Internet Privacy and the European Union

Final Report by Ni'Gere Epps

I. Introduction

I will be investigating how Europeans' feelings about the European Union, under which they are regulated, affect their concerns about online privacy and personal information. I hypothesize that those who have higher levels of satisfaction with their nation's union will be less concerned with privacy risks involved with sharing personal information online as well as the amount of control they have over this information. The dataset I am using was collected by the European Commission in 2015 through survey questionnaires, by the means of in-person interviews, proportionally sampled to represent the population of the 28 countries in the European Union at the time. Through using survey responses by the means of an ordinal feeling thermometer / scale toward the European Union and an ordinal concern thermometer toward sharing personal information online, I can directly measure the theory in an empirical manner. In my regression, I control for each participant's trust in their government and country, their internet usage, age, gender, education, marital status, political affiliation, and satisfaction with life. I find that having better feelings and satisfaction toward the current state of the European Union decreases concern for providing and having control over online personal information in a statistically significant way. However, some problems of causal inference exist which I will address.

II. Research Problem

In the growing age of technology, there have been many concerns with the use of personal information online. This has also lead to organizations researching ways in which to reduce concern for online privacy in order for consumers to be more willing to use the internet and provide personal information. This is because perceived risk is an important aspect of the consumer decision-making process nd thus understanding the barriers to internet use can help e-marketers understand the consumers. In the years and months leading up to this survey, regulators and courts in the EU were increasingly vigilant in relation to privacy practices and policies of large online companies, not only through privacy-specific regulations and enforcement, but also through the application of consumer legislation. As a result, the European Union proposed new data protection laws and regulations in 2015 that would apply to all 28 member countries of the EU. These include the following [1]:

- Allowing national watchdogs to issue fines, potentially totaling the equivalent of hundreds of millions of dollars, if companies misuse people's online data, including obtaining information without people's consent.
- Enshrining the right to be forgotten into European law, giving people in the region the right to ask that companies remove data about them that is either no longer relevant or out of date.
- Requiring companies to inform national regulators within three days of any reported data breach, a proposal that goes significantly further than what is demanded by American authorities.

- Obliging anyone under 16 to obtain parental consent before using popular services like Facebook, Snapchat and Instagram, unless any national government lowers the age limit to 13.
- Extending the new rules to any company that has customers in the region, even if the company is based outside the European Union.

This report seeks to address the question of which aspects of one's life affects their concern for internet privacy. Liebermann et al. tested internet interactions in regards to one's concern with sharing personal information online and found that there was a significant difference between people who used the internet to buy items and people who didn't, as well as heavy users of the internet vs light users [2]. However in this study the sample size is very small and biased (465 employed adults). In using the data from the European Commission, I am able to conduct a more ecologically valid study because the data is much larger and diverse (wide range of countries, demographics, internet use, etc).

A 2006 study conducted by Bowie and Jamal titled "Privacy Rights on the Internet: Self-Regulation or Government Regulation?", compared the US's current system with the Europeans Union's system of regulating government and concluded that there was no merit in the US adopting a similar system of regulation [3]. The study suggests that assurance seals within individual sites is an adequate measure to secure internet privacy, which is likely representative of US attitudes towards government control of privacy, and thus it is helpful to apply the same test to members of the European Union and their attitudes toward the EU.

Bergström notes in her study that "questions regarding personal privacy are becoming increasingly relevant, and the discussion continuously arises regarding what digital tracks we

leave. Intrusive use and manipulation of personal information not only affect people's behaviour, but also they could have important implications for political and civic society. Previous research has relied on convenience samples and has often focused on one or only a few areas of use. The presented study, based on a probability sample, gives an overall picture of how privacy concerns are perceived in different online contexts and how socio-demography, internet experience, trust, and political orientation contribute to the understanding of privacy concerns in different settings. The results clearly point to privacy concerns as being very diverse and dependent on the application in question. All dimensions that are used to explain privacy concerns are partly supported in the study. But their explanatory powers differ and not all areas of concern are affected by the same explanatory factors. Trust in other people is the single most important factor explaining privacy concerns when using digital media and applications. The more people trust others, the less concern they have for misuse of personal information" [4].

