Introduction

- 33 London Boroughs (including City of London as a borough)

- □ Hypothesis 1 (H1) : There is no correlation between environmental factors and health
- □ Hypothesis 2 (H2) : There is no correlation between socioeconomic factors and health
- □ Null hypothesis 1 (NH1) : There is a correlation between environmental factors and health
- □ Null hypothesis 2 (NH2) : There is a correlation between socio-economic factors and health

Inspired by the model of analysing ecological regressions in *Natário and Knorr-Held's (2003)* study which analyses changes in 'exposure variables' quantified at an area unit level to predict risks, further akin to *Congdon's (2008)* similar study on London Boroughs, we developed a simplified structural equation model (SEM) to analyse the external factors that impact London boroughs.

Structural equation models (SEMs) show connections among data through a general framework while using latent variables, where latent variables are variables which cannot be directly analysed and require deduction through mathematical models from other observed variables *(Kaplan, 2004)*.

Making use of inductive reasoning, we explore the relationship among factors of London Boroughs, on an environmental, socio-economic and health level, in accordance with the documented Literature Review. Our SEM uses Borough Scores on health, socio-economic and environmental factors as latent variables.

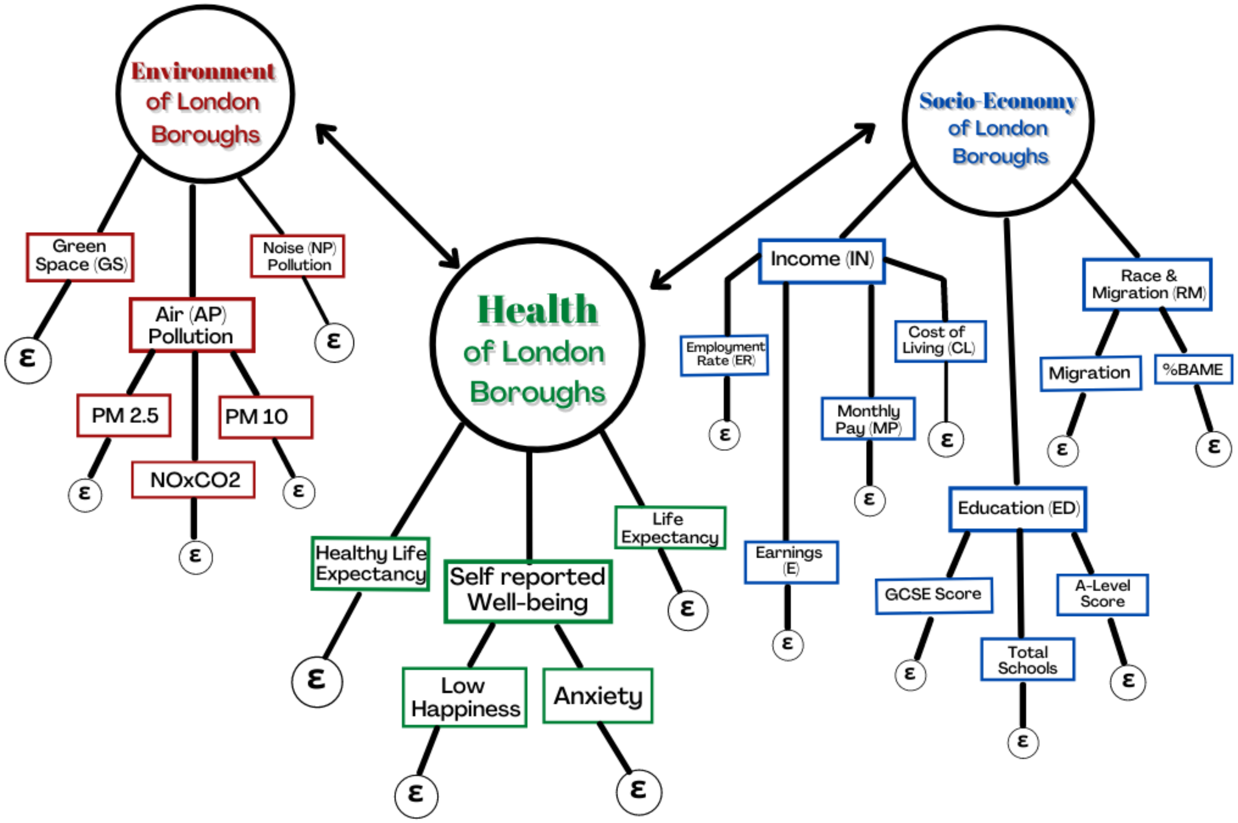


Figure X: *Structural Equation Model* (SME) *of London Boroughs' Environmental, Health and Socio-Economic factors* (modelled manually in *Canva (2021)*, a design platform that provided open source elements to manually complete and design the SME diagram.

Hot Deck imputation is a statistical method for missing data processing where compromised or non-existent data is replaced with observed responses from *"similar"* units *(Andridge and Little, 2010).* The core of the method is to replace the missing values with plausible data for further analysis not to generate fully accurate missing data replacement *(Little, Rubin and Bayes, 2002).* To replace missing or incorrect data, we used techniques similar to Hot Deck imputation, assuming that most data of London boroughs will turn out similarly if we estimate based on the missing boroughs' neighbours' data or by approximating data that was 2-4 years apart.

Making sense of observed patterns is the core driving force of SMEs and theories *(Babbie, 2007).* By proving or disproving the hypotheses, by the end of our study we will have observed patterns in the health status of London boroughs and predict future outcomes and strategies for improving their overall health, by identifying what affects it.

Process

The process to corroborate or disprove our hypotheses, incorporating additional visualisation, will be as follows:

- First, we will calculate an overall Borough Score for each factor we are going to study (health, BS_H, socio-economic, BS_{SE}, and environmental factors, BS_E). The scores will be obtained through the processing of related variables and indicators, giving each of them a sub-score and then adding or subtracting these, depending on the formalised equation. All of this is done using Microsoft Excel. Each variable's data will be processed differently according to its individual characteristics and will be explained further in the following page.

- Then we will plot choropleth maps for each of the three Borough Scores, through the usage of a *London Datastore (2022)* template of London boroughs. This is because choropleth maps have historically been impactful in cartography analysis based on their effectiveness, being a powerful tool of exploratory visualisation *(Brewer et. al, 1997)*. Indeed, these visualisations will be helpful in clearly demonstrating how health, environmental, and socioeconomic outcomes vary across London. This will help us better identify disparities.

- We will further plot heatmaps and pair-plots on Python using the Seaborn library, to show how individual socioeconomic and environmental variables are correlated with health outcomes *(Waskom, a, 2022; Waskom, b, 2022).* These visualisations will also be useful later, in the final step of our methodology, to see how individual socioeconomic variables are correlated with each other, and how individual environmental variables are correlated with one another.

- To visualise the correlation between BSh and BSe, and the correlation between BSh and BSse, simple linear regressions will be plotted on Python *(Waskom, c, 2022).*

- Lastly, we will conduct multi-linear regressions (MLRs) taking into account the combined effect of multiple socioeconomic variables, and of multiple environmental variables on health (BSh). Based on previous calculations and mapping for single variables, the aim of conducting a multi-linear regression is to generate a holistic analysis of the overall correlation between different independent variables on a dependent variable *(Pennsylvania State University, a, 2022).* However, before MLRs may be carried out, it is important to test for multicollinearity, to ensure that none of the independent variables are strongly correlated with one another, which may affect the accuracy of results *(Pennsylvania State University, b, 2022)*. If two independent variables are strongly correlated, it is recommended to remove one variable *(Pennsylvania*

State University, c, 2022). To test for multicollinearity, we will use the heatmaps and pairplots plotted beforehand.

- The results from the two multi-linear regressions will be compared, in order to determine which of the two independent variables (socioeconomic or environmental) has a bigger impact on the health of London citizens.

Variable	**Description**
ε	Error Margin
BSH	Borough Score Health
BSE	Borough Score on Environment
BSSE	Borough Score on Socio-Economic

BSE **=** GS + NP + AP + ε

Variable	**Description**
BSE	Borough Score on Environment
AP	Air Pollution
NP	Noise Pollution
GS	Green Space

BSH **=** LH + A + LE + HLE + ε

Variable	**Description**
BSH	Borough Score Health
A	Anxiety
LH	Low Happiness
LE	Life Expectancy
HLE	Healthy Life Expectancy

BSSE = IN + ED + RM + ε

Variable	Description
BSSE	Borough Score on Socio-Economic
IN	Income
ED	Education
RM	Race & Migration
ER	Employment Rate
CL	Cost of Living
MP	Monthly Pay
TS	Total Schools
ALVL	A-level attainment
GCSE	GCSE attentiment
BAME	Black, Asian, and minority ethnic % score
M	Migration Rate Score

Variable	Calculation
BSH	LH + A + LE + HLE + ε
BSE	GS + NP + AP + ε
BSSE	IN + ED + RM + ε

Calculation of the Borough score for environmental factors (BSE)

In order to calculate the Borough Score for environmental factors, we will use the following variables for each borough:

- Green space
- Noise pollution
- Air pollution (NOx CO2, PM2.5, PM10)

After the data processing of each variable, we will obtain scores between 0 and 1 for each variable (in the case of air pollution, the scores will be individual for each greenhouse gas). To achieve these values, we will normalise the data using equation 1, the data normalisation formula for linear scaling *(Google Developers, 2021).*

$$X - Xmin/Xmax - Xmin$$

Equation 1: Data normalisation formula *(Google Developers, 2021)*

We will use Microsoft Excel to obtain these scores, as well as to plot the clustered bar charts for each score and create boxplots to visualise the distribution of BSe for all boroughs

Air Pollution - (AP)

The data contains the CO2, NOx, PM10 and PM2.5 emissions in tonnes per year per borough, for the year 2016 *(London Datastore, a, 2021).* To obtain the air pollution score we will process the data:

1. First we divide the emissions per borough between the surface of each borough, (as a bigger surface will emit more pollutants), to get the emissions per square km.

2. We normalise the emissions per km2 (using equation 1) so the data is between 0 and 1.

After ordering the boroughs from highest to lowest score, we can observe that City of London is the borough with the most emissions overall, followed by Hillingdon. Regarding individual values for each pollutant, for CO2 and NOx, Hillingdon is the most polluting borough, while for PM2.5 and PM10 it is City of London. The borough with the least emissions is Haringey, which scored 0 for all four pollutants through normalisation.

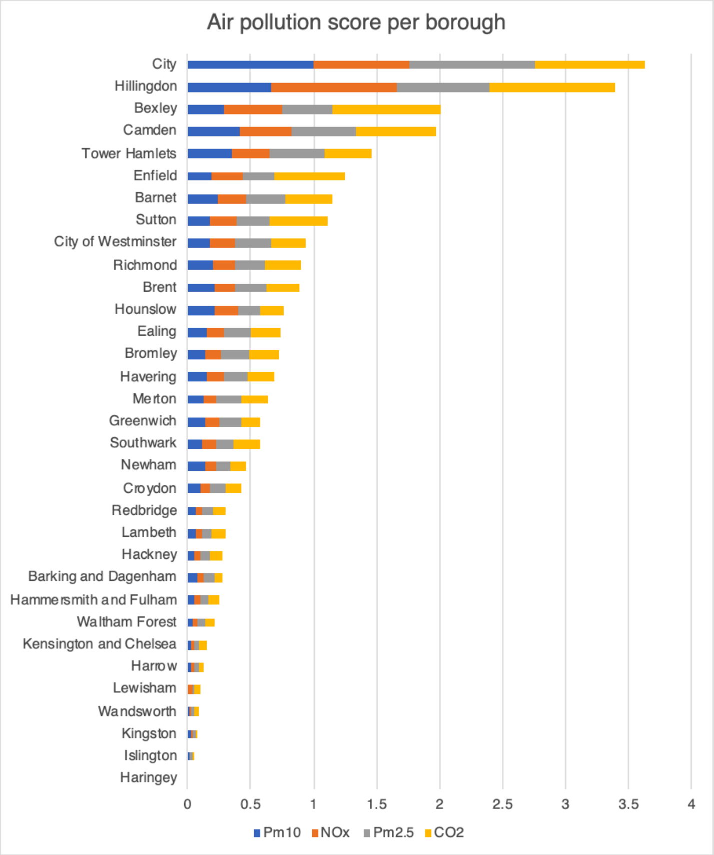


Figure 2: Air pollution score per London borough in 2016.

