Was the Brexit vote rooted in English Nationalism?

Introduction:

This report will look at English nationalism within the United Kingdom, regarding the 2016 European Union referendum vote, which has been defined as being driven by English Nationalism (O'Toole, 2016). To answer this effectively the report will be structured into different sections. Firstly, a literature review which will look at existing literature regarding the European Union referendum and the role of English nationalism. The second section will outline the different methods, variables and data used. The third will report the findings and analyse the substantive meaning; is there a statistically significant link between the English nationalism and the Brexit vote?

Literature Review:

On the 23rd June 2016, the United Kingdom and Gibraltar cast their votes in the European Union referendum, to decide if they should leave or remain in the European Union (EU). With the UK voting to leave the EU by 51.9%, the results clearly reflect the disparities within the UK (Black, 2016). In Scotland 63%, Northern Ireland 55.8% and Gibraltar 95.9% voted remain whilst in England 53.4% and Wales 52.5% voted leave (UK Electoral Commission, 2016). Considering the clear divide in votes it has been suggested that 'Brexit is an English Nationalist Movement' (O'Toole, 2016). This forms the focus of this report; to what extent does a link exist between English identification and views towards Britain's policy regarding the EU referendum?
Firstly, to understand English nationalism it is important to look at Britain's imperial past which conceptualises the argument that; English nationalism is conceived as undeniably imperial nationalism, that can be traced back to England's establishment of the British Empire (Black, 2018). England was defined as the 'inner creator' of the United Kingdom (Kumar, 2006), whilst English nationalism was linked to a history of imperial prestige and the emergence of Britain as an imperial and industrial superpower (Black, 2018). This is where factors such as age and national identity are useful when it comes to understanding the EU referendum vote, firstly because the referendum results clearly highlight the disparities between age and EU membership; younger voters were more likely to vote remain than older voters, only 27% of 18-24-year-olds voted to leave compared to 60% of people aged 65 years and over (BBC News, 2016). Secondly, if English nationalism was established through imperial nationalism, by England's success in establishing the British Empire. The referendum can be defined as the consequence of English discontent, an attempt to recapture a sense of stability and community that the empire once provided (Gilroy, 2005).

Gerald Newman (1987) identifies that; nationalism is an ideology with a psychological nature, the main element of this nature regards the importance of an 'out group' in the formation of 'in group' consciousness and discipline (Newman 1987). This is an important feature to identify when observing the nationalist undertones in EU referendum regarding debates on immigration (Brown, 2017). Which played a large role in the Leave Party's campaign, primarily led by the United Kingdom Independence Party (UKIP), who's support surged to 27.5% in the European election of 2014, a considerable growth from the 16.5% they received in the 2009 vote, it appears that UKIP have struck an alliance with many voters on the issue of immigration (BBC News 2014). Concerns regarding immigration can involve a variety of different factors which is why certain socio-economic factors will be taken into consideration. For example, a large proportion of northern, labour-held constituencies received a high leave turnout, such a Middlesbrough and Stoke-on-Trent (Goodwin, Heath, 2016), this is also true for traditionally labour held areas in Wales, which shows how the Leave support manifested in areas which were more economically disadvantaged, where education levels are low and the local population is heavily white (Goodwin, Heath, 2016). Whilst cities such as London and Edinburgh saw some of the highest votes to remain (BBC News, 2016), this highlights how locality and the economic stability of an area played a significant role regarding the EU referendum.

Data, Methods and Variables:

The dataset I have chosen is the 2016 British Social Attitudes survey. The population was surveyed from July to November 2016 and the referendum was held on the 23rd June 2016, so it will be the best representation of the attitudes of the British public at the closest time to the referendum. This survey is conducted annually, which is a considerable strength as the survey is kept up to date each year to identify any changes in attitudes. The survey uses multi-stage stratified random sample to collect the data, to make it representative to the population. It is designed to produce a representative sample of adults aged 18 or over (British Social Attitudes, 2016), this age restriction would normally be identified as a limitation but because of the focus on the EU referendum vote, however this can be accepted as under 18-year-olds are not eligible to vote. The sampling frame is confined to those living in private households (British Social Attitudes, 2016). This is a significant limitation of this sample as it does not include people living in institutions, such as university halls, which means that a proportion of students, primarily young adults in their first year of university will not be included in the report.

The dependant variable I have chosen for my research is ‘What should Britain’s long-term policy be? To leave or remain in the EU?’ as this represents the main component of my study, to find out if there is a link between English identification and the Brexit vote. The independent variable/main predictor will be ‘Do you think of yourself as more English or British (England only)’, this will be recoded into three sections, to make the results easier to interpret. In response to the literature there are a variety of other controlled variables I’m going to use to find out to what extend a link exists, these include:

