Ellen Talbot

24/10/2018

Chapter 5. Reconsidering fuel poverty; The EUC, a new classification

5.1 Introduction

5.1.1 Establishing a fuel poverty definiton

The issue of global fuel poverty and energy security has been addressed in earlier chapters. Here the focus is concentrated on the UK and the need for the ways in which fuel poverty is defined, understood, addressed and legislated for to be clearer.

Residential fuel poverty has historically been difficult to define, and there is no single unified example. The broadly accepted definition is that of Brenda Boardman (1991) :

"The inability to afford adequate heat because of energy efficiency in the home."

It exists as the product of three aggravating factors; low income, high energy prices and energy inefficient housing stock. It is the last of these which is critical in differentiating fuel poverty from other types of deprivation as certain types of dwellings will undeniably cost more to heat than others, as a function of their physical properties. Fuel poverty exists where low income houses pay inflated energy costs because they live in inefficient dwellings and is a very real concern for many households in the UK due the the comparatively low quality of the national housing stock (Guertler and Royston 2013). This inefficiency coupled with a temperate national climate where external temperatures regularly dip below those required for healthy living present a very real risk of households subsisting in conditions which range from thermally uncomfortable to terminally debilitating.

Isolation of fuel poverty as a distinct form of deprivation is usually traced back to the 1973 oil crisis, when soaring domestic fuel prices resulted in many households facing difficulties affording fuel (Bradshaw and Hutton 1983). The issue began to garner attention and in 1975 the National Right to Fuel Campaign (NRFC) was formed with the objective of ending fuel poverty in the UK and securing a warm, dry and well lit home for all, regardless of income and location (NRFC 2013).

Despite this promise, major advancements in the fuel poverty vernacular were not made until the publishing of Brenda Boardman's 'Fuel Poverty' in 1991, which offered the first quantitative definition and multi-disciplinary account of the problem. She offered the 10% threshold definition, whereby fuel poverty was the situation where expenditure on energy services is equal to or greater than 10 percent of income (Boardman 1991, pp 201). This figure was derived from then contemporary data as to the energy expenditure in the lowest three income deciles.

Even still, fuel poverty did not become a formal concern of the UK government until 2000, when the Warm Homes and Energy Conservation Act 200 (WHECA) required that the

Government "specify a target date for achieving the objective of ensuring that as far as is reasonably practicable persons in England and Wales do not live in fuel poverty". Subsequently, a target was established that fuel poverty should be eradicated in England by 2016, and in vulnerable households by 2010 (DEFRA and DTI 2005; DTI 2001). A complimentary strategy was born and a version of Boardman's fuel poverty definition written into policy for monitoring purposes (DTI 2001).

In the subsequent decade, a range of policies both economic and technical were implemented with the goal of tackling fuel poverty. However on the face of it, these were a resounding failure. Fuel poverty steadily rose year on year and both the 2010 and 2016 poverty eradication targets were missed; which can be construed as evidence of an ineffective policy approach on the part of the Government. In 2010, the October spending review included a commitment to re-evaluate the use of the 10pc definition as part of a drive to reduce state expenditure and a subsequent report which later became widely knows as the Hills Review reaffirmed fuel poverty as a serious problem distinct from income poverty (Hills 2012).

5.1.2 The Hills Fuel Poverty Report

In his review, Hills reconsidered the difference between poverty and fuel poverty; it is not a new distinction, but the Hills review represented a further entrenchment of this 'dividing practise'. The distinction has highly important policy implications, chiefly because it distances discussions of fuel poverty from those of poverty and was identified with reference to the interaction between low incomes and high required spending. In doing so it foregrounds energy efficiency measures as an appropriate response to fuel poverty above measures that address low incomes or cost of living. Whilst Hills praised the 10pc definition for it's ability to capture the interactions of the drivers of fuel poverty, he found fault with its ability to effectively represent the nature of that problem and identified a multitude of weaknesses, some of which are highlighted in Table 5.1.2 below.

Table 5.1.2 Issues and Problematisations with the 10pc definition

Issue	Problematisation
The fixed threshold	A fixed threshold means that the definition of fuel poverty is extremely sensitive to that choice.
High Income High Cost	Under the 10pc definition, those with high incomes and high fuel costs can be considered fuel poor if their energy costs are sufficient.
Treatment of housing cost	For the purposes of measurement, incomes have been considered before housing costs are subtracted, i.e., inclusive of income that is not truly disposable as it is apportioned to a specific, unavoidable purpose.

As an alternative, Hills proposed a different conceptualisation of fuel poverty, which reconfigures fuel poverty in relative terms; the Low Income High Cost definition (LIHC). It differs from the 10pc definition which is based on an absolute threshold for fuel costs and is instead relative; a household is fuel poor if it's fuel expenditure is comparatively high and its income is comparatively low. Figure 5.1.2a illustrates the LIHC fuel poverty definition.

Table : Hills issues and problematisations with the 10pc definition.

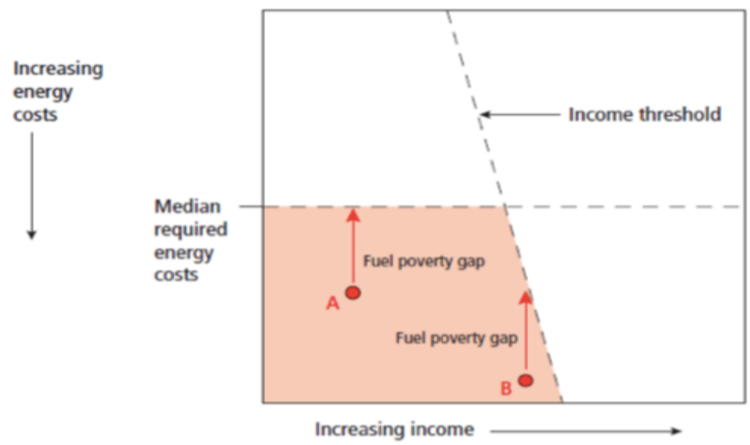


Figure 5.1.2a The Low Income High Cost Fuel Poverty Indicator (Energy Saving Trust 2019)

The thresholds used are as follows; the income threshold falls where subtraction of required equivalised energy costs from income leaves the household at the Department for Work and Pensions' (DWP) official poverty line, as defined in Households Below Average Income (HBAI) analysis; less than 60 per cent of median equivalised household income, after housing costs (Hills 2012, pp 53) . Effectively, this defines a low income household as one that, having paid required energy costs, is below the official poverty line. The cost threshold lies at the point whereby equivalised household bills equal the national median (Hills 2012, pp 59-60) and whilst this does not entirely eliminate the failure of policy to that date it does go some way to mitigating the distortionary impact of price rises upon official figures. Hills also defined another measure of fuel poverty known as the fuel poverty gap - the reduction in required spending which would take a household out of fuel poverty as can be seen in Figure 5.1 (Energy Saving Trust 2019).

Under the 10pc definition, the Department for Energy and Climate Change (DECC) measured fuel poverty under both Before Housing Cost (BHC) and After Housing Cost (AHC), but used BHC for official statistics. Considering AHC results is a reduction in considered income for those with higher housing costs, which manifests itself in a shift away from pensioners who are more likely to own their homes outright towards working age adults, including families with children. This was a popular move as it was argued that AHC more accurately reflects the composition of the fuel poor group, where housing costs are high.

When the LIHC definition was written into official policy in 2013 (DECC 2013) the number of fuel poor households did decreased from 4 million to 2.7 million but the number of fuel poor individuals increased from 7.4 million to 7.8 million. This came as a result of the elimination of some Low Income Low Cost (LILC) and High Income High Cost (HIHC) households, but equivalised energy usage meant more larger households with higher occupation were considered fuel poor. This new strategy marked a shift in the problematisation of fuel poverty from a condition that should and can be eradicated to a condition than can at best be alleviated. It was chosen partly for this very reason; it has a tendency to show a consistent population of fuel poor households over time due to its equivalsation of fuel costs. It also shifted the narrative to one with emphasis on "targeting the most vulnerable" and the prioritisation of the most severely fuel poor.

The official target was given as 'ensuring that as many fuel poor homes as is reasonably practicable achieve a minimum energy efficiency rating of Band C by 2030', with interim milestones of 'Band E by 2020' and 'Band D by 2025', therefore placing the entire focus of the strategy on energy efficiency (HM Government 2014). The Committee on Fuel Poverty (CFP) whose key role is to monitor and report on progress towards these milestones commented in November 2018 that despite the average fuel poverty gap closing by 14% over the last 4 years, progress towards achieving even the smallest improvements is "slow and flat-lining" (Committee on Fuel Poverty 2018). This reinforces the need to refocus the definitionof fuel poverty as a multifaceted issues requiring improved targetting strategies.

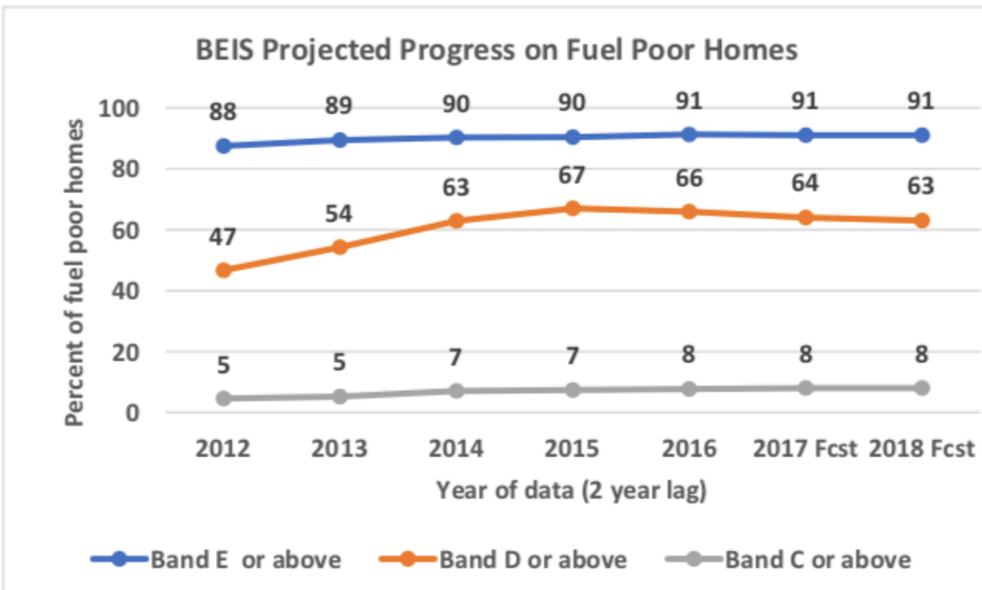


Figure 5.1.2b Energy Efficienct Rating Improvements (Committee on Fuel Poverty 2018)

5.1.3 Policy implications

This austerity lead redefinition also meant a shifting of goalposts in terms of responsibility. Government schemes such as 'Warm Front' were concluded, leaving only supplier-led improvement schemes, making them entirely accountable for the delivery of energy efficiency measures to fuel poor households. This means that the role of the state in supporting the fuel poor has been significantly diminished in favour of a model based on supplier obligations, funded via energy bills. The fuel poor become the subjects of the energy market, and certainly from a lived perspective, gains in energy efficiency can easily be overshadowed by changing welfare policy and energy prices (Middlemiss and Gillard 2015). The Centre for fuel poverty reported that not all will take any cost reduction from energy efficiency improvements as monetary gain and will instead trade off for increased thermal comfort. This is typically dependent on their annual net household income and the pattern of comfort taking is described as "thermostat settings increasing from 18 degrees c until either the thermostat reaches 21 degrees c or half of the financial gain is spent on additional heating cost, whichever occurs first". The fact that some in fuel poverty systematically underheat their homes is hidden in the current definition of fuel poverty, as it used a predefined required heating pattern to obtain an adequate level of warmth. Those vulnerable to welfare policy changes and in receipt of other benefits are also likely to be reluctant to increase their consumption for fear of increasing bills and unexpected costs which could lead to them returning to a state of fuel poverty, or faced with the heat-or-eat dilemma.

The governments austerity drive means that only the most vulnerable or most impacted fuel poor subjects can have help in meeting their needs, whether that means those with the largest fuel poverty gaps or those with physical vulnerabilities. Low Income Low Cost (LILC) households are not a priority yet the Office for Gas and Electricity Market's (OFGEM) Energy Supply Probe identified that low income households were less likely to change tariffs, switch suppliers, compare offers, have the ability to access on-line offers and be more likely to be prevented from switching by existing debt (OFGEM 2008, pp 11 and pp 57). They are also more likely to have been given pre-payment meters or are unable to pay via direct debit; both of which incur higher costs, all of which amount to more undetected inequality in the LIHC indicator and need clarification.