Greenspace - (GS)

The data used to obtain the green space score consists of the km2 of public green space each borough has *(London Datastore, c, 2021).* To obtain the green space score we will process the data:

1. First we will divide the area of green space each borough has between the total area of the borough, to get the area of green space per km2.

2. We normalise the area of green space per km2 (using equation 1) so the data is between 0 and 1.

3. In this particular case, we will invert the data, as in our study the higher the score the least environmentally friendly a borough is. However, through the previous calculations, boroughs with a high score have a bigger area of green space per km2, therefore being more environmentally friendly. To invert the data, we will apply (1-x) to all the scores obtained.

After ordering the boroughs from highest to lowest score, we can observe that the City of London is the borough with the least area of greenspace per km2, followed by Islington. The borough with the most area of greenspace per km2 is Havering.

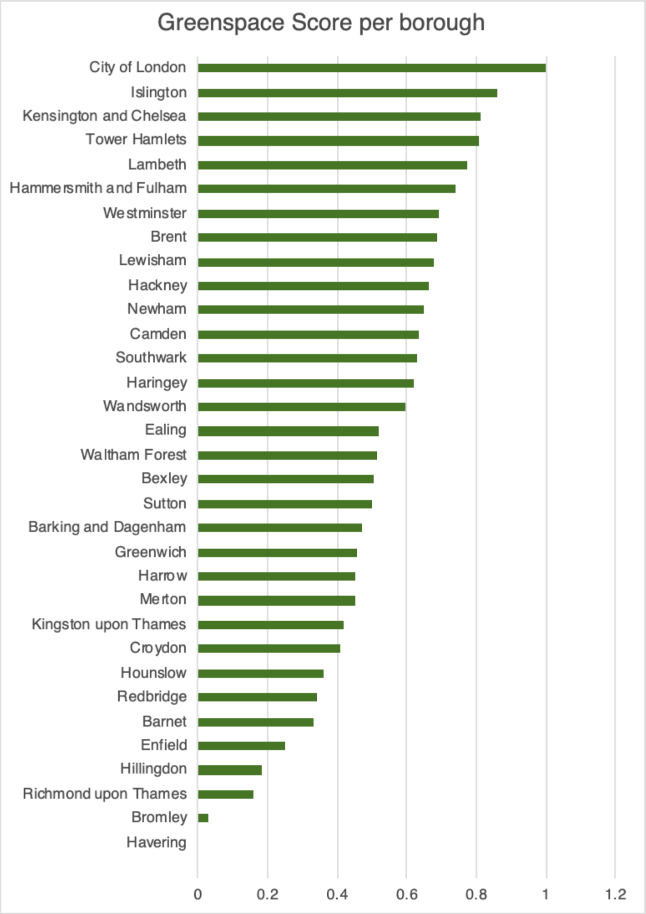


Figure 3: Greenspace score per London borough.

Noise pollution - (NP)

The data contains the number of complaints about noise per thousand population per borough, for the year 2019 *(London Datastore, d, 2021)*. To obtain the noise pollution score we will process the data:

1. First we will have to generate the data for the boroughs of Hackney and City of London, as in the chosen dataset the data on noise complaints for these two boroughs was merged. To generate them we will use the surface of each borough and the total noise complaints:

Total noise complaints/Total surface x Hackney surface = City noise complaints per thousand population.

Total noise complaints/Total surface x City surface = City noise complaints per thousand population.

Note: The limitations of this data will be further discussed in the limitations section.

2. Then, we normalise the number of noise complaints per thousand population for each borough (using equation 1) hence the data is between 0 and 1.

After ordering the boroughs from highest to lowest score, we can observe that Westminster is the borough with the most noise complaints per thousand population, followed by Kensington and Chelsea. The borough with the lowest noise pollution score is City of London.

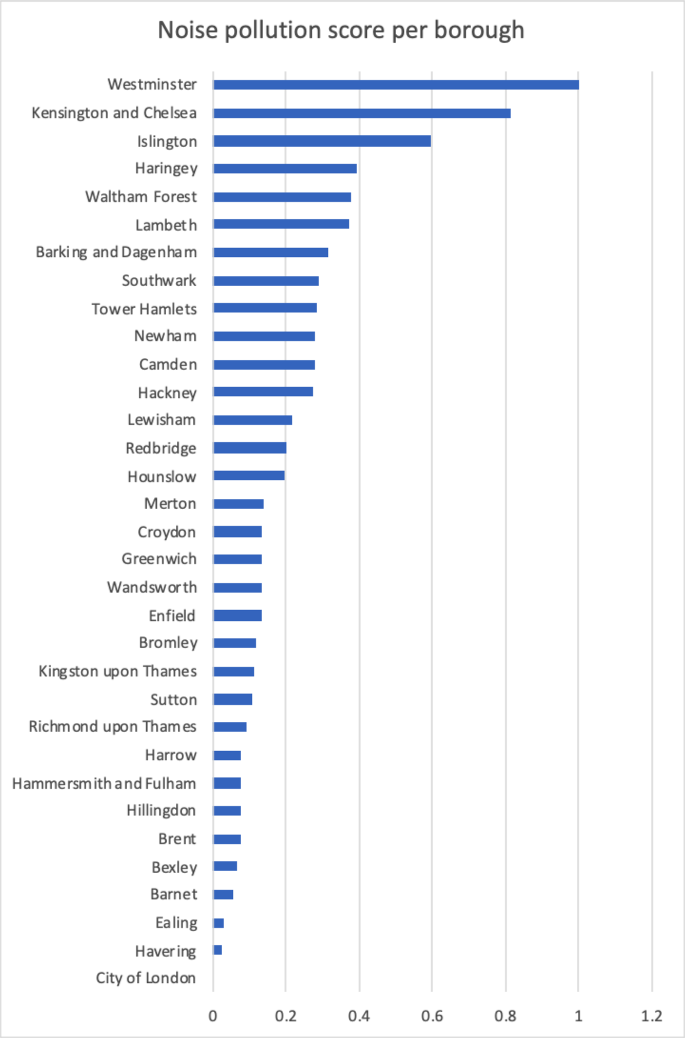
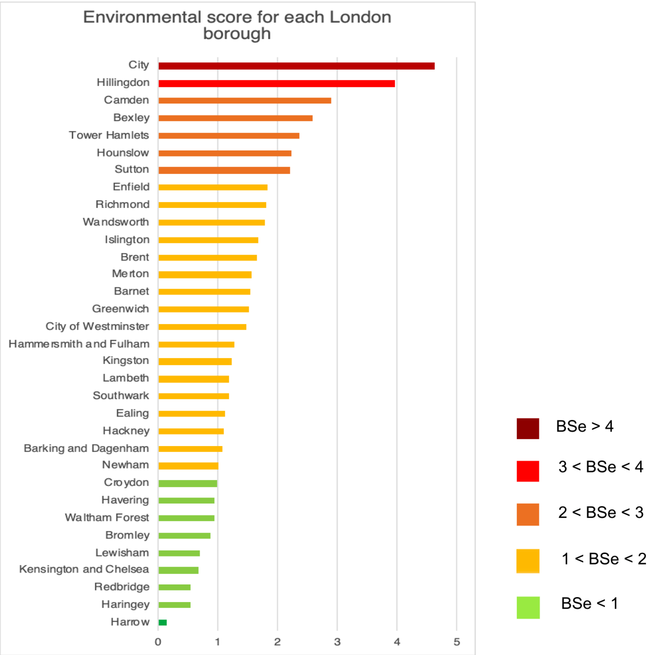


Figure 4: Noise pollution score per London borough in 2019.

TOTAL BSE

In conclusion, for the Borough Score on environmental factors, we can conclude that City of London is the least environmentally friendly borough, with the biggest scores on emissions and lack of greenspace, followed by Hillingdon. When we represent this data in a box plot, only these two boroughs stand out from the data as outliers over the higher tukey fence.



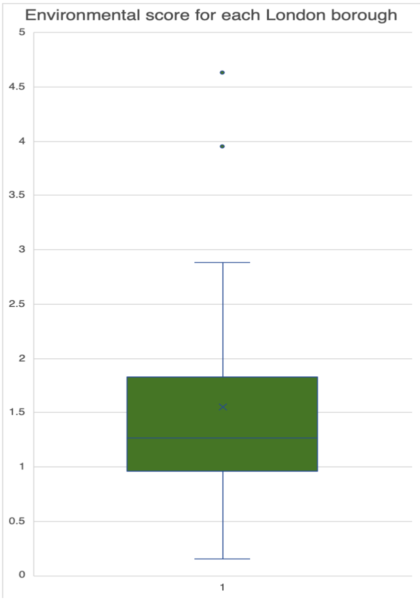


Figure 5: Borough Scores for environmental factors for all London boroughs.

Figure 6: Boxplot with borough Scores for environmental factors for all London boroughs.

Table 4: Boxplot values		BSe
Boxplot Statistics	Minimum	0.15
	Maximum	2.88
	Lower quartile	0.96
	Upper quartile	1.82
	Interquartile range	1.26
	Lo. outlier limit	0.15
	Hi. outlier limit	2.88
Tukey fences	Lo. Tukey fence (LQ - 1.5 * IQR)	- 0.93
	Hi. Tukey fence (UQ + 1.5 * IQR)	3.71
Outliers	City of London and Hillingdon	

Calculation of the Borough score for health factors (BSH)

In order to calculate the Borough Score for environmental factors, we will use the following variables for each borough:

- Self reported well-being (anxiety).
- Self reported well-being (low happiness).
- Life expectancy.
- Healthy life expectancy.

After the data processing of each variable, we will obtain scores between 0 and 1 for each variable. To achieve these values, we will normalise using the following equation:

X-Xmin/Xmax-Xmin

Equation 1: Data normalisation formula.

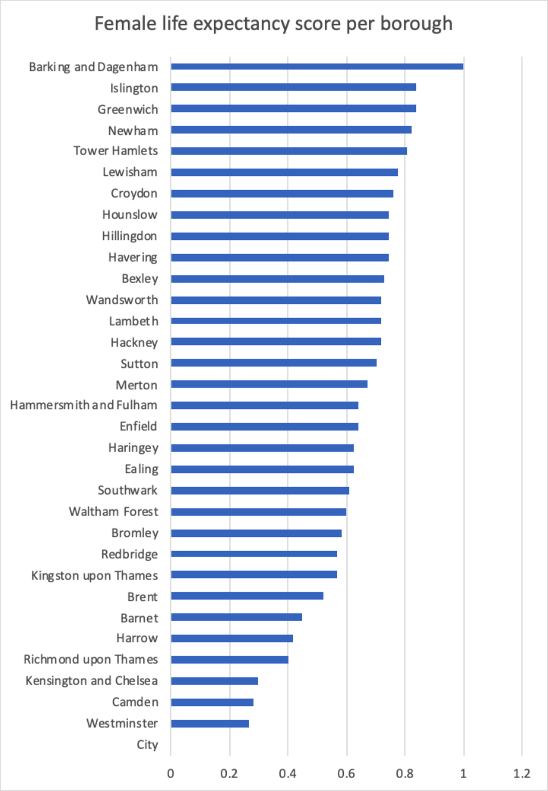
We will use Microsoft Excel to obtain these scores, as well as to plot the clustered bar charts for each score and create boxplots to visualise the distribution of BSh for all boroughs.

