PLOS COMPLEX SYSTEMS

RESEARCH ARTICLE

Food purchase data reveals the locations of London's 'food deserts'

Tayla P. Broadbridge[1,2]*, J. Edward F. Green[1], Simon P. Preston[2], Nabil T. Fadai[2], John Maclean[1]

1 School of Computer and Mathematical Sciences, University of Adelaide, Adelaide, South Australia, Australia, **2** School of Mathematical Sciences, University of Nottingham, Nottingham, United Kingdom

* tayla.broadbridge@adelaide.edu.au

OPEN ACCESS

Citation: Broadbridge TP, Green JEF, Preston SP, Fadai NT, Maclean J (2025) Food purchase data reveals the locations of London's 'food deserts'. PLOS Complex Syst 2(11): e0000072. https://doi.org/10.1371/journal.pcsy.0000072

Editor: Zhen Zhu, University of Kent, UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND

Received: May 1, 2025

Accepted: September 17, 2025

Published: November 6, 2025

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pcsy.0000072

Data availability statement: All data used for analysis is publicly available and is included in the reference list. Boundary shapefiles are

Abstract

A nutritious diet is essential for preventing diet-related diseases. In the UK, obesity and related diseases are leading causes of death, with more than half of London's residents classified as overweight or obese. 'Food deserts' refer to areas where residents are unable to access a nutritious diet, where barriers to obtaining healthy foods are thought to underpin dietary behaviour. Previous attempts to identify 'food deserts' have relied on assumptions about the relationships between store locations, sociodemographic factors, and access to healthy food. These methods typically classify areas as 'food deserts' without any direct, quantitative link to food purchase data or dietary patterns. By utilising food purchase records from Tesco transactions, we explore the relationship between food purchasing patterns and sociodemographic factors in London, with a focus on identifying food deserts and their drivers. Food purchasing patterns vary spatially, with significant spatial clustering of nutritionally deficient food purchases across London's boroughs. These clusters are statistically explained by sociodemographic factors using a geographically weighted regression model, which enables the exploration of how the influence of sociodemographics, walk time, and car ownership varies across different areas of London. Our findings demonstrate the potential of analysing food purchase data to identify food deserts and their drivers, and suggest that area-specific, context-sensitive interventions are necessary for the implementation of local public health strategies.

Author summary

Poor diets are a major risk to health, contributing to 13% of deaths in the UK, and over half of London's residents are now overweight or obese. While the concept of 'food deserts' are commonly thought of as areas lacking nearby supermarkets, in urban environments the issue is more complex—as barriers such as affordability and availability are also prevalent. In this study, we analyse supermarket transaction data from 1.6 million London customers to understand how food purchase patterns vary across the city. We apply an unsupervised statistical method to identify a dominant purchasing pattern,

sourced from the Greater London Authority at https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london. Shapefile licence information: Contains National Statistics data Crown copyright and database right [2015]. Contains Ordnance Survey data Crown copyright and database right [2015]. Code used to generate results can be found on GitHub at https://github.com/taylabroadbridge/paper-fooddeserts-gwr.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

which distinguishes between diets high in sugar and carbohydrates to those richer in fibre and protein. We then explore how these patterns relate to local conditions using geographically weighted regression, finding that both the drivers of food deserts and the demographics most affected by them vary widely across London. This study highlights the value of large-scale consumer data for understanding urban health challenges and provides a new data-driven way to identify areas where barriers to healthy diets exist.

1. Introduction

Poor diet and nutrition are a leading global risk to health [1], and accounts for 13% of deaths in the UK [2]. Diets high in processed foods, sugars, and fats are known to cause obesity and are associated with a range of negative health outcomes including hypertension, heart disease, and diabetes [2]. In recent years, 61% of London's adult population were classed as overweight or obese [3], and childhood obesity rates (for ages 10–11) were higher than the England average [4]. Increased production of processed foods, rapid urbanisation, and changing lifestyles have led to a shift in dietary patterns [1]. Surveys indicate that only 29% of England's adults consumed the recommended portions of fruit and vegetables per day, whilst 7% consumed none at all [5]. As of 2025, children are consuming less than half the recommended amount of fruit and vegetables but over twice the recommended amount of sugar [6]. Barriers to accessing a nutritious diet vary from food availability and affordability, to lack of knowledge about what constitutes a healthy diet [7]. Governments play a central role in creating and maintaining a healthy food environment that enables people to adopt and maintain healthy dietary practices [1]. However, to effectively target interventions to areas where residents face barriers to accessing healthy food, area identification is necessary.

Areas where residents face barriers to a healthy and affordable diet are often referred to as 'food deserts' [8]. The 'food desert' metaphor became prevalent throughout the late 1990s in social exclusion and health inequalities debates in the UK. Major food retailers were considered responsible for the emergence of poor dietary behaviour due to their failure to establish supermarket locations in low-income communities, and thereby denying residents access to affordable food [9]. From the early 2000s, improving retail access to food in urban areas was the focus of dietary intervention research in the UK [10–13]. Several studies have since argued that supermarket provision and reducing the distance from residents' homes to supermarkets does not have a significant affect on food purchasing patterns unless combined with other factors such as price and availability [14]. Further, treating food access as only a store distribution problem ignores critical factors such as racial and economic landscapes that shape residents' urban life and mobility, and consequently, their shopping behaviour [15]. A study by authors of [16] found that exposing low- and high-income households to the same food only reduced nutritional inequality by 10%, and, a scoping review of adult food choices identified high food prices and lack of transportation to be the major barriers to food access [17]. Modern studies have motivated the need for the 'food desert' metaphor to encapsulate a multidimensional description of food access, which is classified into five key domains: spatial accessibility (configuration of food stores), affordability (price of healthy food options), availability (quantity of retailers stocking healthy foods), accommodation (ability of retailers to accept alternative payments), and acceptability (stock of culturally appropriate foods) [18–20]. This idea is particularly important in urban areas, where residents generally live within close proximity to supermarkets, and since poverty rates are consistently higher than in rural areas in the UK [21].

Although recent research has identified the barriers that impair food access and health outcomes, there remains no consistent methodology to characterise areas as 'food deserts'. A widely used approach to identify food deserts was developed by the United States Department of Agriculture (USDA), which characterised food deserts as census tracts that are low-income and have limited spatial access (1 mile for urban and 10 miles for rural areas) to grocery stores [22]. The USDA Food Access Research Atlas identifies limited spatial access to healthy food based on distinct distance thresholds for urban and rural areas, although, a variety of thresholds are available [23]. Researchers at the Consumer Data Research Centre (CDRC) developed a food desert mapping tool which assigns areas as food deserts using weightings for different access domains such as income, transport, store distance and e-commerce access, where each domain of access has an equal weighting [24]. This method has since been extended upon to incorporate financial barriers, such as fuel poverty, in response to the cost of living crisis, to highlight areas where intervention has become a priority [25]. Such methods rely on predefined thresholds and weightings of sociodemographic and environmental characteristics, rather than being directly driven by data on food purchases. The variety of thresholds, weightings, and dimensions used in these approaches limit consensus in food desert identification, and hence, the understanding of their drivers. Before the relative contribution of such drivers can be quantified, a food purchase data-driven approach is important for identifying food deserts in alignment with food consumption behaviours.

This manuscript addresses the gap in the literature owing to two points from the previous paragraphs: that urban food deserts are acknowledged to be caused by multiple factors, and that there is no consistent way to define them. To fill this gap, we identify food deserts in London, defined as areas where residents are purchasing nutritionally deficient foods. We then perform a statistical analysis to discover the factors that are most associated with each 'food desert'. The conclusion is that these factors are strongly suggestive of the drivers of food desert behaviour.

The use of geographically aggregated consumer data offers valuable insights into dietary behaviours and an advantage for analysis. A study of Leeds, UK used geographically aggregated food purchase records to identify areas where residents purchased a lack of fruit and vegetables, and linked such purchases to sociodemographic characteristics, shopping frequency, and distance to supermarkets [26]. Similarly, food-sharing data has recently enabled the development of models to map food insecurity in London [27,28]. A study by authors of [29] has linked diet-related diseases to food purchases using Tesco loyalty card data in London, finding that areas with higher prevalence of metabolic syndrome (hypertension, high cholesterol and diabetes) also purchase considerable amounts of carbohydrates and sugar. Health outcomes vary across London's boroughs [29,30], suggesting that sociodemographic and environmental factors may play a role in these disparities. While there is a growing body of literature exploring the quantification of food deserts [24,31,32], such methods have not been directly linked to dietary or food purchase records. More recent approaches use mobility data and online food delivery records to infer dietary behaviour based on exposure or visitation patterns [64,65]. These studies offer valuable insights into out-of-home eating environments but do not capture in-home food purchasing. Supermarket transactions—such as those analysed in our study—provide a complementary perspective to out-of-home purchase data by offering direct evidence of foods residents buy for home consumption. The Tesco dataset represents purchases from 1.6 million loyalty card owners across London, offering fine geographic granularity and thorough insights into the food purchases of London's residents [33]. Such large-scale supermarket transaction data provides an objective measure of food purchasing behaviour, limiting biases associated with traditional survey methods which rely on self-reporting [34]. Since Tesco has the leading market share of supermarkets across the UK (27.6%) [35], it is the most extensive purchase data publicly available.

The spatial analysis of purchasing patterns is a valuable tool in linking sociodemographic and environmental factors to food consumption at a local level. A national survey from the Netherlands was used to extract dietary patterns on the basis of food categories, which found that dietary patterns were strongly spatially clustered [36]. This method, which links dietary data to sociodemographics such as age, income, and education, identified key factors influencing food behaviour, aligning with the broader recommendation to use dietary patterns rather than the consumption of individual foods or nutrients for population dietary recommendations [37]. Additionally, while the use of ordinary least squares regression provides valuable insights [26,36], its assumption of spatial homogeneity limits the ability to account for spatially clustered food purchase behaviour. The Tesco food purchase dataset has been shown to exhibit spatial heterogeneity consistent with patterns in educational attainment [38], however, the contribution of other sociodemographic factors was not considered.

