Determinants of Interest in the EVs in the UK

Claire Chang Liu & Noriko Magara

Nov. 13th, 2015

This paper examines the determinants of interest in electric vehicles (EVs) in the UK. First, we will describe the process of data gathering. Then we descriptively analyze the dependent variable and independent variables, and use logistic regression models to indentify the correlations. Last part will be a brief discussion of the results of the analysis.

1 Data Gathering

This analysis is based on survey data in the UK called "Opinions and Lifestyle Survey, Electric Vehicles Module, February 2014 and February 2015" (Office for National Statistics. Social Survey Division 2015). Note that registration is required to download the dataset from the UK Data Service website above. After logging in and agreeing with the terms of conditions, the datasets are available for download in STATA or other format. The registration process may take a few days. However, other documents (e.g. questionnaire, brief summary) are available at the link above without registration. This module comes with two datasets: survey results from 2014 and 2015. For our study, we use both survey years by combining both datasets together (Office for National Statistics. HM Revenue and Customs 2015).

2 Descriptive Analysis

2.1 Variables

The list of variables are shown in the table below.

Table 1: List of Variables

Names	Types	Descriptions
EVinterest	dummy	Whether the respondent is interested in EVs or not
RAGE	continuous	Respondent's age
Male	dummy	Respondent's sex
inccat	categorical	Respondent's income in 4 categories (low, low-mid, high-mid, high)
degree	dummy	Whether the respondent is a college graduate or not
licence	dummy	Whether the respondent has a valid drivers licence or not
NumCar	continuous	Number of cars in the respondent's household
DVHsize	continuous	Respondent's household size
havechildren	dummy	Whether the Respondent has children or not
Scotland	dummy	Whether the Respondent lives in Scotland or not

2.2 Summary statistics

- **EVinterest**: 18.48 % of respondents are interested in EVs
- **RAGE**: Mean=52.03, Standard deviation=17.92. The distribution of age is shown in the figure below.
- **Male**: 46.84 % of respondents are male

- **inccat**: distribution is shown in the table and the figure below. The thresholds of each category are based on 25, 50, 75 percentile gross income from the UK census.

Table 2: Distribution of Income

Income	Range	Percentage
Low	Below 14,559GDP	50.53
Low-Mid	14,560GBP~20,799GBP	15.09
High-Mid	20,800GBP~25,999GBP	8.95
High	26,000GBP and more	25.44

- **degree**: 24.39 % of respondents have college degree
- **licence**: 76.55 % of respondents have valid drivers licence
- **NumCar**: Mean=1.14, Standard deviation=0.86. The distribution of the number of cars in respondent's household is shown in the figure below.
- **DVHsize**: Mean=2.28, Standard deviation=1.24. The distribution of household size is shown in the figure below.
- **havechildren**: 28.07 % of respondents have children
- **Scotland**: 8.71 % of respondents lives in Scotland