The positioning of energy efficiency as a key technology under the LIHC definition has implications for what is possible in fuel poverty policy and beyond. A focus on energy efficiency reduces attention to other structural problems which exacerbate fuel poverty, particularly fuel costs and pricing, and income inequality. This redefinition lends itself to a technical assessment of the need for help to do with the efficiency of the housing stock, as opposed to a behavioural one and it can only be reiterated that a focus on cost effectiveness on the governments behalf will lead to some households with vulnerable members slipping through the cracks (Middlemiss 2017).

5.1.4 Why changes are needed

Indeed, a focus on energy efficiency makes a great deal of sense, especially in the UK given the quality of the housing stock. Many fuel poor homes are poorly insulated and relying on inefficient appliances and investment in energy efficiency is a cost effective approach in both the long and short term as the benefits of home improvements remain for many years (Boardman 2013) and as a relatively new industry, energy efficiency creates jobs and stimulates economic growth (Middlemiss 2017). But as previously mentioned, changes in policy which are seemingly unrelated to fuel poverty may have significant ramifications for those vulnerable to change and while fuel poverty is considered in isolation, changes are likely to cancel each other out. For instance, as far as fuel poverty is concenered, the most vulnerable are those in poor efficiency, hard to heat homes. When theses charcterstics are coupled with being rental properties, there is a dichotomy between the obligation of the landlord and the tenant where neither will see any benefit to making substanstial improvements to the home and is cited as one of the biggest barriers to improving energy efficiency in the rental sector; landlords see little incentive to invest as it is their tenant who will benefit from the lower bills, and the teant is either not inclined to invest in improvements as they won't live in the property long enough to see real financial reward, or are prevented from doing minor improvements through lack of consent from the landlord.

The rest of this chapter goes on to explain through use of the EPC data and census data that the new problematisation and definition of fuel poverty at a national policy level does not go far enough in terms of acting from the ground up (Moore 2012). Fuel poverty can be driven by many factors other than energy inefficiency such as high costs of energy, ill health and unreliable income (Middlemiss 2017). As such it is important to consider this multifaceted issue outside of its technical and structural limitations in order to better understand and legislate for it and issues surrounding it.

5.1.5 Conceptual framework for a Classification of Energy use

The multi-dimensionality of the fuel poverty issue is important, as the literature presented at the beginning of this chapter and previous chapters analyses have identified. It is apparent that there exists a breadth of spatially referenced data that could be utilised for classification which broadly cover the following domains:

- Access patterns
- Household level energy usage (aggregated to small area statistics)
- Structural and physical fixtures and fittings of households
- Demographic and contextual attributes

In the previous chapter, patterns of access to upgraded energy efficiency technologies such as smart meters were investigated and a Morans I test revealed geographical disparities. Profiling by Output Area Classification (OAC) revealed that propensity was not evenly distributed across all groups, suggesting that this access is likely to be influenced by a number of factors as well as a broadly suggesting inequity in the prioritisation of infrastructure upgrades, forming a basis to support a conceptual framework for an energy based classification. Attributes pertaining to age; population density and occupation as well as some more physical attributes such as accommodation type, building type and ownership have all been discussed prior as having links to overall consumption and would likely assist in building a typology by introducing measures of demographic characteristics to supplement consumption data, and as such, create a broader view of energy efficiency and fuel poverty.

5.2 Geodemographic specification

Geodemographic classifications in their most basic sense, can be described as 'the analysis of people by where they live' (Sleight 2004) and involve analysis of attributes relating to the socio-economic and built environment characteristics of small geographic areas. There are numerous approaches to constructing what is often referred to as part art and part science; the bespoke geodemographic classification. There is still significant variation in the methods employed to do so, despite recent developments aiding understanding within the field. This variation should be expected however, given that a particular set of methods or procedures suitable for one classification may not be suitable for another. Harris et al. (2005) present the most comprehensive overview of the various stages involved in constructing a geodemographic classification and as such this section will follow a similarly segmented approach, presenting a short overview of:

- Selecting potential measures
- Data evaluation
- Transformation and normalisation
- Weighting
- Standardisation
- Clustering
- Cluster hierarchy
- Textual and visual summaries

The first stage in the building of a bespoke geodemographic classification is the evaluation of potential measures. A typical K-Means cluster analysis requires measures to be numeric and continuous as against discrete or categorical in order for the algorithm to work correctly, which often requires extensive data manipulation. Other clustering algorithms exist, but are less commonly used, such as Expectation-Maximisation or hierarchical clustering (Harris, Sleight, and Webber 2005; Vickers and Rees 2007; Singleton and Spielman 2014). Data inputs are generally aggregated to a predefined geographic resolution, as dictated by the scales at which all data sources are available. These resolutions in the UK on which a classification can be built vary, but common aggregate geographies include postcode (or aggregation of; postcode sector/postcode district) boundaries, Output Area (OA), Lower Suer Output Ares (LSOA) or ward. The data structure prior to clustering is typically a table whereby rows represent one of these geographies and columns represent the attribute information on which they will be clustered.

Exploratory Data Analysis (EDA) is typically the next stage, evaluating input variables to examine issues such as missingness and correlation, assess distributions and skew, and more generally, gain an overview of relationships between variables. At this stage a classification builder may choose which, if any, variables should be removed if they present duplicate information (typically those variables that show high correlations with others) but there are no firm rules and choices can be largely subjective.

Following assessment of the input data, typically the next stage involves data normalisation and transformation, which aims to limit skew and standardising the data so they share a common scale. In an ideal scenario, all variables would exhibit normal distributions - some clustering algorithms are optimised to find spherical clusters, which can be problematic with skewed inputs. In practise however, there are very few situations where this holds true. There are a number of normalisation practises that can be implemented, including log10, Box-Cox and cube-root transformations (Gale et al. 2016). Whether or not skewed data would be transformed at all is a subject that has generated academic debate. There are both advantages and disadvantages; cluster formation is less likely to be adversely affected with normalisation versus the loss of the important information disseminated into distinctive clusters by outliers, which are reduced by normalisation (Singleton and Spielman 2014). In some commercial classifications this can be overcome by employing a weighting technique to reduce the impact of skew on cluster formation, but how the weights are derived is typically subjective and open to criticism (Harris, Sleight, and Webber 2005). The standardisation of this data is important because input variables will typically display different distribution, and in order to assess how large or small a particular geographic area's variance is from the mean, and to draw comparisons between variables, a common scale must be used. A common method is to transform data values to z-scores - calculated by subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation.

$$z = \frac{x - \mu}{\sigma}$$

Where μ is the mean of the population and σ is the standard deviation of the population.

This results in a set of scores that are positive if they fall above the mean and negative if they fall below, i.e. all standardized variables will have an adjusted population mean of 0. Using z-scores can be problematic, for example if an input variable is highly skewed with many outliers, the resulting z-score may be large enough to influence an area's cluster membership regardless of the area's other attributes. Again, weighting and variable normalisation techniques can be utilised to alleviate this issue.

Following standardisation across the dataset of final variables, the next stage is to run a cluster analysis. Typically an iterative allocation-reallocation method (K-means) is used, although other methods such as hierarchical also exist. The hierarchical method essentially treats each area as a separate cluster in the first instance and merges these 'clusters' based on measures of similarity. After similar clusters are merged, average values for the new clusters are computed and the process repeats until convergence, where an appropriate number of clusters (that exhibit minimum intra-cluster variance and maximum inter-cluster variance) are found. Although methodologically simplistic, this method can be expensive in terms of both time and computational effort due to the assessment and reassignment of cluster pairs whilst holding the intermediary results in memory. This can be particularly problematic when datasets are extremely large.

An iterative allocation-reallocation method uses a different technique to compute cluster assignments. A K-means algorithm works by setting seeds, which can be pre-defined or random observations within a dataset. The number of initial seeds is equal to the user-defined value for k (the number of clusters to be output). The algorithm then begins to assign observations to each seed based on proximity, typically measured by Euclidean distance. This initial allocation represents the first iteration of the algorithm. The centroids of the newly formed clusters are then calculated and become the seeds for the next iteration of assignments. The algorithm aims to minimise the Within Cluster Sum of Squares (WCSS), which is the cumulative sum of all the squared Euclidean distances from observations to cluster centroids. Smaller WCSS values represent more homogeneous (or similar) clusters. The algorithm repeats for many iterations until convergence, when assignments no longer change and WCSS values have been minimised. Figure 5.2.1 describes how the algorithm works on a 2 dimensional dataset.

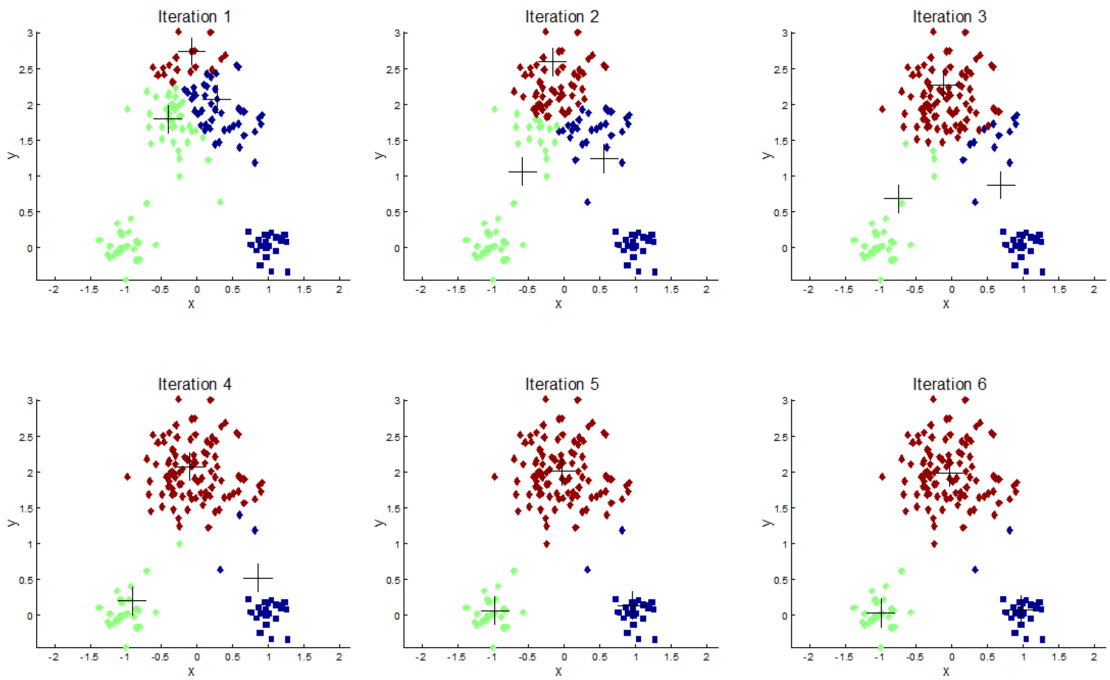


Figure 5.2a Iterative allocation clustering Source: http://www.practicaldb.com/data-visualization-consulting/cluster-analysis/

After the initial cluster run, it is possible to design either a top-down or bottom-up hierarchical structure to the classification, as described in Figure 5.2b. Top-down hierarchies involve clustering input data into predefined numbers of clusters that will form the highest tier of the resulting classification, which are then re-clustered within themselves to generate what would typically be called subgroups. This method can be repeated as many times as it is desired, although it would be logical to stop when sub-clusters begin to display no obvious differences from the parental clusters. Finding the optimum number of clusters at each stage is largely down to trial and error, although methods exist to aid evaluation. One such method involves the use of 'Clustergrams'; visualisations of the assignment and re-assignment of observations to clusters across a range of values for k. This visual method can assist in the selection of an optimum k value as it is possible to identify which clusters split to form new clusters and assess similarity or 'closeness' of newly formed clusters.

A bottom-up hierarchy involves clustering the data into k clusters, which are then merged to form a higher tier within the hierarchy. The number of tiers, initial number of clusters and grouping of clusters is however, largely down to trial and error and even in large datasets, cluster sizes and characteristics can vary significantly. As an example, the most granular tier of a three-tier classification may be referred to as 'type'. If the data is capable of supporting 50 'types' (essentially 50 distinct clusters) it can be clustered to this extent in the first instance. Similar 'types' may then be merged to form a less granular mid-level tier in the classification, call these 'groups'. Similarly, these 'groups' could be merged, again based on similarity, to form a coarse top tier of the classification or 'Supergroups' (Harris, Sleight, and Webber 2005).