Bergstrom found that significant co-variations were found for age, level of education, political orientation and general internet habits, but not for all measured areas. However, the correlations were weak and could be a result of many of the previous studies being older, especially since privacy concerns are a rather contextualised issue. Her study was conducted in 2013 using studies from before that time period and thus my study can be used to build upon Bergstrom's. Though the data was collected in 2015, it is newer, updated, and more representative of the current population than the studies used by Bergstrom (some conducted) as early as 1974). Thus, this study provides a more contemporary and ecological analysis of perceived risks of providing personal information online and barriers to internet use, especially in regards to "trust in other people", specifically the European Union.

III. Theory

I hypothesize that those who have higher levels of satisfaction for their nation's union will be less concerned with sharing personal information online. I believe that if one is satisfied with their nation's union, then they will likely be less suspicious of their government and its use of online information and will have more trust in sharing information online In general, there is a link between trust and sharing information, and this can be applied to trust / satisfaction for one's government and sharing private or personal information online. In the case of the European Union, the mechanism here is the EU's commitment to online privacy. If residents have better feelings / trust toward the EU, they will have more confidence in its implementation of privacy regulations, thus decreasing their concern or providing and controlling personal information online. There may be a variety of other mechanisms such as "business policies and knowledge about privacy settings" [4] or other variables such as previous experiences with the protection or misuse of one's online personal information, or other mechanisms / variables not measured in the dataset.

IV. Dataset and Variables

The data is downloaded from ICPSR (Inter-university Consortium for Political and Social Research). It was collected by the European Commission through survey questionnaires, by the means of in-person interviews. The samples are random but proportional to represent the population of the 28 countries in the European Union in 2015. Here the methodology and purpose of the survey is described in greater detail:

"The Eurobarometer series is a unique cross-national and cross-temporal survey program conducted on behalf of the European Commission. These surveys regularly monitor public opinion in the European Union (EU) member countries and consist of standard modules and special topic modules. The standard modules address attitudes towards European unification, institutions and policies, measurements for general socio-political orientations, as well as respondent and household demographics. The special topic modules address such topics as agriculture, education, natural environment and resources, public health, public safety and crime, and science and technology."

"This round of Eurobarometer surveys covers the following special topics: (1) Europeans in 2015 and (2) Data Protection and the Internet. Regarding these two topics, respondents were asked about their Internet activity, personal data disclosure, online data disclosure reasons, government data collection revelations, online data disclosure risks, social web privacy, and data protection complaints. In addition, respondents were asked their opinions on the economic situation in their countries, how much they trusted certain institutions, and how often they discuss political matters with friends or relatives."

"The basic sample design applied in all states is a multi-stage, random (probability) one. In each country, a number of sampling points was drawn with the probability proportional to population size (for a total coverage of the country) and to population density. In order to do so, the sampling points were drawn systematically from each of the 'administrative regional units', after stratification by individual unit and type of area. They thus represent the whole territory of the countries surveyed according to the EUROSTAT NUTS II (or equivalent) and according to the distribution of the resident population of the respective nationalities in terms of metropolitan,

urban and rural areas. In each of the selected sampling points, a starting address was drawn, at random (following the ‘closest birthday rule’). All interviews were conducted face-to-face in people's homes and in the appropriate national language. CAPI (Computer Assisted Personal Interview) was used in those countries where this technique was available” [5].

The dataset consists of 27,980 respondents / observations (individual unit of analysis) before the regression. Missing values occurred because “Prefer not to answer” and “I don’t know” responses were coded as such however I did not observe a specific pattern or sample bias with these missing values. This is likely because “Don’t know” responses are based on personal knowledge rather than a specific trend / trait of the respondent.