In this work, we leverage consumer behaviour data to classify areas as food deserts according to whether purchases are nutritionally deficient at a statistically significant geographical level. We use the Tesco food purchase data [33] to uncover shopping behaviours according to food category purchases, and identify where London's residents are consuming nutritionally deficient diets. We first employ an unsupervised statistical method to identify dominant purchasing patterns, which separates purchases of high-sugar, high-carbohydrate, and processed foods from those rich in fibre and protein (see Sect 2.2). Our findings, detailed in Sect 2.4, show that nutritionally deficient food purchasing patterns are spatially clustered, enabling us to locate 'food deserts'. We approach the issue from the perspective of consumer behaviour and classify areas as food deserts based on whether purchases are nutritionally deficient at a statistically significant geographical level.

A critical element of this research is accounting for the spatial variation in food purchasing patterns and their non-stationary relationship with sociodemographic factors. Understanding how these factors vary across different neighbourhoods is vital for identifying the specific barriers to food access that residents face. We therefore apply a spatially dependent statistical regression technique—geographically weighted regression (GWR)—a well-established technique for modelling spatial non-stationarity in relationships between variables [39]. Our findings demonstrate that the relationship between food purchasing behaviour and drivers of food access vary across space, uncovering the nuanced connections between these factors and food purchases across neighbourhoods. For each neighbourhood, we identify the key factors associated with food deserts, providing clear insights into the specific barriers to food access across different areas (see Sect 3.4).

2. Purchasing pattern analysis

2.1. Data overview: Tesco Grocery 1.0

The Tesco Grocery 1.0 dataset [33] is a record of 420 million food items purchased by 1.6 million Tesco Clubcard owners who shopped at 411 Tesco stores across Greater London in 2015. Transactions were recorded and linked to the residential address associated with the Tesco Clubcard account, and have been aggregated at geographical administrative areas to preserve customer anonymity. The group of variables we consider are the food product classes, whereby each food and drink item purchased was grouped into a distinct food or drink category.

The geographical units used throughout our analysis are Lower Layer Super Output Areas (LSOAs): units of approximately equal population size defined by the Office for National

Statistics [40]. Using the 2011 boundaries, there are 4,833 LSOAs nested within Greater London's 33 local authority districts, each comprising between 1,000 and 3,000 residents. LSOAs are the smallest official geographical census units in England and provide a more detailed understanding of spatial purchasing patterns at a local level, compared to a larger geographical scale such as the local authority level.

We use an Open Access 'LSOA Atlas' dataset from the London Datastore which provides a summary of demographic and related data for each LSOA in London using the 2011 boundaries [50]. For each LSOA, we have the fraction of all items scanned at the register that fall into each food product category, based on item counts. A summary of the 12 food categories considered in our analysis are shown in Table 1. Alcohol and non-calorific beverages were removed from the dataset prior to analysis, as they are not directly relevant to the study of food deserts or nutritional food purchasing patterns focussed on food quality and availability.

We restrict our analysis to a contiguous area comprising 23 London local authorities, selected to support spatial analysis (see Sect 3.3) and shown in Fig 1. The selection was based on a balance of three key factors: store coverage, transaction volume, and Tesco customer representativeness. Representativeness is defined as the number of unique Tesco Clubcard customers in an LSOA, expressed as a proportion of the LSOA's total population. Following recommendations from the original dataset authors [33], areas where this proportion fell below approximately 0.1 were generally avoided. Transaction volume was also considered to reduce the inclusion of areas with sparse purchasing data. To support spatial analysis, we limited the study area to a set of geographically contiguous, complete local authorities. Due to the unavailability of store location data, spatial contiguity was approximated using a reference map from the original dataset publication. Maps showing variation in customer representativeness and transaction volume across LSOAs are provided in the Supporting Information (S1 Appendix). This approach helps to minimise sampling bias by excluding areas where Tesco customers are not representative of the broader population, ensuring spatial bias does not distort the analysis of food category variables. LSOAs provide the finest granularity that allows for analysis of anonymised data, and our restricted study area represents a population of approximately 5–6 million London residents.

Table 1. Food and drink categories extracted from the Tesco Grocery 1.0 dataset [41].

Food categories	Description
Grains	Bread, rice, pasta
Sweets	Candies, chocolate
Soft drinks	Carbonated sodas
Fruit & veg	Fruits and vegetables
Fish	
Red meat	Beef, pork
Poultry	
Sauces	Tomato sauce, soups
Fats & oils	Butter, olive oil
Eggs	
Dairy	Milk, cheese
Readymade meals	Pre-cooked meals

https://doi.org/10.1371/journal.pcsy.0000072.t001

Fig 1. Map of our study area with LSOA boundaries included. The area of focus is highlighted in green. Shapefiles for LSOA boundaries were obtained from the Greater London Authority via the London Datastore [63]: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london. Contains National Statistics data Crown copyright and database right 2015. Contains Ordnance Survey data Crown copyright and database right 2015. Licensed under the UK Open Government Licence v2.0 (https://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/).

https://doi.org/10.1371/journal.pcsy.0000072.g001

2.2. Identifying purchasing behaviour

We employ Principal Component Analysis (PCA), an unsupervised statistical method that finds linear combinations of the data that successively capture the maximum variance [42]. PCA is widely used for identifying dietary patterns in various countries such as Australia [43], Iran [44], and the Netherlands [36]. PCA has also been applied in health studies to explore dietary patterns in adults with high blood pressure [45].

We define X as our standardised $n \times p$ matrix of food category purchases, where $n = 3361$ represents the number of LSOAs and $p = 12$ is the number of food categories. We perform PCA on the standardised matrix, X, to identify dominant patterns in dietary purchasing behaviour across London's LSOAs. Standardisation ensures each food category contributes equally to the analysis, preventing those with low absolute values but high relative variance from disproportionately influencing results [46]. This approach corresponds with correlation-based PCA. We extracted the first principal component (PC1), which captures the largest proportion of variance in the food purchasing data. The corresponding component scores reflect how closely each LSOA's purchase profile aligns with this dominant dietary pattern. Factor loadings indicate the contribution of each food category to this pattern, and are proportional to the correlation between food categories and the PC1 scores. These scores form the basis of subsequent spatial modelling to identify areas exhibiting nutrient-deficient purchasing behaviours.

The food category variables are compositional. The extraction of *multiple* principal components from compositional data may be inappropriate, because each vector of proportions is restricted to a space where the elements are non-negative and sum to unity, so the assumption that components are orthogonal to one another is questionable [47]. To avoid potential issues in the inference of multiple components, we use only the first component.

First, we examine the factor loadings to understand the relationship between food categories and the dominant purchasing pattern. Factor loadings are shown in Fig 2, and indicate the strength of the correlation between food categories and the dominant purchasing pattern. Purchases of sweets, grains, soft drinks and readymade meals are positively correlated with $PC1$, whereas purchases of fruits, vegetables and fish are negatively correlated with $PC1$. The first eigenvector, A_1, has an eigenvalue of $\lambda_1 = 0.304$, meaning that the purchasing pattern explains 30.4% of the total variance in food purchases. There is a distinguishable separation in purchases of high-sugar, high-carbohydrate, and processed food categories to fresh, high-fibre, and high-protein options. Therefore, $PC1$ scores represent the adherence of purchases to the high-sugar and high-carbohydrate pattern; a higher positive score indicating higher adherence. We focus our analysis on $PC1$ as it explains a substantial proportion of variance in food purchasing behaviour and offers a clear nutritional interpretation, capturing the contrast between nutrient-poor and nutrient-dense food categories. The variance explained by subsequent components is considerably lower and is shown in S2 Appendix.

2.3. Spatial analysis of purchasing behaviour

Building on our PCA results, we now examine the spatial relationships between LSOAs to explore geographic variations in purchasing behaviour. To quantify the relationship between purchases and their location, we use spatial autocorrelation. Spatial autocorrelation considers the degree to which the similarity in values between observations in a dataset are related to the similarity in locations of such observations. We define the relationship between neighbouring LSOAs in terms of spatial weights, according to the *queen contiguity criterion*,

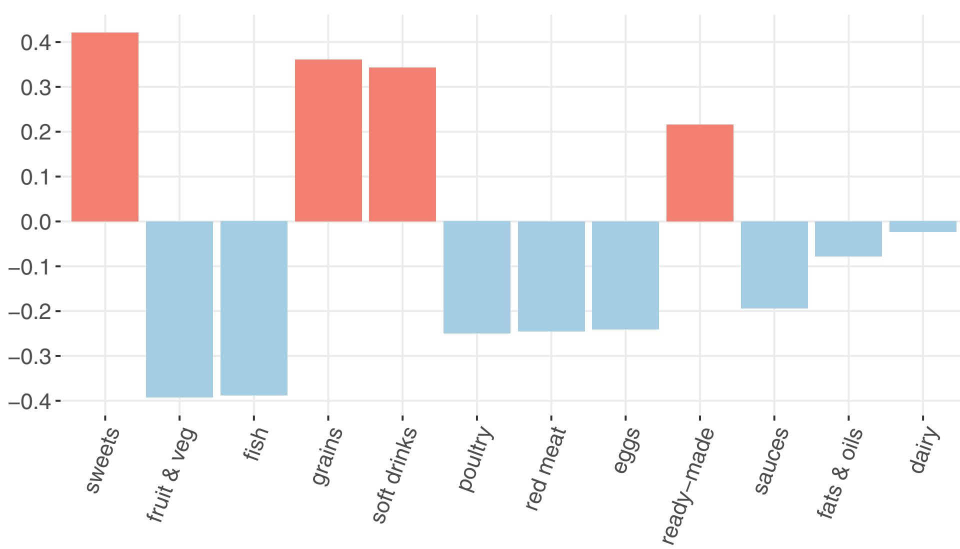

Fig 2. Principal component 1 loadings as defined in Sect 2.2 against their respective food categories. Sweets, fruit and vegetables, fish, grains, and soft drinks have the highest positive loadings. Purchases of sweets, grains and soft drinks are inversely correlated with purchases of fruit and vegetables and fish.