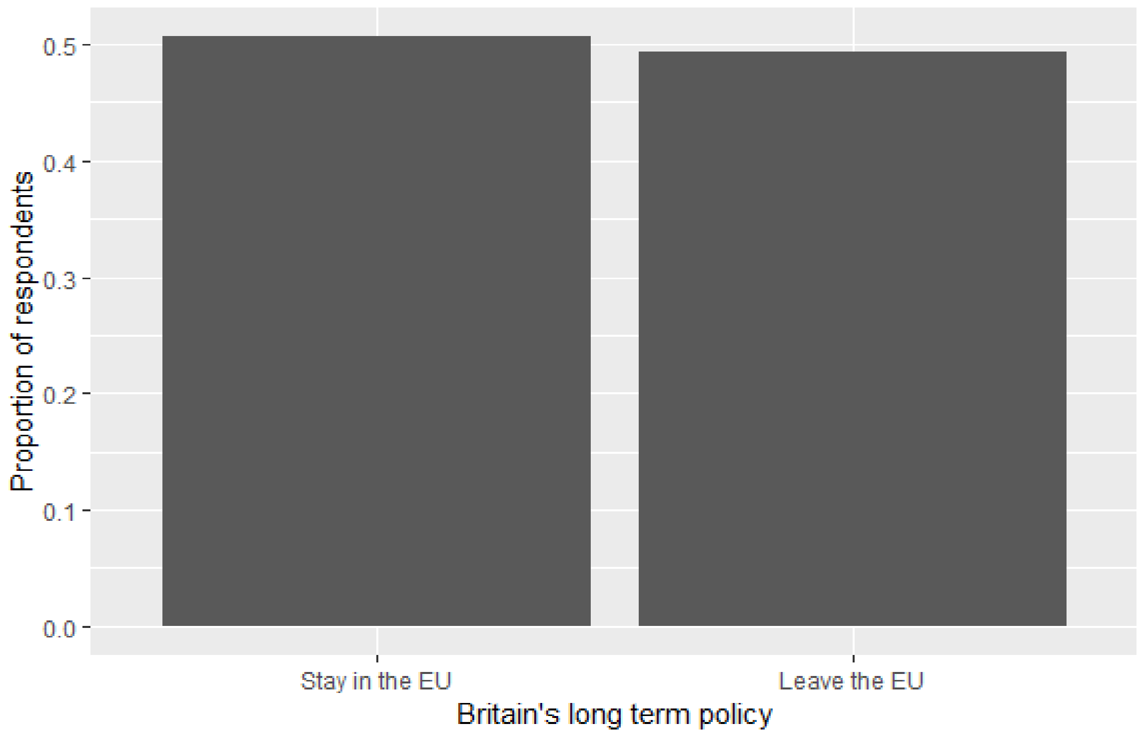
- Party identification
- Education
- Employment status
- Immigration concern
- Occupation
- How would you describe the place where you live?
- Age

The methods I am going to use in my report include tables with the counts/percentages and bar charts because my dependent variable is categorical. For the bivariate analysis a variety of methods will be used. These include crosstabulation tables to analyse the distribution, chi-squared tests to test if there is significant evidence to reject the null hypothesis, clustered and stacked bar charts to represent the data. The model which will be used in this report is a logistic regression model because the dependant variable is a binary categorical variable; a variable with only two values.

Findings:

An analysis of the British Social Attitudes Survey 2016 highlights the division regarding the EU referendum vote in 2016, when asked the question ‘What do you think Britain’s long-term policy should be?’, 50.6% answered to stay in the EU, whilst 49.4% answered to leave (Table A1), you can see this in the bar chart below:

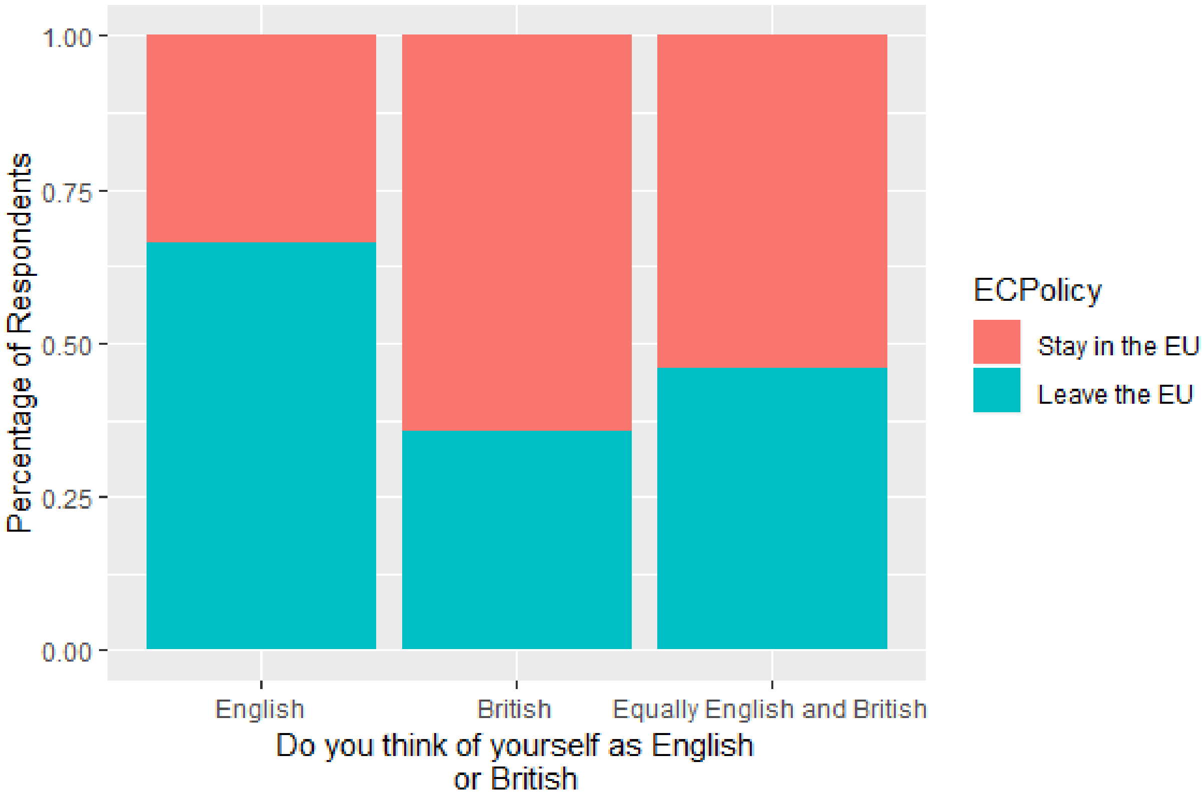
Figure 1: What should Britain’s long-term policy be regarding the EU Referendum?