Summary of Variables:

Name of Variable	Type of Variable	Description of Variable	Central Tendency Measurement	Dispersion Measurement	Number & % of Missing Values
Concern for Providing Online Personal Information	Ordinal	4 is most concern, 1 is least concern	Median: 3	IQR = 2	2512 8.98%
Concern for Control Over Online Personal Information	Ordinal	4 is most concern, 1 is least concern	Median: 3	IQR = 1	13574 48.51%
Current Feelings Toward EU	Ordinal	5 is very positive, 1 is very negative	Median: 3	IQR = 1	371 1.33%
Country Member of Eurozone	Nominal / Binary	1 is member, 0 is non-member	Mode: 1	IQR = 1	0

Country Joined EU Post 2004	Nominal / Binary	1 is joined 2004 and later, 0 is before 2004	Mode: 0	IQR = 1	0
Trust in the National Government	Nominal / Binary	1 is tend to trust, 0 is tend not to trust	Mode: 1	IQR = 1	1538 5.50%
Internet Use Frequency	Ordinal	6 is everyday, 0 is never	Median: 6	IQR = 2	3171 11.33%
Uses Internet to Purchase	Nominal / Binary	1 is uses internet to purchase, 0 is doesn't	Median: 1	IQR = 1	7738 27.66%
Age	Ratio	Self-reported Age	Mean: 51 Median: 52	Range: 15-99 SD: 18.2 Var: 331.6	0
Male / Gender	Nominal / Binary	1 is male, 0 is female	Median: 0	IQR = 1	0
Married	Nominal / Binary	1 is married, 0 is non-married	Median: 1	IQR = 1	255 0.91%
Education Past High School	Nominal / Binary	1 is education past high school, 0 is high school education or below	Median: 1	IQR = 1	0
Political Affiliation	Ordinal	10 is right, 1 is left	Median: 5	IQR = 3	5356 19.14%
Satisfaction with Life	Ordinal	4 is very satisfied, 1 is not at all satisfied	Median: 3	IQR = 2	74 0.26%

The distribution of each variable is located in the Appendices section.

Internal Validity:

My study seeks to measure the concepts of one's trust in the European Union with their trust in online privacy. Concepts: I am using satisfaction with the EU as an indicator for trust because I believe these two are correlated (a high satisfaction often leads to a higher trust). The dataset I'm using has a "Do you tend to trust the EU" variable however it is only a binary variable. King [6] warns against using dichotomous variables as a main independent variable therefore I didn't use this variable. In order to test the association of my current feelings thermometer and trust in the EU, I ran a correlation test on the two variables and found that they were highly correlated with a score of 0.58. Thus, the internal validity of my main independent variable is not influenced too much, though it is not a direct measurement of trust. In terms of my dependent variables, I am using concern for sharing and controlling personal information online as indicators for (lack of) trust in one's internet privacy and information. This is a direct measure of my concept thus I have high internal validity.

V. Exploratory Analysis

Expected Signs

Name of Variable	Sign	Reasoning
Current Feelings Toward EU	-	Better feelings toward EU will increase trust in its regulation of personal online information
Country Member of Eurozone	-	Greater connection to EU will decrease concern due to greater trust of the EU
Country Joined EU Post 2004	+	Less established connection to EU will increase concern due to less trust of the EU
Trust in the National Government	-	More trust in national government will decrease concern due to decreased suspicion of government use of information
Internet Use Frequency	-	Using the internet more frequently will decrease concern because of more exposure to the internet
Uses Internet to Purchase	-	More exposure to the internet and comfort with supplying personal (financial) information will decrease concern
Age	+	Increased age will increase concern due to less knowledge / exposure to the internet
Male / Gender	?	It is hard to predict because various gender roles and concepts associated with gender may affect one's concern for providing personal information
Married	+	Married people will be more concerned because providing personal information could affect both them and their spouse
Education Past High School	-	Having more education past high school will cause less concern due to more knowledge about the internet
Political Affiliation	+	More conservative people will have more concern and less willingness to provide personal information due to values
Satisfaction with Life	-	Higher satisfaction with life will lead to less concern due to more fulfillment & trust with country lifestyle

Correlation matrix:

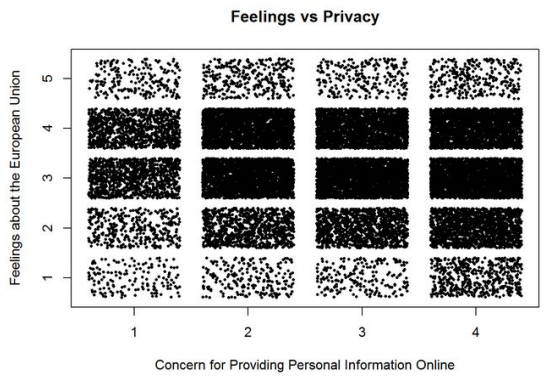
In exploring the relationship between each variable in the regression, I conducted a Pearson correlation matrix as shown below:

