Effect of household income on student performance

Candidate numbers: 2210077, 2206664, 2200287
Student ID numbers: 100913283, 100925146, 10091480

Table of Contents

The Research Question

How does household income affect student performance in secondary education and their chances for getting into higher education?

The Hypothesis

Students from lower income households are more likely to underperform in secondary education and therefore not meet entry requirements for university.

Introduction

Household income has a large impact on a variety of aspects of a child's life. One of the major aspects could be their academic performance. The most popular view is that more "wealthy" parents provide a better academic environment for their children, whilst parents with lower income tend to lead to early school leaving. This may be due to children of lower income parents not having enough funds to continue their higher education or them perceiving a responsibility to earn money, which leads to them getting a job at a young age instead. In most cases, these young individual's wind-up never going back to education despite having the funds later in life. This emphasises the need of identifying impoverished / disadvantaged students at danger of underachievement as early as possible, so that more resources may be directed towards them, who will someday become full participants in all aspects of social life and the financial system.

In this project, we will investigate the causality between Household income and academic performance amongst a set group of pupils from three London boroughs.
We will use housing prices as an indicator of household income and GCSE grades in Maths and English as determinants of academic performance.

Literature review

A substantial amount of research has been done on the effects of parental background on their children's cognitive skills, education, health, and later income. Adolescents' behaviour and decisions are heavily influenced by their parents and the family environment in general. The paper by (Colm et al., n.d.) investigated how much early school leaving (around the age of 16) may be attributed to changes in permanent income, parental education levels, or economic shocks at this age. Using OLS, they indicated that the parental education level is more significant for their children's further education rather than the parental income. However, the estimates based on IV did not show this effect. When exogenous variance in permanent income is controlled for, it appears that parental education has no independent effect, and that permanent income remains crucial even when credit constraints at age 16 are considered. Similarly, the paper by (Carneiro & Heckman, 2009), suggests, children with

a better family background, in terms of parental behaviour, quality time as a family while the children are growing up, have a much larger impact on their choices for the future rather than the household income or parental education. In addition to that, (Cooper & Stewart, 2018) estimated that a £7000 increase in household income for children in receipt of free school meals (FSM), raising them to the average income for the rest of the population could be expected to remove approximately half the gap in Key Stage 2 outcomes between FSM children and their non-FSM peers. (Cooper & Stewart, 2018)

Most of the previous literature found on this field include two or more independent variables such as parental education, income shock to children (at 16) alongside with household income. There is little to no literature which addresses the causal effect of only household income on the children's academic performance. Therefore, we intend to make progress in this matter in question and only focus on indicators of household income as an independent variable, with the purpose that young adults are capable of continuing their higher education.

Data

We obtained our official Housing price dataset for Croydon, Bromley, and Richmond from the Office for National Statistics website. The housing price dataset includes house prices for 2018 and 2019. The academic data for Richmond, Bromley and Croydon is acquired from the government website, which includes school type, student grades, student gender etc.

All datasets are based on 2018 and 2019 as those are the most recent data, we were able to access. The reason behind choosing the most recent dataset available is to ensure better accuracy and relevance to the current economy.
The table below includes the appropriate descriptive statistics for the average house prices of the three councils over the two years.

	Median house price 2018	Median house price 2019
count	6.800000e+01	6.800000e+01
mean	4.905095e+05	4.905095e+05
std	1.773749e+05	1.773749e+05
min	2.900000e+05	2.900000e+05
25%	3.718750e+05	3.718750e+05
50%	4.500000e+05	4.500000e+05
75%	5.562500e+05	5.562500e+05
max	1.370000e+06	1.370000e+06

Table 1 Summary statistics

Methodology

For this project we have selected three boroughs; Croydon, Bromley, and Richmond; in order to test our hypothesis in distinct parts of London. Croydon is ranked one of the lowest at 23 in London's "richest" boroughs, whereas Richmond and Bromley are ranked 6 and 10 respectively.(Peracha, 2020) This will allow us to look at three very different regions of the city and see how household income affects student performance and their chances at getting into university. We selected non-private schools with mixed gender students to ensure no external factors affect the test results. We have used "Number of pupils achieving EBacc Maths subject area with a standard 9-4 pass" and "Number of pupils achieving EBacc English subject area with a standard 9-4 pass" as the dependent variables as these are the minimum requirements for students to qualify for admission into university. We also have variables such as "Number of pupils at the end of key stage 4 with English as their first language" and "Number of pupils at the end of key stage 4 with English as additional language (EAL)" in order to remove biases towards disadvantaged students. The independent variable is the median house price which is an indicator of household income.