Life expectancy

The data contains the life expectancy for males and females separately per borough *(Trust for London, a, 2021)*. This dataset was missing a value for the City of London, so we added the value from the *City of London Health Profile (Public Health England, 2017).* To obtain the life expectancy score we will:

1. Normalise the data separately for males and females using equation 1.

2. Invert the data, as in our study the higher the score the least healthy a borough is. However, without the inversion, the high values correspond to the boroughs with a higher life expectancy. To correct this, we will perform (1-x) to each value, so that boroughs with a higher life expectancy have a lower score.



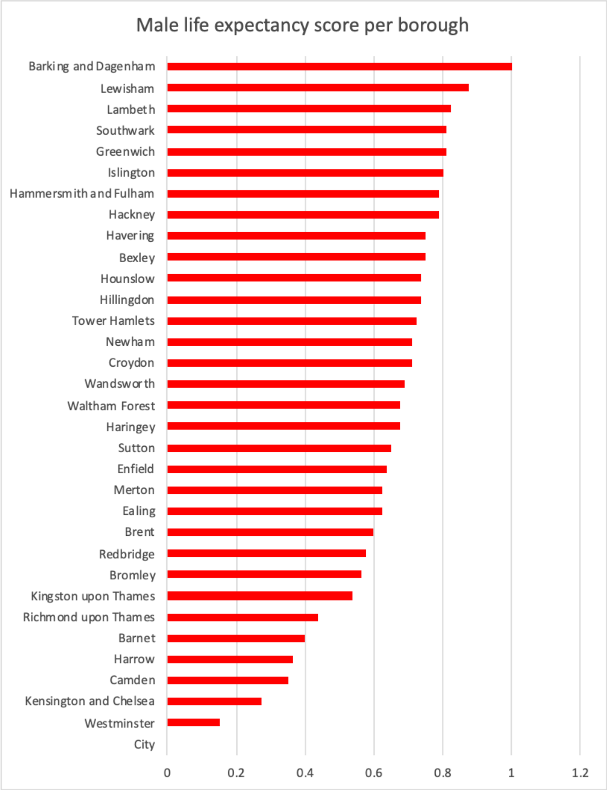


Figure 7: Female Life expectancy score per London borough.
Figure 8: Male Life expectancy score per London borough.

After the normalisation, we can observe that the borough with the highest life expectancy score is Barking and Dagenham, for both males and females. However, the second highest score belongs to different boroughs depending on sex; Islington for females and Lewisham for males. For both sexes, the two boroughs with the lowest score, and thus highest life expectancy, are Westminster and City of London

Healthy life expectancy - (HLE)

The data contains the healthy life expectancy (HLE) for males and females separately per borough *(Trust for London, a, 2021).* This dataset was missing a value for the City of London, so we calculated it in the following way:

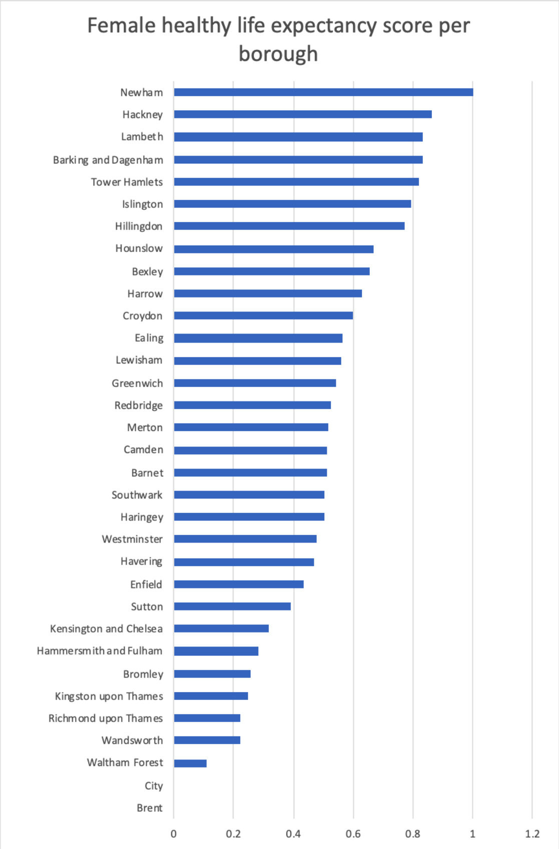
In order to estimate the City of London's HLE we will use two values, the life expectancy and the percentage of residents with very good health for the City of London. The latter value was obtained from the *2011 Census: Health and Disability in London Report (London Datastore, i, 2021).* As we observed in our previous analysis on life expectancy, the City of London was the borough with the highest life expectancy. Regarding the percentage of healthy residents, City of London is the borough with the highest percentage, with 88% of residents deemed healthy. Considering that the City of London is the borough with both the healthiest residents and the residents with the biggest life expectancy, we can estimate that it will be the borough with the highest HLE as well.

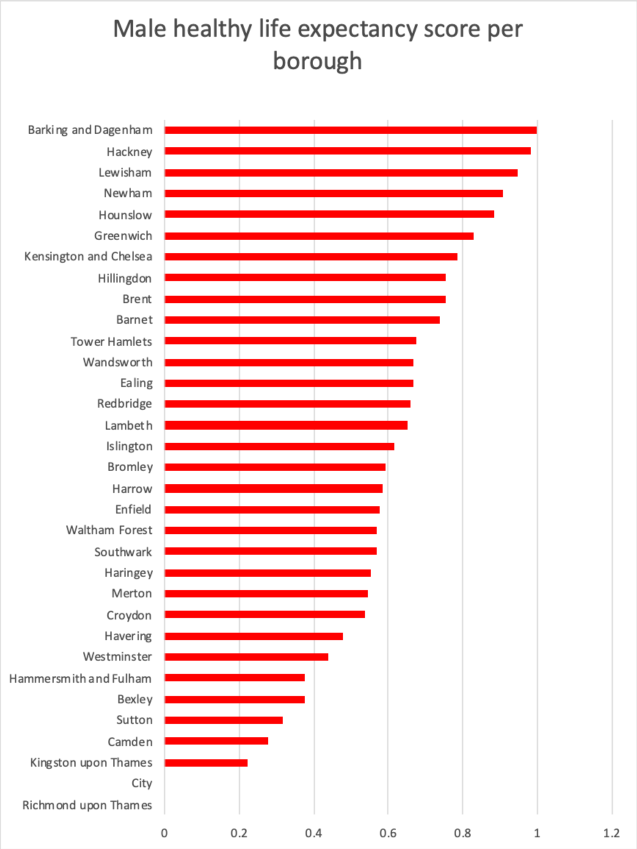
Thanks to the normalisation process, we do not have to award this borough a particular age for the HLE value, but it will be sufficient to award it with the 0 value after the process of normalisation. However, because of this, it will not be part of the normalisation process, so another borough will be the Xmin and therefore have a score of 0 as well.

After we have all the values, we will:

1. Normalise the data separately for males and females using equation 1.

2. Inverse the data, as in our study the higher the score the least healthy a borough is; (as explained in the Life expectancy section).

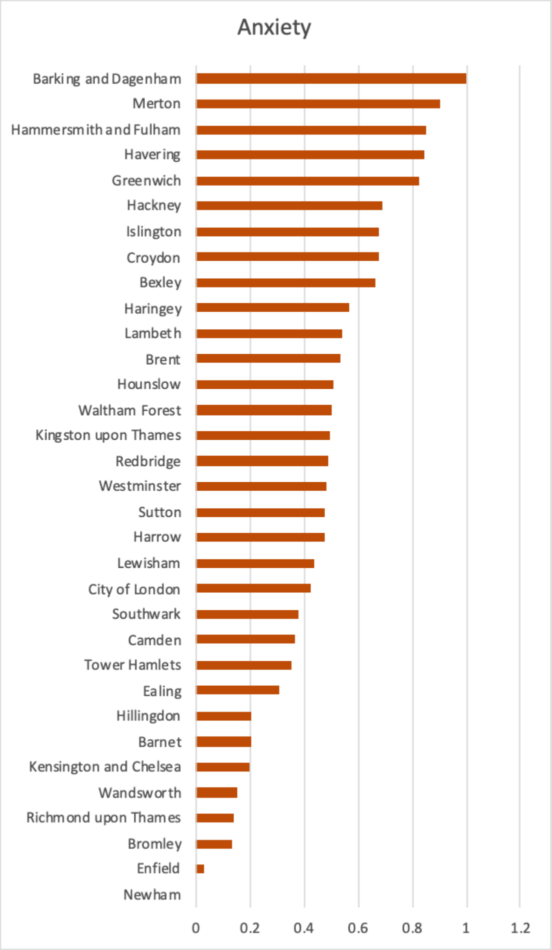


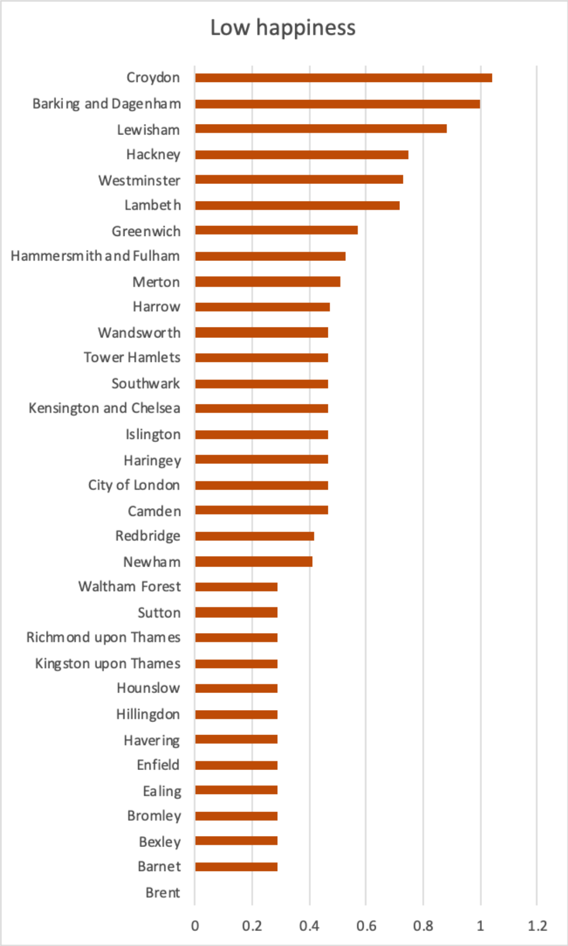


The boroughs with the lowest healthy life expectancy and therefore highest scores in this category are Newham for females and Barking and Dagenham for males, followed by Hackney for both sexes. For each sex, there are two boroughs with the lowest score: City (as we previously explained) and Brent for females and Richmond for males.

Self reported well-being - Anxiety (A) and Low Happiness (LH)

The data contains the overall self reported well-being in relation to high anxiety and low happiness for each London borough for the year 2015 *(London Datastore, d, 2021).* To obtain the self reported well-being scores we will simply normalise the data separately for males and females using equation 1.

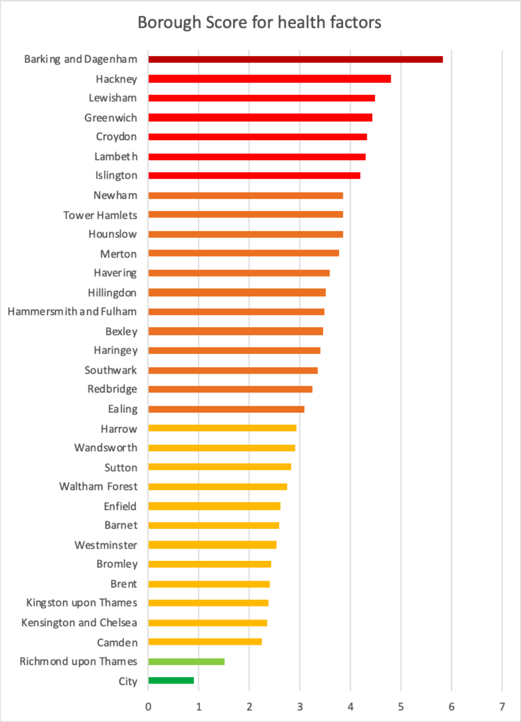




The boroughs with the lowest healthy life expectancy and therefore highest scores in this category are Newham for females and Barking and Dagenham for males, followed by Hackney for both sexes. For each sex, there are two boroughs with the lowest score: City (as we previously explained) and Brent for females and Richmond for males.

