INDIVIDUAL DATA ANALYSIS PROJECT: BREXIT

Content

List of Tables

Introduction

Britain's Exit or Brexit from one of the most powerful trading blocs will happen to brits vote to leave the National Referendum on June 23rd June 2016. The dominant part of U.K. residents voted in favour of the U.K. to leave the European Union. British people voted to leave European Union (EU) by 52% to 48%. This sudden decision was an explanation behind stress for most U.K and E.U nationals which clearly could negatively influence the U.K economy when all is said in done and besides it left by and large E.U. subjects indeterminate about their future Citizenship status. **The cocktail of ingredients included a revolt against the consequences of globalisation by those who have not benefited from it, a revolt which found its most potent voice in the anger of working-class communities about depressed wages and pressure on local services from rising immigration (Shipman, 2016).** Some of the Key issues for EU Referendum were immigration, security and economy. Britain is set to officially leave the EU on march 29th 2019, A transition period will carry to the end of 2020.

The main objective of this project is to analyse and explore different factors (variables) such as Country of Birth, Occupation, Age, Occupation and Proficiency of English Language. Various reports have been distributed from studies experienced to discover what precisely could have caused this choice.

The report provided by Matthew Goodwin and Oliver Heath states that lack of opportunity across the country led to Brexit. The Relationship between Level of Education, Age, Country of Birth and Employment status of a citizen are some of the reasons for Brexit. As per Christopher Rocks Country of Birth is one of the important variables defined for the Brexit in his prepositions.

As most of these studies had considered different variables for Brexit, we will be considering a combine data for analysis from the Nomis and the data.gov.uk so as to check if there were some other variable affecting the referendum results. The data that is present on Nomis and Data.gov.in is not reliable since national census survey occurs every ten years and lastly the census data was collected by a survey in 2011. I have downloaded the data for 5 variables as I consider them to be the key factors leading to the Brexit votes, i.e., Age, Qualification, Proficiency in English and Country of Birth.

DS7001 Data Ecology

Data Acquisition: -

First, the Data was downloaded from the below link mentioned in the table

Data	Link
Age	https://www.nomisweb.co.uk/census/2011/ks102uk
Occupation	https://www.nomisweb.co.uk/census/2011/ks608uk
Qualification	https://www.nomisweb.co.uk/census/2011/ks501ew
Country of birth	https://www.nomisweb.co.uk/census/2011/qs203uk
Proficiency in English	https://www.nomisweb.co.uk/query/construct/submit.asp?menuopt=201&subcomp=

The Variables names were further shortened to use for R programming

Variable Names	Abbreviation (variable names changed)
Qualification	
Level 1	Level1
Level 2	Level2
Apprenticeship	Appnship
Level 3	Level3
Level 4	Level4Plus
Other Qualification	
Full time students age 18 and over	FTSA18plus
Full time students and employed age 18 to 74	FTSEA18to74
Full time students and unemployed age 18 to 74	FTSUA18to74
Full time students age 18 to 74	FTSA18to74
Proficiency in English (ProficiencyEng)	
Main Language is English	MLE
Main Language is not English can speak English very well	MLNECSEVW
Main Language is not English can speak English well	MLNECSEW
Main Language is not English cannot speak English well	MLNECNSEW
Main Language is not English cannot speak English	MLNECNSE
Occupation	
Managers, Directors and Senior Officials	MDSO
Professional	Professional
Associate Professional and Technical	APT
Administrative and Secretarial	AdminSec
Skilled Traders	Skilledtraders
Carin, Leisure and Other Services	CLOS
Sales and Customer Services	SCS
Process Plant and Machine Operatives	PPMO

AGE	
All Usual Residents	**TotPop**
Age 0 to 4	Removed from data and reduced from total population
Age 5 to 7	Removed from data and reduced from total population
Age 8 to 9	Removed from data and reduced from total population
Age 10 to 14	Removed from data and reduced from total population
Age 15	Removed from data and reduced from total population
Age 16 to 17	Removed from data and reduced from total population
Age 18 to 19	Age18to29
Age 20 to 24	
Age 25 to 29	
Age 30 to 44	Age30to44
Age 45 to 59	Age45to59
Age 60 to 74	Age60to74
Age 70 and Above	Age70andabove

The aim of the data exploration is to check whether the above-mentioned variables contribute to the Brexit vote. For that, I have downloaded the data for 5 variables as I consider them to be the key factors leading to the Brexit votes, i.e., Age, Qualification, Proficiency in English, Occupation and Country of Birth. I will be explaining the data bit by bit so that the idea to the reader will be cleared on how we analysed the data for the study.

Age: -

The data for age consists of different groups such as, 0-4, 5-7,8-9, 10-14, 15, 16-17,18-19,30-44, and so on. I have removed the age group from 0 to 17 and reduced that from the total population for the data analysis as the legal age to caste vote is 18. I have merged few data columns (age groups) such as 18-19, 20-24,25-29 as 18-29. The other groups were minimised to different category using DB Browser for SQLite with the help of SQL queries.

The final age groups in the data after merging them and reducing the population of age group 0-17 from the total population are mentioned below: -

Total age, Age 18 to 29, age 30 to 44, age 45 to 59, age 60 to 74, age 74 and above and the primary key was set to the AreaCode11 for the SQL queries.

Using the DB Browser for SQLite application six tables were created named Refdum (Referendum), Age, Qualification, Country of Birth, Proficiency in English and Occupation.

In DB Browser for SQLite application we created a database and made the tables filled with entries as shown below

```
CREATE TABLE `Refdum` (
	`regioncode`	TEXT,
	`region`	TEXT,
	`areacode11`	TEXT,
	`areacode14`	TEXT,
	`area`	TEXT,
	`centroidX`	INTEGER,
	`centroidY`	INTEGER,
	`electorate`	INTEGER,
	`Expectedballots`	INTEGER,
	`VerifiedBallotPapers`	INTEGER,
	`Pct_Turnout`	REAL,
	`Votes_Cast`	INTEGER,
	`Valid_Votes`	INTEGER,
	`Remain`	INTEGER,
	`Leave`	TEXT,
	`Rejected_Ballots`	INTEGER,
	`No_official_mark`	INTEGER,
	`Unmarked_or_void`	INTEGER,
	`Writing_or_mark`	INTEGER,
	`Pct_Remain`	REAL,
	`Pct_Leave`	REAL,
	`Pct_Rejected`	REAL,
	`Outcome`	TEXT,
	`Diff`	REAL,
	PRIMARY KEY(`areacode11`)
);
```

Once these tables were created the CSV files for every variable was imported into their respective columns as shown below

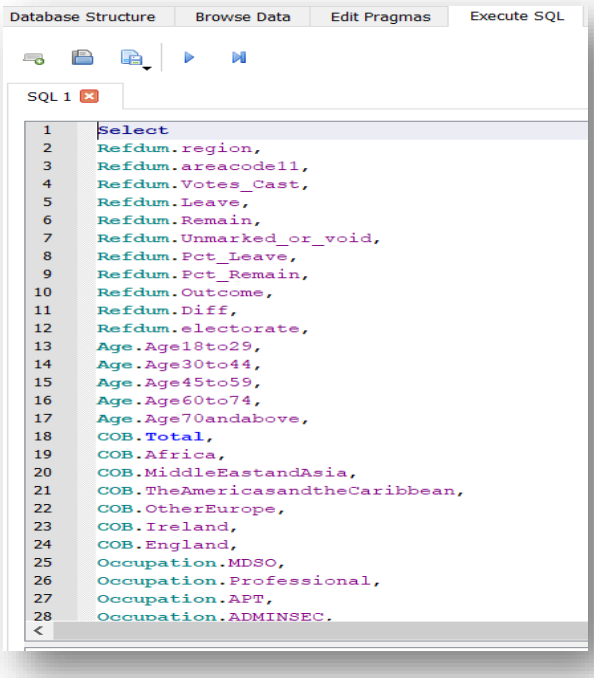
Database Structure | Browse Data | Edit Pragmas | Execute SQL

Table: Refdum — New Record | Delete Record

	regioncode	region	areacode11	areacode14	area	centroidX	centroidY	electorate	Expectedballots	rifiedBallotPape	Pct
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	E12000001	North East	E06000001	E06000001	Hartlepool	447878	530729	70341	46137	46134	65.5
2	E12000001	North East	E06000002	E06000002	Middlesbrough	450413	516581	94612	61395	61393	64.8
3	E12000001	North East	E06000003	E06000003	Redcar and Cl...	463446	517816	103529	72741	72741	70.2
4	E12000001	North East	E06000004	E06000004	Stockton-on-...	443277	518710	141486	100462	100460	71.0
5	E12000001	North East	E06000005	E06000005	Darlington	429039	517141	77662	55194	55195	71.0
6	E12000002	North West	E06000006	E06000006	Halton	352685	383373	95289	65047	65047	68.2
7	E12000002	North West	E06000007	E06000007	Warrington	362678	389287	157042	115206	115206	73.3
8	E12000002	North West	E06000008	E06000008	Blackburn wit...	369223	421898	100117	65416	65408	65.3
9	E12000002	North West	E06000009	E06000009	Blackpool	332337	436221	102354	66959	66959	65.4
10	E12000003	Yorkshire and...	E06000010	E06000010	Kingston upo...	509854	431037	180230	113439	113439	62.9
11	E12000003	Yorkshire and...	E06000011	E06000011	East Riding of...	499259	443072	266047	199056	199039	74.8
12	E12000003	Yorkshire and...	E06000012	E06000012	North East Li...	523453	406715	116302	79016	79013	67.9
13	E12000003	Yorkshire and...	E06000013	E06000013	North Lincoln...	493163	411392	123611	88912	88907	71.9
14	E12000003	Yorkshire and...	E06000014	E06000014	York	461682	451813	155157	109695	109691	70.6
15	E12000004	East Midlands	E06000015	E06000015	Derby	435850	335144	171246	120807	120798	70.5
16	E12000004	East Midlands	E06000016	E06000016	Leicester	459075	304820	213819	139319	139309	65.1
17	E12000004	East Midlands	E06000017	E06000017	Rutland	491089	307850	29390	22989	22986	78.2
18	E12000004	East Midlands	E06000018	E06000018	Nottingham	455413	340674	195394	120792	120792	61.8
19	E12000005	West Midlands	E06000019	E06000019	Herefordshire...	349072	245278	138247	108336	108336	78.3
20	E12000005	West Midlands	E06000020	E06000020	Telford and ...	367089	314838	124338	89707	89704	72.1
21	E12000005	West Midlands	E06000021	E06000021	Stoke-on-Trent	389060	346760	179010	117691	117680	65.7
22	E12000009	South West	E06000022	E06000022	Bath and Nort	367003	161986	136522	105300	105298	77.1

1 - 23 of 651 — Go to: 1

The Two Columns, Area and AreaCode was present in every table and AreaCode was selected as the primary key for the tables to join different tables and extract the data present in each variable table.



