which is Cluster 3 is consisting employed young families with kids until the age of 14. On the other hand, Cluster 5 is a composition of more mature families mostly with non-dependant kids. The residents are generally over the age of 45 with higher education levels and working in transport and commerce. Public transport for journey to work is a preferred option. Cluster 8 is another group that is formed by younger generation including students, singles and couples. High education levels, renting privately and living mostly in flats are the other featured characteristic for this group. Cluster 7 differs from Cluster 8 and the rest of the groups by its dominant characteristics of varying ethnicity within the age of 25 to 44. This group constructed by young internationals with high employment rates, higher education backgrounds. Double income and living in flats are the other properties that can be noticed for this group. In general, each of these eight groups have some unique characteristics which combine to produce distinctive spatial distributions. In order to explore the geographical patterns, all of the groups are mapped, some of these visual outputs are shown below (Fig. 3). These map outputs not only show the spatial distribution of the classifications, they also provide some understanding on the spatial concentrations of each cluster. Figure 3 is a collection of maps showing the classification for each small area in Ireland for the selected groups of 1, 2, 3, and 5 which are the groups with the highest population

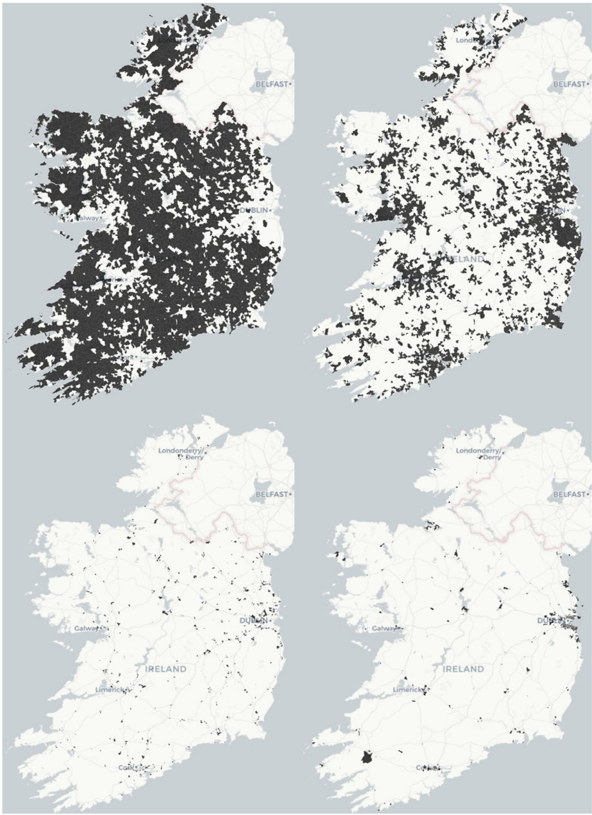
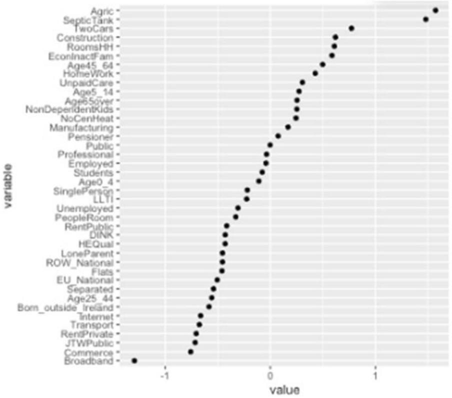


Fig. 3 Cluster outputs (Top left: Group 1, Top right: Group 2, Bottom left: Group 3, Bottom right: Group 5)

percentages. The spatial concentration - especially the ones for Group 1 and 2 – gives the reader some idea about some of the characteristics of that group as well. The gaps seen around the cities narrow down as the view changes from Group 1 to Group 2. It is also visually noticeable from the spatial distributions presented on the maps that Group 1 has the highest number of small areas included in that category. However, Group 2 has the highest population percentage among all the clusters (20.8%).

To support the understanding of each group, Cluster Characteristics Cards are prepared for each group. They include the spatial distribution of the classification in map form, as well as some descriptive explanations and statistics of each group. These cards and more detailed information about each classification can be found in Appendix 1. Figure 4 represents the part of these cards which focuses on variable score distribution plot and some descriptive values in a table format for each cluster. The representative variables for each group - with the highest and lowest 5 values – is shown and mean values of these variables are comparable with Dublin means.

The next visual (Fig. 5) shows the classifications from selected groups by focusing on Dublin City, Cork and Galway in order to examine the pattern of the classifications in the city context of Ireland. The groups 7 and 8 are selected for this representation since the highest percentages of spatial distribution of these



Variable	min	mean	max	Dublin mean
Rooms HH	4.19	5.76	7.02	4.61
Septic Tank	1.72	2.57	100.00	1.58
Two Cars	12.20	56.12	83.13	32.03
Agriculture	0.00	16.04	53.33	0.29
Construction	0.00	7.13	33.33	3.74
Rent Private	0.00	5.68	26.09	24.32
JTW Public	0.00	1.15	10.74	20.74
Commerce	0.00	17.69	45.45	28.53
Transport	0.00	4.80	20.18	12.28
Broadband	27.59	75.65	100.00	95.36

Fig. 4 Cluster properties (Group 1)

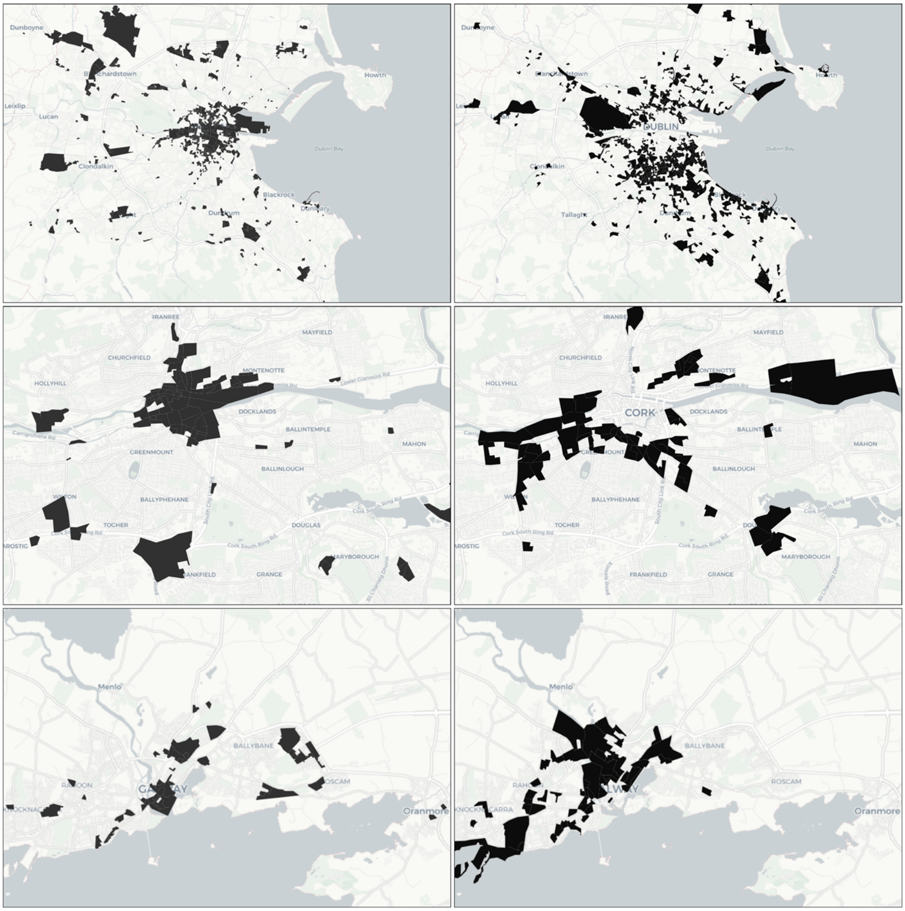


Fig. 5 Geodemographic distributions in cities (Top to bottom: Dublin, Cork, Galway; Left: Group 7 - Right: Group 8)