TOTAL BSH

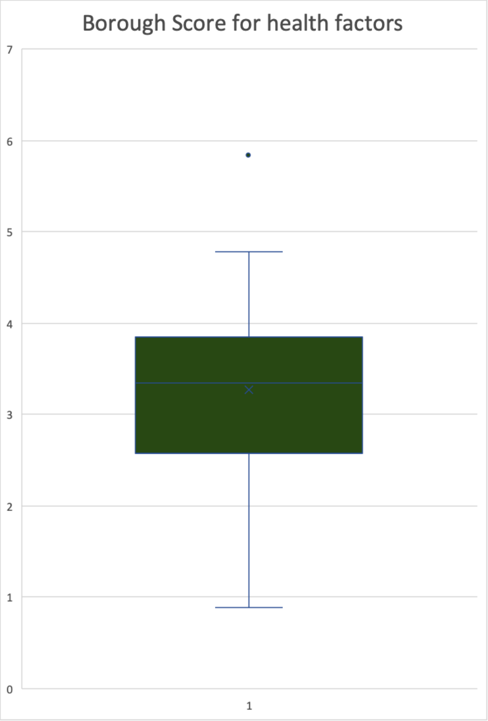
After obtaining the scores for each variable, to calculate the Borough Score for health factors we simply add up all the individual scores for each borough:



Borough Score for Health (BSh)

🟥	BSh > 5
🟥	4 < BSh < 5
🟧	3 < BSh < 4
🟨	2 < BSh < 3
🟩	BSh < 2

Table 3: Boxplot values		BSe
Boxplot Statistics	Minimum	0.89
	Maximum	5.83
	Lower quartile	2.56
	Upper quartile	3.85
	Interquartile range	3.27
	Lo. outlier limit	0.89
	Hi. outlier limit	4.78
Tukey fences	Lo. Tukey fence (LQ - 1.5 * IQR)	- 2.34
	Hi. Tukey fence (UQ + 1.5 * IQR)	8.75
Outlier	Barking and Dagenham	



The borough with the highest Borough score for health factors, and therefore the borough with the least overall health is Barking and Dagenham, followed by Hackney and Lewisham. On the other hand, the borough with the lowest score and highest health is the City of London, followed by Richmond and Thames.

Calculation of the Borough score for socio-economic factors (BSSE)

In order to calculate the Borough Score for socio-economic factors, we will use the following variables for each borough:

- Income - (IN)
- Education - (ED)
- Migration & Race - (MR)

After the data processing of each variable, we will obtain scores between 0 and 1 for each variable. All variables' scores will be divided into other socio-economic factors which will be further detailed below. To achieve these values, we will normalise the data using equation 1, the Data Normalisation Formula.

We will then plot the clustered bar charts for each score and create boxplots for the final socio economic borough scores using the software Microsoft Excel.

Income

The data contains the Monthly Pay (MP) and the Employment Rate (ER) separately per borough for the year 2015 from London Borough Profiles *(London Datastore, f, 2021),* the Earnings per Head per Borough (E) from 2018 from the *London Datastore (e, 2021)* and the Cost of Living (CL) from where we updated and corrected crowd-sourced data *(Numbeo, 2022).*

Our equation for calculating the income variable (IN):

$IN = MP + ER + E - CL + \varepsilon$, where IN is Income, MP is Monthly Pay, ER is Employment Rate, E is Earnings per Head per Borough, CL is Cost of Living, and ε is a margin of error which will not be counted numerically towards the final calculations.

- **MP (Monthly Pay)**

    1. We normalise the data for monthly payment in each London borough for the year 2015, using equation 1.
    2. In this case we count the highest score as the borough with the most monthly payment.

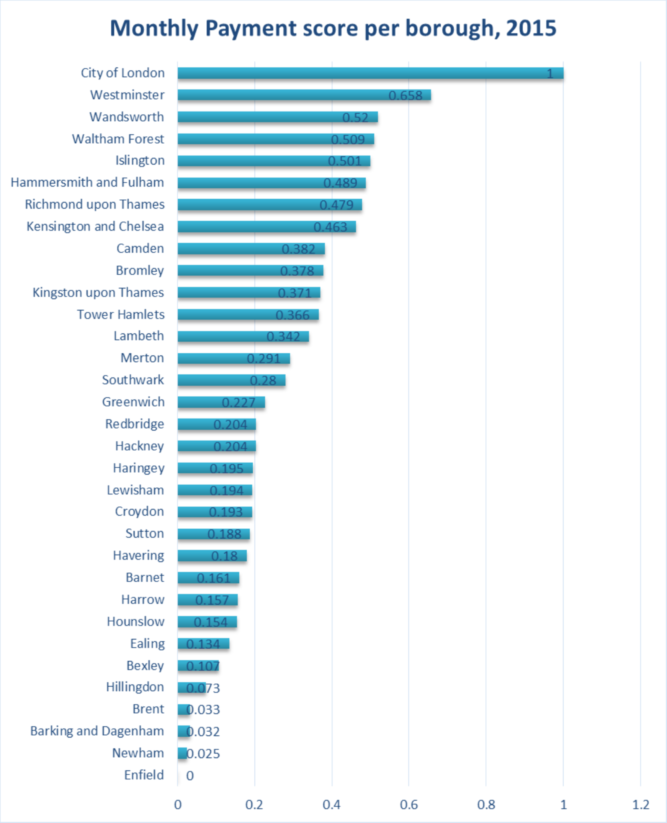


Figure 1: Monthly Payment score per London borough in 2015.

After calculations, we notice that the City of London has the highest monthly payment score while Enfield has the lowest one, City of London being central and Enfield being far north on the London boroughs' map.

- **E (Earnings per Head per Borough)**

  1. We normalise the data of Earnings per Head per borough in each London borough for the year 2018, using equation 1.
  2. In this case we count the highest score as the borough with the most Earnings per Head per borough.

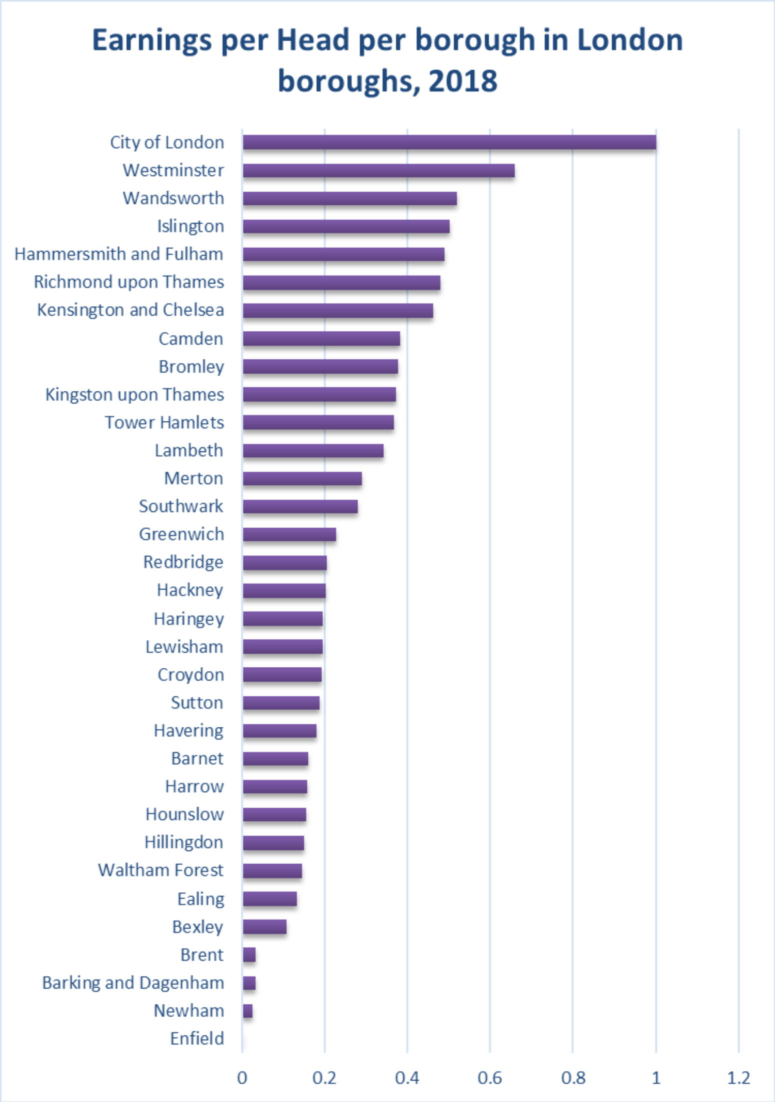


Figure 20: Earnings per Head per borough score per London borough in 2018.

In the figure we notice how the City of London and Westminster along other central boroughs of London rank the highest in terms of Earnings per Head per borough, whilst out-skirting boroughs rank lower.

- **ER (Employment Rate)**

    1. We normalise the data of monthly payment in each London borough for the year 2015, using equation 1.
    2. In this case we count the highest score as the borough with the most employment rate.

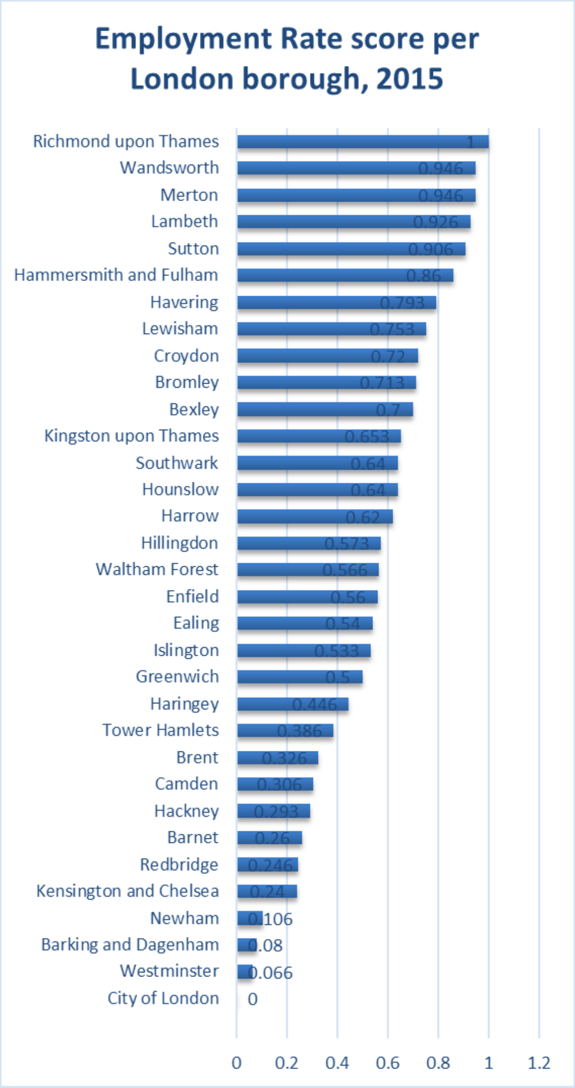


Figure 21: Employment Rate score per London borough in 2015.

After calculations, we notice that the City of London and Westimnster which have the highest Earnings (E), have, inversely, the lowest Employment rate. West neighbouring boroughs such as Richmond upon Thames, Wandsworth, and Merton have the highest employment rate score.

- **CL (Cost of Living)**

CL = Rent/ month + Utilities/ month + Transportation/ day + Supermarket Cost/ week

The data for Rent/ month for a single apartment in the city centre of the borough, the cost of a travel ticket/ day, the cost of monthly utilities/ month and some supermarkets cost prices/ week were taken from *Numbeo (2022)*. However, half of the data regarding the cost of living (CL) in each borough for each variable (rent, utilities, transportation and supermarket cost) was missing since *Numbeo (2022)* offers crowd-sourced data.