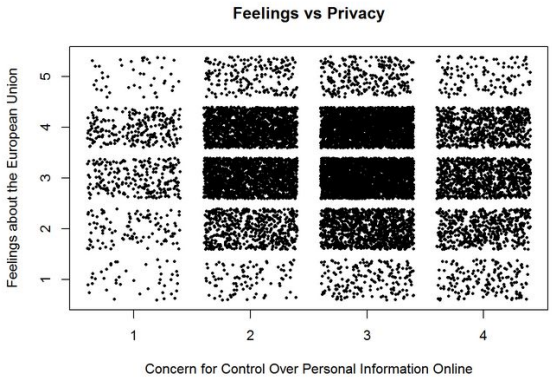
As seen in the correlation matrix, none of the variables have a high correlation with each other. The highest correlation is between the two dependent variables (0.31) which are expected to have a high correlation because they are measuring the same concept. Age and marriage are also equally correlated which is expected (older people tend to be married). Because of these overall low correlation values, multicollinearity is not one of the issues with causal inference.

Analysis of main independent and dependent variables:

Data: Concern for Providing Personal Information Online and Feelings About EU
t = -14.34, df = 25216, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval: (-0.10216861 , -0.07768363)
Correlation coefficient: -0.08993971

Data: Concern for Control Over Personal Information Online and Feelings About EU
t = -8.307, df = 14337, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval: (-0.08548158 , -0.05290237)
Correlation coefficient: -0.06921043





Based on the correlation coefficients and p-values, we would expect to find a negative relationship between the main independent variable and dependent variables. This means that as one's feelings about the European Union increases, their concern for their personal information online will decrease in a statistically significant way.

VI. Regression Analysis

Here is my final regression, the bivariate regression on the right and the controls included in the left model:

Feelings and Trust Toward European Union on Concern for Online Privacy

	Concern for Online Personal Information (OPI)			
	Providing OPI	Control Over OPI	Providing OPI	Control Over OPI
Current Feelings Toward EU	-0.060*** (0.009)	-0.028*** (0.009)	-0.101*** (0.007)	-0.061*** (0.007)
Country Member of Eurozone	0.056*** (0.016)	0.017 (0.016)		
Country Joined EU Post 2004	-0.017 (0.017)	-0.060*** (0.017)		
Trust in National Government	-0.162*** (0.016)	-0.109*** (0.015)		
Internet Use Frequency	0.005 (0.005)	-0.010 (0.008)		
Uses Internet to Purchase	-0.097*** (0.018)	-0.138*** (0.017)		
Age	0.001*** (0.0005)	0.001 (0.0005)		
Male	-0.079*** (0.015)	-0.049*** (0.014)		
Married	0.039** (0.015)	0.071*** (0.015)		
Education Past High School	-0.002 (0.016)	0.008 (0.016)		
Political Affiliation (Toward Right)	-0.004 (0.003)	-0.009*** (0.003)		
Satisfaction with Life	-0.046*** (0.011)	-0.042*** (0.011)		
Constant	3.182*** (0.055)	3.207*** (0.069)	3.146*** (0.023)	2.971*** (0.025)
Observations	18,891	11,559	25,218	14,339
R^2	0.022	0.023	0.008	0.005
Adjusted R^2	0.021	0.022	0.008	0.005
Residual Std. Error	0.998 (df = 18878)	0.766 (df = 11546)	1.010 (df = 25216)	0.776 (df = 14337)
F Statistic	34.991*** (df = 12; 18878)	23.143*** (df = 12; 11546)	205.639*** (df = 1; 25215)	69.006*** (df = 1; 14337)

Note: *p<0.1; **p<0.05; ***p<0.01

Based on the results, I find that positive feelings toward the European Union result in a significant decrease for one's concern for online privacy and personal information, after controlling for other factors. This is consistent with my theory. Some inconsistencies with my expectations are that members of the Eurozone had a significant increase in concern rather than decrease, and countries that have joined the European Union in 2004 and later had a significant decrease in concern. Individuals who leaned more towards the right on the political spectrum had significantly decreased concern for controlling personal information, and the results showed that

males had less concern than women. For the most part, my expectations were supported by the data, though there were some inconsistencies and resolved trends in my controls.

