Modelling correlation between access to green space and demographic and health variables in London.

Titilayo Ayeni, Shenzhangyin Yang, Niamh French, James Edwards

2021-01-21

This report is best read as a gitbook. To read it in this format, open the _book folder that has also been submitted to Moodle and click on the index.html file. From there you can navigate between chapters using the side bar.

1. Introduction

Green space can defined as an open area of natural vegetation set apart for recreational or aesthetic purposes embedded within an urban environment (Taylor and Hochuli (2017)), which for this study is the Greater London Area. There is a large volume of research indicating the social, economic and health benefits of good access to green spaces (Zhou and Parves Rana (2012), De Ridder et al. (2004),Douglas, Lennon, and Scott (2017)). The importance of this is recognised in the United Nations Sustainable Development Goals by Goal 11 'Sustainable Cities and Communities' (Choi et al. (2016)). This study illustrates that spatial variation exists within demographic groups across London as a whole and within sub-divisions of London. This study also proposes that inequalities in access to green spaces exist between these demographic groups, this reflects similar trends in other UK cities (Comber, Brunsdon, and Green (2008)). This has important implications for planning professionals both in London and elsewhere.

Planning professionals can use this and other research to understand why these spatial variations exist and implement policies to improve the equity of access to green space across all demographic groups. This study examines quality of green space access at the ward level, and therefore potential interventions can be focused on small areas, whereby the access to green space could be significantly improved upon. Previously suggested policies include prioritizing areas of poor access for future development of green spaces, and improving availability of safe and appealing walking and cycling routes to green spaces (Dai (2011)). It is important to note that the benefits of green spaces vary not just with their availability but also with their quality and their use (Zhang et al. (2017)). For example, a public football pitch will bring different benefits than a public woodland or community garden.

This study aims to model the statistical relationship between key demographic and health variables and green space access in London. The success of the proposed models will be evaluated across London as a whole and at a more local scale in North, East, South and West London in order to understand if similar patterns are present at different scales and in different regions of the city. The definition of green space used includes parks, green corridors, outdoor sports fields, allotments, and cemeteries. Variables of average income, life expectancy, wellbeing, age and ethnicity were incorporated into regression models to act as predictors of green space in ward. These variables have been proven correlated with access to green space in other study areas. For example, a study from Astell-Burt et al. (2014) examined whether low-income neighbourhoods have the least green space, and found that green space availability was substantially lower in boroughs of Australian cities with a higher percentage of low-income residents. This study proposes that a similar pattern exists in London, as those areas close to green space such as Hampstead and Richmond in London are often highly desirable Hur, Nasar, and Chun (2010) and therefore have inflated house prices (Smith (2010)). Heynen, Perkins, and Roy (2006) found that areas that have had enhanced green space development, such as Stratford in East London, “unintentionally increase adjacent residential property values” and “drive out residents of lower socio-economic status.” A further example of the correlation proposed is present in a study by Mitchell and Popham (2008) which concluded that in areas showing lower than average life expectancy due to income deprivation, the difference from average was smaller in areas with good access to greenspace compared to those who have poor access.

Inequality in access to green space between different ethnicities has been shown in literature. A study of the spatial distribution of quality of green space access between different ethnicities in Leicester, UK by Comber, Brunsdon, and Green (2008) found that the Indian population had 26.3% less access to green space than the white population, with a host of other inequalities also present. This is important for policy-makers to consider when seeking to address health inequalities, and potentially improve the equity of provision of green space between ethnic groups.

2. Data Description

2.1 Green Space Data

There are a number of ways to measure access to green space, each with their own limitations. For this study, the percentage of ward land area used for green space was taken as a proxy for access to green space for residents in that ward. The limitations of this proxy are discussed in detail later in this report. The data for this proxy was sourced from Ministry of Housing (2005). This is part of the Generalised Land Use Database published by the Ministry of Housing, Communities & Local Government (MHCLG) in 2005. The dataset includes two forms of data: the land area of each London ward and the percentage of this area used for each possible land use. A clear pattern in the spatial distribution (Figure 2.1) of this data is the greater percentage of green space per ward in the 'Green-Belt' of suburban London (a statutory ring of green areas surrounding the capital). This is particularly clear in East and West London.

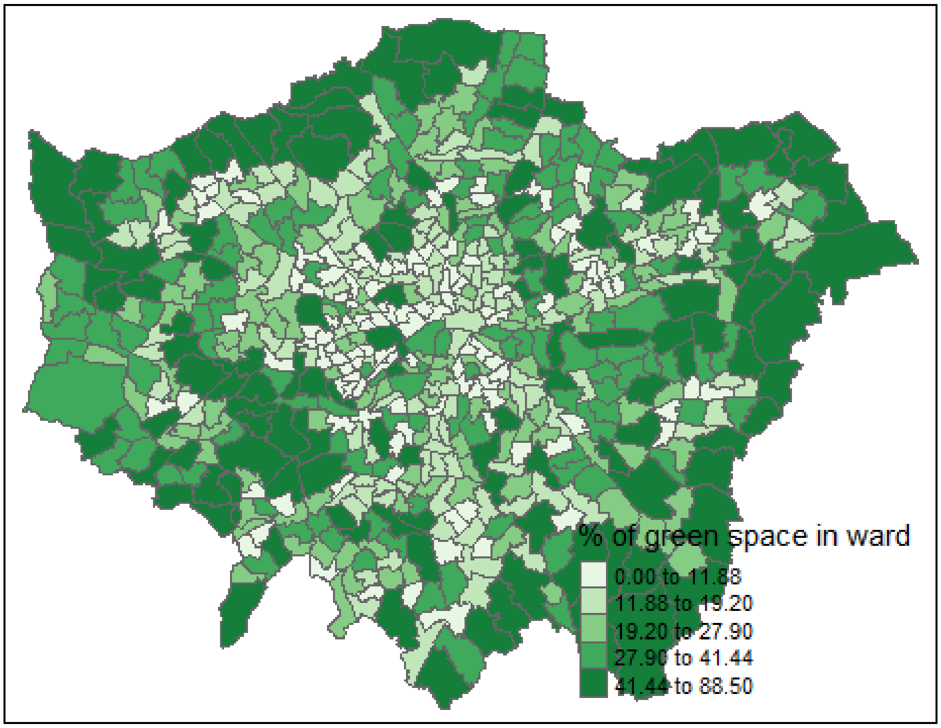

Figure 2.1: % of green space per ward in Greater London

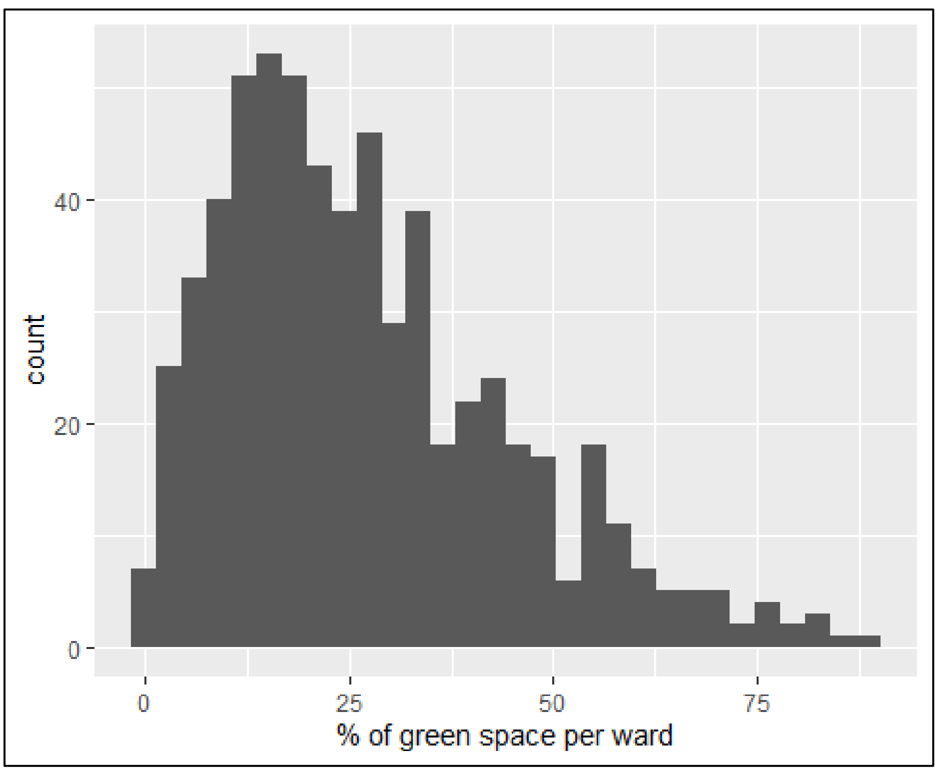

Figure 2.2: Histogram of % of green space per ward in Greater London

Figure 2.2 shows that the data is not normally distributed, which could reduce the integrity of our models. The data can be normalised by taking the natural logarithm of the percentage of the area used after finding the sum of twelve and the variable (ln(green space + 12)). The result is shown below in Figure 2.3.

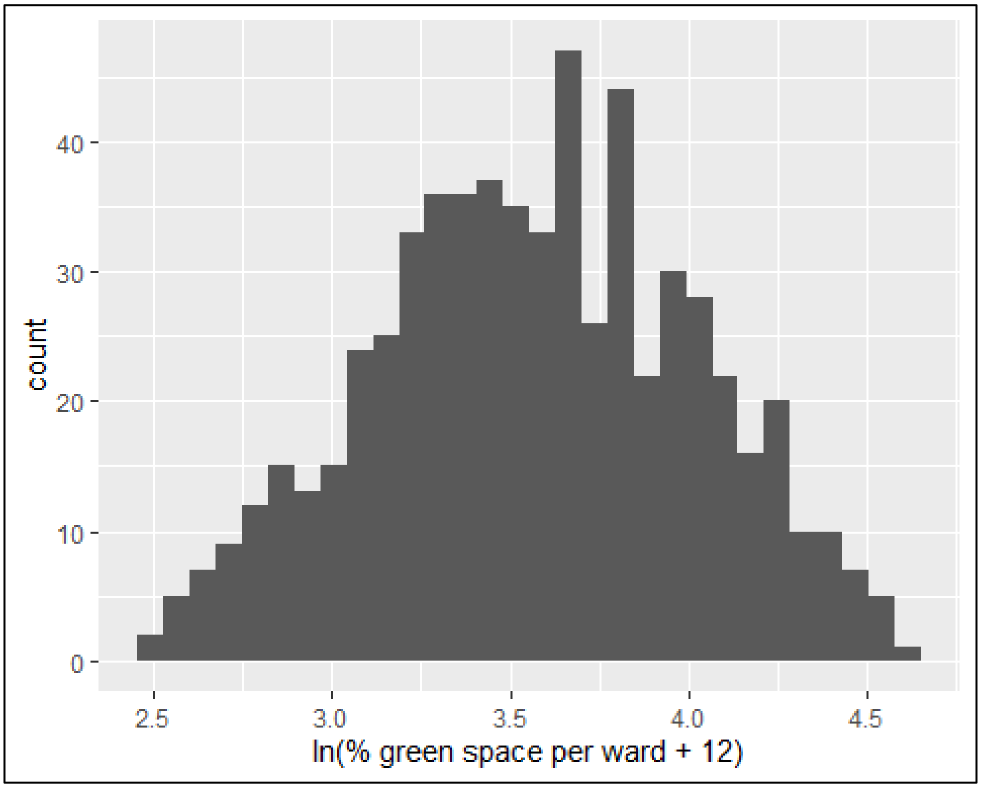

Figure 2.3 Histogram of normalised % of green space per ward in Greater London

It is important to note the limitation of the measure of access to green space used in this study, particularly the lack of consideration of network access, or access to green space in neighbouring wards. A number of datasets were considered for the study. A possible choice was a dataset Access to Public Open Space and Nature (Greater London CIC (GiGL) (2015)). This dataset contains the percentage of residential households per ward that have sufficient access to green space in line with UK governmental standards (London (2017)). However, this data was found to be extremely skewed, which made it impossible to define an appropriate model. After analysis of a number of other data sources, the source used in this study was decided to be most appropriate.

2.2 Demographic Data

After a literature review, a shortlist was created containing independent variables likely to be significantly correlated with access to green space. These were all aggregate statistics for wards including average income, average age, ethnicity, average wellbeing and average life expectancy. The data for these variables was sourced from the UK 2011 Census dataset (National Statistics (2011)). The UK Office for National Statistics provided a rigorous data collection system to capture the data on the questionnaires, interpret it and confirm it, making the census data trustworthy and reliable. As the data for these variables share a source, they are consistent in time. However, since these data were collected ten years ago, they may no longer be representative. The census data required refining and cleaning before use. Unnecessary data was removed, and the data was examined for null values. The data was then compiled in a comma separated values file for use in R. This file was joined to a shapefile of London ward boundaries, producing a shapefile containing geographic boundaries, demographic and health information. The following variables required transformation with a natural logarithm in order to fit a normal distribution: number of black people in the ward, number of Asian people in the ward, average income of the ward, number of people of 'Other' ethnicity.

2.3 Data Multicollinearity

Multicollinearity refers to the precise correlation or high correlation between explanatory variables in the linear regression model (Daoud (2017)). Multicollinearity between independent variables may significantly reduce the regression model's fit. Figure 2.4 and Table 2.1 show the correlation between independent variables, and the VIF values for independent variables, respectively.