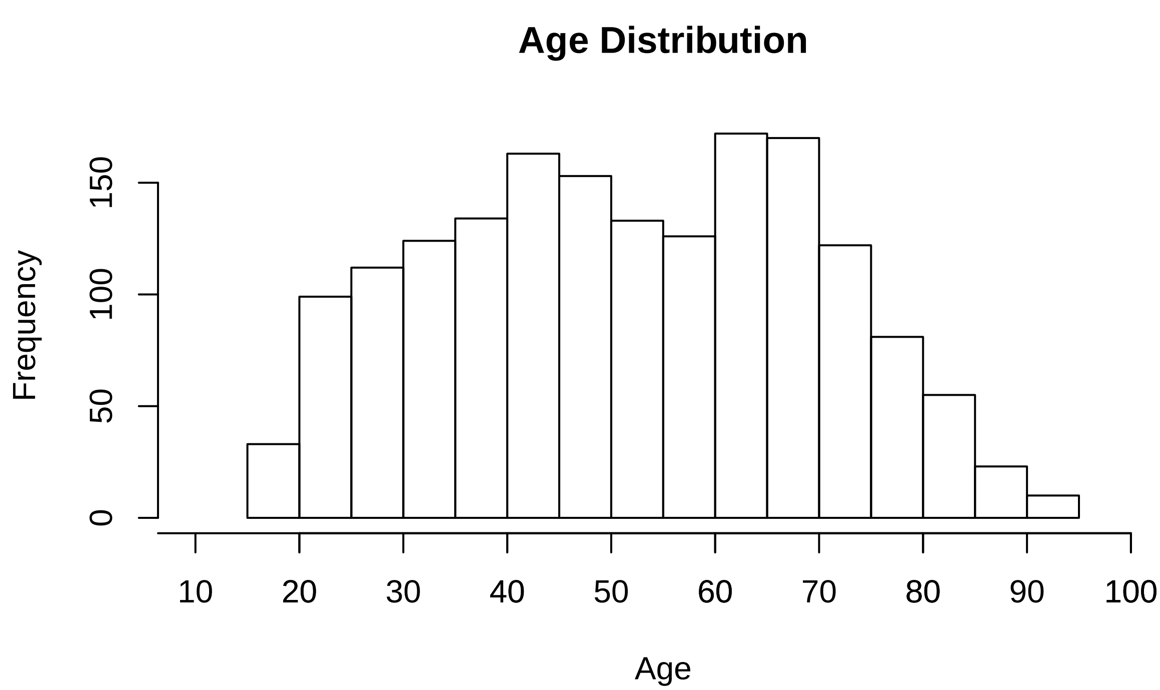

Figure 1: Distribution of respondent's age

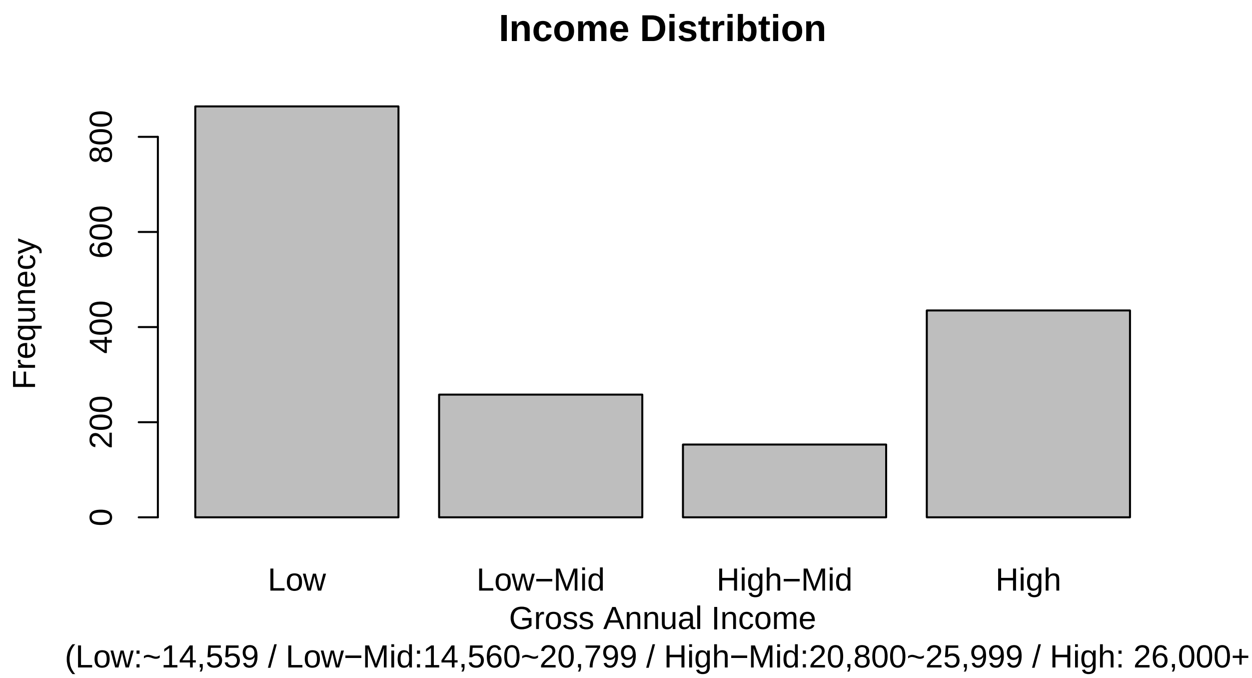

Figure 2: Distribution of respondent's income

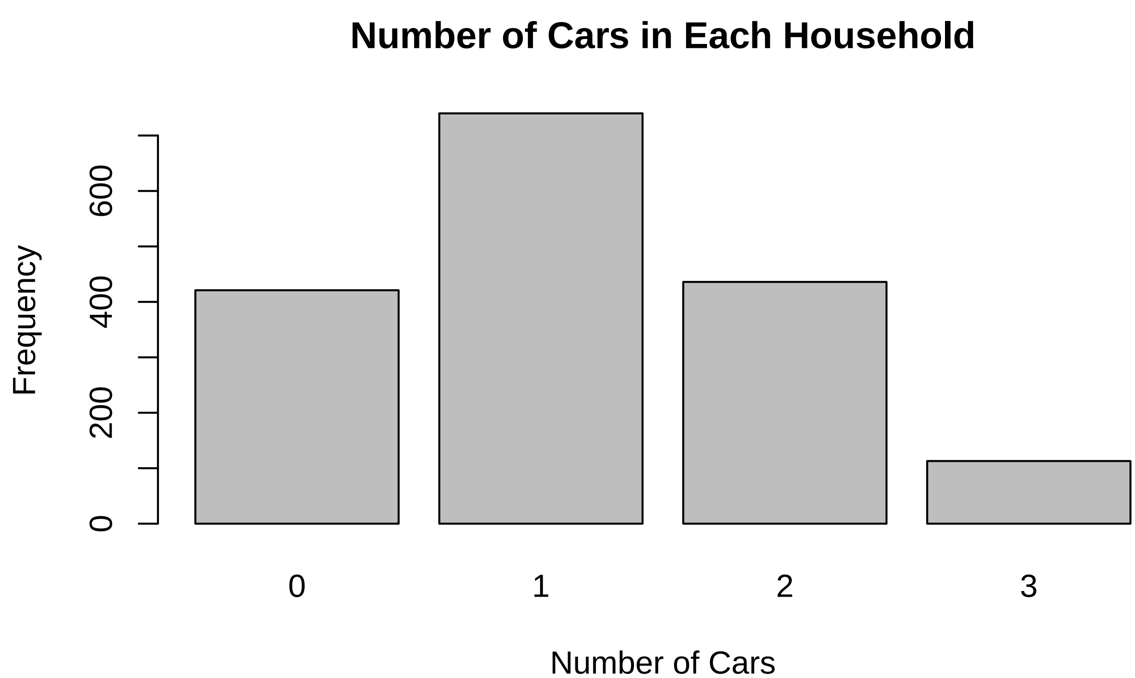

Figure 3: Distribution of number of cars

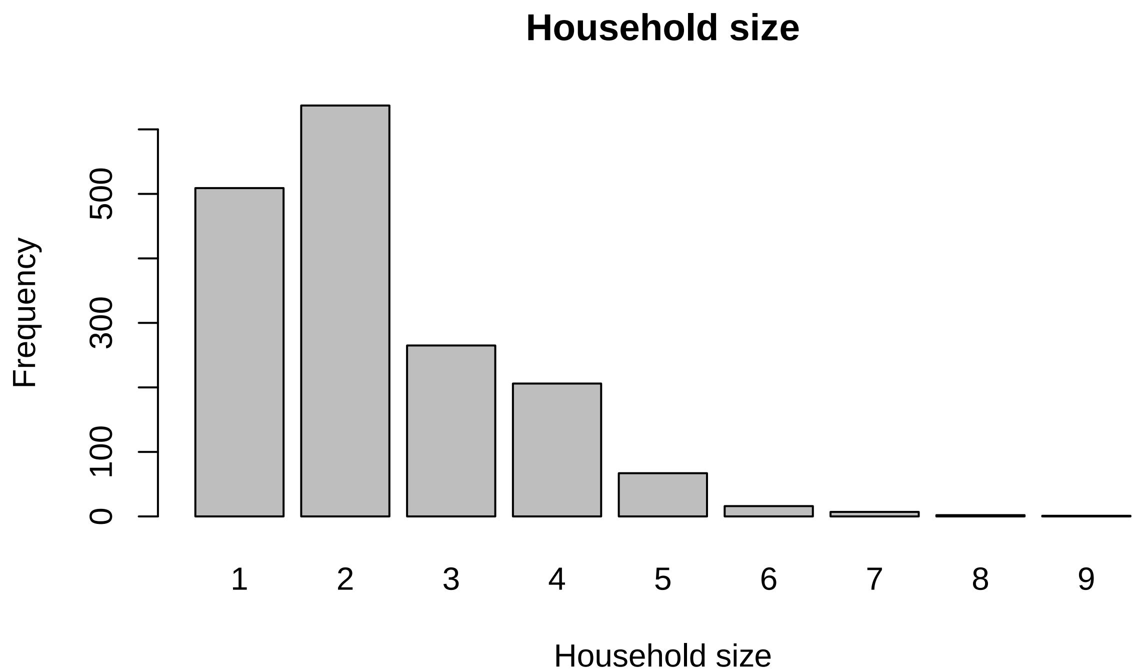

Figure 4: Distribution of household size

3 Logistic Regression

3.1 Modelling

This section presents our efforts to answer the question, "Who are more/less likely to be interested in EVs". Since the dependent variable is a binary variable, i.e. respondents having an interest in EVs or not, we adopt logistic regression models here.

Nine independent variables are categorized into four groups. Socio-economic variables include age, sex, gross annual income level, and education level. The second group includes having driver's license or not, and the number of cars available to the household, which to some extent describe the respondents' potential to buy a new car. The third one is household-level variables, including household size and whether or not having depedent children. The last one is region. (The UK has 11 regions in total. However, our preliminary analysis suggests that, using London as the baseline, only Scotland is significantly different. Therefore, we only include "Scotland" in the regression.)

We only include socio-economic variables in our base model (regression 1) and add in a new group of independent variables each time. As the results in table XX show, all models are significant at 99% level. Except for the two variables, size of household and having dependent children or not, all other variables are statistically signifcant as well. The fact that household-related variables are effective predictors at macro level, but not at micro level, may suggest a neighborhood effect, but it is beyond the scope of this paper.