This division is further amplified when looking at identification as British or English, 66.2% of people who identified as English voted to leave the EU, whilst 64.4% of people who identified as British voted to remain in the EU, which indicates that English nationalism is a contributing factor regarding the Brexit vote, you can see this represented in the graph below:

Figure 2: Attitudes towards Britain's EU Policy and identification as English or British (%)

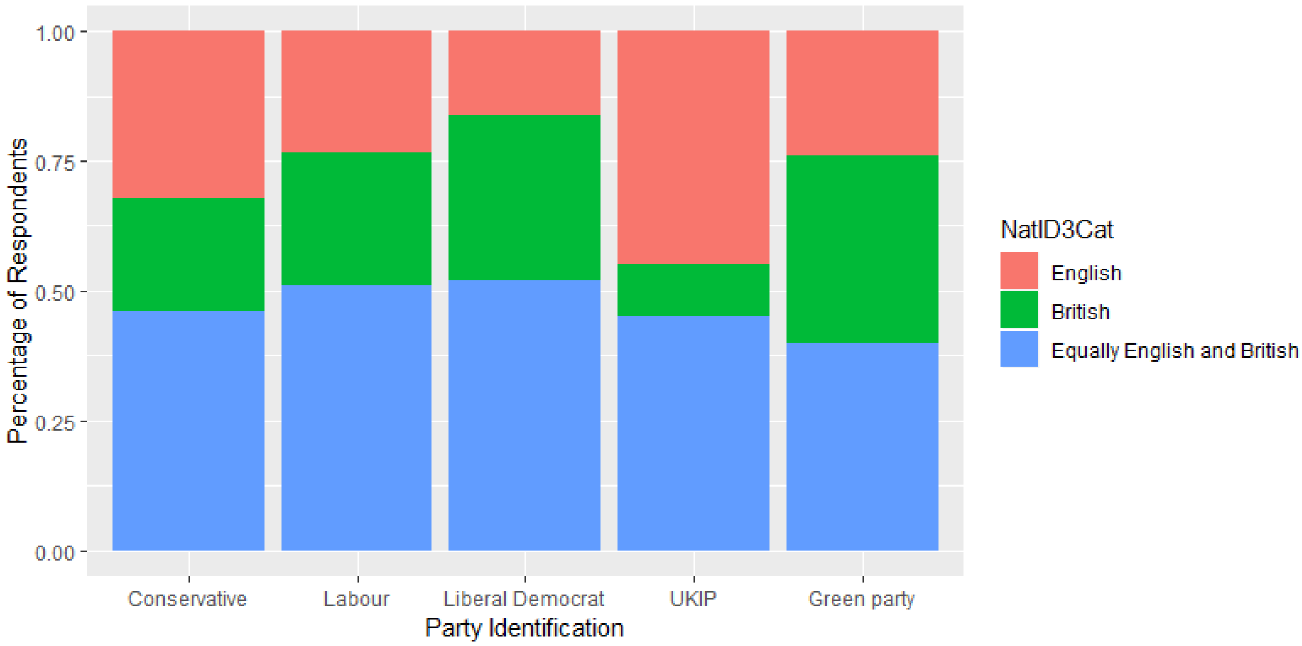
However, other factors partially explain the voting behaviour in the referendum, such



as age. Young voters aged 18-25 were more likely to vote to stay in the EU compared to those over 66 (72% compared with 36.4%) (TableA2). Age categories also follow similar trend when looking at national identification, those over 66 are the highest percentage of people who identified as English, whilst 18-25-year olds were the highest to identify as English and British (36.8% compared to 56%) (Table A3). The distribution of this variable is a considerable weakness as there are far more people in the over 66 category compared to the 18-25 year olds (Table A3), however these results illuminate what has been identified in the literature, and therefore could represent the changing attitudes of different generations, however it might not be very useful in further analysis due to the disparity in the categories.

Another interesting factor which can link English nationalism to the EU referendum is the concern of immigration. Those who identified as English were more likely to have a concern regarding immigration than those who identified as British (47% compared to 27.6%) (TableA4). Similarly, party identification to UKIP and identifying as English instead of British had a significant relationship (44.9% compared to 10.1%) (Table A5). Whilst it had already been identified in the literature UKIP's active role in the leave campaign, this further identification of Englishness among the party members makes the research question more plausible.

Figure 3: Identifying as English or British within political parties



Education was another factor which had a large impact, the more educated you were the more likely you are to vote to stay in the EU. Graduates were nearly 3 times more likely to vote to stay in the EU compared to those with no qualification (78.7% compared to 27.3%) (Table A6). It was identified in the literature that economically disadvantaged areas were more likely to vote leave, therefore the preliminary analysis included variables identifying wealth distribution such as: employment status, occupation and county of current residence. When these variables were tested against the main predictor, the results varied from one category to another and did not follow any trend (Table A7, A8, A9), suggesting that there is no causal link between national identification and these variables.

I fitted two logistic regression models both with Britain's long-term policy; to leave or stay in the EU as the binary dependent variable, with English or British identification as the main predictor.