The above query will store the data in one table by joining the different tables using the primary key.

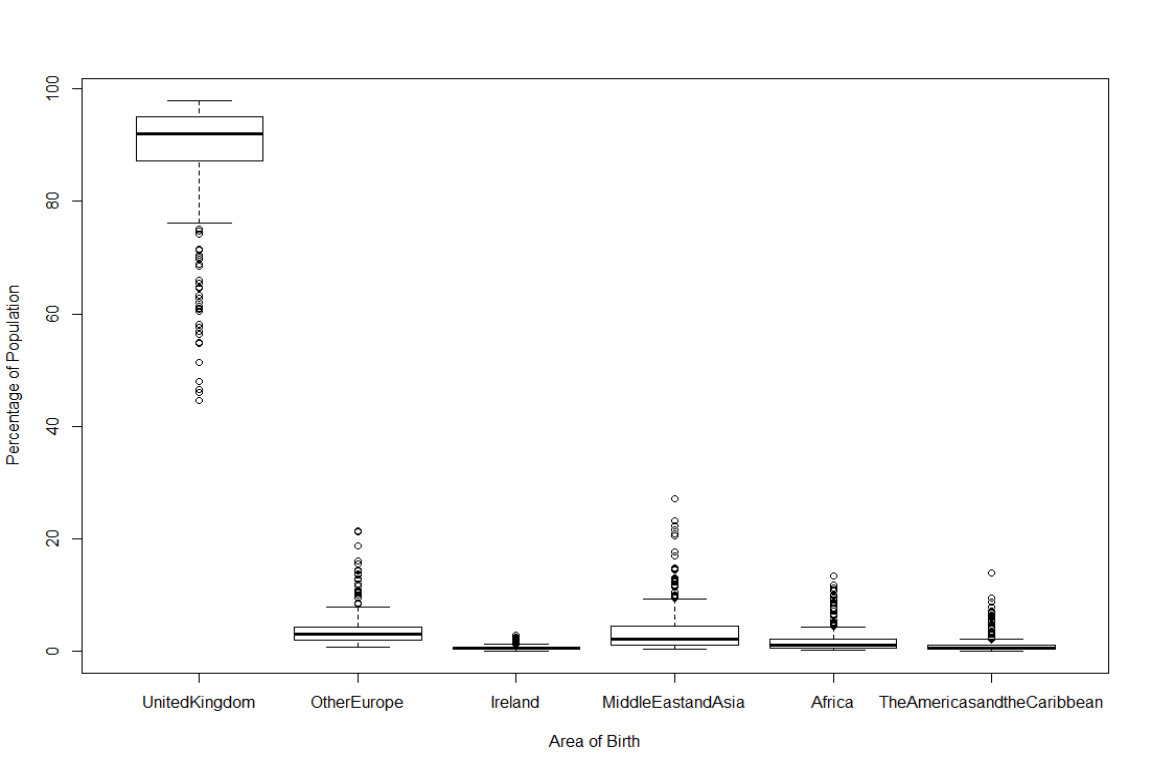
Another Query of SQL was used to convert the Variables to percentage and were rounded to two decimal points. The data was converted to a discrete variable so that we can perform the correlation test and find the relation between different variables with the Leave variable and will check whether that variable had a strong influence on the Leave variable. Later a new csv file was created using the SQL queries for data exploration. The downloaded data from Nomis and data.gov.uk was as per the requirements.

The Independent variable in the csv file are: - PLevel4plus, PUnitedKingdom and some of the variables from the ProficiencyEng Table. We will perform different test and hypothesis and examine the results for the strongest correlation between different variables and will reject the null hypothesis. We will examine the relation between the % of people born in UK that voted for the referendum and the Leave variable for the Brexit data.

Data Exploration: -

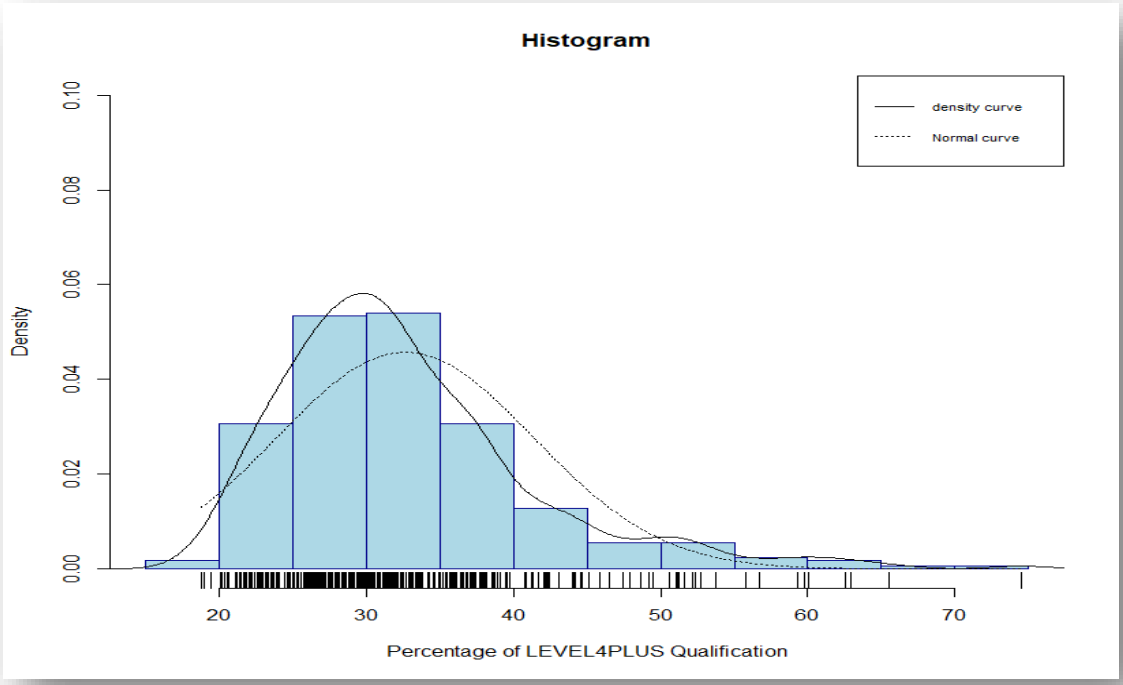
To start with data exploration, we have to find some relationship between the dependent variables and the independent variables. We have used box plot to check if the values are normally, linearly or exponentially distributed. Box plot shows us the outliers in the data so it is easily visible from the values.

The codes below show the boxplot of the percentage population and the Region of birth.

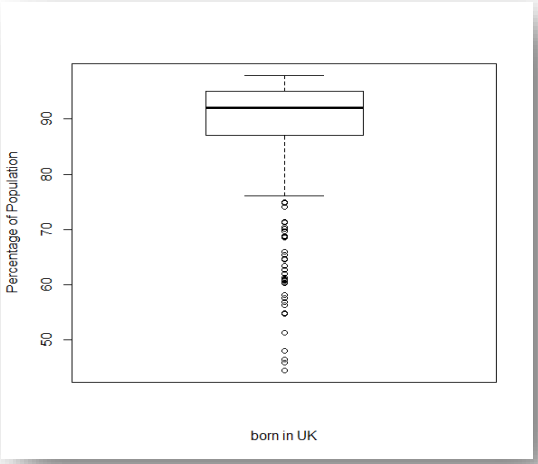


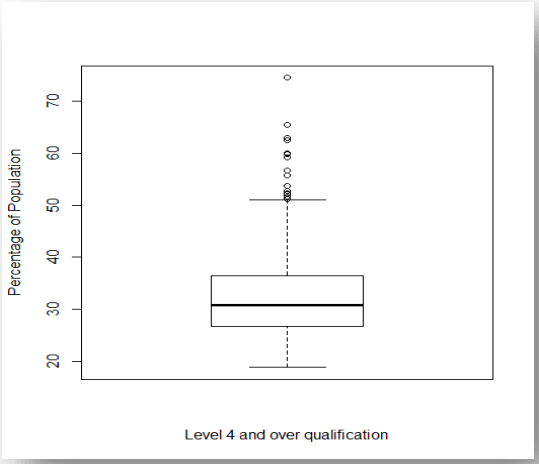
It is evident from the above box plot image that the people born in United Kingdom were the highest among all others as per the area of birth. It is crucial to figure this out as a matter of high importance else there will be no base that this was not the situation, there would be no reason for our exploration in any case.

Further, we will examine a density histogram of the Level4PLUS qualification and other education if the numbers are normally distributed and we will check the mean value for the level4PLUS and other qualified population.