As the housing price dataset on excel had a lot of sheets irrelevant to our investigation, we only extracted the data for Croydon, Bromley and Richmond into python using pandas. In addition, the academic information datasets included schools that are not located in Croydon, Bromley, or Richmond. With the aim to only use unbiased data, we refined the datasets using the .loc[ ] function, which removed any secondary school that was not located

in areas of interest and was not public or mixed gender. We also renamed the column names of the variables to make it easier to interpret.

Some of the numerical values in the data frame were operating as objects and strings which made it difficult to do calculations with them. Therefore, we changed them into integers and floats using pandas. We then calculated the descriptive statistics which show the mean, median, standard deviation etc. of each variable to check for any patterns in the different datasets as well as visualise what the data is showing.

As we had two datasets for the academic information with different variables, we merged them using pd.merge, and then we merged that data frame with the average house prices in the corresponding area. This allowed us to run OLS regressions using "Number of pupils achieving EBacc Maths subject area with a standard 9-4 pass" as the dependent variable and "Average house value in 2018" as the independent variable.

$$Y_i = \beta_0 + \beta_1 X_i + u$$

$Y_i = Student\ performance\ at\ time\ i$
$X_i = Measure\ of\ household\ income\ at\ time\ i$
$\beta_0 = constant$
$\beta_1 = magnitude\ of\ the\ effect\ of\ housing\ prices\ on\ student\ performance$
$u = error\ term$

$H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$

Results

The box plot below represents the average distribution of average house prices in 2018 and 2019 across the three local authorities in the data set. It is important to note down that Richmond's lowest average price is higher than Bromley's maximum average price. However, there are two outliers for Richmond with values significantly higher than the rest.