clusters are occurring in city borders. With this study here some interactive tools are added to the visuals to make exploration of spatial distribution of the classifications easier and efficient for the user – although this cannot be represented explicitly in the paper. Therefore, some of the zoomed in versions of these maps are presented here especially for the groups that are mainly visible in the city context.

Comparison of Classifications - 2011 Census Versus 2016 Census

In this section of the paper a comparison of the classifications based on the 2011 and 2016 Census data for Ireland is provided. To assess this, we will be focusing on several different perspectives including the background details for each study, social structure, spatial pattern and the reproducible research perspective. The potential reflections of all from a policy perspective will be discussed in the following section. In order to compare studies, the important issue is to have matching spatial and other datasets and also similar methodologies. Spatial and social attributes as well as the methods applied should be overlapping in each study. In their study Brunsdon et al. (2016) analysed the geodemographic structure of Ireland based on the 2011 Census results. While doing that they also provided an open outline for further studies. Census data for Ireland is open data but the authors made the code and steps to run the analysis publicly available as well (Brunsdon et al. 2011b). Besides all the other advantages the authors believed in reproducible research approach, one that they highlight in their paper was 'future-proofing' (Brunsdon et al. 2016). This study which is focusing on the latest possible cluster structure of Ireland is following the steps of the previous study on purpose, not only for the reproducible research objective but also to facilitate a comparison of the outcomes.

After the introduction of small areas into Ireland's national statistical boundary system, the 2011 Census is the first dataset to have data corresponding to Small Area scale. Therefore, the 2011 Census classification study was able to use this spatial scale and there were 18,488 Small Areas in total. In 2016 small area was still existing as a spatial unit in the census but the latest number for this statistical boundary was increased to 18,641. Since the change was mostly in quantity, splitting or amalgamating a few small areas and not massively in the geographical forms of the units, it was still possible to carry out a comparison. The other component to check for is the variables. Although the same set of variables were available in both census dataset, some small adjustments was needed for the latest study. This was mainly because first, for some of the variables the existence of 'not stated' column should be taken into account. The formulations for the variable calculations that were provided with the 2011 classification study were modified regarding to this issue. Second, some of the formulations were resulting in minus values with the 2016 census variables, those formulas were adjusted to the new variable setting. After all these small transforms the same 40 selected variables for 2011 classification were applicable for the 2016 census geodemographics study. Data quality was another topic to deal with the new dataset. It was discovered that some of the small areas had missing information which was misleading the analysis and the results. These small areas were excluded from the 2016 study but there was not any recorded decision like that for 2011 study. Five small areas that were extracted from the study were the areas where student accommodation facilities were concentrated. The effect of this on the classifications will be discussed further in this section. The method for clustering applied by the earlier study was a PCA approach followed by a PAM clustering algorithm. After a few performance tests with other clustering algorithms the same method was used for the new grouping for Ireland. After all these similar courses of actions, the attention can be given to the general outputs of these two geodemographics studies for different years.

In the geodemographics study completed with the 2011 Census population there were 16 clusters that were organised under 8 main groups. In the 2016 Census geodemographics study the final number of the clusters were 8. One of the main differences of the latest study from the earlier one as well as the rest of the previous geodemographics studies is the 'not labelling' approach. Different aspects of this approach will be highlighted in the next section of this paper. Since there are no 'catchy short labels' proposed for 2016 classifications the group will be represented by their group number only. When the general characteristics of the main groups of 2011 and 2016 are examined you can mostly see an overlapping distribution. For the main groups from the 2011 study, 'Rural communities; Mature suburbs; Struggling peripheries; Young families; University; City centre diversity', their characteristic reflections can be seen in some of the Groups from the 2016 outcomes. The other dimension is the groups that are existing in the 2011 output but not in the 2016 one such as the groups represented under the names of 'Urban poor' and 'Commuterland'.

On the other hand, there are other some interesting points to mention and investigate. One of these differences is related to one of the subgroups of the 'rural communities' in the 2011 study which is named 'Mature rural communities'. This is a group consisting of people living in the rural setting but not actively involved in agricultural activity. In the 2011 study one of the other highlighted characteristics of this group is the age range which is between 45 to 64. When we look at today's conditions this group is not necessarily represented by this age range. This is a finding that could be explained with the existing conditions recently that are experienced in Ireland. The residents that are not active in agriculture in rural areas are not only over 45 s anymore. Rural areas, especially the ones close to the town or village centre are some of the desirable destinations for the younger generation too. The other one is in 2016 students don't create a strong distinctive group as they do in 2011 classifications. There might be several reasons for this outcome but one of them is possibly the small areas which were excluded from the study. Removing those areas from the analysis could contribute to this outcome, since these were the areas where students were densely accommodated. The other potential reason is the new way of living – shared accommodations. Sharing is not necessarily an activity among

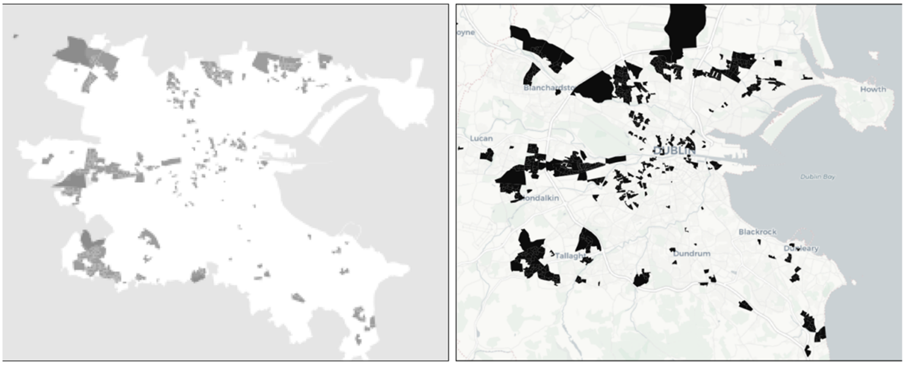


Fig. 6 Comparison of 2011 and 2016 census classification results (Left: Struggling peripheries - Single parent urban deprivation 2011, Source: Brunsdon, C.; Right: Group 6–2016)

the same social status. Cross status sharing accommodation is getting more popular especially in the urban settings.

