Understanding London using Open Data

Introduction

This research project aims to investigate the correlative relationship of how the amount of annual household income may consequently affect a child's attainment. This analysis will be conducted by examining the spatial correlative patterns between gross annual household income and GCSE pupil scores in 2007-2008. It has been well documented that a family's household income directly correlates influences the attainment levels of the pupils from that area (Isaacs and Magnuson, 2011; Hirsch, 2007; Blanden and Gregg, 2004). Therefore, a further area of analyse that will be undergone, is the exploration as to if gender of the pupil will affect the results of the correlation.

Many studies access the correlation between a households annual income and a child's attainment, many authors use the term 'School readiness' which characterises how prepared a child is for the socio and academic challenges that school may present. 'School readiness' can comprise of many different variables, therefore it cannot simply be attributed to the income of a pupil's household. However, there is still many authors who believe household conditions and characteristics can be direct attributed to a child, and therefore be a direct influencer of their academic potential (Akee et al., 2010). This is an important area of research because it has been observed that these this positive correlation between the two variables can lead to lower levels of economic success in adulthood and reduced social mobility (Isaacs and Magnusen, 2011; Hirsch, 2007). Therefore, this socio-economic trend is often continued across generations, which is why the spatially distribution of this trend should be significant. This research is particular important in the U.K. as it has been observed it contains a particularly high correlation between low income and reduced levels of academic achievements (Blanden and Gregg, 2004), as well as, being recognised as having one of the greatest socio-economic correlations in education among similar countries (Hirsch, 2007). This is despite the U.K. has a notion of 'opportunity for all', which has been stated by all three major political parties (Gregg, Macmillan, 2010).

This study tested the correlation between household income and child education by using Gross annual household income data (£) as the dependent variable, from which study the correlative differences made from percentage of pupils attaining between 5 A*-C grades in 2007-2008. To investigate potential gender variation, it was chosen to test the boys score, the girls score, and the combined score's (independent variables) correlation with household income.

The data used to perform this analysis were originally from the London Datastore, and are as follows(further information within the MetaData spreadsheet is available):

- The "london-borough-atlas.xlsx" excel spreadsheet – "iadatasheet 5"
  - All Pupils at the End of KS4 Achieving 5+ A* - C Including English and Mathematics

-   Boy Pupils at the End of KS4 Achieving 5+ A*-C Including English and Mathematics
-   Girl Pupils at the End of KS4 Achieving 5+ A*-C Including English and Mathematics
- The "london-borough-atlas.xlsx" excel spreadsheet – "iadatasheet 4"=
    -   Gross Annual Pay - Full Time - Total (£)

