

Spatial Statistics
Assignment No 4
OLS Regression Analysis

Muhammad Bilal
Matriculation: - 12214473

Q.No.1

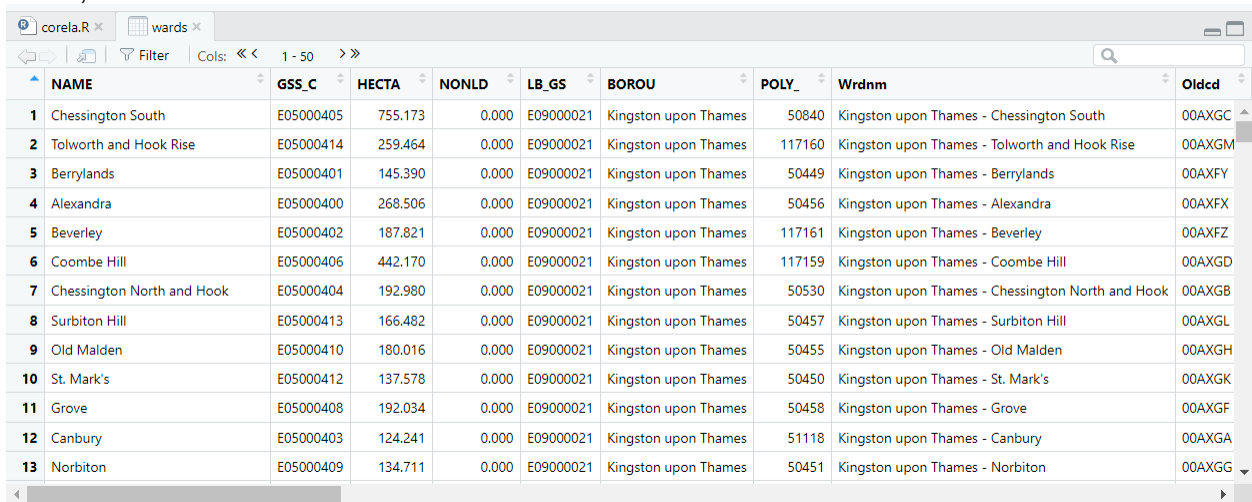
Investigate the effect of the explanatory variable Median Age 2013 (Mda_2013) on the dependent variable Average GCSE Score (AGc_2) by means of a linear regression model.

Ans: -

Loading Libraries and Adding London Ward Data save in ward variable.

```
# Loading Libraries
library(sf)
library(ggplot2)
library(dplyr)
#Load London Data save in Ward variable
wards <- sf::st_read("LondonWards.shp")
```

Ward Data looks like this that contains different information related to wards in London like their name, GCSE score and etc.



	NAME	GSS_C	HECTA	NONLD	LB_GS	BOROU	POLY_	Wrdnm	Oldcd
1	Chessington South	E05000405	755.173	0.000	E09000021	Kingston upon Thames	50840	Kingston upon Thames - Chessington South	00AXGC
2	Tolworth and Hook Rise	E05000414	259.464	0.000	E09000021	Kingston upon Thames	117160	Kingston upon Thames - Tolworth and Hook Rise	00AXGM
3	Berrylands	E05000401	145.390	0.000	E09000021	Kingston upon Thames	50449	Kingston upon Thames - Berrylands	00AXFY
4	Alexandra	E05000400	268.506	0.000	E09000021	Kingston upon Thames	50456	Kingston upon Thames - Alexandra	00AXFX
5	Beverley	E05000402	187.821	0.000	E09000021	Kingston upon Thames	117161	Kingston upon Thames - Beverley	00AXFZ
6	Coombe Hill	E05000406	442.170	0.000	E09000021	Kingston upon Thames	117159	Kingston upon Thames - Coombe Hill	00AXGD
7	Chessington North and Hook	E05000404	192.980	0.000	E09000021	Kingston upon Thames	50530	Kingston upon Thames - Chessington North and Hook	00AXGB
8	Surbiton Hill	E05000413	166.482	0.000	E09000021	Kingston upon Thames	50457	Kingston upon Thames - Surbiton Hill	00AXGL
9	Old Malden	E05000410	180.016	0.000	E09000021	Kingston upon Thames	50455	Kingston upon Thames - Old Malden	00AXGH
10	St. Mark's	E05000412	137.578	0.000	E09000021	Kingston upon Thames	50450	Kingston upon Thames - St. Mark's	00AXGK
11	Grove	E05000408	192.034	0.000	E09000021	Kingston upon Thames	50458	Kingston upon Thames - Grove	00AXGF
12	Canbury	E05000403	124.241	0.000	E09000021	Kingston upon Thames	51118	Kingston upon Thames - Canbury	00AXGA
13	Norbiton	E05000409	134.711	0.000	E09000021	Kingston upon Thames	50451	Kingston upon Thames - Norbiton	00AXGG