Table 2: Logistic Regression Models Summary

	Leave the EU (Model1)				**Leave the EU (Model2)**			
Predictors	*Log-Odds*	*std. Error*	*CI*	*p*	*Log-Odds*	*std. Error*	*CI*	*p*
Intercept	0.67	0.12	0.44 – 0.91	**<0.001**	-0.84	0.21	-1.25 – -0.43	**<0.001**

British	-1.27	0.18	-1.62 – -0.92	**<0.001**	-0.79	0.20	-1.19 – -0.39	**<0.001**
Equally English and British	-0.83	0.15	-1.13 – -0.54	**<0.001**	-0.68	0.17	-1.02 – -0.35	**<0.001**
Immigration concern					1.21	0.15	0.91 – 1.51	**<0.001**
A Level					0.99	0.20	0.60 – 1.38	**<0.001**
O Level/CSE					1.39	0.21	0.98 – 1.80	**<0.001**
No Qualification					1.90	0.24	1.44 – 2.37	**<0.001**
Labour					-0.46	0.16	-0.78 – -0.15	**0.004**
Liberal Democrat					-0.59	0.28	-1.14 – -0.04	**0.037**
UKIP					2.39	0.61	1.20 – 3.59	**<0.001**
Green Party					-0.68	0.47	-1.60 – 0.24	0.149
Observations	1078				1078			
Cox & Snell's R^2 / Nagelkerke's R^2	0.052 / 0.069				0.269 / 0.359			
AIC	1442.741				1177.904			

In model 1, the coefficients show that the log odds of voting to leave the EU are significantly smaller for those who identify as British (b=-1.27, z=-7.1, p<0.001) (how to calculate: Log odd/SD) and who identify as equally as both (b=-0.8, z=-5.5, p<0.001). When expressing the coefficients as odds-ratio, you can see people who identify as British are 72% less likely to vote to leave the EU, along with people who identify equally as both are 57% less likely to vote to leave than those who identify as English. This confirms what we identified previously in the literature that you are more likely to vote leave if you identify yourself as English. The second model was fitted with immigration concern, party identification and education as independent variables. You can see that after controlling these variables the log odds of voting to leave the EU have increased for both British (b=-0.8, z=--3.9, p<0.001), and those who identify equally as both ($b=-0.7$, $z=-4$, $p<0.001$). However, they are still significantly less likely to vote to leave compared to English voters, this expressed as a odds-ratio shows that 55% of British identifiers and 45% of people who identify as equally British and English are less likely to vote to leave the EU compared to English identifiers.

Looking at the coefficients for the controlled variables you can see that there is a high positive association between UKIP and voting to leave the EU ($b=2.4$,$z=3.9$, $p<0.001$) which agrees with what we have identified previously, that UKIP voters are more likely to vote leave than any other party. Immigration concern is also significantly related, those who mentioned having a concern about immigration were more likely to vote to leave the EU ($b=1.2$, $z=7.8$, $p<0.001$). Lastly, education was significantly related, the model finds that people who had no educational qualifications were more likely to vote leave than their counterparts ($b=1.9$, $z=7.9$, $p<0.001$) compared to those with high school education ($b=1.4$, $z=6.6$, $p<0.001$) and individuals with A-Levels or equivalent ($b=1$, $z=5$, $p<0.001$). Overall, the results from both models suggest that there is a significant relationship between national identity and voting to leave/stay in the EU, this remains true when controlling the independent variables.

References

BBC News. (2016). EU Referendum: The result in maps and charts. Available Online: https://www.bbc.co.uk/news/uk-politics-36616028

BBC News. (2016). Lincolnshire records UK's highest Brexit vote. Available Online: https://www.bbc.co.uk/news/uk-politics-eu-referendum-36616740

Black, J. (2018). From mood to movement: English nationalism, the European Union and taking back control. The European Journal of Social Science.

Brown, H. (2017). Post-Brexit Britain: Thinking about 'English Nationalism' as a factor in the EU referendum. International Politics Reviews.

British Social Attitudes Survey. (2016). User Guide. Available Online: http://doc.ukdataservice.ac.uk/doc/8252/mrdoc/pdf/8252_bsa_2016_user_guide.pdf

Gilroy, P. (2005). Postcolonial Melancholia. New York. Columbia University Press.

Goodwin, M. Heath, O. (2016). The 2016 Referendum, Brexit and the Left Behind: An Aggregate-Level Analysis of the Result.

Kumar, K. (2006). English and French National Identity: Comparisons and Contrasts.

Newman, G. (1987). The Rise of English Nationalism. A Cultural History 1740-1830. New York. St. Martin's Press.

O'Toole, F. (2016). The Guardian. Brexit is being driven by English Nationalism. And will End in Self-rule. Available Online: https://www.theguardian.com/commentisfree/2016/jun/18/england-eu-referendum-brexit

The Electoral Commission. (2016). Available Online: https://www.electoralcommission.org.uk/

Appendices A.

Table A1: What should Britain's long term policy be?

Val	*frq*	*raw.prc*	*valid.prc*	*cum.prc*
Stay in the EU	546	50.65	50.65	50.65
Leave the EU	532	49.35	49.35	100

total N=1078 · valid N=546 · $\bar{x}$=1.49 · σ=0.50

Table A2: Britain's EU Policy and Age Categories

ECPolicy	*RAgecat3*						***Total***
	18-25	26-35	36-45	46-55	56-65	>66	
Stay in the EU	36 72 %	76 58.5 %	115 68.9 %	102 54.8 %	90 45.7 %	127 36.5 %	546 50.6 %
Leave the EU	14 28 %	54 41.5 %	52 31.1 %	84 45.2 %	107 54.3 %	221 63.5 %	532 49.4 %
Total	50 100 %	130 100 %	167 100 %	186 100 %	197 100 %	348 100 %	1078 100 %

χ^2=65.599 · df=5 · Cramer's V=0.247 · p=0.000

Table A3: National Identity and Age Categories

NatID3Cat	*RAgecat3*						***Total***
	18-25	26-35	36-45	46-55	56-65	>66	
English	16 32 %	25 19.2 %	35 21 %	41 22 %	63 32 %	128 36.8 %	308 28.6 %
British	6 12 %	41 31.5 %	41 24.6 %	50 26.9 %	38 19.3 %	77 22.1 %	253 23.5 %
Equally English and British	28 56 %	64 49.2 %	91 54.5 %	95 51.1 %	96 48.7 %	143 41.1 %	517 48 %