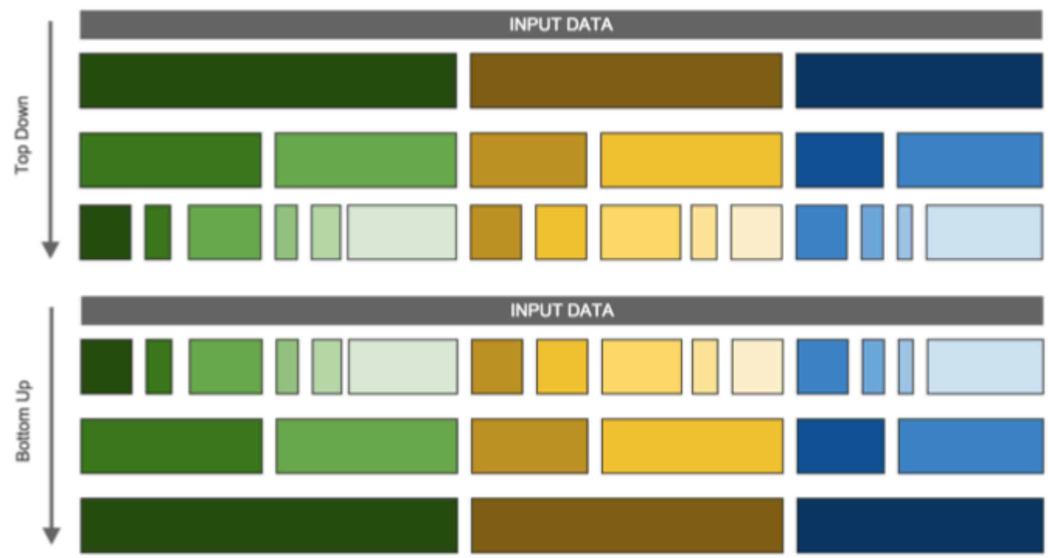


Figure 5.2b Hierarchical clustering design (cite Riddlesden, D)

5.3 Describing Energy

5.3.1 Current Fuel Poverty Statistics

In order to draw comparisons and critique the current fuel poverty defintion, it is imperative that the data it generates is properly understood. The Government produced openly available fuel poverty statistics for the year 2016 at the LSOA level based on the Low Income High Cost Indicator, which was again reweighted to Postcode Sector level for consistency. Figure 5.3.1a shows the spread of fuel poverty across England and Wales for all areas where data was available, by proportion of houses per PCS. Much of Wales is missing, but what is available falls into the higher categories of fuel poverty.

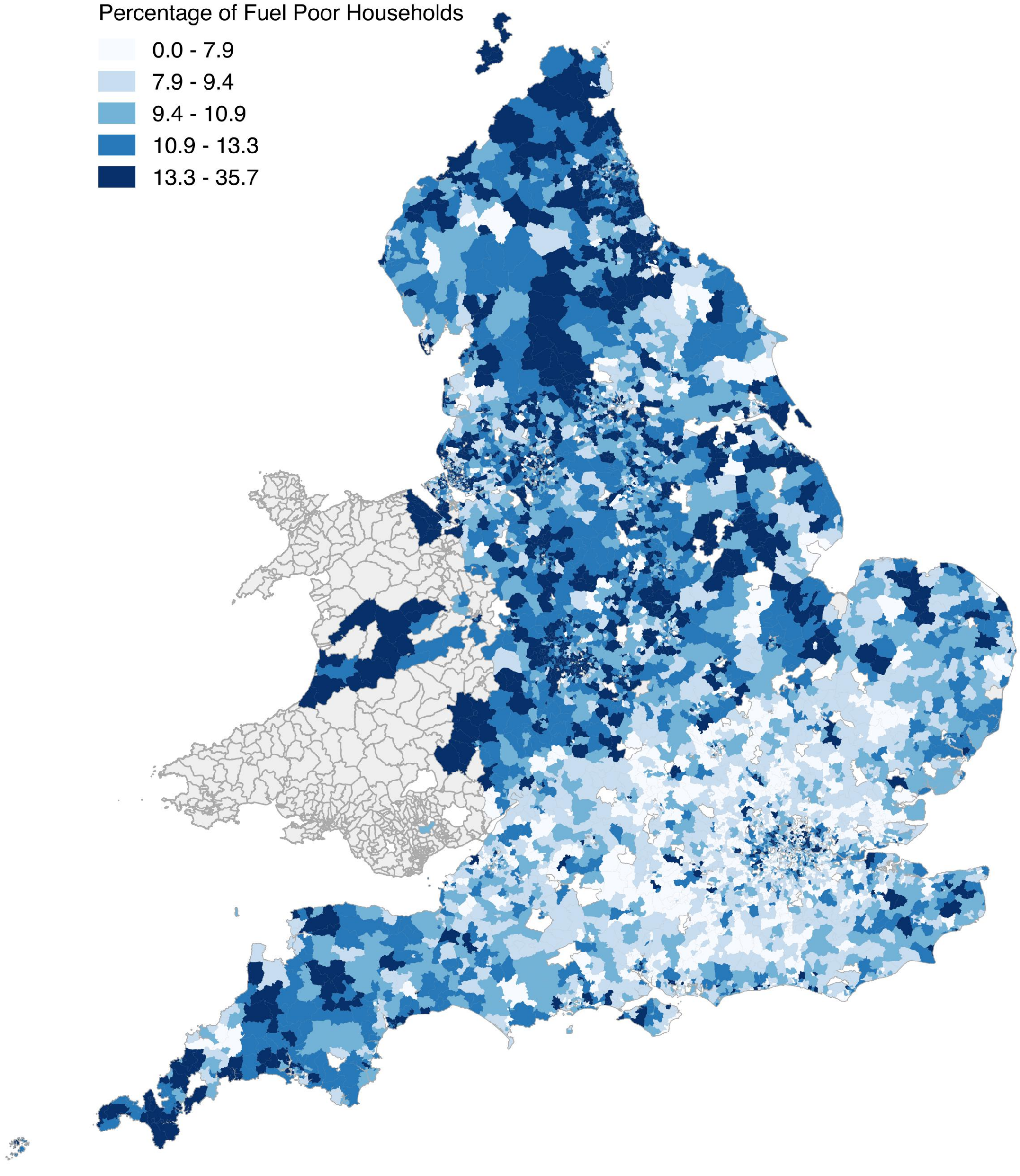


Figure 5.3.1a : Current Distribution of Fuel Poverty in England and Wales

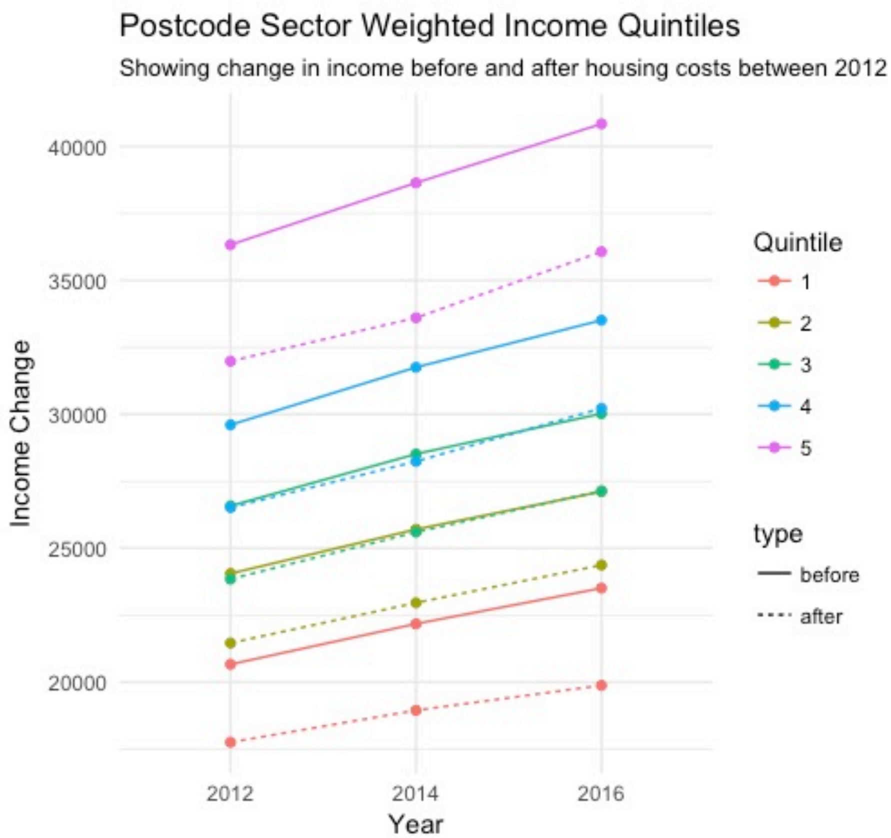
Figure 5.3.1a shows there is a clear geographic divide between the North and South of England, and even in Central London, where costs tend to be expectionally high, there are far fewer Postcode Sectors with a high percentage of fuel poor properties. This is possibly due to the UK legacy of northern mill towns lending itself to poor quality, low efficiency terrace housing, coupled with higher earning power towards the capital. There is also some evidence in Figure 5.3.1a of a disparity between urban and rural areas, possibly linked to the much lower density of houses in rural areas, as well as the fact that they are typically older, larger harder to heat properties that are underoccupied by older citizens [This needs a cite].

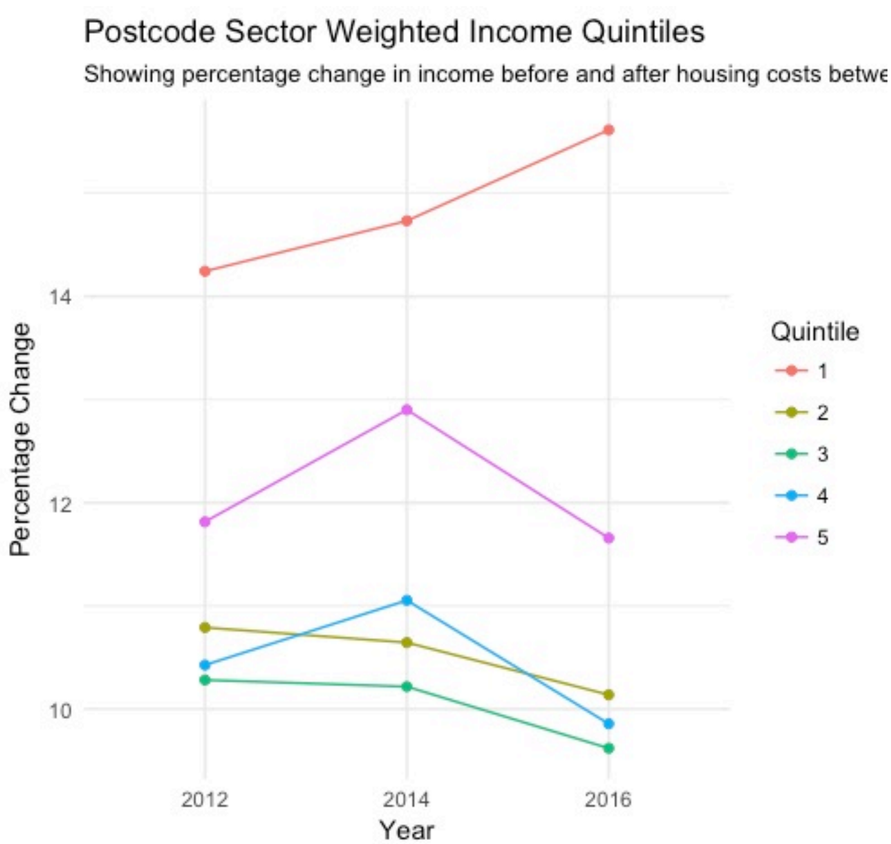
5.3.2 Before and After Housing Costs

To investigate the assumption that fuel poverty is caused by having a low income and high energy cost, the data from the UK government which provides information on net income before and after housing costs over a four year period was investigated. The data was split into income quintiles and the change over time compared to 2012 before housing cost figure.

Figure 5.3.2a below shows that for the lowest income quintile in 2016, the gap created by housing cost is widening and Figure 5.3.2b shows that this data found the lowest income quintile to have seen the sharpest percentage increase in housing costs over the four years, as well as having the highest percentage of income accounted for by housing costs overall. As discussed in existing literature, facets of poverty overlap those of fuel poverty. This analysis shows a degree of unpredictability for those in lower income quintiles, suggesting that budgeting and planning for future housing costs is likely to be difficult. This applies to both LIHC households and LILC households. This uncertainty may lead to those households finding themselves in either short term fuel poverty while their finances recover, or perpetually unable to meet their energy costs once their housing costs have risen. They may choose to maintain inadequate heating and lighting in their homes, or forgo other necessities such as food or transport to provide thermal comfort and in order to minimise the shortfall.