Therefore we conducted a Hot Deck imputation strategy in which we calculated each missing data of London boroughs by doing the arithmetic mean of their neighbouring boroughs of which the data was available online. Hot Deck imputation is a statistical method for missing data processing where compromised or non-existent data is replaced with observed responses from *"similar"* units *(Andridge and Little, 2010).* The core of the method is to replace the missing values with plausible data for further analysis not to generate fully accurate missing data replacement *(Little, Rubin and Bayes, 2002).*

For Supermarket Cost we used the data provided by *Numbeo (2022)* but did not use all variables provided by them, only the sum of the prices for: Milk (regular, 1 litre), Loaf of Fresh White Bread, Rice (white), Eggs (regular), Local Cheese (1kg), Chicken Fillets (1kg), Apples (1kg), Banana (1kg), Oranges (1kg), Tomato (1kg), Potato (1kg), Onion (1kg), Lettuce (1 head), Water (1.5 litre bottle), Bottle of Wine (Mid-Range), Domestic Beer (0.5 litre bottle), and Imported Beer (0.33 litre bottle).

1. We hot-deck imputed missing data by replacing it with the already available data from Numbeo (2022) and then through arithmetic means of the neighbouring missing boroughs' data constructed our data set.
2. Then adding up all variables for Cost of Living (CL) and normalising the data using equation 1.
3. In this case we count the highest score as the borough with the highest cost of living.

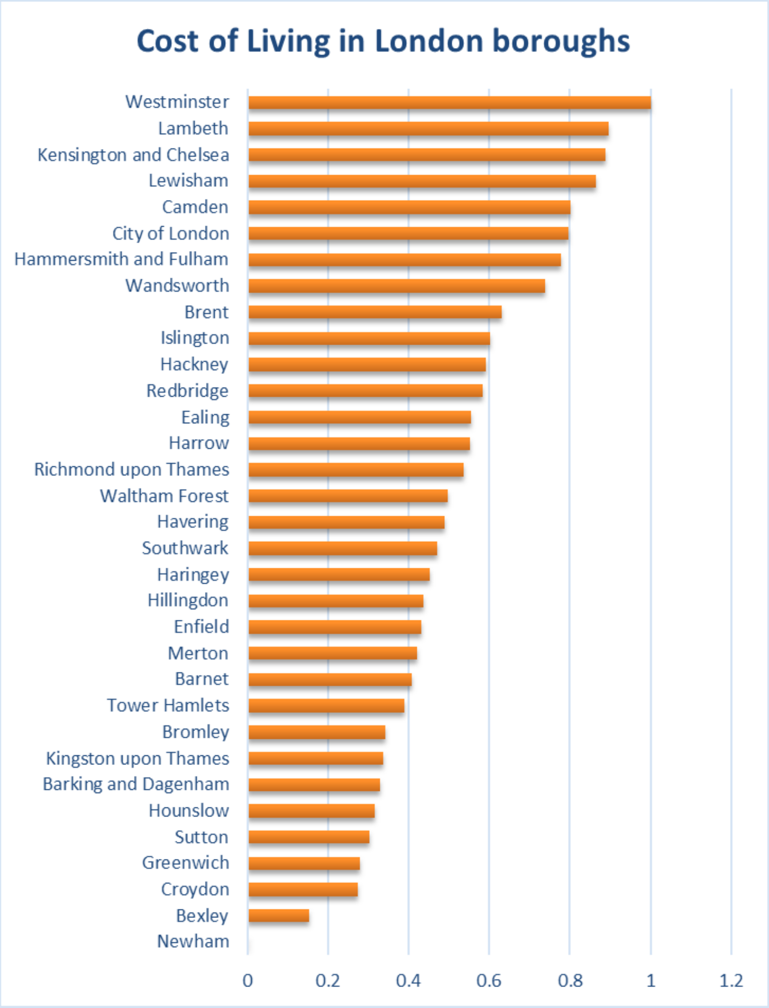


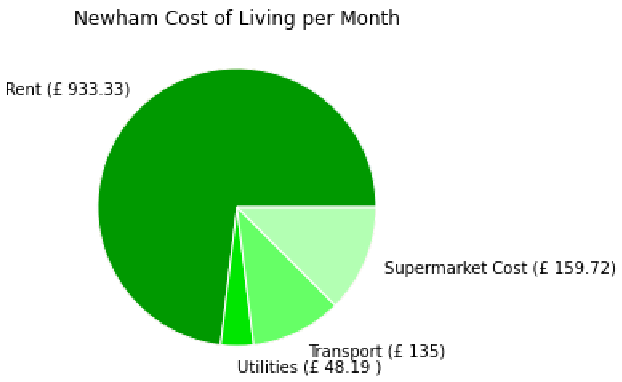
Figure 3: Cost of Living in London boroughs

After calculations, we notice similarly, as with Earnings (E), that central boroughs have the highest cost of living (CL) whilst out-skirting boroughs rank lower having a more affordable cost of living.

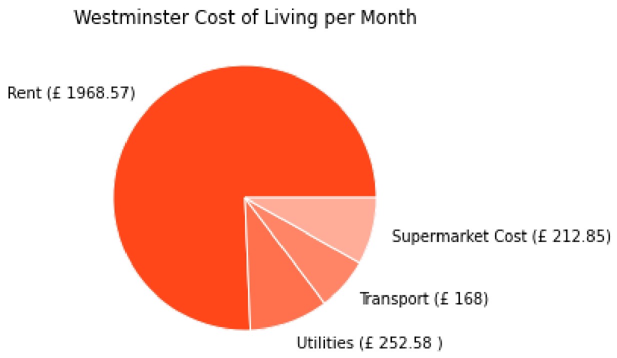
With help from *Python Graph Gallery (2018)* which offers free sources, we picked pie-charts to illustrate in Python, how the Cost of Living is split across the lowest (Newham), highest (Westminster), average (Harringay) and most central (City of London) borough. For the data to be even we calculated the monthly cost for each variable (Rent, Utilities, Transportation, and Supermarket Cost) where a different time-frame was given.

To calculate the monthly Transportation score we assumed that a person uses only two tickets per day and then multiplied the price by 30 (the approximate number of days in a month).

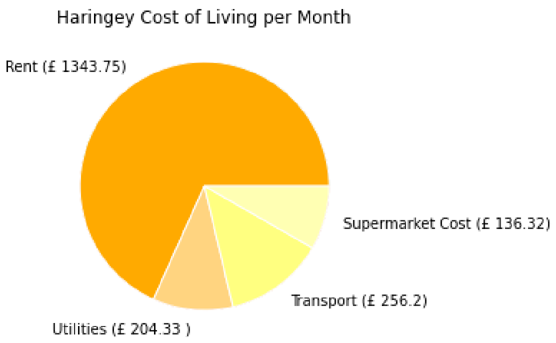
To calculate the monthly Supermarket Cost, we multiply the weekly estimated payment by 52 and then divide the result by 12.



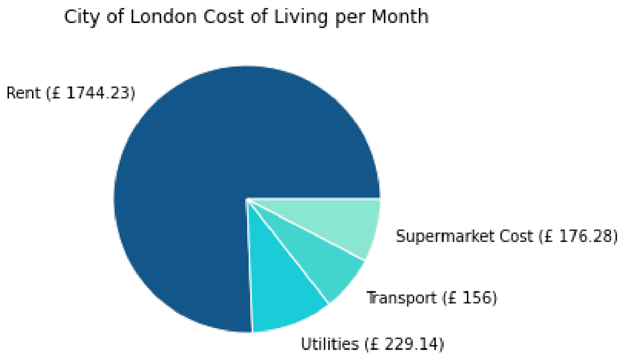
Newham - the most affordable borough



Westminster - the most expensive borough



Haringey - the average (CL)



City of London - the most central borough

Note: A detailed step-by-step iPython notebook for this visualisation is available in our GitHub Repository: [*https://github.com/Ftracy/QMGroup22/blob/main/Cost%20of%20Living%20(CL)%20Pie-charts.ipynb*](https://github.com/Ftracy/QMGroup22/blob/main/Cost%20of%20Living%20(CL)%20Pie-charts.ipynb)
Sources: Python Graph Gallery, 2018

We notice how the rent in the central boroughs such as Westminster and City of London is substantially higher than their utilities or transport. Also, Newham has the lowest utilities payment per month with a quarter or more cheaper than utilities across other London boroughs.

Final Income (IN) Variable Calculation

1. We calculate Income using our Income variable formula:

   IN = MP + ER + E - CL + ε
   where IN is Income, MP is Monthly Pay, ER is Employment Rate, E is Earnings per Head per Borough, CL is Cost of Living, and ε is a margin of error which will not be counted numerically towards the final calculations.
2. Then since some data after calculation gave negative results, we normalised the Income score again using equation 1 so the data is between 0 and 1.
3. In this case, we will invert the data so that the boroughs with a high score will be the ones which will be the least Income-friendly. To invert the data, we will apply (1-x) to all the scores obtained.

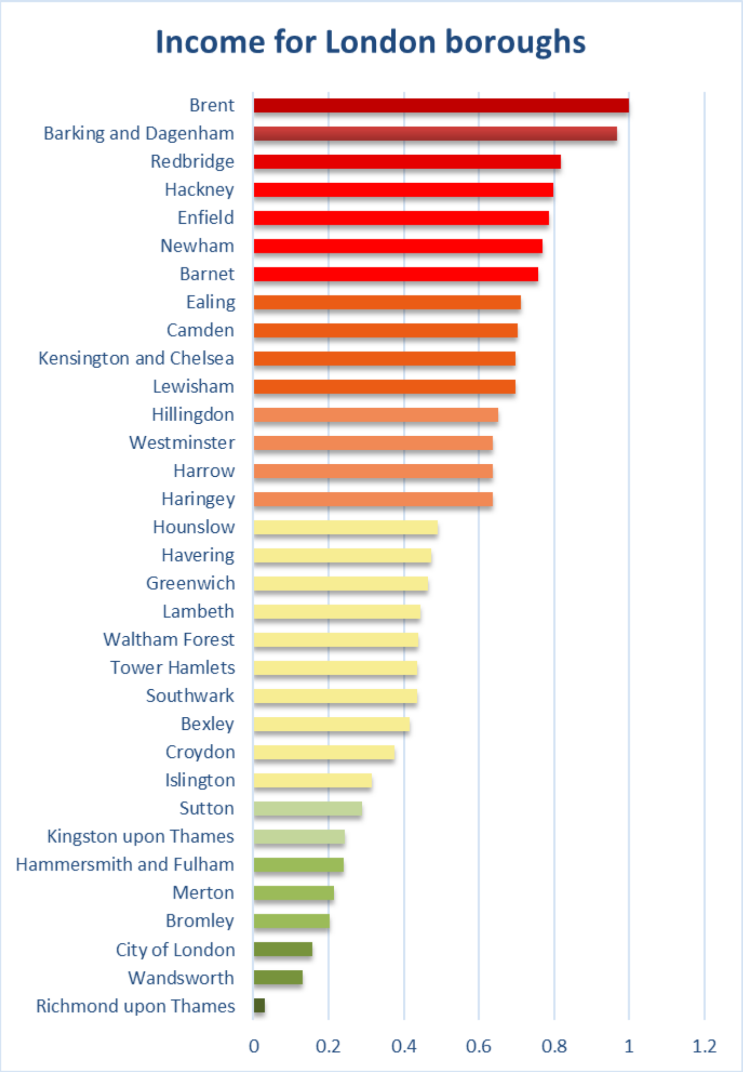


Figure 4: Income for London boroughs

Legend for Income (IN)

🟥 1 ≥ IN > 0.81

🟥 0.81 ≥ IN > 0.71

🟧 0.71 ≥ IN > 0.49

🟨 0.49 ≥ IN > 0.28

🟩 0.28 ≥ IN ≥ 0.003

In this figure we notice how most central London boroughs rank either at the bottom or lower bottom meaning they are more income-friendly while most out-skirting boroughs rank higher, being less environmentally friendly. Some exceptions include central boroughs like Hackney ranking higher, meaning they are less income-friendly than other central boroughs.

