Data Essay

Table of Contents

Legend

Terms that have been abbreviated for convenience:

GE = 2017 U.K. General Election
Economic Activity = Percentage of economically active population within British Constituencies as of the 2011 Census
House Ownership = Percentage of house-owning population within British Constituencies as of the 2011 Census.

Introduction

Trends in voter behaviour have played a central role in the study of elections cross-nationally. In this regard, Pulzer famously stated that "class is the basis of British party politics[1]". This a notion extensively supported by the political science literature. Most notably, Andersen et al. highlights how the majority of research has rested on explanations regarding "the individual's class membership and his or her vote[2]". In conducting his research, Andersen found that voting in the British General Elections[3] was affected by class at a constituency level and not at the level of the individual – with the Conservative vote dominating upper-class constituencies and the Labour vote dominating working-class constituencies. Inspired by these findings, my essay will attempt to draw a causal link between wealth and employment as indicators of socioeconomic status within constituencies, and trends in Conservative vote share in the 2017 General Election. To test a possible relationship between the former, being the determinants and the latter, being the dependent variable, a regression model will be implemented. I will then seek to improve my first model through a second - which will feature an analysis of the determinants' interaction term and its implications on the model.

Data

I have chosen to rely on the 2017 British Election Study Result's file to conduct my research[4]. In total it includes 632 observations. The BES internet panel data is

---

[1] Pulzer, Peter (1967). *Political Representation and Elections in Britain.* London: Allen and Unwin

[2] Andersen, R., & Heath, A. (2002). Class Matters: The Persisting Effects of Contextual Social Class on Individual Voting in Britain.

[3] Between 1964 and 1997

[4] "2017 BES Constituency Results With Census And Candidate Data - The British Election Study". 2019. *Britishelectionstudy.Com*. https://www.britishelectionstudy.com/data-object/2017-bes-constituency-results-with-census-and-candidate-data/.

collected by YouGov using random probability sampling[5]. This means that every unit in the population has a chance of being selected for the sample; and the probability of selection for any unit in the population is either known or could be calculated.

Methodology

According to the Office for National Statistics, significant indicators of socioeconomic status include employment and wealth[6]. To capture these measures, I used variables from the BES dataset to act as proxies of the indicators themselves. In this regard, I derived information on employment from the c11EconomicActive variable – mapping out the population of economically active individuals within British constituencies as of the 2011 Census. Moreover, I used the c11HouseOutright variable – measuring the population of house owners within British constituencies, as a proxy for wealth. **In turn, c11EconomicActive and c11HouseOutright acted as my independent variables. Conversely, the variable Con17 – measuring Conservative vote share in the 2017 GE, was selected as my dependent variable**. All the variables used in my analysis have not been recoded and have thus remained as in the original BES dataset.

Formalised hypotheses

I have formulated the hypotheses for my first model in the following manner:

Model 1: Effect of Economic Activity and House Ownership on 2017 GE Conservative Vote Share

Null Hypotheses:

- $H_{0(1)}$ = There is no relationship between Conservative vote share and percentage of the economically active population in a constituency while holding the percentage of house owners constant.
- $H_{0(2)}$ = There is no relationship between Conservative vote share and the percentage of the house-owning population in a constituency while holding the percentage of the economically active population constant.

---

[5] How do you gather your data? - The British Election Study. (2019). Retrieved 8 December 2019, from https://www.britishelectionstudy.com/qa_faqs/how-do-you-gather-your-data/

[6] Socioeconomic inequalities in England and Wales - Office for National Statistics. (2019). Retrieved 10 December 2019, from https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/articles/measuringsocioeconomicinequalitiesinavoidablemortalityinenglandandwales/latest

Alternative Hypotheses:

- $H_{A(1)}$ = There is a positive relationship between Conservative vote share and the percentage of the economically active population in a constituency while holding the percentage of house owners constant.
- $H_{A(2)}$ = There is a positive relationship between Conservative vote share and percentage of the house-owning population in a constituency while holding the percentage of the economically active population constant.

Regression Model I

<div align="center">

Table 1

	Dependent variable:
	Conservative Vote Share (2017 General Election)
Outright House Ownership (Census 2011)	1.234***
	(0.050)
Economic Activity (Census 2011)	1.697***
	(0.099)
Constant	-113.695***
	(7.042)
Observations	632
R^2	0.598
Adjusted R^2	0.597
Residual Std. Error	9.770 (df = 629)
F Statistic	468.441*** (df = 2; 629)
Note:	*p<0.1; **p<0.05; ***p<0.01

</div>

Table 1: Regression Model I

Interpreting the Model

Table 1 depicts a multivariate regression model – testing for the effect of house ownership and economic activity on the 2017 GE Conservative vote share.

<u>Constant term:</u>

The constant represents the estimated average proportion of Conservative vote share in the 2017 GE without the effect of house ownership and economic activity (holding all other variables constant).

This figure is listed as -113.70. However, this is not realistically possible as a constituency cannot have not had negative Conservative vote share and is thus not of much significance.

Coefficients:

Regression coefficients are estimates of the unknown population parameters and describe the relationship between a predictor variable and the response[7].

As shown in Table 1, for every unit increase in the proportion of house ownership within constituencies, the expected proportion of Conservative vote share increased by 1.23 on average, adjusting/controlling for economic activity.

In a similar light, the respective coefficient for economic activity shows that for every unit increase in the proportion of economic activity within constituencies, the expected proportion of Conservative vote share increased by 1.68 on average, adjusting/controlling for house ownership.

As the coefficients were both positive values, this suggests that there is a positive causal relationship between conservative vote share in the 2017 GE and the proportion of house ownership and economic activity in constituencies at the 95% confidence level.

R^2/Adjusted R^2 value:

In mathematical terms, the R^2 is a statistical measure that represents the proportion of the variance for a dependent variable explained by a(n) independent variable(s) in a regression model[8].

In the context of Model I, this can be interpreted as the following: house ownership and economic activity explains 60% of the variance in Conservative vote share.

However, as stated by Gujarati, the R^2 value cannot always be trusted in multivariate regression models[9]. This is because every time a predictor is added, the R^2 increases. Thus the adjusted R^2 should rather be used as it has been adjusted for the number of predictors in the model.

This is not an issue in Model I as there is a very insignificant discrepancy between the R^2 and Adjusted-R^2 (0.598 vs 0.597).