Other groups have seen their housing cost relative to their income reduce or remain steady, meaning they are less likely to find themselves at risk of fuel poverty. They are also more likely to have the ability to swallow any increase as part of their outgoings without seeing a detrimental effect on their disposable income due to their higher earnings, they may have the ability to save some of their income to deal with unexpected expenses and may also be able to afford to make improvements to their houses to decrease their energy costs further.


Postcode Sector Weighted Income Quintiles
Showing change in income before and after housing costs between 2012
Income Change
40000
35000
30000
25000
20000
2012
2014
2016
Year
Quintile
1
2
3
4
5
type
before
after


Postcode Sector Weighted Income Quintiles
Showing percentage change in income before and after housing costs betwe
Percentage Change
14
12
10
2012
2014
2016
Year
Quintile
1
2
3
4
5

5.3.3 Energy Performance Certificates

The Energy Performance Certificate Dataset is an open access dataset published on-line by the UK government detailing all energy certificates within their records. It can be freely downloaded in full by providing an email address, or searched via an API. It was first published on-line in 2017 and is updated every six months. The full dataset used in this analysis was downloaded in November 2018, containing 176 variables and X records in total. The full definitions and methodology for the creation for each variable can be viewed in the accompanying guidance notes, an extract of which is included in appendix 5, covering the subset of variables used in this analysis.

Energy performance certificates were introduced in stages from 2007 and stems from the EU directive on the energy performance of buildings. It was intended that the energy efficiency of buildings was made transparent through these certificates and provided information showing how the energy efficiency of the building could be improved. They give each building a linear rating from A (the most efficient) to G (the most inefficient) (MHCLG, 2018). EPCs also predict how costly it will be to heat and light, what its carbon dioxide emissions are likely to be as well as stating what the energy efficiency rating could be if improvements are made, and highlights cost-effective ways to achieve a better rating (Energy Saving Trust 2019).

They are valid for ten years but can be renewed sooner if the house is part of a transaction which requires one, for instance a sale or rental, or if improvements have been made. Other transactions also require the generation of an EPC such as certain types of government energy improvement funding schemes to prove eligibility such as an assessment for Green Deal; following a Green Deal; FIT (Feed in Tariff) application; RHI (Renewable Heat Incentive) application or ECO (Energy Company Obligation) assessment.

By examining the EPCs alongside demographic characteristics in the later clustering analysis, we are able to see the detailed profile of those who may typically be defined as fuel poor, and are able to examine more thoroughly the indicators that make them so. This would improve targeting of help and resources, as is the governments overarching aim.

5.3.4 Output Area Classification

To investigate the aggregated socio-spatial structure of energy efficiency, the Office for National Statistics Output Area Classification was appended to the EPC data. This classification was created for England and Wales from 2011 census data. As already distinguished, geodemographic classifications are aggregate categorical summary measures of the social and built environment of small areas. The OAC is aggregate to Output Area zones, which comprise a minimum of 40 houses; however the optimal size is 125. The OAC was appended to the EPC energy efficiency rating and the mean energy efficiency score per OAC group was calculated.

Despite the relatively small range in energy efficiency ratings between the highest and lowest groups of just under 20, there are clear disparities between the groups. The three groups within Supergroup 1; 1a - Farming Communities, 1b - Rural Tenants and 1c - Ageing Rural Dwellers take all three of the lowest energy efficiency ratings with between 51.46 and

56.58. This could be due to the fact that these rural households are less likely to be connected to the mains gas network, using wood, coal or oil as their main fuel source and live in large, old buildings typical of rural areas. The highest rated group are 2b - Inner City Students with a rating of 71.04, likely due to them living in very modern and efficient, newly built halls of residence style accommodation. It is fair to say that it is typical of this type of living arrangement to be inclusive of energy bills, so it is on the interest of the owners and investors to ensure that bills are as low as possible to maximise their profits on the room rates. The groups characterised by living in terraced accommodation such as 4b - Challenged Asian Terraces and 8b - Challenged Terrace Workers typically score lower than others placing in 4th and 5th lowest respectively after the rural dwellers, likely due to the poor quality housing stock symptomatic of the UK, but also because of, as previously mentioned their inability to seek out and afford energy improvements. Figure 5.3.4a shows the average current energy efficiency rating arranged by OAC group.

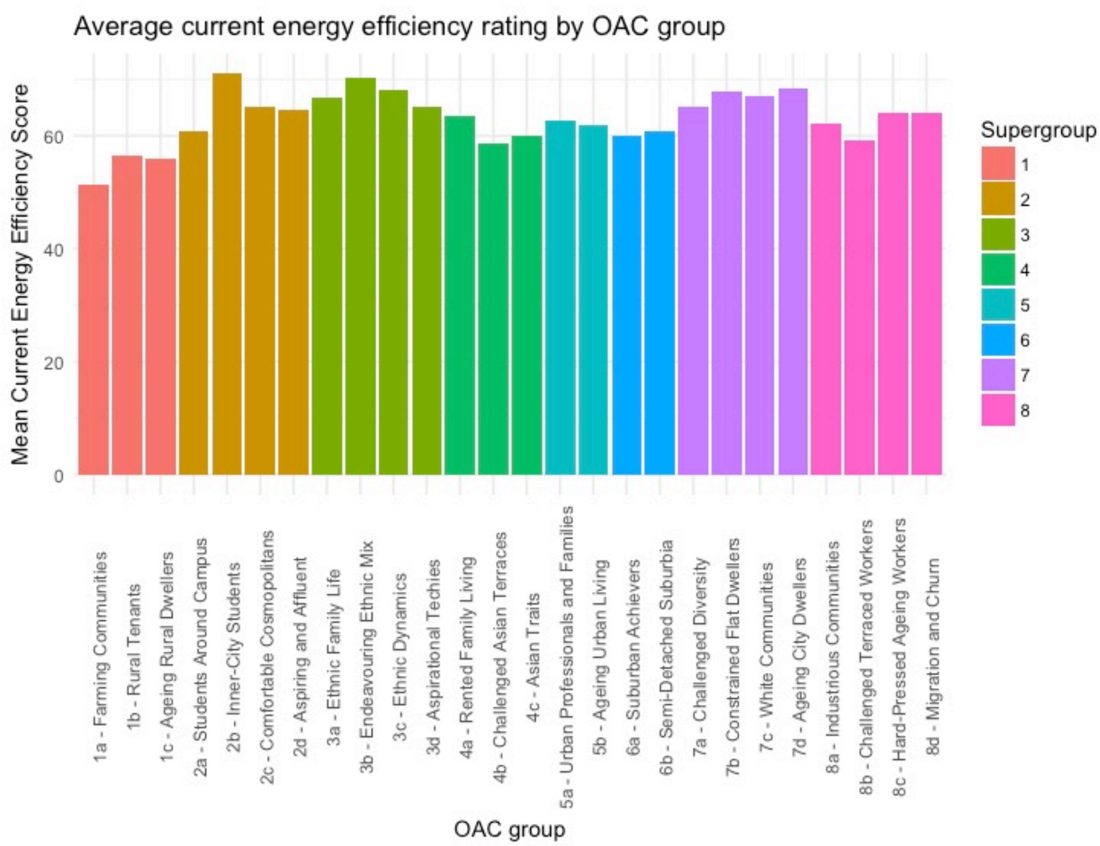


Figure 5.3.4a : Energy efficiency by Output Area Classification.

5.4 Building a classification of energy consumption

Data from the 2011 census was downloaded from the UK data service, at a household level and manipulated into the required geographic scale. This was done using the same reweighing technique as mentioned in Chapter 4, employing the National Statistics Postcode Lookup database and a calculated aggregate postcode sector weighted population headcount.

5.4.1 Selecting Measures

As a first step, consideration was required to identify those variables that would form useful inputs into the classification. Table 5.4.1a provides a summary of the elements that were initially selected to guide the classification process and included a number of domains relating to property features, energy efficiency, levels of consumption and demographic indicators. The majority of measures were derived from the Energy Performance Certificate (EPC) dataset and the 2011 census. In total 176 preliminary attributes were selected over the three taxonomical elements (Table 5.4.1a). To keep this chapter concise, the full table of variables, their descriptors and domains has been included in appendix 5. Both these datasets contain a wide variety of possible candidate variables, a large number are highly correlated or homogenous across space and so deemed less effective in classification building; for instance any variation in sex is considered to be of lower importance since overall the ratio in small areas is the same. Furthermore, some variables in the EPC dataset were simply unsuitable and did not address the needs of the end-user such as those linked to potential efficiency and consumption.

Due to the sparse nature of some variables at the individual level and for consistency throughout, all measures from the EPC dataset have been aggregated to postcode sector level. At this level there is coverage across the majority of England and Wales. Contextual and demographic indicators were obtained from the census and broadly represent attributes that are known to correspond with levels of energy consumption and fuel poverty such as age, income and accommodation type. As before, the reweighting of census variables to the postcode sector level was done utilising the NSPD. Before undertaking any data evaluation methods, the dataset was checked for missingness and as a result, 17 variables were removed due to being over 80% missing. In the case of EPC variables these pertained to measures providing very high level descriptors affecting the overall energy rating of very few households and in the case of census variables it was found that missingness was due to the variables being associated only with Scottish census responses, meaning our study area was not covered.

Table 5.4.1a Chosen measures

Domain	Sub-domain
Geo-locator	Postcode sector
Energy Information	
	Current energy efficiency
	Current environmental impact
	Current energy consumption
	Current costs
	Number of storeys
	Habitable rooms
	Energy Certificate rating
	Transaction type

Energy tariff

Mains fuel source

Physical Attributes

Property type

Built Form

Mains Gas Flag

Floor level (flats only)

Top Storey Flat (flats only)

Extensions

Wind turbines

Solar Water Heating

Photovoltaics

Accommodation type

Number of Rooms

Fixtures and Fittings

Glazing

Hot Water

Secondary Heating

Central Heating

Demographic

Economic Activity

NSSEC

Marital Status

Tenure

Age group

Cumulatively, the data totalled 1,568,688 PCS level observations (approximately 160 x 7500). It was apparent that a number of these attributes displayed skewed distributions which required further data mining steps prior to clustering, discussed in the subsequent section of this chapter.

5.4.2 Assessing Skew

A test of skewness was applied in favour of a visual inspection of a histogram for each variable to reduce the likelihood of interpretative error and to provide a quantitative measure for comparative purposes. Skewness is a measure of the asymmetry of the distribution of a random variable about its mean and generally a measure of skewness is either positive or negative. A positive skew gives the distribution of a random variable is where the right tail of the distribution is longer and the mean is greater than the median, whereas a negative skew is the inverse, with a long left tail and a mean lower than the median

(Papoulis and Pillai 2002). Three general rules apply when interpreting skew, which were applied to the dataset, with results recorded in Table 5.1 in the appendix; they are as follows:

- If a skewness value falls below -1 or above 1, the distribution is thought to be highly skewed.
- If a skewness value is between -1 and -0.5 or between 0.5 and 1, the distribution is moderately skewed.
- If a skewness value is between -0.5 and 0.5 then the distribution is approximately symmetric.

Table 5.4.2a summarises the information displayed in the appendix, giving the number of variables which fall into each of these categories.

Table 5.4.2a: Skew Distribution Summary.

Skew	Frequency	Percentage
Highly Negative	14	8.9
Moderate Negative	2	1.3
Approximately Symmetric	25	15.8
Moderate Positive	8	5.1
Highly Positive	109	69

The majority of measures that were assessed were highly skewed in their distribution, but it is argued that measures exhibiting skew would either be normalised using power transformations to reduce skewness or used regardless of skew, as the outliers within these measures may assist in producing distinct clusters, as discussed in Singleton and Spielman (2014). For these reasons, no variables were eliminated based on their skewness - especially as some of the measures would be expected to display skewed distributions given the domain. For example, the distribution of elderly populations is not even, because certain areas attract these groups. The same can be said for local area characteristics such as building and property type - inner city areas tend to have more high rise flats than suburban areas and so on.

5.4.3 Data Evaluation

In addition to an assessment of skewness, Exploratory Data Analysis (EDA) was used to understand the extent of correlation between variables. A correlation matrix was generated for the entire dataset to assess this. It is generally discouraged to include highly correlated measures as it can result in duplicate information (Harris, Sleight, and Webber 2005) where multiple measures adequately capture the same relationship. This correlation can be addressed in one of two ways; by omitting multiple highly correlated measures to leave a single variable that is correlated with the largest number of other measures in order to ensure robustness, or alternatively, all measures can be included with or without applying weights. Weighting can be problematic as the process of selecting weights for individual measures can be argued to be subjective (Harris, Sleight, and Webber 2005). The correlation matrix at 165X165 was too large to include in its entirety, but an extract is included as an appendix.