Figure 1

Visualizing Ward data based on Average GCSE Score

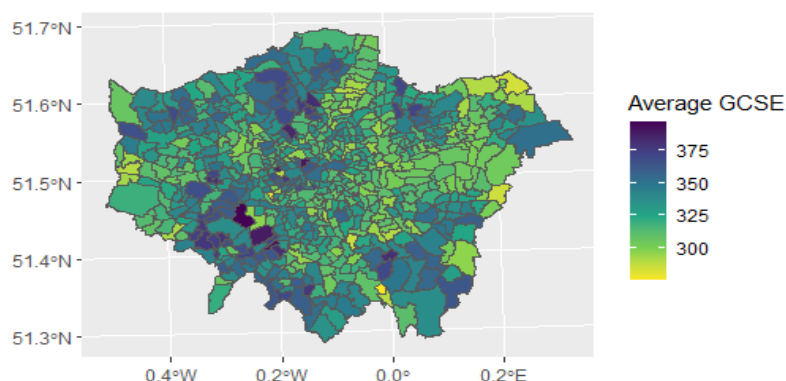


Figure 2

Implementing Linear Regression Model on to see correlation between median Age 2013 and Average GCSE score

```
#Apply Linear Regression Method to see correlation between two parameters
ggplot(data = wards, aes(x = MdA_2013, y = AGC_2)) +
  geom_point() +xlab("Median Age - 2013") +
  ylab("Average GCSE - 2014") +
  geom_smooth(method=lm , color="red", fill="#69b3a2", se=FALSE) +
  theme_minimal()
```

Output of the Model as shown below

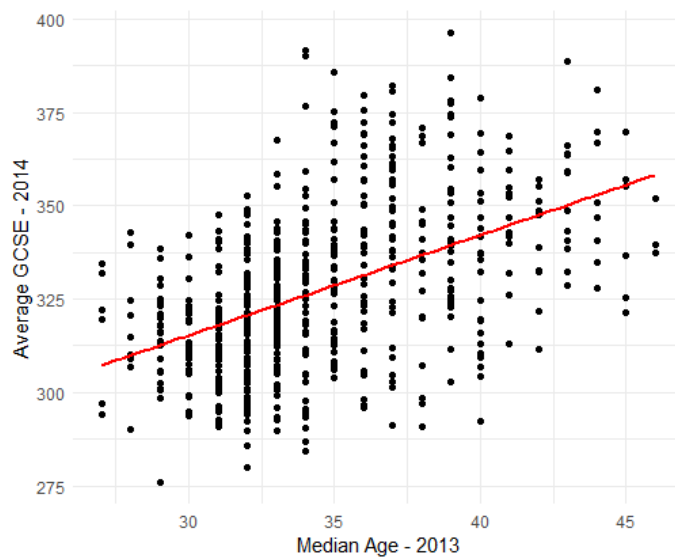


Figure 3

Summary of the Model

Residuals:

Min	1Q	Median	3Q	Max
-49.834	-12.915	-1.517	11.265	65.587

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	234.6639	6.6078	35.51	<2e-16 ***
MdA_2013	2.6867	0.1904	14.11	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.8 on 624 degrees of freedom

Multiple R-squared: 0.2419, Adjusted R-squared: 0.2407

F-statistic: 199.2 on 1 and 624 DF, p-value: < 2.2e-16

Figure 4

Interpretation: -

The output of the model shows positive correlation between Average GCSE Score and Median Age because as the value of median age increase on the other hand GCSE score also increases as shown in Figure 3. R squares show the percentage of variation explained by the relationship between two variables so in our case 24% of variation in Average GCSE Score can be explained by the variable Median Age. Less P value indicates more confidence depending on the amount of data. This model shows that there is a significant relationship between these variables in the ward dataset.

Q No 2: -

Find a causal explanation for the observed correlation. Does your regression model show causality? What other variables could influence the dependent variable. Explain in a few sentences.

Ans: - Through Causal explanation we can explain the cause of increase or decrease of dependent variable values based on the independent variable that can be determined by different experiments or research on the other end correlation only shows that positive or negative relation between variable does not show the cause means what derive those variables to change. So, our model does not show causality because we don't know the reason why average GCSE score increase with median age increase its only show data relation.

other variables that could influence the Average GCSE variable would require correlation matrix of each variable with GCSE Score. Generally, in social context, there can be multiple factors that influence a dependent variable. Some potential variables that may influence Average GCSE population density, Crime rate, employment ratio, education and many more.