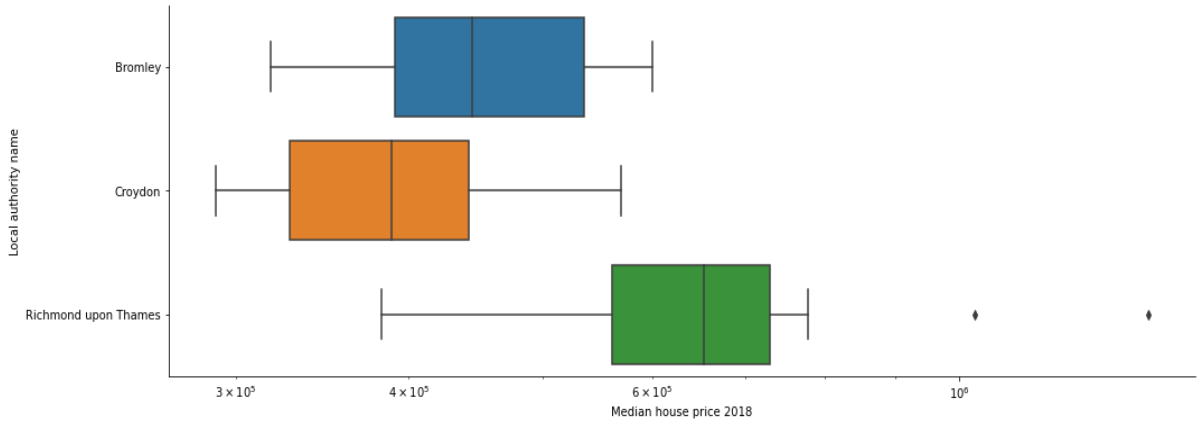

Figure 1 boxplots of house prices in each borough in 2018

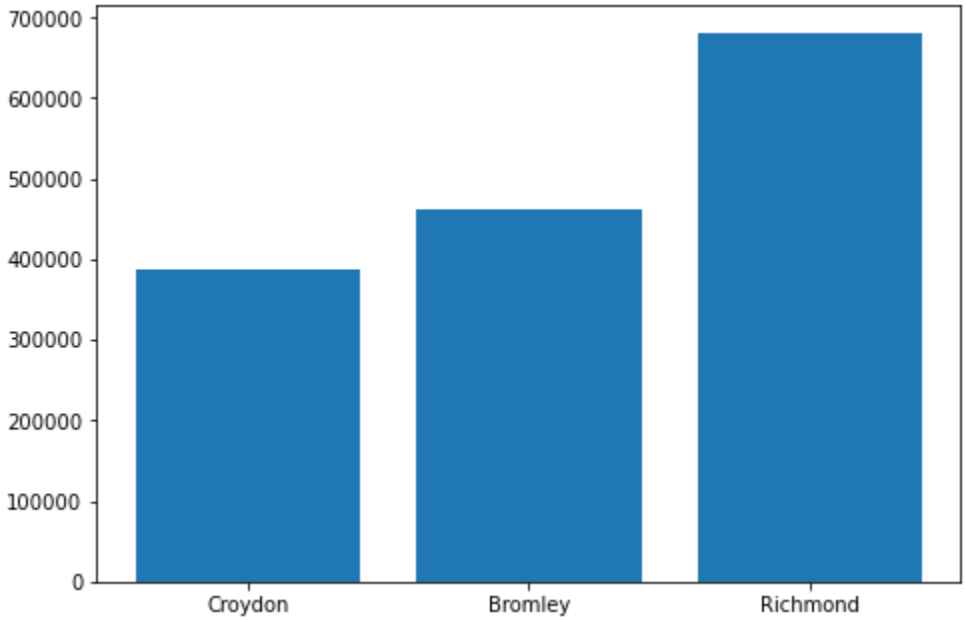

Figure 2 Bar chart of house prices in each borough

The categorical scatter plot plots the categorical variable Town with the data points being Number of pupils achieving EBacc Maths subject area with a standard 9-4 pass in 2019 in each school. Due to having a finite sample of observation it is clear that students in Croydon tend to have a lower rate of passing the basic subject of Maths. The number of students in Bromley who have the highest pass rate out of the three authorities, is treated as an outlier as it is significantly different from the rest. The same scatter plot but for the number of students who passed the subject of English in 2019 can be found in the appendix.

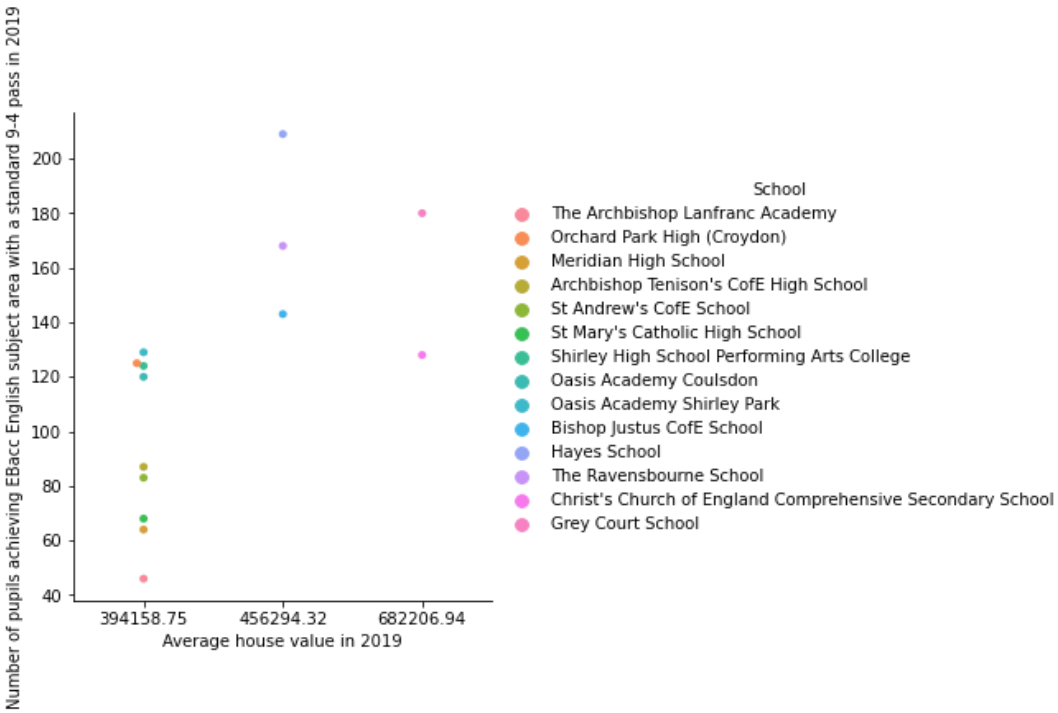

Figure 3 Scatter plot of Students GCSE English score in 2019 against the average house price in each borough

Following up on the previous scatter plot, this one establishes a vague but positive correlation between the average household income for 2019 and students GCSE results in 2019. As seen on the graph, students in lower household income areas are more likely to underachieve and thus have lower passing rate.