Results

Figure 1: Shows the results from the different correlation's performed within R.
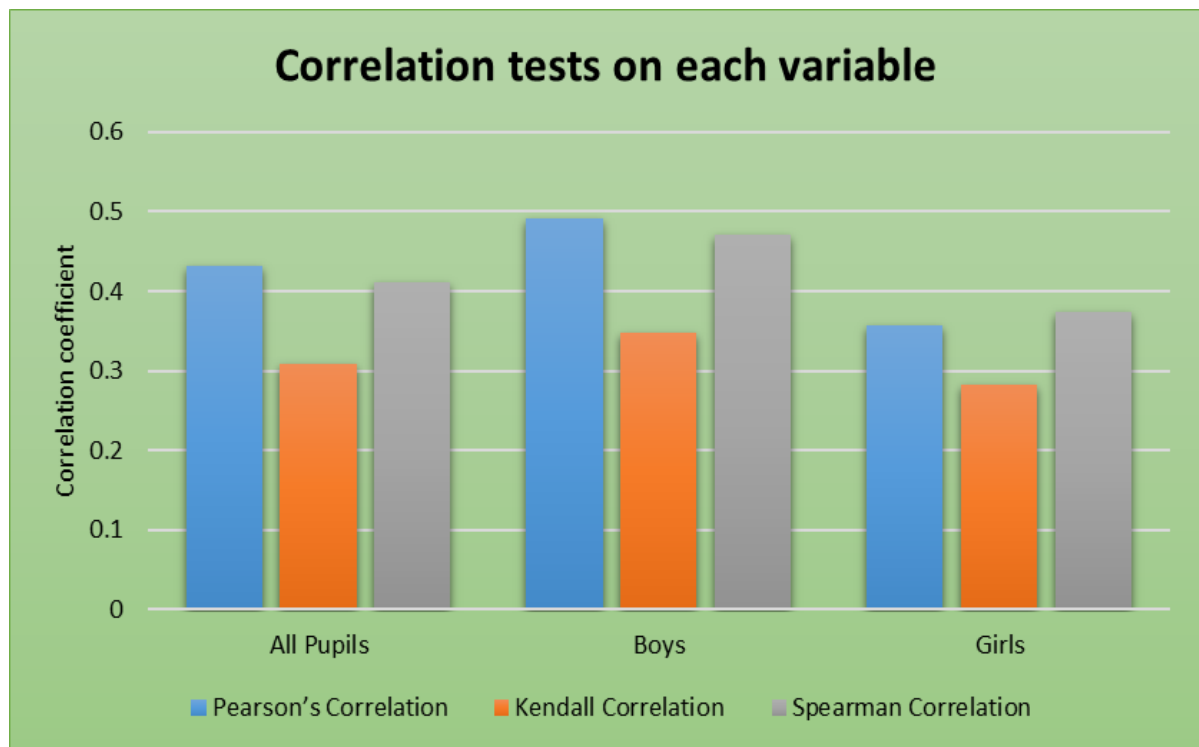
	All Pupils	Boys	Girls
Pearson's Correlation	0.4321613	0.4905545	0.3579569
Kendall Correlation	0.3095929	0.3469196	0.2824646
Spearman Correlation	0.4106979	0.4712961	0.3743628

Figure 1 shows the comparison of using different correlation analysis algorithms, the use of which is dependent on the parameters of your research objectives and the data used. For the purpose of this project the Pearson's correlation coefficient was deemed most suitable. The Pearson's correlation is a statistical measure of the strength of a linear relationship between a pair or set of data. Positive values from the Pearson's correlation coefficient signify a direct positive linear correlation, the closer to 1 or -1, being the stronger the linear correlation.

It can be seen that all the correlation analysis has a moderate significance positive correlation between Annual household income and all three tested variables, ranging from 0.36 with Girls GCSE scores and 0.49 with Boys GCSE scores. Whereas the girls can be seen to have consequently significantly less of a statistical correlation with the dependent variable, the annual household income, only resulting 0.36, a relatively weak correlation coefficient value.

Figure 2: Linear regression statistics for the three study variables when measured again Gross Household Income to show correlative analysis.

Figure 2a: All pupil linear regression statistics.
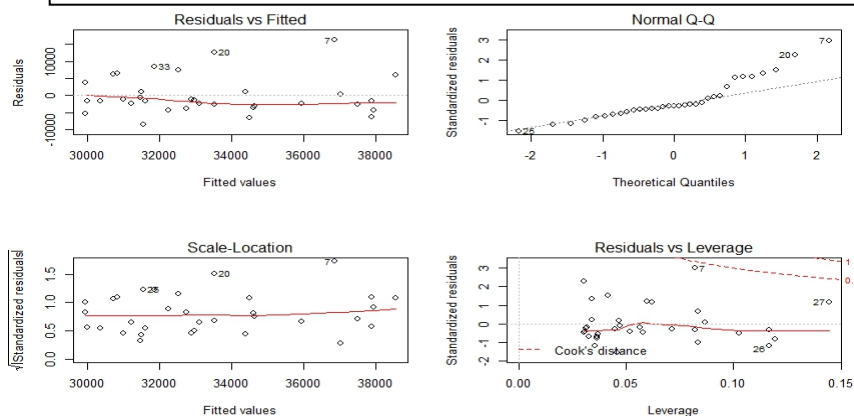