---

[7] Frost, J. (2019). Regression coefficients - Statistics By Jim. Retrieved 8 December 2019, from https://statisticsbyjim.com/glossary/regression-coefficient/
[8] Gujarati N. Damodar (2009). *Basic Econometrics*. New York. The McGraw-Hill Companies
[9] Ibid

F-value:

The F-statistic tests the null hypothesis that *all* of the regression coefficients are equal to zero[10]. The further the F-statistic is from 1 the better it is. As shown in Table 1, the F-statistic of 468 thus indicates that there is a statistically significant relationship between house ownership, economic activity and Conservative vote share and the null hypothesis that the regression coefficients are equal to zero is refuted. Furthermore, it also indicates that the R^2 value can be trusted.

Summation of findings:

Thus, we can conclude from the results illustrated in this model suggest that the variables of house ownership and economic activity independently share a **positive and statistically significant causal relationship** with the Conservative vote share in the 2017 GE at the **95% confidence level**.

Therefore, these results refute the claims $H_{0(1)}$ and $H_{0(2)}$ and support $H_{A(1)}$ and $H_{A(2)}$ made in the 'Hypotheses' section.

Diagnostic plots

Diagnostic tests are of paramount importance in assessing the validity of a model. Without exploring these tests, the results of regression models lose all significance as the results may be skewed or even totally invalid.

In this regard, Figure 6, 7 and 8 act as references to the assumption tests I will conduct for Model I.

---

[10] Frost, J. (2019). Regression coefficients - Statistics By Jim. Retrieved 8 December 2019, from https://statisticsbyjim.com/glossary/regression-coefficient/

Residuals vs Fitted

Fitted values
lm(data$Con17 ~ data$c11HouseOutright + data$c11EconomicActive)
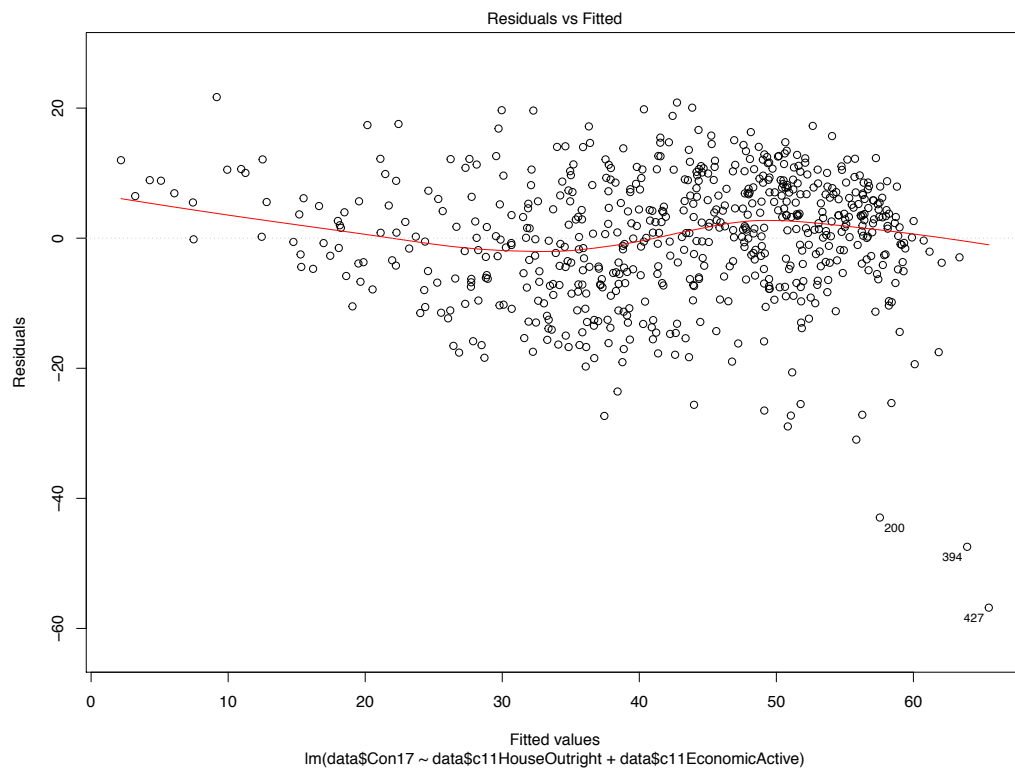
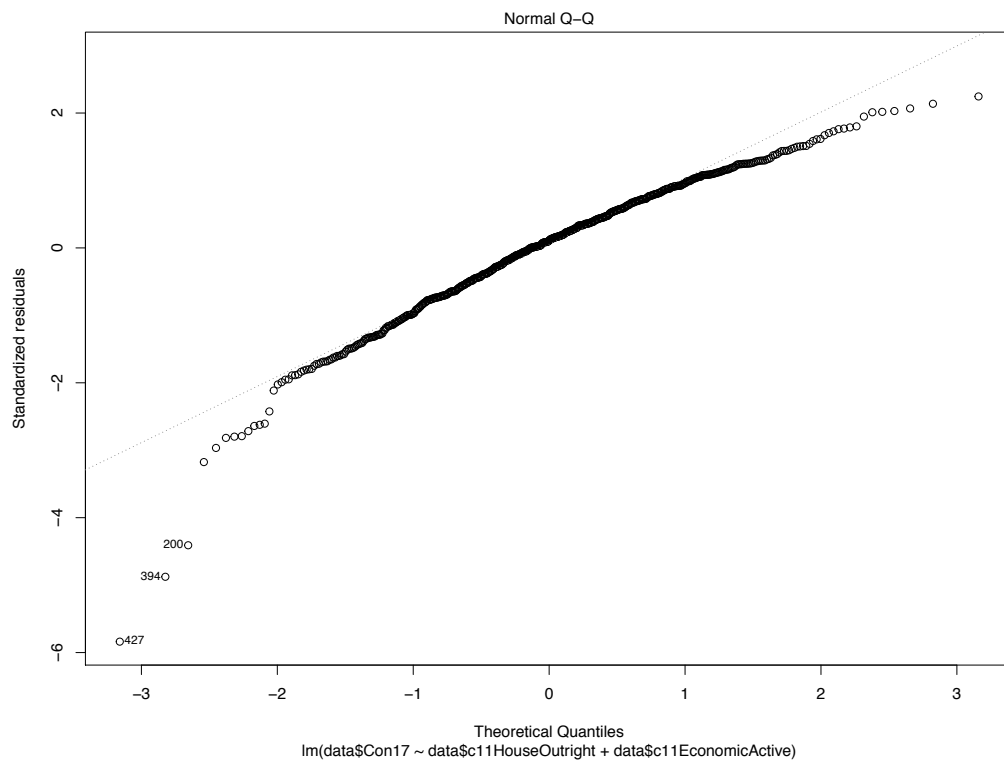*Figure 1: Residuals vs Fitted Plot of Regression Model I*
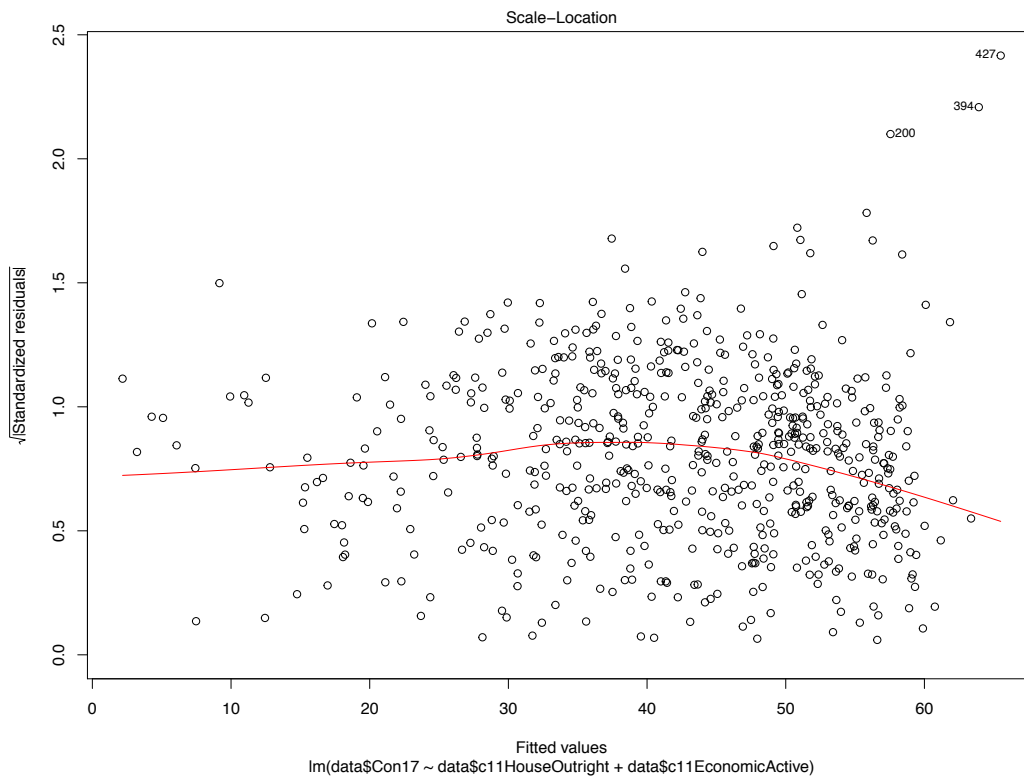
Figure 2: Q-Q Plot of Regression Model I

