CASA0006 Assessment 3
Student ID: 21197822

Association of Greenspace Type and Access with Personal Wellbeing

Reproducible Analysis via Github: https://github.com/sunny-netizen/Quant_A3.git

Introduction & Literature Review

Greenspace is an important consideration for urban planning. A review of literature finds that demands for greenspace are under-met for most cities and that a diversity of greenspace types provides the greatest benefit; and that greenspace should be 'accessible, available and adaptable' (Boulton et al., 2019). Other studies have evaluated the impact of greenspace on personal wellbeing. Houlden et al. (2019) found that the amount of greenspace within 300m of residents in Greater London was significantly associated with satisfaction and worth but not happiness, three wellbeing metrics from the Annual Population Survey by the UK's Office for National Statistics, using multiple linear regression models to account for potential confounders from census data. They further detected that the associations had similar patterns of geographic variation (were spatially autocorrelated) with surprising negative associations in some areas and suggest additional factors, such as individual preference, walking distance, and greenspace size, type, accessibility, or use, could be at play (Houlden et al., 2019). Inspired by this research, my project will explore if some of these suggested greenspace factors influence local variations in personal wellbeing, providing insight for urban planning.

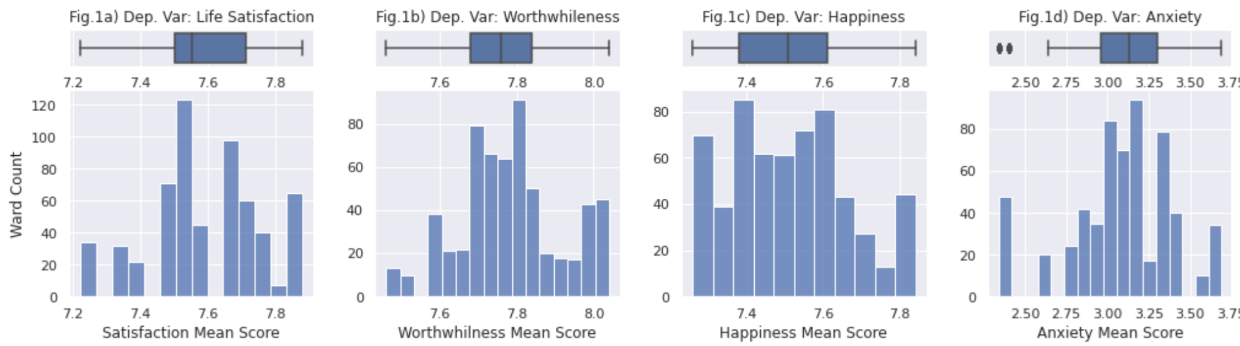
Research Question & Hypothesis

My research question is what factors relating to greenspace type and access impact personal wellbeing in London on the local scale and how can this be quantified? My alternate research hypothesis is that at least one of the chosen metrics of greenspace types are associated with wellbeing indicators, whereas my null research hypothesis is that none of these metrics are associated.