It is possible to summarise the most notable correlations observed between sets of input measures. In general:

- Measures relating to physical properties of and within homes were strongly correlated to one another. These measures also had strong correlations to measures of life stage and age profile.
- Measures relating to current energy ratings were also highly correlated with physical properties of buildings
- Measures relating to the instance of renewable energy sources were not strongly correlated with one another, suggesting that there is little demand for more than one type of renewable energy source once one is present.

It was decided that six variables related to two measures should be removed on the basis of the correlation matrix. These relate to *Top Storey Flat Yes, No* or *Not applicable '(FTS_Y, FTS_N, FTS_NA), 'Solar Water Heating Flag Yes, No or Not applicable' (SWHF_Y, SWHF_N, SWHF_NA)*. The information that can be disseminated from the results of the top storey flat identifier are already covered by another measure *Floor_Level* , likewise solar water heating flags are covered by *Photo_supply* - you cannot have solar water heating if photovoltaic panels aren't in place. Both flat top storey and solar water heating give such high level information that they are attributed to very few certificates overall and relate to so few other variables that it is fair to say that their removal is unlikely to make any significant difference to the final clusters.

Broadly speaking, other correlated variable were not unexpected and have been observed in previous literature. As such, no other measures were removed as a result of this data evaluation step. This decision was made on the basis that removing correlated variables based on the analysis of global statistical relationships could potentially mask local variation and lead to the smoothing of important non-linear patterns at a more granular level, again as discussed by Singleton and Spielman (2014).

5.4.4 Transformation and Normalisation

The next stage of the data evaluation was the consideration of transformation and normalisation procedures. Variable normalisation methods are in simplistic terms adjustments where the aim is to bring distributions into alignment, effectively reducing skew. There are two existing arguments; normalise to minimise outlier effects or embrace 'natural' distributions and allow outliers to influence cluster formation. Two variable normalisation methods are frequently referred to in geodemographic literature; Box-Cox and log10. These were tested alongside a cube-root transformation before a decision was taken. The former two methods require values to be positive and greater than one and as some variables had values that fell below one, a constant of 100 would have to be applied to ensure that transformations could take effect.

log10 transformations, whilst reducing skew in most instances, apply a globally standard method of Normalisation across a dataset, leading to compressed differences between large values and increasing differences between small values to artificially reduce variance. The Box-Cox method uses an exponent lambda to transform a variable (y) and normalise its distribution. Multiple lambda values are tested and the one which results in the most normal distribution and is therefore the best performing is used for the power transformation. This means that the extent to which a variable is transformed is dependent on its level of skew. The Box-Cox method can be expressed as below and is implemented programatically using the forecasts package in R.

$$Yi(\lambda) = \begin{cases} Yi^{\lambda} - 1/\lambda & (\lambda \neq 0) \\ log(y) & (\lambda = 0) \end{cases}$$

Finally the cube-root (x to x^(1/3)), is a fairly strong transformation. Whilst not as strong as the log transformation, it has utility in that it can be applied to zero and negative values without the need to include a constant. The three methods of skew reduction were compared and Table 5.4.4a details the results for a subset of the top ten most highly skewed variables.

Table 5.4.4a Normalisation method results for ten most skewed variables.

variable	Skew (Raw)	Skew (log10)	Skew (Cuberoot)	Skew (Box-Cox)
CER_A	79.12	77.95	1.21	0.07
CER_INVALID!	94.34	94.34	30.53	7.01
HWD_gas_other	91.15	90.84	0.77	-0.14
HWD_heat_pump	66.95	66.90	9.59	7.31
HWD_none	66.42	65.30	0.11	-0.91
HWD_oil	65.40	64.00	2.96	1.55
MF_Community_scheme	75.98	75.11	27.71	15.42
SHD_hot_water_only	79.59	79.58	17.51	14.12
SHD_NA	91.39	91.12	2.34	1.06
WTC_TRUE	92.30	92.10	0.95	-0.10

It is evident that the Box-Cox method significantly improves the overall symmetricity of the variables. Whilst some are still highly skewed, they are much closer to zero than prior to transformation and others are more moderately skewed or approximately symmetrical than before. As such, the Box-Cox method was applied to all skewed variables and two datasets were output from this evaluation; this and the natural (non-manipulated) distribution dataset.

5.4.5 Variable Weighting

An alternative method to normalisation is weighting. In essence, applying weights lessens the impact of heavily skewed variables in cluster formation by reducing influence. The extent to which a variable should be down (or up) weighted is subjective and heavily reliant on the preferences and objectives of the classification builder. As such, this method has been criticised in academic literature (Harris, Sleight, and Webber 2005; Singleton and Spielman 2014). Given the subjective nature of variable weighting, this method was not assessed for implementation.

5.4.6 Standardisation

In the final stage of data evaluation, it was necessary to standardise input variables to ensure that they all fall on a common scale. As such, both the transformed and naturally distributed datasets were standardised using z-scores. This is the most common method for data standardisation and scores are calculated by subtracting the population mean from an individual raw score then dividing the difference by the population standard deviation. This results in a set of scores that are positive if they fall above the means and negative if they fall below, meaning that all standardised variables have an adjusted population mean of 0.

$$z = \frac{x - \mu}{\sigma}$$

Where μ is the mean of the population and σ is the standard deviation of the population.

5.4.7 Initial Clustering

The first stage of the clustering process was to cluster both the transformed and naturally distributed inputs to observe the effects of the transformation on cluster assignments. At this stage, a full classification is unnecessary and an initial run of k = 5 clusters was completed; an arbitrary value. The initial output summary tables as shown in Tables 5.4.7a and 5.4.7b revealed apparent differences in terms of cluster sizes, aggregate characteristics, geographic distribution and ease of interpretability. These summaries are based on a visual, subjective interpretation of the outputted cluster means and have been condensed such that they can be presented below.

Table 5.4.7a Natural Distribution Clusters

Cluster	Certificate A-G	Property Type	Built Form	Transaction	Glazing	Main Fuel	Economic Activity	Tenure	Status	Age
1	E/F/G	Houses	Detached	Sale	Mixed	Solid	Retired	Owned	Married	45 and over
2	D	Houses	Terrace	Social Rent	Double	Gas	Unemployed	Social rent	Separated	0 - 16
3	Mixed	Houses	Semi-Detached	Mixed	Double	Gas	Full time	Mortgaged	Divorced	Mixed
4	Mixed	Houses/Bungalows	Semi-Detached	Sale	Mixed	Mixed	Full time	Mixed	Married	60 and over
5	B/C	Flats	Terrace	Private Rent	Single	Electric	Student	Private rent	Single	18 - 44

Table 5.4.7b Box-Cox Distribution Clusters

Cluster	Certificate A-G	Property Type	Built Form	Transaction	Glazing	Main Fuel	Economic Activity	Tenure	Status	Age
1	Mixed	Mixed	Terrace	Social Rent	Mixed	Mixed	Unemployed	Social rent	Divorced	0 - 19
2	Mixed	Mixed	Mixed	Sale	Mixed	Mixed	Mixed	Owner/Mortgaged	Married	75 and over
3	C	Flats	Terrace	Private Rent	Single	Mixed	Retired	Private rent	Single	20 - 44
4	A	Mixed	Detached	Sale	Mixed	Bio-fuels	Mixed	Mixed	Married	45 and over
5	Mixed	Mixed	Mixed	Mixed	Mixed	Mixed	Mixed	Mixed	Single	Mixed

Initial interpretations suggested that the naturally distributed data produced the most homogeneous and interpretable assignments, with more distinctive clusters formed. The full aggregate statistics support the summarised outputs in Tables 5.4.7a and 5.4.7b. Based on the representation it is apparent that in the naturally distributed cluster assignments :

- Cluster 1 is made up of retired home-owners in large, inefficient housing.
- Cluster 2 represents mostly young struggling families in middling efficiency properties.
- Cluster 3 is a mixture of ages living in semi-detached properties with expensive fuel and heating types but relatively efficient buildings.
- Cluster 4 is made up of middle aged home-owners in energy efficient housing.
- Cluster 5 is mostly students and young professionals renting privately in newly built or renovated properties where energy efficiency is better.

In comparison, the transformed data did not form such easily interpretable clusters. For many of the key variables the cluster assignments are not easily distinguishable and are heavily mixed. This could in part be due to the power transformations, which appear to suppress potentially interesting patterns within the data. The resulting output is a mix; some homogeneous clusters alongside some that are heavily mixed or close to average in the majority of measures, which hinders interpretability. Despite the fact that the transformed dataset produced more evenly sized clusters than the naturally distributed dataset, because more homogeneous clusters offer greater interpretability and ultimately result in a more defined classification, it was deemed that the data displaying the natural distribution would be used to build the final classification.

5.4.8 Construction and Hierarchical Design

The final stage of classification building was to cluster the input measures to form a conclusive classification. As discussed, one method of selecting the initial number of clusters is to use 'Clustergrams'; plots which aid the interpretation of an optimum value of k, by visualising the distribution and redistribution of observation between values for a range of k values. As seen in Figure 5.2.2.9a, the Clustergram tested the value of k from 2 through 10 and allows an optimum value to be selected by enabling identification of assignment and reassignment and also allows visualisation of the relationship between new and existing clusters as the value of k increases. For each iteration the method works by multiplying the cluster centres by the first loading of the principal components of the original data, thus offering a weighted mean of each cluster's centre dimensions, as indicated by the red point.

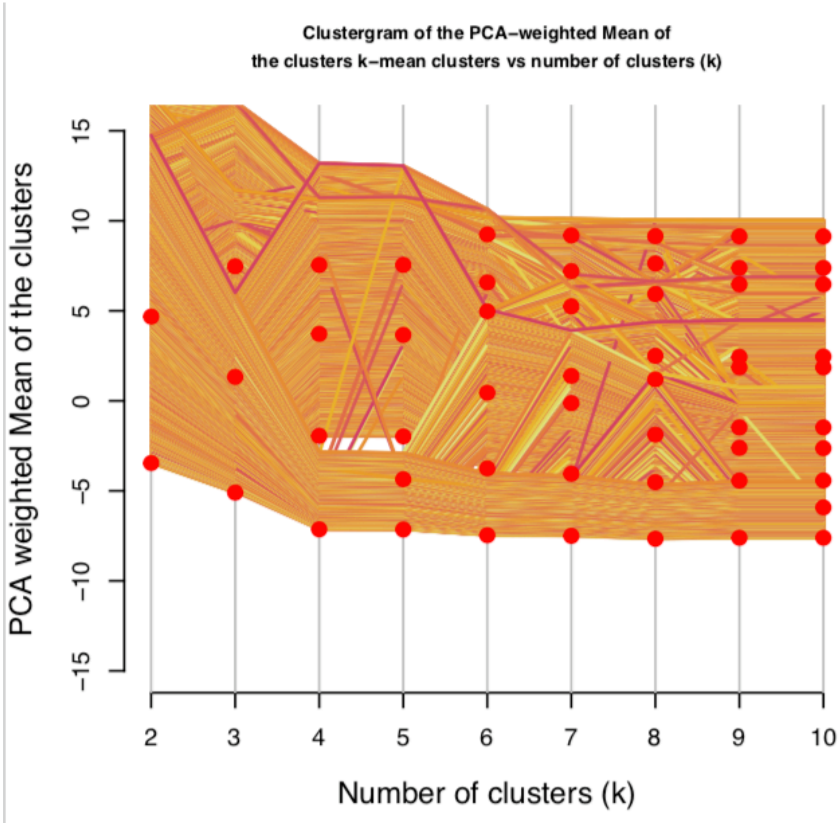
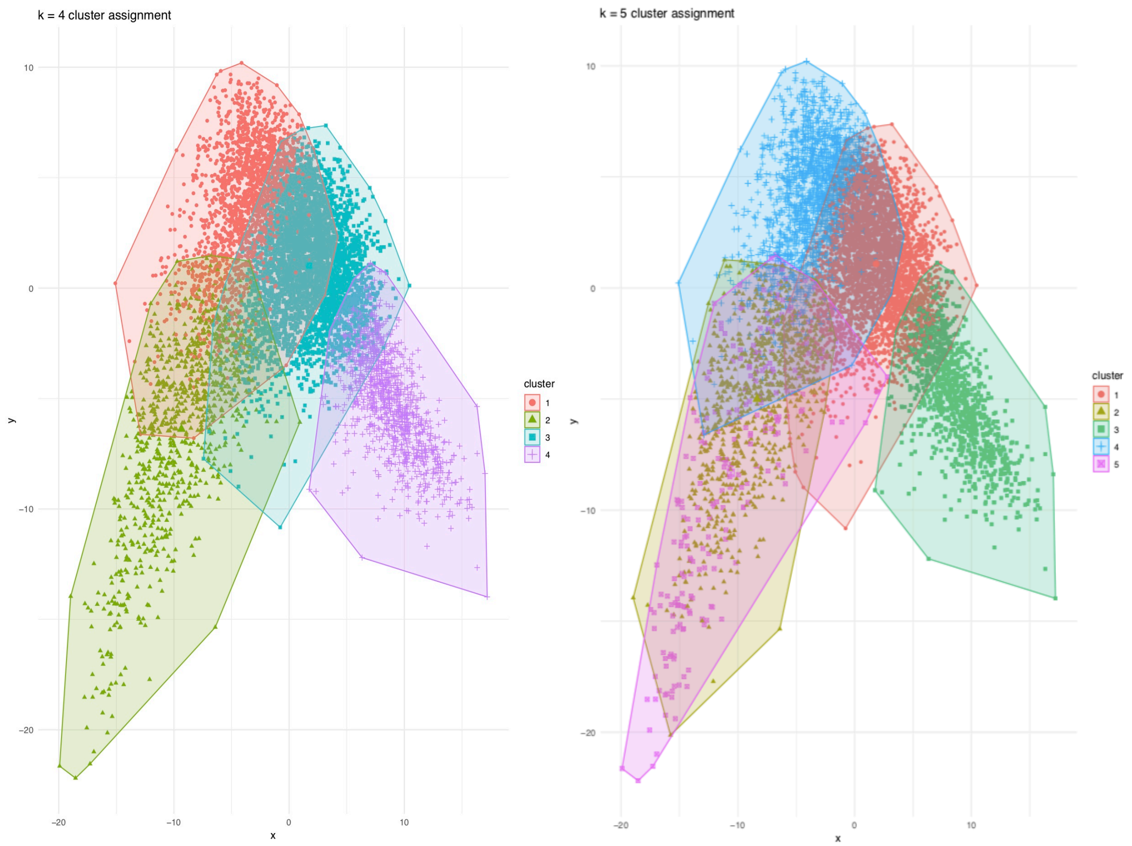