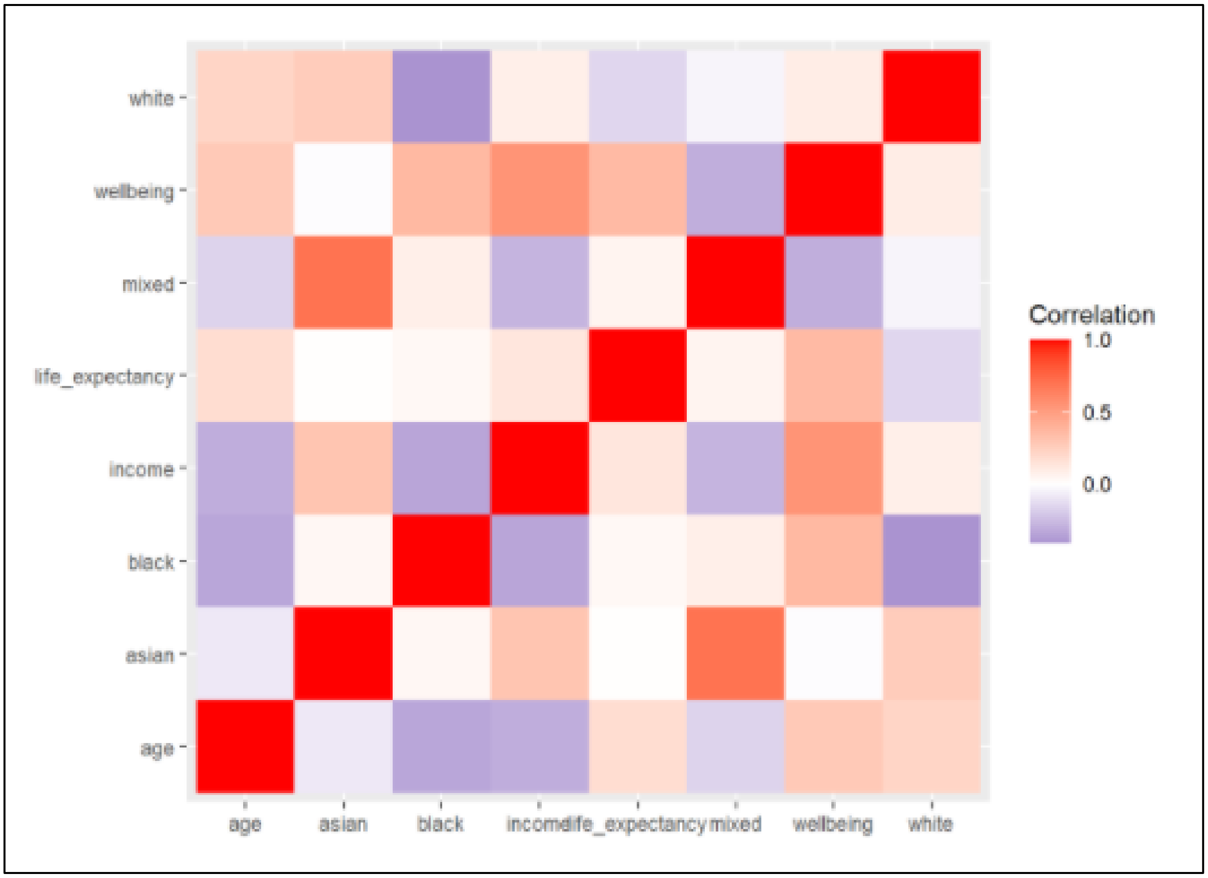

Figure 2.4 Correlation matrix for dependent variables

Independent variable	VIF
age	2.628899
life expectancy	2.142507
income	2.971197
asian	1.779267
black	6.177844
mixed	3.613875
other	2.365799

Table 2.1: VIF values for independent variable

The results of tests for multicollinearity revealed high correlations between a number of the independent variables. In general, the correlation coefficient between 0 and 0.5 means that the model is less likely to have multicollinearity problems, while the correlation coefficient between 0.5 and 1 means that the model may have serious multicollinearity problems. If the VIF is greater than 5, there could be obvious multicollinearity in the model (Daoud (2017)). Combined with these two parameters, the data of wellbeing and white ethnicity were removed to reduce the multicollinearity.

3. Exploratory Spatial Data Analysis

Exploratory spatial data analysis (ESDA) was performed to gain an insight into the spatial patterns that are present in the data. This ESDA was used to support our decision when choosing the most appropriate models.

3.1 London Greenspace distribution

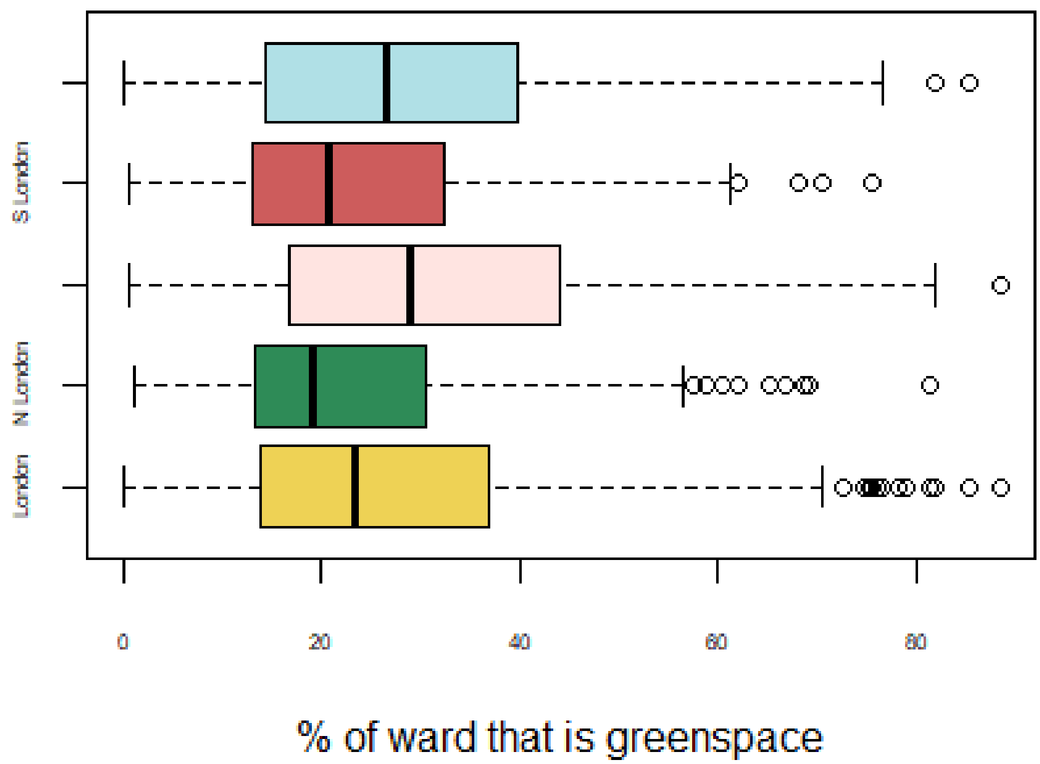

Figure 3.1: Box plots of green space distribution in sub-sections of London

London is a highly urban city with a considerable number of wards comprised of less than 11.9% green space. The median percentage of green space across Greater London was 23.5%.

The study area has been evenly divided into 4 sections: North, East, South and West London. The boundaries of these areas were decided by considering adjacency and similarity of characteristics. Figure 3.1 illustrates how the four regions of London differ in terms of access to green space. North London has the lowest median % of green space, followed by South London, West London and finally East London. North London displays the highest positive skew, with South London also being positively skewed, whilst East and West London are more normally distributed.

3.2 Demographic distribution across London

3.2.1 Income

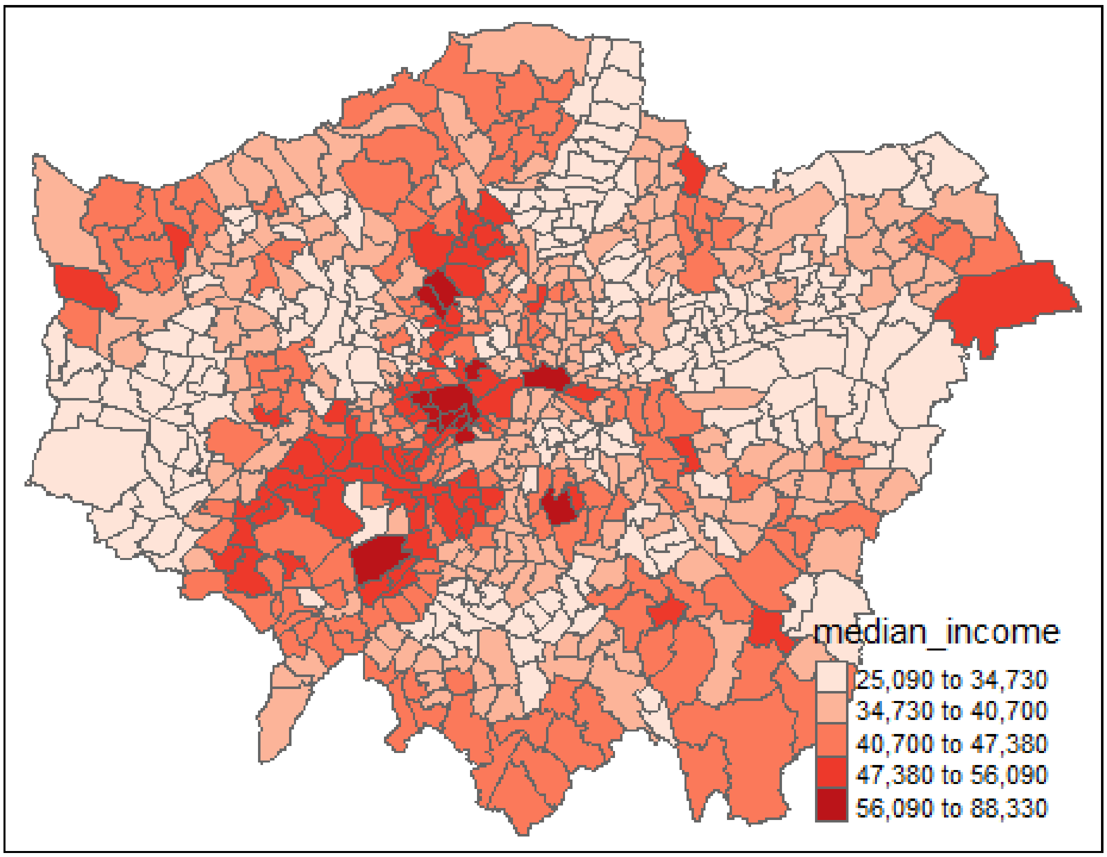

Figure 3.2: Plot of Median Income per ward in Greater London

As seen in Figure 3.2, the wards with the highest median incomes (above £56,000) are found in areas of London such as Kensington and Chelsea, Richmond and Hampstead. In comparison, those with the lowest median incomes are located on the periphery of the city, particularly towards the West and East, with a section of low-income wards along the River Lea valley in the North.

3.2.2 Age

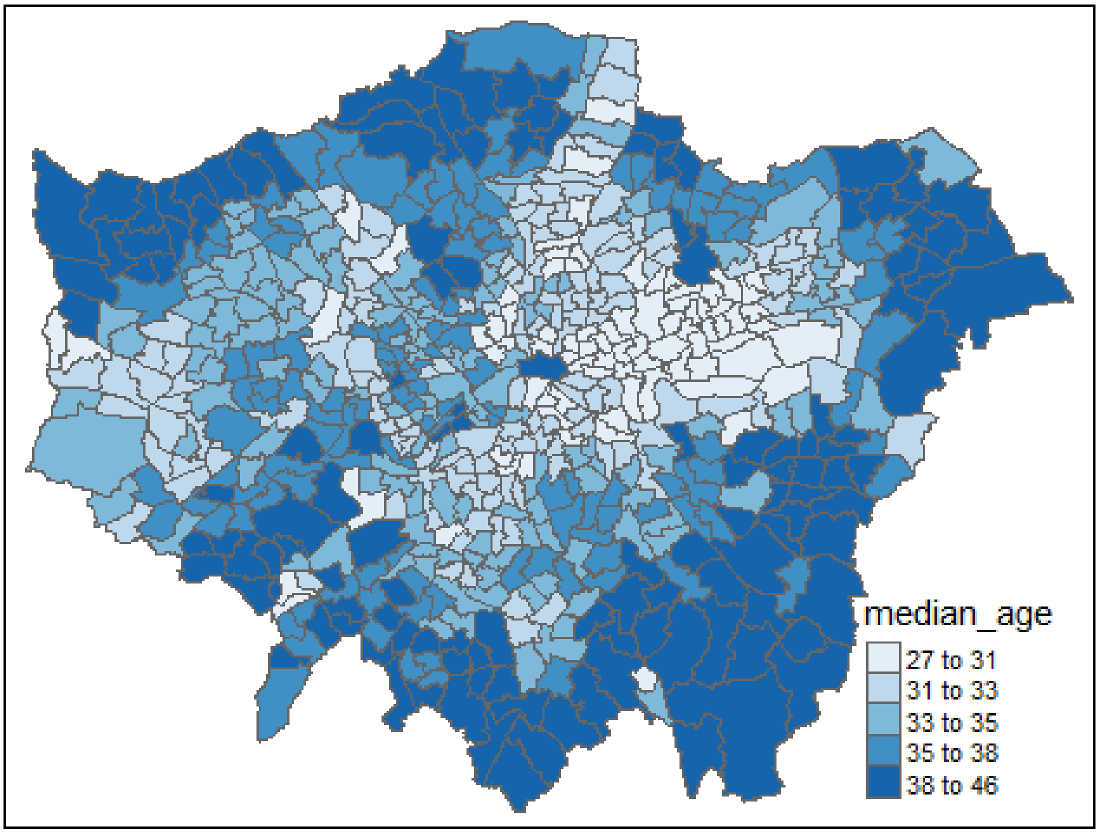

Figure 3.3: Plot of Median Age per ward in Greater London

Evident in Figure 3.3, younger populations are seen in more central areas, and particularly the "East End" of London. The median age is higher in those wards on the edge of Greater London that are perhaps quieter and where access to nature may be increased. Many of the wards with the highest median ages in the South of London, on the borders of Kent, Surrey and Sussex, as well as the North West, North and North East boundaries. There are some exceptions, notably areas of higher median income in the South-West such as Richmond also have higher median ages.

3.2.3 Life Expectancy

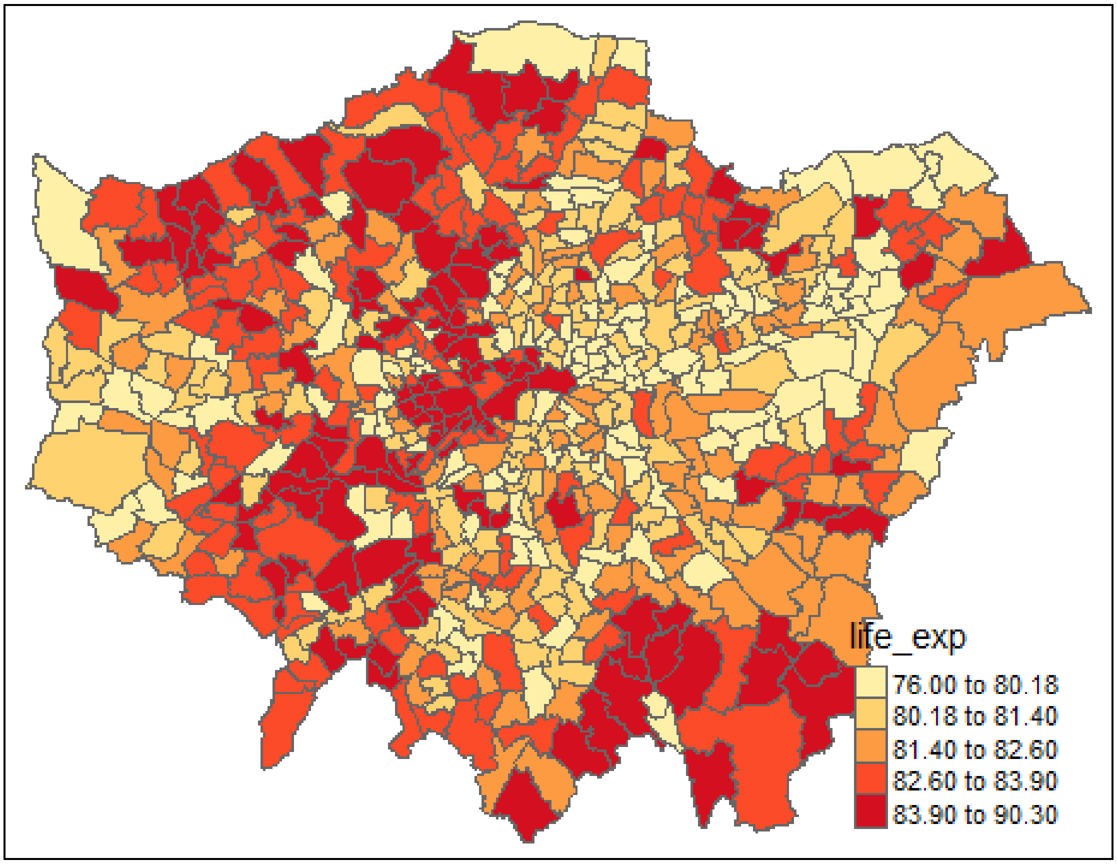

Figure 3.4: Plot of Life Expectancy per ward in Greater London

As seen in Figure 3.4, life expectancy is generally higher in the West than the East, with much of the East having a life expectancy below 80. The highest life expectancies are within Central London, towards affluent areas in the North such as Highgate, and the South West such as Kingston. Higher life expectancies can also be in areas of the South East on the border of Kent and East Sussex.