Table 3: Step-wise Modelling

	Dependent variable:			
	EVinterest			
	(1)	(2)	(3)	(4)
Age	-0.01^{*}	-0.01^{*}	-0.01^{*}	-0.01^{*}
	(0.004)	(0.004)	(0.004)	(0.004)
Male	0.49^{***}	0.46^{***}	0.46^{***}	0.47^{***}
	(0.14)	(0.14)	(0.14)	(0.14)
Income: low-middle	0.07	-0.09	-0.09	-0.10
	(0.20)	(0.21)	(0.21)	(0.21)
Income: high-middle	0.27	0.05	0.05	0.03
	(0.23)	(0.24)	(0.24)	(0.24)
Income: high	0.57^{***}	0.31^{*}	0.30^{*}	0.29^{*}
	(0.17)	(0.17)	(0.17)	(0.17)
College degree	0.69^{***}	0.65^{***}	0.65^{***}	0.65^{***}
	(0.15)	(0.15)	(0.15)	(0.15)
Drivers licence		0.92^{***}	0.92^{***}	0.90^{***}
		(0.24)	(0.24)	(0.24)
Number of cars		0.18^{**}	0.19^{**}	0.19^{**}
		(0.08)	(0.09)	(0.09)
Size of household			-0.02	-0.03
			(0.08)	(0.08)
Having dependent children			0.04	0.02
			(0.21)	(0.21)
Scotland				-0.70^{**}
				(0.29)
(Intercept)	-1.79^{***}	-2.63^{***}	-2.59^{***}	-2.49^{***}
	(0.23)	(0.29)	(0.36)	(0.36)
Observations	1,710	1,710	1,710	1,710
R^2	0.08	0.11	0.11	0.12
χ^2	90.15^{***} (df = 6)	124.56^{***} (df = 8)	124.64^{***} (df = 10)	131.59^{***} (df = 11)
Note:				*p<0.1; **p<0.05; ***p<0.01

Based on this analysis, we leave out the two household-related variables and use the following formula.

$$\ln\left(\frac{p}{1-p}\right) = \beta_0+\beta_1 Age+\beta_2 Male+\beta_3 Lowmid+\beta_4 Highmid+\beta_5 High+\beta_6 College+\beta_7 license+\beta_8 NumCar+\beta_9 Scotland$$

Where p is the probability of being interested in EVs. The estimated model is shown in the table below.

Table 4: Interests in EVs

	Dependent variable:
	EVinterest
Age	-0.01^{*}
	(0.004)
Male	0.47^{***}
	(0.14)
Income: low-middle	-0.10
	(0.21)
Income: high-middle	0.03
	(0.24)
Income: high	0.29^{*}
	(0.17)
College degree	0.65^{***}
	(0.15)
Drivers licence	0.91^{***}
	(0.24)
Number of cars	0.18^{**}
	(0.09)
Scotland	-0.70^{**}
	(0.29)
(Intercept)	-2.56^{***}
	(0.29)
Observations	1,710
R^2	0.12
χ^2	131.47^{***} (df = 9)
Note:	*p<0.1; **p<0.05; ***p<0.01

The model is statistically significant at 99% level and five out of the seven variables are significant at 95% level. On average, males are more likely to be interested in EVs. Having a college degree or driver's license also increases the likelihood of being interested. The number of cars is positively relative to the likelihood, suggesting that people who already have the "basics" are more likely to consider EVs. However, being in Scotland reduces this likelihood.

The less significant variables are age and income level. Age is negatively relative to the likelihood of being interested, meaning that younger people are more likely to have an interest in EVs. Interestingly, income level seems not to matter much. Using the lowest quartile group as the baseline, only the highest quartile group is significantly different at 90% level. Nevertheless, results from regression 1 shows that income level (at least when comparing the highest quartile to the lowest quartile) is statistically significant at 99% level. The decrease in significance is due to the correlation between income level, and having driver's license and number of cars available.

3.2 Comparison of Probabilities by Group

In addition to the odds ratio provided in the regression output, we want to be more specific with the size of effects of the seven independent variables. Therefore, the ranges of probability based on the variation of different variables are provided as following.