Figure 3: Scale-Location Plot of Regression Model I

Testing the Assumptions

Assumption 1: Linearity

The assumption of linearity stipulates that the coefficients are linear in the parameters. This diagnostic assessment examines whether the regression model is fundamentally biased or if the coefficients are biased in some way.

In examining the residuals vs. fitted plot in Figure 6, it is clear to see that the residuals roughly form a horizontal band around the 0 line – suggesting that the variance of the error terms are equal to a large extent. **Based on these findings the assumption of linearity is not violated in the model.**

Assumption 2: Covariance of x and u is zero

This diagnostic assessment tests whether the x-values are independent and are not correlated with the error. If violated, the assumption stipulates that all coefficients in the model are fundamentally biased.

To check this, I have conducted a correlation test mapping the association between the residuals and fitted values of the model.

Pearson's correlation coefficient:

	Residuals vs. fitted values
t-statistic	-2.2258e-15
p-value	1
correlation coefficient	-8.86789e-17

As the p-value is greater than 0.05 and the correlation value is statistically close to 0, the null hypothesis is held and this assumption is not violated.

Assumption 3: Mean of the residuals is zero

In a self-explanatory fashion, this assumption holds that the mean of residuals in x should be 0. If violated, the assumption asserts that all coefficients are fundamentally biased. I have conducted a one-sample t-test to check this assumption.

One-sample t-test:

	Residuals of Model I
t-statistic	-7.6661e-16
p-value	1
mean of x	-2.974639e-16

As the p-value is greater than 0.05 and the mean of *x* is statistically close to 0, the null hypothesis is held and this assumption is not violated.

Assumption 4: Variance of the Residual Error Term is Constant

This assumption tests for homoscedasticity – the condition in which the variance of the residual/error term, in a regression model is constant[11]. If the assumption is violated, the standard errors are seen as biased and the p-values cannot be trusted.

When checking the Scale-Location Plot in Figure 8, it is evident that there is no discernible pattern in the distribution of residuals and that they are roughly clustered around 0. **Thus, the variance of the residuals is homoscedastic and the assumption is not violated.**

Assumption 5: *X* Observations are Independent

In a self-explanatory fashion, this assumption holds that every observation, being independent, carries information that cannot be inferred, wholly or partly, by any other observation in the sample.

This is shown in Figure 8 by the fact that the residuals in *x* are dispersed randomly with no discernible pattern with the mean roughly forming a horizontal band around 0.

Assumption 6: The Number of Observations Must Be Greater Than the Number of Parameters to be Estimated

This assumption asserts that the sample size (*n*) of the observations must be greater than the number of parameters (independent variables) to be estimated. Here, the rule of thumb' is that there should be at least 30 observations per parameter (not including the intercept)[12]. If the assumption does not hold, there could be a significant relationship between the variables of interest, but there is not nearly enough data to observe it.

Model I makes use of two parameters, i.e.: proportion of house ownership and economic activity. Furthermore, the sample size in the 2017 BES dataset is 632.

Thus, there are over 300 observations for every parameter and the assumption is not violated.

---

[11] Homoscedasticity. (2019). Retrieved 8 December 2019, from https://www.investopedia.com/terms/h/homoskedastic.asp

[12] Gujarati N. Damodar (2009). *Basic Econometrics*. New York. The McGraw-Hill Companies

Assumption 7: *X* Must Vary and Outliers Do Not Affect Data

This assumption is layered in its stipulations. Firstly it states that the distribution of observations in the independent variables should be varied. Secondly, it asserts that outliers in the parameters do not affect the data. If violated, there could be a significant relationship between the variables of interest, but there is not nearly enough data to observe it.

As shown by the histograms in Figure 2 and 3 already discussed, there is extensive variation of observations within the distributions of the variables of economic activity and house ownership. This is shown by the fact that both variables do not have a majority of observations lying on one particular region of the *x*-axis.