Within this general framework, the next focus is the differentiation in the geographical structure of the clusters. For this purpose, one of the groups from the 2011 census classification which is labelled as 'Single parent urban deprivation' is examined. The characteristics of this cluster and the Group 6 from the 2016 census classification share similarities. When we look at the spatial distribution (Fig. 6) in Dublin city, it is easily recognisable that the small areas categorised under this group still keep their characteristics. Furthermore, the footprint is expanding among the highly spatially contagious locations. It is noticeable that the characteristics are carried to the neighbourhood boundaries. This situation is not particular to Dublin city. It is valid for the other cities of Ireland as well, such as Galway and Cork.

Lastly, it will be useful to mention the comparison remarks from a reproducible research perspective. What was provided with the 2011 census based geodemographics study was excellent. It not only led researchers for further studies; it gave an opportunity to apply the analysis as well as use it and build upon it. After going through this process and having the full experience, it can be stated that the basic principles of the previous study will be followed, and each step will be kept open. The only difference that the newer study will bring up for the reproducible research will be providing a documentation for the whole procedure.

Discussion

This paper started with the motivation to examine the latest geodemographics structure of Ireland based on the latest available Census population data. While doing so, the main drivers were to add some more value to geodemographic studies in the Irish context; to make geodemographics studies more valuable for a wider spectrum; to support openness for discussions on the general criticisms over geodemographics; and to support reproducible research in every possible extent.

In order to ensure additional value in geodemographic studies for the Irish context, this study is not only providing the latest socio-spatial classifications for the country, it is also comparing the two different set of geodemographic outcomes temporarily and spatially. There are several advantages to this approach, one of which could be expanding the use of these applications to a wider audience. Updated and extended forms of information can be crucial in geodemographics studies especially for the policy perspectives of countries or regions. Geodemographic classifications are known as strong tools for politics, marketing, service distribution, etc. Hence, the value of these studies is not limited and should not be limited only to marketing and advertising agencies. The social structural changes of some specific geographical destinations from year 2011 to 2016 has a great importance for every level of authorities in charge of policy making. The outcome is not necessarily only important for a theoretical development but also in terms of practical actions. The influence of geodemographics study can cover a big range from planners from local authorities to decision makers at every level.

Specific to this study's findings, attention should be given to the expanding borders of low economic and health conditions in inner cities especially from a policy perspective. The distribution of supporting services and locational choices for essential amenities can play an important role to lead this spatial and social transformation. One of the other findings which is again important from a policy perspective is the formation of the rural settings in the recent years. Rural areas are attractive destinations not only for mature families or for agricultural activities anymore particularly in Ireland. In the Irish context, non-affordable house prices and high demand in urban areas are pushing the citizens to live in the outskirts as well as in the rural areas which are in a close proximity to an urban centre. Potential policy implementations to deal with this situation will differ based on the national, regional and local scale strategies as well as the economic strategic plans of the country. In any case, it is an important outcome of this study to point out as a reference for potential actions. For further policy development, the geodemographics studies can be useful as a base and in combination with other studies can be beneficial for a multi-dimensional evaluation. There are already many examples in Ireland of the application of outputs from similar types of spatial classifications in the preparation of national and local development strategies. For example, the preparation of the National Spatial Strategy in 1999/2000 was guided by the detailed demographic analyses and several county strategies were guided by the framework provided by the national classification.

The other remark that we want to stress is the criticisms that geodemographics studies face and how to overcome some of these issues. In general, these criticisms can be categorised under four different titles. First, the whole process of geodemographics studies being a 'black box'. The steps followed in most of the studies are closed and hidden. In simple terms what variables and what sort of algorithms are preferred for that particular study is a mystery. Briefly speaking that is why open approach is addressed in this study. Second is the procedure for the evaluation of the classifications. Some of the researchers are concerned that most of the time geodemographics are automated systems that are done by statisticians or software developers. The concern is then these applications ending up missing useful insights. For this purpose, this paper is a product of a team effort, consisting of individuals coming from different backgrounds with an aim of creating a multi-perspective approach for geodemographics studies. Third, the algorithm selected for the purpose of the cluster study. There are several different choices and decision on the appropriate cluster method is left to the researcher. In this paper we are applying a hard clustering method which assigns each small area into one cluster. As an alternative, soft clustering algorithms are being used in geodemographic studies which give the flexibility to a cluster to be represented by membership values and not necessarily assigned to a one particular group. The potential advantages and the disadvantages of these techniques in geodemographics studies is the topic to explore in our further studies. The last one is the practice of naming the clusters which is one of the other focal points of this paper. "Some academic researchers critized the absence of any substantial theoretic basis to the taxonomy and were uncomfortable with what they felt to be subjective, qualitative descriptions of the categories that emerged" (Webber and Burrows 2018) and some were concerned by the possibility of inferred social discrimination generated by these labels. We believe that as academics we can bring new insights to the labelling. In order to do that we are not giving any labels to any of the groups on purpose - to start with. Non-experts can build some sort of a language to represent their judgements and this language can refer to the social groups based on where they live. This was discussed in the book of Webber and Burrows (2018) and some of the examples given by the authors to this category were 'the Notting Hill Set' or 'East End Hipsters' or 'Hamstead Intellectuals'. If non-experts with a lack of evidence-based knowledge would make these stereotypical judgements, then the question is why with some evidence-based knowledge we - as academics - should still keep putting people in stereotypical groups. Going through some of these examples - without mentioning the classification program they are belonging to -such as "Symbols of success; Elderly needs; Liberal opinion; Clever capitalists; Rising materialists; Business class; Asian identities; Meals on wheels; Bank of mum and dad; Families with needs" can provide some kind of a frame to understand the extent of the issue. As Webber and Burrows (2018) observe "catchy enough to be memorable and specific enough for the cluster to be distinguishable from others" may not be what we necessarily need in academic geodemographic studies since they should differ from commercial uses or advertising purposes. Labels constrained to a maximum of 20 characters just to be memorable can cause unnecessary generalisation as well as discrimination. More discussions on this issue in geodemographics studies is another crucial part of our further work.

This study does not only use reproducible research to achieve something new, but also used it in order to compare two studies from different time periods. In addition, it used the previous study experiences and built upon it. With this study a new form of visualisation tool is tested and added to the open code directory for geodemographics studies (Yazgi Walsh et al. 2017). Rather than a static representation, a dynamic interactive approach is adapted to the visual representation of the clusters in Ireland. The aim here is to provide the user with an easier means of engagement with the results. There are several different interactive interfaces included in the process. Querying and hovering events are the initial steps applied. Later on, some of the interactive visualization tools such as dynamic linking and brushing are illustrated through a spatial map of classification groups of Ireland. Although they cannot be presented as part of this paper, the code and those interactive maps are available through RPubs open portals.

Concluding Remarks

This study is a demonstration of a reproducible research practice. It also is a good example of supporting open research and open science approach by making every task in the study publicly available. As well as providing the latest socio-spatial pattern of Ireland, it is also comparing different studies to bring up a policy discussion based on the geodemographic dynamics in Ireland. Last but not least it is challenging academics to rethink and consider dropping the labelling practices in geodemographics applications.