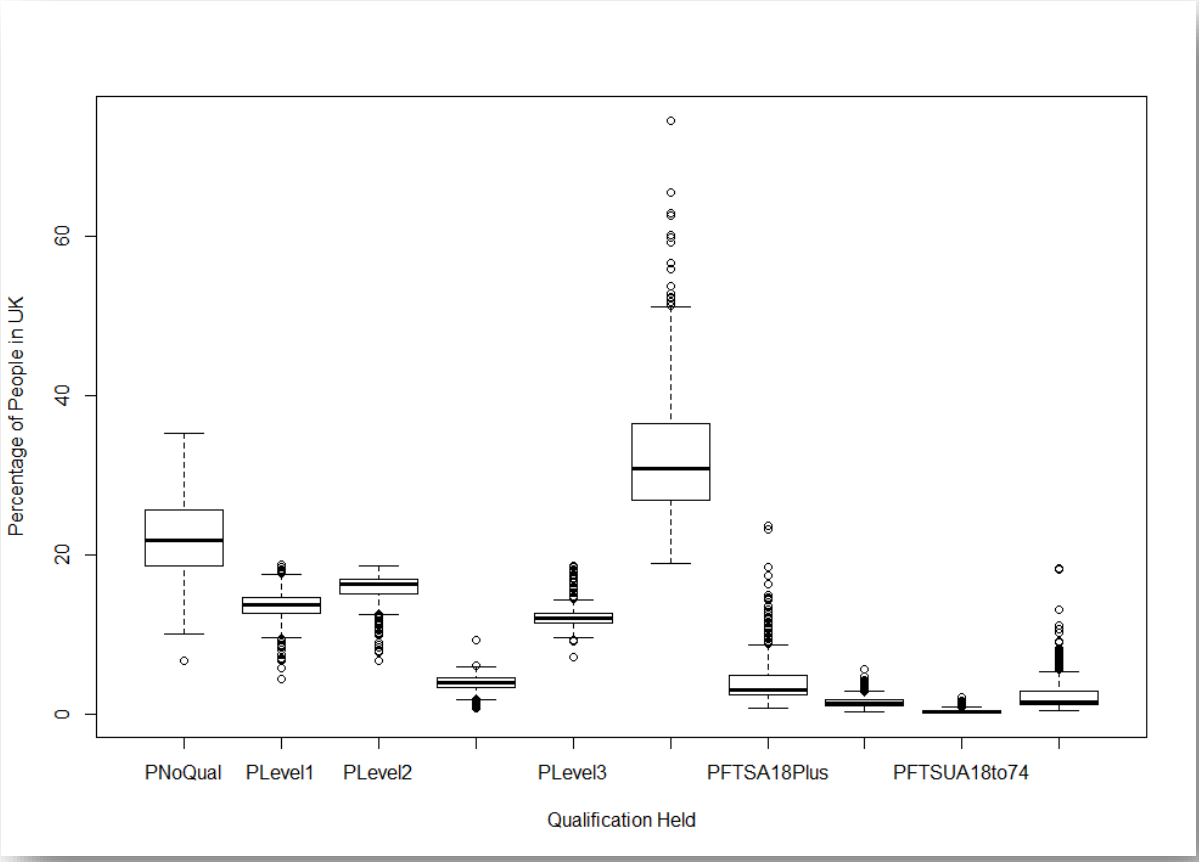


33 percent of total population is having a level 4 qualification and is evident from the mean value. The values are not normally distributed, resulting to a conclusion that there is a large number of educated population resident in different. From the normal curve and the density curve, the values for the mean are close to the values for median.As per the data, we will now examine the voters for Level4Plus Qualification and the voters that are born in UK. A large number of values are outside the distribution and hence considered as outliers. It is possible that the British citizen prefer to live in community in the same area. **Labour insiders say that as they talk to people in their homes and when out canvassing, Europe is not a priority issue. Jobs, cost of living, the NHS, affordable housing, cuts to public services are more important (MACSHANE, 2015).**

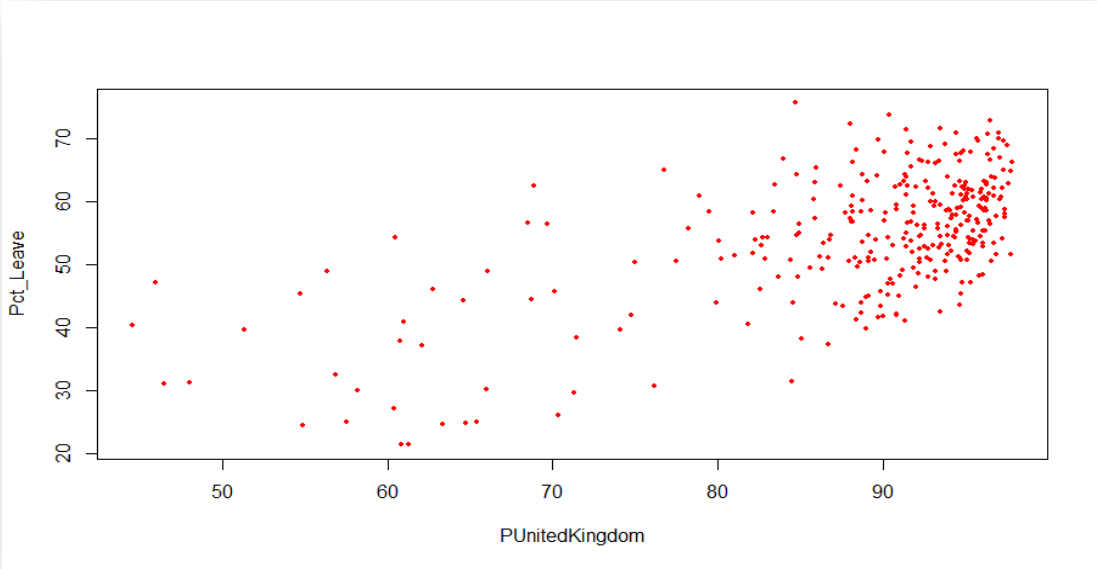


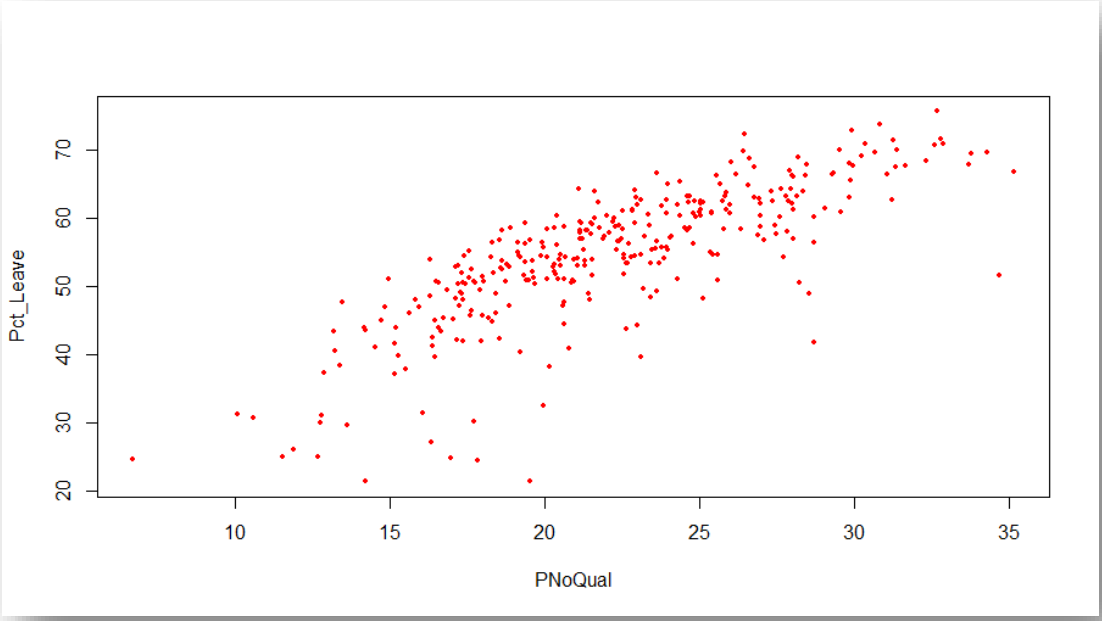


Before moving forward with the exploration, we will find the distribution of the level of education for citizens born in United Kingdom. Below is the box plot of the Qualification table with all the variables. 20 percent of electors born in United Kingdom have no Qualification and it is clearly evident from the box plot that the population for Level 4 plus education is about 30 percent.



Now, I will be using the scattered plots as using the scattered plots we will be able to check the correlation between different variables. We will find the dependability of the variables. The 2 scatterplots below show a distribution of Voters born in UK and Percentage of Population with No Qualification with the Pct_Leave variable. **Steven Hill has also made the point that the difficulty of finding a job is encouraging young people to stay in education, and that discouragement from entering the workforce can severely distort the true picture (Merritt, 2016).**





It is evident that the UK born no qualified people have a very strong correlation. In contrary, people born in United Kingdom alone has a moderate positive correlation with the Leave variable.

As the variables are discrete continuous, we are going to perform Spearman Rank's correlation test. From the scattered plot it is clearly visible that the distribution is monotonic and we can run Spearman Rank's correlation test so as to measure the strength and direction of the monotonic relationship. If from the scatter plots the variations were linear then we would have run the Pearson's correlation.