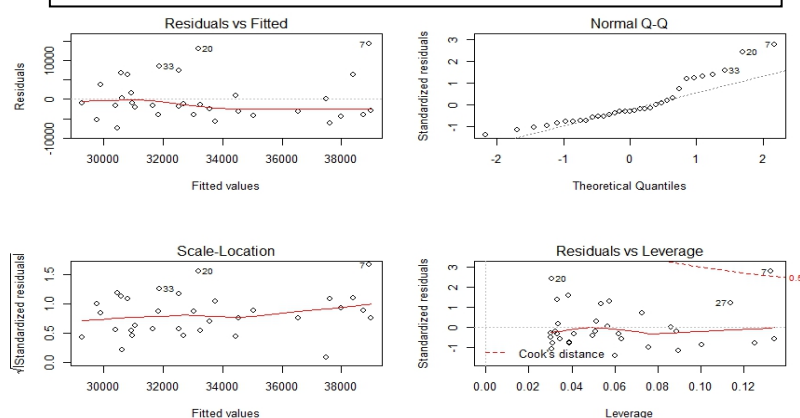Figure 2b: Boys linear regression statistics.

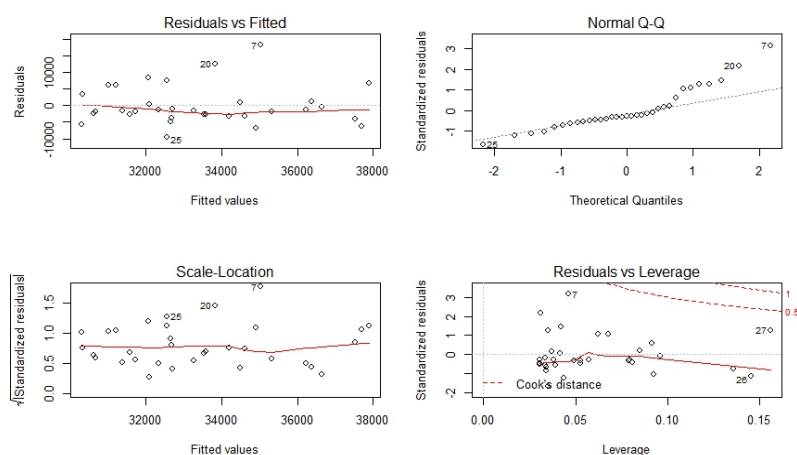Figure 2c: Girls linear regression statistics.

Figure 3: Show the Linear Regression relationship between the three study variables and House hold income. Including 95% confidence interval.

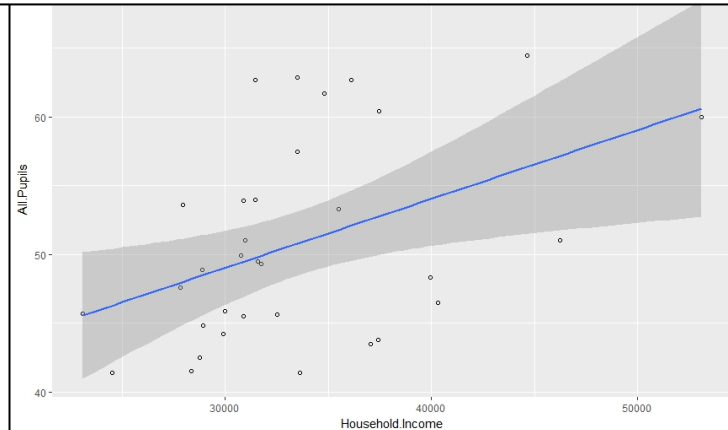Figure 3a: All pupils plotted against the household income.

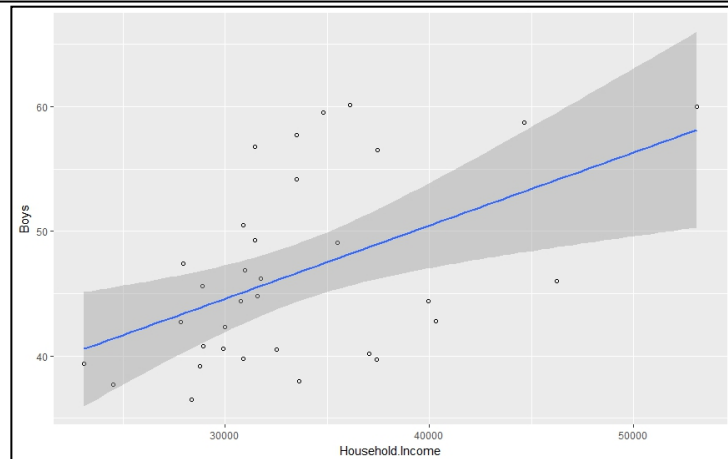Figure 3b: Boy pupils plotted against the household income.