https://doi.org/10.1371/journal.pcsy.0000072.g002

whereby neighbours are defined as areas that share a common boundary point [48]. We use a frequently used measure of spatial autocorrelation, Global Moran's I (GMI) [49], a summary statistic that quantifies the extent of spatial clustering for a particular characteristic. GMI is defined as

$$GMI = \frac{n}{S_0}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{ij}(Y_{1i}-\bar{Y}_1)(Y_{1j}-\bar{Y}_1)}{\sum_{i=1}^{n}(Y_{1i}-\bar{Y}_1)^2},$$

where $n = 3361$ is the number of LSOAs, Y_1 is the vector of $PC1$ scores, Y_{1i} is $PC1$ score at LSOA i, ω_{ij} are the elements of a mostly sparse matrix of spatial weights ($\omega_{ii} = 0$), and $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{ij}$ is the aggregate of all the spatial weights. Row weights have been standardised so that $\sum_{j=1}^{n}\omega_{ij} = 1$, to avoid bias towards areas with a larger number of neighbours (denser areas), so $n/S_0 = 1$. GMI takes a value between −1 and 1, where a high positive GMI indicates clustering of similar values, meaning neighbouring areas share similar values, hence $GMI = 1$ indicates perfect spatial clustering. $GMI = -1$ indicates perfect dispersion in the data, meaning that the neighbouring areas have dissimilar values. A GMI close to 0 means that the data is distributed randomly and exhibits no spatial clustering.

The GMI can be decomposed into its local contributions to show the individual contribution of LSOA i to the overall global spatial autocorrelation, $GMI = \frac{1}{n}\sum_{i=1}^{n} I_i$, where I_i represents Local Moran's I,

$$I_i = n\frac{Y_{1i}-\bar{Y}_1}{\sum_{i=1}^{n}(Y_{1i}-\bar{Y}_1)^2}\sum_{j=1}^{n}\omega_{ij}(Y_{1j}-\bar{Y}_1).$$

Having outlined the method for quantifying spatial autocorrelation and examining spatial relationships, we now move on to explore the dominant purchasing pattern and its spatial distribution.

2.4. Identifying food deserts

$PC1$ scores are mapped to their corresponding LSOAs in Fig 3, which shows that high-sugar and high-carbohydrate food purchases are located across London's east and west boroughs, with highest concentration in Newham, Redbridge, and Barking and Dagenham. LSOAs with stronger adherence to the high fibre and high protein purchasing pattern are seen in London's inner north-west.

We consider the $PC1$ score a useful metric for measuring how closely each area aligns with the characteristics of a food desert. We will refer to such food items as nutritionally deficient foods, and areas exhibiting this behaviour as those with a nutrient deficient purchasing pattern. This quantity provides a reasonable distinction between more and less nutritious food purchases, whilst capturing more than 30% of the variance in the food purchase data.

The $PC1$ score exhibits strong spatial clustering with $GMI = 0.70$ and p-value < 0.001, indicating that residents exhibit similar purchasing patterns to their neighbouring LSOAs. This measure provides us with a global estimate for autocorrelation, but not the location of the clusters. In Fig 4, we visualise the Local Indicators of Spatial Autocorrelation (LISA)[49]; locations whose $PC1$ score significantly contributes to the overall GMI. Red highlighted areas have positive values of $PC1$ and have neighbours that also have positive values (i.e. *Hot Spots*). Conversely, blue highlighted areas have negative values of $PC1$ and have neighbouring LSOAs that also have negative values (i.e. *Cold Spots*). These are the areas that contribute significantly to the overall positive global spatial autocorrelation of $GMI = 0.70$.

We consider Hot Spots as indicators for food deserts. That is, food deserts are classified as areas where purchases are nutritionally deficient, and whose neighbouring LSOAs exhibit a

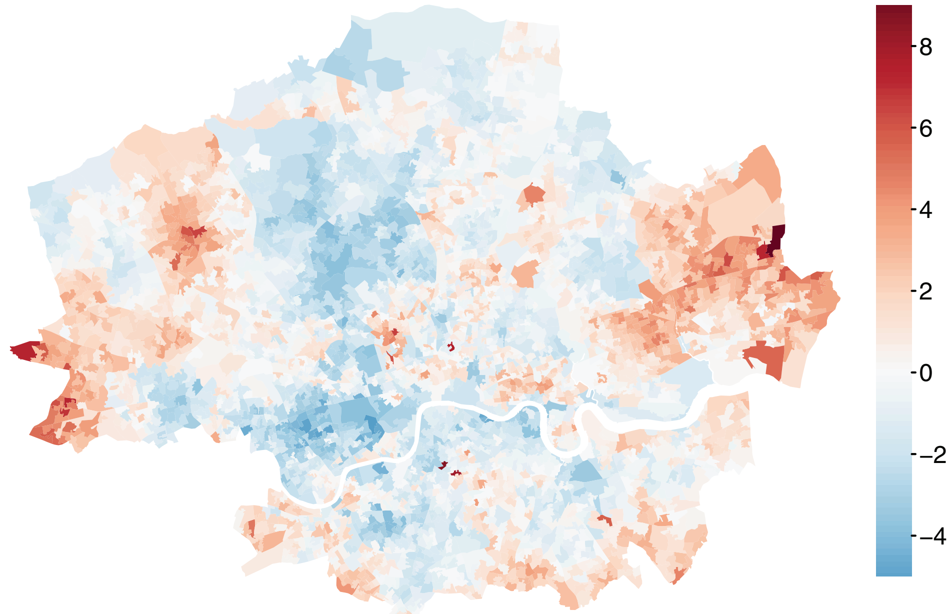

Fig 3. Principal component scores across our study area. Adherence to purchases high in sugar and carbohydrates are most prevalent in London's east, west and north-west (red). Adherence to high-fibre and high-protein purchases are predominantly in London's inner-west (blue). Shapefiles for LSOA boundaries were obtained from the Greater London Authority via the London Datastore: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london. Contains National Statistics data Crown copyright and database right 2015. Contains Ordnance Survey data Crown copyright and database right 2015. Licensed under the UK Open Government Licence v2.0 (https://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/).

https://doi.org/10.1371/journal.pcsy.0000072.g003

similar purchasing behaviour. On the other hand, Cold Spots are those where residents are purchasing more nutritious options. We could attribute such areas to *food oases*, areas where nutritious foods are plentiful and affordable. The food oases are located in central and western London. The largest food desert areas are located in the east and west, whilst a few smaller food deserts are located in inner-city and south of the Thames.

We hypothesise that the spatial variation in purchasing behaviour, as measured by $PC1$ score, can be substantially explained by other covariates such as sociodemographic characteristics and barriers to food access. This hypothesis is the topic of Sect 3.

3. Potential drivers of London's food deserts

3.1. Data overview

Linking food purchases to sociodemographic data provides greater insight into how residents' interactions with food environments may be shaping their purchasing habits, and which barrier of access may be impeding their access to healthy foods. We use an Open Access 'LSOA Atlas' dataset from the London Datastore which provides a summary of demographic and related data for each LSOA in London using the 2011 boundaries [50], and journey time estimates from the Department for Transport (DfT) [51] to investigate the relationships between purchasing patterns and sociodemographic factors at LSOA level.

We consider census data that has been summarised at LSOA level, including average age, Black, Asian, and minority ethnic (BAME) population, household median income, journey

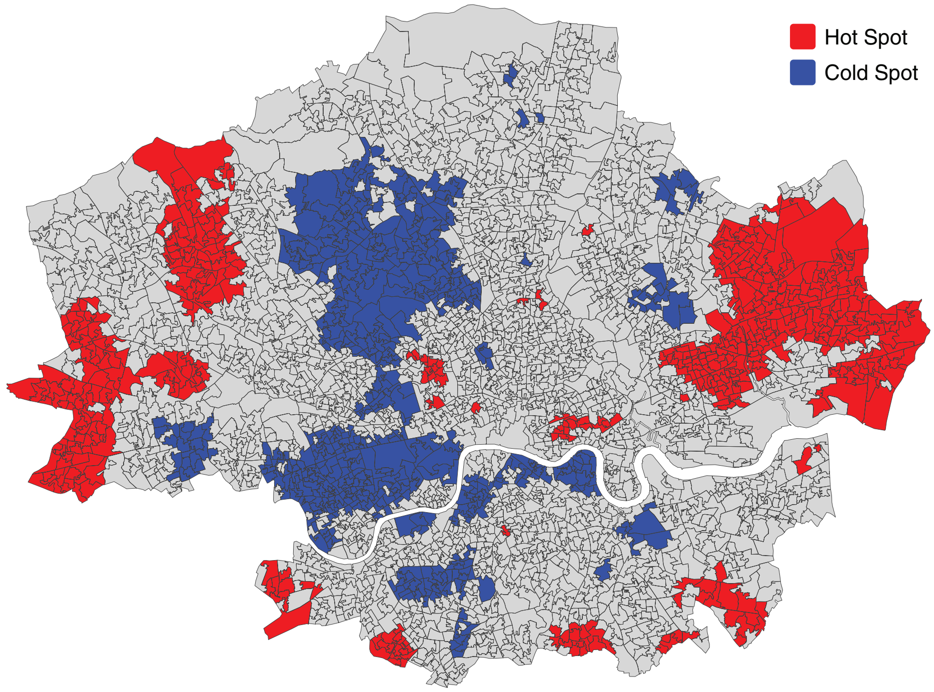

Fig 4. Local Indicators of Spatial Association (LISA) clusters of PC1 score at the 95% significance level. Hot Spots (red) represent areas where purchases are nutritionally deficient, and are our indicator of 'food deserts'. Cold Spots (blue) represent areas that have purchasing habits reflecting that of a 'food oasis'. Shapefiles for LSOA boundaries were obtained from the Greater London Authority via the London Datastore: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london. Contains National Statistics data Crown copyright and database right 2015. Contains Ordnance Survey data Crown copyright and database right 2015. Licensed under the UK Open Government Licence v2.0 (https://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/).

https://doi.org/10.1371/journal.pcsy.0000072.g004

time to the nearest store (by walking), and car ownership. These variables are important for linking food purchases to domains of inaccessibility, such as socioeconomic factors, food affordability, and transportation challenges, which have been highlighted in existing literature on the multidimensional nature of food deserts. Public transport accessibility and education level metrics have been excluded from the analysis due to collinearity with walk time and income estimates, respectively. An overview of these variables and their spatial autocorrelation is provided in Table 2, which shows that all the variables are spatially clustered.