Total	50 100 %	130 100 %	167 100 %	186 100 %	197 100 %	348 100 %	1078 100 %

χ^2=34.525 · df=10 · Cramer's V=0.127 · p=0.000

Table A4: National identity and Immigration Concern

NatID3Cat	*CrPImm*		***Total***
	Not mentioned	Mentioned	
English	163 52.9 %	145 47.1 %	308 100 %
British	183 72.3 %	70 27.7 %	253 100 %
Equally English and British	322 62.3 %	195 37.7 %	517 100 %
Total	668 62 %	410 38 %	1078 100 %

χ^2=22.246 · df=2 · Cramer's V=0.144 · p=0.000

Table A5: National identity and Party alignment

NatID3Cat	*PartyID3*					***Total***
	Conservative	Labour	Liberal Democrat	UKIP	Green party	
English	172 32.3 %	86 23.2 %	13 16 %	31 44.9 %	6 24 %	308 28.6 %
British	116 21.8 %	95 25.7 %	26 32.1 %	7 10.1 %	9 36 %	253 23.5 %
Equally English and British	245 46 %	189 51.1 %	42 51.9 %	31 44.9 %	10 40 %	517 48 %
Total	533 100 %	370 100 %	81 100 %	69 100 %	25 100 %	1078 100 %

χ^2=30.120 · df=8 · Cramer's V=0.118 · p=0.000

Table A6: Britain's EU Policy and Educational Qualifications

ECPolicy	*HEdQual3* Degree	Higher educ below degree/A level	O level or equiv/CSE	No qualification	***Total***
Stay in the EU	226 78.7 %	164 51.9 %	104 36.5 %	52 27.4 %	546 50.6 %
Leave the EU	61 21.3 %	152 48.1 %	181 63.5 %	138 72.6 %	532 49.4 %
Total	287 100 %	316 100 %	285 100 %	190 100 %	1078 100 %

χ^2=154.890 · df=3 · Cramer's V=0.379 · p=0.000

Table A7: National Identity and Economic Class

NatID3Cat	*RClassGp* Managerial & professional occups	Intermediate occupations	Employers in small org; own account workers	Lower supervisory & technical occupations	Semi-routine & routine occupations	***Total***
English	125 24.9 %	48 29.8 %	38 37.6 %	26 34.7 %	71 29.8 %	308 28.6 %
British	146 29 %	35 21.7 %	22 21.8 %	11 14.7 %	39 16.4 %	253 23.5 %
Equally English and British	232 46.1 %	78 48.4 %	41 40.6 %	38 50.7 %	128 53.8 %	517 48 %
Total	503 100 %	161 100 %	101 100 %	75 100 %	238 100 %	1078 100 %

χ^2=24.338 · df=8 · Cramer's V=0.106 · p=0.002

Table A8: National Identity and Economic Position

NatID3Cat	*EcoPos6* Employee	Self-employed	Unemployed	Looking after home	In f-t education	Retired	***Total***

English	113 23.7 %	27 28.7 %	9 23.7 %	11 20.4 %	5 23.8 %	143 36.2 %	308 28.6 %
British	113 23.7 %	26 27.7 %	6 15.8 %	17 31.5 %	4 19 %	87 22 %	253 23.5 %
Equally English and British	250 52.5 %	41 43.6 %	23 60.5 %	26 48.1 %	12 57.1 %	165 41.8 %	517 48 %
Total	476 100 %	94 100 %	38 100 %	54 100 %	21 100 %	395 100 %	1078 100 %

χ^2=24.582 · *df=10 · Cramer's V=0.107 · p=0.006*

Table A9: National Identity and Area of Residence

NatID3Cat	*EcoPos6* Employee	Self-employed	Unemployed	Looking after home	In f-t education	Retired	***Total***
English	113 23.7 %	27 28.7 %	9 23.7 %	11 20.4 %	5 23.8 %	143 36.2 %	308 28.6 %
British	113 23.7 %	26 27.7 %	6 15.8 %	17 31.5 %	4 19 %	87 22 %	253 23.5 %
Equally English and British	250 52.5 %	41 43.6 %	23 60.5 %	26 48.1 %	12 57.1 %	165 41.8 %	517 48 %
Total	476 100 %	94 100 %	38 100 %	54 100 %	21 100 %	395 100 %	1078 100 %

χ^2=24.582 · *df=10 · Cramer's V=0.107 · p=0.006*

Table A10: Logistic Regression Model 1

	Leave the EU (Model1)			
Predictors	*Log-Odds*	*std. Error*	*CI*	*p*
Intercept	0.67	0.12	0.44 – 0.91	**<0.001**

British	-1.27	0.18	-1.62 – -0.92	**<0.001**
Equally English and British	-0.83	0.15	-1.13 – -0.54	**<0.001**
Observations	1078			
Cox & Snell's R^2 / Nagelkerke's R^2	0.052 / 0.069			
AIC	1442.741			

Table A11: Logistic Regression Model 2

	Leave the EU (Model2)			
Predictors	*Log-Odds*	*std. Error*	*CI*	*p*
Intercept	-0.84	0.21	-1.25 – -0.43	**<0.001**
British	-0.79	0.20	-1.19 – -0.39	**<0.001**
Equally English and British	-0.68	0.17	-1.02 – -0.35	**<0.001**
Immigration Concern	1.21	0.15	0.91 – 1.51	**<0.001**
A-Level	0.99	0.20	0.60 – 1.38	**<0.001**
O-Level/Equiv	1.39	0.21	0.98 – 1.80	**<0.001**
No Qualification	1.90	0.24	1.44 – 2.37	**<0.001**
Labour	-0.46	0.16	-0.78 – -0.15	**0.004**
Liberal Democrat	-0.59	0.28	-1.14 – -0.04	**0.037**
UKIP	2.39	0.61	1.20 – 3.59	**<0.001**
Green Party	-0.68	0.47	-1.60 – 0.24	0.149
Observations	1078			
Cox & Snell's R^2 / Nagelkerke's R^2	0.269 / 0.359			
AIC	1177.904			