VII. Problems of Causal Inference

I encounter a few problems of causal inference. The first being the internal validity of my independent variable. Though highly correlated, feelings toward the Europen Union aren't an exact measurement of one's trust in it. Thus the regression coefficient isn't a precise measurement of my concepts however it is still statistically significant.

Another problem is the type of data used in the dataset. Besides age, all of my variables are nominal or binary which may result in a lack of variance. Along with this, a majority of the variables are Likert scales which are self-reported and can be subjective based on someone's mood while taking the survey, trust in the interviewer, or other confounding factors. Thus I may have spuriousness bias such as someone being a trustworthy person, which may lead to tell having better feelings toward online privacy and the EU without a specific mechanism such as proposed data policies and regulations.

Lastly, I may have a problem with missing values. Though there was no clear pattern with missing values / "Don't know" responses, I did lose over half of the responses after including my controls in the regression.

Some problems of causal inference that I expect to have is that there may be a variety of variables that influence both independent and dependent variable that aren't measured in this dataset and cannot be controlled. For example, distrust in public institutions may cause worse feelings toward the European Union as well as a higher concern for sharing personal information

online. Because there may exist a range of factors which affect both the independent variable (feelings toward the European Union) and dependent variable (concern for online privacy), it is not possible to conclude definitively that there is causality between the two. However, by controlling for variables provided in the dataset, I can limit these issues of causal inference.

VIII. Conclusion

In conclusion, I found that current feelings toward the European Union result in a significant decrease for one's concern for their online privacy and personal information, after controlling for other factors. This is consistent with my theory. Some inconsistencies with my expectations are that members of the Eurozone had a significant increase in concern rather than decrease, and countries that have joined the European Union in 2004 and later had a significant decrease in concern. Individuals who were more right on the political spectrum had significantly decreased concern for controlling personal information, and the results showed that males had the least concern as well.

This study adds onto the ongoing research of perceived privacy risks and barriers to internet use and providing personal information online. It provides a more contemporary and ecological analysis of perceived risks because the data is larger and newer than that used in previous studies. In terms of future research, it would be helpful to conduct a study of this magnitude in other countries or groups of countries in order to test how the results compare to those of the members of the European Union. It would also be beneficial to re-conduct the surveys after the passing of new proposed laws and regulations for the European Union such as those discussed in the "Research Problem" section (they were not enacted until 2017). Lastly, the

internal validity of the study and it's findings would be improved by addressing some of the issues of causal inference such as the heavy use of Likert scales.

IX. References

[1] Scott, Mark. “Europe Approves Tough New Data Protection Rules.” *The New York Times*, The New York Times, 15 Dec. 2015, www.nytimes.com/2015/12/16/technology/eu-data-privacy.html.

[2] Liebermann, Yehoshua, and Shmuel Stashevsky. "Perceived risks as barriers to Internet and e-commerce usage." *Qualitative Market Research: An International Journal* 5, no. 4 (2002): 291-300.

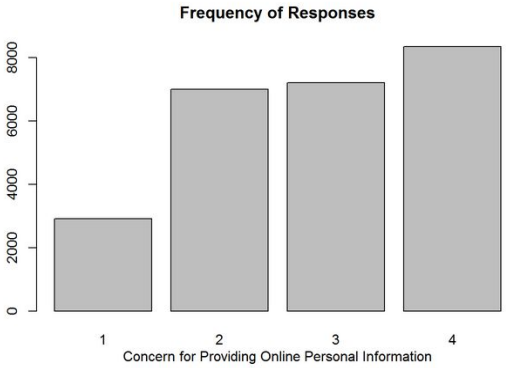
[3] Bowie, Norman E., and Karim Jamal. “Privacy Rights on the Internet: Self-Regulation or Government Regulation?” Business Ethics Quarterly 16, no. 3 (2006): 323–42. doi:10.5840/beq200616340.