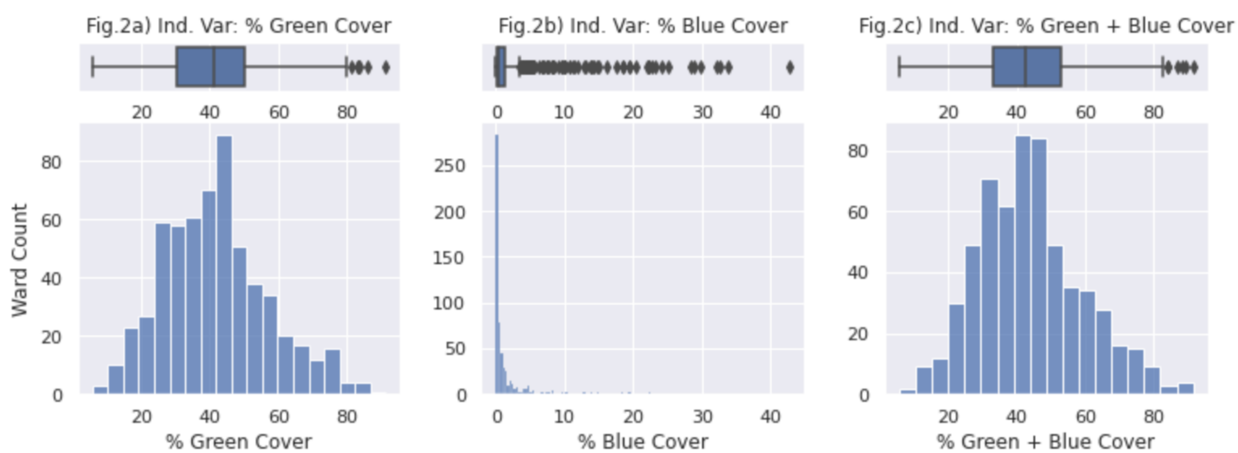
Since the eight independent variables may not all be mutually exclusive in the greenspace coverage they represent, such as with percent greenspace and types of parks, the benefit of multiple regression in this study is to assess the explanatory power of each metric of greenspace type, including discerning if any are confounding, rather than creating models of complete explanatory power. My alternate hypothesis suggests all metrics being significant, non-collinear, and non-confounding, rather than directly referring to a high amount of variance explained by the model (high R-squared value).

Presentation of Data

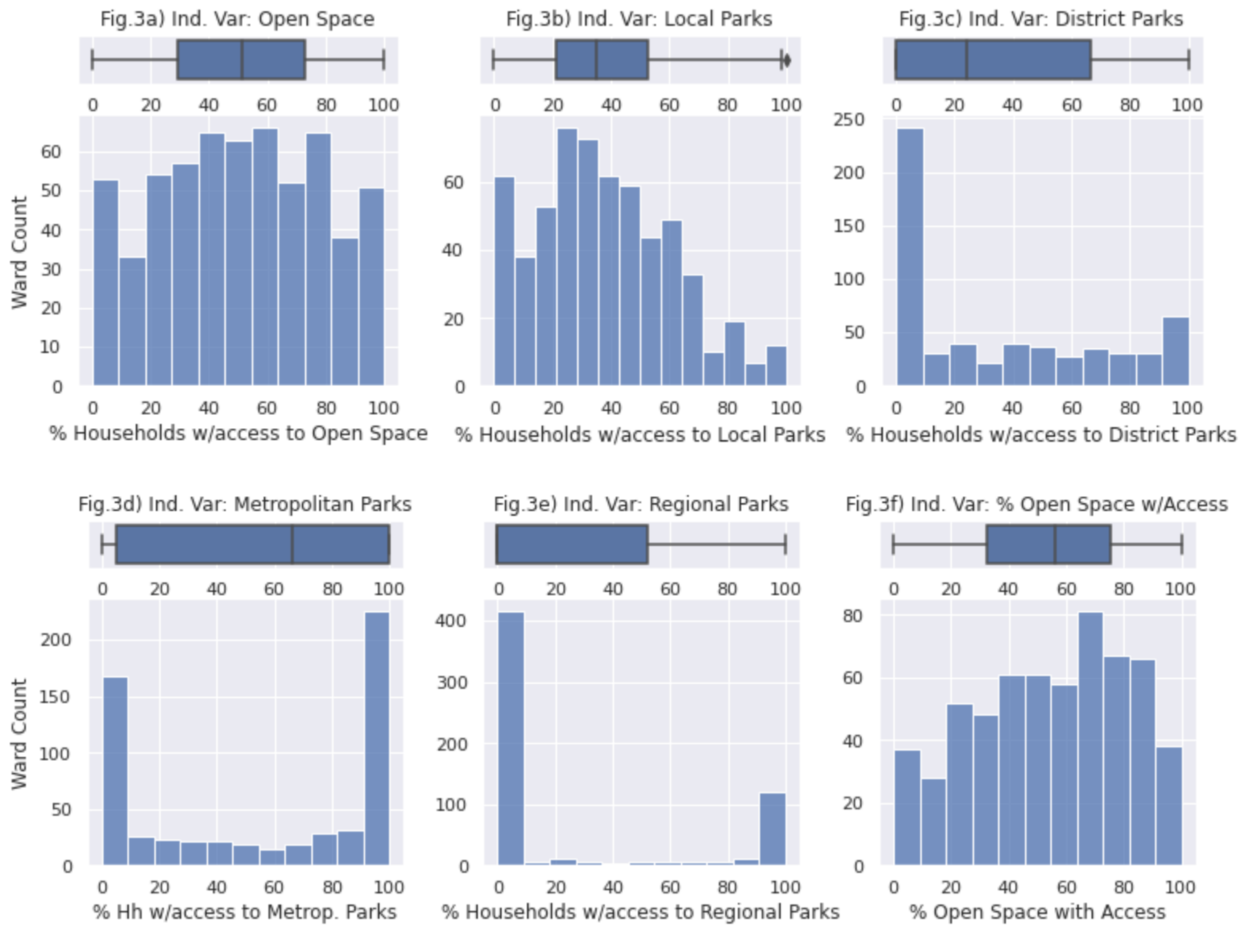
Paralleling Houlden et al. (2019), I used life satisfaction, worth, and happiness as my dependent variables and analyzed them in separate multiple regression models (Fig.1abc). For additional perspective, I also included anxiety as a dependent variable (Fig.1d). I also did not account for confounding variables from census data as they did, but I suggest this step for future research. Their study looked at the area individual residents instead of bounded spatial units, but my data lent towards using Greater London wards as a spatial unit (Houlden et al., 2019). Since the personal wellbeing scores (scoring 1-10) only had borough-level data, I extrapolated each mean borough score to each of its wards. I excluded the City of London from my analysis since not all data included it.