Appendices B.

```
#Load data and libraries
library(foreign)
bsa16<-read.dta("bsa16_to_ukda.dta")
library(psych)
```

```
library(MASS)
library(car)
library(ggplot2)
library(sjPlot)
library(sjlabelled)
library(sjmisc)
library(sjstats)

#Run table to see the distribution of national identity in England
table(bsa16$NatId)

#Create new variable with national identity in 3 categories
bsa16$NatID3Cat <- NA
bsa16$NatID3Cat[bsa16$NatId=="English not British"|bsa16$NatId=="More English
than British"] <- 1
bsa16$NatID3Cat[bsa16$NatId=="More British than English"|bsa16$NatId=="British
not English"] <- 2
bsa16$NatID3Cat[bsa16$NatId=="Equally English and British"] <- 3
bsa16$NatID3Cat[bsa16$NatId=="Not applicable"|bsa16$NatId=="Other description
(WRITE IN)"|
                 bsa16$NatId=="(None of these)"|bsa16$NatId=="Don't
know"|bsa16$NatId=="Refusal"] <- NA
bsa16$NatID3Cat <- as.factor(bsa16$NatID3Cat)
levels(bsa16$NatID3Cat)<-c("English", "British", "Equally English and British")
table(bsa16$NatID3Cat)

#Run table to see the distribution of 'What should Britain's long-term policy be?'
table(bsa16$ECPolicy2)

#create new variable 'ECPolicy' in 2 categories
bsa16$ECPolicy <- NA
bsa16$ECPolicy[bsa16$ECPolicy2=="stay in the EU and try to reduce the EU
powers"|
                bsa16$ECPolicy2=="stay in the EU and try to keep the EU powers as they
are"|
                bsa16$ECPolicy2=="stay in the EU and try to increase the EU powers"] <-
1
bsa16$ECPolicy[bsa16$ECPolicy2=="leave the European Union,"] <- 2
bsa16$ECPolicy[bsa16$ECPolicy2=="Schedule not
applicable"|bsa16$ECPolicy2=="Item not applicable"|
                bsa16$ECPolicy2=="work for the formation of a single European
government"|
                bsa16$ECPolicy2=="Don't Know"|bsa16$ECPolicy2=="Refusal"] <- NA
bsa16$ECPolicy <- as.factor(bsa16$ECPolicy)
levels(bsa16$ECPolicy)<-c("Stay in the EU", "Leave the EU")
table(bsa16$ECPolicy)

#Run table to see the distribution of age categories
table(bsa16$RAgecat3)
```

```
#recode age to exclude DK/Refusal
bsa16$RAgecat3[bsa16$RAgecat3=="DK/Ref"] <-NA
bsa16$RAgecat3<-droplevels(bsa16$RAgecat3)
table(bsa16$RAgecat3)

#run table to see the distribution of occupation
table(bsa16$RClassGp)

#recode to exclude 'Not aplicable' and 'Not classifiable'
bsa16$RClassGp[bsa16$RClassGp=="Not classifiable" |bsa16$RClassGp=="Not
applicable"] <-NA
bsa16$RClassGp<-droplevels(bsa16$RClassGp)
table(bsa16$RClassGp)

#run table to see the distribution of current concern about immigration
table(bsa16$CrPImm)
#recode occupation to define missing values
bsa16$CrPImm[bsa16$CrPImm=="Schedule not applicable"
|bsa16$CrPImm=="Item not applicable"|
bsa16$CrPImm=="Don't Know"|bsa16$CrPImm=="Refusal"] <-NA
bsa16$CrPImm<-droplevels(bsa16$CrPImm)
table(bsa16$CrPImm)

#run table to see the distribution of employment status
table(bsa16$REconPos)
#create new variable for employment status with 6 categories
bsa16$EcoPos6 <- NA
bsa16$EcoPos6[bsa16$REconPos=="Employee (full-
time)"|bsa16$REconPos=="Employee (part-time)"|
bsa16$REconPos=="In work (status not known)"]<-1
bsa16$EcoPos6[bsa16$REconPos=="Self-employed (p-
t)"|bsa16$REconPos=="Self-employed (f-t)"]<-2
bsa16$EcoPos6[bsa16$REconPos=="Unemployed"|bsa16$REconPos=="Waiting to
take up work"]<-3
bsa16$EcoPos6[bsa16$REconPos=="Looking after the home"]<-4
bsa16$EcoPos6[bsa16$REconPos=="In f-t education"]<-5
bsa16$EcoPos6[bsa16$REconPos=="Retired"]<-6
bsa16$EcoPos6[bsa16$REconPos=="Other"|bsa16$REconPos=="Don't
know"|bsa16$REconPos=="Refusal"]<-NA
bsa16$EcoPos6 <- as.factor(bsa16$EcoPos6)
levels(bsa16$EcoPos6)<-c("Employee","Self-employed","Unemployed","Looking
after home",
"In f-t education","Retired")
table(bsa16$EcoPos6)

#run table to see the distribution of education
table(bsa16$HEdQual3)
#recode education to define missing values
bsa16$HEdQual3[bsa16$HEdQual3=="DK/Refusal/NA"]<-NA
bsa16$HEdQual3<-droplevels(bsa16$HEdQual3)
```

```
table(bsa16$HEdQual3)

#run table to see the distribution of party identification
table(bsa16$PartyID3)
#recode to define missing values
bsa16$PartyID3[bsa16$PartyID3=="Not
applicable"|bsa16$PartyID3=="Other/DK/Ref"|
bsa16$PartyID3=="Other party"|bsa16$PartyID3=="None"]<-NA
bsa16$PartyID3<-droplevels(bsa16$PartyID3)
table(bsa16$PartyID3)

#run table to see the distribution of where people live
table(bsa16$ResPres)