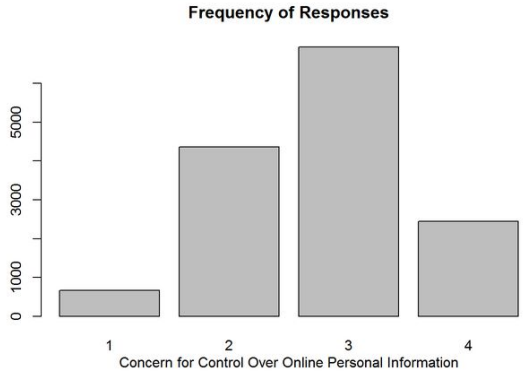
[4] Bergström, Annika. "Online privacy concerns: A broad approach to understanding the concerns of different groups for different uses." *Computers in Human Behavior* 53 (2015): 419-426.

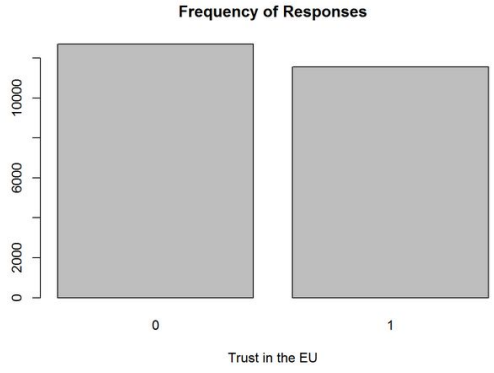
[5] European Commission. Eurobarometer 83.1: Europeans in 2015, Data Protection and the Internet, February-March 2015. Ann Arbor, MI: GESIS [distributor], Inter-university Consortium for Political and Social Research [distributor], 2017-04-05. https://doi.org/10.3886/ICPSR36665.v1

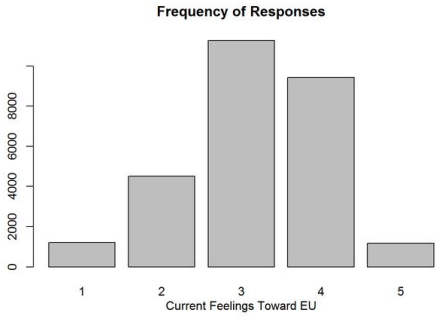
[6] King, Gary, and Eleanor Neff Powell. "How not to lie without statistics." In *delivery at the 2008 Annual Meeting of the American Political Science Association*. 2008.