3.2.4 Ethnicity

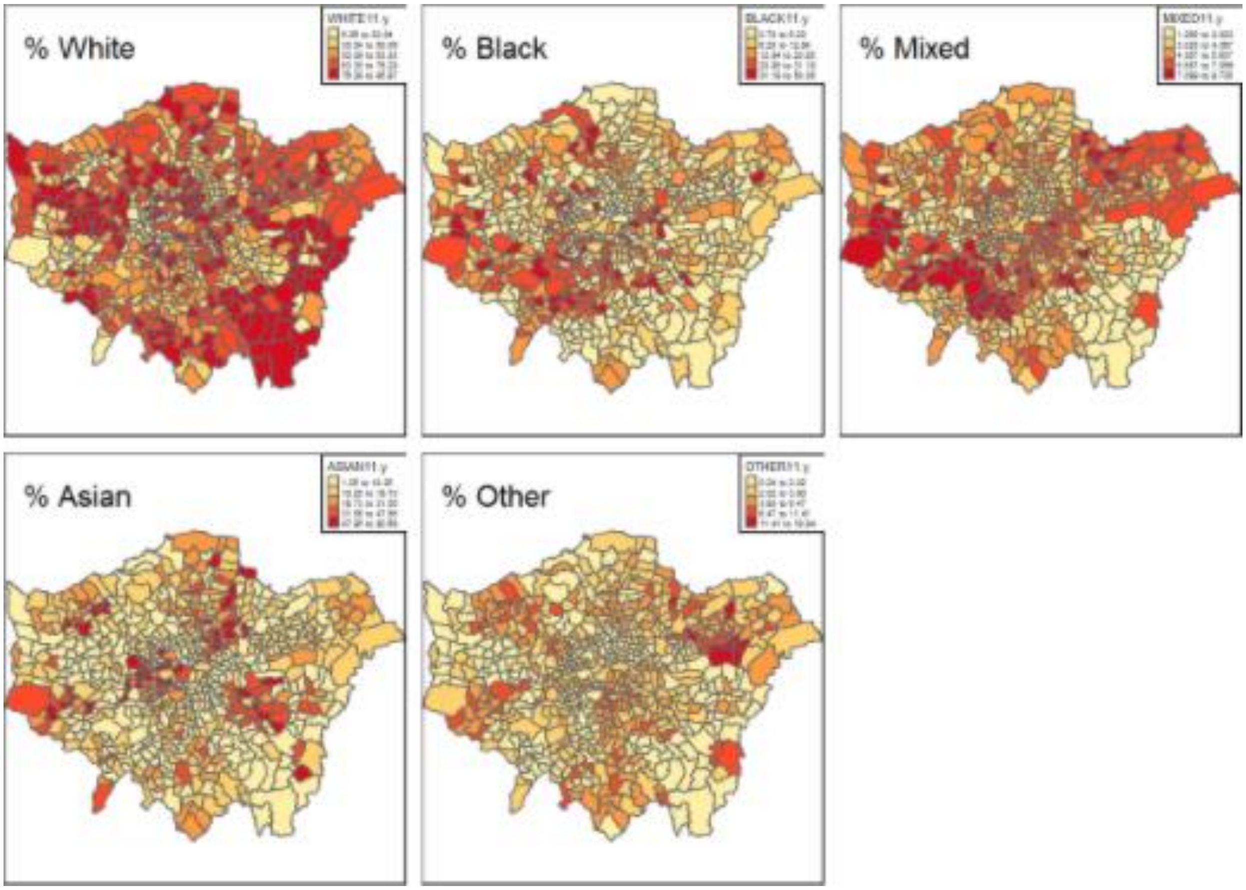

Figure 3.5: Plots of the ethnicity distribution per ward in Greater London

As evident in Figure 3.5, at the time of the 2011 census, Greater London has a large and diverse ethnic minority population with 18.4% of Londoners identifying as Asian/Asian British, 13.3% identifying as Black, 5% identifying as Mixed, and 3.4% identifying with Other ethnicities. The remaining 59.8% of Londoners identify as White. Areas with larger Black populations are concentrated around the east and south of London. For the Asian population, there is a great concentration around central London as well as the Lea Valley towards the North of London, with another cluster around South east London in the Lewisham area. The Mixed population represent less than 10% in all wards, but the distribution tends to be highly concentrated around south-west London and east London, with a significantly smaller Mixed population in the south-east. There are clusters of Other ethnicities around East London, particularly in the Stratford area and around the North West and Hillingdon area.

3.3 Spatial Autocorrelation

3.3.1 Global Moran’s I

Tests of spatial autocorrelation enable clear understanding of the extent to which neighboring wards in Greater London are significantly clustered spatially. The Global Moran’s I was used to measure spatial autocorrelation between green space across every ward in London, by comparing the actual value for green space to a distance-weighted matrix of neighbours, returning a value between -1 and 1 (Houlden et al. (2019)). This is calculated through the following formula:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

Where n is the number of spatial locations; x_1 and x_j are the values of the spatial process measured at every location i and j, respectively; $\bar{x}$ represents the mean of x; w_{ij} is the element of a spatial weight matrix W giving the spatial weight between locations i and j.

The Global Moran’s I across Greater London was 0.285, suggesting positive autocorrelation. For the Global Moran’s I statistic, the null hypothesis states that the spatial processes promoting the observed pattern of green space occur by random chance. To test this, we can use a Monte-Carlo simulation which uses a permutation test to examine how often the value of the observed statistic is seen through randomly chance.

The Monte-Carlo simulation of Moran’s I shows that the observed value of Moran’s I was higher than any of the 999 random permutations, and the observed value has a p-value of 0.001. This allows us to reject the null hypothesis. We can therefore conclude that, at the ward level in Greater London, green space is significantly positively autocorrelated.

3.3.2 Local Moran's I

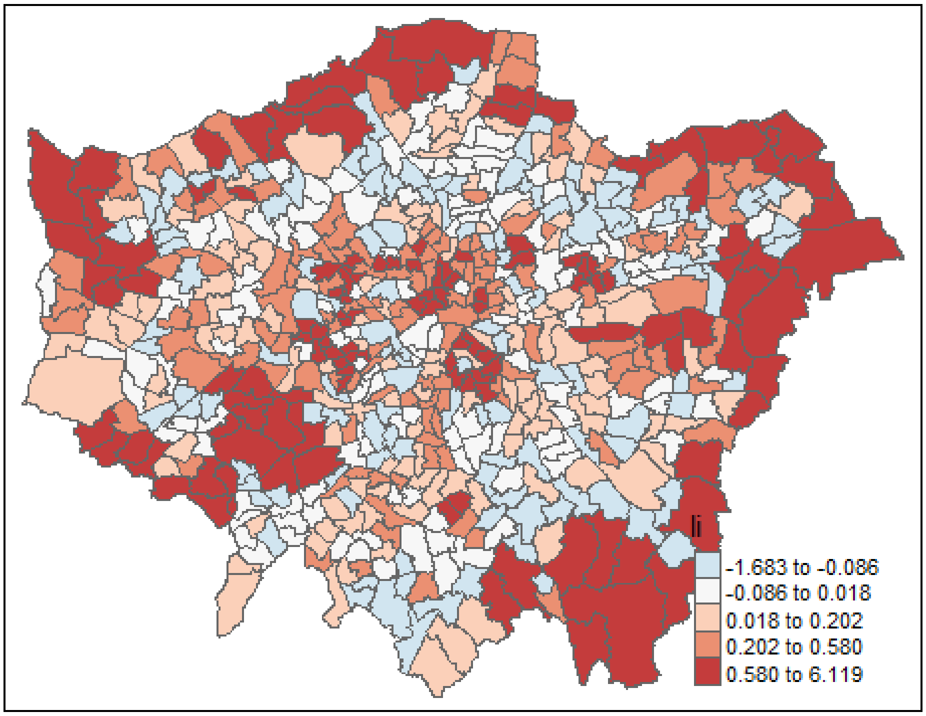

Figure 3.6: Plot of Local Moran's I of Green Space in Greater London

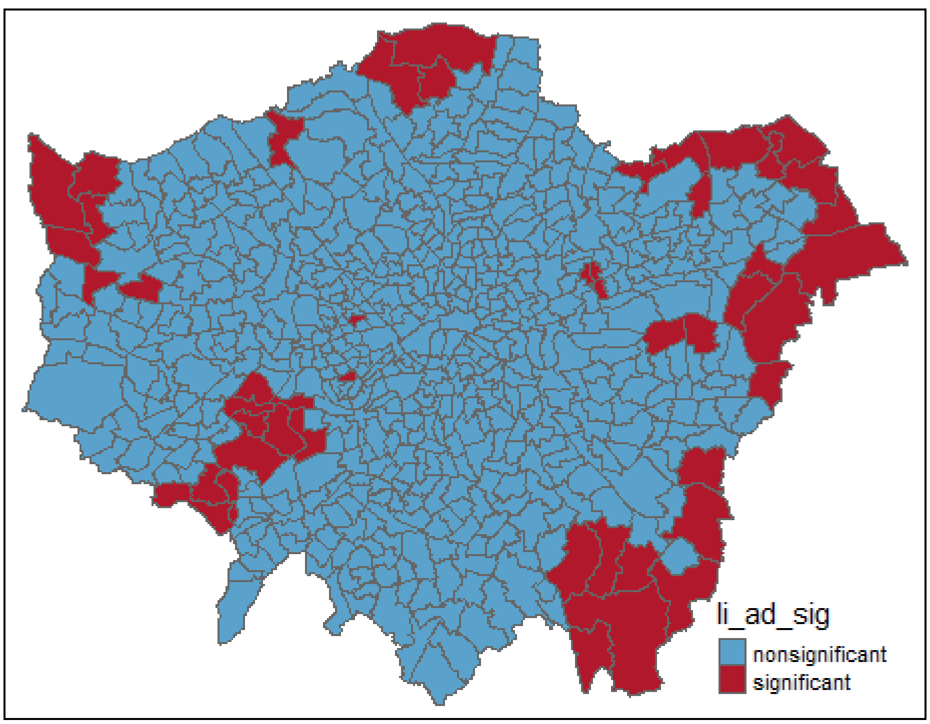

Figure 3.7: Plot of Significant and Non-Significant Local Moran's I Results in Greater London

In addition to the Global Moran's I, it is useful to examine the Local Moran's I (Figure 3.6, Figure 3.7) which is a method to identify local clusters and local spatial outliers.

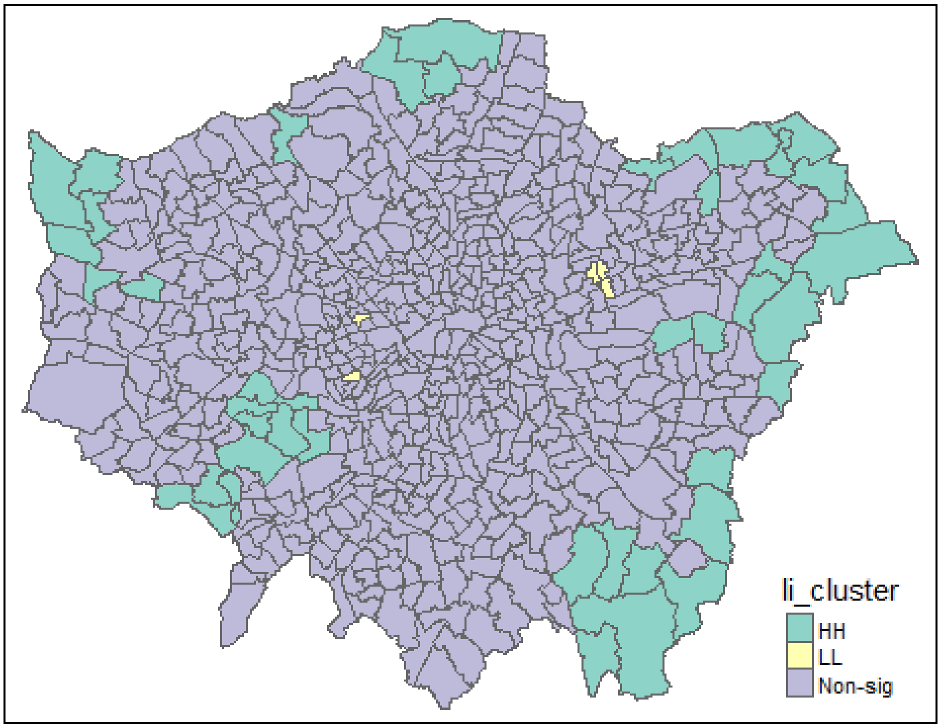

Figure 3.8: Plot of Greenspace Clusters in Greater London

Figure 3.8 shows significant High-High clusters around the perimeter of East, North and North-West London. The area around Richmond in the South West of London has a High-High cluster, presumably due to the presence of one of London's largest parks (Richmond Park). In addition, there were clusters of Low-Low green space access, including around the highly urban wards of Boleyn, Green Street East and Green Street West.