A reproducible and open approach to research aims to make research auditable and adaptable, allowing future researchers to build upon it. It has to be also mentioned that there are challenges in open research. As well as generally known concerns like copyright issues, confidentiality and quality (Stodden et al. 2013) there are some other difficulties that researchers can face. Even when the study is done for the same geography, variable definitions or spatial borders of areas can still change over time. All these and other dynamic entities within the analysis should be examined carefully and adjustments made (also reproducibly) to allow for the new situation. In this study the comparison was based on a national framework but if researchers want to use or compare studies based on a different country framework, as mentioned by Brunsdon et al. (2016), in that case specification of the dataset and the variables should be carefully taken into account. There are other precautions that have to be taken over data which could be specifically related to geodemographic studies - particularly ones using open data. It is important to emphasize the six key issues as mentioned by Kitchin and McArdle's (2016) epistemology: scope and access, veracity and validity, usability and literacy, uses and utility, and ethics. Open data providers should also be focusing on the quality of the data as well as the number of the dataset that are publicly available.

One of the main issues that this paper wants to highlight is openness; supporting the use of open data and open source software and promoting open research. With that approach, this study also encourages updating these classifications in further census years and enrich the comparison to feed into the policy decisions. Reproducible steps will eventually increase the level of openness and help to uncover frequently unreported facts and procedures behind the creation and analysis of data. This might be useful as a bridging effect between the different understandings of social science disciplines and could bring more collaborations into existence. The authors would like to believe in geodemographics should be a "common language which might stimulate greater cross-fertilization of insights between different professional discipline" as asked by Webber and Burrows (2018).

Acknowledgements This paper has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number 15/IA/3090. We would also like to thank the two anonymous referees for the helpful comments on the earlier version of this paper.

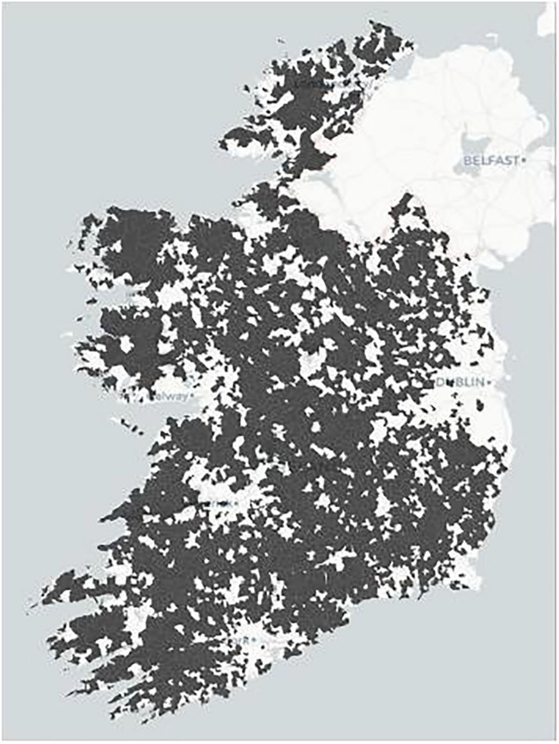
Compliance with Ethical Standards

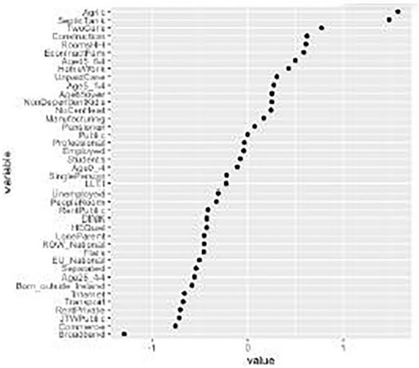
Conflict of Interest The authors declare that they have no conflict of interest.

Research Involving Human Participants and/or Animals This article does not contain any studies with human participants or animals performed by any of the authors.

Appendix 1 – Cluster characteristic cards

Group 1

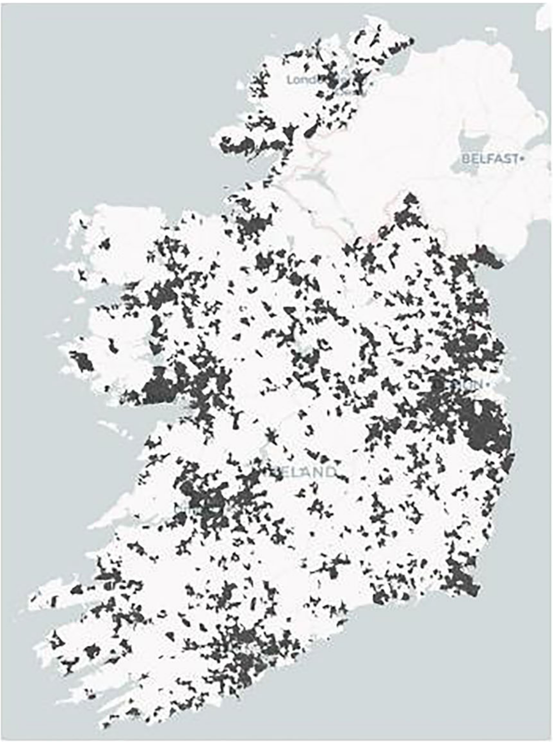


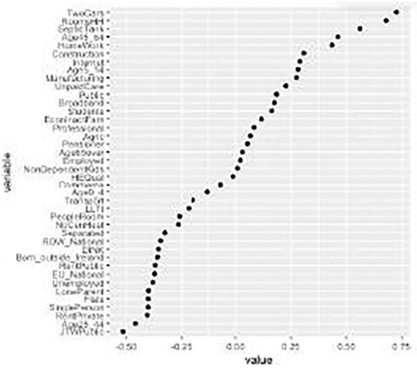


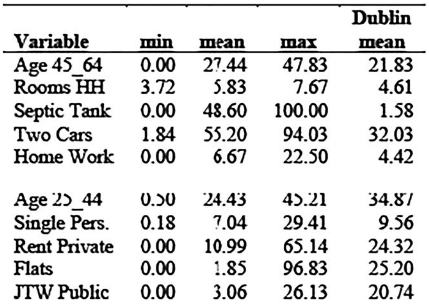
Variable	min	mean	max	Dublin mean
Rooms HH	4.19	5.76	7.02	4.61
Septic Tank	1.72	2.57	100.00	1.58
Two Cars	12.20	56.12	83.13	32.03
Agriculture	0.00	16.04	53.33	0.29
Construction	0.00	7.13	33.33	3.74
Rent Private	0.00	5.68	26.09	24.32
JTW Public	0.00	1.15	10.74	20.74
Commerce	0.00	17.69	45.45	28.53
Transport	0.00	4.80	20.18	12.28
Broadband	27.59	75.65	100.00	95.36

The main economic activities of this group are construction and agriculture. The high numbers of car ownership rate, septic tank usage and number of rooms per household are some of the main characteristics of this group and they are much higher than Dublin average. This group has the highest number of small areas and second highest population percentage (18.3%). From a spatial perspective, this group will be highly distributed in the counties and none in cities.