Education

The data contains the GCSE attainment per borough (GCSE) for the year 2013/2014 from London Borough Profiles *(London Datastore, f, 2021)*, the A-level results per borough for the year 2013/2014 from the London Datastore, *(London Datastore, g, 2021)* and the number of total schools (TS), private and independent, for the year 2019 in each London borough from the London Datastore *(London Datastore, h, 2021).*

Our equation for calculating the education variable (ED):

ED = GCSE + ALVL + TS + ε , where ED is Education, GCSE is the score for GCSE attainment, ALVL is the score for A-level attainment, TS is the number of total schools and ε is a margin of error which will not be counted numerically towards the final calculations.

- **Total Schools (TS)**

    1. We normalise the data for Total Schools, Independent Schools and Public Schools in each London borough for the year 2019, using equation 1.
    2. We analyse the data for Independent and Public Schools but use only the Total Schools data (TS) in our final calculations
    3. In this case we count the highest score as the borough with the best education.

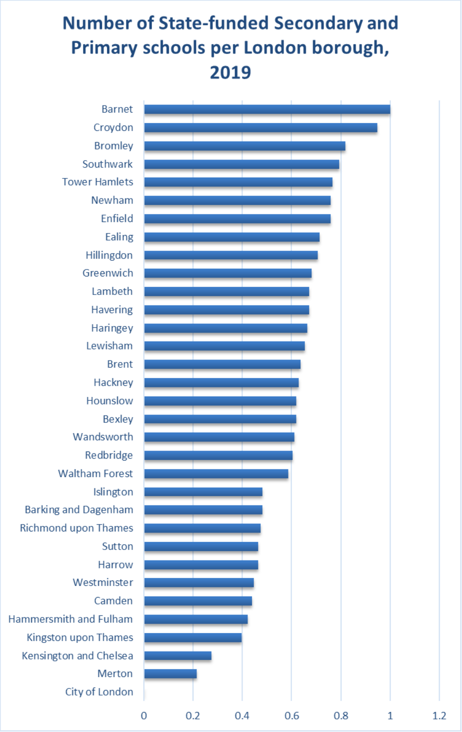


Figure 5: State-funded Secondary and Primary schools per London borough for the year 2019

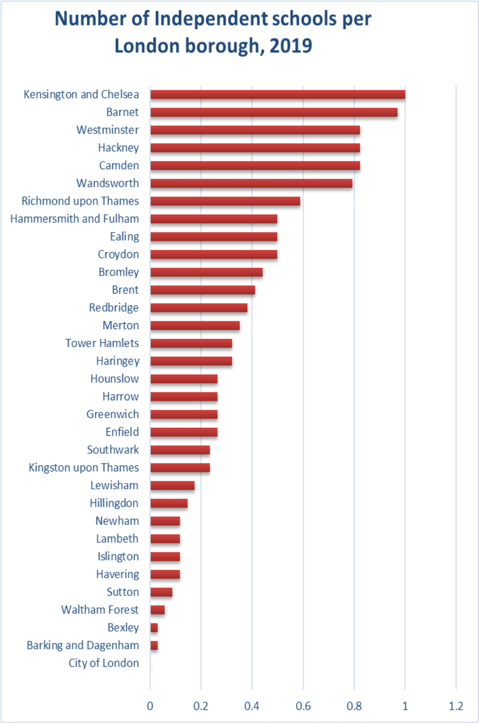


Figure 6: Number of Independent schools per London borough, 2019

From these figures we notice first that there are substantially more state funded schools than independent ones across London boroughs as well as that there are more independent schools in central London boroughs whilst having less state funded schools than outskirting boroughs.

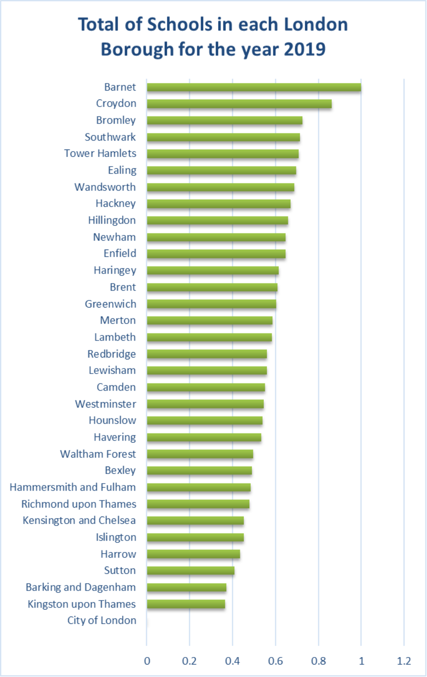


Figure 7: Total of Schools across London Boroughs for the year 2019

We notice that apart from the City of London which has a considerably lower score in the number of schools either public or independent, most other London boroughs have a proportional mix of independent and public schools therefore most having a steady average score.

- **GCSE Attainment (GCSE)**

1. We normalise the data for the GCSE attainment score for each London borough for the year 2013/2014 using equation 1.
2. In this case we count the highest score as the borough with the highest GCSE attainment.

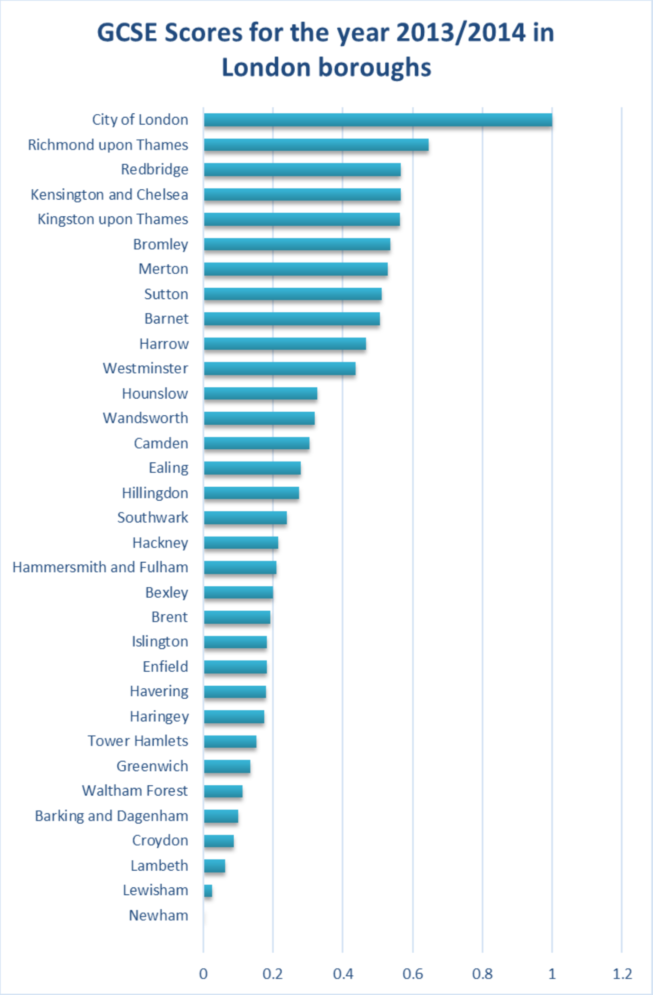


Figure 8: GCSE attainment per London borough for the year 2013/2014

The City of London has the highest GCSE score, considerably higher than the rest of scores which appear steady across most central London boroughs whilst the out-skirting boroughs have a lower score.

This is surprising since the City of London as seen in the calculation of (TS) the Total of Schools has the least number of public or independent schools in contrast to other London boroughs.

- **A-level Attainment (ALVL)**

1. We normalise the data for the A-level attainment score for each London borough for the year 2013/2014 using equation 1.
2. In this case we count the highest score as the borough with the highest A-level attainment.

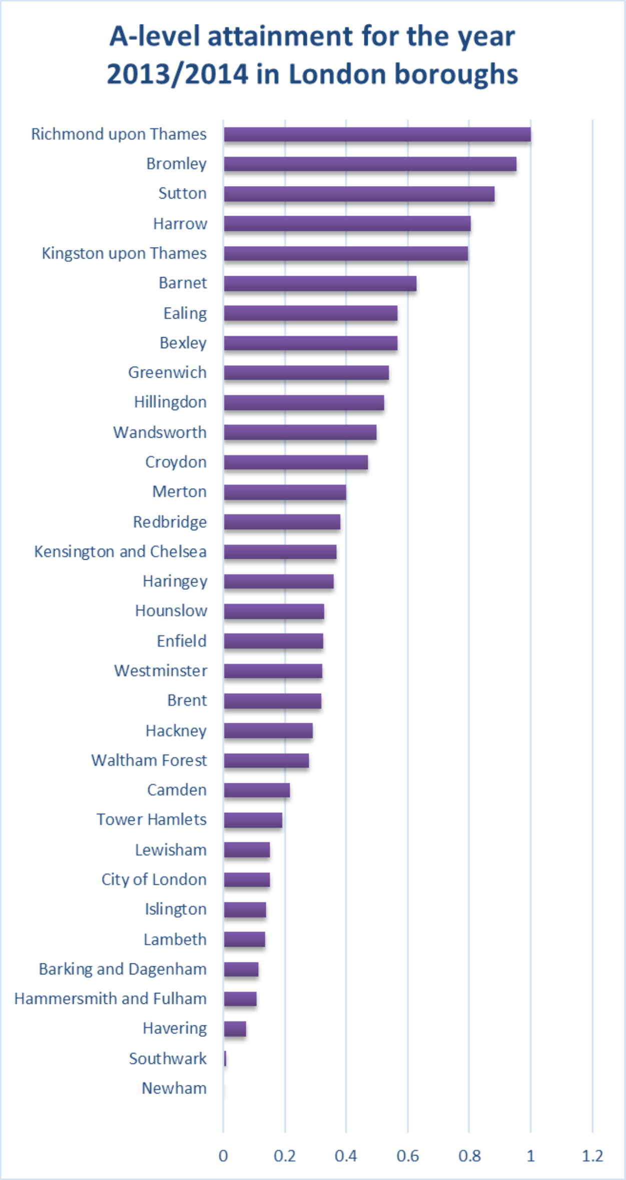


Figure 9: GCSE attainment per London borough for the year 2013/2014

We notice that apart from the City of London, which dropped considerably from its ranking on GCSE's scores, all other boroughs remained relatively steady on their exam attainments, where in the case that they increased or decreased the change was not considerably high from their GCSE attainment score.

Final Education (ED) Variable Calculation

1. We calculate Education using our Education variable formula:

   ED = ALVL + GCSE + TS + ε
   where ED is Education, ALVL is the A-level attainment score, GCSE is the GCSE attainment score, TS is the total of schools score, and ε is a margin of error which will not be counted numerically towards the final calculations.
2. Then we normalised the Education score again using equation 1 so the data is between 0 and 1.
3. In this case, we will invert the data so that the boroughs with a high score will be the ones which will have the least developed Education score. To invert the data, we will apply (1-x) to all the scores obtained.