I selected my independent variables from on Houlden et al. (2019)'s suggestion of greenspace type as an additional factor associated with wellbeing. I could not ascertain if the greenspace Houlden et al. (2019) used from Greenspace Information for Greater London CIC, 2017 included water, and a review study by Britton et al. (2020) mentions how bluespace (water) may oftentimes be lumped with greenspace. Therefore, I interrogate green cover and blue cover as separate types, looking at percent green cover, percent blue cover, and their combined percentages in each ward determined by the Greater London authority from near-infrared aerial imagery (NDVI) and land use datasets (Fig.2) (GLA GIS Team, 2019).



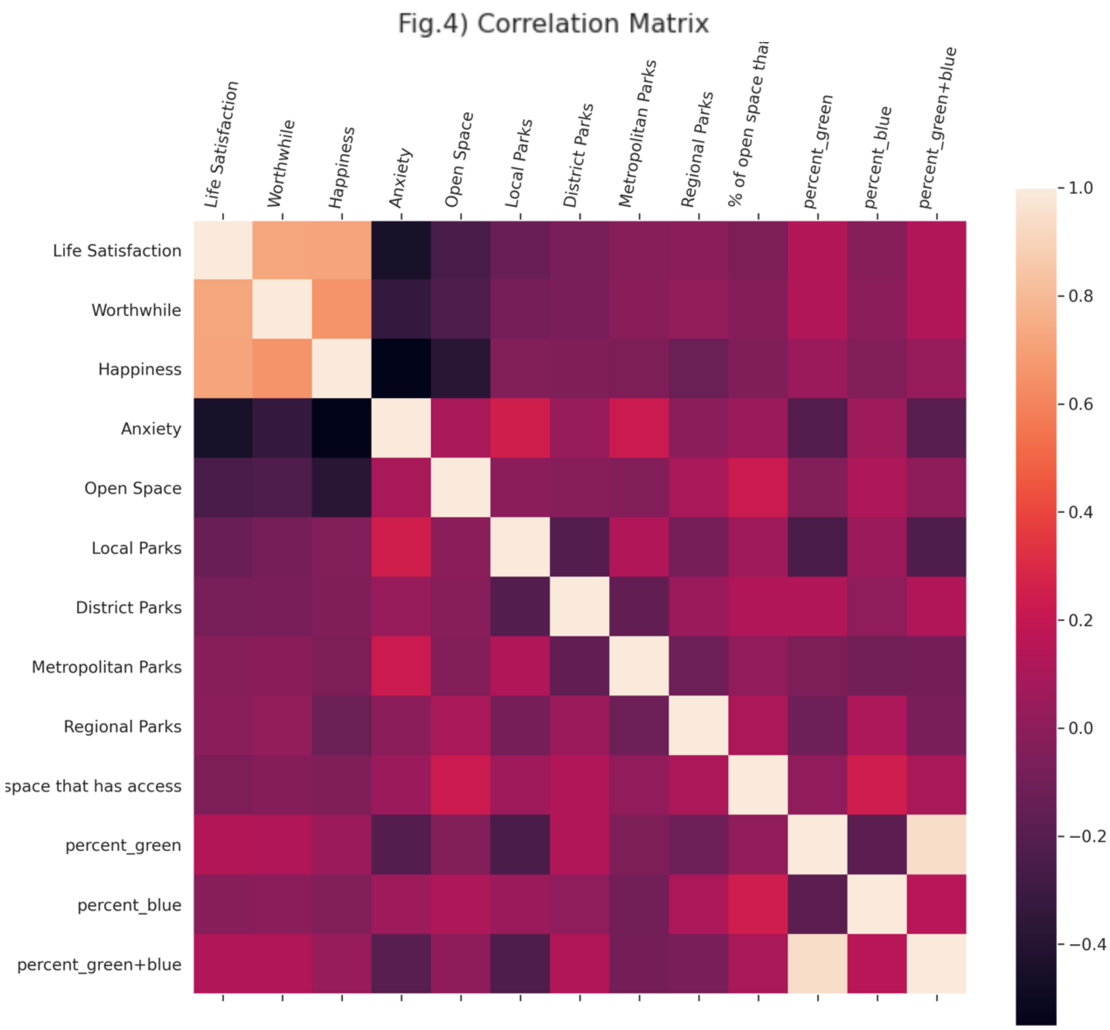
I also looked at the percentage of households with access to different types of public open spaces. These categories include open space, local parks, districts parks, metropolitan parks, and regional parks per ward, as well as the percent of open space with access in each ward, as defined by the London Plan 2011 (Fig.3) (GiGL, 2014). The data source defines access as the recommended walking distance from households to greenspace types (GiGL, 2014). A limitation of this data is that it is from 2013 prior to boundary changes for three wards in 2014, which may not accurately compare with other independent variables that represent wards after 2014.



Methods & Results

Upon evaluating the distributions of the variables, I decide to not transform any variables. After cleaning the variables and dropping NaNs, the sample size represented 600 wards. The three metrics of personal wellbeing had close to normal distributions. I retained the two outliers in the anxiety data because they did not lay far from Tukey's fences. Some of the independent variables, such as percent access to local parks and district parks, were more skewed, but log transform generated negative infinity values that invalidated too many samples (Fig.3bc). Percent access to metropolitan parks and regional parks appeared bimodal, but I also did not transform them (Fig.3de). Percent blue had the strongest skew; log-transformation normalizes its distribution but decreases the R-squared value of multiple regression models, so I do not transform it (Fig.2b).

Prior to multiple regression, I evaluated multicollinearity in the variables with correlation matrices and Variation Inflation Factor (VIF). The correlation matrix shows a strong correlation (< 0.937) between percent green & blue and percent green, reflecting how the greater amount of greenspace in greater London overwhelms the combined value (Fig.5). VIF on the independent variables unsurprisingly dropped percent green & blue. Moderately correlated pairs include percent green & blue and percent blue, anxiety and local parks, and anxiety and metropolitan parks (> 0.200). The positive dependent wellbeing variables of life satisfaction, worthwhileness, and happiness are highly correlated between each other (> 0.600), whereas anxiety is strongly uncorrelated with the three, especially happiness (0.348).