DS7001 Data Ecology

We will run the Spearman Rank's correlation test with the leave decision set to 95% confidence level to check the correlation between the two independent variables.

```
        Spearman's rank correlation rho

data:  PUnitedKingdom and Pct_Leave
S = 2996300, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.4810866

Warning message:
In cor.test.default(PUnitedKingdom, Pct_Leave, method = "spearman",  :
  Cannot compute exact p-value with ties
> cor.test(PNoQual, Pct_Leave, method = "spearman", conf.level = 0.95)

        Spearman's rank correlation rho

data:  PNoQual and Pct_Leave
S = 1097700, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.8098967

Warning message:
In cor.test.default(PNoQual, Pct_Leave, method = "spearman", conf.level = 0.95) :
  Cannot compute exact p-value with ties
> cor.test(PUnitedKingdom, Pct_Leave, method = "pearson", conf.level = 0.95)

        Pearson's product-moment correlation

data:  PUnitedKingdom and Pct_Leave
t = 14.156, df = 324, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5462434 0.6810780
sample estimates:
      cor
0.6181878

> cor.test(PNoQual, Pct_Leave, method = "pearson", conf.level = 0.95)

        Pearson's product-moment correlation

data:  PNoQual and Pct_Leave
t = 22.991, df = 324, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7422408 0.8254105
sample estimates:
      cor
0.7873824
```

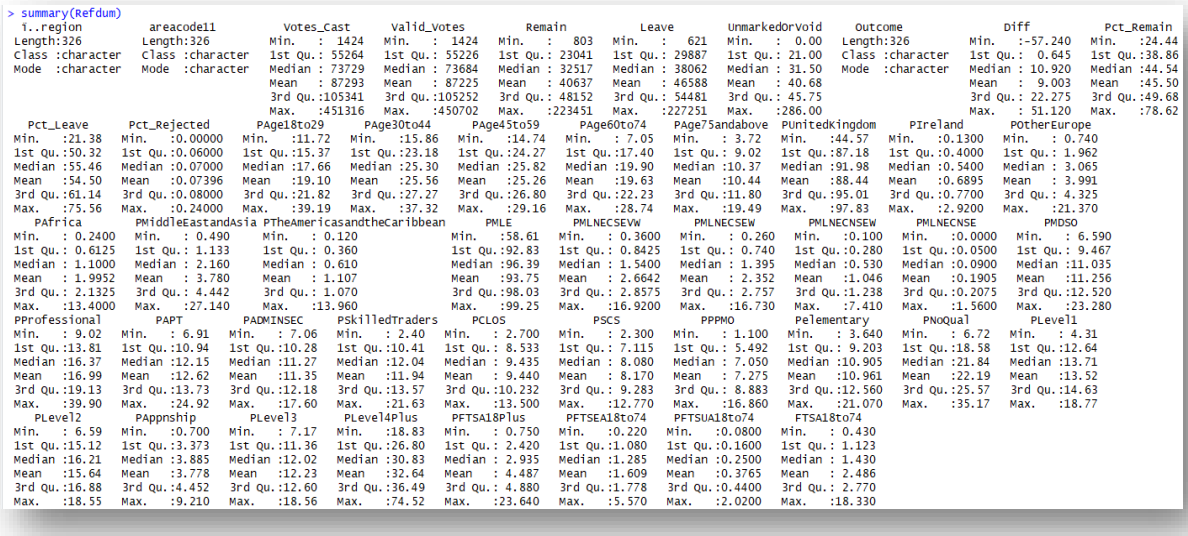
As per the above test for spearman and Pearson, it can be concluded that the correlation between these two variables is very strong and positive. This can be verified by the values of rho.

DS7001 Data Ecology

Rho value of spearman test for PNoQual with leave is 0.81 and the p value is less than 2.2e-16 and value for Pearson test for PNoQual with leave is 0.79 and the p value is less than 2.2e-16 respectively. We can safely reject the null hypothesis as the value of p is less than 0.01 and conclude that there is no relation between the population of people born in UK with the decision of leave. There is a moderate relationship between the variables and with this reference we cannot prove that with 99 percent confidence level, the Brexit decision depend on population of born in United Kingdom.

As the data is 5years ago the Brexit referendum we cannot make an exact conclusion.

Now we will summarize all the values for the variables in the dataset and will explore them for distribution.



```
> summary(Refdum)
  i..region          areacode11         Votes_Cast      Valid_Votes         Remain           Leave        UnmarkedorVoid     Outcome               Diff            Pct_Remain
 Length:326         Length:326         Min.   :  1424   Min.   :  1424   Min.   :   803   Min.   :   621   Min.   :  0.00   Length:326         Min.   :-57.240   Min.   :24.44
 Class :character   Class :character   1st Qu.: 55264   1st Qu.: 55226   1st Qu.: 23041   1st Qu.: 29887   1st Qu.: 21.00   Class :character   1st Qu.:  0.645   1st Qu.:38.86
 Mode  :character   Mode  :character   Median : 73729   Median : 73684   Median : 32517   Median : 38062   Median : 31.50   Mode  :character   Median : 10.920   Median :44.54
                                       Mean   : 87293   Mean   : 87225   Mean   : 40637   Mean   : 46588   Mean   : 40.68                      Mean   :  9.003   Mean   :45.50
                                       3rd Qu.:105341   3rd Qu.:105252   3rd Qu.: 48152   3rd Qu.: 54481   3rd Qu.: 45.75                      3rd Qu.: 22.275   3rd Qu.:49.68
                                       Max.   :451316   Max.   :450702   Max.   :223451   Max.   :227251   Max.   :286.00                      Max.   : 51.120   Max.   :78.62
   Pct_Leave      Pct_Rejected       PAge18to29      PAge30to44      PAge45to59      PAge60to74     PAge75andabove  PUnitedKingdom     PIreland       POtherEurope
 Min.   :21.38   Min.   :0.00000   Min.   :11.72   Min.   :15.86   Min.   :14.74   Min.   : 7.05   Min.   : 3.72   Min.   :44.57   Min.   :0.1300   Min.   : 0.740
 1st Qu.:50.32   1st Qu.:0.06000   1st Qu.:15.37   1st Qu.:23.18   1st Qu.:24.27   1st Qu.:17.40   1st Qu.: 9.02   1st Qu.:87.18   1st Qu.:0.4000   1st Qu.: 1.962
 Median :55.46   Median :0.07000   Median :17.66   Median :25.30   Median :25.82   Median :19.90   Median :10.37   Median :91.98   Median :0.5400   Median : 3.065
 Mean   :54.50   Mean   :0.07396   Mean   :19.10   Mean   :25.56   Mean   :25.26   Mean   :19.63   Mean   :10.44   Mean   :88.44   Mean   :0.6895   Mean   : 3.991
 3rd Qu.:61.14   3rd Qu.:0.08000   3rd Qu.:21.82   3rd Qu.:27.27   3rd Qu.:26.80   3rd Qu.:22.23   3rd Qu.:11.80   3rd Qu.:95.01   3rd Qu.:0.7700   3rd Qu.: 4.325
 Max.   :75.56   Max.   :0.24000   Max.   :39.19   Max.   :37.32   Max.   :29.16   Max.   :28.74   Max.   :19.49   Max.   :97.83   Max.   :2.9200   Max.   :21.370
    PAfrica         PMiddleEastandAsia PTheAmericasandtheCaribbean      PMLE          PMLNECSEVW       PMLNECSEW       PMLNECNSEW      PMLNECNSE          PMDSO
 Min.   : 0.2400   Min.   : 0.490     Min.   : 0.120              Min.   :58.61   Min.   : 0.3600   Min.   : 0.260   Min.   :0.100   Min.   :0.0000   Min.   : 6.590
 1st Qu.: 0.6125   1st Qu.: 1.133     1st Qu.: 0.360              1st Qu.:92.83   1st Qu.: 0.8425   1st Qu.: 0.740   1st Qu.:0.280   1st Qu.:0.0500   1st Qu.: 9.467
 Median : 1.1000   Median : 2.160     Median : 0.610              Median :96.39   Median : 1.5400   Median : 1.395   Median :0.530   Median :0.0900   Median :11.035
 Mean   : 1.9952   Mean   : 3.780     Mean   : 1.107              Mean   :93.75   Mean   : 2.6642   Mean   : 2.352   Mean   :1.046   Mean   :0.1905   Mean   :11.256
 3rd Qu.: 2.1325   3rd Qu.: 4.442     3rd Qu.: 1.070              3rd Qu.:98.03   3rd Qu.: 2.8575   3rd Qu.: 2.757   3rd Qu.:1.238   3rd Qu.:0.2075   3rd Qu.:12.520
 Max.   :13.4000   Max.   :27.140     Max.   :13.960              Max.   :99.25   Max.   :16.9200   Max.   :16.730   Max.   :7.410   Max.   :1.5600   Max.   :23.280
 PProfessional        PAPT          PADMINSEC      PSkilledTraders      PCLOS            PSCS            PPPMO          Pelementary       PNoQual          PLevel1
 Min.   : 9.02   Min.   : 6.91   Min.   : 7.06   Min.   : 2.40   Min.   : 2.700   Min.   : 2.300   Min.   : 1.100   Min.   : 3.640   Min.   : 6.72   Min.   : 4.31
 1st Qu.:13.81   1st Qu.:10.94   1st Qu.:10.28   1st Qu.:10.41   1st Qu.: 8.533   1st Qu.: 7.115   1st Qu.: 5.492   1st Qu.: 9.203   1st Qu.:18.58   1st Qu.:12.64
 Median :16.37   Median :12.15   Median :11.27   Median :12.04   Median : 9.435   Median : 8.080   Median : 7.050   Median :10.905   Median :21.84   Median :13.71
 Mean   :16.99   Mean   :12.62   Mean   :11.35   Mean   :11.94   Mean   : 9.440   Mean   : 8.170   Mean   : 7.275   Mean   :10.961   Mean   :22.19   Mean   :13.52
 3rd Qu.:19.13   3rd Qu.:13.73   3rd Qu.:12.18   3rd Qu.:13.57   3rd Qu.:10.232   3rd Qu.: 9.283   3rd Qu.: 8.883   3rd Qu.:12.560   3rd Qu.:25.57   3rd Qu.:14.63
 Max.   :39.90   Max.   :24.92   Max.   :17.60   Max.   :21.63   Max.   :13.500   Max.   :12.770   Max.   :16.860   Max.   :21.070   Max.   :35.17   Max.   :18.77
    PLevel2         PAppnship        PLevel3        PLevel4Plus      PFTSA18Plus      PFTSEA18to74     PFTSUA18to74      FTSA18to74
 Min.   : 6.59   Min.   :0.700   Min.   : 7.17   Min.   :18.83   Min.   : 0.750   Min.   :0.220   Min.   :0.0800   Min.   : 0.430
 1st Qu.:15.12   1st Qu.:3.373   1st Qu.:11.36   1st Qu.:26.80   1st Qu.: 2.420   1st Qu.:1.080   1st Qu.:0.1600   1st Qu.: 1.123
 Median :16.21   Median :3.885   Median :12.02   Median :30.83   Median : 2.935   Median :1.285   Median :0.2500   Median : 1.430
 Mean   :15.64   Mean   :3.778   Mean   :12.23   Mean   :32.64   Mean   : 4.487   Mean   :1.609   Mean   :0.3765   Mean   : 2.486
 3rd Qu.:16.88   3rd Qu.:4.452   3rd Qu.:12.60   3rd Qu.:36.49   3rd Qu.: 4.880   3rd Qu.:1.778   3rd Qu.:0.4400   3rd Qu.: 2.770
 Max.   :18.55   Max.   :9.210   Max.   :18.56   Max.   :74.52   Max.   :23.640   Max.   :5.570   Max.   :2.0200   Max.   :18.330
```