Group 2



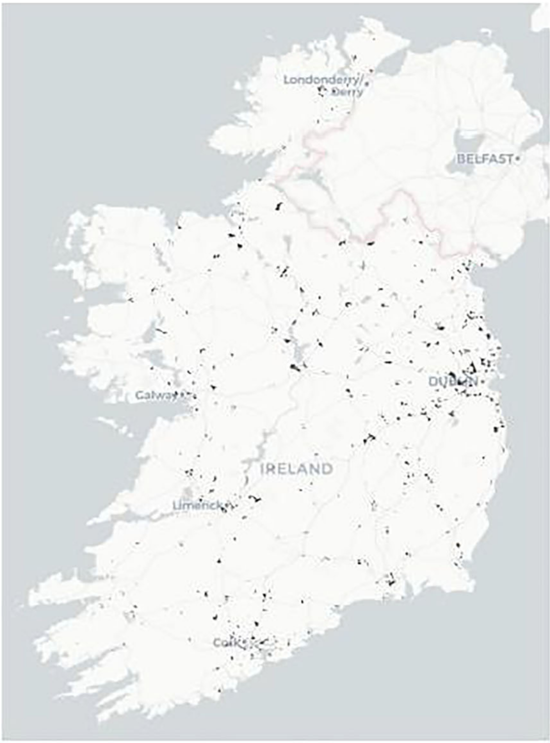


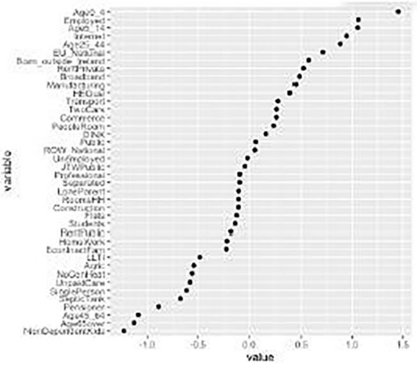


Variable	min	mean	max	Dublin mean
Age 45_64	0.00	27.44	47.83	21.83
Rooms HH	3.72	5.83	7.67	4.61
Septic Tank	0.00	48.60	100.00	1.58
Two Cars	1.84	55.20	94.03	32.03
Home Work	0.00	6.67	22.50	4.42
Age 25_44	0.50	24.43	45.21	34.87
Single Pers.	0.18	7.04	29.41	9.56
Rent Private	0.00	10.99	65.14	24.32
Flats	0.00	1.85	96.83	25.20
JTW Public	0.00	3.06	26.13	20.74

The composition of the households is around the middle age groups with two or more cars. Living in slightly bigger houses probably with septic tank facility. The tendency of the households in this group are more working from home. Group 2 consists the highest population compared to other groups (20.8%). Cork County, Galway County and Donegal County are the locations with the highest spatial distribution of this group.

Group 3

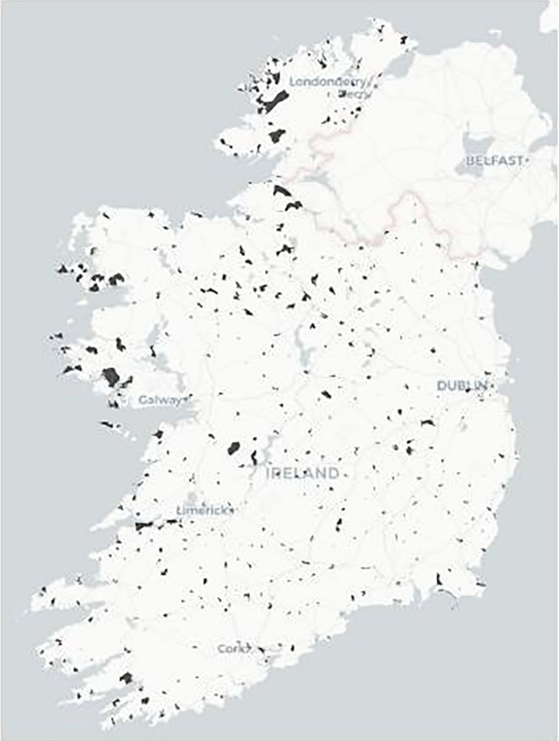


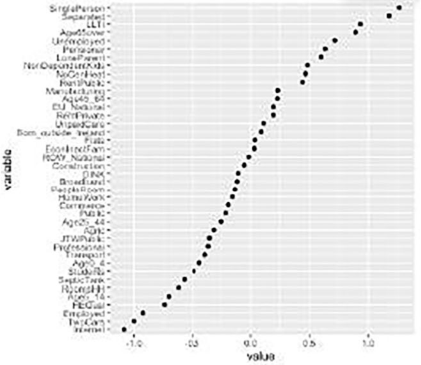


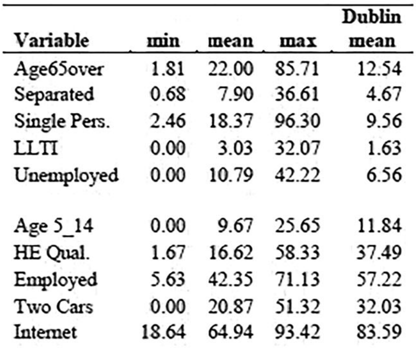
Variable	min	mean	max	Dublin mean
Age 0_4	1.61	11.35	21.46	6.57
Age 5_14	4.70	19.41	34.17	11.84
Age 25_44	20.65	39.36	66.83	34.87
Employed	34.34	65.84	90.08	57.22
Internet	63.89	89.00	100.00	83.59
Age 45_64	2.99	16.31	36.81	21.83
Age65over	0.00	4.10	25.90	12.54
Pensioner	0.00	1.87	6.11	6.19
Non-Depend. Kids	0.00	9.55	37.50	25.87
Septic Tank	0.00	2.83	86.67	1.58

Household composition in this group is mostly young families with dependent children. Employment rates are high. From a population perspective 16.4% of Ireland are residents of this group which is the third highest in rank. Fingal, South Dublin, Kildare and Cork County are the areas where this group is highly concentrated.

Group 4



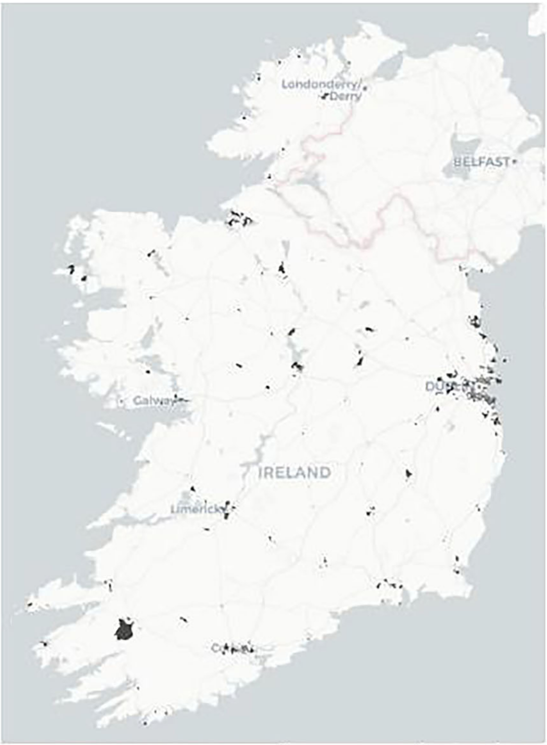


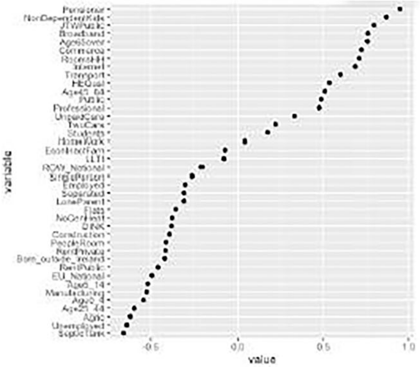


Variable	min	mean	max	Dublin mean
Age65over	1.81	22.00	85.71	12.54
Separated	0.68	7.90	36.61	4.67
Single Pers.	2.46	18.37	96.30	9.56
LLTI	0.00	3.03	32.07	1.63
Unemployed	0.00	10.79	42.22	6.56
Age 5_14	0.00	9.67	25.65	11.84
HE Qual.	1.67	16.62	58.33	37.49
Employed	5.63	42.35	71.13	57.22
Two Cars	0.00	20.87	51.32	32.03
Internet	18.64	64.94	93.42	83.59