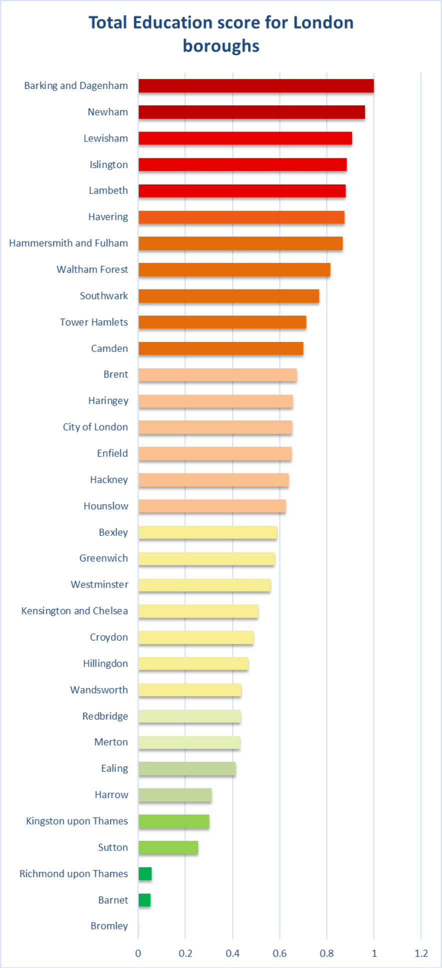


Figure 10: Total Education score for London boroughs

The figure shows how by analysing both the number of schools in a borough and its exam attainment out-skirting boroughs outperform central London boroughs, therefore we have outskirting Bromley and Barnet in the south and respectively north bound of London boroughs whilst The City of London has an average and Westminster and above average score.

Legend for Education (ED)

🟥 1 ≥ ED ≥ 0.87

🟧 0.87 > ED ≥ 0.7

▨ 0.7 > ED ≥ 0.6

🟨 0.6 > ED ≥ 0.43

🟩 0.43 > ED ≥ 0.3

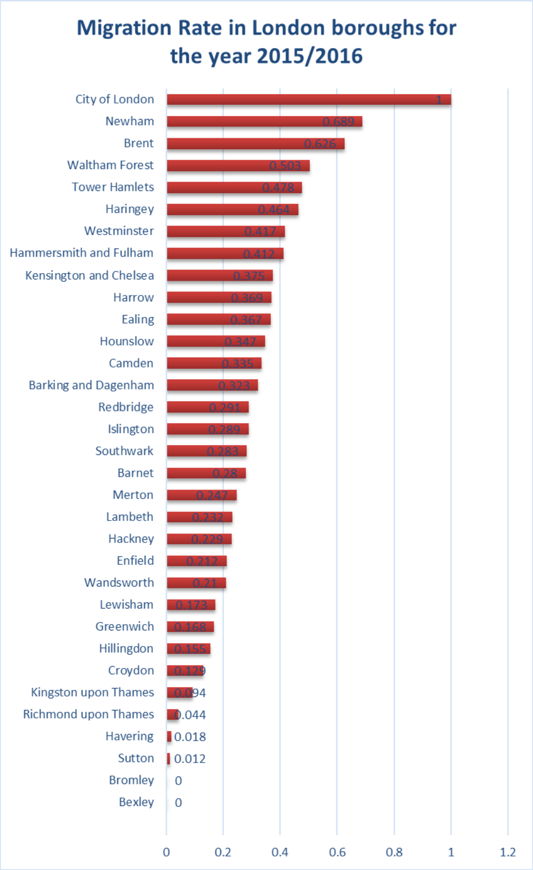
Race & Migration

The data contains the Black, Asian, and minority ethnic (BAME) percentage and the Migration (M) Rates per borough for the year 2015/2016 from London Borough Profiles *(London Datastore, f, 2021).*
Our equation for calculating the education variable (RM):

RM = BAME + M + ε , R

- **Migration Rate (M)**

    1. We normalise the data for Migration Rates in each London borough for the year 2015/2016, using equation 1.
    2. In this case we count the highest score as the borough with the highest migration



.Figure: Migration Rate in London boroughs for the year 2015/2016

We notice how central London boroughs and most neighbouring boroughs of The City of London have the highest migration rates whilst out-skirting boroughs such as Bromley and Sutton have the lowest Migration rates.

- Percentage of Black, Asian, and minority ethnic groups **(BAME)**

1. We normalise the data for BAME percentage in each London borough for the year 2015/2016, using equation 1.
2. In this case we count the highest score as the borough with the highest percentage of BAME groups.

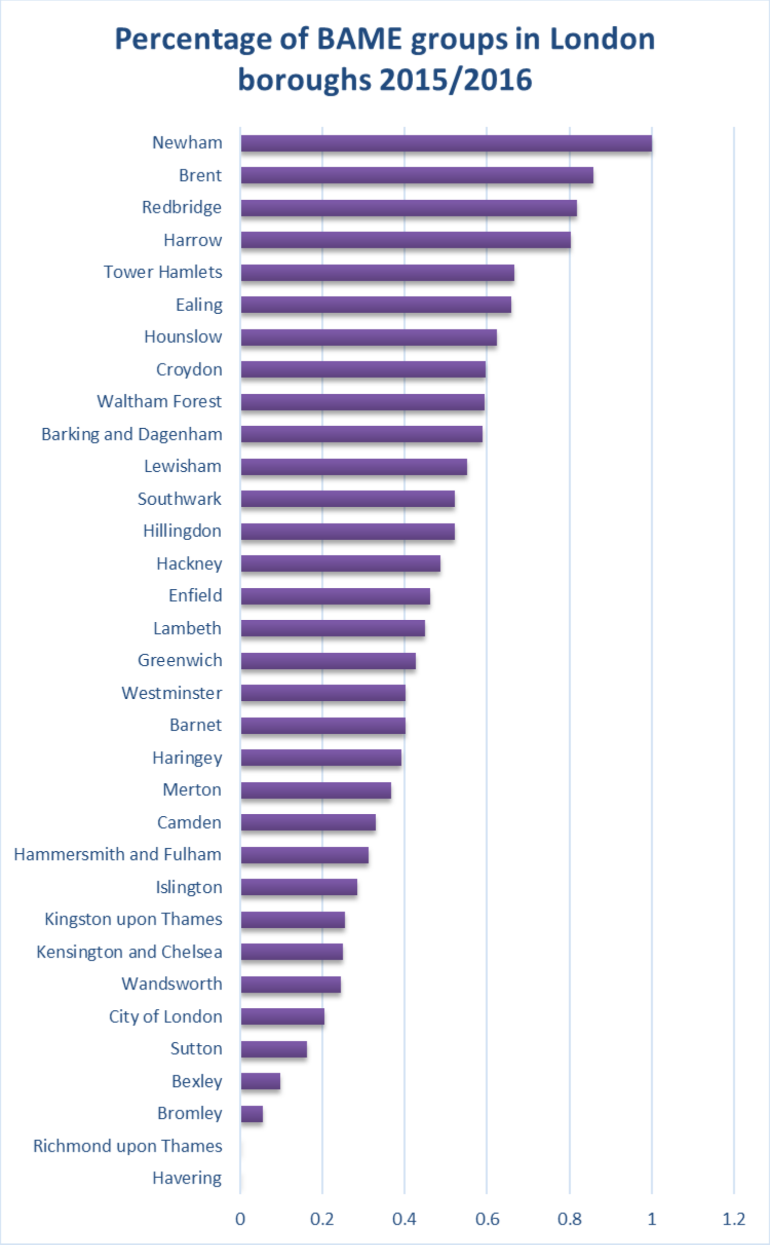


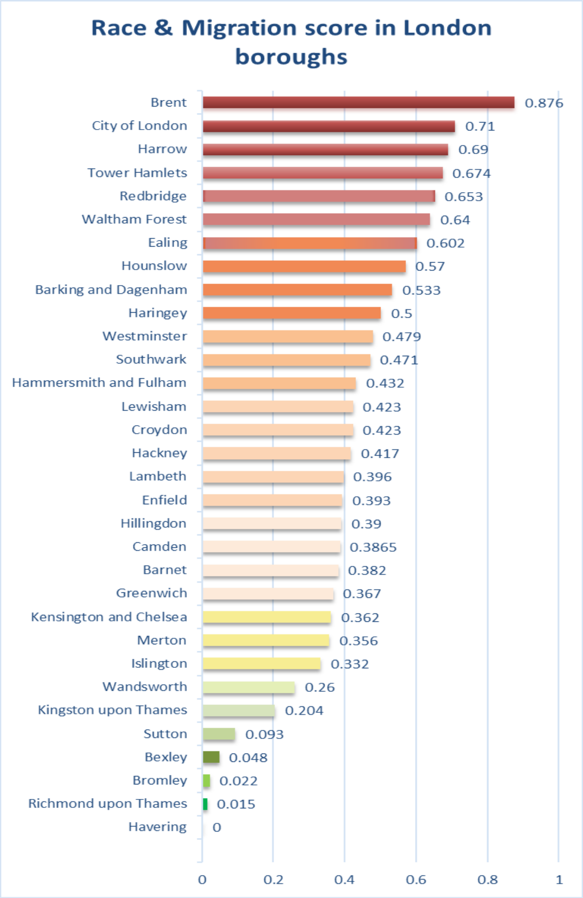
Figure: Percentage of BAME groups per London borough in the year 2015/2016

We notice how the neighbouring boroughs of central boroughs such as Newham, and some out-skirting boroughs have a higher percentage while central boroughs have an average towards low or low percentage; the City of London being the 6th lowest in percentage.

Final Race & Migration (RM) Variable Calculation

1. We calculate Race and Migration using our Race and Migration variable formula:

   RM = BAME + M + ε , where RM is the Race and Migration score, BAME is the score f for the percentage of Black, Asian, and minority ethnic groups in London boroughs, M is the score for Migration rates, TS is the number of total schools, and ε is a margin of error which will not be counted numerically towards the final calculations.
2. Then we normalised the Race and Migration score again using equation 1 so the data is between 0 and 1.
3. In this case, we will not invert the data because the higher the score of Migration and Race the more improvement the borough needs in terms of supporting the BAME and migrant communities, which typically have poorer health, as expanded on in the Literature Review.



Legend for Race & Migration (RM)

🟥 0.87 ≥ RM ≥ 0.6

🟧 0.6 > RM ≥ 0.5

⬜ 0.5 > RM ≥ 0.39

🟨 0.39 > RM ≥ 0.3

🟩 0.3 > RM ≥ 0

Figure: Race & Migration score in London boroughs

We notice how the neighbouring boroughs of central boroughs such as Brent, central boroughs such as City of London and Tower Hamlets,and out-skirting boroughs such as Harrow all have a high Migration & Race score, meaning that there is room for supporting and welcoming migrant and BAME communities across all London boroughs.

TOTAL BSSE

After obtaining the scores for each variable, to calculate the Borough Score for socio-economic factors we add up all individual scores and variables for each borough through the BSSE equation:

$$\text{BSSE} = \text{IN} + \text{ED} + \text{RM} + \varepsilon$$

Where BSSE is the borough score for socio-economic factors, IN is Income, ED is Education and RM is Race & Migration as further detailed above. The lower the score the better the borough is situated in terms of Socio-Economic factors.