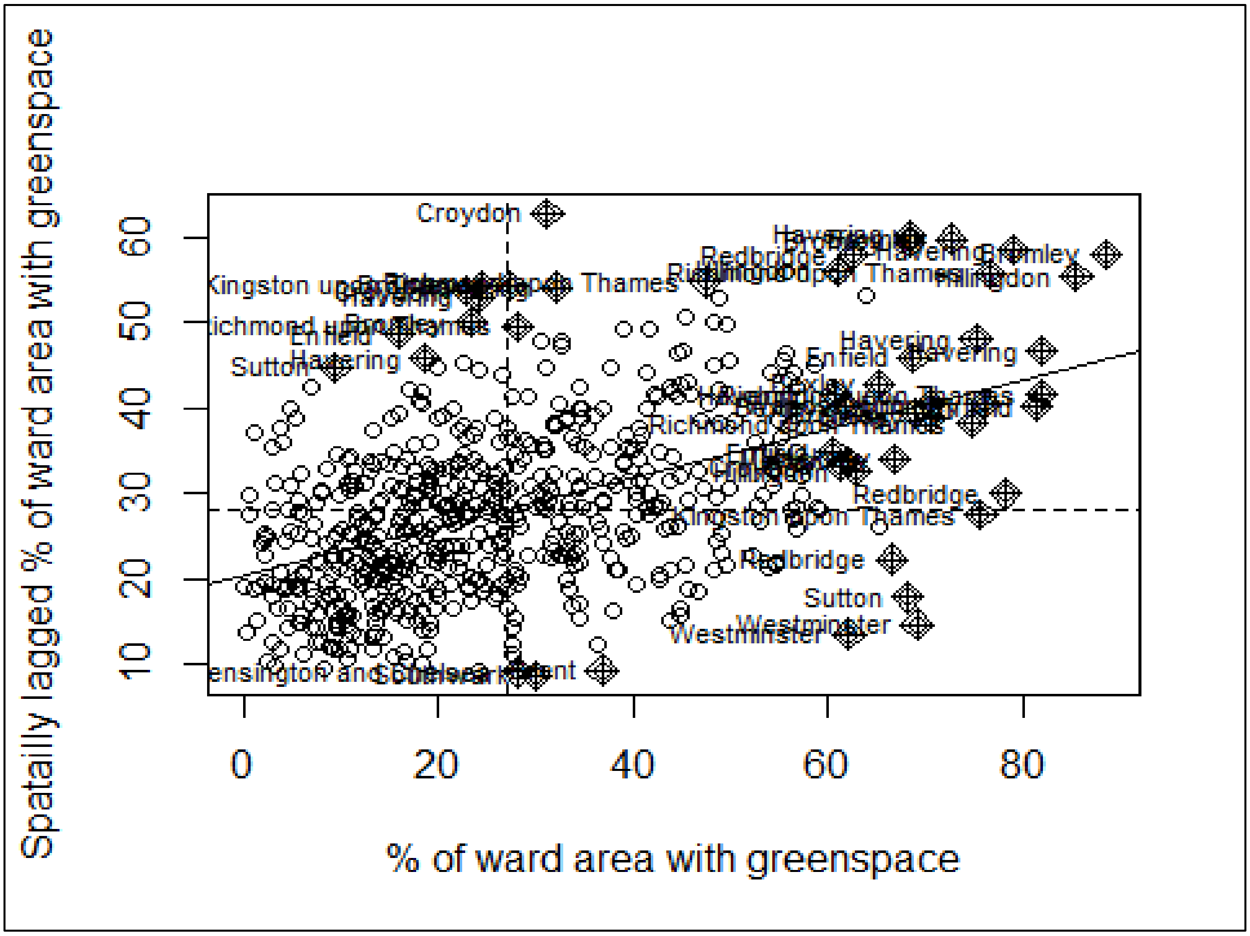

Figure 3.9: Moran Scatterplot of Green Space

The Moran scatter plot shown in Figure 3.9 displays the % green space and London borough of each ward. Clustering of boroughs in areas of the scatterplot can be seen. This supports the High-High, Low-Low map in Figure 3.8. The scatterplot displays those High-High areas in the top right, with the Low-Low areas in the bottom left. You can see that boroughs on the periphery of London such as Havering and Redbridge where access to greenspace is high are plotted in the High-High corner of the graph in Figure 3.9.

3.3.3 Semivariogram

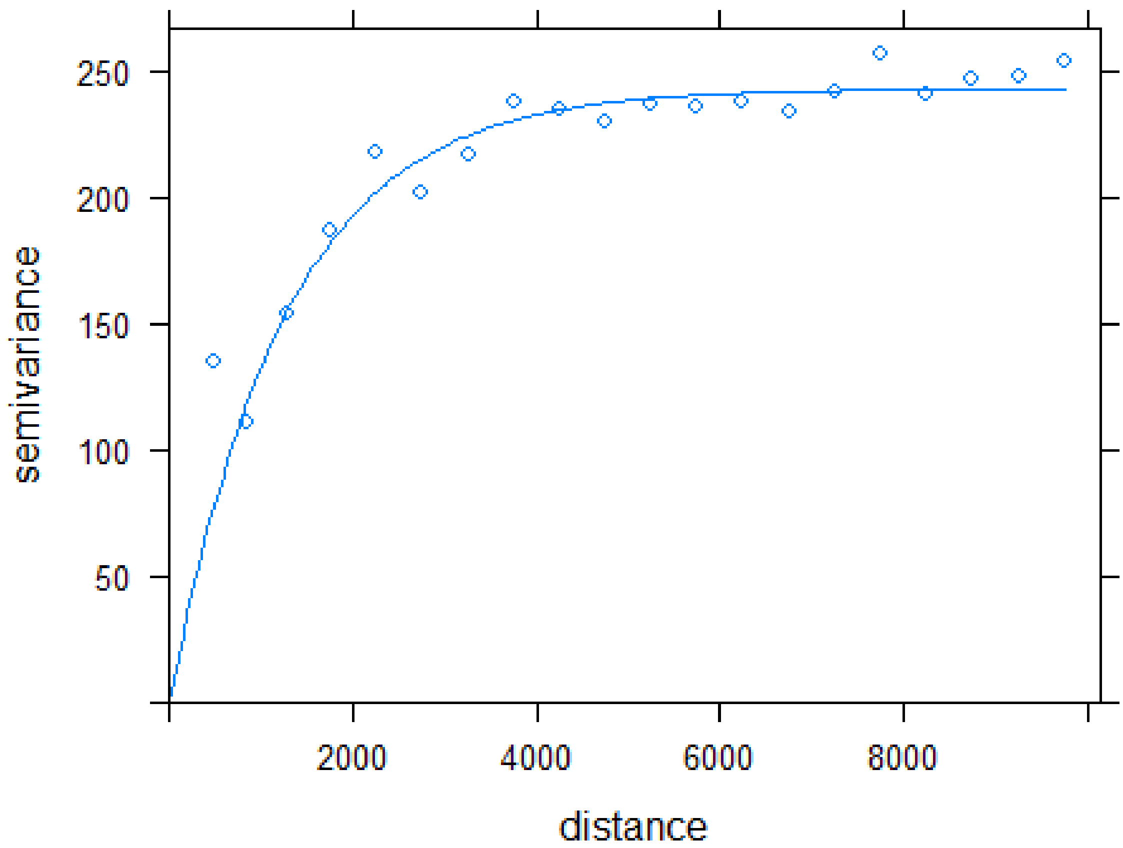

Figure 3.10: Greenspace Semivariogram

A Semivariogram examines how semi-variance increases with distance. From the Semivariogram in Figure 3.10 we can see that the semi-variance increases up to a distance of around 4000m, at which point it levels off.

4. Methodology

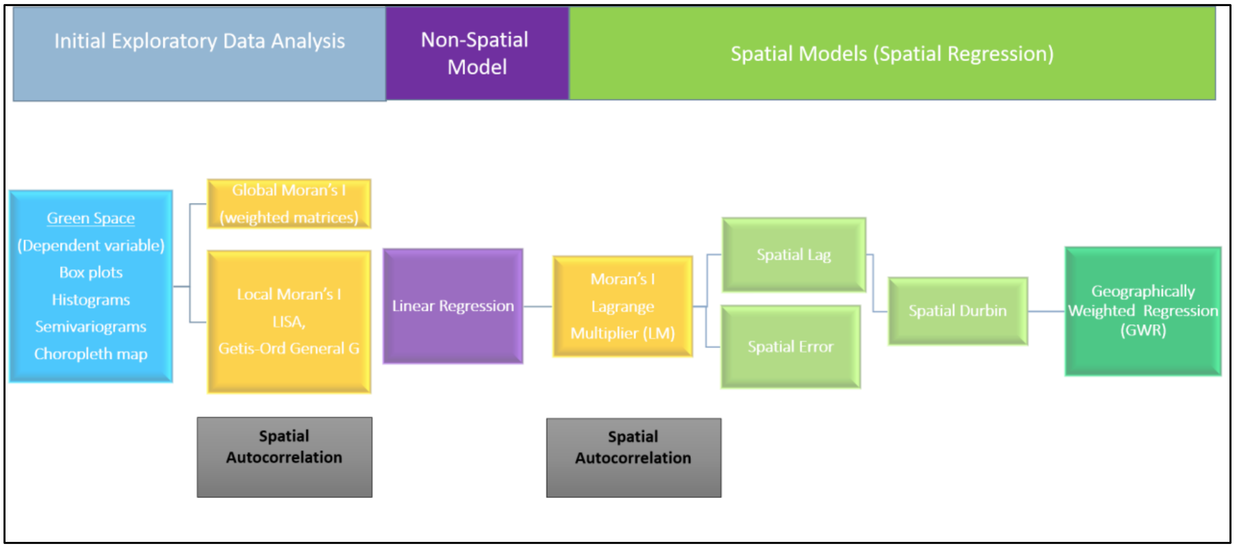

Figure 4.1 The analyse, evaluate, decision-making approach used for our study.

A literature review was completed in order to gain domain knowledge for the study, establish gaps in the literature and understand the methods used by other researchers for analysing variations in access to green space. From this appropriate methods and variables were established. Data was sourced with careful consideration to reliability and completeness. Exploratory analysis was performed on this data, and variables were normalised where required. Tests for local and global spatial autocorrelation were performed to establish whether spatial regression models would be required; these included Global and Local Moran's I, and Getis and Ord's Gi and Gi*. Spatial autocorrelation can be described as the extent to which attributes of objects are significantly clustered spatially, which can result in the possibility of underestimating errors and overestimating the statistical significance of regression coefficients in a model (Haining (2003)). Testing and modelling were performed on Greater London and sub-sections of London to understand variation in results in different areas and at different scales.

An Ordinary Least Squares (OLS) multiple linear regression model was the first model created. The residuals of this model were examined for spatial autocorrelation using

Lagrange Multiplier tests and Moran's I. The results of these tests guided the procedure for model testing and selection. After each stage of modelling the results were analysed and evaluated before deciding whether further modelling was required, this procedure is illustrated in Figure 4.1. Even with this guidance, all models for each study area and dataset were run to provide supporting evidence for model selection and highlight a model's unsuitability. The spatial models were tested in the order Spatial Lag, Spatial Error, Spatial Durbin, Geographically Weighted Regression, OLS with Spatial Filtering. Finally, the most suitable model for each study area was chosen.

The spatial lag model takes significant autocorrelation in the dependent variable as the assumption but may also take varying spatial scales in the data into account (Chi and Zhu (2008)). On the contrary, the spatial error model suggests spatial autocorrelation in errors, as a result of key independent variables that have not been included in the model. The third spatial regression model used was the Spatial Durbin model, which implies that autocorrelation may be present in one or more independent variables, as well as the dependent variable. A Geographically Weighted Regression (GWR) model allows for the model to vary over space, by passing a search window from a multitude of points in the study area and fitting a distance weighted regression model each time (Brunsdon (2008)). A benefit of the GWR over more basic spatial regression models is that it accounts for spatial heterogeneity, which can be particularly important to represent parts of London that are on the boundaries of wards, for example.

The OLS with Spatial Filtering model uses Eigenvector Spatial Filtering Regression (ESFR) to explain the underlying spatial processes in the regression model and account for the autocorrelation present. The spatial filter is included into the OLS regression model to represent spatial autocorrelation as a synthetic variable obtained from a linear combination of selected spatial weight matrix eigenvectors (Griffith and Paelinck (2011)). The eigenvectors in a linear format allow for the filtering of spatial autocorrelation in the regression residuals, as well as increased model accuracy and reduced uncertainty as each eigenvector is uncorrelated with each other (Getis and Griffith (2002)).

The study was completed using the R programming language within the RStudio application, enabling the incorporation of tools to compute our results and display them in graphical form. The codes used have been stored for reproducibility.

5. Results

Base Formula:

$$ln(g + 12) = age + le + ln(income) + ln(ap/tp) + ln(bp/tp) + (mp/tp) + ln(op/tp)$$

where:

g = % green space in ward
age = mean age in ward
le = life expectancy in ward
$income$ = mean income in ward
tp = total population in ward
ap = asian population in ward
bp = black population in ward
mp = mixed ethnicity population in ward
op = population in ward of ethnicities other than white or those previously stated

Spatial Durbin, Error and Lag models for each respective study area used the same base formula, predictive variable, dependent variable, and spatial weight matrix (first-order queen criterion) as above. The wards immediately adjacent were considered neighbours whether they touched the corners or the flat edge of a ward's polygon. An identical strategy was used in the GWR model though varied by selecting a relevant bandwidth for the isotropic Gaussian Kernel. In this study, when a result is deemed statistically significant, it will refer to a p-values ≥0.05, correspondingly <0.05 is not significant.

5.1 OLS

The Ordinary Least Squares (OLS) models fitted to each region were significant (at the 95% level) and returned an adjusted R^2 0.1052-0.2722, West London returning the highest result. The OLS models only explained 10-27% of the green space variation for each location. Combined with the High Residual Standard Error (RSE) 0.3804 – 0.4365, it can be concluded that OLS was a poor fit for modelling green space both in the subdivisions and the whole of London. This can be observed in the large difference between the actual value of green space and the predicted value.

Figures 5.1, 5.2 and 5.3 display the Q-Q plots for all datasets. All plots showed that points were above the theoretical normal line at the lower values near the beginning and below the line near the end. The plots exhibited a thin-tailed distribution; deviation at the ends was considered negligible. Consequently, the datasets were considered close to normal distribution: the Shapiro-Wilk test provided further evidence of normality.