The composition of the households is mostly a combination of separated people, people living on their own and people over the age of 65. Unemployment rate is quite high, and the health conditions of the households are particularly bad. This group's spatial concentration is high in Cork county, Tipperary and Donegal and very low in Dun Laoghaire. 10.4% of population of Ireland is in this cluster.

Group 5

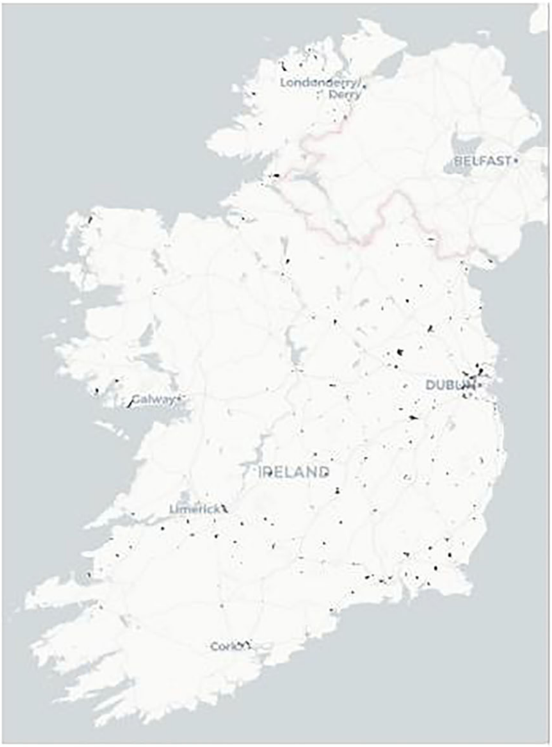


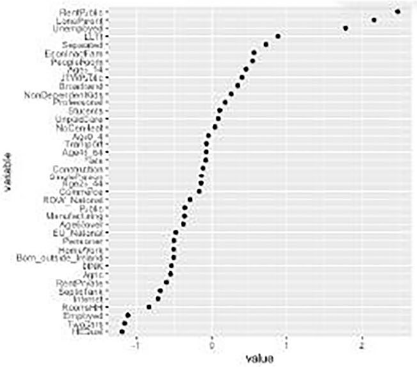


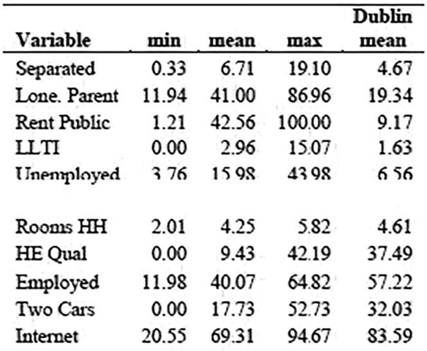
Variable	min	mean	max	Dublin mean
Age65over	1.02	20.80	55.29	12.54
Pensioner	0.00	11.93	41.14	6.19
Non-Depend. Kids	2.50	37.28	71.43	25.87
JTW Public	0.00	15.59	49.23	20.74
Broadband	73.68	96.63	100.00	95.36
Age 0_4	0.00	5.04	12.65	6.57
Age 25_44	6.33	22.87	42.15	34.87
Septic Tank	0.00	3.35	93.55	1.58
Unemployed	0.00	4.14	15.29	6.56
Agriculture	0.00	0.49	12.07	0.29

The composition of the population in this group is mostly elderly households; either families with nondependent children or people who are living alone. The percentage of the pensioners in this group is higher than the other groups and Dublin. Based on the households' declarations who are part of this group, their health conditions are not good. 14% of inhabitants are part of this group in Ireland. Dublin with its four local authorities (Dublin City, Dun Laoghaire, South Dublin and Fingal) has the 64.5% of the small areas assigned to this group.

Group 6



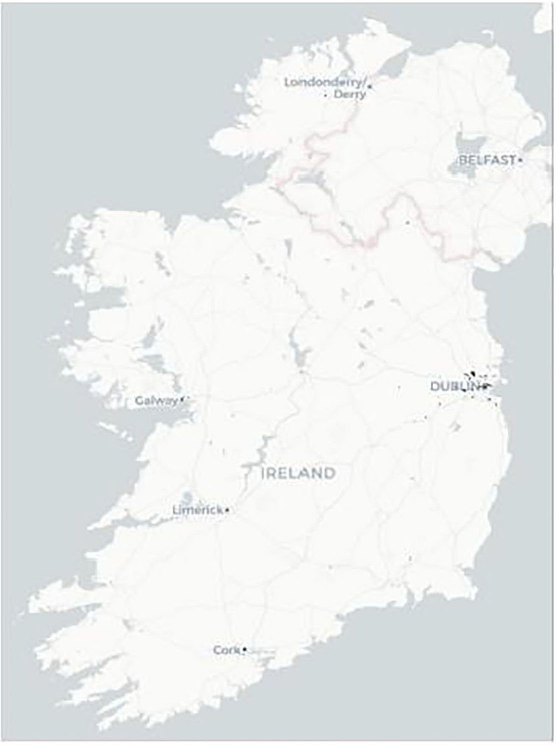


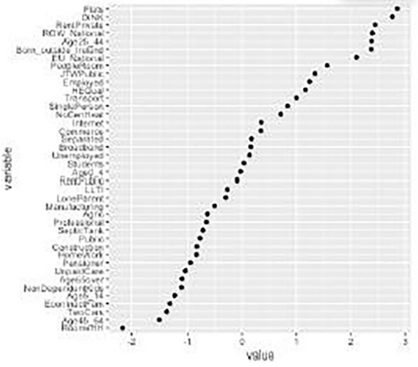


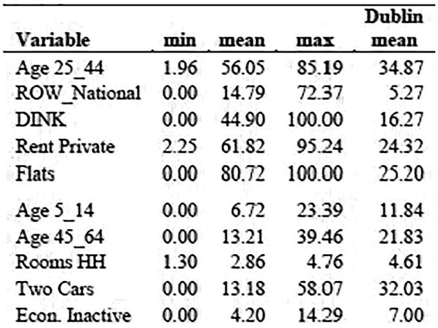
Variable	min	mean	max	Dublin mean
Separated	0.33	6.71	19.10	4.67
Lone. Parent	11.94	41.00	86.96	19.34
Rent Public	1.21	42.56	100.00	9.17
LLTI	0.00	2.96	15.07	1.63
Unemployed	3.76	15.98	43.98	6.56
Rooms HH	2.01	4.25	5.82	4.61
HE Qual	0.00	9.43	42.19	37.49
Employed	11.98	40.07	64.82	57.22
Two Cars	0.00	17.73	52.73	32.03
Internet	20.55	69.31	94.67	83.59

This group is consisting separated people and single parents. The unemployment rate is high, and the health conditions of people are low. In this group households are mainly renting their houses from local authorities. This group holds the 9% of the total population. Dublin city has the highest spatial concentration of this group followed by South Dublin.