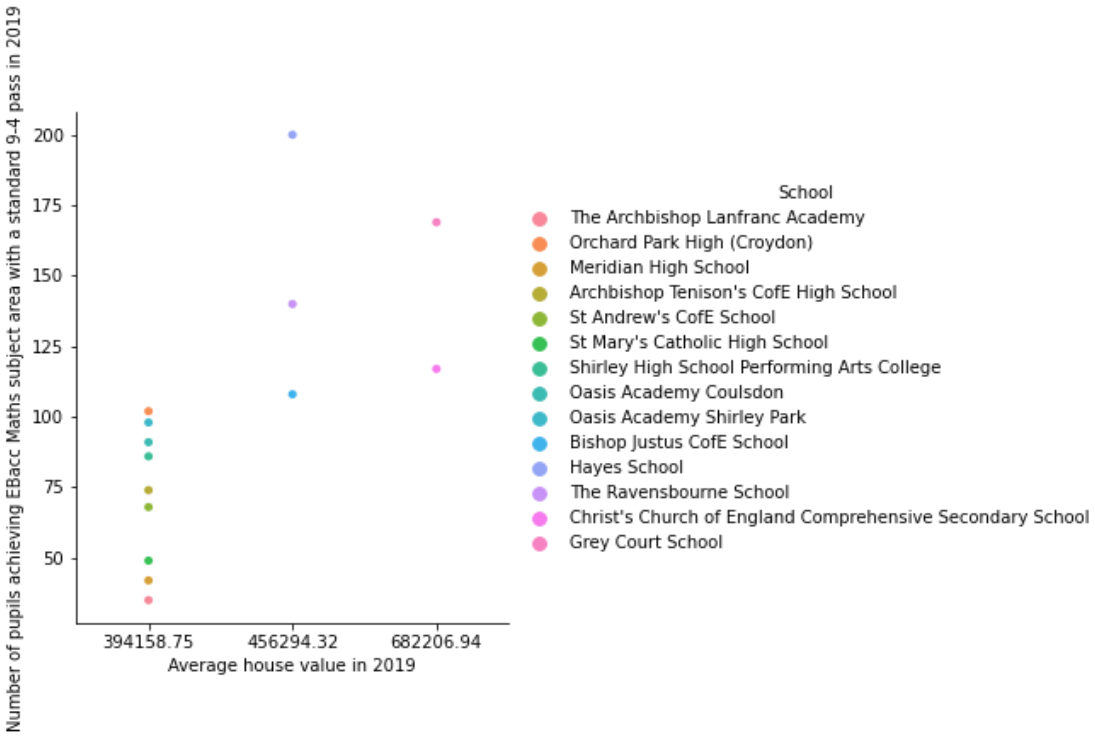

Figure 4 Scatter plot of Students GCSE Maths score in 2019 against the average house price in each borough

In addition, we executed four regression analysis with four different dependent variables against house prices to test for causality. The dependent variables were pupils achieving 9-4 in Maths and English in the years 2018-2019. The results in **Table 2** to **5** display very similar results as they all confirm our null hypothesis. The F statistics values (108.5, 94.12, 117.1 and 121.7) are large for every test and the correlation coefficients are positive values which suggests significance at the 5% significance level. The simple linear regression focuses only on the impact of average household price on GCSE scores, without considering any omitted variables. The regression coefficient is 0.0003/0.0002 in our four models. Thus, the magnitude of the effect of average household prices on the results of the students is estimated to be fairly small. The lack of p value in our model is due to the finite sample size and thus cannot prove statistical significance.

Limitations

Analysing school performance and household income in the same year may mean the mechanisms of household income have not had a chance to take full effect on young pupils' academic performance, hence, the results might not be very accurate.

Furthermore, there are only 2 observations for schools in Richmond because most of the schools in that area are independent and we wanted to remove any biases, therefore, we ought to eliminate any schools that are not public.

Our investigation could be better by using data from more areas instead of only two boroughs, as it would provide us with more schools to analyse student performance which would indicate better accuracy. We could also use a wider time period to test our hypothesis to examine any difference in the results throughout different years.

Conclusion

All things considered; we have established that household income has a significant impact on a child's academic performance. The OLS regression results on **Table 2** to **5** and the graphs in **Figure 1** to **4** show a strong relationship between the dependent and independent variable. Consequently, this indicates that a change in household income changes student performance correspondingly.