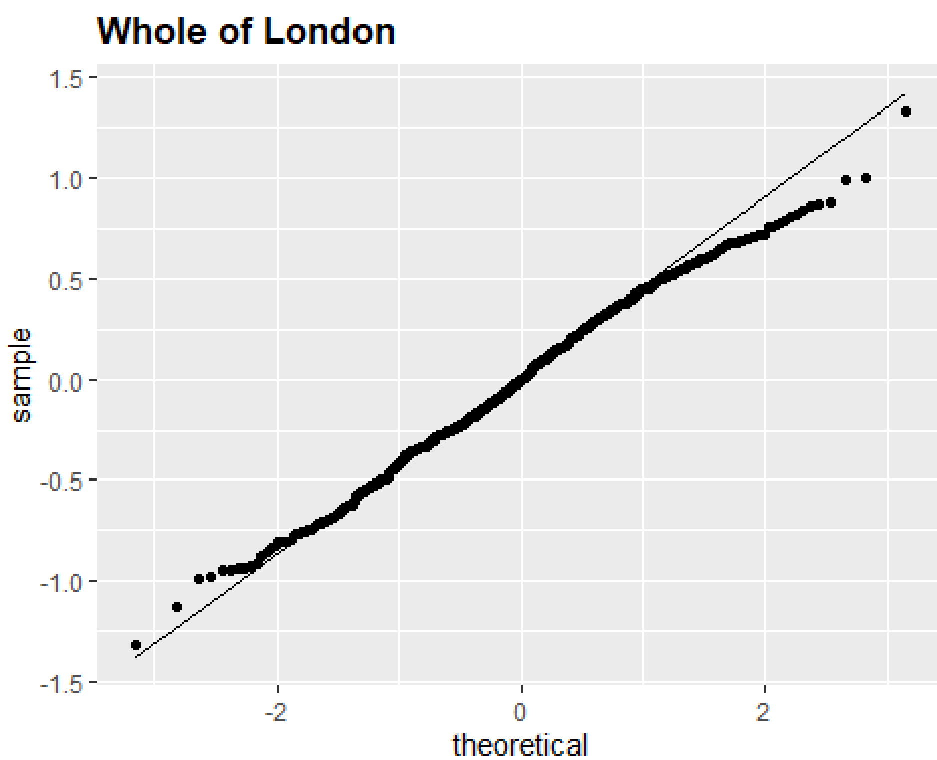

Figure 5.1: Q-Q plot of linear model residuals, Greater London

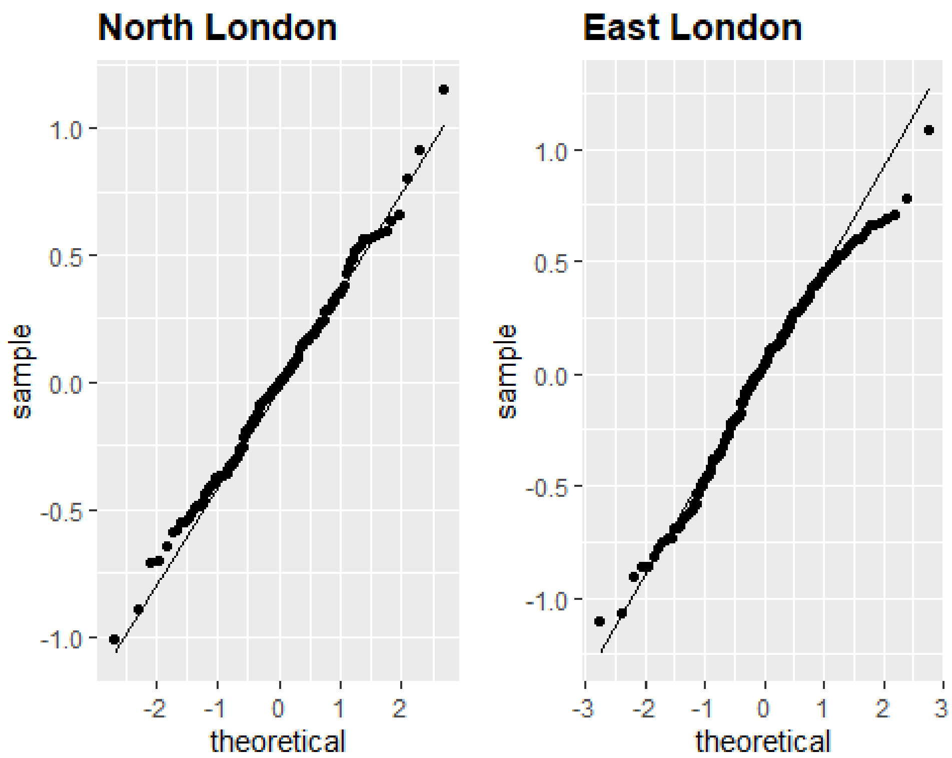

Figure 5.2: Q-Q plot of linear model residuals, North and East London

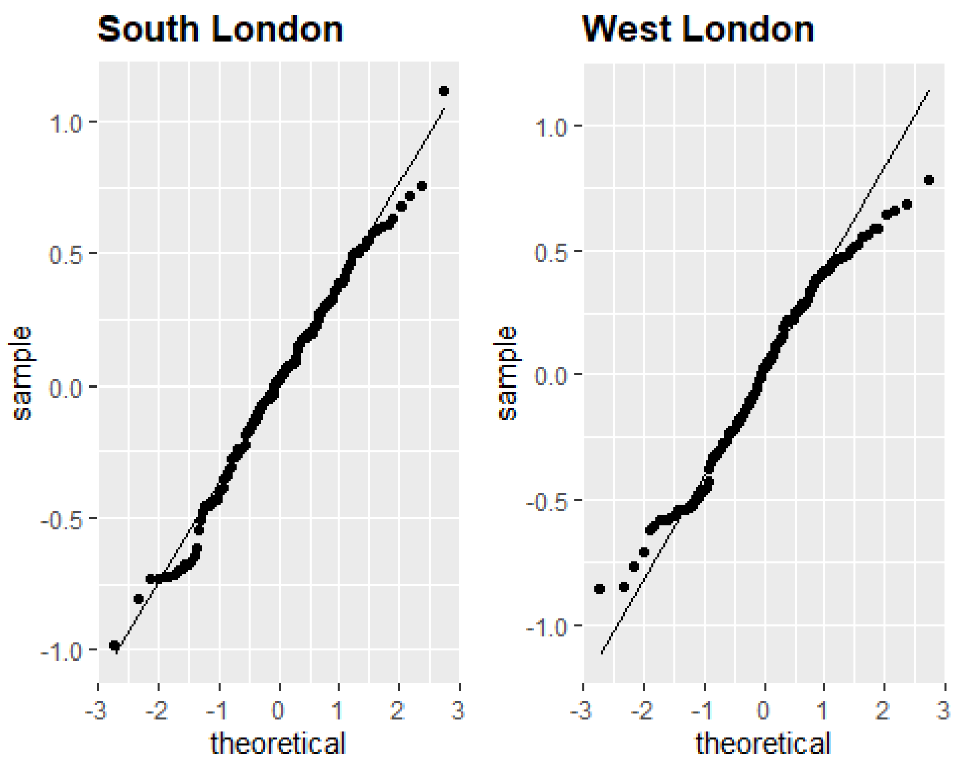

Figure 5.3: Q-Q plot of linear model residuals, South and West London

5.2 Shapiro-Wilk Normality Test

The Shapiro-Wilk (Shapiro and Wilk (1965)) results for the OLS residuals for North, East and South areas were not significant at a 95% confidence level (p-value > 0.05) and had a high W value. Thus, the null hypothesis of a normal distribution was not rejected, and normality in the residuals was assumed. Conversely, London as a whole and the West were weakly significant at a p-value of 0.02431 and 0.00907. For this study these values were deemed acceptable enough for the assumption of normality. Similarly, the natural logarithm transformation and addition of 12 to the dependent variable showed a significant increase in assumed normality.

5.3 Moran's I on OLS residuals

The Global Moran's I test results on the OLS residuals detected statistically significant and positive spatial autocorrelation; similar green space values were clustered together in all areas apart from the South (Table 5.1) The whole of London and the West had the largest Moran's I of 0.171590363 and 0.177676307 and a p-value 5.826E-15 and 2.65E-06. The results stated that, with the exception of the South, the null hypothesis of spatial randomness could be rejected with confidence. The results also suggested that the unaccounted spatial autocorrelation in the OLS residuals was a key factor causing the high RSE and low R^2 value.

Table 5.1: Global Moran I for Regression Residuals

		London	North	East	West	South
1	Observed Moran's I	0.1715903627	0.087117028	0.108292348	0.177676307	-0.03053761
2	Expectation	-0.0083988916	-0.0352447	-0.025650785	-0.030140504	-0.03025517
3	p-value	5.826E-15	0.007698	0.001336	2.65E-06	0.5023

5.3.1 South

Contrary to all other results, the South had a negative observed Moran's I -0.03053761 nearly identical to the expectation value of -0.03025517: the value if no spatial autocorrelation was present. The Moran's I value when combined with the statistical insignificance, indicated that there was not enough evidence to suggest that spatial autocorrelation was present. The fitting of a spatial regression model would not improve the prediction of green space, therefore, despite its limitations, the OLS model was concluded to be the best fit available for South London.

5.4 Lagrange Multiplier Tests

Table 5.2 shows all the Lagrange Multiplier test results.

Table 5.2: Lagrange Multiplier Diagnostics for Spatial Dependence

		London	North	East	West	South
1	LMlag	56.262	7.5719	7.2718	1.85E+01	0.96829
2	p-value (LMlag)	6.34E-14	0.005928	0.007005	1.73E-05	0.3251
3	LMerr	52.139	2.5679	5.2787	13.204	0.36006
4	p-value (LMerr)	5.17E-13	0.1091	0.02159	0.0002793	0.5485
5	RLMlag	4.1866	11.009	3.1031	6.2264	2.7989
6	p-value (RLMlag)	0.04074	0.0009069	0.07814	0.01259	0.09433
7	RLMerr	0.063807	6.0046	1.11	0.96381	2.1907
8	p-value (RLMerr)	0.8006	0.01427	0.2921	0.3262	0.1388

5.4.1 London

For London as a whole the LMerr and LMlag values were high, 56.262 and 52.139, and significant at a 95% confidence. However, the robust test indicated a Spatial Lag model should be fitted as the RLMerr was low and insignificant.

5.4.2 North

Likewise, for the North, the results indicated that a Spatial Lag model would best fit the data. The LMlag value was significant, whereas the LMerr was not, the robust test produced similar results.

5.4.3 East and West

The East and West demonstrated statistical significance for both regular LM tests. However, the robust test for RLMerr values was lower than LMlag and not significant for the West, signifying that a Spatial Lag model should be fitted. In contrast, both robust LM tests for the East had a p-value > 0.05; both Error and Lag would be run.

5.4.4 South

Both LM tests was insignificant for South London, this was expected considering the Moran's I result concluded that there was not enough statistical evidence to stipulate the presence of autocorrelation.

5.5 Spatial Regression Models

The LM test and Global Moran's I guided the procedure for model testing and selection. Even with this guidance, all models for each study area and dataset were run to provide supporting evidence for model selection and highlight a model's unsuitability.

5.5.1 Spatial Lag Model

The Spatial Lag model's Rho value ranged from 0.29468-0.45669 with significant p-values for London, North, East and West. The results imply that there was a lagged relationship between green space values in neighbouring wards. The spatial autocorrelation detected in the Moran's I was modelled out for all three locations; the p-value exceeded the 0.05 threshold. Therefore, the null hypothesis of spatial randomness could not be rejected.

Compared to the OLS, the AIC of the Lag model demonstrated a distinct decrease, across all study regions: the whole of London showed the most considerable change

695.5 to 648.57. The reduction signified the model's better fit for predicting green space values. The Rho and residual autocorrelation test's lack of statistical significance in the South was reflected in the increased AIC, from 159.21 to 160.9, this was interpreted as the worst fit.

5.5.2 Spatial Error Model

The LM test results suggest that only the East dataset be modelled with the Spatial Error; despite this, all datasets were passed through each model and tested to allow for complete comparison. The Spatial Lag's unique coefficient, lambda, had a high value and was statistically significant in all regions apart from the North and South; which highlighted potential clustering in the residuals. Nevertheless, the log-likelihood was more negative; also, the AIC was higher than the Spatial Lag results. The Akaike Information Criteria rule of thumb states that if the AIC value of models falls within two or less, they have the same explanatory power (Fabozzi et al. 2014). None of the Spatial Errors results fell within the parameter; for that reason, the model was deemed a less adequate fit relative than the Lag.

5.5.3 Spatial Durbin

The Rho values for the Durbin model in the North was not significant at a 95% confidence level. The reduced AIC and less negative Log-likelihood for the London dataset provided evidence that Durbin model would be a better fit than the Lag. The AIC of the East had a difference greater than two in comparison to the Lag; therefore, the model was interpreted as weaker in explaining power. Across all study areas, spatial autocorrelation had been modelled out similar to the Lag. In the results for the West, conflict arose between the Log-likelihood and AIC value between the Durbin and Lag model. Durbin had a greater AIC 143.31 and less negative log-likelihood -54.65 compared to 137.71 and -58.85 of the Lag. Despite the contradiction, Lag was interpreted as the best fit, based on the principle that AIC was used as the primary model selection parameter. When similar conflicts arose throughout the study, the same approach was taken.

5.5.4 GWR

The spatial autocorrelation evident in the OLS model's residuals was reduced in all study areas in which the GWR model was run, though not wholly modelled out. Excluding the South, the observed Moran's I ranged 0.087117-0.177676 (OLS residuals) compared to 0.0175191-0.08093797 (GWR) all with an associated p-value < 0.05. In accordance with the reduction in observed spatial autocorrelation, the AIC for all study regions was considerably lower for GWR than the OLS, Lag, Durbin and Error models. The R^2 value significantly increased across all areas when compared to the OLS.

5.5.5 OLS and Spatial Filtering

The OLS model is nested within the OLS and Spatial Filtering model, hence comparison of their results represents the direct impact of Spatial Filtering function upon the model's strength. Apart from the South, every study area observed a noticeable decrease in RSE 0.3804-0.4365 to 0.3218-0.4047 compared to the initial OLS results. The adjusted R^2 rose considerably from 0.1286-0.2722 to 0.2922-0.4791.

The transition of the results demonstrated how each model's predictive power increases with spatial autocorrelation being modelled out; the Global Moran's I was statistically insignificant in each study area. Across London, East, West and North the AIC was the smallest when compared to the OLS, Lag, Durbin and Error models. However, in all areas apart from Greater London GWR had a lower AIC. There was a difference in AIC value less than one for Greater London. According to Akaike Information Criteria rule of thumb, the minor difference indicates that the two models have the same explanatory power (Fabozzi et al. 2014).

The full list of AIC values for each model is displayed in Table 5.3The GWR model for all areas, including the South, had the smallest AIC value overall. This was closely followed by OLS and Spatial Filtering, then the Lag model. Despite the lower GWR AIC value, OLS with Spatial Filtering was concluded to be the best fit in all areas except the South. This is due to the limitations of the GWR, most notably the inability of GWR to remove spatial autocorrelation in residuals. The OLS with Spatial Filtering model was the only model to fully remove spatial autocorrelation in residuals. Excluding the North and South, the Durbin model for the other locations had the fourth-highest AIC. All results, not including the South, indicated Error being the second least powerful spatial regression model. Finally, the OLS was the model deemed weakest at providing accurate predictive greenspace values.

Full results for all models are available in the Appendix.

Table 5.3: AIC Results

		London	North	East	West	South
1	Number of Observations:	625	135	173	159	158
2	OLS	695.9519	134.9908	213.9736	153.6169	159.2142
3	Error	650.3795	133.7242	210.6101	141.3932	160.7782
4	Lag	648.5731	129.6145	208.956	137.711	160.0925
5	Durbin	639.2799	131.1229	213.5265	143.3125	165.3068
6	GWR	564.501	98.46093	196.742	93.19053	144.7065
7	OLS and Spatial Filtering	564.4889	118.6283	192.471	110.4278	159.4238

6. Discussion and Conclusions

6.1 Model Selection

For most of the study areas, it is evident that a spatial regression model is necessary to account for the spatial autocorrelation and reduce errors. The principle of model selection based solely on AIC is useful but not entirely sufficient. Geographically Weighted Regression has the lowest AIC in every area apart from London (whole) and the South, indicating the best fit. However, for models in all areas there remains statistically significant spatial autocorrelation in the residuals. The Ordinary Least Squares with Spatial Filtering was the only model remove this autocorrelation, reducing the model's error and increasing its explaining power. For this reason, Ordinary Least Squares and Spatial Filtering were chosen as the best model for predicting green space in all areas where spatial autocorrelation was observed (i.e., all areas except South London).

Below are details of the models selected as most appropriate for each area. Figures 6.1, 6.2, 6.3, 6.4 and 6.5 display side by side choropleths of London, North, East, West and South. The left map represents the dataset's values for the percentage of green space; the map adjacent shows fitted or predicted values for the most appropriate model for the area.

g = % green space in ward

age = mean age in ward

le = life expectancy in ward

$income$ = mean income in ward

tp = total population in ward

ap = asian population in ward

bp = black population in ward

mp = mixed ethnicity population in ward

op = population in ward of ethnicities other than white or those previously stated

$spatial\ filter$ = fitted result of Roger Bivand's semi-parametric spatial filtering function (Bivand (2007)) applied to the Ordinary Least Squares model for the area. Documentation for the spatial filtering function can be found at https://www.rdocumentation.org/packages/spatialreg/versions/1.1-5/topics/SpatialFiltering.