Figure 5.4.8a Clustergram iterations for k = 2:10

The figure above represents one iteration of the clustergram algorithm. The analysis was iterated six times to test for robustness, and each plot inspected. In this excerpt it is apparent that when k = 2, observation assignments are relatively even and the two clusters are well spaced. The spacing suggests that the two clusters are sufficiently homogeneous in terms of their characteristics that they are easily distinguishable from one another. In this context, they would likely represent two very different groups of energy consumption characteristics. As the number of k is increased to 3, it is possible to track the reassignment of observations. In this example, a number of observations from the upper and lower are reassigned to form a central cluster, and some observations from the upper and lower clusters are reassigned to each other. At k = 4, some iterations see a new middle cluster formed, and some see the lower cluster and middle cluster split to form a cluster overlapping the lower, which may impact the interpretability and reinforcing the need for multiple iterations of the clustergram. At k = 5 the cluster assignments are well spaced, which is also the case for some iterations of k = 6, but in other cases and for most iterations of k = 7, cluster assignments become less evenly spaced, with some overlap. This suggests the optimum value for k has been exceeded.

Given this evaluation, a K-Means clustering algorithm was undertaken for k values of 4 and 5. As the dataset is relatively compact, this was not expensive in terms of computational effort. The two cluster assignments were visualised as can be seen in Figure 5.2.2.9b.



The final number of clusters was decided after the visual interpretation of these figures. As shown in Figure 5.4.8c, at k = 5 the clusters have a significant amount of overlap, with, for example cluster 2 being almost completely covered by cluster 5. This is likely to make interpreting the characteristics of these clusters difficult, so given the relatively small amount of overlap present in k = 4 (Figure 5.4.8b), this assignment appears the most likely to provide easily interpretable and homogeneous clusters. The clustering run was repeated over 10,000 iterations to ensure an optimal result. Each of the four initial clusters were then separated and re-clustered in an attempt to build a second tier within the classification to give a more granular final typology. However, on investigation it was apparent that most clusters were unable to support results significantly different from the parent cluster and so the second tier was not investigated further. Following this final categorisation, the resulting classification was a single tier typology containing four clusters. For the purpose of this classification and for consistency with similar literature, clusters were renamed 'Supergroups'. The next stage was to study the characteristics of each and translate this information into a set of descriptive summaries.

5.5 Cluster summation

The process of summarising Supergroup characteristics was achieved using a number of methods. Firstly, summary tables describing the mean for each input attribute were recorded and inspected visually. To aid interpretation, handwritten notes of cluster characteristics were made and have been included in appendix 5. From these, key characteristics were recorded and provided the basis for the resulting textual summaries ('Pen Portraits'). Secondly the clusters were mapped using a Geographic Information System (GIS) to reveal their geographic distributions and were classified as one of the following; deep rural, rural fringe, major urban, towns or urban fringe areas. By combining this information, provisional names were created for the Supergroups and full Pen Portraits compiled.

5.5.1 Supergroup 1 : Cold and Costly

Figure 5.5.1 shows the national distribution of the first Supergroup in the classification. It is apparent that this Supergroup is concentrated around major urban and suburban areas and towns, which typically attract younger populations and families. Analysis of the cluster means resulted in the Supergroup name “Supergroup 1: Constanty Cold and Costly” and the following Pen Portrait:

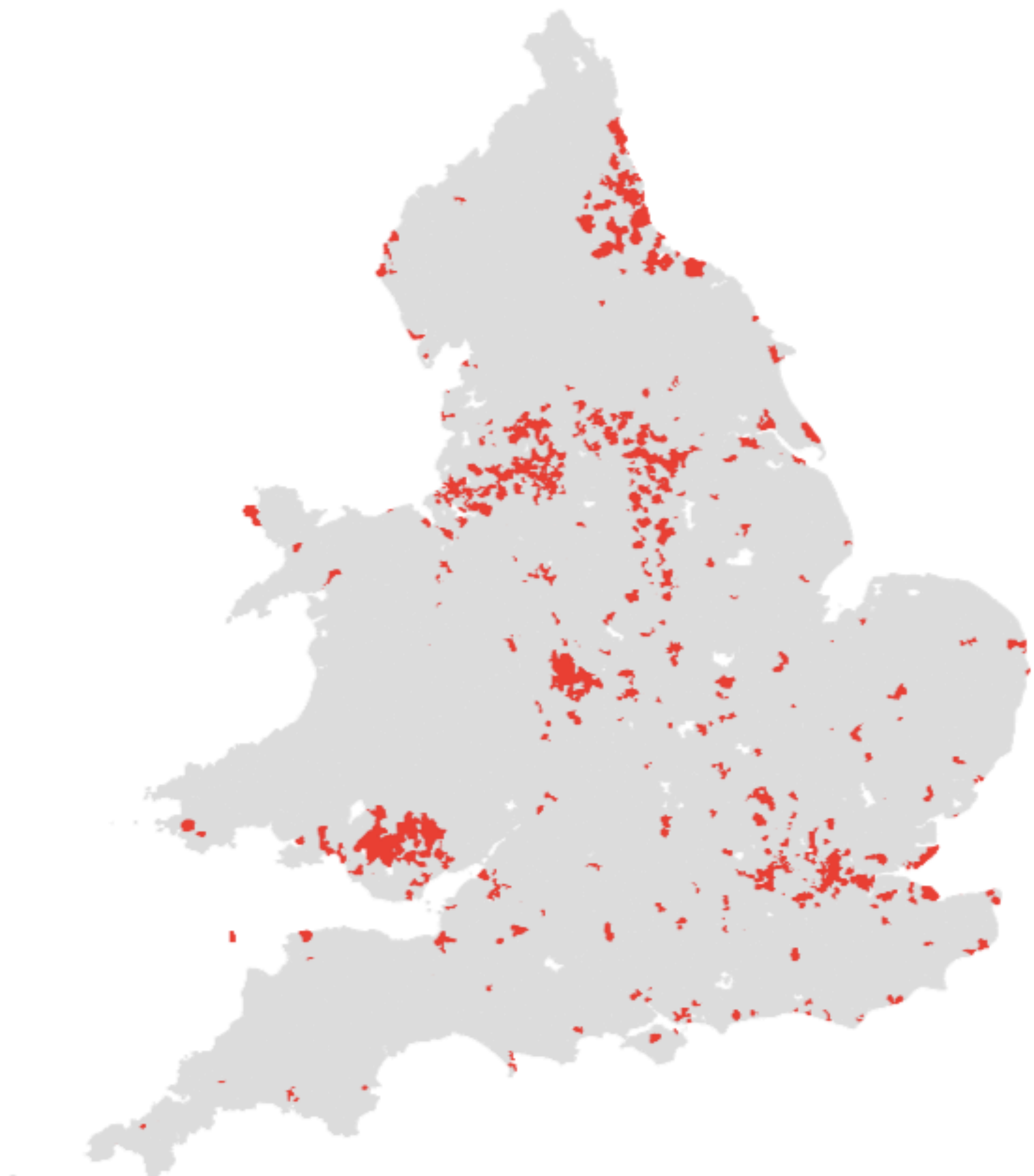


Figure 5.5.1 Supergroup 1 National Distribution

"The Constantly Cold and Costly Supergroup are characterised by the likelihood of their homes being underheated or costly to heat. Members of this Supergroup are most likely to be housed in socially rented accommodation and therefore suffer from the tenant/landlord dichotomy, giving them little to no autonomy over the cost of their consumption. A high proportion use the more expensive single energy tariffs which are associated with pre-payment meters and despite the fact there is some evidence of minor structural improvements being made such as high levels of double glazing and EPCs generated from upgrade projects, the majority of homes are still most likely to be Band D or below. They are typically terraced or semi-detached houses, occupied by families with children. The adults in this Supergroup are most likely to be long term unemployed, disabled or working in routine and semi-routine occupations, all of which combine to make this group the most likely to struggle to heat their homes consistently without becoming fuel poor, especially during colder months when they have been unable to build up any credit with their energy supplier to cover the increased usage. Comprising of 1921 postcode sectors (25%) and 25% of the population, it is the 2nd biggest cluster.

5.5.2 Supergroup 2 : Winter Fuel Allowance

Figure 5.5.2 shows the national distribution of the second Supergroup within the classification. Unlike the first, it is clear that this Supergroup is widespread nationally, covering most deep rural and rural fringe areas. This Supergroup does not cluster around urban areas and towns those areas associated with younger populations. Analysis of the cluster means resulted in the Supergroup name "Winter Fuel Allowance" and the following Pen Portrait:

Figure 5.5.2 Supergroup 2 National Distribution

The “Winter Fuel Allowance” Supergroup are characterised by their age and location, as well as varying levels of engagement with their energy usage. They are most likely to be aged over 45 or elderly and married home-owners with few young children. They typically live in detached houses with a large floor area and are the most likely to be working part time, or for themselves, or are retired. Of those that are energy aware, there is significant evidence of them exercising their autonomy over their energy consumption and its costs by making long term investments in their properties. They undertake general and efficiency based home improvements such as constructing extensions, installing solar panels and having energy upgrade work to improve their homes. This leads to this Supergroup having a high proportion of properties with an A energy rating. However, those who are disengaged and living in very remote areas rely on inefficient and expensive non-standard means of energy consumption such as oil or wood and are less likely to be connected to the mains gas network. These large, rural homes are likely to be inefficient and underoccupied and as such, costly to heat to a comfortable level. It is the second half of this Supergroup who may find themselves with increased costs during colder months; the energy engaged are in a position to have some credit with their fuel supplier as they have dual fuel tariffs characterised by monthly direct debit payments to absorb the change. The “Winter Fuel Allowance” Supergroup is made up of 1066 postcode sectors (14%) and accounts for 10% of the population.