Across all personal wellbeing metrics, the multiple regression models were significant (Prob. F-statistic <0.050) but only accounted for only 5 – 20% of residuals (low R-squared values) (Table 1). Statistically significant coefficients (p-value < 0.05) have very small values (|coefficient| < 0.01). Adjusted R-squared values were only slightly lower, suggesting that the number of predictors was not an issue.

Table 1: Multiple Regression Summary Results

Summary Statistics	**Life Satisfaction**	**Worthwhile**	**Happiness**	**Anxiety**
R-sq	0.107	0.085	0.167	0.155
Adj. R-sq	0.095	0.072	0.156	0.143
F-statistic	8.824	6.792	14.76	13.43

<table>
<tr><td>Prob(F-statistic)
Signif <0.05</td><td>1.94e-11</td><td>1.52e-08</td><td>9.12e-20</td><td>6.21e-18</td></tr>
<tr><td>Statistics for significant coefficients where P>|t|
Signif <0.05

Reported values are Coef
&
95% CI
[0.025, 0.975]</td><td>Open Space
-0.0016
[-0.002, -0.001]

Local Parks
-0.0009
[-0.002, -0.000]

District Parks
-0.0006
[-0.001, -0.000]

%Green&Blue
0.0013
[0.000, 0.002]</td><td>Open Space
-0.0012
[-0.002, -0.001]

District Parks
-0.0004
[-0.001, -0.000]

%Green&Blue
0.0011
[0.000, 0.002]</td><td>Open Space
-0.0022
[-0.003, -0.002]

District Parks
-0.0004
[-0.001, -4.08e-05]

Metropolitan Parks
-0.0003
[-0.001, -2.61e-05]

Regional Parks
-0.0004
[-0.001, -0.000]</td><td>Open Space
0.0012
[0.000, 0.002]

Local Parks
0.0030
[0.002, 0.004]

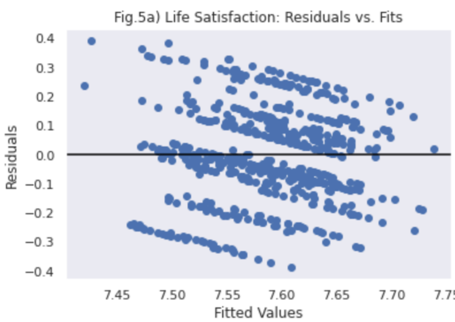
District Parks
0.0013
[0.001, 0.002]

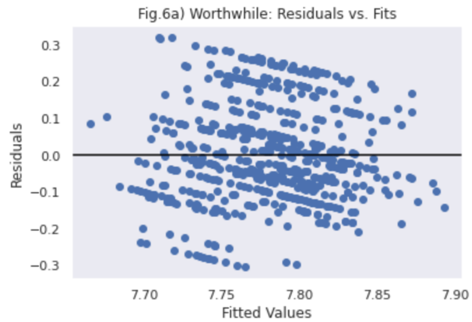
Metropolitan Parks
0.0017
[0.001, 0.002]

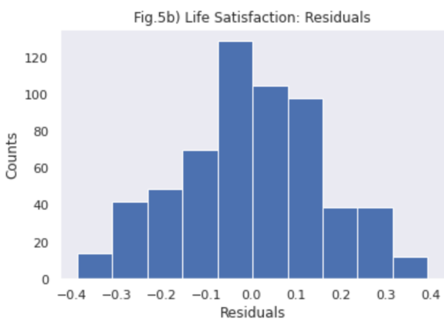
%Green&Blue
-0.0031
[-0.005, -0.001]