Dimensionality Reduction: -

The main Objective of this project is to find the variable which is having the “very strong” relationship with the Pct_Leave variable. We have to reject null hypothesis by concluding that citizens that are born in UK are in strong opposition of the immigrants and so that was the main reason of Brexit.

We will use the below criteria to find very strong relationship between variables.

Range (Rank r)	Relation type
0-0.19	very weak
0.20– 0.39	Weak
0.40 – 0.59	Moderate
0.60 – 0.79	Strong
0.80 – 1	very strong

We will now use the scattered plot analysis between dependent and the independent variable and will remove the variables which are not beneficial to us. In the appendix I have attached all the codes for R programming which will perform the scattered plot for visualization. The r vales will rank the variables and we will be able to decide the which variables will be affecting the Brexit voting pattern.

Now we will perform the spearman’s rank test with a confidence level of 95 percent on the below mentioned variable

PFTSEA18to74 with Pct_Leave

PAge95andabove with Pct_Leave

Page18to29 with Pct_Leave

PMLNECSEW with Pct_Leave

PNoQual with PUnitedKingdom

PMLNECSEW with PUnitedKingdom

```
        Spearman's rank correlation rho

data:  PFTSEA18to74 and Pct_Leave
S = 8006400, p-value = 4.619e-13
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.3865663

Warning message:
In cor.test.default(PFTSEA18to74, Pct_Leave, method = "spearman",  :
  Cannot compute exact p-value with ties
> cor.test(PAge75andabove, Pct_Leave, method = "spearman", conf.level = 0.95)

        Spearman's rank correlation rho

data:  PAge75andabove and Pct_Leave
S = 4517200, p-value = 7.395e-05
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.2176988

Warning message:
In cor.test.default(PAge75andabove, Pct_Leave, method = "spearman",  :
  Cannot compute exact p-value with ties
> cor.test(PAge18to29, Pct_Leave, method = "spearman", conf.level = 0.95)

        Spearman's rank correlation rho

data:  PAge18to29 and Pct_Leave
S = 6572400, p-value = 0.01248
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.1382268

Warning message:
In cor.test.default(PAge18to29, Pct_Leave, method = "spearman",  :
  Cannot compute exact p-value with ties
```

```
        Spearman's rank correlation rho

data:  PNoQual and PUnitedKingdom
S = 3211100, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.4439033

Warning message:
In cor.test.default(PNoQual, PUnitedKingdom, method = "spearman",  :
  Cannot compute exact p-value with ties
> cor.test(PMLNECSEW, PUnitedKingdom, method = "spearman", conf.level = 0.95)

        Spearman's rank correlation rho

data:  PMLNECSEW and PUnitedKingdom
S = 11197000, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.9391386

Warning message:
In cor.test.default(PMLNECSEW, PUnitedKingdom, method = "spearman",  :
  Cannot compute exact p-value with ties
> cor.test(PMLNECSEW, Pct_Leave, method = "spearman", conf.level = 0.95)

        Spearman's rank correlation rho

data:  PMLNECSEW and Pct_Leave
S = 7661200, p-value = 1.5e-09
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.3267879

Warning message:
In cor.test.default(PMLNECSEW, Pct_Leave, method = "spearman", conf.level = 0.95) :
  Cannot compute exact p-value with ties
```

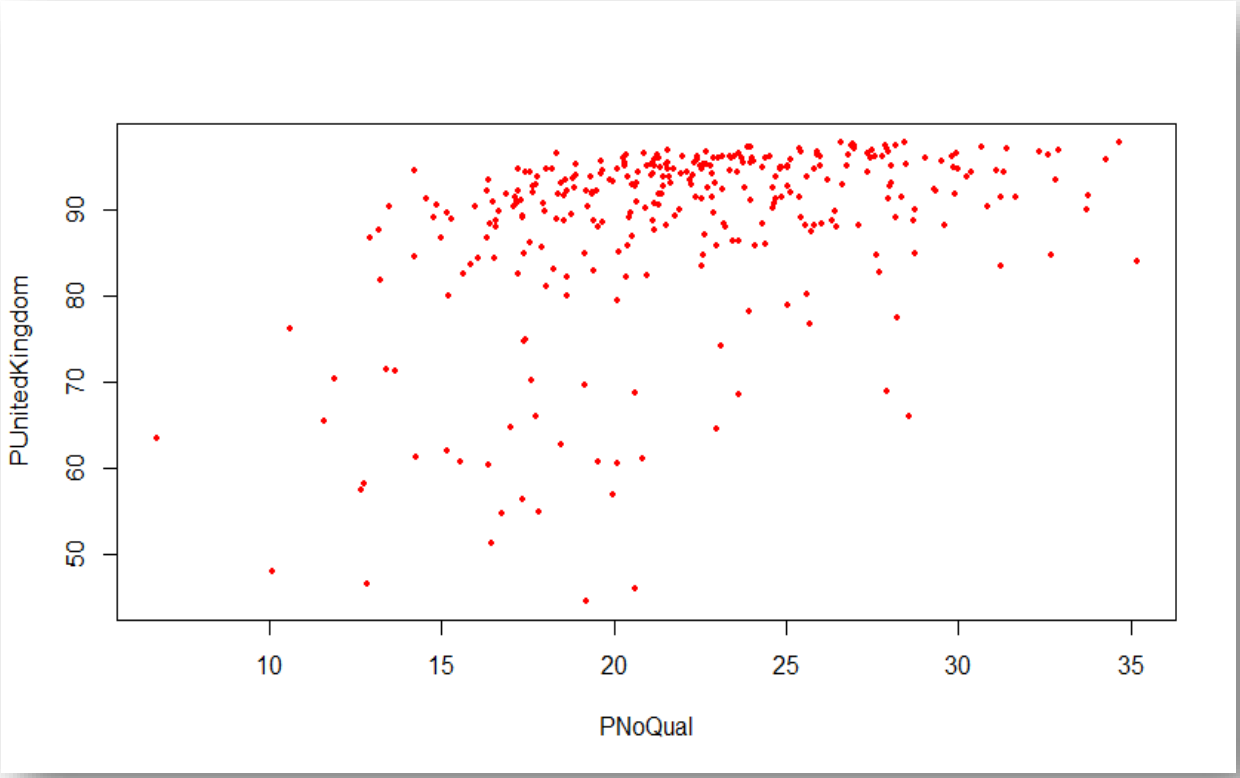
From the above 6 Spearman's rank correlation tests, 4 will conclude the independent variables and 2 internal tests for independent variable to remove the variables.

We performed scattered plot test for the variable PFTSE18to74 with Pct_Leave and concluded a weak negative correlation. We can drop the variable as the r value is -0.386. These people most likely voted to remain but was neither very influential on that decision either. As the r value is negative so we can say that the full-time students and employed under the age of 18 to 74 were not very much influential on the leave or remain decision

From the scatter plot of the age group of 75 plus it is evident that the senior citizens were not so determine about their decision on leaving the UK. The r value is 0.21 which shows a weak positive correlation. As it is a weak positive correlation, we will keep PAge74andabove to see how it can affect the other variables to Referendum.

The variable PAge18to29 has a negative weak correlation with Pct_leave and r value of -0.12 so we can neglect this variable. The reason may be the increase in number of young people due to birth from immigrant parents which might have resulted the young people to vote to remain in UK. **The current focus in most of the EU is on refugees: the tide of desperate people seeking a new life in Europe as an escape from the horrors of war and dire poverty in Africa and the Middle East (Liddle, 2016).**

Now we will perform scatter plot analysis on the independent variables so that we can check how these variables affect the Brexit voting patter as they had a moderate positive correlation. Figure below shows a scatter plot analysis of People born in Uk with the uneducated people. Further we performed spearman's rank correlation test and the r value resulted in 0.49 which concludes that there was some influence of educated people on the Leave decision.



The variable MLNECSPEW (Main Language Not English Can Speak English well) also had a strong negative correlation with the PUnitedKingdom (Percentage of People born in United Kingdom) and the r value for the spearman's rank test came out to be -0.93. So, we can drop the MLNECSPEVW variable.

At last we will perform the spearman's rank test and scatter plot analysis for the variable MLNECSEVW (Main Language Not English Can Speak English Very Well) with Pct_leave variable and it concluded to be a moderate negative correlation. This result can be the immigrants born in UK but with another ethnicity voting for remain. This category may be the people who have migrated long back from Europe in UK and are now citizens of UK.