To prepare the $PC1$ scores for modelling and to reduce skewness, we apply a Yeo-Johnson transformation, an extension of the Box-Cox power transformation to accommodate for negative values. The transformation parameter was optimised for best fit. Further details are provided in S2 Appendix.

3.2. Linear regression modelling

We identify factors associated with nutritionally deficient food purchases at LSOA level using an ordinary least squares (OLS) linear regression model. The model predicts the transformed first principal component score ($PC1$) using centred and scaled sociodemographic variables as predictors. This model serves as a baseline for comparison with the geographically weighted regression (GWR), which accounts for spatial variation in the relationships.

Table 2. Summary statistics for our response variable, *PC*1, and LSOA-level sociodemographic data across our study area. BAME refers to Black, Asian, and minority ethnic population (%). Data is collected alongside the Tesco Grocery 1.0 data, London Datastore (summarised data from the 2011 Census at LSOA level), and the Department for Transport (DfT).

Characteristic	Source	Mean	GMI
Food purchasing behaviour (*PC*1)		0	0.70
Average age (years)	Tesco 1.0	35.4	0.59
Household median income (£)	London Datastore	35558	0.61
BAME population (%)	London Datastore	44.3	0.80
Households with cars (%)	London Datastore	53.6	0.78
Walk time (minutes)	DfT	5.95	0.49

https://doi.org/10.1371/journal.pcsy.0000072.t002

The OLS regression model estimates that lower median income and higher Black, Asian, and minority ethnic populations are associated with greater adherence to the nutritionally deficient food purchasing pattern (see Table 3). The relationship between car ownership and $PC1$ is positive, and shows that an increase in car ownership corresponds with an increased adherence to nutritionally deficient purchases. The model explains approximately 38% of the variance in our response, with a residual standard error of 1.493 on 3355 degrees of freedom. Although the model explains some of the variance in the response variable, the sociodemographic variables are not able to explain its spatial autocorrelation. The model residuals still exhibit strong spatial clustering, with $GMI = 0.60$ (reduced from $GMI = 0.70$). We therefore extend our analysis to account for localised relationships between the covariates and response.

3.3. Modelling with Geographically Weighted Regression (GWR)

Geographically Weighted Regression (GWR) is a useful tool for exploring spatial heterogeneity [39]. GWR offers a more nuanced analysis with which we can incorporate spatial context into the analysis of purchasing patterns and sociodemographic characteristics. In this section we will extend our analysis to investigate the non-stationary spatial relationships between London's purchasing patterns and sociodemographic characteristics, with use of the `GWModel` package [52].

GWR, an application of kernel regression, builds upon the basic linear regression model by estimating the model coefficients in terms of spatial locations. The GWR regression equation is

$$y_i = \beta_{i0} + \sum_k \beta_{ik} x_{ik} + \epsilon_i, \quad (1)$$

where β_{ik} is the value of the kth parameter at LSOA centroid i for $i = 1, \cdots, n$. This method allows for local differences in how variables change, so that the model's coefficients are

Table 3. OLS regression estimates and associated *p*-values.

Variable	Estimate	*p*-value
Intercept	-0.14580	1.64×10^{-8}
Average age (years)	-0.21209	4.02×10^{-11}
Household median income (£)	-0.72683	$< 2 \times 10^{-16}$
BAME population (%)	0.50379	$< 2 \times 10^{-16}$
Households with cars (%)	0.61992	$< 2 \times 10^{-16}$
Walk time (minutes)	-0.15759	3.34×10^{-7}

https://doi.org/10.1371/journal.pcsy.0000072.t003

specific to LSOA i, rather than being equal across all areas. Rather, one regression is fit to each LSOA, producing locally specific parameter results. GWR results generate a complete map of the spatial variation of the parameter estimates.

GWR fits a model around each regression point (LSOA centroid), i, where each set of regression coefficients is estimated using weighted least squares (WLS). The WLS approach is an extension to OLS whereby a weighting factor is applied to each squared difference before minimising, so that the inaccuracy of some predictions carries more of a penalty than others. The estimated coefficients satisfy

$$\hat{\beta}_i = (Z^T W_i Z)^{-1} Z^T W_i y,$$

where Z is the matrix of standardised sociodemographic variables as defined in Sect 3.2 with a column of ones for the intercept, y is our response variable vector of transformed $PC1$ scores, $\hat{\beta}_i = (\beta_{i0}, \beta_{i1}, \ldots, \beta_{im})^T$ is the vector of $m + 1$ local regression coefficients, and W_i is the $n \times n$ weight matrix for LSOA i, which scales the rows of Z and y according to their spatial relevance to LSOA i. Each diagonal element of W_i, denoted w_{ij}, represents the geographical weighting assigned to LSOA j when estimating the model at LSOA i, with off-diagonal elements set to zero.

These weightings, w_{ij}, are defined according to a kernel density function, such that observations nearer to LSOA i have more influence in determining the regression coefficients than observations further away. The weighting of LSOA j on LSOA i is given by the bi-square kernel function,

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b}\right)^2\right]^2 & \text{if } d_{ij} < b, \quad \text{and} \\ 0 & \text{otherwise,} \end{cases}$$

where d_{ij} is the Euclidean distance between the centroids of LSOA i and j (British National Grid coordinates in metres), and b denotes the bandwidth size (cut-off point, beyond which LSOAs have zero weighting on the parameter estimate). This weighting function is continuous and near-Gaussian up to distance b, and assigns a zero weight for LSOAs beyond b.

The size of our neighbourhoods for which each local regression model is fit, b, can either be fixed (distance) or adaptive (discrete number of neighbours). We opt for an adaptive bandwidth which quantifies bandwidth according to number of discrete neighbours (LSOAs), since LSOAs are all of different sizes depending on the population density. The model is therefore fit to a smaller geographical area in densely populated areas (such as inner London), and large geographical areas for suburban areas.

One aspect of GWR is that the estimated parameters are, in part, dependent on the weighting function or kernel selected. The problem is therefore how to select an appropriate bandwidth or decay function in GWR. There are a number of criteria that can be used for bandwidth selection. Methods to determine the optimal bandwidth (and the sensitivity of bandwidth on model results) are studied extensively in the literature [53]. A large bandwidth means more data points are included in the model, which reduces variance. On the other hand, a small bandwidth confines the regression to a local area and parameter estimates will be dependent on observations closer to the regression point.

We determine an optimal bandwidth, b, with an automatic optimisation algorithm which selects the bandwidth that produces the best model. AICc and cross-validation (CV) are the most frequently used methods in the literature. We consider the two kernel shapes, bi-square and Gaussian, along with tuning methods, CV and AICc, to determine a bandwidth which optimises model performance. Table 4 shows the bandwidths and their resulting GWR model fit (according to AIC, AICc and R squared). Due to the risk of overfitting, which can occur when GWR is fit to a small number of neighbourhoods, we choose the bandwidth that minimises AICc. As shown in Table 4, the bandwidth that minimises AICc, $b = 72$, is larger than the bandwidth which maximises R^2; but the value we choose is almost as accurate ($R^2 = 0.711$), and significantly reduces the risk of overfitting. We visualise the bandwidth, $b = 72$, in Fig 5, and in the Supporting Information (S3 Appendix). This size of bandwidth allows us to capture variation within our food desert areas shown in Fig 4.

As shown in Eq (1), GWR typically fits an intercept term that varies spatially across the study area, meaning that each LSOA is assigned a unique baseline value for the predicted response ($PC1$). This behaviour is undesirable when our aim is to enhance the interpretability of the model and focus on the key predictors. Accordingly, we re-centre the response prior

Table 4. Performance of the different spatial weighting functions and bandwidth selection schemes. We opt to use a bi-square kernel with bandwidth of 72, which minimises the AICc of the resulting model.