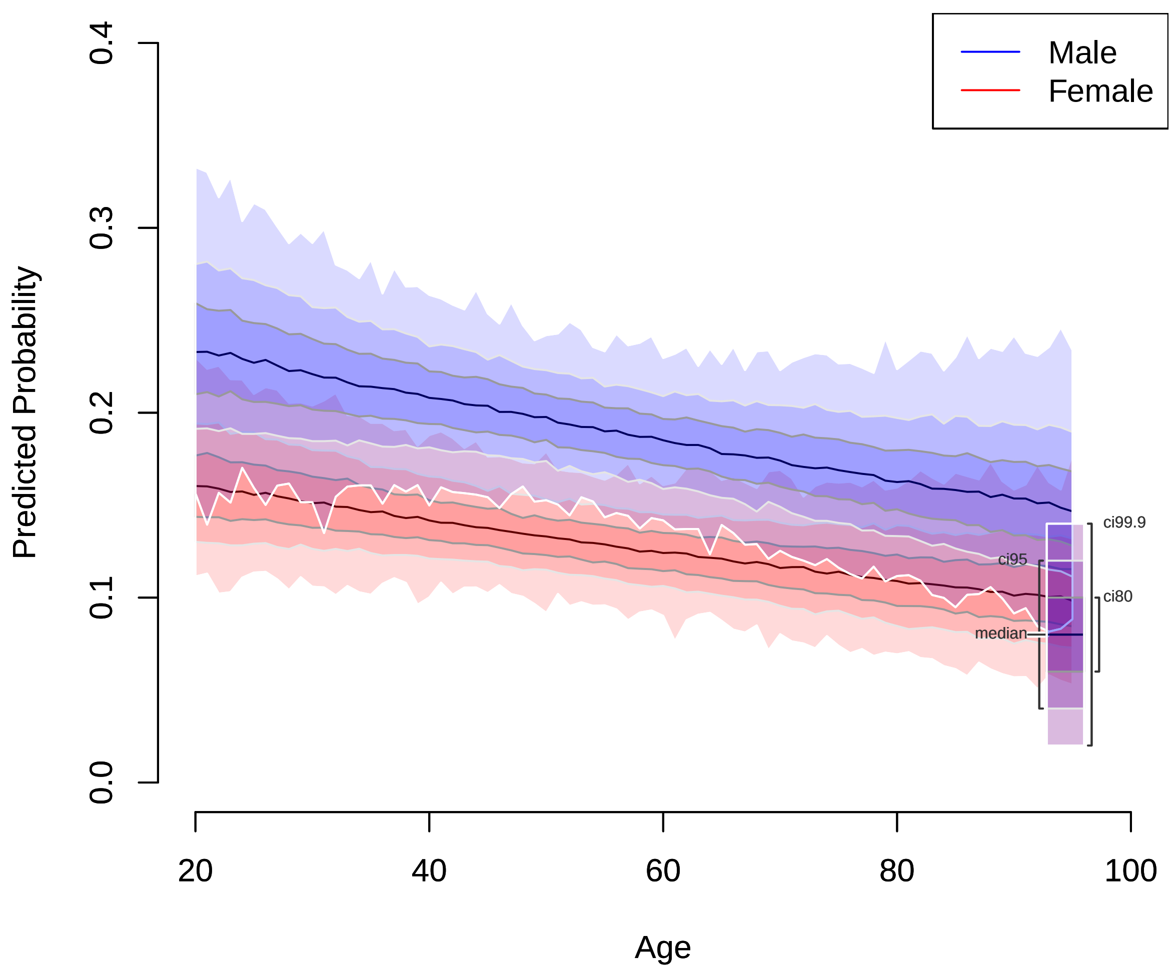

Figure 5: Predicted Probability of Having an Interest in EVs by Age and Sex

Figure 1 shows the predicted probability of having an interest in EVs by age and sex. Regardless of age, males show a greater interest than females. Even though the variation among males is larger too, the difference in probability is on average about 5 percentage points. The linear relationship between age and this probability is clearly demostrated here as well. Younger people are more likely to be interested than older people.

The following tables summarize probabilities of having an interest in EVs of different groups and allow us to know the exact differences between groups.