Bi-variate correlations

As the data Present in the Referendum CSV file is discretely ordinal, we can use spearman rank's test. We will be able to perform the test for strong correlation among different variables. As most of the variables in the dataset are not showing a strong correlation with the leave decision, we will further proceed with partial correlation and check how the variables are internally correlated. The partial correlation method measures the degree between two variables, with the impact of a lot of controlling irregular variable exploration.

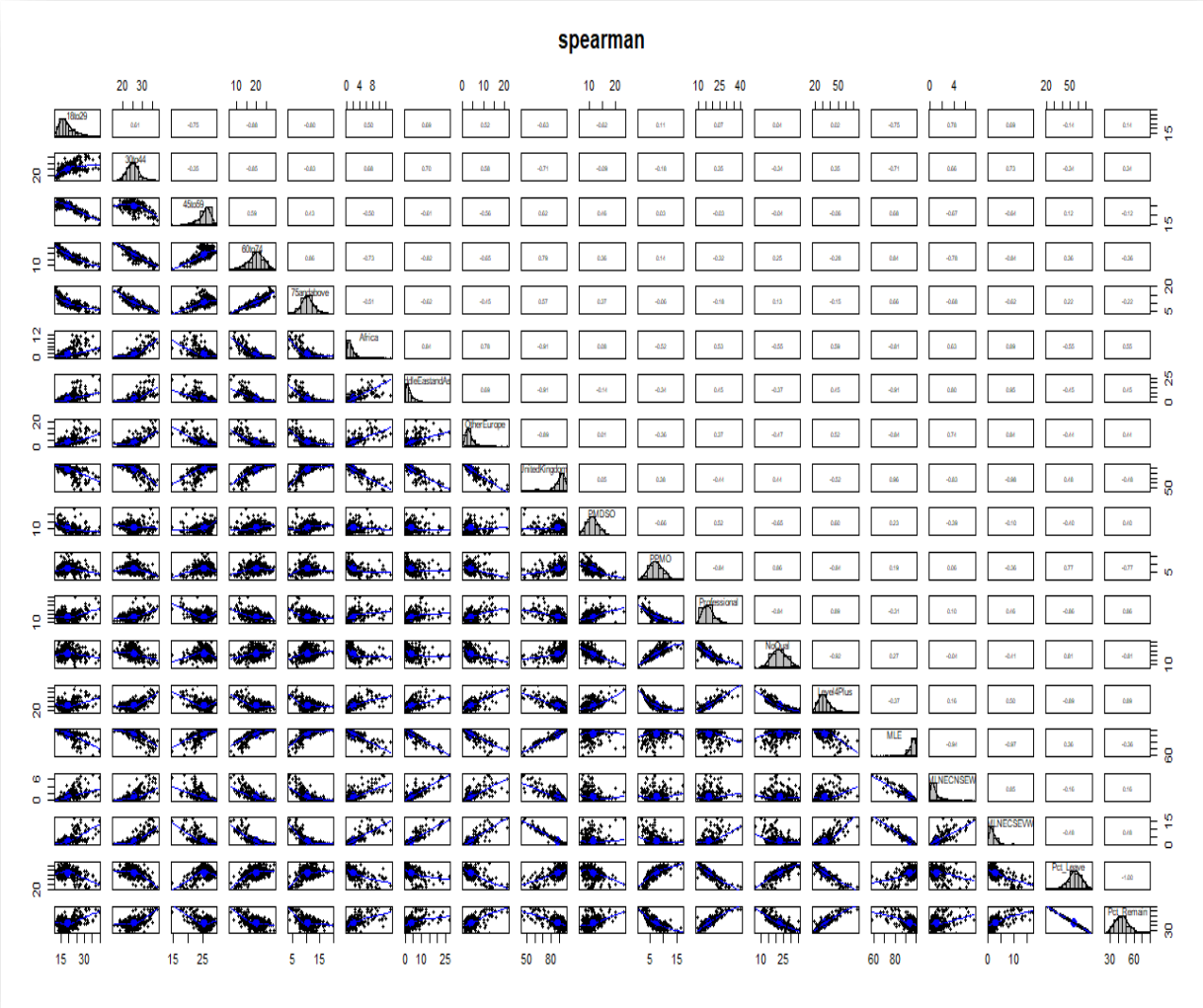
We have taken into consideration those major population variables which seems to be feasible from the dataset and removed those which are lesser in numbers or whose percentage is very less. For example, we have only considered the 4 Ethnic groups such as United Kingdom, Africans, Middle east and Asians and Other Europeans as they concord a large number of population group in UK. **If we look at nationality provisions in the Independence acts of countries such as the commonwealth country Nigeria, Ghana, the Malayan federation, Zambia or Malawi we recognise the same pattern: The question of who kept British nationality largely depended on whether one would acquire nationality of the newly independent state (Mindus, 2017).**

We have selected occupation as a variable because it is a reflection of education class in different industrial sector. The outcome of Referendum may also be affected due to the occupation of people resident in UK. So, we have considered the variables as MDSO (Managing directors and senior officials), Professionals and PPMO (Power Plant and Machine Operatives).

From the table, we will select the variables for Proficient in English as MLE (Main Language English), MLNECNSEW (Main Language is Not English and Cannot Speak English Well) and MLNECSEVW (Main Language Not English And Can Speak English Very Well). We selected these 3 variables as we consider them to be having influence on the Leave decision.

Below is the spearman's correlation for all the variables present in the dataset for Brexit. It is difficult to read the correlation as there is large number of variables used to prove our hypothesis.

DS7001 Data Ecology



Below is the spearman's rank correlation test and the results for different variables.

```
> cor.test(Refdum$PAge18to29, Refdum$Pct_Leave, method = "spearman")

        Spearman's rank correlation rho

data:  Refdum$PAge18to29 and Refdum$Pct_Leave
S = 6572400, p-value = 0.01248
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.1382268

Warning message:
In cor.test.default(Refdum$PAge18to29, Refdum$Pct_Leave, method = "spearman") :
  Cannot compute exact p-value with ties
> cor.test(Refdum$PAge30to44, Refdum$Pct_Leave, method = "spearman")

        Spearman's rank correlation rho

data:  Refdum$PAge30to44 and Refdum$Pct_Leave
S = 7734500, p-value = 3.1e-10
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.3394727

Warning message:
In cor.test.default(Refdum$PAge30to44, Refdum$Pct_Leave, method = "spearman") :
  Cannot compute exact p-value with ties
> cor.test(Refdum$PAge45to59, Refdum$Pct_Leave, method = "spearman")

        Spearman's rank correlation rho

data:  Refdum$PAge45to59 and Refdum$Pct_Leave
S = 5066500, p-value = 0.02689
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.1225796

Warning message:
In cor.test.default(Refdum$PAge45to59, Refdum$Pct_Leave, method = "spearman") :
  Cannot compute exact p-value with ties
```

By looking at results for correlation of dependent variable with all independent variables with results set at a 95% confidence level;

Variables	Correlation with Pt_Leave
PAge18to29	negative correlation
PAge30to44	negative correlation
PAge45to59	**weak positive correlation**
PAfrica	negative correlation
PMiddleEast	weak negative correlation
POtherEurope	negative correlation
PUnitedKingdom	**moderate positive correlation**
PMDSO	moderate negative correlation
PPMO	**strong positive correlation**
Professionals	negative correlation
PNoQual	**strong positive correlation**
PLevel4Plus	very strong negative correlation
PMLE	**weak positive correlation**
PMLNECNSE	weak negative correlation
PMLNECSEVW	moderate negative correlation

We have performed correlation test and found the below variables to have a positive correlation with the Leave Variable

- PAge45to59
- PUnitedKingdom
- PPMO
- PNoQual
- PMLE

We can draw different verbal inferences depending on the variable that have a positive correlation on the leave variable. So, the below questions come into picture depending on the outcome of the spearman's rank correlation test

Was the population of people who have no qualification and between the age group of 45 to 59 influenced for the Leave decision? Or the people with high level of education and are at good at their occupation level voted against the Leave decision? Or the people who are proficient in English but their main language is not English and their ethnicity is different might have voted for the Remain decision of Brexit.

These are some interesting questions that we will be looking forward to answer from the above correlations test.

Partial correlations

From the above table of correlation, we will perform partial correlation test to test the influence on the Leave variable.

```
> library(ppcor)
> pcor.test(Refdum$PUnitedKingdom, Refdum$PMLNECSEVW, Refdum$Pct_Remain)
    estimate       p.value statistic   n gp  Method
1 -0.9726308 8.446692e-207 -75.23075 326  1 pearson
> pcor.test(Refdum$PUnitedKingdom, Refdum$PPPMO, Refdum$Pct_Leave)
    estimate      p.value statistic   n gp  Method
1 -0.2399696 1.222618e-05 -4.442592 326  1 pearson
> pcor.test(Refdum$PNoQual, Refdum$PAge45to59, Refdum$Pct_Leave)
    estimate      p.value statistic   n gp  Method
1 -0.4175432 3.831694e-15 -8.258528 326  1 pearson
> pcor.test(Refdum$PUnitedKingdom, Refdum$PAge45to59, Refdum$Pct_Leave)
   estimate      p.value statistic   n gp  Method
1 0.6218819 3.598655e-36  14.27203 326  1 pearson
> pcor.test(Refdum$PUnitedKingdom, Refdum$PNoQual, Refdum$Pct_Leave)
    estimate     p.value statistic   n gp  Method
1 -0.1492046 0.007048162 -2.711892 326  1 pearson
```

The above outcomes states that

The value of R is very close to -1 for the first test. The correlation between PMLNECSEVW and PUnitedKingdom have a very strong negative correlation, i.e., people born in United Kingdom and who have ethnicity other than British voted against the Leave decision of Brexit.