Optimisation method	Bandwidth	AIC	AICc	R^2
Bi-square, CV	55	9091.409	10196.27	0.726
Bi-square, AICc	72	9375.27	10155.24	0.711

https://doi.org/10.1371/journal.pcsy.0000072.t004

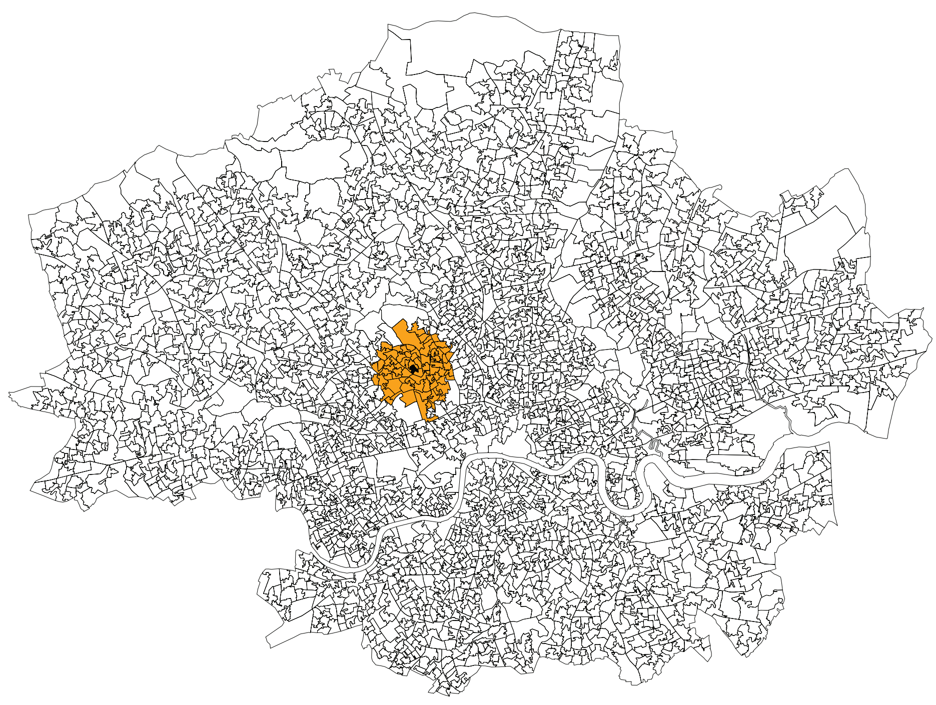

Fig 5. Map of neighbourhood (bandwidth) size, $b = 72$ nearest LSOAs. The black LSOA is the target LSOA, and the red LSOAs are considered its neighbours. The chosen bandwidth appears approximately valid for resolving geographical features of the scale of the food deserts visible in Figs 3 and 4. Shapefiles for LSOA boundaries were obtained from the Greater London Authority via the London Datastore: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london. Contains National Statistics data Crown copyright and database right 2015. Contains Ordnance Survey data Crown copyright and database right 2015. Licensed under the UK Open Government Licence v2.0 (https://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/).

https://doi.org/10.1371/journal.pcsy.0000072.g005

to fitting GWR by subtracting the OLS intercept term (see Table 3), and focus our subsequent analysis on the effects of covariates relative the mean behaviour of the response variable (at zero).

3.4. Main results

GWR offers an improvement to the OLS model. GMI has decreased from 0.70 to 0.14, so the GWR model is successful in capturing a large portion of the spatial variation in food purchases. A plot of local R^2 is shown in Fig 6, where darker blue coloured areas represent those where the model has a better fit. We overlay food desert areas with a white border, and see that the model fits well in all food desert areas (R^2>0.5), with the exception of a small handful of LSOAs in central London. Adjusted global R^2 has also increased from 0.38 to 0.71.

The most common approach to presenting the results of GWR is to generate maps of the parameter estimates, β_{ik}. Such maps allow the visualisation of the estimated model coefficients, from which we can deduce the direction and magnitude of the relationship between variable k and the predicted response variable for each area, i. We choose to visualise a quantity that is similar to the predicted coefficients, but provides some additional insight into how much the predictor influences the outcome for each area. The contribution of each predictor to the predicted $PC1$ score $\hat{y}_i$, is calculated by

$$C_{ik} = Z_{ik}\beta_{ik},$$

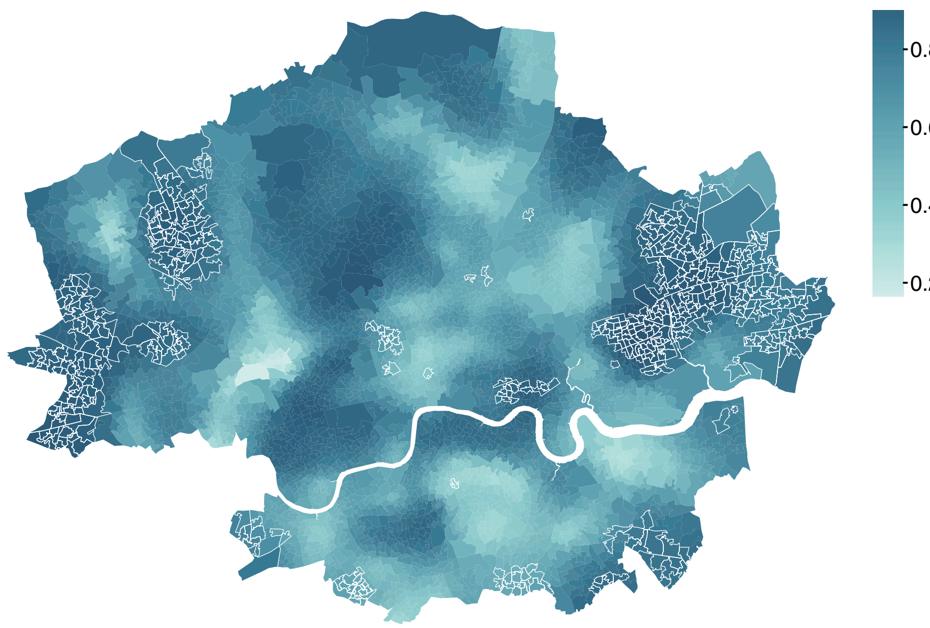

Fig 6. Map of local R^2 for our GWR model. Darker areas correspond with a better model fit, and LSOAs classified as food deserts according to LISA clusters are outlined in white. Shapefiles for LSOA boundaries were obtained from the Greater London Authority via the London Datastore: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london. Contains National Statistics data Crown copyright and database right 2015. Contains Ordnance Survey data Crown copyright and database right 2015. Licensed under the UK Open Government Licence v2.0 (https://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/).

https://doi.org/10.1371/journal.pcsy.0000072.g006

where β_{ik} is the GWR coefficient and Z_{ik} is the standardised census data for variable k at LSOA i. We will refer to this quantity as the *contribution*, where the sum of the contribution over all predictors is the predicted $PC1$ score, $\hat{y}_i = \sum_{k=1}^{p} C_{ik}$.

The contributions measure how much of the food purchasing behaviour is associated with each predictor across the study area. The contribution with the greatest magnitude is interpreted as being the most important predictor for a given LSOA, where the sign of the contribution determines whether the predictor is contributing to nutrient deficient (positive sign) or nutritious (negative sign) foods. For example, if longer walk time in a region is associated with purchasing more nutrient deficient foods, this would show up as a positive contribution. Hence, the longer the walk time in that region, the more 'nutrient-deficient' the purchasing behaviour becomes.

Fig 7 (left column) shows the local census estimates for each of our predictor variables prior to centring and scaling, and the right column shows the local contribution values, C_{ik}, for each predictor k in LSOA i. Along with the value of C_{ik}, the LSOAs outlined in black denote those where the LISA cluster is classified as a Hot Spot, and, the β_{ik} coefficient is significantly different from zero at the 95% confidence level, rejecting the null hypothesis $H_0 : \beta_{ik} = 0$.

Household median income is associated with nutrient deficient purchases in the east and west, whereas income contributes to more nutritious purchases in the inner-west. Higher car ownership rate contributes to more nutritious purchases in a small region of London's north-west, but contributes to nutrient deficient purchases in the east, west and north-west. Black, Asian, and minority ethnic (BAME) population is associated with nutrient deficient purchases and has a high contribution in the west, north east and inner-east food deserts. The contribution of age is weak in comparison to the other terms, but has a slightly positive contribution in a few LSOAs. Walk time has negligible contribution to the predicted response across space, and the figure has been excluded.

We quantify the most influential predictor (MIP) using the contribution term, C_{ik}, to highlight the important features or demographics that are associated with food deserts across London. This metric offers a simple way to identify which sociodemographic factor has the greatest impact on the predicted response for each LSOA, i. The MIP is defined as the factor with the largest positive contribution term to nutrient deficient purchases (positive $PC1$) in each LSOA i, i.e.,

$$\mathrm{MIP}_i = \arg\max_k C_{ik}.$$

The MIP is set to null in the case where i) none of the β_{ik} coefficients are different from zero at the 95% significance level, or ii) where none of the contribution terms C_{ik} are greater than zero. If there are no positive C_{ik} for LSOA i, then we deduce that none of the sociodemographics in our model are contributing to nutrient deficient purchases. Rather, they are contributing to nutritious purchases instead.

Fig 8 shows the MIP for each LSOA, where food desert areas are plotted with full opacity. The maximal contribution of each explanatory variable to the predicted response varies considerably across the study area. Income, Black, Asian, and minority ethnic (BAME) population, and car ownership contribute the most to nutrient deficient purchases in food deserts, whereas age and walk time are the highest contributor for only a few LSOAs. In Ealing, income has the largest contribution to nutrient deficient purchases. In areas such as Newham and Brent, Black, Asian, and minority ethnic population is associated with nutrient deficient purchases. Car ownership is associated with nutrient deficient purchases in Barking and Dagenham. This result is perhaps unintuitive in the context of food deserts, as high car ownership is associated with nutrient deficient purchases, which we discuss in Sect 4. For