5.5.3 Supergroup 3: Efficient City Living

Figure 5.5.3 shows the national distribution of the third Supergroup in the classification. This Supergroup has a tendency to cluster around major urban areas; predominantly London, but cities such as Manchester, Liverpool, Birmingham and Bristol can also be picked up on the map. Analysis of the cluster means resulted in the Supergroup name "Efficient City Living" and the following pen-portrait:

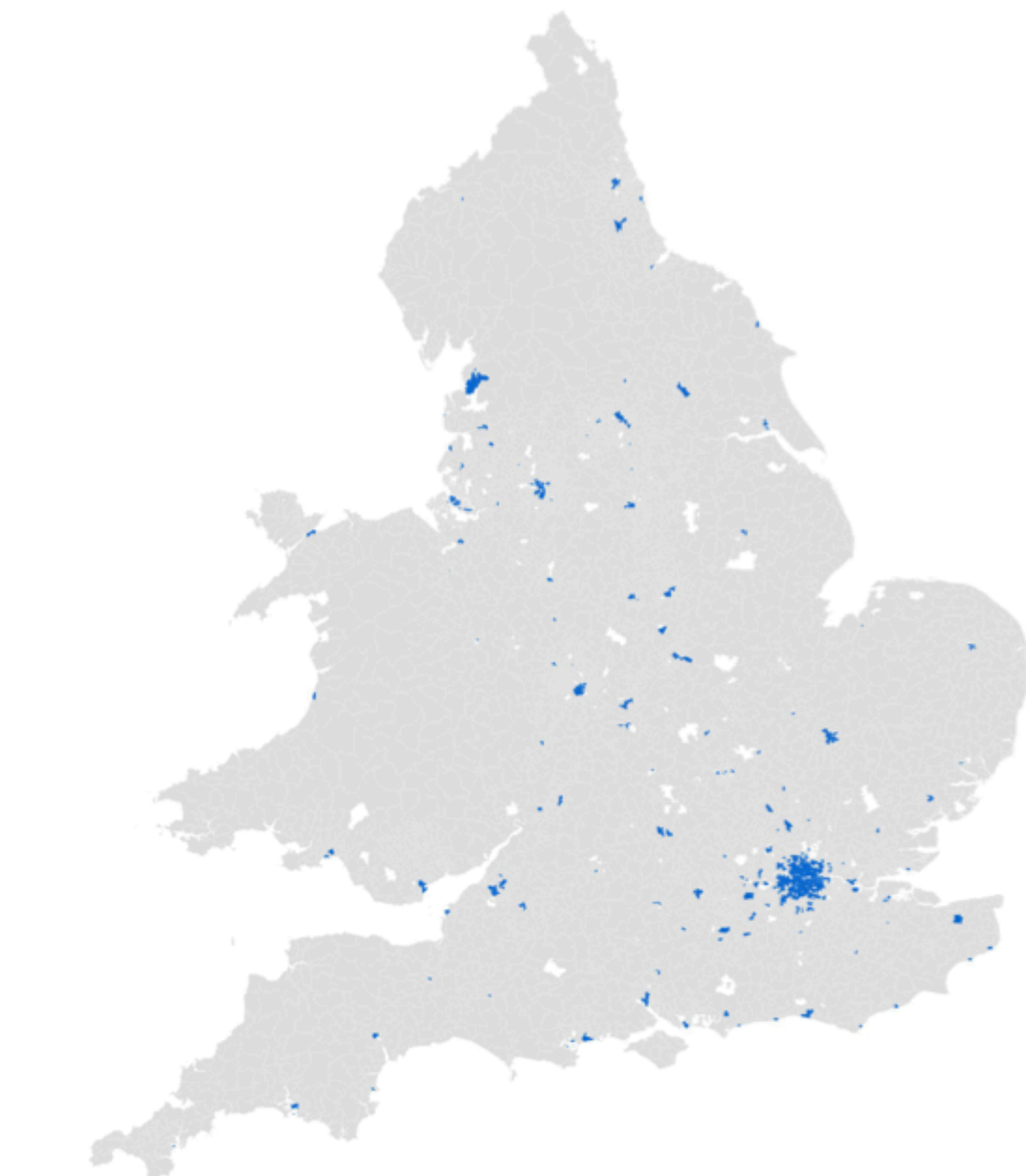


Figure 5.5.3 Supergroup 3 National Distribution

The “Efficient City Living” Supergroup are characterised by their age and geographical location. They are young, working in higher managerial and professional occupations, and are living in major cities across the country. They typically live in privately rented flats or houses, which are newly built or purposely converted with updated, efficicent fixtures and fittings such as triple glazing and modern boilers. This coupled with the fact that they typically have a lower square footage to heat and light means this Supergroup are the least likely to find themselves with high energy bills they are unable to alter. Their relatively high income allows them to absorb shocks to their bills and so even in the cases where expensive fittings such as immersion heaters are found, this group are the least likely to find themselves in fuel poverty. Some of this cluster are students (Lancaster University is clearly visible on the map) living in purpose built halls of residence, whose all inclusive living arrangement means that whilst they have no autonomy over their energy efficicency, they also do not need to consider energy bills as an extra cost and so, will not find themselves in fuel poverty. They are the smallest group, made up of 912 postcode sectors (12%), and are 10% of the population.

5.5.4 Supergroup 4: Typical Tariff

Figure 5.5.4 shows the national distribution of the fourth Supergroup of the classification. This Supergroup is mostly split between towns and urban areas, but it is not uncommon for rural areas to be included. Analysis of the cluster means resulted in the Supergroup name "Typical Tariff" and the following Pen Portrait:

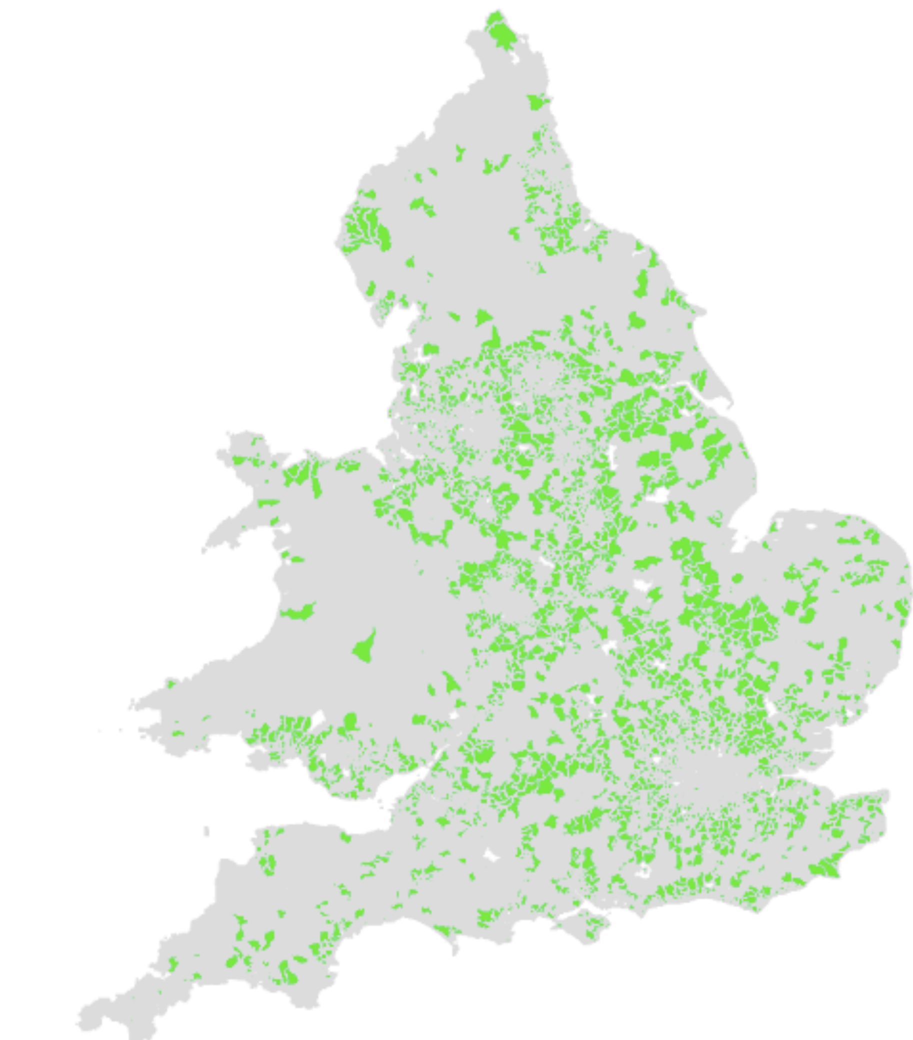


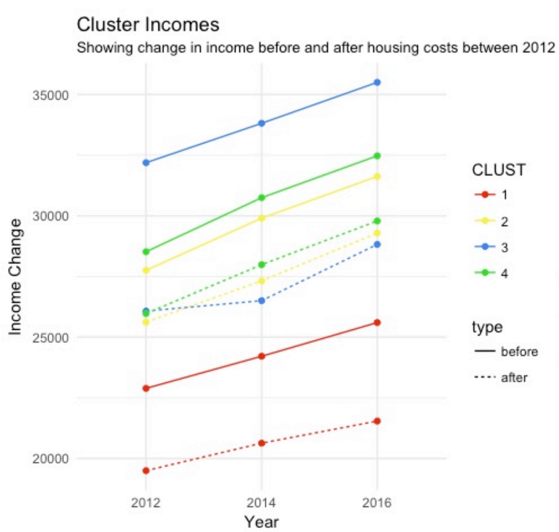
Figure 5.5.4 Supergroup 4 National Distribution

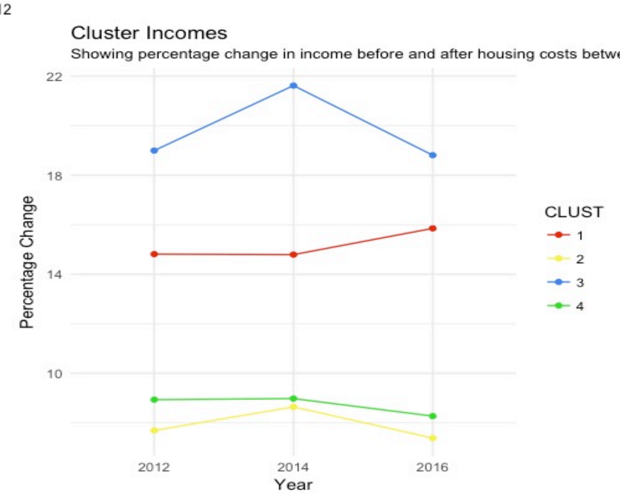
The “Typical Tariff” Supergroup displays characteristics that are closest to the overall averages. Areas are characterised by mixed energy efficiency and average floor area, and display a variety of fixtures, fittings and physical property attributes. Homes are typically semi-detached and are mostly mortgaged. There is a higher proportion of elderly people but an overall mix of ages and family types. Members of this Supergroup who are of working age are typically in middle or lower supervisory jobs. There are fewer shared houses and private rentals than other clusters. It is the largest cluster, accounting for 49% of postcode sectors (3762) and 52% of the population.

5.6 Using the EUC to rethink Fuel Poverty

From this new classification it can be seen that demographic indicators as well as the energy efficiency characteristics of homes could be used to understand and define fuel poverty from a ground up approach to improve targeting and allow the fuel poverty alleviation to remain as low cost as possible whilst also effectively targeting those most vulnerable. Under the current definition, low income is one of the prevelent factors in defining the fuel poor and so far has not been considered as part of the new classification. To conclude this analysis, the intersection between clusters and income has been investigated and then compared to the current fuel poverty distribution to understand the similarities and differences from that which already exists, in order to prove utility in a multifaceted approach in redefining fuel poverty.

5.6.1 Cluster incomes





It is clear that the Constantly Cold and Costly Supergroup are the lowest earners by some margin. They are also seeing the gap between BHC and AHC increase over time and whilst they are not the group with the largest financial burden of housing cost, relative to their overall income they are the only group to have seen an increase in the percentage of housing cost. Their demographic characteristics lend themselves to having other significant costs such as those associated with disabilities and expensive energy tariffs.

The Supergroup 'Efficient City Living' find themselves paying a substantially bigger proportion of their income towards housing cost due to their location in major towns and cities where rentals and sales command considerably higher prices. They have seen the gap begin to close over time, giving them a larger proportion of their income available for other costs, but they are unlikely to consider using this money for improvements or energy bills as they already live in efficient housing, and do not generally struggle to cover their energy costs.

Both the 'Winter Fuel Allowance' and 'Typical Tariff' SUpergroups have seen a small decrease in terms of percentage of income accounted for by housing cost, and a steady rise in both incomes and housing costs overall leads them to be the most stable of the clusters, possibly able to react better to both housing and energy cost price changes. The slight percetage decrease in housing cost may allow those who previously would have been in short term fuel poverty to plan for seasonality and reduce their energy bills by making some investment into the efficiency in their homes; for example, by considering replacing boilers or the prevelent single glazing, or by moving away from solid fuels.

5.6.2 Cluster comparisons to current Fuel Poverty

When the geographic distribution of the clusters is compared to the current fuel poverty map, and in particular the highest fuel poverty quintile, interesting patterns begin to occur.