#recode to define missing values
bsa16$ResPres[bsa16$ResPres=="Schedule not
applicable"|bsa16$ResPres=="Item not applicable"|
bsa16$ResPres=="Don't
Know"|bsa16$ResPres=="Refusal"|bsa16$ResPres=="(Other answer (WRITE
IN))"]<-NA
bsa16$ResPres<-droplevels(bsa16$ResPres)
table(bsa16$ResPres)

#Univariate Analysis

#create object called tabECPolicy which contains the univariate distribution of
variable ECPolicy
tabECPolicy<-table(bsa16$ECPolicy)
tabECPolicy
addmargins(tabECPolicy)
#as a proportion
prop.table(tabECPolicy)
#as a percentage
prop.table(tabECPolicy)*100
#as a bar chart which excludes missing values
ggplot(bsa16[!is.na(bsa16$ECPolicy),],aes(x=ECPolicy, y= ..prop.., group = 1))+
geom_bar(stat = "Count") + xlab("Britain's long term policy")+ ylab("Proportion of
respondents")

#create object called tabID which contains the univariate distribution of variable
NatID3Cat
tabID<-table(bsa16$NatID3Cat)
tabID
addmargins(tabID)
#as a proportion
prop.table(tabID)
#as a percentage
prop.table(tabID)*100
ggplot(bsa16[!is.na(bsa16$NatID3Cat),], aes(x = NatID3Cat,y= ..prop.., group = 1)) +
geom_bar(stat = "count")+
```

```
xlab("Do you think yourself as more English or British?")+ ylab("Proportion of respondents")

#create an object called tabAge3 which contains the univariate distribution of variable AgeCat3
tabAge3<-table(bsa16$RAgecat3)
tabAge3
addmargins(tabAge3)
#as a proportion
prop.table(tabAge3)
#as a percentage
prop.table(tabAge3)*100

#create an object called tabOcc which contains the univariate distribution of variable RClassGp
tabRClassGp<-table(bsa16$RClassGp)
tabRClassGp
addmargins(tabRClassGp)
#as a proportion
prop.table(tabRClassGp)
#as a percentage
prop.table(tabRClassGp)*100

#create an object called tab which contains the univariate distribution of variable CrPImm
tabImm<-table(bsa16$CrPImm)
tabImm
addmargins(tabImm)
#as a proportion
prop.table(tabImm)
#as a percentage
prop.table(tabImm)*100

#create an object called tabEcoP which contains the univariate distribution of variable EcoPos6
tabEcoP<-table(bsa16$EcoPos6)
tabEcoP
addmargins(tabEcoP)
#as a proportion
prop.table(tabEcoP)
#as a percentage
prop.table(tabEcoP)*100

#create an object called tabEdu which contains the univariate distribution of variable HedQual3
tabEdu<-table(bsa16$HEdQual3)
tabEdu
addmargins(tabEdu)
#as a proportion
prop.table(tabEdu)
```

```
#as a percentage
prop.table(tabEdu)*100

#create an object called tabPartID which contains the univariate distribution of
variable PartID3
tabPartyID<-table(bsa16$PartyID3)
tabPartyID
addmargins(tabPartyID)
#as a proportion
prop.table(tabPartyID)
#as a percentage
prop.table(tabPartyID)*100

#create an object called tabRes which contains the univariate distribution of variable
ResPres
tabRes<-table(bsa16$ResPres)
tabRes
addmargins(tabRes)
#as a proportion
prop.table(tabRes)
#as a percentage
prop.table(tabRes)*100

#Bivariate Analysis
#Analysis of dependant variable ECPolicy and main predictor NatID3Cat

#create a crosstab which represents the disribution of NatID3Cat and ECPolicy
IDEC<-table(bsa16$NatID3Cat, bsa16$ECPolicy)
addmargins(IDEC)
#as a percentage
prop.table(IDEC,1)*100

#stacked bar chart
ggplot(data = bsa16[!is.na(bsa16$NatID3Cat)&
!is.na(bsa16$ECPolicy),],aes(x=NatID3Cat))+
geom_bar(stat = "Count", aes(fill=ECPolicy), position = "fill")+xlab("Do you think of
yourself as English
or British")+ylab("Percentage of Respondents")

#run chi-squared test
chi_IDEC<-chisq.test(IDEC, correct = F)
chi_IDEC

#Analysis of dependent variable across the independent variables

#create a crosstab which represents the disribution of ECPolicy and RAgeCat3
AgeEC<-table(bsa16$ECPolicy, bsa16$RAgecat3)
addmargins(AgeEC)
#as a percentage
prop.table(AgeEC,2)*100
```

```
#run chi-squared test
chi_AgeEC<-chisq.test(AgeEC, correct = F)
chi_AgeEC

#create a crosstab which represents the disribution of ECPolicy and RClassGp
EClass<-table(bsa16$ECPolicy, bsa16$RClassGp)
addmargins(EClass)
#as a percentage
prop.table(EClass,2)*100
#run chi-squared test
chi_EClass<-chisq.test(EClass, correct = F)
chi_EClass

#create a crosstab which represents the disribution of ECPolicy and CrPImm
ImEC<-table(bsa16$ECPolicy, bsa16$CrPImm)
addmargins(ImEC)
#as a percentage
prop.table(ImEC,1)*100
#run chi-squared test
chi_ImEC<-chisq.test(ImEC, correct = F)