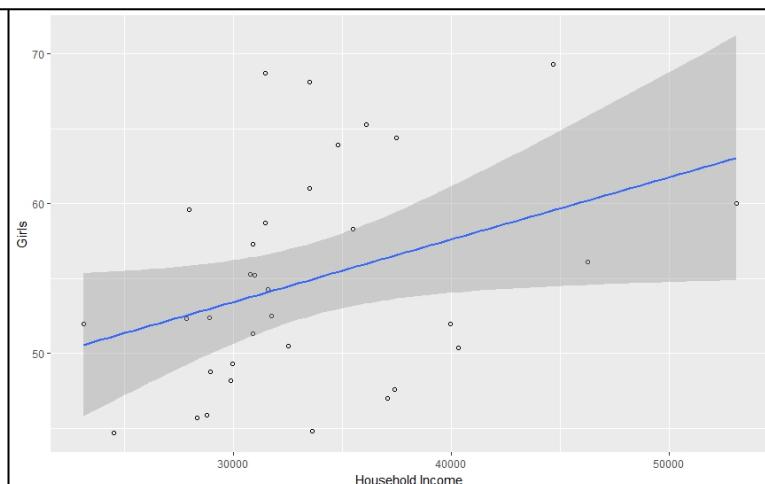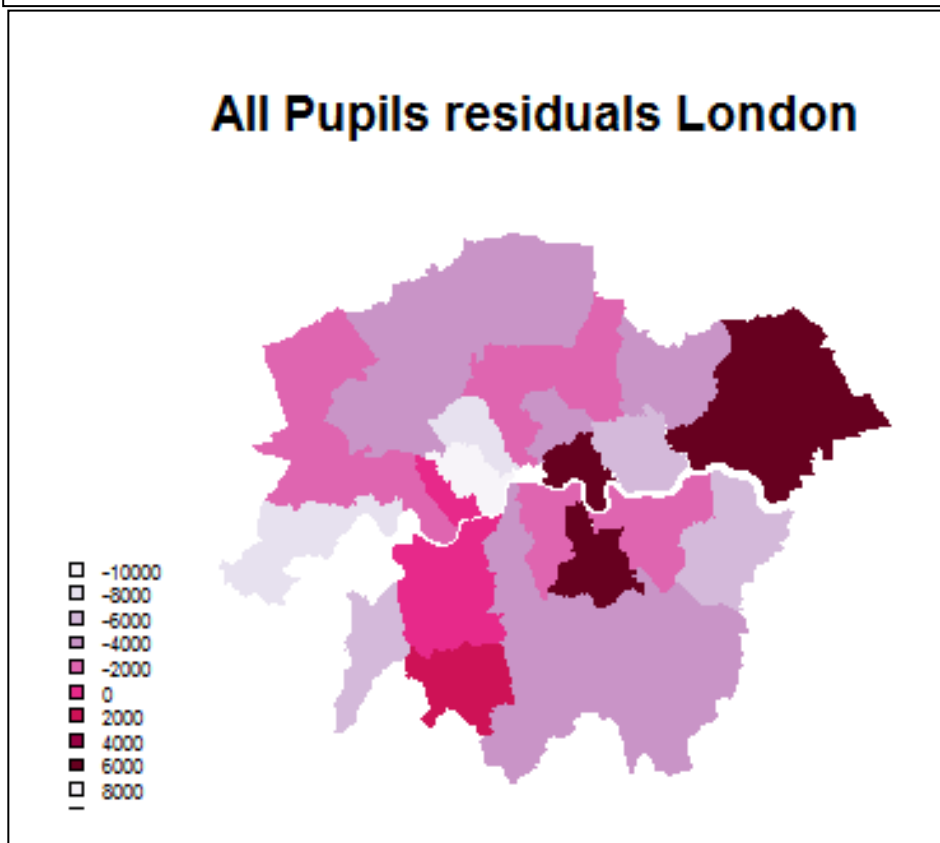Figure 3c: Girl pupils plotted against the household income.