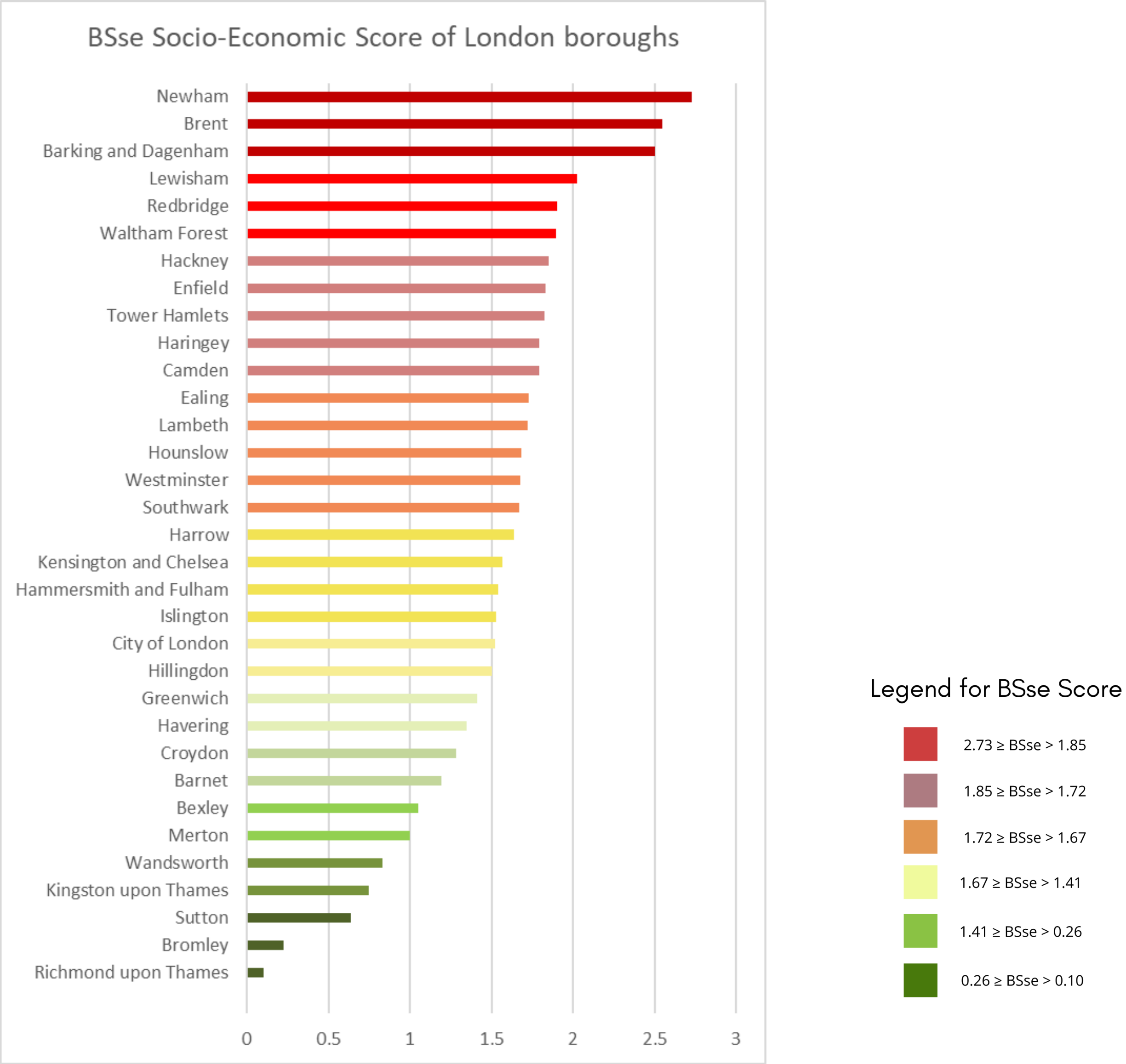
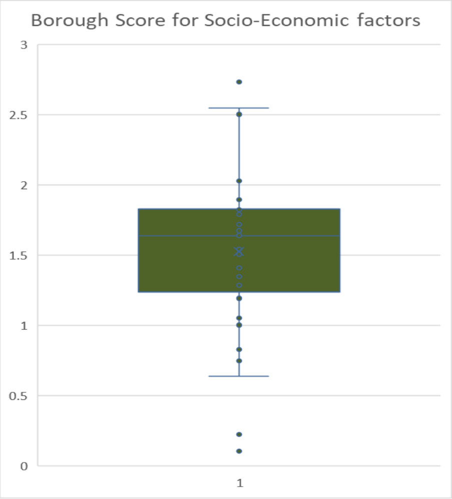


Figure: The BSse Socio-Economic Score of London boroughs

Table 5: Boxplot values		BSse
Boxplot Statistics	Minimum	0.103
	Maximum	2.732
	Lower quartile	1.239
	Upper quartile	1.826
	Interquartile range	0.587
	Lo. outlier limit	0.103
	Hi. outlier limit	2.732
Tukey fences	Lo. Tukey fence (LQ - 1.5 * IQR)	0.358
	Hi. Tukey fence (UQ + 1.5 * IQR)	2.706
Outlier	Newham, Bromley, and Richmond	



Adding our data up we notice how out-skirting boroughs have the lowest and therefore better score regarding socio-economy whilst central London boroughs rank median and neighbouring central boroughs rank the highest having the worst socio-economic scores. We also notice that Richmond, a west out-skirting borough, has the lowest score, and therefore represents the best socio-economic profile while Newham, a central London borough, has the highest score and therefore represents a low socio-economic profile.

References:

Andridge, R. and Little, R., 2010. A Review of Hot Deck Imputation for Survey Non-response. International Statistical Review, [online] 78(1), pp.40-64. Available at: <https://doi.org/10.1111/j.1751-5823.2010.00103.x> [Accessed 4 January 2022].

Babbie, E., 2007. The Practice of Social Research. 11th ed. [ebook] Belmont: Thomson/Wadsworth. Available at: <https://7b0a75e0-a-62cb3a1a-s-sites.googlegroups.com/site/fspacburean/home/quantitative-methods/postarefaratitlu/chapter%201%20and%202%20babbie.pdf?attachauth=ANoY7cqduS5VsK7B9OtLnt51hAyshJlaIYfsvUokIEE_uuP8O_eC_6DeKqRGr65GRR1Jymjsk4q8bAhqC5-kSSalxKIWRWbpe2Rmq2yP6YbZpqkmbuRLMsqs2c6iod8BSQeEhAWE0piP-HsljuE0OsisHeIZue2O_7Xx2yFq1pVvhoDi7IsWLrZ4Yc2uiOLyhfI5as3ySnKbKs4t2IRlWh69VDr4f3U86A5BT_xtL8bmMbLEVJ5Zf_CQDxDbDO6_uQHzToZfjVRP7tfzxv1Dh1xvCqW3OzDSkoJQAonSIwlFIhaBuTt8DgY%3D&attredirects=0> [Accessed 4 January 2022].

Brewer, C., MacEachren, A., Pickle, L. and Herrmann, D., 1997. Mapping Mortality: Evaluating Color Schemes for Choropleth Maps. Annals of the Association of American Geographers, [online] 87(3), pp.411-438. Available at: <https://doi.org/10.1111/1467-8306.00061> [Accessed 4 January 2022].

Canva, 2021. Canva. [online] Canva.com. Available at: <https://www.canva.com/en_gb/> [Accessed 4 January 2022].

Congdon, P., 2008. A spatial structural equation model for health outcomes. Journal of Statistical Planning and Inference, [online] 138(7), pp.2090-2105. Available at: <https://doi.org/10.1016/j.jspi.2007.09.001> [Accessed 4 January 2022].

Google Developers, 2021. Normalization | Data Preparation and Feature Engineering for Machine Learning | Google Developers. [online] Google Developers. Available at: <https://developers.google.com/machine-learning/data-prep/transform/normalization> [Accessed 4 January 2022].

Kaplan, D., 2004. The Sage Handbook of Quantitative Methodology for the Social Sciences. [ebook] Thousand Oaks: Sage Publications, pp.106-109. Available at: <https://www.researchgate.net/profile/Nguyen_Trung_Hiep3/post/Can_you_provide_me_some_good_books_papers_on_quantitative_research/attachment/59d62c9679197b807798ae98/AS%3A347459084668928%401459852109528/download/The+SAGE+Handbook+of+Quantitative+Methodology+for+the+Social+Sciences.pdf> [Accessed 4 January 2022].

Little, R. and Rubin, D., 2014. Bayes and Multiple Imputation. Statistical Analysis with Missing Data, [online] pp.200-220. Available at: <https://doi.org/10.1002/9781119013563.ch10> [Accessed 4 January 2022].

London Datastore, a, 2021. London Atmospheric Emissions Inventory (LAEI) 2016 – London Datastore. [online] Data.london.gov.uk. Available at: <https://data.london.gov.uk/dataset/london-atmospheric-emissions-inventory--laei--2016> [Accessed 30 December 2021].

London Datastore, c, 2021. Land Use by Borough and Ward – London Datastore. [online] Data.london.gov.uk. Available at: <https://data.london.gov.uk/dataset/land-use-ward> [Accessed 30 December 2021].

London Datastore, d, 2021. Public Health Outcomes Framework Indicators – London Datastore. [online] Data.london.gov.uk. Available at: <https://data.london.gov.uk/dataset/public-health-outcomes-framework-indicators> [Accessed 30 December 2021].

London Datastore, e, 2021. Earnings by Place of Residence, Borough – London Datastore. [online] Data.london.gov.uk. Available at: <https://data.london.gov.uk/dataset/earnings-place-residence-borough> [Accessed 30 December 2021].

London Datastore, f, 2021. London Borough Profiles and Atlas – London Datastore. [online] Data.london.gov.uk. Available at: <https://data.london.gov.uk/dataset/london-borough-profiles> [Accessed 30 December 2021].

London Datastore, h, 2021. Schools and Pupils by Type of School, Borough – London Datastore. [online] Data.london.gov.uk. Available at: <https://data.london.gov.uk/dataset/schools-and-pupils-type-school-borough> [Accessed 30 December 2021].

London Datastore, i, 2021. 2011 Census Health & Care – London Datastore. [online] Data.london.gov.uk. Available at: <https://data.london.gov.uk/dataset/2011-census-health-care> [Accessed 30 December 2021].

Natário, I. and Knorr-Held, L., 2003. Non-Parametric Ecological Regression and Spatial Variation. Biometrical Journal, [online] 45(6), pp.670-688. Available at: <https://doi.org/10.1002/bimj.200390041> [Accessed 4 January 2022].

Numbeo, 2022. Cost of Living. [online] Numbeo.com. Available at: <https://www.numbeo.com/cost-of-living/> [Accessed 30 December 2021].

The Pennsylvania State University, a, 2022. Lesson 5: Multiple Linear Regression | STAT 501. [online] PennState: Statistics Online Courses. Available from: https://online.stat.psu.edu/stat501/lesson/5 [Accessed 4 January 2022].

The Pennsylvania State University, b, 2022. Lesson 12: Multicollinearity & Other Regression Pitfalls | STAT 501. [online] PennState: Statistics Online Courses. Available from: https://online.stat.psu.edu/stat501/lesson/12 [Accessed 4 January 2022].

The Pennsylvania State University, c, 2022. 12.5 - Reducing Data-based Multicollinearity | STAT 501. [online] PennState: Statistics Online Courses. Available from: https://online.stat.psu.edu/stat501/lesson/12/12.5 [Accessed 4 January 2022].

Public Health England, 2015. Health inequalities in London. [ebook] London: Public Health England. Available at: <https://www.gov.uk/government/publications/health-inequalities-in-london> [Accessed 20 December 2021].

Python Graph Gallery, 2018. [online] Python-graph-gallery.com. Available at: <https://www.python-graph-gallery.com/pie-plot-matplotlib-basic> [Accessed 5 January 2022].

Waskom, M., a, 2021. seaborn.boxplot — seaborn 0.11.2 documentation. [online] Seaborn.pydata.org. Available at: <https://seaborn.pydata.org/generated/seaborn.boxplot.html#seaborn.boxplot> [Accessed 30 December 2021].

Waskom, a, 2022. seaborn.heatmap — seaborn 0.11.2 documentation. [online] Seaborn.pydata.org. Available at: https://seaborn.pydata.org/generated/seaborn.heatmap.html [Accessed 4 January 2022].

Waskom, M., b, 2021. seaborn.swarmplot — seaborn 0.11.2 documentation. [online] Seaborn.pydata.org. Available at: <https://seaborn.pydata.org/generated/seaborn.swarmplot.html#seaborn.swarmplot> [Accessed 30 December 2021].

Waskom, b, 2022. seaborn.pairplot — seaborn 0.11.2 documentation. [online] Seaborn.pydata.org. Available at: <https://seaborn.pydata.org/generated/seaborn.pairplot.html> [Accessed 4 January 2022].

Waskom, c, 2022. Linear regression with marginal distributions — seaborn 0.11.2 documentation. [online] Seaborn.pydata.org. Available at: https://seaborn.pydata.org/examples/regression_marginals.html [Accessed 6 January 2022].