VC = coefficients for the vectors of the spatial filter, as shown in Tables 7.4.5.3-7.4.5.6

6.1.1 Greater London

Ordinary Least Squares with Spatial Filtering:

$$g = exp(0.0297(age) - 0.0028(le) - 0.5967(ln(income)) - 0.0341(ln(ap/tp)) - 0.0271(ln(bp/tp)) - 0.0001(mp/tp) - 0.0498(ln(op/tp)) + VC(fitted(spatialfilter))) - 2.129$$

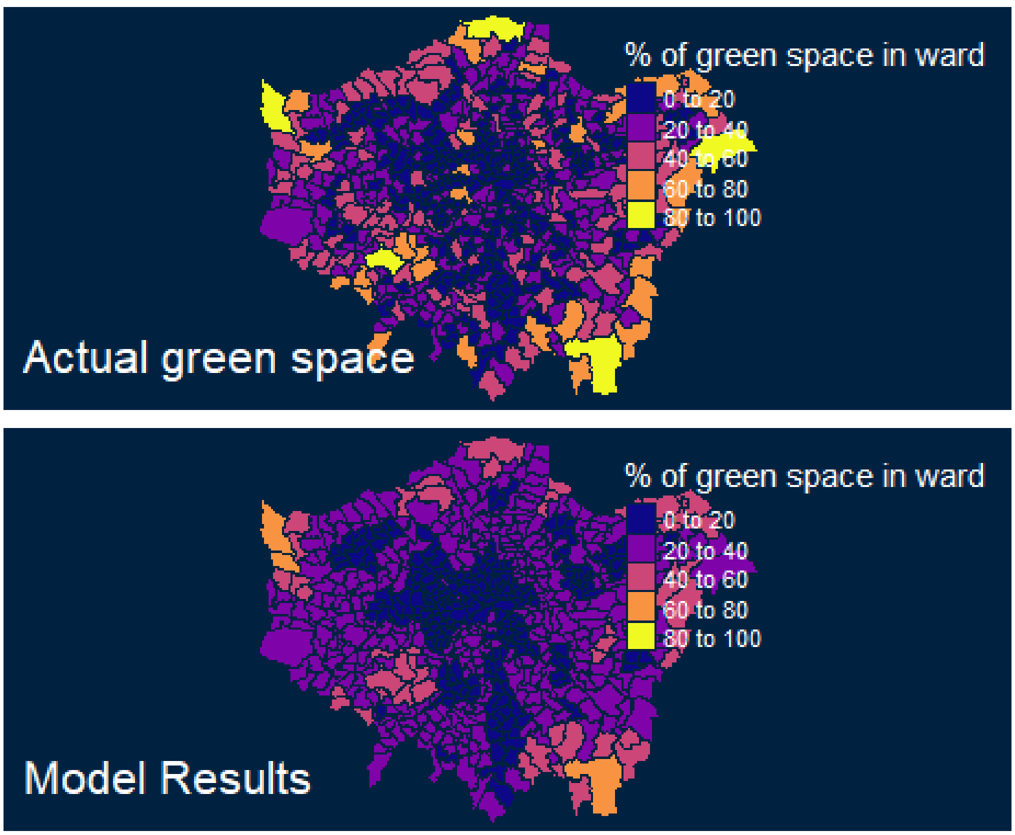

Figure 6.1: Fitted and observed values of % green space per ward in Greater London

6.12 North London

Ordinary Least Squares with Spatial Filtering:

$$g = exp(0.0724(age) + 0.0474(le) - 0.1172(ln(income)) - 0.0008(ln(ap/tp)) + 0.4206(ln(bp/tp)) - 0.0650(mp/tp) - 0.0498(ln(op/tp)) - 3.7756 + VC(spatialfilter)) - 12$$

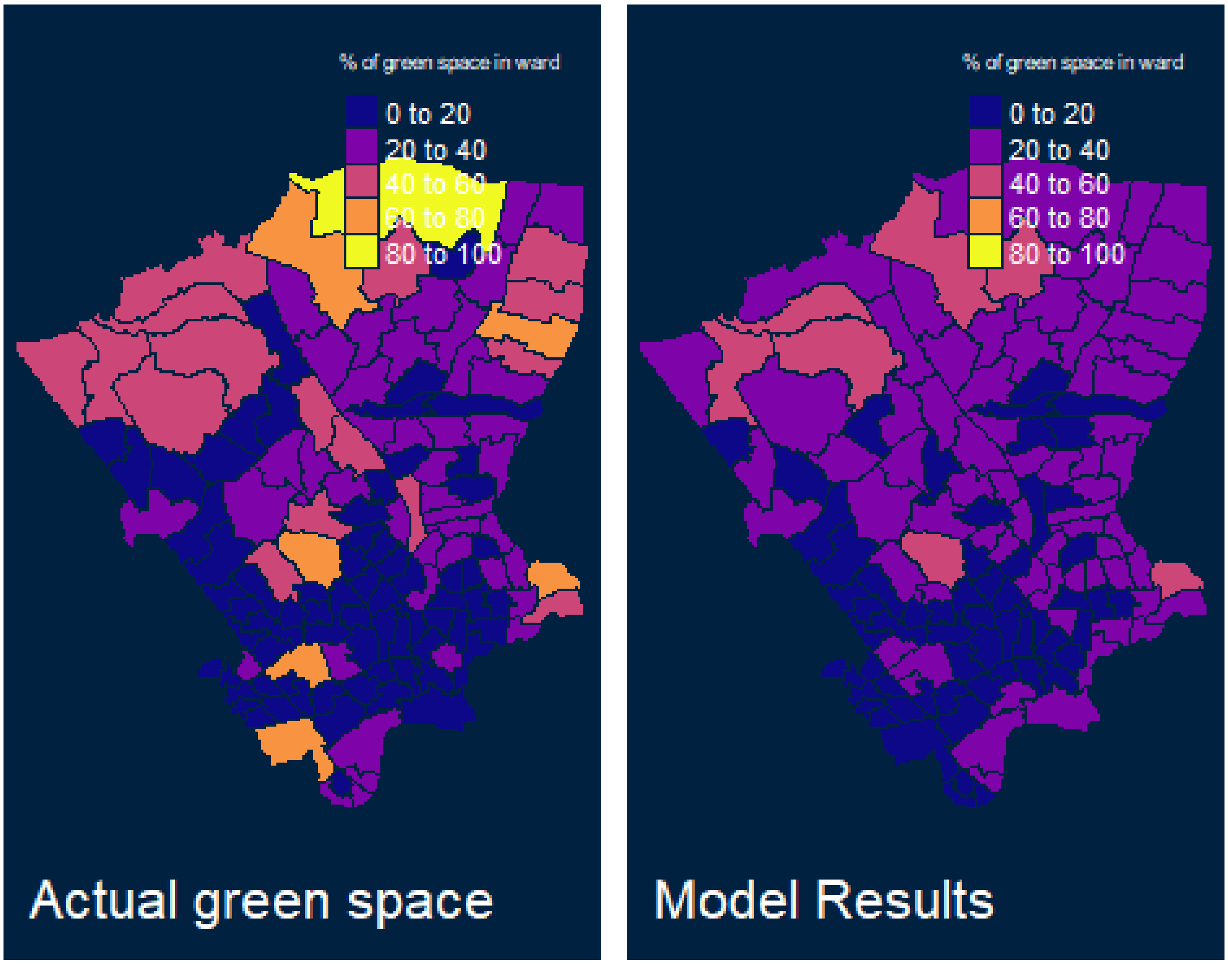

Figure 6.2: Fitted and observed values of % green space per ward in North London

6.1.3 East London

Ordinary Least Squares with Spatial Filtering:

$$g = exp(0.0119(age) - 0.0252(le) - 0.1075(ln(income)) - 0.1683(ln(ap/tp)) + 0.0236(ln(bp/tp)) + 0.0005(mp/tp) - 0.0860(ln(op/tp)) + 7.7045 + VC(spatialfilter)) - 12$$

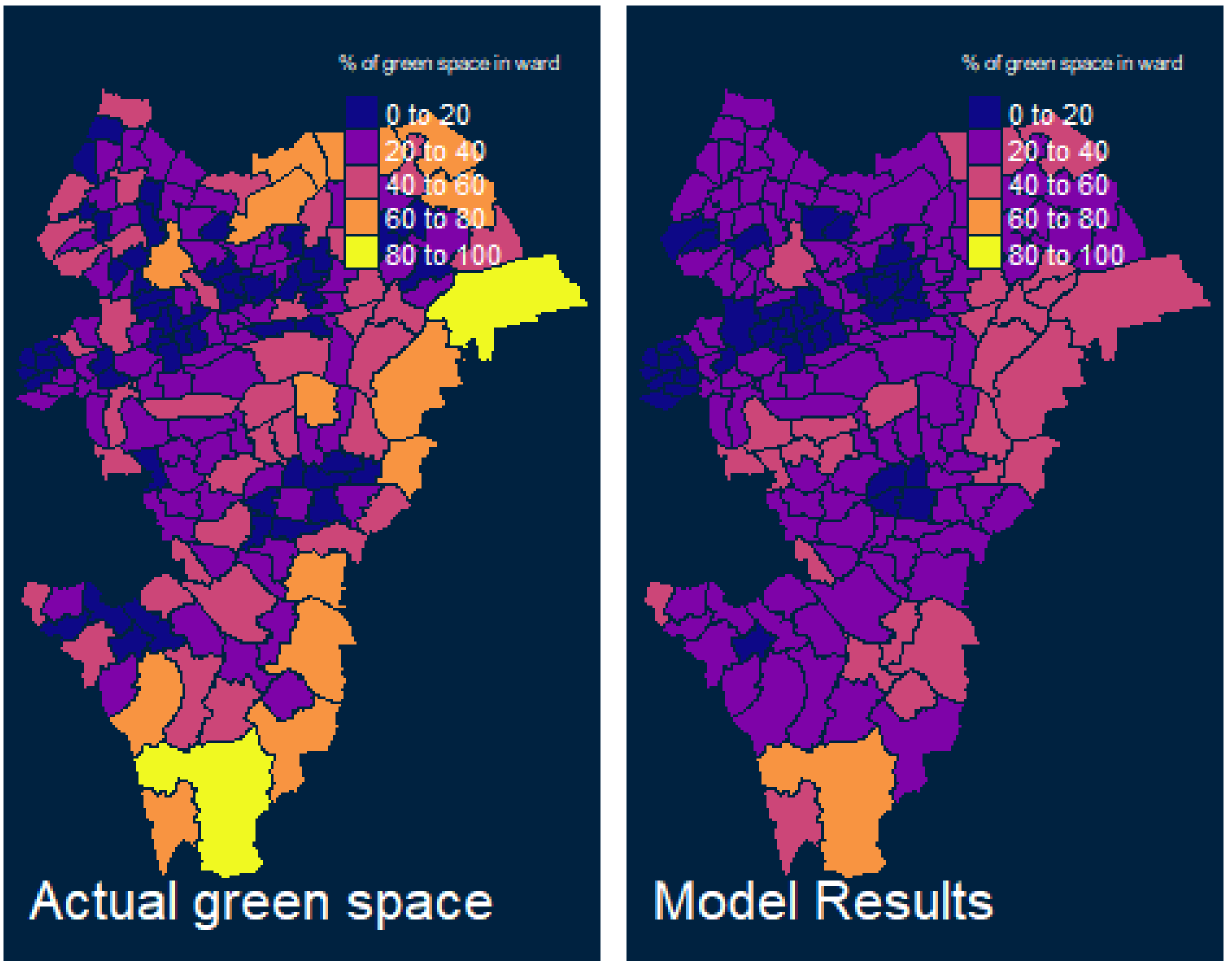

Figure 6.3: Fitted and observed values of % green space per ward in East London

6.1.4 West London

Ordinary Least Squares with Spatial Filtering:

$$g = exp(0.0504(age) - 0.0168(le) - 1.4774(ln(income)) + 0.0137(ln(ap/tp)) - 0.3282(ln(bp/tp)) + 0.0003(mp/tp) - 0.0110(ln(op/tp)) + 20.8318 + VC(spatialfilter)) - 12$$

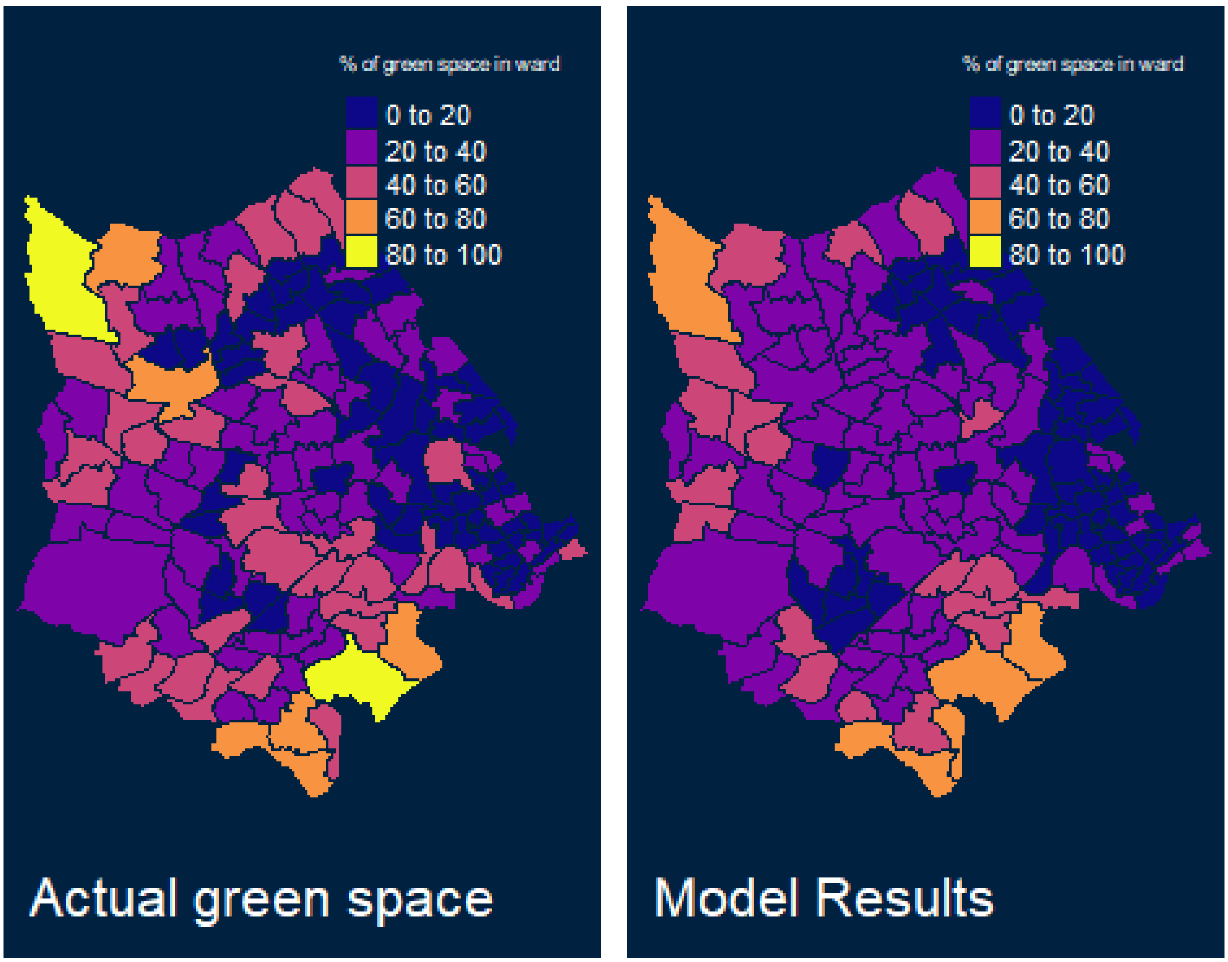

Figure 6.4: Fitted and observed values of % green space per ward in West London

6.1.5 South London

Ordinary Least Squares:

$$g = exp(0.0122(age) + 0.0155(le) - 0.3715(ln(income)) - 0.0415(ln(ap/tp)) + 0.0263(ln(bp/tp)) - 0.0002(mp/tp) - 0.1623(ln(op/tp)) + 6.9080) - 12$$