Group 7



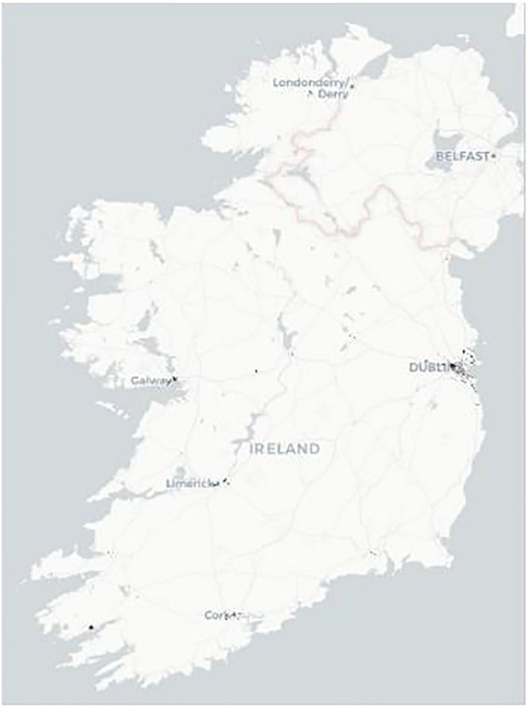


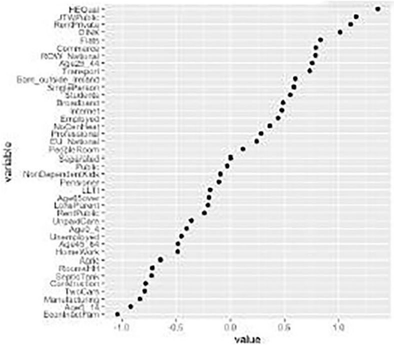


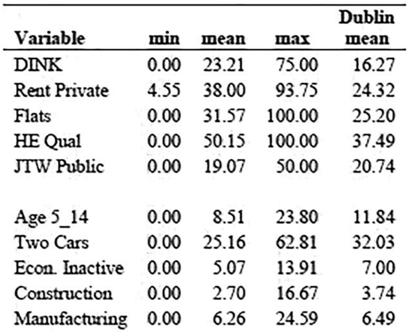
Variable	min	mean	max	Dublin mean
Age 25_44	1.96	56.05	85.19	34.87
ROW_National	0.00	14.79	72.37	5.27
DINK	0.00	44.90	100.00	16.27
Rent Private	2.25	61.82	95.24	24.32
Flats	0.00	80.72	100.00	25.20
Age 5_14	0.00	6.72	23.39	11.84
Age 45_64	0.00	13.21	39.46	21.83
Rooms HH	1.30	2.86	4.76	4.61
Two Cars	0.00	13.18	58.07	32.03
Econ. Inactive	0.00	4.20	14.29	7.00

Households in this group are mostly pre-families. Young couples both employed and with no children is one of the main characteristics of this group. The general tendency among the households in this group is to rent from private landlords and specially to rent flats. The composition of the population is based on international households mostly born out of Europe and Ireland. This group has 5.9% of population which spatially gathered in Dublin (63.3%).

Group 8







Variable	min	mean	max	Dublin mean
DINK	0.00	23.21	75.00	16.27
Rent Private	4.55	38.00	93.75	24.32
Flats	0.00	31.57	100.00	25.20
HE Qual	0.00	50.15	100.00	37.49
JTW Public	0.00	19.07	50.00	20.74
Age 5_14	0.00	8.51	23.80	11.84
Two Cars	0.00	25.16	62.81	32.03
Econ. Inactive	0.00	5.07	13.91	7.00
Construction	0.00	2.70	16.67	3.74
Manufacturing	0.00	6.26	24.59	6.49

One of the noticeable characteristics of this group is higher education levels. The mean value of the higher education quality is over the Dublin average. The households who are part of this group are highly participating in using public transport to travel to work. The house type that the households are living in are mostly flats and rental. Double income households with no kids are the other property of this group that emerges. Compared to other groups small part of the population is residing in this group (5.2%) but high part of the small areas assigned to this group are in Dublin (78.1%) – Dublin city (52.3%), Dun Laoghaire (15.2%), Fingal (5.6%). Cork city and Galway city are the other spots that this group is highly concentrated. On the other hand, counties like Cavan, Longford, Monaghan, Tipperary and Leitrim do not show any characteristics related to this group.

Appendix 2 – Variables

Variable	Census Theme	Description	Analysis Theme
Age0_4	Sex, Age and Marital Status	Age 0 to 4	Demographic
Age5_14	Sex, Age and Marital Status	Age 5 to 14	Demographic
Age25_44	Sex, Age and Marital Status	Age 25 to 44	Demographic
Age45_64	Sex, Age and Marital Status	Age 45 to 64	Demographic
Age65over	Sex, Age and Marital Status	Age 65 and over	Demographic
EU_National	Migration, Ethnicity, Religion and Foreign Languages	EU Nationality	Demographic
ROW_National	Migration, Ethnicity, Religion and Foreign Languages	Nationality - Rest of world	Demographic
Born_outside_Ireland	Migration, Ethnicity, Religion and Foreign Languages	Birthplace out of Ireland	Demographic
Separated	Sex, Age and Marital Status	Separated and Divorced	Household Composition
SinglePerson	Sex, Age and Marital Status	1 person households	Household Composition
Pensioner	Families	Retired households	Household Composition
LoneParent	Families	One parent family with children	Household Composition
DINK	Families	Pre-family	Household Composition
NonDependentKids	Families	Families with youngest child aged 20 and over	Household Composition
RentPublic	Housing	Rented from private landlord	Housing
RentPrivate	Housing	Rented from Local Authority	Housing
Flats	Housing	Flat/apartment	Housing
NoCenHeat	Housing	No central heating	Housing
RoomsHH	Housing	Average no. of rooms for household	Housing
PeopleRoom	Housing	Average no. of people for room	Housing
SepticTank	Housing	Individual septic tank	Housing
HEQual	Education	Higher education - including bachelor, postgraduate, doctorate	Socio Economic

References

Adnan, M., Longley, P. A., Singleton, A. D., & Brunsdon, C. (2010). Towards real-time geodemographics: Clustering algorithm performance for large multidimensional spatial databases. *Transactions in GIS, 14*(3), 283–297.

Batey, P., Brown, P., & Corver, M. (1999). Participation in higher education: A geodemographic perspective on the potential for further expansion in student numbers. *Journal of Geographical Systems, 1*(3), 277–303.