chi_ImEC

#create a crosstab which represents the disribution of ECPolicy and EcoPos6
ECPos<-table(bsa16$ECPolicy, bsa16$EcoPos6)
addmargins(ECPos)
#as a percentage
prop.table(ECPos,2)*100
#run chi-squared test
chi_ECPos<-chisq.test(ECPos, correct = F)
chi_ECPos

#create a crosstab which represents the disribution of ECPolicy and HedQual3
EduEC<-table(bsa16$ECPolicy, bsa16$HEdQual3)
addmargins(EduEC)
#as a percentage
prop.table(EduEC,2)*100
#run chi-squared test
chi_EduEC<-chisq.test(EduEC, correct = F)
chi_EduEC

#create a crosstab which represents the disribution of ECPolicy and PartyID3
ECPID<-table(bsa16$ECPolicy, bsa16$PartyID3)
addmargins(ECPID)
#as a percentage
prop.table(ECPID,2)*100
#clustered bar chart
ggplot(data = bsa16[!is.na(bsa16$ECPolicy)&
!is.na(bsa16$PartyID3),],aes(x=PartyID3))+
  geom_bar(stat = "Count", aes(fill=ECPolicy), position = "fill")+xlab("Party
Identification")+ylab("Percentage of Respondents")
```

```
#run chi-squared test
chi_ECPID<-chisq.test(ECPID, correct = F)
chi_ECPID

#create a crosstab which represents the disribution of ECPolicy and ResPres
ECRes<-table(bsa16$ECPolicy, bsa16$ResPres)
addmargins(ECRes)
#as a percentage
prop.table(ECRes,2)*100
#run chi-squared test
chi_ECRes<-chisq.test(ECRes, correct = F)
chi_ECRes

#Analysis of the main predictor across the controlled variables

#create a crosstab which represents the disribution of NatID3Cat and RAgeCat3
AgeID<-table(bsa16$NatID3Cat, bsa16$RAgecat3)
addmargins(AgeID)
#as a percentage
prop.table(AgeID,2)*100
#run chi-squared test
chi_AgeID<-chisq.test(AgeID, correct = F)
chi_AgeID

#create a crosstab which represents the disribution of NatID3Cat and RClassGp
IDOC<-table(bsa16$NatID3Cat, bsa16$RClassGp)
addmargins(IDOC)
#as a percentage
prop.table(IDOC,2)*100
#run chi-squared test
chi_IDOC<-chisq.test(IDOC, correct = F)
chi_IDOC

#create a crosstab which represents the disribution of NatID3Cat and CrPImm
ImmID<-table(bsa16$NatID3Cat, bsa16$CrPImm)
addmargins(ImmID)
#as a percentage
prop.table(ImmID,1)*100
#run chi-squared test
chi_ImmID<-chisq.test(ImmID, correct = F)
chi_ImmID

#create a crosstab which represents the disribution of NatID3Cat and EcoPos6
EcoP<-table(bsa16$NatID3Cat, bsa16$EcoPos6)
addmargins(EcoP)
#as a percentage
prop.table(EcoP,2)*100
#run chi-squared test
chi_EcoP<-chisq.test(EcoP, correct = F)
chi_EcoP
```

```
#create a crosstab which represents the disribution of NatID3Cat and HedQual3
Edu<-table(bsa16$NatID3Cat, bsa16$HEdQual3)
addmargins(Edu)
#as a percentage
prop.table(Edu,2)*100
#run chi-squared test
chi_Edu<-chisq.test(Edu, correct = F)
chi_Edu

#create a crosstab which represents the disribution of NatID3Cat and PartyID3
PID<-table(bsa16$NatID3Cat, bsa16$PartyID3)
addmargins(PID)
#as a percentage
prop.table(PID,2)*100
#clustered bar chart
ggplot(data = bsa16[!is.na(bsa16$NatID3Cat)&
!is.na(bsa16$PartyID3),],aes(x=PartyID3))+
geom_bar(stat = "Count", aes(fill=NatID3Cat), position = "fill")+xlab("Party
Identification")+ylab("Percentage of Respondents")
#run chi-squared test
chi_PID<-chisq.test(PID, correct = F)
chi_PID

#create a crosstab which represents the disribution of NatID3Cat and ResPres
Res<-table(bsa16$NatID3Cat, bsa16$ResPres)
addmargins(Res)
#as a percentage
prop.table(Res,2)*100
#run chi-squared test
chi_Res<-chisq.test(Res, correct = F)
chi_Res

#Modelling - Logistic Regression

#restrict the data to only include complete cases
bsa16<-bsa16[complete.cases(bsa16$NatID3Cat,
bsa16$ECPolicy,bsa16$RAgecat3, bsa16$RClassGp,
bsa16$CrPImm, bsa16$EcoPos6, bsa16$HEdQual3, bsa16$PartyID3,
bsa16$ResPres),]

#model 1 - analysis of dependent variable and main predictor
model1 <-glm(ECPolicy~NatID3Cat, data = bsa16, family = "binomial")
summary(model1)
exp(cbind(OR=coef(model1), confint(model1)))


#model 2 - analysis of dependent variable, main predictor and all the conrolled
variables
```

```
model2 <-glm(ECPolicy~NatID3Cat + RAgecat3 + RClassGp + CrPImm + EcoPos6
+ HEdQual3 +
PartyID3 + ResPres, data = bsa16, family = "binomial")
summary(model2)

#model 3 - analysis of dependent variable, main predictor and significant variables
model3 <-glm(ECPolicy~NatID3Cat+ CrPImm + HEdQual3 + PartyID3,
        data = bsa16, family = "binomial")
summary(model3)
exp(cbind(OR=coef(model3), confint(model3)))

model1$aic
model2$aic
model3$aic
```