The variables PunitedKingdom and PPPMO have a negative correlation which concludes that the machine operatives that are born in United Kingdom voted for Remain in the Referendum.The population of uneducated age group 45 to 59 with partial correlation test on the leave variable also shows a negative correlation. This explains that the illiterate population between the age group 45 to 59 voted for EU to remain in the UK.

The population born in UK and is uneducated also shows a weak negative correlation so it can be inferred that they voted for the remain decision. **Exit polling reported that 49 per cent of those who voted Leave were like her, and said they were for Brexit so that decisions about the UK should be taken in the UK. A further 33 per cent said they voted so that the country would have control over its borders when it came to immigration (Barnett, 2017).**

Exceptionally is the result of PAge45to59 and PUnitedKingdom with Leave variable is a strong correlation. It is clearly seen that the percentage of population born in United Kingdom and in the age group of 45 to 59 voted for the Leave decision of Brexit Referendum.

A regression analysis would eventually be able to help us draw a conclusion to the hypothesis testing.

Hypothesis testing & Regression Analysis

The data we have analysed till now is a nominal data and some of variables are not distributed normally so we will be able to apply non-parametric test to the Referendum Data and analyse the results to its best. First, we will apply multiple regression to the five variables that might be affecting the Brexit data. The lm() function will generate a simple regression model of the Leave variable and the summary is shown below

```
Call:
lm(formula = Refdum$Pct_Leave ~ Refdum$PAge45to59 + Refdum$PUnitedKingdom +
    Refdum$PNoQual + Refdum$PPPMO + Refdum$PMLNECSEVW)

Residuals:
     Min       1Q   Median       3Q      Max
-21.4116  -2.6213   0.1273   2.6368  10.9515

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           8.61381   14.38776   0.599    0.550
Refdum$PAge45to59     0.82985    0.17723   4.682 4.20e-06 ***
Refdum$PUnitedKingdom -0.01331    0.13876  -0.096    0.924
Refdum$PNoQual         0.90277    0.10907   8.277 3.47e-15 ***
Refdum$PPPMO           1.09424    0.20291   5.393 1.35e-07 ***
Refdum$PMLNECSEVW     -0.70925    0.48968  -1.448    0.148
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.789 on 320 degrees of freedom
Multiple R-squared:  0.775,     Adjusted R-squared:  0.7714
F-statistic: 220.4 on 5 and 320 DF,  p-value: < 2.2e-16
```

As per the above output from the lm() function test we found that the R squared value is 0.77 and proves to be strong and very high. The variables such as PunitedKingdom (citizens born in United Kingdom) and PNoQual (People with No Qualification) have a high importance level and this has a certainty dimension of 95% and even near 99% from our past tests. The variable PAge45to59 (Age Distribution between 45 and 59) is of not much importance.

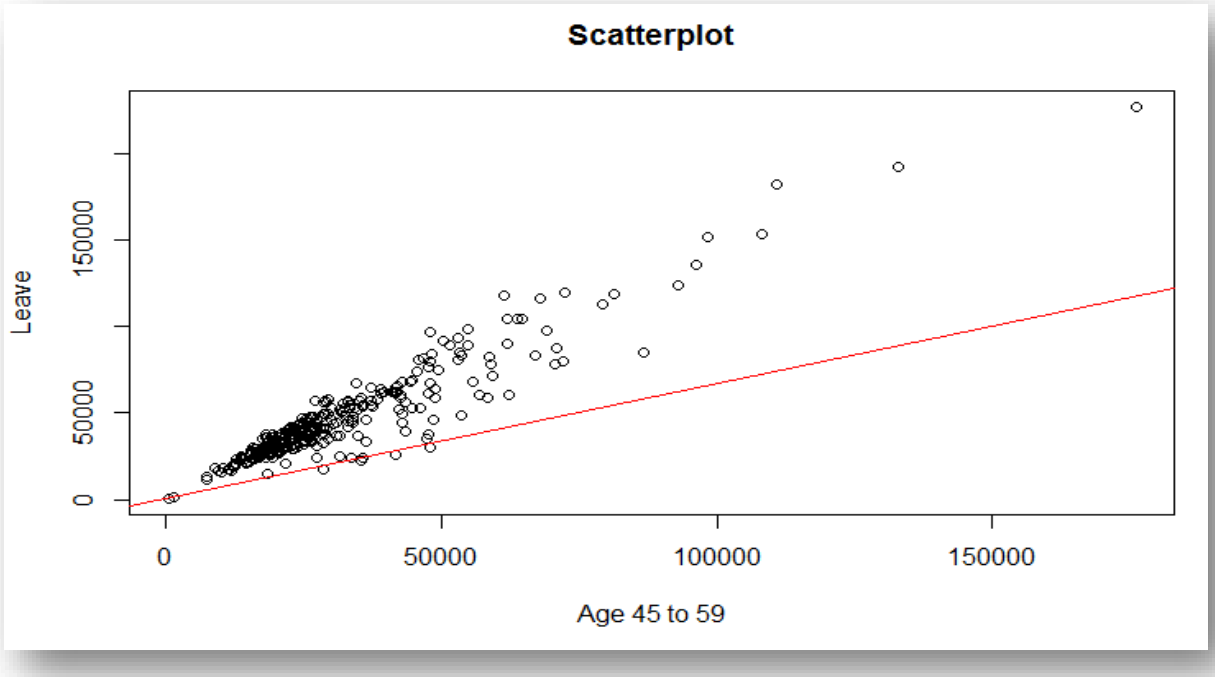
In the dataset that we created from the Brexit Referendum, altogether we have 326 observations with region codes and the area name with 47 different variables, there are only 5 variables that came out to be affecting the Brexit voting patter as per the past data exploration.

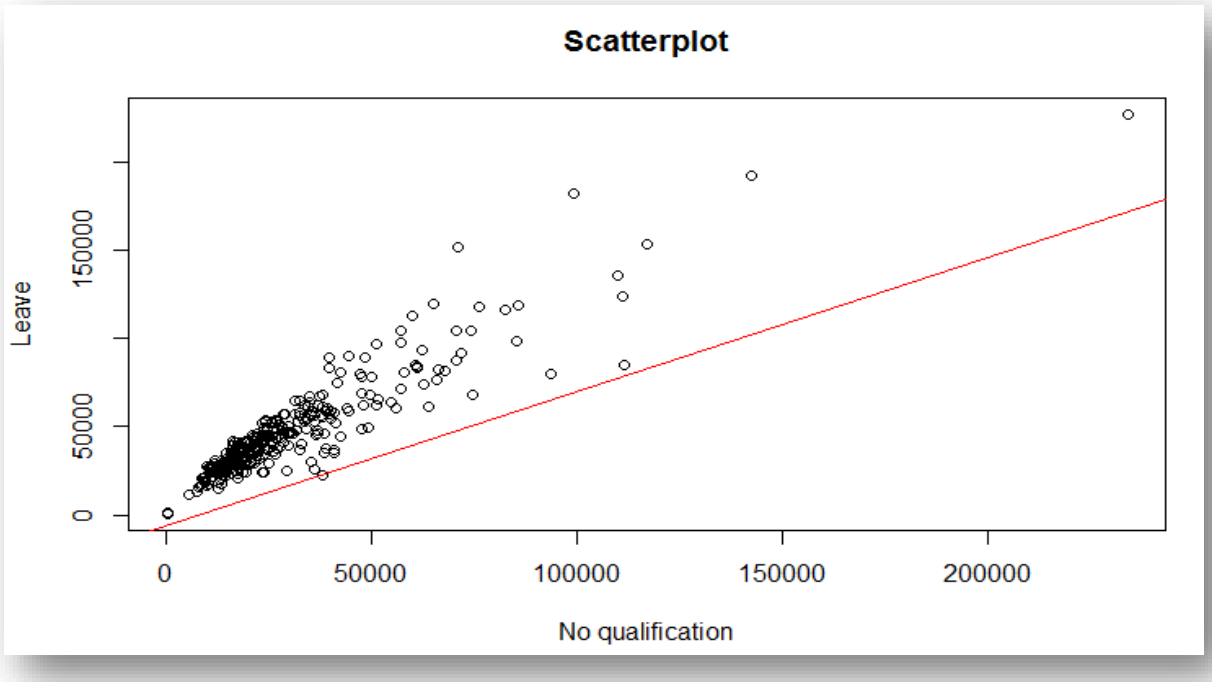
Now, we will perform the correlation test and summarise the result between different variables on the Leave variable and compare their R squared values to find potential outcome. As per the correlation test, the R squared values for PNoQual (Percentage of people with No Qualification, i.e., 0.62) is significantly larger than the value of PUnitedKingdom (Percentage of People born in Unitedkingdom, i.e., 0.38).