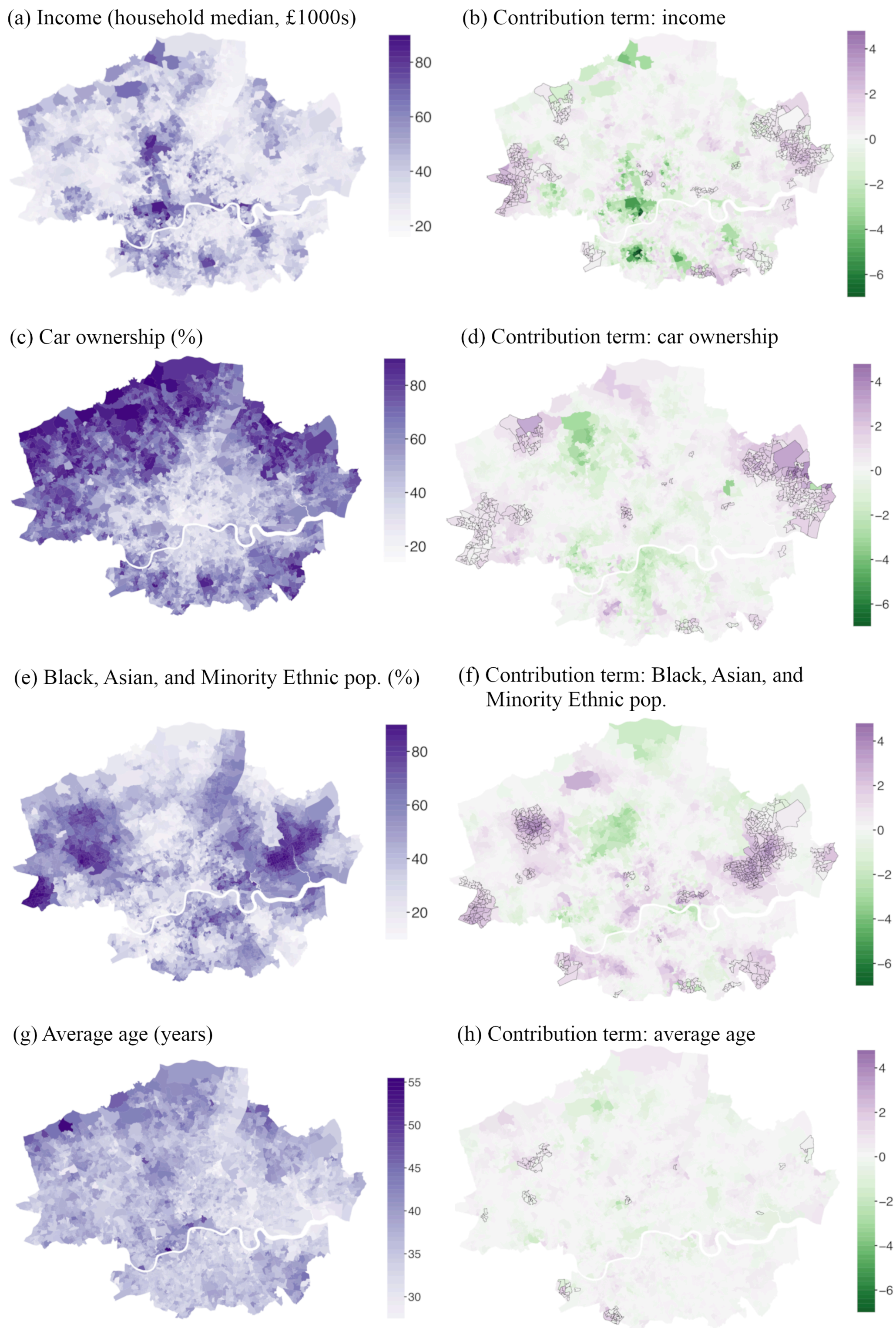

Fig 7. Sociodemographic estimates (left column) and corresponding GWR model contributions C_{ik} (right column) mapped across London for (a–b) household median income, (c–d) car ownership, (e–f) Black, Asian, and Minority Ethnic population, and (g–h) average age. LSOAs outlined in black denote areas where the LISA cluster is a Hot Spot (food purchases are reflective of 'food deserts'), and the β_{ik} coefficient for location i and covariate k is statistically significant. Household income; Black, Asian, and minority ethnic population; and car ownership contribute to nutrient

deficient purchases in food desert areas. Shapefiles for LSOA boundaries were obtained from the Greater London Authority via the London Datastore: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london. Contains National Statistics data Crown copyright and database right 2015. Contains Ordnance Survey data Crown copyright and database right 2015. Licensed under the UK Open Government Licence v2.0 (https://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/).

https://doi.org/10.1371/journal.pcsy.0000072.g007

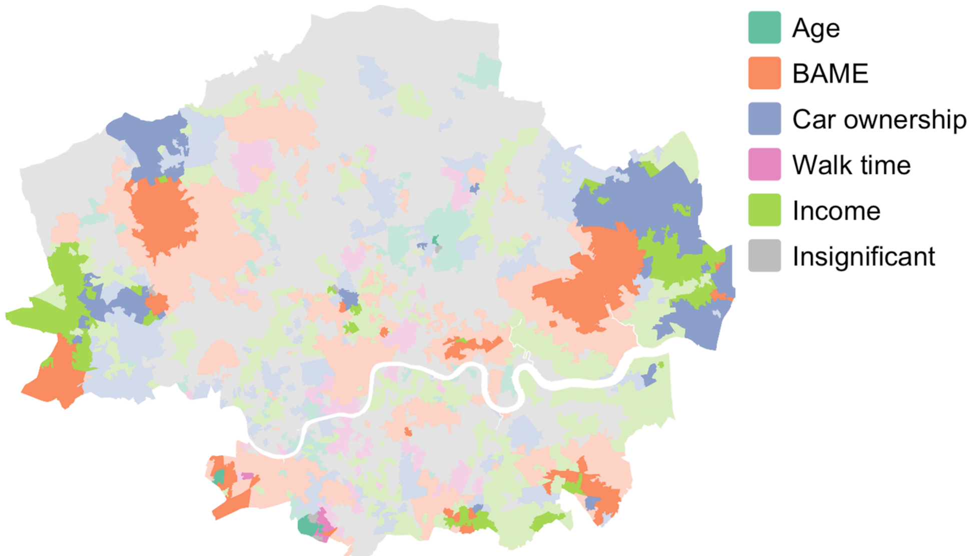

Fig 8. Map of Most Influential Predictor (MIP) in food deserts: variables that have the largest contribution to nutrient deficient purchasing behaviour. Opaque polygons represent food desert areas. 'Insignificant' areas refer to LSOAs where there are only negative contributions (contributions toward healthy behaviour), or, areas where there is not a statistically significant β_{ik} coefficient. BAME refers to Black, Asian, and minority ethnic population (%). Shapefiles for LSOA boundaries were obtained from the Greater London Authority via the London Datastore: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london. Contains National Statistics data Crown copyright and database right 2015. Contains Ordnance Survey data Crown copyright and database right 2015. Licensed under the UK Open Government Licence v2.0 (https://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/).

https://doi.org/10.1371/journal.pcsy.0000072.g008

comparison with MIP in food deserts, the Supporting Information (S4 Appendix) shows the corresponding plot of the most significant predictors of a food oasis.

4. Discussion and conclusions

We developed a statistical model to identify London neighbourhoods where residents' purchases follow a nutrient deficient pattern, highlighting areas that are characteristic of 'food deserts'. The significance of area-specific adherence to high-sugar and high-carbohydrate purchases suggests that residents in these areas are not accessing a nutritionally adequate diet. We attribute such behaviours to 'food deserts', areas where residents face barriers to accessing more nutritious foods such as high-fibre and high-protein options. Such areas are located in London's east (e.g. Newham; Barking and Dagenham), and some areas of London's north west (e.g. Ealing; Brent), and are associated with sociodemographic factors. This method not only highlights areas where residents are purchasing nutritionally deficient foods, but also narrows the scope of the problem by focussing on the areas most affected by food desert conditions and potential barriers to access.

We linked food purchasing patterns to sociodemographic and environmental factors using GWR, revealing spatial variation across London's boroughs. Our study finds that household income and the proportion of Black, Asian, and minority ethnic populations are strongly associated with food deserts in the London area. Our analysis identifies significant associations, but the underlying mechanisms are likely to vary and require further investigation. These spatial differences in associations may reflect a combination of socioeconomic conditions, local food retail environments, or varying food preference. For example, Black, Asian, and minority ethnic populations are disproportionately affected by poverty, being twice as likely to experience poverty than white Londoners [54]. Further work involving local experts is needed to better understand these context-specific relationships and to develop appropriately targeted interventions. Understanding these nuanced local dynamics requires collaboration with local experts and policymakers to develop tailored and culturally sensitive interventions. Our findings highlight the importance of such localised approaches, as city-wide analyses may mask important within-city heterogeneity in food access and dietary behaviours.

In the study, we also identified a statistically significant association between high car ownership and food desert areas in London's east. This association is unexpected. A potential explanation is that urban areas manifest convenient and accessible stores that respond to the needs of carless households, and more accessible public transport reduces the need for personal vehicles. Residents in such areas may choose to give up cars or live in urban areas when they cannot afford one. This dynamic may not hold in rural areas, where store dispersion and limited public transport require residents to travel long distances to access food. Interestingly, walk time to the nearest store did not significantly influence food purchasing patterns in food desert areas, with a minor and localised contribution observed only in London's north-west and south-west. These results suggest that differences in walk times and car ownership rates are less important than sociodemographic factors in the context of urban food deserts.

Our analysis also has limitations. First, we rely exclusively on Tesco Clubcard data, which captures only supermarket purchases and excludes other dietary sources such as fast food visits, food delivery services, restaurants, or purchases from other retailers (e.g., other supermarkets, specialty shops, or local markets). Due to customer anonymity, we were also unable to identify the specific Tesco locations used by shoppers, limiting our ability to assess differences in food access or pricing across store formats (e.g., Tesco Superstore versus Tesco Express) or geographic areas. Second, the limited coverage of temporal data means we cannot adequately observe changes in purchasing behaviour over time. This limits our ability to detect seasonal variation, long-term dietary trends, or responses to events such as food supply shortages. If made available, time-series data could help distinguish short-term behaviours from persistent patterns, allowing for modelling of how food purchasing responds to economic, policy, or environmental shifts. Further, while our analysis operates at the LSOA level which is the finest granularity available for purchase records, it may mask finer-scale disparities in food access. Utilising street-level or postcode-level data would enable more fine-grained spatial modelling, where food access issues may be localised within small urban pockets. Such granularity would be particularly valuable in the most densely populated areas of London, where average indicators at LSOA level may conceal concentrated areas of deprivation or limited access. It is also important to note that potential confounding factors may influence our observed associations. For example, pricing strategies or the availability of culturally preferred foods across different neighbourhoods could affect purchasing behaviour independently of income or ethnicity. These influences may partially explain spatial variation in food purchasing patterns and should be explored in future research. Further research using finer temporal and spatial data should explore these causal pathways to design more effective interventions aimed at improving dietary behaviours and health outcomes: Finally, while we identify associations between food purchasing and sociodemographic factors, causal inferences cannot be drawn from cross-sectional data. Further research using individual-level or longitudinal data is needed to explore causal pathways and to design more targeted and effective interventions aimed at improving dietary behaviours and health outcomes.