Effect of outliers on data:

	House Ownership
Outliers identified	7 from 632 observations
Proportion (%) of outliers	1.10759493670886
Mean of the outliers	8.46111225329303
Mean without removing outliers	31.0540690673224
Mean if we remove the outliers	31.3071101836395

	Economic Activity
Effect of outliers	"Nothing changed" (output returned)

As shown above there is no significant change when considering outliers in the parameters. This is illustrated most effectively in comparing the mean with and without outliers.

Assumption 8: The Residuals Are Normally Distributed

This assumptions stipulates that the residuals should be normally distributed. To check this, the Q-Q-plot – plotting the residuals (*y*) against a perfectly normal distribution (*x*), is used. If the assumption does not hold, the standard errors in the model are seen as biased and the p-values cannot be trusted.

Figure 7 displays the Q-Q plot for Model I. Although the residuals in the lower tail stray from the normal distribution, the coincide with the outliers in the model. Upon further inquiry, it is clear to see that the majority of residuals perfectly align with the normal distribution. **Thus this assumption is not violated and the p-values and standard errors can be trusted in the model.**

Assumption 8: The Residuals Are Normally Distributed

This assumptions stipulates that the residuals should be normally distributed. To check this, the Q-Q-plot – plotting the residuals (y) against a perfectly normal distribution (x), is used. If the assumption does not hold, the standard errors in the model are seen as biased and the p-values cannot be trusted.

Figure 7 displays the Q-Q plot for Model I. Although the residuals in the lower tail stray from the normal distribution, the coincide with the outliers in the model. Upon further inquiry, it is clear to see that the majority of residuals perfectly align with the normal distribution. **Thus this assumption is not violated and the p-values and standard errors can be trusted in the model.**

Assumption 9: Checking for Multicollinearity

Multicollinearity is a state of very high intercorrelations or inter-associations among the independent variables[13]. It is therefore a type of disturbance in the data, and if present in the data, the statistical inferences made about the data may not be reliable.

For example, the greater the multicollinearity, the greater the standard errors.

In checking for multicollinearity, the t-values and f-value should be found as significant, and correlation between the independent variable coefficients should not exceed 0.7[14] .

Correlation table: % of House Ownership vs % of Economic Activity:

Variables	% of House Ownership	% of Economic Activity
% of House Ownership	1	0.0289277
% of Economic Activity	0.0289277	1

T-values in Model I:

Variable	T-value
Intercept (Con17)	-16.14
% of House Ownership	24.89
% of Economic Activity	17.08

As shown above there is no significant change when considering outliers in the parameters. This is illustrated most effectively in comparing the mean with and without outliers.

---

[13] Multicollinearity - Statistics Solutions. (2019). Retrieved 8 December 2019, from https://www.statisticssolutions.com/multicollinearity/
[14] Gujarati N. Damodar (2009). *Basic Econometrics*. New York. The McGraw-Hill Companies

As seen from the tables above, our f-value and t-values are significant, and correlation coefficients are less than 0.7. **Thus, we can assert that the assumption of no multicollinearity is not violated.**

Assumption 10: Checking for Specification Bias

As stated by Zeileis, "omitted variable bias arises when regressors correlate with an omitted variable." For omitted variable bias to arise, two things must be true: at least one of the included regressors must be correlated with the omitted variable and the omitted variable must be a determinant of the dependent variable[15].

When checking for regressors that were correlated with the parameters chosen, the relationship between the % of 65-to-74-year-olds in constituencies and house ownership was the most notable with a correlation of 89%.

However the second condition did not hold true as the correlation between Conservative vote share and % of 65-to-74-year-olds in constituencies was not notably strong – falling below the 0.7 benchmark. Thus, the age group variable cannot be a determinant for conservative vote share and there it is unlikely that the assumption is violated.

Regression Model II: Addition of Interaction Term (House Ownership x Economic Activity)

In formulating a model that could possibly improve the original, I decided to explore the interaction between house ownership and economic activity. As affirmed by Doling, "throughout the past, home ownership has been associated with the economic prosperity of full employment[16]". Moving a step further from this, he continues in stating that "home ownership appeals particularly to the increasing numbers of professional employees and non-manual workers[17]".

In this light, given that the literature holds that more affluent groups vote Conservative in the U.K[18]. – and that there is a significant link between more profitable, professional employment and house ownership, it follows that Conservative vote share could be more aptly accounted for in the interaction of house ownership and economic activity.

---

[15]Zeleis, A. (2019). Omitted Variable Bias | Introduction to Econometrics with R. Retrieved 8 December 2019, from https://www.econometrics-with-r.org/6-1-omitted-variable-bias.html

[16] Doling, J., & Ronald, R. (2010). Home ownership and asset-based welfare. *Journal Of Housing And The Built Environment*, *25*(2), 165-173.

[17] Ibid

[18]British Election Study (Wave 4, March 2015)

Visualising the interaction

Estimated Coefficient for Outright House Ownership on Conservative Vote Share by Level of Economic Activity
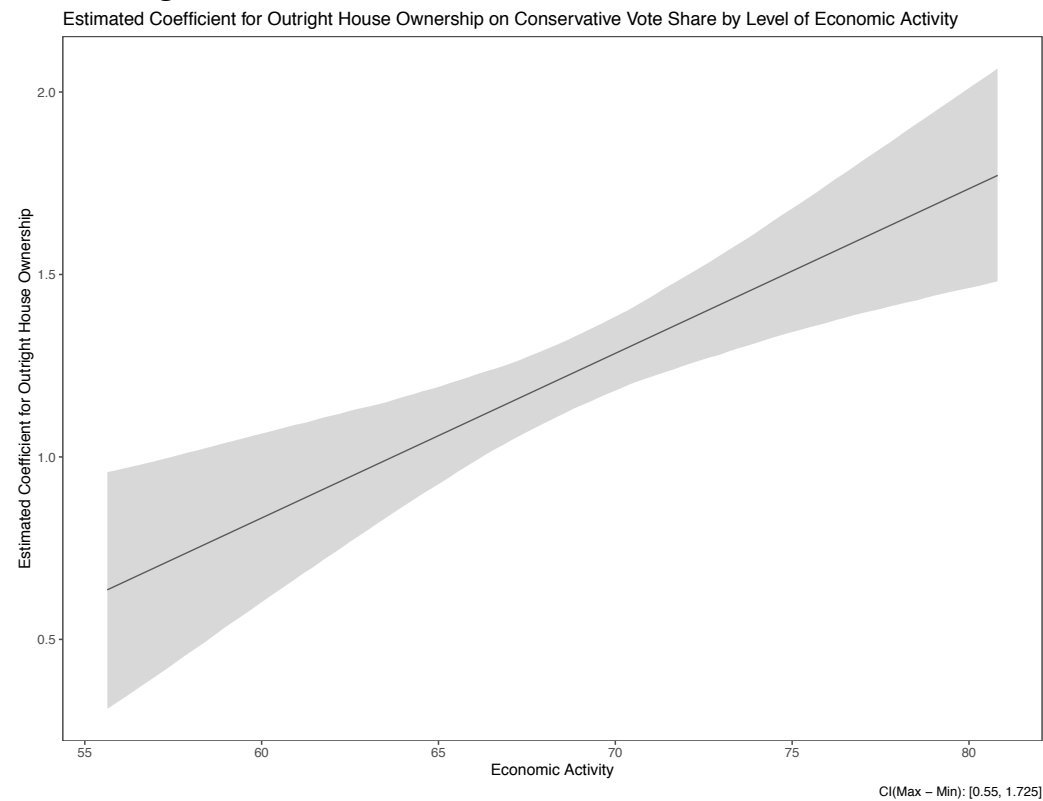