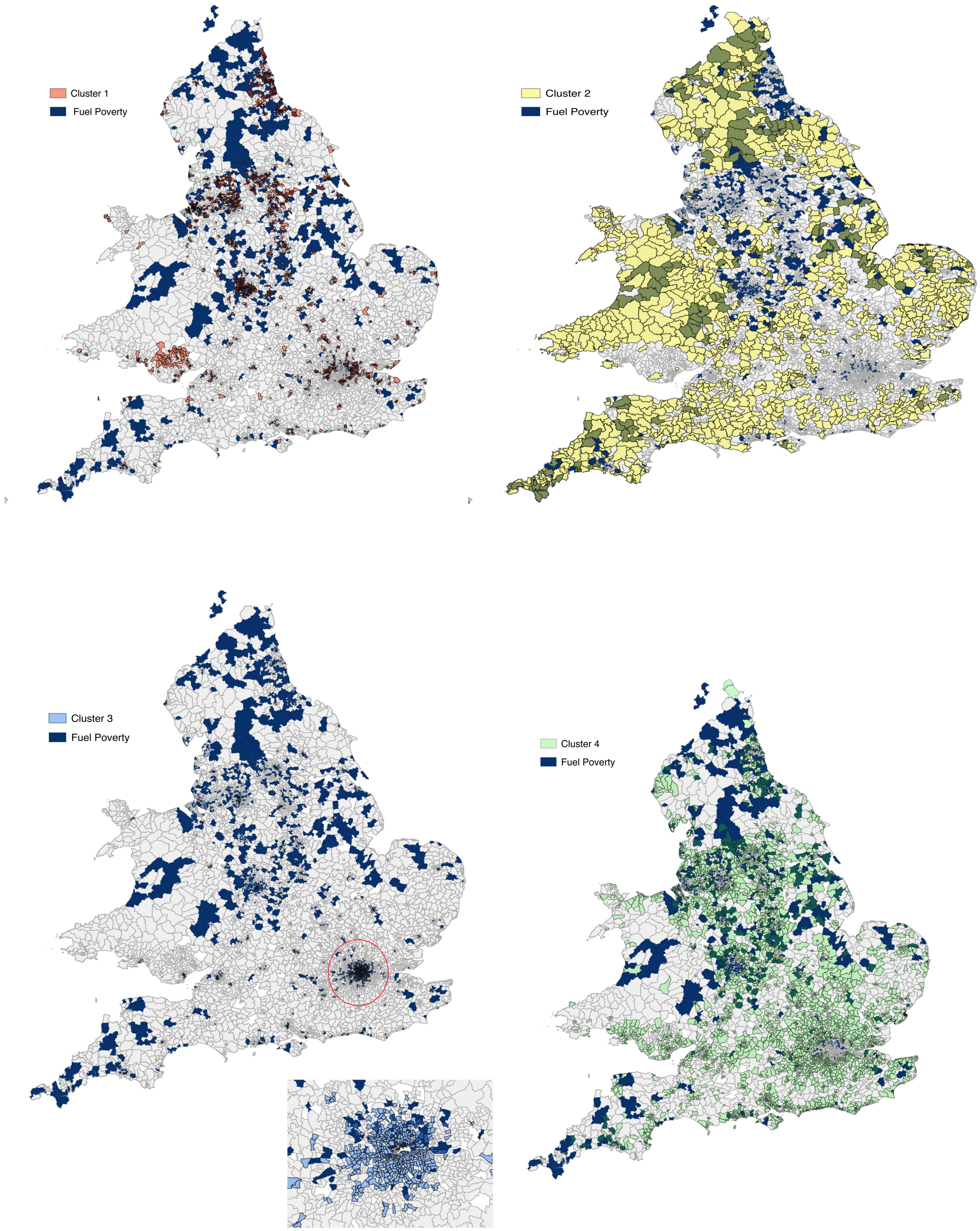

Cluster 1
Fuel Poverty
Cluster 2
Fuel Poverty
Cluster 3
Fuel Poverty
Cluster 4
Fuel Poverty

Table 5.6.1a Cluster Income Quintile Frequency

Cluster	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
1	0.62	0.17	0.11	0.09	0.02
2	0.08	0.25	0.27	0.25	0.15
3	0.15	0.09	0.11	0.16	0.49
4	0.08	0.22	0.22	0.24	0.24

The fact that the group of areas with the highest percentage of fuel poor households as understood by the current fuel poverty definition are represented (in varying degrees) within every cluster of our classification indicates that the current definition of "Low Income High Cost" is too simplistic and leads to other factors being overlooked.

Supergroup 'Constantly Cold and Costly' are largely in the lowest income quintile, and it is fair to say that it is highly likely that income is a factor in those areas which are already classified as fuel poor. Figure 5.6.1c shows that there is some overlap with the current FPI (fuel poverty indicator), especially in major urban areas and most clearly in the suburbs of central London. These areas are typically characterised by hard to heat, inefficient homes, those who find themselves unemployed and young families. It is true that this group are the best fit for the Low Income, High Cost indicator, but there are still areas in the which are considered fuel poor which do not appear in the EUC (Energy User Classification) cluster. It is possible that it is their inability to make changes and their disengagment from the the consumption process which leads to them being trapped in homes which they cannot afford to heat, on energy tariffs they are unable to change.

The 'Winter Fuel Allowance' Supergroup have a low percentage of people in the lowest income quintile, but as can be seen in Figure 5.6.1d, there is a significant amount of overlap with the current FPI, especially in more rural areas. Most households in this cluster are evenly spread between income quintiles 2,3 and 4, suggesting that income is less relevent in causing their fuel poverty, and it is more likely to be linked to old, inefficient and underoccupied buildings. These conclusions can be drawn from what we know about the demographic characteristics of this group - they are typically middle aged or retired, living in properties with a high floor area. They are much less likely than other groups to be connected to mains source gas and rely on oil or solid fuels as their main source of energy. These households may have higher incomes and relatively higher costs, and so are ignored under the current definition, but nonetheless struggle to heat their homes to a comfortable temperature.

The 'Efficient City Living' Supergroup are the highest earners, as seen in both Figure 5.6.1a and Table 5.6.1a and there is very little overlap with the fuel poverty layer, except for in North London, as can be seen in the insert map in Figure 5.6.1e. This may suggest that despite some people being high earners, they are still succeptible to struggling to cover their energy costs, possibly as such a high percentage of their income is taken up in the unavoidable cost of housing. Those in this cluster who do fall into the lower income quintiles are more likely to be students, living in new, purpose built accomodation and so are earning very little, but do not fall into fuel poverty because of access to student loans and all inclusive living arrangements where their bills are taken care of. They may also be The exception in this cluster, who are struggling and unemployed and finding it hard to meet the higher costs of inner city living. The current fuel poverty definition overlooks those who are considered High Income, High Cost. The lack of overlap across the rest of England and Wales could be attributed to the energy and demographic characteristics uncovered during the clustering process - this cluster are the most likely to have energy efficient properties, a low floor area, and access to the cheapest tariffs indicating lower energy costs.

Despite the 'Typical Tariff' having a small percentage of households in the lowest income quintile, there is still a significant amount of overlap visible between the group and the fuel povery layer in Figure 5.6.1f, especially in the North and Midlands. The majority of these households are evenly spread between the top 4 income quintiles, and the pen profile of this cluster suggests that they are the homes with the most average characteristics, reiterating that the characterisation of fuel poverty as a 'Low Income, High Cost" is a misleading one. It could possibly be said then that most people with the most average characteristics are on some level at risk of fuel poverty at any time, and only those with the highest wages and lowest costs are unlikely to struggle to meet costs, but one or the other is not enough to guarentee thermal comfort.

5.7 Conclusions

This chapter has shown that both the 10pc and Low Income High Cost definitions of fuel poverty are lacking in terms of the lived experience of fuel poverty and it has re-evaluated the ways this multifaceted issue should be considered outside of it's technical and structural problematisation. The income analysis showed the greatest levels of instability for the lowest income earners for their housing costs, which has numerous repercussions; it is harder for them firstly to plan for these changes and increases, but also harder for them to recover from financial shocks as they are the least likely to be able to save any of their disposable income to account for these. This leads to debt, and restricts energy payment options which are available to them, often leaving only the most expensive. As such, they are more likely to find themselves in short term fuel poverty in the winter as their costs change, and may find themselves in perpetual fuel poverty if the gap between before and after housing cost continues to grow. All other income quintiles have seen the gap between before and after housing cost begin to close, giving them greater stability. When the cluster income spread is considered, it is clear that the lowest earners are again the most vulnerable to instability, although cluster 3 as the highest earners spend a significantly higher percenatge of their income on housing costs.

The variations in energy efficiency rating between the OAC groups suggested that energy efficiency cannot be soley linked to the structural and physical aspects of a household. There were clear differences in the energy ratings of varying demographics and whilst building type and structural properties affecting energy efficiency were present, other factors such as rurality, ethnicity, employment and age were also at play.

The clustering provided a novel insight into the Energy Performance Certificate Data as well giving extra detail on the charcteristics most likely to apply those who may find themselves in fuel poverty other than energy efficiency.

By combining the EPC, census and income data, the visualisations show the intersection between the clusters and income and provide clarification that fuel poverty is a multidimensional issue by showing that fuel poverty characteristics are present in areas which are currently overlooked by the fuel poverty definition. It also shows that in each cluster of the EUC, there are areas which are currently considered fuel poor, suggesting that the definition does not encompass all facets of fuel poverty.

Overall it is true to say that energy efficiency does have an effect on a households likelihood of being fuel poor, but it is not the only factor and by using the current definition, there are many other characteristics that are overlooked and left unresolved by the current targeting. The EUC provides utility in defining these other characteristics in order to improve targeting and suggesting ways other than improvements in energy efficiency which may go some way to alleviating fuel poverty both in the most vulnerable and overall.

5.8 References

Boardman, Brenda. 1991. *Fuel poverty: from cold homes to affordable warmth*. Pinter Pub Limited.

———. 2013. *Fixing fuel poverty: challenges and solutions*. Routledge.

Bradshaw, Jonathan, and Sandra Hutton. 1983. “Social policy options and fuel poverty.” *Journal of Economic Psychology* 3 (3-4). Elsevier: 249–66.

Committee on Fuel Poverty. 2018. “Third Annual Report.” London.

DECC. 2013. “Annual report on fuel poverty statistics 2013.” London.

DEFRA, and DTI. 2005. “The UK Fuel Poverty Strategy: 3rd Annual Progress Report 2005.” London: DTI.

DTI. 2001. “The UK Fuel Poverty Strategy.” London.

Energy Saving Trust. 2019. “Energy Performance Certificates.” http://www.energysavingtrust.org.uk/home-energy-efficiency/energy-performance-certificates.

Gale, Christopher G, A Singleton, Andrew G Bates, and Paul A Longley. 2016. “Creating the 2011 area classification for output areas (2011 OAC).” *Journal of Spatial Information Science* 12. University of Maine: 1–27.

Guertler, Pedro, and Sarah Royston. 2013. “Fact-file: The cold man of Europe.” *Association for the Conservation of Energy, London*.

Harris, Richard, Peter Sleight, and Richard Webber. 2005. *Geodemographics, GIS and neighbourhood targeting*. Vol. 7. John Wiley; Sons.

Hills, John. 2012. "Getting the measure of fuel poverty: Final Report of the Fuel Poverty Review." Centre for Analysis of Social Exclusion, London School of Economics and~?

HM Government. 2014. "Cutting the cost of keeping warm: A new fuel poverty strategy for England." London: DECC.

Middlemiss, Lucie. 2017. "A critical analysis of the new politics of fuel poverty in England." *Critical Social Policy* 37 (3). SAGE Publications Sage UK: London, England: 425–43.

Middlemiss, Lucie, and Ross Gillard. 2015. "Fuel poverty from the bottom-up: Characterising household energy vulnerability through the lived experience of the fuel poor." *Energy Research and Social Science* 6: 146–54. https://doi.org/10.1016/j.erss.2015.02.001.

Moore, Richard. 2012. "Definitions of fuel poverty: Implications for policy." *Energy Policy* 49: 19–26. https://doi.org/10.1016/j.enpol.2012.01.057.

NRFC. 2013. "About Us." http://right2fueluk.com/about-us/.

OFGEM. 2008. "Energy Supply Probe–Initial Findings Report." Ofgem London.

Papoulis, Athanasios, and S Unnikrishna Pillai. 2002. *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education.

Singleton, Alexander D, and Seth E Spielman. 2014. "The past, present, and future of geodemographic research in the United States and United Kingdom." *The Professional Geographer* 66 (4). Taylor & Francis: 558–67.

Sleight, Peter. 2004. *Targeting customers: How to use geodemographic and lifestyle data in your business*. World Advertising Research Center Henley-on-Thames.

Vickers, Dan, and Phil Rees. 2007. "Creating the UK National Statistics 2001 output area classification." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170 (2). Wiley Online Library: 379–403.