Bibliography:

(Carneiro & Heckman, 2009), Carneiro, P. M., & Heckman, J. J. (2009). Human Capital Policy. Ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=434544

(Colm et al., n.d.), Colm, A., Vincent O', H., & Walker, S. (n.d.). THE IMPACT OF PARENTAL INCOME AND EDUCATION ON THE SCHOOLING OF THEIR CHILDREN. https://ifs.org.uk/wps/wp0505.pdf

(Cooper & Stewart, 2018) Cooper, K., & Stewart, K. (2018, July 27). Does money affect children's outcomes? JRF. https://www.jrf.org.uk/report/does-money-affect-children%E2%80%99s-outcomes

(Peracha, 2020) Peracha, Q. (2020, April 19). The average income where you live in London compared to other boroughs. MyLondon. https://www.mylondon.news/news/zone-1-news/londons-richest-boroughs-average-income-18114728

(Download Data - GOV.UK, n.d.) Download data - GOV.UK. (n.d.). Find and Compare Schools in England. Retrieved April 29, 2022, from https://www.compare-school-performance.service.gov.uk/download-data?currentstep=datatypes®iontype=la&la=318&downloadYear=2018-2019&datatypes=spine&datatypes=ks4&datatypes=ks5&datatypes=ks4destination&datatypes=pupdestprov

Download data - GOV.UK. (n.d.). Find and Compare Schools in England. Retrieved April 29, 2022, from https://www.compare-school-performance.service.gov.uk/download-data?currentstep=datatypes®iontype=la&la=306&downloadYear=2018-2019&datatypes=spine&datatypes=ks4&datatypes=ks4prov&datatypes=ks4destination&datatypes=Census

https://www.ons.gov.uk/peoplepopulationandcommunity/housing/datasets/medianpricepaidbywardhpssadataset37

Appendix

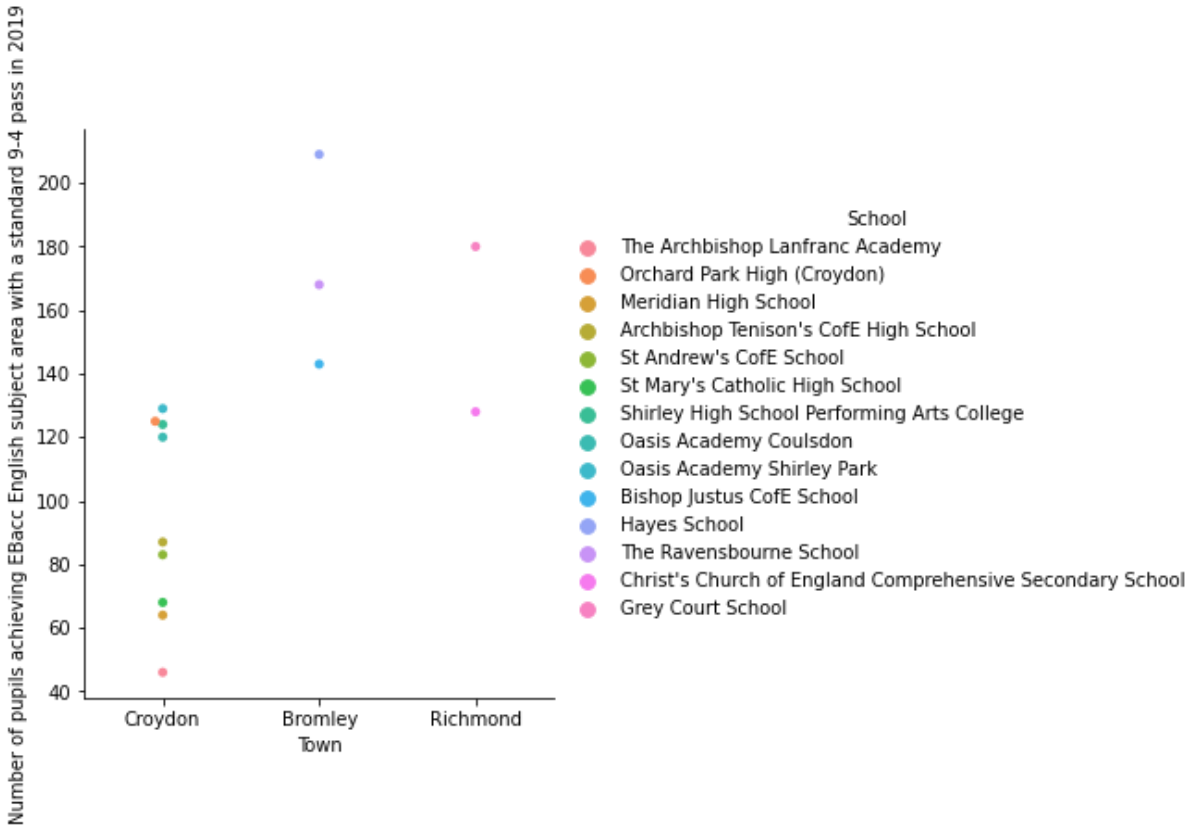

Figure 5 Scatter plot of Students GCSE English score in 2019 against the borough they live in