CI(Max – Min): [0.55, 1.725]

Figure 4: Interaction between House Ownership and Economic Activity

Figure 9 visualises the changes in the coefficient of house ownership in a two-way interaction with economic activity. The plot also includes simulates the 95% confidence intervals of these coefficients (specified in the bottom right-hand corner). The graph ultimately shows us that with increasing economic activity, the magnitude of house ownership correspondingly increases.

Interpreting Model II

Table 2

	Dependent variable:
	Conservative Vote Share (2017 General Election)
Outright House Ownership (Census 2011)	-1.854**
	(0.808)
Economic Activity (Census 2011)	0.418
	(0.348)
Interaction between Economic Activity and Outright House Ownership	0.045***
	(0.012)
Constant	-25.681
	(24.008)
Observations	632
R^2	0.607
Adjusted R^2	0.606
Residual Std. Error	9.666 (df = 628)
F Statistic	323.976*** (df = 3; 628)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2: Regression Model II (with addition of interaction term - House Ownership x Economic Activity)

Upon running Model II, the partial slope coefficients of the independent variables drastically changed. For outright house ownership the coefficient value switched its direction of significance from positive to negative at the 95% confidence level. For economic activity, the coefficient lost its significance altogether. On the other hand, there proved to be a significant, positive causal relationship between the interaction term of economic activity and outright house ownership at the 95% confidence level.

This is because the coefficients now only show the effects from the percentage of economically active in a constituency while excluding (i.e. controlling separately) for the effects from the percentage of economically active homeowners. **This suggests that it is not percentage of the economically active population that pushes the Conservative vote share up (indeed the effect of that independent predictor is much smaller in the second model), but how many of them also own their houses outright.** The same logic applies to the interpretation of the homeownership effect.

Furthermore, ANOVA F-Tests are used to compare between the fits of models. As shown by the table above, Model II provides a better fit in accounting for Conservative vote share. This is because the f-value is larger than 1 and the p-value is less than 0.05, we can state that there is a significant improvement from at the 95% confidence level.

	Residual Degrees of Freedom	Residual Sum of Squares	Sum of Squares	F value	P value
1	629	60043			
2	628	58672	1371.2	14.676	0.0001405

Conclusion

In relaying the results of my analysis back to the original investigation, it becomes evident that independently, the effects of wealth and employment on populations within constituencies are significant socioeconomic indicators of one's tendency to vote Conservative. However, in exploring the interaction between the two predictors, wealth subordinated the effect of employment in its significance as shown in the results of Model II.

As a recommendation for future research, I would suggest focusing on occupation instead of employment in understanding how class affects the Conservative vote share. This is because occupation is a far better measure in its capacity to capture one's income, and thus, economic status. As unemployment has been consistently under 5% in the U.K.[19], the discrepancies in income levels between groups cannot be analysed using any measure of employment.

Word count without appendix: 3240 words

---

[19] Unemployment - Office for National Statistics. (2019). Retrieved 9 December 2019, from https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment

Appendix

Descriptives

Trends in Conservative Vote Share (2017 GE)
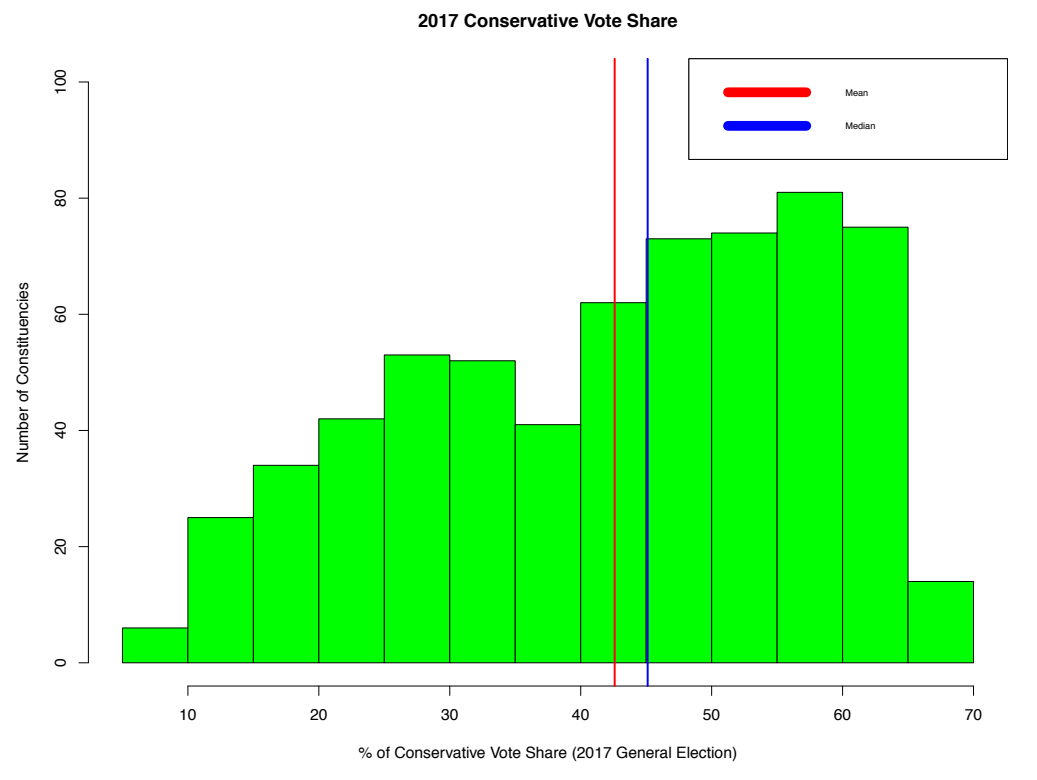
2017 Conservative Vote Share

Figure 5: Trends in Conservative Vote Share (2017 U.K. GE)

As shown in Figure 1, the 2017 GE Conservative vote share was not normally distributed. We know this because there was a significant disparity between the mean, represented as the red line and the median, represented as the blue line (42.6% and 45.1% respectively). As the mean value is lower than the median, we can deduce that the distribution of the data is skewed to the left.