A promising approach to determine effective location-specific interventions is the use of predictive simulation models, which could account for the unique characteristics of different food deserts. Policy interventions on populations cannot rely on controlled trials, and therefore predictive modelling and surveillance data (e.g. food purchase data) must feature more heavily in the standards of evidence for policy change [55]. For example, such methods have been used to test various interventions aimed at increasing households' fruit and vegetable intake through initiatives such as financial subsidies [56], increasing shopping frequency [57], expanding public transit options [58], improving health eduction [59], and eliminating residential segregation [60]. Our analysis, by linking food purchasing to sociodemographic and environmental factors, can help inform such models and guide future research into which interventions may be most effective in reducing dietary inequality. Incorporating additional datasets, such as transport networks or commuting patterns, could further enhance this work. Such integration would also allow researchers to examine food choice constraints related to commuting schedules or transport connectivity—particularly in areas where access to affordable and nutritious food is limited by travel time or cost.

Our findings identify key sociodemographic and environmental factors associated with nutritionally deficient food purchasing in London, offering a foundation for future work exploring targeted interventions. To address all barriers to nutritious foods—spatial access, affordability, accommodation, acceptability, and appropriateness—we need a deeper understanding of how the type of food store influences local purchasing patterns. One area of interest is the difference in shopping behaviours between large Tesco stores and their smaller convenience store format, Tesco Express, particularly in terms of price and availability. A comparison of grocery prices in 2022 found that Tesco Express was 10% more expensive, with a yearly price difference of £817.91 [61]. Given that financial barriers were found to be a key driver of food deserts, particularly in areas like Barking and Dagenham and Ealing, our findings provide a basis for identifying areas where income-directed interventions should be prioritised. Additionally, while our current study does not explore the introduction of generally lower-priced grocery stores like Aldi or Lidl, the methodology we have developed can be applied to such datasets if they become available, supporting future research on the potential impact of these stores on residents' nutrition.

This study offers a more holistic method of identifying food deserts, moving beyond the reliance on sociodemographic and environment characteristics alone. Our approach, which relates purchasing behaviour to sociodemographic characteristics in a way that allows for spatial heterogeneity, is novel in the context of food deserts. While earlier studies typically infer dietary risk from spatial access or socioeconomic indicators, our approach begins with observed purchasing behaviour. This data-driven method allows us to identify patterns of nutritional vulnerability that traditional methods may overlook, particularly in urban areas where food retail is abundant but dietary health remains poor. For example, we find boroughs such as Barking and Dagenham, Newham, and Ealing to be characterised by nutritionally deficient purchases, despite these areas appearing well-served by food retailers. Our findings suggest that behavioural data can reveal food access challenges that would otherwise be masked in spatial accessibility focussed approaches. Our analysis reveals significant spatial variation in food purchasing behaviours, with strong clustering of nutritionally deficient food purchases. The prevalence of both high-carbohydrate and high-sugar purchases in certain areas, and high-protein, high-fibre diets in others further highlights the inequality in dietary nutrition across London's boroughs. Our findings indicate that the sociodemographic factors influencing food purchasing vary across space, emphasising the need for targeted interventions that address local food access issues according to specific socioeconomic characteristics. As a result, interventions need to be strategically targeted to areas exhibiting nutritionally deficient purchasing behaviours, and with sociodemographic characteristics (indicators of barriers) in mind. A 'one-size-fits-all' approach is therefore not the most effective way of addressing food deserts in London, and intervention techniques should be tailored to the local level to maximise the effectiveness and compliance of such interventions. This study offers a modern way to classify London's food deserts—by starting with the data that reflects residents' behaviour—and subsequently linking nutrient deficient purchases with their potential drivers.

Data accessibility

Tesco purchase data is available from [41] and summarised London census data is available from [50]. All choropleth maps were created using `tmap` [62]. Shapefiles are available from the Greater London Authority https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london and contains National Statistics data Crown copyright and database right 2015. Contains Ordnance Survey data Crown copyright and database right 2015. Licensed under the UK Open Government Licence v2.0. Code used to generate results can be found on GitHub https://github.com/taylabroadbridge/paper-fooddeserts-gwr.

Supporting information

S1 Appendix. Data coverage.
(PDF)
S2 Appendix. Technical information.
(PDF)
S3 Appendix. Bandwidth visualisation.
(PDF)
S4 Appendix. Most Influential Predictor (MIP) for Food Oases.
(PDF)

Author contributions

Conceptualization: Tayla Broadbridge, J. Edward F. Green, John Maclean.

Formal analysis: Tayla Broadbridge.

Investigation: Tayla Broadbridge.

Methodology: Tayla Broadbridge, Simon P. Preston, John Maclean.

Project administration: J. Edward F. Green, Simon P. Preston, Nabil T. Fadai, John Maclean.

Supervision: J. Edward F. Green, Simon P. Preston, Nabil T. Fadai, John Maclean.

Validation: Tayla Broadbridge.

Visualization: Tayla Broadbridge.

Writing – original draft: Tayla Broadbridge.

Writing – review & editing: Tayla Broadbridge, J. Edward F. Green, Simon P. Preston, Nabil T. Fadai, John Maclean.

References

1. World Health Organisation. Healthy Diet. 2020. [cited 2023 Mar 15]. https://www.who.int/news-room/fact-sheets/detail/healthy-diet
2. Food Standards Agency. Chapter 1: The Nation's Plate, Our Diet and Food Choices Today; 2022. [cited in 18 Feb 2025]. https://www.food.gov.uk/our-work/chapter-1-the-nations-plate-our-diet-and-food-choices-today
3. NHS Digital. Health Survey for England, Part 2: Overweight and Obesity. National Health Service; 2022. [cited 18 Feb 2025].
4. Stiebahl S. Obesity Statistics. House of Commons Library; 2025.
5. NHS Digital. Health survey for England, part 1. National Health Service; 2022.
6. The Food Foundation. The broken plate. 2025. https://foodfoundation.org.uk/publication/broken-plate-2025f
7. Corfe S. What are the barriers to eating healthily in the UK. The Social Market Foundation. 2018.
8. Beaulac J, Kristjansson E, Cummins S. A systematic review of food deserts, 1966–2007. Preventing Chronic Disease; 2009.
9. Beaumont et al. Report from the Policy Sub-Group to the Nutrition Task Force Low Income Project Team of the Department of Health. Radlett, Hertfordshire: Institute of Grocery Distribution; 1995.
10. Wrigley N, Warm D, Margetts B. Deprivation, diet, and food-retail access: findings from the leeds 'food deserts' study. Environ Plan A. 2003;35(1):151–88. https://doi.org/10.1068/a35150
11. Wrigley N, Guy C, Lowe M. Urban regeneration, social inclusion and large store development: the Seacroft development in context. Urban Studies. 2002;39(11):2101–14.
12. Clarke G, Eyre H, Guy C. Deriving indicators of access to food retail provision in british cities: studies of cardiff, leeds and bradford. Urban Studies. 2002;39(11):2041–60. https://doi.org/10.1080/0042098022000011353
13. Whelan A, Wrigley N, Warm D, Cannings E. Life in a "Food Desert". Urban Studies. 2002;39(11):2083–100. https://doi.org/10.1080/0042098022000011371
14. Cummins S, Flint E, Matthews SA. New neighborhood grocery store increased awareness of food access but did not alter dietary habits or obesity. Health Aff (Millwood). 2014;33(2):283–91. https://doi.org/10.1377/hlthaff.2013.0512 PMID: 24493772
15. Shannon J. Beyond the supermarket solution: linking food deserts, neighborhood context, and everyday mobility. Annals of the American Association of Geographers. 2015;106(1):186–202. https://doi.org/10.1080/00045608.2015.1095059
16. Allcott H, Diamond R, Dubé J-P, Handbury J, Rahkovsky I, Schnell M. Food deserts and the causes of nutritional inequality*. The Quarterly Journal of Economics. 2019;134(4):1793–844. https://doi.org/10.1093/qje/qjz015
17. Madlala SS, Hill J, Kunneke E, Lopes T, Faber M. Adult food choices in association with the local retail food environment and food access in resource-poor communities: a scoping review. BMC Public Health. 2023;23(1):1083. https://doi.org/10.1186/s12889-023-15996-y PMID: 37280606
18. Caspi CE, Sorensen G, Subramanian SV, Kawachi I. The local food environment and diet: a systematic review. Health Place. 2012;18(5):1172–87. https://doi.org/10.1016/j.healthplace.2012.05.006 PMID: 22717379
19. Usher K. Valuing all knowledges through an expanded definition of access. JAFSCD. 2015. p. 1–6. https://doi.org/10.5304/jafscd.2015.054.018
20. Widener MJ. Spatial access to food: retiring the food desert metaphor. Physiol Behav. 2018;193(Pt B):257–60. https://doi.org/10.1016/j.physbeh.2018.02.032 PMID: 29454842
21. Vera-Toscano E, Shucksmith M, Brown DL, Brown H. The rural–urban poverty gap in England after the 2008 financial crisis: exploring the effects of budgetary cuts and welfare reforms. Regional Studies. 2023;58(6):1264–81. https://doi.org/10.1080/00343404.2023.2235374
22. United States Department of Agriculture. Definition of a Food Desert. 2017. [cited 2022 Dec 2]. https://www.ers.usda.gov/webdocs/DataFiles/80591/archiveddocumentation.pdf?v=41332
23. United States Department of Agriculture. Food Access Research Atlas Documentation [cited 2025 Oct 2]. https://www.ers.usda.gov/data-products/food-access-research-atlas/documentation
24. Consumer Data Research Centre. E-Food Desert Index (E-FDI). [cited 2023 June 4]. https://data.cdrc.ac.uk/dataset/e-food-desert-index