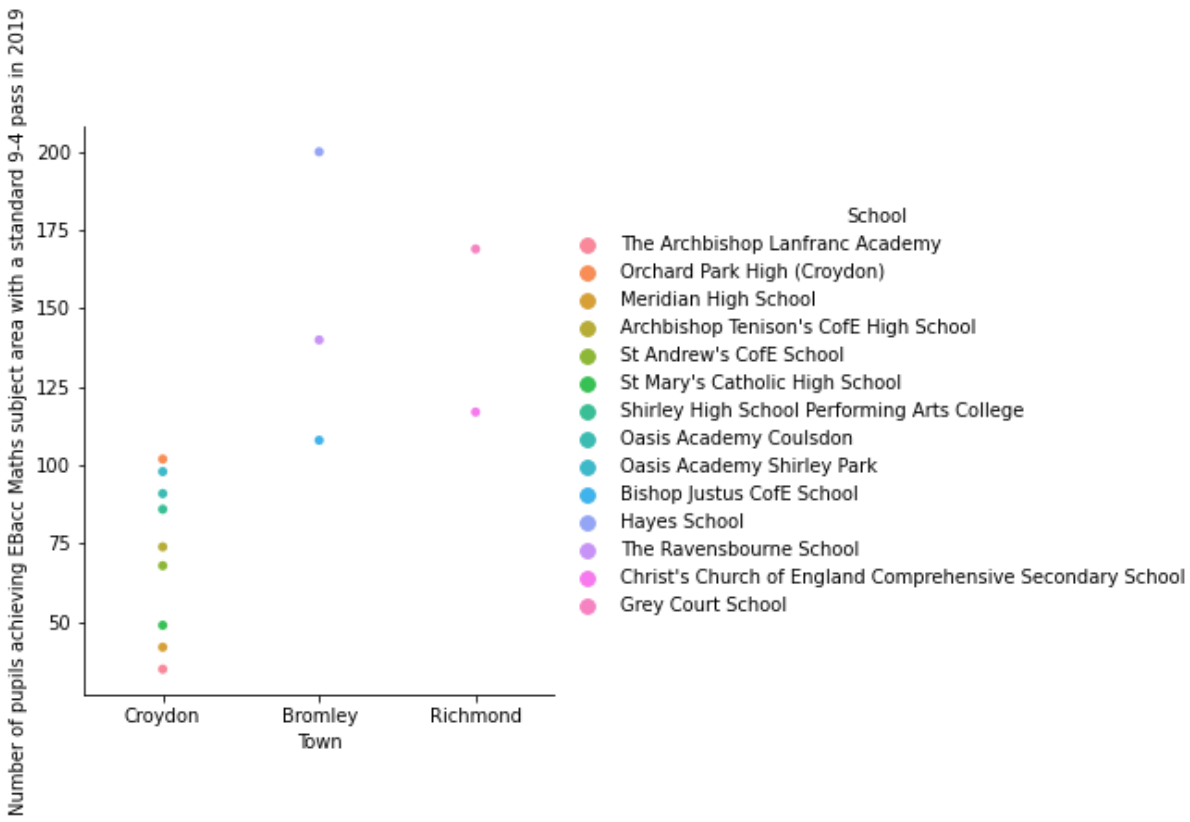

Figure 6 Scatter plot of Students GCSE Maths score in 2019 against the borough they live in

OLS Regression Results

Dep. Variable:	Number of pupils achieving EBacc Maths subject area with a standard 9-4 pass in 2018	**R-squared (uncentered):**	0.893
Model:	OLS	**Adj. R-squared (uncentered):**	0.885
Method:	Least Squares	**F-statistic:**	108.5
Date:	Fri, 29 Apr 2022	**Prob (F-statistic):**	1.12e-07
Time:	15:35:15	**Log-Likelihood:**	-69.873
No. Observations:	14	**AIC:**	141.7
Df Residuals:	13	**BIC:**	142.4
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Average house value in 2018	0.0002	2.16e-05	10.416	0.000	0.000	0.000