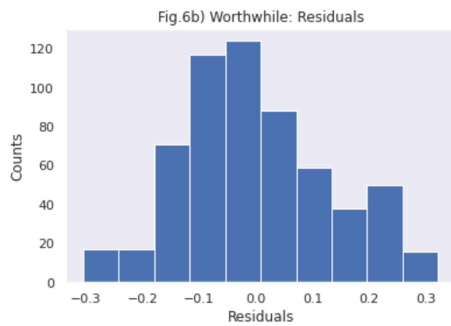
%Blue
0.0052
[0.000, 0.010]</td></tr>
</table>

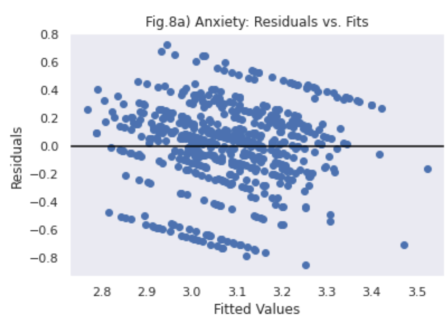
From visual inspection, all four models relate to assumptions for multiple linear regression similarly. They all exhibit diagonal bands in their residual vs. fits plots due to extrapolating borough data of personal wellbeing scores to wards. The assumption of linearity is met; the Residuals vs. Fits plot shows the mean of residuals plausibly stays close to zero for most fitted values (Fig.5a-8a). Due to the shape created by the diagonal bands and their cutoffs, smaller x-values have a fanning effect, and larger x-values have a funneling effect, but the since the effects are subtle, this study considers the assumption of homoscedascity met (Penn. State U., 2018, Glen_b, 2016**).** The assumption of normally distributed residuals is also met, although anxiety residuals have a slight left skew (Fig.5b-8b). Otherwise, there are no outliers, and without investigating spatial autocorrelation, the data meets the assumption for independent errors, given its random sampling.

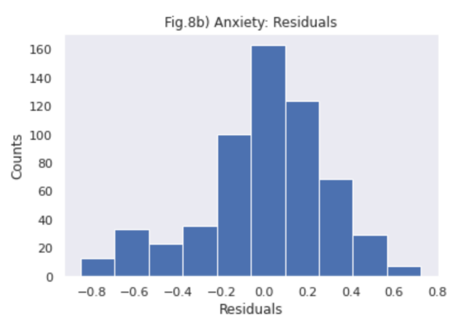












Discussion

In answer to my research question, what factors relating to greenspace type and access impact personal wellbeing in London on the local scale, I have found several variables associated with life satisfaction, worthwhileness, happiness, and anxiety, as all four of these models had significant F-statistics and some terms with statistically significant coefficients, allowing me to reject the null hypothesis that none of the greenspace type variables are associated with personal wellbeing. However, these results are in the context of low R-squared values for each model, which is not unexpected, since this study did not address other confounding factors. However, it highlights how greenspace type accounts for less than 20% of personal wellbeing measures.

Given the imprecision of these models and low magnitudes of the coefficients, I interpret significant coefficients with skepticism. Based on confidence intervals, and if all other independent variables are fixed, the greatest estimated magnitude of rate of change is 0.004 in percent households with access to a park type and 0.005 for percent blue and/or green cover for one unit of change in wellbeing scores. This means a very small amount of cover or a fraction of a household with increased park access (for an average ward population of 5500) is associated with a one-point score change (ONS, 2016). This seems unreasonably drastic. Interestingly, access to most park types is inversely related with positive wellbeing metrics and directly related the negative wellbeing metric, anxiety. On the other hand, the composite percent green and blue coverage have a positive relationship with life satisfaction and happiness and the opposite for anxiety, whereas just percent blue coverage is directly related to anxiety. Percent of open space with access was the only variable that did not have a significant coefficient in any model.