Furthermore, most of the data points lie to the right of median. As the x-axis scale depicts lower proportions of Conservative vote share to the left, and higher proportions of Conservative vote share to the right, **the implication here is that a greater number of constituencies had higher percentages of Conservative vote share as opposed to lower**.

Trends in Economic Activity

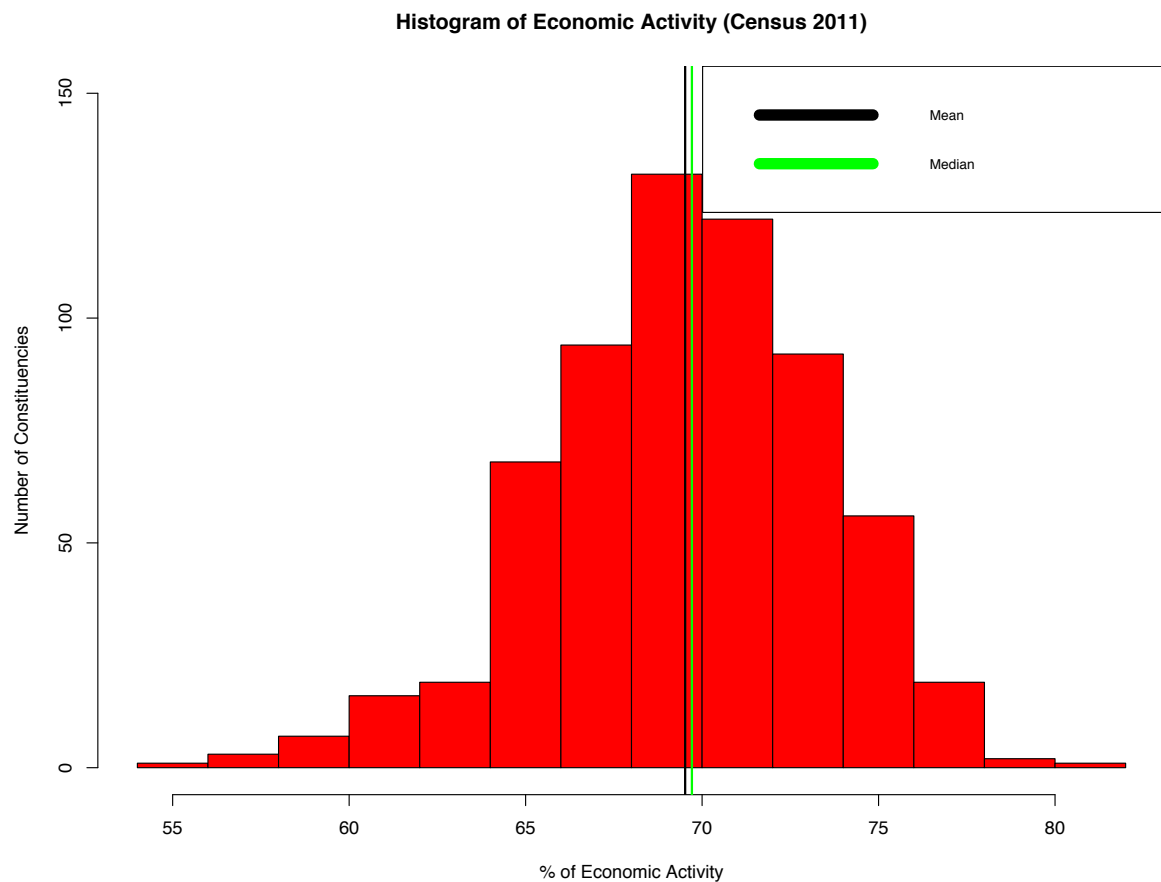**Histogram of Economic Activity (Census 2011)**

Figure 6: Histogram illustrating the trends in Economic Activity Across U.K. Constituencies

As shown in Figure 2, the proportion of economic activity was normally distributed to a great extent. We know this because the mean and the median were significantly similar in value ($\cong$ 69.52% and 69.71% respectively).

From the graph we can deduce that the average proportion of economic activity lies between the most popular rates of economic activity (67.5% - 70%) amongst constituencies in the U.K. **This suggests that to a large degree, the levels of economic activity were largely uniform and high (approx. 70%).**

Trends in House Ownership

Histogram of Outright House Ownership (Census 2011)
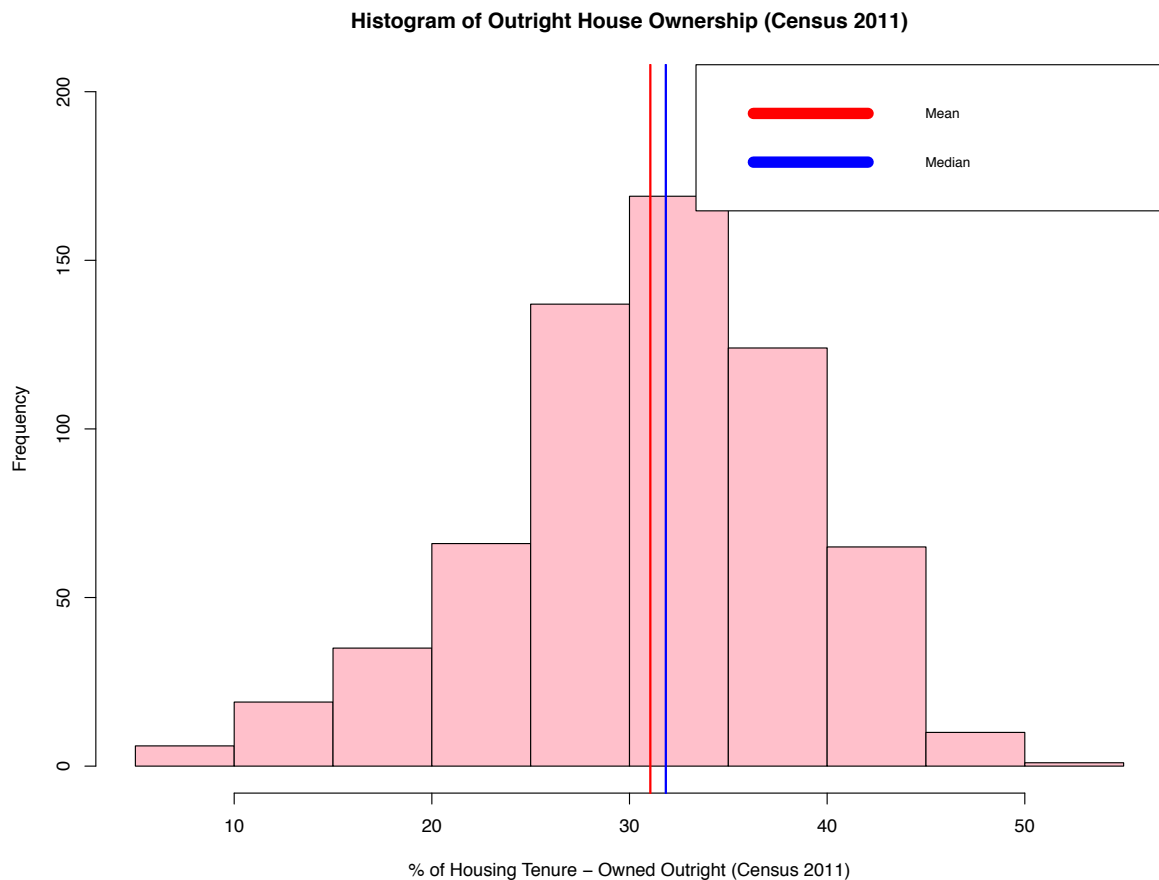