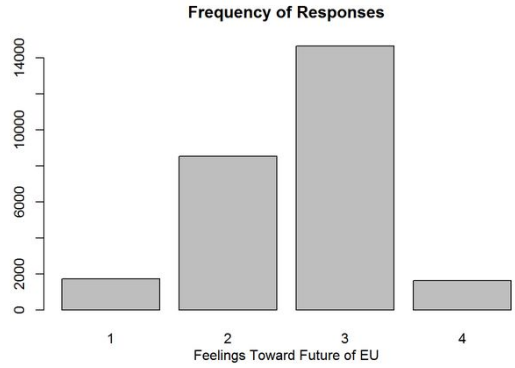
X. Appendices

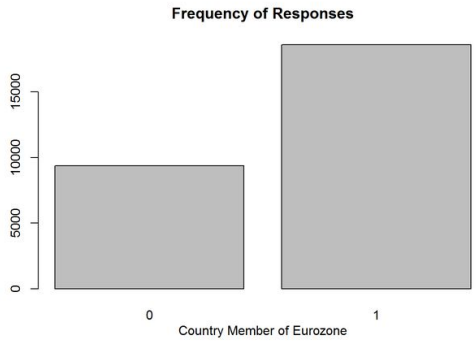


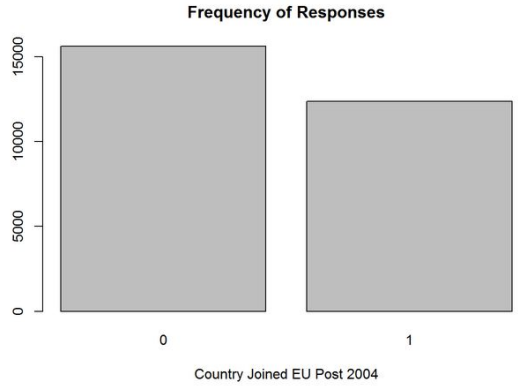


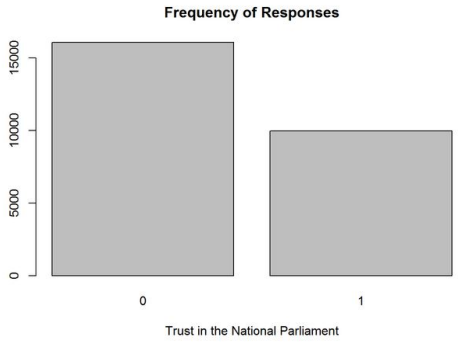


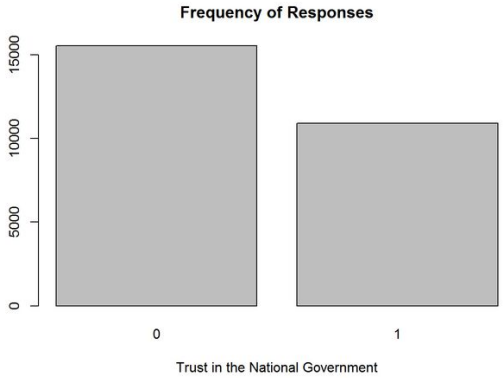


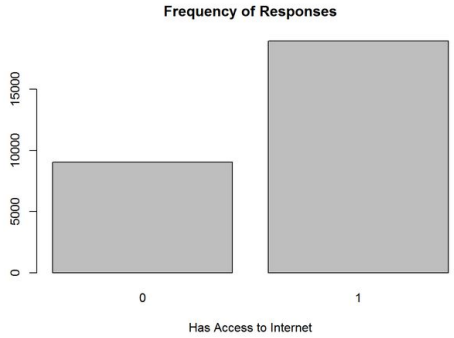


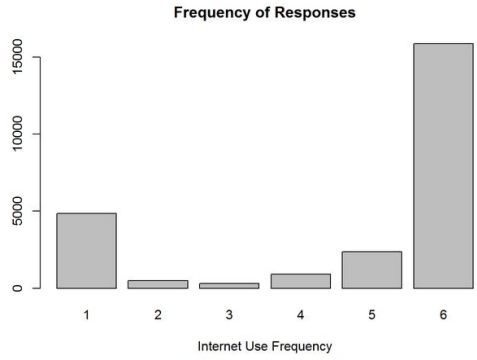


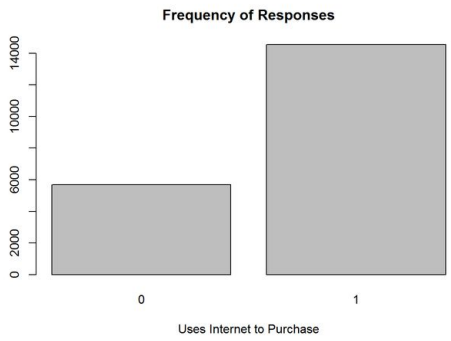

Frequency of Responses
0
5000
10000
15000
0
1
Country Joined EU Post 2004

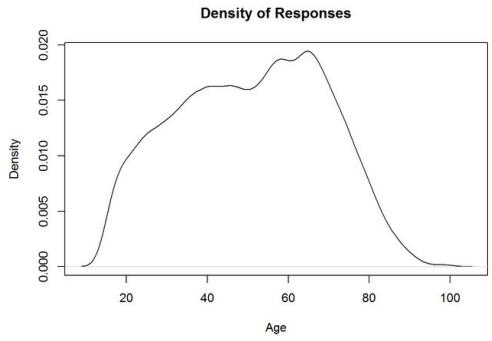

Frequency of Responses
0
5000
10000
15000
0
1
Trust in the National Parliament

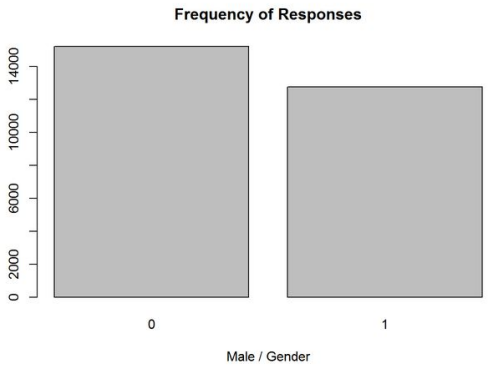

Frequency of Responses
0
5000
10000
15000
0
1
Trust in the National Government

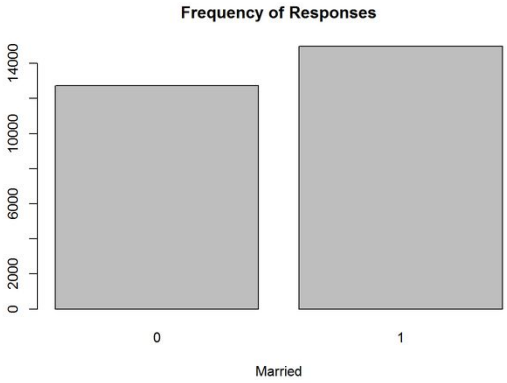

Frequency of Responses
0
5000
10000
15000
0
1