Conclusion & Limitations

Although satisfaction, worthwhileness, and happiness seem to have greater association similarities than anxiety, the final picture of how factors relate to greenspace type and access remains inconclusive for urban planning, primarily for the low R-squared values and small coefficients, and secondarily for the mixed messages of the coefficients. Houlden et al. (2019) notes that studies based on administrative boundaries had mixed results for associating greenspace and wellbeing. The wards-based approach in this study may have similar outcomes. Perhaps many types of parks are detrimental to wellbeing, or perhaps only the composite coverage of green and blue space adequately reflects the positive benefits of nature, as many studies have shown previously. Or perhaps urban waterbodies are related to anxiety.

Future studies should address limitations in this study. Houlden et al. (2019) consider greenspace near individual residents rather than composition of greenspace in spatial units because greenspaces beyond spatial boundaries may influence residents within. This approach could be applied to greenspace types as well. Another major issue with my study is standardization. Since the personal wellbeing metrics were mean score per borough (extrapolated to wards), I maintained independent variables as percentages per ward to maintain the meaning of mean wellbeing scores. However, future studies may standardize data to be proportional among wards. I suggest dividing percent green and/or percent blue coverage by ward area, wellbeing mean scores by ward population above age 16 (matching survey population), and percent access to park types by number of ward residences. I also suggest adding a population weighting when extrapolating mean score borough data to wards. Finally, future studies may examine all independent variables and dependent variables with k-means or hierarchical cluster analysis to explore subgroups or patterning within the data to draw additional insights on greenspace types and personal wellbeing across wards.

Word Count: 1733

Bibliography

Boulton, C., Dedekorkut-Howes, A. and Byrne, J. (2018). 'Factors shaping urban greenspace provision: A systematic review of the literature'. *Landscape and Urban Planning*, 178, pp. 82–101. doi: 10.1016/j.landurbplan.2018.05.029.

Britton, E., Kindermann, G., Domegan, C. and Carlin, C. (2020). 'Blue care: a systematic review of blue space interventions for health and wellbeing'. *Health Promotion International*, 35 (1), pp. 50–69. doi: 10.1093/heapro/day103.

Glen_b. (2016). 'scatterplot - Does this residual plot indicate heteroscedasticity?' *Cross Validated*. Available at: https://stats.stackexchange.com/questions/206999/does-this-residual-plot-indicate-heteroscedasticity (Accessed: 18 January 2022).

Greater London Authority (GLA) GIS Team. (2019). *London Green and Blue Cover*. London Datastore. Available at: https://data.london.gov.uk/dataset/green-and-blue-cover (Accessed: 13 January 2022).

Greenspace Information for Greater London CIC (GiGL). (2014). *Access to Public Open Space and Nature by Ward*. London Datastore. Available at: https://data.london.gov.uk/dataset/access-public-open-space-and-nature-ward (Accessed: 16 January 2022).

Houlden, V., Porto de Albuquerque, J., Weich, S. and Jarvis, S. (2019). 'A spatial analysis of proximate greenspace and mental wellbeing in London'. *Applied Geography*, 109, p. 102036. doi: 10.1016/j.apgeog.2019.102036.

Office for National Statistics (ONS). (2016). *[ARCHIVED CONTENT] Electoral wards/divisions*. *The National Archives*. Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20160108131156/http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/administrative/england/electoral-wards-divisions/index.html (Accessed: 18 January 2022).

Personal well-being frequently asked questions - Office for National Statistics. (no date). Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/perso nalwellbeingfrequentlyaskedquestions#what-is-measuring-national-well-being-about (Accessed: 15 January 2022).

The Pennsylvania State University. (2018). *4.4 - Identifying Specific Problems Using Residual Plots | STAT 462*. Available at: https://online.stat.psu.edu/stat462/node/120/ (Accessed: 18 January 2022).

Smith, Martha K. (2014). *Common Mistakes in Interpretation of Regression Coefficients*. Available at: https://web.ma.utexas.edu/users/mks/statmistakes/regressioncoeffs.html (Accessed: 18 January 2022).