Figure 3: Histogram showing trends in House Ownership Across Constituencies in the U.K. (Census 2011)

As was the case in Figure 2, the distribution of Figure 3's data was normally distributed to a large extent. Once again, this is because the value of the mean and the median were significantly similar ($\cong$ 31.05 % and 31.84 % respectively).

From the graph we can deduce that the average proportion of house ownership coincided with the peak distribution of the data, as the mean (represented by the blue line) lies on the highest bar of the histogram. **This suggests that to a large degree, the levels of house ownership amongst constituencies were largely uniform and low (approx. 31%).**

Correlation

Correlation between Economic Activity and Conservative Vote Share

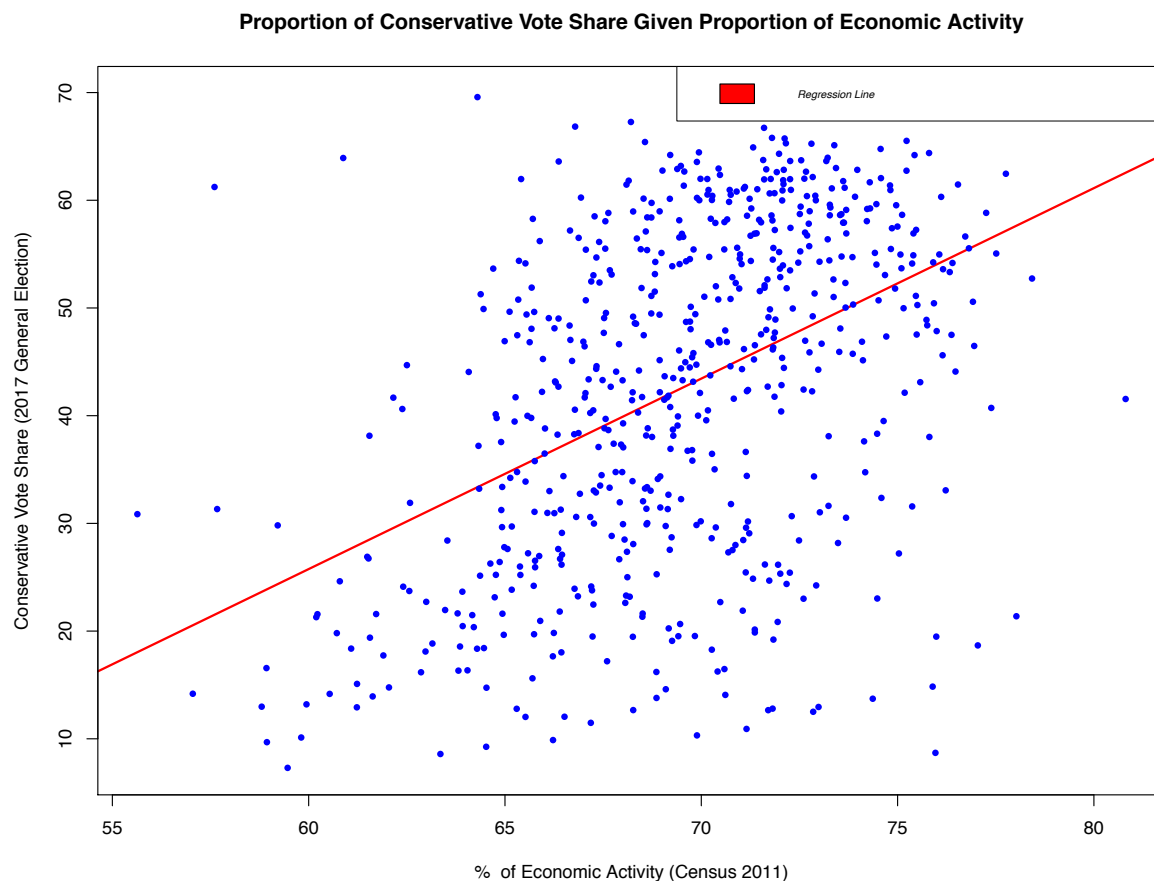**Proportion of Conservative Vote Share Given Proportion of Economic Activity**

Figure 4: Scatterplot illustrating the relationship between % of Economically Active Individuals in a Constituency and Conservative Vote Share in the 2017 GE

Figure 4 illustrates the relationship between proportions of economic activity and conservative vote share in the 2017 U.K. GE.

Upon analysing the scatterplot, one can see that the observations increase along the *x*-axis from left to right. Additionally, there are a substantial amount of outliers to the far-right bottom end of the *x*-axis.

This suggests that there is a positive correlation between economic activity and 2017 Conservative vote share. However as mentioned, this relationship might be weakened by the numerous outliers

This is reflected in the correlation table of the two variables:

	Conservative vote share vs. Economic Activity
t-statistic	12.65
p-value	2.2e-16
correlation coefficient	0.450072

As the p-value is less than 0.05 and the correlation value is 0.450072, the relationship between the two variables is positive and moderately strong.