Figure 4: The residuals of all pupils georeferenced in the London Shapefile.

Figures 2 and 3 presents the results of the Linear Regression relationship, with a linear regression line and 95% confidence interval displayed on the plot. The points which lie either side of the regression line (line of best fit) are known as the residuals. The results from the linear regression models (figure 2) coincide with the correlation statistics as shown in figure 1. Where the models such as Normal Q-q and Scale-location both show distribution, where the less normally distributed residuals from the line of best fit, this results in a lower correlation coefficient. An example of this is comparing figure 2 model plots with girl and boy variables, as the boy variable models show far more normally distributed residuals than that of the girl models. This corresponds with figure 1 which states the boy variable having a notably great correlation coefficient.

The results from these linear regression plots coincide with the results from Figure 1. This is as it can clearly be seen that the girls have a significant worse residual distribution around the linear regression line, which can be directly compared to Figure 1 that presents similarly that girls appear to have less of a correlation with the dependent variable of annual household income. This can be directly compared to the boy's variable data which has a notably more well-distributed points, which is correlated to the correlation results as presented in figure 1. Also, there is no apparent geographic pattern to the residual map (figure 4), and therefore it can be inferred that there is unlikely to be other unobserved variables influencing our dependent variable in the linear regression model.

Discussion

Figure 1 shows the results of the correlation statistical analysis, where each test variable was correlated with the household annual income for 2007 – 2008. When considering the Pearson's correlation coefficient (this studies chosen correlation algorithm), it can be noted that as the boys has noticeably higher correlation coefficients for all correlations tested (0.49), it can be attributed that Boys GCSE results from 2007 - 2008 were more statistically dependent with the annual household income from 2007 – 2008. The results from this data can be attributed to many reasons as there is no way to accurately infer that whether household income is a direct influence on child's test due to too many other potential variables being influenced. However, it can be noted that the there is a relationship of some significance, and therefore it is possible to suggest that income could be, as well as many others, a key influencer on a pupils academic potential.

One potential reason as to why the results suggest a stronger correlation, or dependence, from boys GCSE results to household income can be accredited to girls being more dependent than boys in their use for ICT (Information and Communications Technology)(Becta, 2008). Becta (2008) go on to state that their studies found males use this form of education research much at home. Therefore I can infer that depending on the economic - which often reflects other household variables - prosperity of a household, a male pupil would be likely to more greatly affect by this factor.

References

Akee, R., Copeland, W., Keeler, G., Angold, A. and Costello, E. (2010). Parents' Incomes and Children's Outcomes: A Quasi-Experiment Using Transfer Payments from Casino Profits. American Economic Journal: Applied Economics, 2(1), pp.86-115.

Becta. How do boys and girls differ in their use of ICT?. (2008). [online] Becta. Available at: http://dera.ioe.ac.uk/8318/1/gender_ict_briefing.pdf [Accessed 7 Jan. 2018].

Blanden, J. and Gregg, P. (2004). Family Income and Educational Attainment: A Review of Approaches and Evidence for Britain. Oxford Review of Economic Policy, 20(2), pp.245-263.

Hirsch, D. (2007). Experiences of poverty and educational disadvantage. Joseph Rowntree Foundation. [online] Available at:

GREGG, Paul; MACMILLAN, Lindsey. Family income,education and cognitive ability in the next generation: exploring income gradients in education and test scores for current cohorts of youth. **Longitudinal and Life Course Studies**, [S.l.], v. 1, n. 3, p. 259-280, july 2010. ISSN 1757-9597. Available at: <http://journals.sfu.ca/llcs/index.php/llcs/article/view/95/86>. Date accessed: 12 jan. 2018.

doi:http://dx.doi.org/10.14301/llcs.v1i3.95.https://www.jrf.org.uk/sites/default/files/jrf/migrated/files/2123.pdf [Accessed 4 Jan. 2018].

Isaacs, J. and Magnuson, K. (2011). Income and Education as Predictors of Children's School Readiness. [online] Available at: https://www.brookings.edu/research/income-and-education-as-predictors-of-childrens-school-readiness/ [Accessed 4 Jan. 2018].