25. Consumer Data Research Centre. Priority Places for Food Index. [cited 2024 Dec 10]. https://priorityplaces.cdrc.ac.uk

26. Jenneson V, Clarke GP, Greenwood DC, Shute B, Tempest B, Rains T, et al. Exploring the geographic variation in fruit and vegetable purchasing behaviour using supermarket transaction data. Nutrients. 2021;14(1):177. https://doi.org/10.3390/nu14010177 PMID: 35011053

27. Nica-Avram G, Harvey J, Smith G, Smith A, Goulding J. Identifying food insecurity in food sharing networks via machine learning. Journal of Business Research. 2021;131:469–84. https://doi.org/10.1016/j.jbusres.2020.09.028

28. Milligan G, Nica-Avram G, Harvey J, Goulding J. Foodinsecurity london: developing a food-insecurity prevalence map for London - a machine learning from food-sharing footprints. IJPDS. 2024;9(4). https://doi.org/10.23889/ijpds.v9i4.2425

29. Aiello LM, Schifanella R, Quercia D, Del Prete L. Large-scale and high-resolution analysis of food purchases and health outcomes. EPJ Data Sci. 2019;8(1). https://doi.org/10.1140/epjds/s13688-019-0191-y

30. Barrett S, Brown R, Belcher E. The London Intelligence: Health and wellbeing. Centre for London. 2019. https://centreforlondon.org/reader/the-london-intelligence-health-and-wellbeing/data

31. Chavis C, Bhuyan IA. Data-driven food desert metric to understand access to grocery stores using chi-square automatic interaction detector decision tree analysis. Transportation Research Record: Journal of the Transportation Research Board. 2022;2676(5):571–89. https://doi.org/10.1177/03611981221089308

32. McEntee J, Agyeman J. Towards the development of a GIS method for identifying rural food deserts: Geographic access in Vermont, USA. Applied Geography. 2010;30(1):165–76. https://doi.org/10.1016/j.apgeog.2009.05.004

33. Aiello LM, Quercia D, Schifanella R, Del Prete L. Tesco Grocery 1.0, a large-scale dataset of grocery purchases in London. Sci Data. 2020;7(1):57. https://doi.org/10.1038/s41597-020-0397-7 PMID: 32071310

34. Jenneson VL, Pontin F, Greenwood DC, Clarke GP, Morris MA. A systematic review of supermarket automated electronic sales data for population dietary surveillance. Nutr Rev. 2022;80(6):1711–22. https://doi.org/10.1093/nutrit/nuab089 PMID: 34757399

35. Tesco PLC. Annual Report and Financial Statements. 2024. https://www.tescoplc.com/investors/reports-results-and-presentations/annual-report-2024

36. Dekker LH, Rijnks RH, Strijker D, Navis GJ. A spatial analysis of dietary patterns in a large representative population in the north of The Netherlands - the Lifelines cohort study. Int J Behav Nutr Phys Act. 2017;14(1):166. https://doi.org/10.1186/s12966-017-0622-8 PMID: 29212502

37. Mozaffarian D. Dietary and policy priorities for cardiovascular disease, diabetes, and obesity: a comprehensive review. Circulation. 2016;133(2):187–225. https://doi.org/10.1161/CIRCULATIONAHA.115.018585 PMID: 26746178

38. Azizi Fard N, De Francisci Morales G, Mejova Y, Schifanella R. On the interplay between educational attainment and nutrition: a spatially-aware perspective. EPJ Data Sci. 2021;10(1). https://doi.org/10.1140/epjds/s13688-021-00273-y

39. Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons; 2002.

40. Office for National Statistics. Census 2021 Geographies. 2021. https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeographies/census2021geographies

41. Aiello LM, Schifanella R, Quercia D, Del Prete L. Tesco Grocery 1.0. figshare. 2020.

42. Jolliffe I. Principal Component Analysis. John Wiley & Sons; 2005.

43. Thorpe MG, Milte CM, Crawford D, McNaughton SA. A comparison of the dietary patterns derived by principal component analysis and cluster analysis in older Australians. Int J Behav Nutr Phys Act. 2016;13:30. https://doi.org/10.1186/s12966-016-0353-2 PMID: 26928406

44. Ebrahimi S, Leech RM, McNaughton SA, Abdollahi M, Houshiarrad A, Livingstone KM. Dietary patterns derived using principal component analysis and associations with sociodemographic characteristics and overweight and obesity: a cross-sectional analysis of Iranian adults. Front Nutr. 2023;10:1091555. https://doi.org/10.3389/fnut.2023.1091555 PMID: 37139453

45. Naja F, Itani L, Hwalla N, Sibai AM, Kharroubi SA. Identification of dietary patterns associated with elevated blood pressure among Lebanese men: a comparison of principal component analysis with reduced rank regression and partial least square methods. PLoS One. 2019;14(8):e0220942. https://doi.org/10.1371/journal.pone.0220942 PMID: 31419246

46. Baxter MJ, Freestone IC. Log-ratio compositional data analysis in archaeometry*. Archaeometry. 2006;48(3):511–31. https://doi.org/10.1111/j.1475-4754.2006.00270.x

47. Aitchison J. Principal component analysis of compositional data. Biometrika. 1983;70(1):57–65. https://doi.org/10.1093/biomet/70.1.57
48. Anselin L. Spatial econometrics. Springer Netherlands; 1988.
49. Anselin L. Local Indicators of Spatial Association—LISA. Geographical Analysis. 1995;27(2):93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x
50. Greater London Authority. LSOA Atlas. 2014. [cited 2022 June 23]. https://data.london.gov.uk/dataset/lsoa-atlas
51. Department for Transport. Journey Times to Key Services by Lower Super Output Area (JTS0507). 2019. https://www.gov.uk/government/statistical-data-sets/journey-time-statistics-data-tables-jts
52. Gollini I, Lu B, Charlton M, Brunsdon C, Harris P. GWmodel: AnRPackage for exploring spatial heterogeneity using geographically weighted models. J Stat Soft. 2015;63(17). https://doi.org/10.18637/jss.v063.i17
53. Fotheringham AS, Yu H, Wolf LJ, Oshan TM, Li Z. On the notion of 'bandwidth' in geographically weighted regression models of spatially varying processes. International Journal of Geographical Information Science. 2022;36(8):1485–502. https://doi.org/10.1080/13658816.2022.2034829
54. Trust for London. Who is in poverty in London? [cited 2024 Dec 10]. https://trustforlondon.org.uk/news/who-is-in-poverty-in-london
55. Mozaffarian D, Angell SY, Lang T, Rivera JA. Role of government policy in nutrition-barriers to and opportunities for healthier eating. BMJ. 2018;361:k2426. https://doi.org/10.1136/bmj.k2426 PMID: 29898890
56. Zhang D, Giabbanelli PJ, Arah OA, Zimmerman FJ. Impact of different policies on unhealthy dietary behaviors in an urban adult population: an agent-based simulation model. Am J Public Health. 2014;104(7):1217–22. https://doi.org/10.2105/AJPH.2014.301934 PMID: 24832414
57. Widener MJ, Metcalf SS, Bar-Yam Y. Agent-based modeling of policies to improve urban food access for low-income populations. Applied Geography. 2013;40:1–10. https://doi.org/10.1016/j.apgeog.2013.01.003
58. Abel KC, Faust KM. Modeling complex human systems: an adaptable framework of urban food deserts. Sustainable Cities and Society. 2020;52:101795. https://doi.org/10.1016/j.scs.2019.101795
59. Blok DJ, de Vlas SJ, Bakker R, van Lenthe FJ. Reducing income inequalities in food consumption: explorations with an agent-based model. Am J Prev Med. 2015;49(4):605–13. https://doi.org/10.1016/j.amepre.2015.03.042 PMID: 26232897
60. Auchincloss AH, Riolo RL, Brown DG, Cook J, Diez Roux AV. An agent-based model of income inequalities in diet in the context of residential segregation. Am J Prev Med. 2011;40(3):303–11. https://doi.org/10.1016/j.amepre.2010.10.033 PMID: 21335261
61. Walsh H. Is it more expensive to shop at Tesco Express and Sainsbury's local convenience stores?. 2023. [cited 2024 Dec 12]. https://www.which.co.uk/news/article/is-it-more-expensive-to-shop-at-tesco-express-and-sainsburys-local-convenience-stores-aOQah2Z8P2tw
62. Tennekes M. Tmap: Thematic Maps in R. Journal of Statistical Software. 2018;84(6).
63. Greater London Authority. Greater London Authority. 2011. https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london
64. García Bulle Bueno B, Horn AL, Bell BM, Bahrami M, Bozkaya B, Pentland A, et al. Effect of mobile food environments on fast food visits. Nat Commun. 2024;15(1):2291. https://doi.org/10.1038/s41467-024-46425-2 PMID: 38480685
65. Keeble M, Adams J, Vanderlee L, Hammond D, Burgoine T. Associations between online food outlet access and online food delivery service use amongst adults in the UK: a cross-sectional analysis of linked data. BMC Public Health. 2021;21(1):1968. https://doi.org/10.1186/s12889-021-11953-9 PMID: 34719382