Has Access to Internet

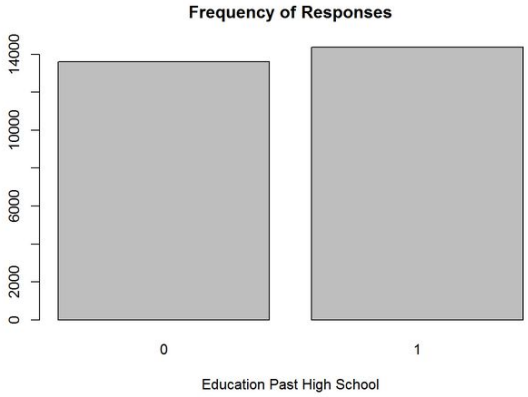

Frequency of Responses
0
5000
10000
15000
1
2
3
4
5
6
Internet Use Frequency

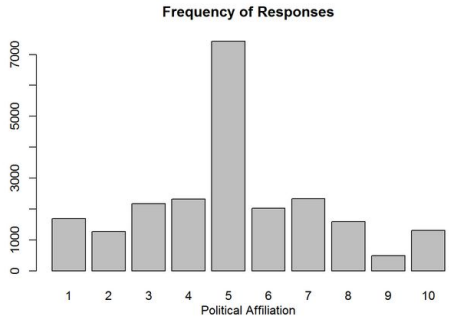

Frequency of Responses
0
2000
6000
10000
14000
0
1
Uses Internet to Purchase

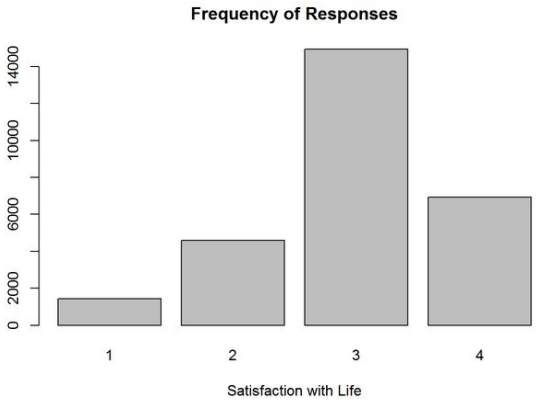

Density of Responses
Density
0.000
0.005
0.010
0.015
0.020
20
40
60
80
100
Age


Frequency of Responses
0
2000
6000
10000
14000
0
1
Male / Gender


Frequency of Responses
0
2000
6000
10000
14000
0
1
Married


Frequency of Responses
0
2000
6000
10000
14000
0
1
Education Past High School


Frequency of Responses
0
1000
3000
5000
7000
1
2
3
4
5
6
7
8
9
10
Political Affiliation


Frequency of Responses
0
2000
6000
10000
14000
1
2
3
4
Satisfaction with Life