Omnibus:	4.131	**Durbin-Watson:**	1.866
Prob(Omnibus):	0.127	**Jarque-Bera (JB):**	1.649
Skew:	0.738	**Prob(JB):**	0.439
Kurtosis:	3.803	**Cond. No.**	1.00

Table 2 Number of pupils achieving EBacc Maths subject area with a standard 9-4 pass in 2018 regressed against House prices in 2018

OLS Regression Results

Dep. Variable:	Number of pupils achieving EBacc Maths subject area with a standard 9-4 pass in 2019	**R-squared (uncentered):**	0.879
Model:	OLS	**Adj. R-squared (uncentered):**	0.869
Method:	Least Squares	**F-statistic:**	94.12
Date:	Fri, 29 Apr 2022	**Prob (F-statistic):**	2.55e-07
Time:	15:53:35	**Log-Likelihood:**	-70.711
No. Observations:	14	**AIC:**	143.4
Df Residuals:	13	**BIC:**	144.1
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Average house value in 2019	0.0002	2.28e-05	9.702	0.000	0.000	0.000

Omnibus:	6.187	**Durbin-Watson:**	1.710
Prob(Omnibus):	0.045	**Jarque-Bera (JB):**	2.969
Skew:	0.992	**Prob(JB):**	0.227
Kurtosis:	4.075	**Cond. No.**	1.00

Table 3 Number of pupils achieving EBacc Maths subject area with a standard 9-4 pass in 2019 regressed against House prices in 2019

OLS Regression Results

Dep. Variable:	Number of pupils achieving EBacc English subject area with a standard 9-4 pass in 2018	**R-squared (uncentered):**	0.900
Model:	OLS	**Adj. R-squared (uncentered):**	0.892
Method:	Least Squares	**F-statistic:**	117.1
Date:	Fri, 29 Apr 2022	**Prob (F-statistic):**	7.14e-08
Time:	15:54:57	**Log-Likelihood:**	-71.219
No. Observations:	14	**AIC:**	144.4
Df Residuals:	13	**BIC:**	145.1
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Average house value in 2018	0.0003	2.38e-05	10.822	0.000	0.000	0.000

Omnibus:	0.270	**Durbin-Watson:**	1.854
Prob(Omnibus):	0.874	**Jarque-Bera (JB):**	0.407
Skew:	0.242	**Prob(JB):**	0.816
Kurtosis:	2.320	**Cond. No.**	1.00

Table 4 Number of pupils achieving EBacc English subject area with a standard 9-4 pass in 2018 regressed against House prices in 2018

OLS Regression Results

Dep. Variable:	Number of pupils achieving EBacc English subject area with a standard 9-4 pass in 2019	**R-squared (uncentered):**	0.904
Model:	OLS	**Adj. R-squared (uncentered):**	0.896
Method:	Least Squares	**F-statistic:**	121.7
Date:	Fri, 29 Apr 2022	**Prob (F-statistic):**	5.68e-08
Time:	15:54:59	**Log-Likelihood:**	-71.397
No. Observations:	14	**AIC:**	144.8
Df Residuals:	13	**BIC:**	145.4
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Average house value in 2019	0.0003	2.4e-05	11.034	0.000	0.000	0.000

Omnibus:	0.530	**Durbin-Watson:**	1.559
Prob(Omnibus):	0.767	**Jarque-Bera (JB):**	0.401
Skew:	0.357	**Prob(JB):**	0.818
Kurtosis:	2.577	**Cond. No.**	1.00

Table 5 Number of pupils achieving EBacc English subject area with a standard 9-4 pass in 2019 regressed against House prices in 2019