Booth, C. (1902). London poverty map legend: 1898-1899. https://booth.lse.ac.uk/map/12/-0.0934/51.4859/100/0. Accessed on 18.10.2018.

Brown, P. J., Hirschfield, A., & Batey, P. W. (1991). Applications of geodemographic methods in the analysis of health condition incidence data. *Papers in Regional Science, 70*(3), 329–344.

Brunsdon, C., Longley, P., Singleton, A., & Ashby, D. (2011a). Predicting participation in higher education: A comparative evaluation of the performance of geodemographic classifications. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 174*(1), 17–30.

Brunsdon, C., Rigby, J., & Charlton, M. (2011b). Ireland census of population 2011: a classification of small areas. https://rpubs.com/chrisbrunsdon/14998.

Brunsdon, C., Charlton, M., Rigby, J., & E. (2016). An open source geodemographic classification of small areas in the republic of Ireland. *Applied Spatial Analysis and Policy, 11*, 183–204. https://doi.org/10.1007/s12061-016-9212-4.

Charlton, M., 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, 13*, 69–98.

CSO, Central Statistics Office Ireland, (2016). Ireland Population Census Results, Small Areas, Accessed in September 2017. https://www.cso.ie/en/census/census2016reports/census2016smallareapopulationstatistics/

Farr, M., Wardlaw, J., & Jones, C. (2008). Tackling health inequalities using geodemographics: A social marketing approach. *International Journal of Market Research, 50*(4), 449–467.

Feng, Z., & Flowerdew, R. (1998). Fuzzy geodemographics: A contribution from fuzzy clustering methods. In S. Carver (Ed.), *Innovations in GIS 5* (pp. 119–127). London: Taylor and Francis.

Feng, Z., & Flowerdew, R. (1999). The use of fuzzy classification to improve geodemographic targeting. In B. Gittings (Ed.), *Innovations in GIS 6* (pp. 133–143). London: Taylor & Francis.

Fisher, P., Tate, N., & Slingsby, A. (2014). Type-2 Fuzzy Sets Applied to Geodemographic Classification. Paper presented at the Eighth International Conference on Geographic Information Science (GIScience 2014), 23-09-2014 - 26-09-2014, Vienna, Austria.

Fisher, P., & Tate, N. J. (2015). Modelling class uncertainty in the geodemographic output area classification. *Environment and Planning. B, Planning & Design, 42*, 541–563.

Goss, J. (1995). "We know who you are and we know where you live": The instrumental rationality of geodemographic systems. *Economic Geography, 71*(2), 171–198.

Grekousis, G., & Hatzichristos, T. (2012). Comparison of two fuzzy algorithms in geodemographic segmentation analysis: The fuzzy C-means and Gustafson–Kessel methods. *Applied Geography, 34*, 125–136.

Harris, R., Sleight, P., & Webber, R. (2005). *Geodemographics, GIS and neighbourhood targeting*. John Wiley and Sons, ISBN: 978-0-470-86413-5.

Hartigan, J. A. (1975). *Clustering algorithms*. London: Wiley.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society: Series C: Applied Statistics, 28*(1), 100–108.

Herbert, D. T. (1967). Social areal analysis: A British study. *Urban Studies, 4*(1), 41–60.

Hoyt, H. (1939). *The structure and growth of residential neighborhoods in American cities*. Washington DC: Federal Housing Administration.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. New York: John Wiley and Sons.

Kitchin, R., & McArdle, G. (2016). Urban data and city dashboards: Six key issues, the Programmable City working paper 21, Maynooth University.

Longley, P. (2005). Geographical information systems: A renaissance of geodemographics for public service delivery. *Progress in Human Geography, 29*(1), 57–63.

Longley, P. (2012). Geodemographics and the practices of geographic information science. *IJGIS, 26*(12), 2227–2237.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1*, 281–297.

Mitchell, V. W., & McGoldrick, P. J. (1994). The role of geodemographics in segmenting and targeting consumer markets: A Delphi study. *European Journal of Marketing, 28*(5), 54–72.

Openshaw, S. (1983). Multivariate analysis of census data: The classification of areas. In D. W. Rhind (Ed.), *A census users handbook* (pp. 243–264). London: Methuen.

Openshaw, S., & Blake, M. (1995). Geodemographic segmentation systems for screening health data. *Journal of Epidemiology & Community Health, 49*(Suppl. 2), S34–S38.

Openshaw, S., Cullingford, D., & Gillard, A. (1980). A critique of the national classifications of OPCS/PRAG. *Town Planning Review, 51*(4), 421–439.

Park, R. E., Burgess, E., & McKenzie, R. (1925). *The City: Suggestions for investigation of human behaviour in the urban environment*. Chicago: Chicago University Press.

Petersen, J., Gibbin, M., Longley, P., Mateos, P., Atkinson, P., & Ashby, D. (2011). Geodemographics as a tool for targeting neighbourhoods in public health campaigns. *Journal of Geographical Systems, 13*(2), 173–192.

Shevky, E., & Bell, W. (1955). *Social area analysis*. Stanford: Stanford University Press.

Shevky, E., & Williams, M. (1949). *The social areas of Los Angeles*. Berkeley: University of California Press.

Singleton, A. D. (2010). The geodemographics of educational progression and their implications for widening participation in higher education. *Environment and Planning A, 42*(11), 2560–2580.

Singleton, A. D., & Longley, P. (2009). Geodemographics, visualisation and social networks. *Applied Geography, 29*, 289–298.

Singleton, A. D., & Spielman, S. (2014). The past, present, and future of Geodemographic research in the United States and United Kingdom. *The Professional Geographer, 66*(4), 558–567.

Son, L. H., Cuong, B. C., Lanzi, P. L., & Thong, N. T. (2012). A novel intuitionistic fuzzy clustering method for geo-demographic analysis. *Expert Systems with Applications, 39*(10), 9848–9859.

Stodden, V., Bailey, D. H., Borwein, J., LeVeque, R. J., Rider, W., & Stein, W., 2013. Setting the default to reproducible reproducibility in computational and experimental mathematics, ICERM workshop report.

Tonks, D. G., & Farr, M. (1995). Market segments for higher education: Using geodemographics. *Marketing Intelligence & Planning, 13*(4), 24–33.

Tryon, R. C. (1955). *Identification of social areas by cluster analysis: General method with application to the San Francisco Bay area*. Berkeley: University of California Press.

Webber, R. J. (1975). *Liverpool Social Area Study, 1971 Data, PRAG technical paper 14*. London: Centre for Environmental Studies.

Webber, R. J. (1977). *An Introduction to the National Classification of Wards and Parishes, PRAG technical paper 23*. London: Centre for Environmental Studies.

Webber, R., & Burrows, R. (2018). *The predictive postcode: The Geodemographic classification of British society*. London: SAGE Publications.

Webber, R., & Craig, J. (1978). *Socio-economic classifications of local authority areas. Studies on medical and population subjects* (p. 35). London: OPCS.

Yazgi Walsh, B., Brunsdon, C., & Charlton, M. (2017). Ireland Census 2016: a Classification of Small Areas, https://www.rpubs.com/burcinwalsh/343141.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.