Table 5: Probability by Income Level

Income Level	Mean Probability
Highest quartile	30.1 %
Second highest quartile	15.9 %
Second lowest quartile	15.9 %
Lowest quartile	13.1 %

Table 6: Probability by Education Level

Education Level	Mean Probability
At least have a college degree	30.7 %
Do not have a college degree	14.5 %

Table 7: Probabilities by Driver's License Status

License Status	Mean Probability
Have driver's license	22.1 %
Do not have driver's license	6.5 %

Table 8: Probabilities by Number of Cars

Number of Cars	Mean Probability
0	7.8 %
1	19.3 %
2	27.3 %
3	18.6 %

Table 9: Probabilities by Region

Regions	Mean Probability
In Scotland	10.1 %
In other regions	19.3 %

4 Discussion and Limitations

In general, interest in EVs is rather limited. Among all respondents, only 18.48% are interested. This, on the other hand, suggests the need to better understand the factors related to the interest in EVs, and this is what motivated our research too.

Above analysis shows that the variations among different groups are quite large. For instance, people having a college degree, on average, have a probability of 30.7% to be interested in EVs, while people without a degree only have a probability of 14.5%. This kind of variation (especially the one related to education level) is actually a good sign, suggesting that interest in EVs could be potentially increased. Promotion of EVs should give more attention to the needs and comsumption patterns of low-interest groups.

However, this also leads to one of the limitations of this paper, not being able answer why different groups show different levels of interest. This question is very hard, not just because we do not have more detailed surveys, but also because different groups may be systematically different in many unobeserved ways. Even though, we will still attempt to make contributions on this question in our final paper.

Another limitation is due to the design of the survey, which did not exclude respondents who are not interested in buying cars in genereal. Therefore, varialbes, including having driver's license and number of cars available, need to be interpreted with great caution. The low interest shown in groups without driver's license or cars might not be caused by their lack of interest in EVs, but by their lack of interest in cars.

Lastly, R^2 of the model is only 0.12, meaning that these variables only explain 12% of the variations in the probability of having an interest in EVs. Partly, this is because of a lack of available information, as it is impossible to survey everything. Partly, attitudes are just not easy to explain, because they could be influenced by countless factors, sometimes as trivial as an image or some careless words. Nevertheless, we believe that there are patterns and these patterns are what we should better understand if EVs are to be further promoted.

5 Reference

R Core Team (2015a) Alathea (2015) Wickham and Francois (2015) R Core Team (2015b) Xie (2015) Gandrud (2015) Jr (2015) Hlavac (2015) Owen et al. (2013)

Alathea, Letaw. 2015. *Captioner: Numbers Figures and Creates Simple Captions.* http://CRAN.R-project.org/package=captioner.

Gandrud, Christopher. 2015. *Repmis: Miscellaneous Tools for Reproducible Research.* http://CRAN.R-project.org/package=repmis.

Hlavac, Marek. 2015. *Stargazer: Well-Formatted Regression and Summary Statistics Tables.* http://CRAN.R-project.org/package=stargazer.

Jr, Frank E Harrell. 2015. *Rms: Regression Modeling Strategies.* http://CRAN.R-project.org/package=rms.

Office for National Statistics. HM Revenue and Customs. 2015. "Percentile Points from 1 to 99 for Total Income Before and After Tax." https://www.gov.uk/government/statistics/percentile-points-from-1-to-99-for-total-income-before-and-after-tax.

Office for National Statistics. Social Survey Division. 2015. "Opinions and Lifestyle Survey, Electric Vehicles Module, February 2014 and February 2015." https://discover.ukdataservice.ac.uk/Catalogue/?sn=7571&type=Data%20catalogue.

Owen, Matt, Kosuke Imai, Gary King, and Olivia Lau. 2013. *Zelig: Everyone's Statistical Software.* http://CRAN.R-project.org/package=Zelig.

R Core Team. 2015a. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

———. 2015b. *Foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, DBase, .* http://CRAN.R-project.org/package=foreign.

Wickham, Hadley, and Romain Francois. 2015. *Dplyr: A Grammar of Data Manipulation.* http://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2015. *Knitr: A General-Purpose Package for Dynamic Report Generation in R.* http://CRAN.R-project.org/package=knitr.