	Total number of students	Number of Boys	Number of Girls	Number of pupils at the end of key stage 4 with English as their first language in 2018	Number of pupils at the end of key stage 4 with English as additional language (EAL) in 2018	Number of pupils achieving EBacc English subject area with a standard 9-4 pass in 2018	Number of pupils achieving EBacc Maths subject area with a standard 9-4 pass in 2018	Number of pupils at the end of key stage 4 with English as their first language in 2019	Number of pupils at the end of key stage 4 with English as additional language (EAL) in 2019	Number of pupils achieving EBacc English subject area with a standard 9-4 pass in 2019	Number of pupils achieving EBacc Maths subject area with a standard 9-4 pass in 2019
count	2.000000	2.000000	2.000000	2.00000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000
mean	1097.000000	644.500000	452.500000	144.50000	26.000000	148.000000	138.500000	147.000000	28.500000	154.000000	143.000000
std	342.239682	185.969083	156.270599	21.92031	9.899495	35.355339	28.991378	38.183766	0.707107	36.769553	36.769553
min	855.000000	513.000000	342.000000	129.00000	19.000000	123.000000	118.000000	120.000000	28.000000	128.000000	117.000000
25%	976.000000	578.750000	397.250000	136.75000	22.500000	135.500000	128.250000	133.500000	28.250000	141.000000	130.000000
50%	1097.000000	644.500000	452.500000	144.50000	26.000000	148.000000	138.500000	147.000000	28.500000	154.000000	143.000000
75%	1218.000000	710.250000	507.750000	152.25000	29.500000	160.500000	148.750000	160.500000	28.750000	167.000000	156.000000
max	1339.000000	776.000000	563.000000	160.00000	33.000000	173.000000	159.000000	174.000000	29.000000	180.000000	169.000000

Table 6 Descriptive statistics Richmond

	Number of pupils at the end of key stage 4 with English as their first language in 2018	Number of pupils at the end of key stage 4 with English as additional language (EAL) in 2018	Number of pupils achieving EBacc English subject area with a standard 9-4 pass in 2018	Number of pupils achieving EBacc Maths subject area with a standard 9-4 pass in 2018	Number of pupils at the end of key stage 4 with English as their first language in 2019	Number of pupils at the end of key stage 4 with English as additional language (EAL) in 2019	Number of pupils achieving EBacc English subject area with a standard 9-4 pass in 2019	Number of pupils achieving EBacc Maths subject area with a standard 9-4 pass in 2019
count	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.0000
mean	199.750000	11.750000	167.250000	153.250000	198.750000	11.500000	175.250000	158.2500
std	25.460754	6.238322	25.539186	37.562171	25.914925	8.582929	27.475747	42.0981
min	171.000000	6.000000	139.000000	121.000000	172.000000	4.000000	143.000000	108.0000
25%	186.750000	7.500000	157.000000	121.750000	184.000000	5.500000	161.750000	132.0000
50%	198.000000	10.500000	164.500000	148.500000	195.000000	9.500000	174.500000	162.5000
75%	211.000000	14.750000	174.750000	180.000000	209.750000	15.500000	188.000000	188.7500
max	232.000000	20.000000	201.000000	195.000000	233.000000	23.000000	209.000000	200.0000

Table 7 Descriptive statistics Bromley

	Total number of students	Number of Boys	Number of Girls	Number of pupils at the end of key stage 4 with English as their first language in 2018	Number of pupils at the end of key stage 4 with English as additional language (EAL) in 2018	Number of pupils achieving EBacc English subject area with a standard 9-4 pass in 2018	Number of pupils achieving EBacc Maths subject area with a standard 9-4 pass in 2018	Number of pupils at the end of key stage 4 with English as their first language in 2019	Number of pupils at the end of key stage 4 with English as additional language (EAL) in 2019	Number of pupils achieving EBacc English subject area with a standard 9-4 pass in 2019	Number of pupils achieving EBacc Maths subject area with a standard 9-4 pass in 2019
count	11.000000	10.000000	11.000000	10.000000	10.000000	10.000000	10.000000	11.000000	11.000000	11.000000	11.000000
mean	811.272727	418.700000	430.636364	93.200000	40.000000	95.900000	82.200000	90.454545	41.818182	100.636364	79.727273
std	362.238897	181.950573	278.978233	30.039973	21.055482	35.075949	32.672789	32.809644	26.064605	31.743575	29.570563
min	424.000000	264.000000	152.000000	48.000000	9.000000	47.000000	35.000000	29.000000	8.000000	46.000000	35.000000
25%	542.500000	278.000000	262.000000	73.000000	26.250000	69.250000	59.750000	86.500000	21.000000	75.500000	58.500000
50%	747.000000	380.000000	356.000000	100.000000	44.000000	99.000000	85.500000	101.000000	41.000000	120.000000	86.000000
75%	983.500000	470.750000	474.500000	108.500000	56.750000	120.000000	96.500000	110.500000	62.000000	124.500000	98.500000
max	1686.000000	851.000000	1049.000000	133.000000	71.000000	144.000000	135.000000	128.000000	83.000000	137.000000	133.000000

Table 8 Descriptive statistics Croydon