Correlation between House Ownership and Conservative Vote Share
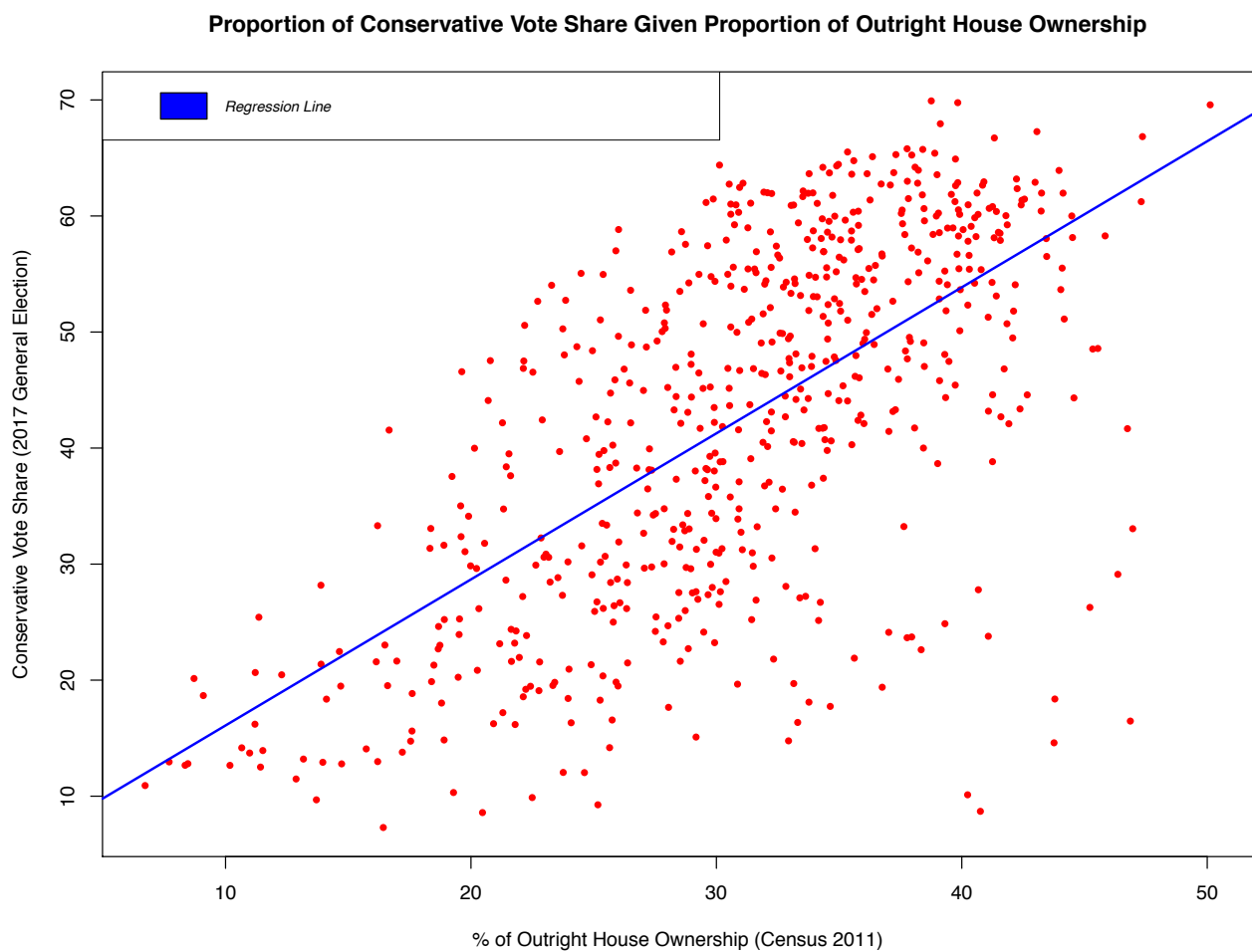
Proportion of Conservative Vote Share Given Proportion of Outright House Ownership

Figure 7: Scatterplot illustrating the relationship between % of House Ownership in a Constituency and Conservative Vote Share in the 2017 U.K. GE

Figure 5 illustrates the relationship between proportions of house ownership in constituencies as of the 2011 Census and conservative vote share in the 2017 U.K. GE.

Upon analysing the scatterplot, one can see that the observations increase along the *x*-axis from left to right. Additionally, there are a substantial amount of outliers to the far-right bottom end of the *x*-axis.

This suggests that there is a positive correlation between economic activity and 2017 Conservative vote share. However as mentioned, this relationship might be weakened by the numerous outliers.

This is reflected in the correlation table of the two variables:

	Conservative vote share vs. Economic Activity
t-statistic	21.008
p-value	2.2e-16
correlation coefficient	0.6418387

As the p-value is less than 0.05 and the correlation value is 0.6418387, the relationship between the two variables is positive and strong.

Bibliography

"2017 BES Constituency Results With Census And Candidate Data - The British Election Study". 2019. *Britishelectionstudy.Com*. https://www.britishelectionstudy.com/data-object/2017-bes-constituency-results-with-census-and-candidate-data/.

Andersen, R., & Heath, A. (2002). Class Matters: The Persisting Effects of Contextual Social Class on Individual Voting in Britain.

Doling, J., & Ronald, R. (2010). Home ownership and asset-based welfare. *Journal Of Housing And The Built Environment*, *25*(2), 165-173.

Frost, J. (2019). Regression coefficients - Statistics By Jim. Retrieved 8 December 2019, from https://statisticsbyjim.com/glossary/regression-coefficient/

Gujarati N. Damodar (2009). *Basic Econometrics*. New York. The McGraw-Hill Companies

How do you gather your data? - The British Election Study. (2019). Retrieved 8 December 2019, from https://www.britishelectionstudy.com/qa_faqs/how-do-you-gather-your-data/

Measures of Socioeconomic Status. (2019). Retrieved 8 December 2019, from https://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/367D3800605DB064CA2578B60013445C/$File/1244055001_2011.pdf

Multicollinearity - Statistics Solutions. (2019). Retrieved 8 December 2019, from https://www.statisticssolutions.com/multicollinearity/

Pulzer, Peter (1967). *Political Representation and Elections in Britain.* London: Allen and Unwin

Socioeconomic inequalities in England and Wales - Office for National Statistics. (2019). Retrieved 10 December 2019, from https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/articles/measuringsocioeconomicinequalitiesinavoidablemortalityinenglandandwales/latest

Unemployment - Office for National Statistics. (2019). Retrieved 9 December 2019, from https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment

Zeleis, A. (2019). Omitted Variable Bias | Introduction to Econometrics with R. Retrieved 8 December 2019, from https://www.econometrics-with-r.org/6-1-omitted-variable-bias.html