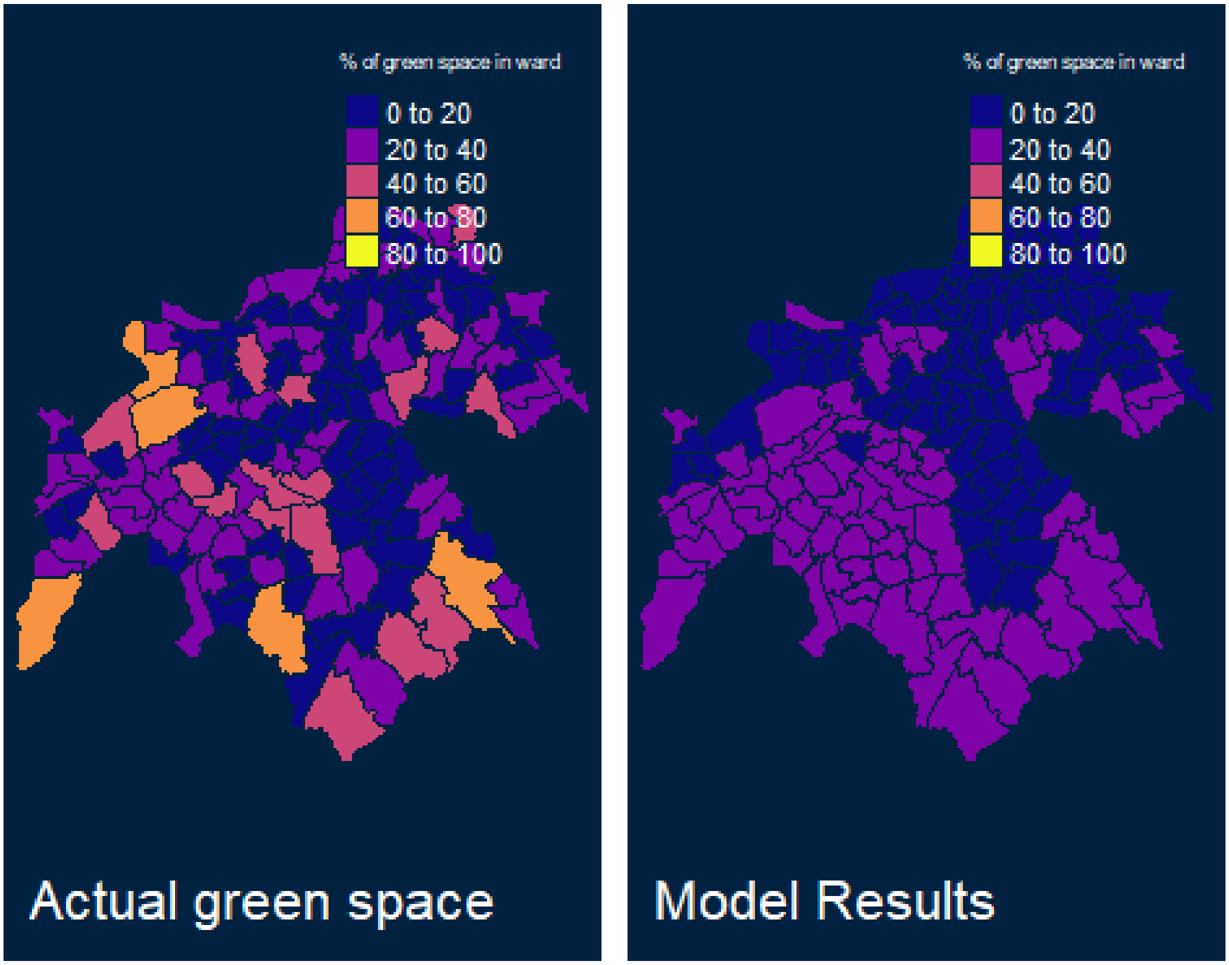

Figure 6.5: Fitted and observed values of % green space per ward in South London

It can be seen that these models are fairly competent for predicting the spatial distribution of access to green space, but underestimate values in the vast majority of cases. It can be concluded that the independent variables considered in this study are somewhat reliable predictors of access to green space, and therefore are closely related. The sign and magnitude of coefficients in the models can be interrogated to understand the relative influence different independent variables have on predicting access to green space in a ward. For example, in most of the models the number of Mixed ethnicity people living in the ward has less influence on predicting access to green space than the median age of the residents of that ward, as demonstrated by the relative magnitudes of their coefficients. The coefficient for the median age variable is positive in all models, suggesting wards with an older population have better access to green spaces. A number of the independent variables have coefficients of different signs in models of different areas, highlighting the importance of modelling at a local scale rather than assuming a pattern in one area will be replicated elsewhere. The requirement of different models for different areas emphasises the complexity of access to green space, further work may create a more robust model that fits data for all the study areas.

6.2 Limitations and future work

Significant multicollinearity between independent variables raises concerns for the results of this study. Regression analysis assumes that a change in the value of an independent variable should change the value of the dependent variable and not affect any other independent variables (Farrar and Glauber (1967)). However, correlation between independent variables (multicollinearity) reduces the precision of coefficient estimates, which in turn reduces the reliability of p-values and increases the difficulty of identifying statistically significant independent variables. It can also cause coefficient estimates to change dramatically if small changes are made to other independent variables in the model (Farrar and Glauber (1967)). Two approaches were considered to reduce multicollinearity. The first approach would be to combine highly correlated independent variables into larger categories; however, this was considered to be an oversimplification and inappropriate considering the nature of the variables involved.

The second approach would be to remove highly correlated variables. As the ‘wellbeing’ variable showed a high correlation with many other independent variables it was removed. The ‘white’ variable was also highly correlated with many other independent variables and had much higher values in each ward than other ethnic groups. Therefore, this variable was also removed. Future work could seek to further reduce correlations between independent variables or create a model unlikely to be influenced by multicollinearity.

A variety of indicators were available for evaluating model fit including log-likelihood, R^2, spatial tests on regression residuals, Lagrange-Multiplier tests, Alkaline Information Criterion, Bayesian Information Criterion and least squares. The Alkaline Information Criterion proved easiest for comparison between models as this indicator was available when using regression functions from the R packages ‘spdep,’ ‘spgwr’ and ‘spatialreg.’ However, other indicators showed different relative best fit between models. Future work could further evaluate the fit of each model in greater detail, and interrogate the merits of each measure of model fit in greater depth.

This study uses the aggregated spatial units of London wards (official subdivisions of London boroughs). These units were used because they are a well-understood subdivision of London and relevant to those working for London councils or planners, the key stakeholders in this study. Wards are the finest spatial units available for the London data required and are standardised across all of the data required for this study. Two problems are introduced due to the aggregation of spatial point data into areal units: the modifiable areal unit problem (MAUP) and the problem of ecological fallacy. The MAUP occurs because the results of statistical analysis can change significantly depending on the size, shape and position of the aggregated areal units used (Goodchild (2011)). The aggregated value will be generated using different point values depending on these three factors. The problem of ecological fallacy occurs when assumptions are made about individual locations or people within a ward based on the characteristics of that ward as a whole (Goodchild (2011)). Large differences in the size of areal units can lead to heteroskedasticity in the residuals from regression analysis. The weight matrix used for spatial regression analysis can also significantly affect the results, for example, this study uses a Queen's weight matrix. This matrix yields different results to a Rook's matrix. Further work could seek to evaluate the most appropriate matrix in greater depth.

There are a number of ways to measure access to green space, each with their own limitations. Perhaps the most reliable measure is to use walkable network distance from each individual home in an area to the nearest entrance to a green space (Greenspace Information for Greater London (2015)). A similar measure was used for a case study of Leicester, the measure was the same except for the use of centres of population census output areas in place of individual homes (Comber, Brunsdon, and Green (2008)). Many planning authorities simply use Euclidean distances from homes to green spaces to measure access (Moseley et al. (2013)). Other studies have used the percentage land use vs. population in an areal unit (Barbosa et al. (2007)). Our study first attempted to use the GIGL dataset for regression analysis, however, the residuals of regression analysis were extremely skewed. Normally distributed residuals are desired for regression analysis (Jarque and Bera (1987)), therefore this dataset was considered inappropriate for our study. Various other datasets and measures were tested for normality of regression residuals, including Euclidean distance to green space for and metric area of green space divided by population per ward. The only data to give normally distributed residuals was a measure of percentage area of green space per ward. Therefore, this data was used for our regression analysis despite the noted lack of nuance in the measure. The data does not consider access to green space across ward boundaries, network walking distances, equality of distribution of green space within a ward or quality of green space. Further work could produce or source a dataset appropriate for regression analysis that considers these important factors.

7. Appendix

7.1 Ordinary Least Squares (OLS) Results

	London	North	East	West	South
Residual Standard Error	0.4189	0.3848	0.4365	0.3804	0.3883
Multiple R-squared	0.1384	0.2322	0.21	0.3045	0.1451
Adjusted R-squared	0.1286	0.1899	0.1765	0.2722	0.1052
p-Value	2.20E-16	1.58E-05	1.59E-06	1.08E-09	0.001183

7.2 Spatial Lag Results

	London	North	East	West	South
Rho	0.40297	0.32206	0.29468	0.45669	-0.13798
LR test value	49.379	7.3763	7.0177	17.906	1.1217
p-Value	2.11E-12	0.0066088	0.0080709	2.32E-05	0.28955
Log Likelihood	-314.2866	-54.80725	-94.47798	-58.85549	-70.04624
AIC	648.57	129.61	208.96	137.71	160.09
AIC for OLS	695.95	134.99	213.97	153.62	159.21
Residual Autocorrelation-Test Value	0.25813	2.8071	0.50144	0.27107	1.5317
Residual Autocorrelation-p Value	0.61141	0.093845	0.45188	0.60262	0.21586

7.3 Spatial Error Results

	London	North	East	West	South
Lambda	0.41608	0.27217	0.27668	0.46645	-0.090527
LR test value	47.572	3.2666	5.3636	14.224	0.43599
p-Value	5.30E-12	0.070703	0.020562	0.00016232	0.50906
Log Likelihood	-315.1897	-56.86209	-95.30504	-70.38909	-70.38909
AIC	650.38	133.72	210.61	141.39	160.78
AIC for OLS	695.95	134.99	213.97	153.62	159.21

7.4 Spatial Durbin Results

	London	North	East	West	South
Rho	0.37479	0.16363	0.24079	0.36453	-0.1142
LR test value	41.266	1.4741	4.4874	9.7043	0.7311
p-Value	1.33E-10	0.22469	0.034146	0.0018383	0.39253
Log Likelihood	-302.6399	-48.56145	-89.76326	-54.65625	-65.65341
AIC	639.28	131.12	213.53	143.31	165.31
AIC for OLS	678.55	134.99	213.9736	151.02	159.21
Residual Autocorrelation-Test Value	1.9248	0.35131	0.18741	0.67823	1.8075
Residual Autocorrelation-p Value	0.16533	0.55337	0.66508	0.4102	0.17881

7.5 Geographically Weighted Regression Results

	London	North	East	West	South
Fixed Bandwidth	5823.581	4941.289	10334.82	4850.866	11140.59
AICc	642.0675	127.0033	215.8062	129.3623	159.2444
AIC	5.65E+02	98.46093	196.742	93.19053	144.7065
Quasi-global R2	0.3487077	0.4189579	0.2682174	0.5426054	0.1824
Moran's I	0.0403409	0.0175191	0.08093797	0.04163799	-0.04694458
p-Value	4.92E-06	0.0217	0.001437	0.001576	0.5284

7.5.1 Shapiro-Wilk Normality Tests before normalisation

Greens pace	London	North	East	West	South
W	0.93963	0.89477	0.96168	0.95177	0.93639
p-Value	3.12E-15	2.59E-08	0.0001102	2.72E-05	1.65E-06

7.5.2 Shapiro-Wilk Normality Test after normalisation

ln(green space + 12)	London	North	East	West	South
W	3.12E-15	0.98383	0.97935	0.98605	0.9928
p-Value	0.003074	0.1109	0.01124	0.1121	0.6184

7.5.3 Shapiro-Wilk Normality Test for OLS Residuals

	London	North	East	West	South
W	0.99453	0.99526	0.03343	0.97689	0.99129
p-Value	0.02431	0.9386	0.98304	0.00907	0.4481

7.6 OLS with Spatial Filtering Results

	London	North	East	West
Residual Standard Error	0.3671	0.3526	0.4047	0.3218
Multiple R-squared	0.3758	0.3959	0.3415	0.5384
Adjusted R-squared	0.3308	0.3197	0.2922	0.4791
p-value	2.20E-16	7.43E-08	5.23E-10	5.762e-16

		London	North	East	West	South
1	Observed Moran's I	-0.044942652	-0.044432333	-0.020959765	-0.067750899	-0.032591742
2	OLS	-0.0508352899	-0.076417741	-0.050377136	-0.082645384	-0.030405068
3	Expectation	0.3952	0.2542	0.2472	0.3631	0.5179
4	p-value	-0.044942652	-0.044432333	-0.020959765	-0.067750899	-0.032591742