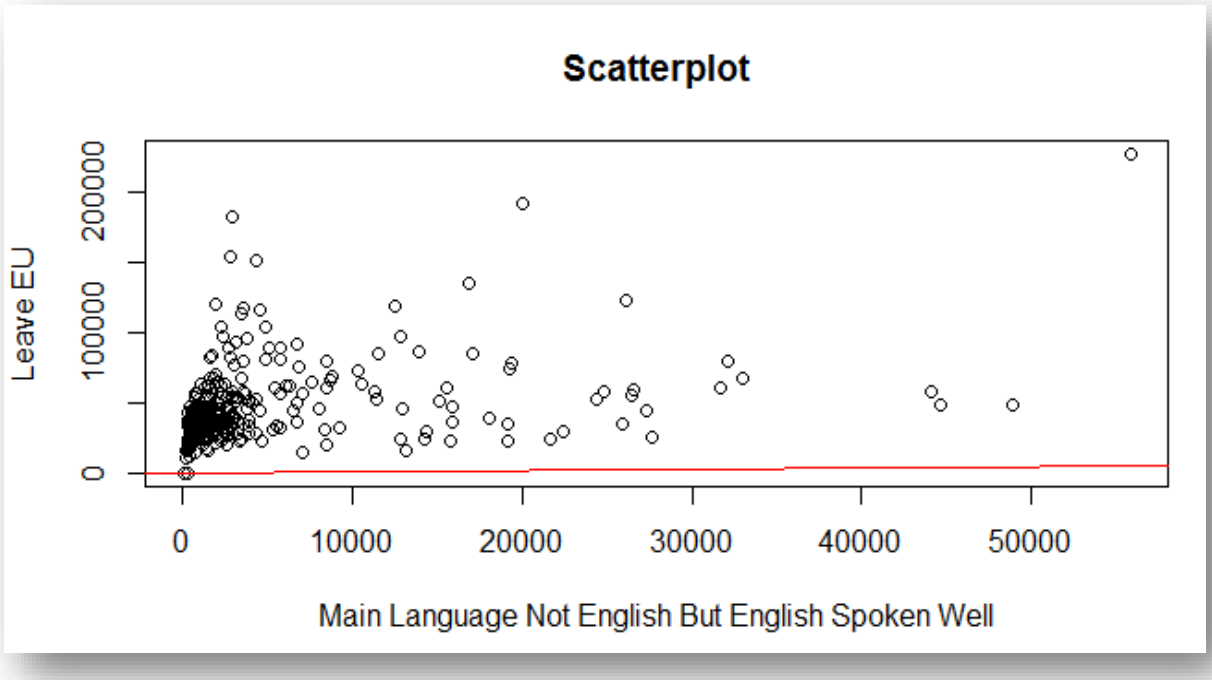
```
Call:
lm(formula = Refdum$PUnitedKingdom ~ Refdum$Pct_Leave)

Residuals:
    Min      1Q  Median      3Q     Max
-37.712  -3.661   2.003   5.674  13.286

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     52.84123    2.55645   20.67   <2e-16 ***
Refdum$Pct_Leave  0.65312    0.04614   14.16   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.331 on 324 degrees of freedom
Multiple R-squared:  0.3822,    Adjusted R-squared:  0.3802
F-statistic: 200.4 on 1 and 324 DF,  p-value: < 2.2e-16
```

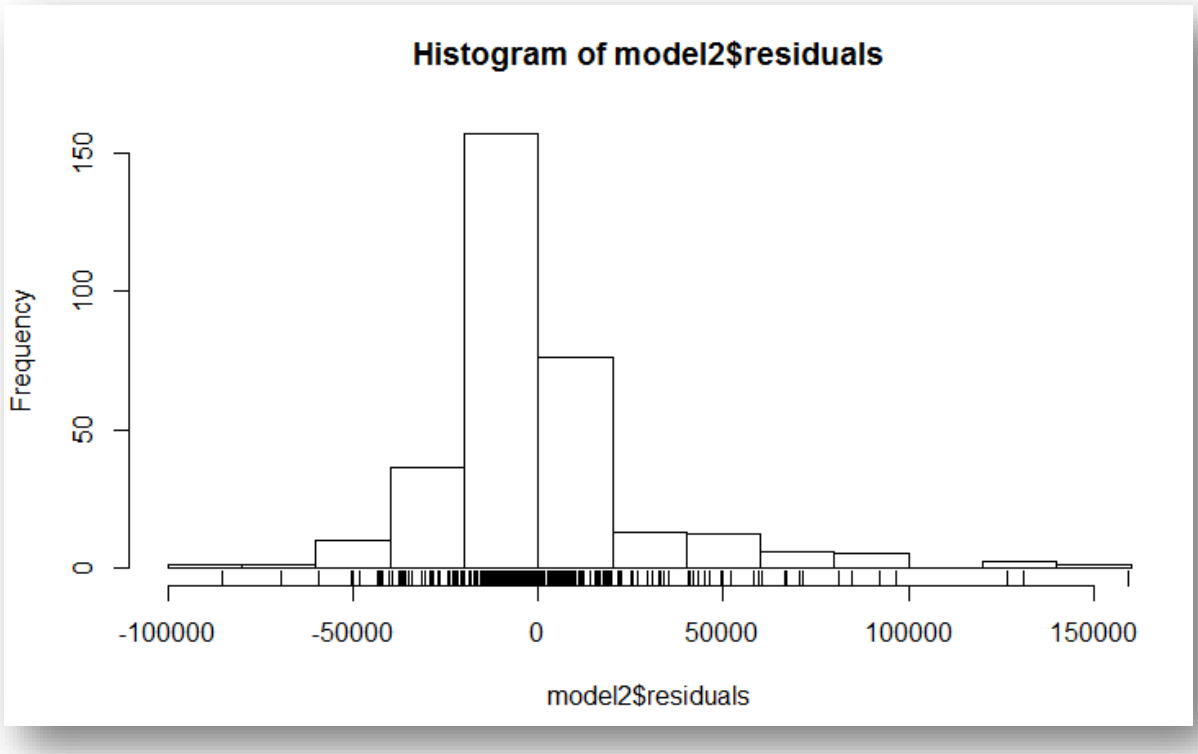
```
Call:
lm(formula = Refdum$PNoQual ~ Refdum$Pct_Leave)

Residuals:
    Min      1Q  Median      3Q     Max
-5.9800 -2.2668 -0.4336  1.7549 13.6378

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.82170    0.94474    0.87    0.385
Refdum$Pct_Leave  0.39198    0.01705   22.99   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.079 on 324 degrees of freedom
Multiple R-squared:   0.62,     Adjusted R-squared:  0.6188
F-statistic: 528.6 on 1 and 324 DF,  p-value: < 2.2e-16
```

There is a strong Positive correlation between the PAge45to59 and PNoQual with the Pct_Leave Variable and is proved by the below scattered plots.




Scatterplot
Leave
0
50000
150000
0
50000
100000
150000
200000
No qualification


Scatterplot
Leave EU
0
100000
200000
0
10000
20000
30000
40000
50000
Main Language Not English But English Spoken Well



Conclusion

As the data is highly unreliable due to the fact that the data used for this literature review is from the nomis data and till year 2011 we cannot predict the real outcome of the Referendum 2016. The data we used is only for the study purpose and to find the different variables that could affect the results.

When we started with the data exploration, we found that the null hypothesis testing resulted as no association between the PUnitedKindgom (% of People born in UK) and the Pct_Leave variable of the Referendum dataset. The PNoQual (people with no qualification) has a remarkable influence over the Pct_Leave variable. From the Previous correlation test we found that the values or R squared and r for the variable PUnitedKigdom (born in UK) are significantly lower than the values of PNoQual (People with No Qualification) variable.

It was additionally intriguing to see that the bi-variate relationships with age and those conceived in the UK and how it influenced the Brexit vote. From the test we have performed, we found that null hypothesis can be rejected for the variable PAge45to59. With 95 percent of confidence and 5 percent of marginal error from the outcome of the programming we can say that there exists a positive coefficient of correlation between PUnitedKingdom and PAge45to59. The variable PNoQual, i.e., Percentage of voters with no Qualification played a dominant role and is the main cause of Brexit as per our findings for this literature.

References: -

Barnett, A., 2017. *The lure of greatness: England's Brexit & Americas Trump*, London: Unbound.

Goodwin, Matthew. "Brexit Vote Explained: Poverty, Low Skills and Lack of Opportunities." *JRF*, 17 Dec. 2018, www.jrf.org.uk/report/brexit-vote-explained-poverty-low-skills-and-lack-opportunities.

KS102UK (Age Structure) - Nomis - Official Labour Market Statistics. 2019. *KS102UK (Age Structure) - Nomis - Official Labour Market Statistics*. [ONLINE] Available at: https://www.nomisweb.co.uk/census/2011/ks102uk. [Accessed 07 January 2019].

KS501EW (Qualifications and students) - Nomis - Official Labour Market Statistics. 2019. *KS501EW (Qualifications and students) - Nomis - Official Labour Market Statistics*. [ONLINE] Available at: https://www.nomisweb.co.uk/census/2011/ks501ew. [Accessed 07 January 2019].

KS608UK (Occupation) - Nomis - Official Labour Market Statistics. 2019. *KS608UK (Occupation) - Nomis - Official Labour Market Statistics*. [ONLINE] Available at: https://www.nomisweb.co.uk/census/2011/ks608uk. [Accessed 07 January 2019].

Liddle, R., 2016. *The Risk of Brexit: The Politics of a Referendum*. 2nd ed. Unit A, Whitacre, 26-34 Stannary Street, London, SE11 4AB: Rowman & Littlefield International Ltd.

MACSHANE, D., 2015. *BREXIT: HOW BRITAIN WILL LEAVE EUROPE*. 1st ed. 6 Salem Road, London W2 4BU: I.B.Tauris & Co. Ltd.

Merritt, G., 2016. *Slippery Slopes: Europe's Troubled Future*. 1st ed. Great Clarendon Street, Oxford, OX2 6DP, United Kingdom: 198 Madison Avenue, New York, NY 10016, United States of America.

Mindus, P., 2017. *European Citizenship after Brexit: Freedom of Movement and Rights of Residence*. 1st ed. Gewerbestrasse 11, 6330 Cham, Switzerland: Springer International Publishing AG.

QS203UK (Country of birth) - Nomis - Official Labour Market Statistics. 2019. *QS203UK (Country of birth) - Nomis - Official Labour Market Statistics*. [ONLINE] Available at: https://www.nomisweb.co.uk/census/2011/qs203uk. [Accessed 07 January 2019].

Rocks, C. (2019). *EEA workers in the London labour market*. [eBook] City Hall the Queens Walk London SE1 2AA: Greater London Authority. Available at: https://www.london.gov.uk/sites/default/files/eea-workers-in-london-cin-56.pdf [Accessed 7 Jan. 2019].

Shipman, T., 2016. *All Out War: The Full Story of How Brexit Sank Britain's Political Class*. 1st ed. 1 London Bridge Street London SE1 9GF: William Collins.