Greater London

	Dependent variable		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	9.87	6.76 – 12.98	**<0.001**
age	0.03	0.02 – 0.04	**<0.001**
life_expectancy	-0.00	-0.02 – 0.02	0.765
income	-0.60	-0.87 – -0.32	**<0.001**
asian	-0.03	-0.08 – 0.01	0.158
black	-0.03	-0.10 – 0.04	0.455
other	-0.05	-0.11 – 0.01	0.079
mixed	-0.00	-0.00 – 0.00	0.398
fitted(spatial_filter)vec2	-1.57	-2.29 – -0.85	**<0.001**
fitted(spatial_filter)vec4	1.58	0.86 – 2.30	**<0.001**
fitted(spatial_filter)vec29	-1.35	-2.07 – -0.63	**<0.001**
fitted(spatial_filter)vec31	-1.19	-1.92 – -0.47	**0.001**
fitted(spatial_filter)vec27	-1.14	-1.86 – -0.42	**0.002**
fitted(spatial_filter)vec19	-1.06	-1.78 – -0.34	**0.004**
fitted(spatial_filter)vec16	-1.03	-1.75 – -0.31	**0.005**
fitted(spatial_filter)vec14	-0.92	-1.64 – -0.20	**0.013**
fitted(spatial_filter)vec58	-1.00	-1.72 – -0.28	**0.006**
fitted(spatial_filter)vec11	-0.78	-1.50 – -0.06	**0.033**
fitted(spatial_filter)vec60	0.96	0.24 – 1.68	**0.009**
fitted(spatial_filter)vec73	0.99	0.26 – 1.71	**0.007**
fitted(spatial_filter)vec28	-0.81	-1.53 – -0.09	**0.027**
fitted(spatial_filter)vec6	0.71	-0.01 – 1.43	0.054
fitted(spatial_filter)vec42	0.80	0.08 – 1.53	**0.029**
fitted(spatial_filter)vec52	0.79	0.07 – 1.51	**0.032**
fitted(spatial_filter)vec65	-0.83	-1.55 – -0.11	**0.024**
fitted(spatial_filter)vec20	0.67	-0.05 – 1.39	0.069
fitted(spatial_filter)vec66	0.80	0.08 – 1.52	**0.030**
fitted(spatial_filter)vec33	-0.70	-1.42 – 0.03	0.059
fitted(spatial_filter)vec71	0.81	0.09 – 1.53	**0.028**
fitted(spatial_filter)vec41	0.70	-0.02 – 1.43	0.055
fitted(spatial_filter)vec126	1.03	0.31 – 1.75	**0.005**
fitted(spatial_filter)vec43	0.69	-0.03 – 1.41	0.062
fitted(spatial_filter)vec21	0.60	-0.12 – 1.32	0.103
fitted(spatial_filter)vec77	0.74	0.02 – 1.46	**0.043**
fitted(spatial_filter)vec102	0.78	0.06 – 1.50	**0.034**
fitted(spatial_filter)vec175	1.17	0.45 – 1.89	**0.001**
fitted(spatial_filter)vec94	-0.74	-1.46 – -0.02	**0.044**
fitted(spatial_filter)vec54	0.63	-0.09 – 1.35	0.087
fitted(spatial_filter)vec112	-0.81	-1.53 – -0.09	**0.028**
fitted(spatial_filter)vec86	0.69	-0.03 – 1.42	0.059
fitted(spatial_filter)vec8	-0.51	-1.23 – 0.21	0.166
fitted(spatial_filter)vec141	-0.87	-1.59 – -0.15	**0.018**
fitted(spatial_filter)vec114	0.64	-0.08 – 1.37	0.080
Observations	625		
R^2 / R^2 adjusted	0.376 / 0.331		

West London

	Dependent variable		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	20.83	14.70 – 26.97	**<0.001**
age	0.05	0.03 – 0.07	**<0.001**
life_expectancy	-0.02	-0.05 – 0.02	0.325
income	-1.48	-2.09 – -0.86	**<0.001**
asian	0.01	-0.09 – 0.12	0.790
black	-0.33	-0.49 – -0.17	**<0.001**
other	-0.01	-0.15 – 0.13	0.878
mixed	0.00	-0.00 – 0.00	0.147
fitted(w_spatial_filter)vec2	1.48	0.84 – 2.11	**<0.001**
fitted(w_spatial_filter)vec11	-1.03	-1.66 – -0.39	**0.002**
fitted(w_spatial_filter)vec4	0.78	0.15 – 1.42	**0.016**
fitted(w_spatial_filter)vec23	1.03	0.39 – 1.67	**0.002**
fitted(w_spatial_filter)vec8	0.63	-0.01 – 1.26	0.053
fitted(w_spatial_filter)vec28	0.82	0.18 – 1.46	**0.012**
fitted(w_spatial_filter)vec24	0.72	0.08 – 1.35	**0.028**
fitted(w_spatial_filter)vec13	-0.53	-1.16 – 0.11	0.104
fitted(w_spatial_filter)vec3	0.41	-0.22 – 1.05	0.201
fitted(w_spatial_filter)vec1	-0.37	-1.00 – 0.27	0.257
fitted(w_spatial_filter)vec45	0.52	-0.11 – 1.16	0.107
Observations	159		
R^2 / R^2 adjusted	0.538 / 0.479		

North London

	Dependent variable		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-3.78	-13.68 – 6.13	0.452
age	0.07	0.05 – 0.10	**<0.001**
life_expectancy	0.05	0.01 – 0.09	**0.017**
income	-0.12	-0.87 – 0.63	0.757
asian	-0.00	-0.16 – 0.16	0.992
black	0.42	0.17 – 0.67	**0.001**
other	-0.06	-0.21 – 0.08	0.386
mixed	-0.00	-0.00 – 0.00	0.051
fitted(n_spatial_filter)vec2	0.77	0.07 – 1.46	**0.032**
fitted(n_spatial_filter)vec6	-0.61	-1.31 – 0.09	0.085
fitted(n_spatial_filter)vec1	0.55	-0.15 – 1.25	0.122
fitted(n_spatial_filter)vec18	0.71	0.01 – 1.41	**0.047**
fitted(n_spatial_filter)vec34	1.00	0.30 – 1.70	**0.005**
fitted(n_spatial_filter)vec19	0.63	-0.06 – 1.33	0.075
fitted(n_spatial_filter)vec3	-0.46	-1.15 – 0.24	0.199
fitted(n_spatial_filter)vec47	0.80	0.10 – 1.50	**0.026**
Observations	135		
R^2 / R^2 adjusted	0.396 / 0.320		

East London

	Dependent variable		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	7.70	-1.75 – 17.16	0.109
age	0.01	-0.02 – 0.04	0.441
life_expectancy	-0.03	-0.07 – 0.02	0.242
income	-0.11	-0.98 – 0.77	0.809
asian	-0.17	-0.31 – -0.02	**0.023**
black	0.02	-0.15 – 0.20	0.792
other	-0.09	-0.29 – 0.11	0.396
mixed	0.00	-0.00 – 0.00	0.065
fitted(e_spatial_filter)vec8	1.15	0.35 – 1.95	**0.005**
fitted(e_spatial_filter)vec9	-1.16	-1.96 – -0.36	**0.005**
fitted(e_spatial_filter)vec13	-1.13	-1.93 – -0.33	**0.006**
fitted(e_spatial_filter)vec17	-1.09	-1.89 – -0.29	**0.008**
fitted(e_spatial_filter)vec1	0.34	-0.46 – 1.14	0.403
Observations	173		
R^2 / R^2 adjusted	0.342 / 0.292		

7.6.1 Lagrange Multiplier Diagnostics for Spatial Dependence

	London	North	East	West	South
LMlag	56.262	7.5719	7.2718	1.85E+01	0.96829
p-Value (LMlag)	6.34E-14	0.005928	0.007005	1.73E-05	0.3251
LMerr	52.139	2.5679	5.2787	13.204	0.36006
p-Value (LMerr)	5.17E-13	0.1091	0.02159	0.0002793	0.5485
RLMlag	4.1866	11.009	3.1031	6.2264	2.7989
p-Value (RLMlag)	0.04074	0.0009069	0.07814	0.01259	0.09433
RLMerr	0.063807	6.0046	1.11	0.96381	2.1907
p-Value (RLMerr)	0.8006	0.01427	0.2921	0.3262	0.1388

References

Astell-Burt, Thomas, Xiaoqi Feng, Suzanne Mavoa, Hannah M Badland, and Billie Giles-Corti. 2014. “Do Low-Income Neighbourhoods Have the Least Green Space? A Cross-Sectional Study of Australia’s Most Populous Cities.” *BMC Public Health* 14 (1): 292. https://doi.org/10.1186/1471-2458-14-292.

Barbosa, Olga, Jamie A. Tratalos, Paul R. Armsworth, Richard G. Davies, Richard A. Fuller, Pat Johnson, and Kevin J. Gaston. 2007. “Who Benefits from Access to Green Space? A Case Study from Sheffield, UK.” *Landscape and Urban Planning* 83 (2): 187–95. https://doi.org/10.1016/j.landurbplan.2007.04.004.

Bivand, Roger. 2007. “The Spdep Package: Comprehensive R Archive Network. Version 0.7-8.” Comprehensive R Archive Network. http://ftp.uni-bayreuth.de/math/statlib/R/CRAN/doc/packages/spdep.pdf.

Brunsdon, Chris. 2008. “Geographical Statistics and the Grid: Grid-Enabled Modelling of Local and Global Geographies.”

Chi, Guangqing, and Jun Zhu. 2008. “Spatial Regression Models for Demographic Analysis.” *Population Research and Policy Review* 27 (1): 17–42. https://doi.org/10.1007/s11113-007-9051-8.

Choi, Junyoung, Myunghwa Hwang, Gayeon Kim, Janghwan Seong, and Jaeseong Ahn. 2016. “Supporting the Measurement of the United Nations’ Sustainable Development Goal 11 Through the Use of National Urban Information Systems and Open Geospatial Technologies: A Case Study of South Korea.” *Open Geospatial Data, Software and Standards* 1 (1): 4. https://doi.org/10.1186/s40965-016-0005-0.

Comber, Alexis, Chris Brunsdon, and Edmund Green. 2008. “Using a GIS-Based Network Analysis to Determine Urban Greenspace Accessibility for Different Ethnic and Religious Groups.” *Landscape and Urban Planning* 86 (1): 103–14. https://doi.org/10.1016/j.landurbplan.2008.01.002.

Dai, Dajun. 2011. “Racial/Ethnic and Socioeconomic Disparities in Urban Green Space Accessibility: Where to Intervene?” *Landscape and Urban Planning* 102 (4): 234–44. https://doi.org/10.1016/j.landurbplan.2011.05.002.

Daoud, Jamal I. 2017. “Multicollinearity and Regression Analysis.” *Journal of Physics: Conference Series* 949 (December): 012009. https://doi.org/10.1088/1742-6596/949/1/012009.

De Ridder, K., V. Adamec, A. Bañuelos, M. Bruse, M. Bürger, O. Damsgaard, J. Dufek, et al. 2004. “An Integrated Methodology to Assess the Benefits of Urban Green Space.” *Science of The Total Environment*, Highway and Urban Pollution, 334-335 (December): 489–97. https://doi.org/10.1016/j.scitotenv.2004.04.054.

Douglas, Owen, Mick Lennon, and Mark Scott. 2017. “Green Space Benefits for Health and Well-Being: A Life-Course Approach for Urban Planning, Design and Management.” *Cities* 66 (June): 53–62. https://doi.org/10.1016/j.cities.2017.03.011.

Fabozzi, Frank J., Sergio M. Focardi, Svetlozar T. Rachev, and Bala G. Arshanapalli. 2014. *The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications*. John Wiley & Sons.

Farrar, Donald E., and Robert R. Glauber. 1967. “Multicollinearity in Regression Analysis: The Problem Revisited.” *The Review of Economics and Statistics* 49 (1): 92–107. https://doi.org/10.2307/1937887.

Getis, Arthur, and Daniel A. Griffith. 2002. “Comparative Spatial Filtering in Regression Analysis.” *Geographical Analysis* 34 (2): 130–40. https://doi.org/10.1111/j.1538-4632.2002.tb01080.x.

Goodchild, Michael F. 2011. “Scale in GIS: An Overview.” *Geomorphology*, Scale Issues in Geomorphology, 130 (1): 5–9. https://doi.org/10.1016/j.geomorph.2010.10.004.

Greater London CIC (GiGL), Greenspace Information for. 2015. “Access to Public Open Space and Nature by Ward.” 2015. https://data.london.gov.uk/dataset/access-public-open-space-and-nature-ward.

Greenspace Information for Greater London, London Datastore. 2015. “Access to Public Open Space and Nature by Ward.” https://data.london.gov.uk/download/access-public-open-space-and-nature-ward/18fc0bef-68af-42c1-b77c-d6babbb66465/access-public-open-space-ward.xls.

Griffith, Daniel A., and Jean H. P. Paelinck. 2011. *Non-Standard Spatial Statistics and Spatial Econometrics*. Advances in Geographic Information Science. Heidelberg: Springer.

Haining, Robert P. 2003. *Spatial Data Analysis: Theory and Practice*. Cambridge, UK ; New York: Cambridge University Press.

Heynen, Nik, Harold A. Perkins, and Parama Roy. 2006. “The Political Ecology of Uneven Urban Green Space: The Impact of Political Economy on Race and Ethnicity in Producing Environmental Inequality in Milwaukee.” *Urban Affairs Review* 42 (1): 3–25. https://doi.org/10.1177/1078087406290729.

Houlden, Victoria, João Porto de Albuquerque, Scott Weich, and Stephen Jarvis. 2019. “A Spatial Analysis of Proximate Greenspace and Mental Wellbeing in London.” *Applied Geography* 109 (August): 102036. https://doi.org/10.1016/j.apgeog.2019.102036.

Hur, Misun, Jack L. Nasar, and Bumseok Chun. 2010. “Neighborhood Satisfaction, Physical and Perceived Naturalness and Openness.” *Journal of Environmental Psychology* 30 (1): 52–59. https://doi.org/10.1016/j.jenvp.2009.05.005.

Jarque, Carlos M., and Anil K. Bera. 1987. “A Test for Normality of Observations and Regression Residuals.” *International Statistical Review / Revue Internationale de Statistique* 55 (2): 163–72. https://doi.org/10.2307/1403192.

London, Mayor of. 2017. “The London Plan.” *London City Hall*. https://www.london.gov.uk//what-we-do/planning/london-plan/current-london-plan/london-plan-2016-pdf.

Ministry of Housing, Communities & Local Government (MHCLG). 2005. “Land Use by Borough and Ward.” 2005. https://data.london.gov.uk/dataset/land-use-ward.

Mitchell, Richard, and Frank Popham. 2008. “Effect of Exposure to Natural Environment on Health Inequalities: An Observational Population Study.” *The Lancet* 372 (9650): 1655–60. https://doi.org/10.1016/S0140-6736(08)61689-X.

Moseley, Darren, Mariella Marzano, Jordan Chetcuti, and Kevin Watts. 2013. “Green Networks for People: Application of a Functional Approach to Support the Planning and Management of Greenspace.” *Landscape and Urban Planning* 116 (August): 1–12. https://doi.org/10.1016/j.landurbplan.2013.04.004.

National Statistics, Office for. 2011. “2011 Census.” 2011. https://data.london.gov.uk/census/.

Shapiro, S. S., and M. B. Wilk. 1965. “An Analysis of Variance Test for Normality (Complete Samples).” *Biometrika* 52 (3/4): 591–611. https://doi.org/10.2307/2333709.

Smith, D. 2010. “Valuing Housing and Green Spaces: Understanding Local Amenities, the Built Environment and House Prices in London.”

Taylor, Lucy, and Dieter F. Hochuli. 2017. “Defining Greenspace: Multiple Uses Across Multiple Disciplines.” *Landscape and Urban Planning* 158 (February): 25–38. https://doi.org/10.1016/j.landurbplan.2016.09.024.

Zhang, Yang, Agnes E. Van den Berg, Terry Van Dijk, and Gerd Weitkamp. 2017. “Quality over Quantity: Contribution of Urban Green Space to Neighborhood Satisfaction.” *International Journal of Environmental Research and Public Health* 14 (5): 535. https://doi.org/10.3390/ijerph14050535.

Zhou, Xiaolu, and Masud Parves Rana. 2012. “Social Benefits of Urban Green Space : A Conceptual Framework of Valuation and Accessibility Measurements.” *Management* *of Environmental Quality: An International Journal* 23 (2): 173–89. https://doi.org/10.1108/14777831211204921.