TO VOTE OR NOT TO VOTE?
Exploring the Impact of Demographics on Voter Behavior

Jeremie Kurgan   3145021   20 December 2022
Nicolas Lockhart   3126373
Filippo Lisanti   3089785
Daniil Tiugaev   3102409

Abstract

This report explores the relationship between a person's voting preferences (to vote or not) against age, education, happiness and political interest, among other factors. The purpose is to identify which variables are most statistically significant and have the greatest impact on voting preferences in the European Union, given the long-standing decline in voter turnout. Multivariate regression analysis is used to assess and conclude whether a relationship exists. Our analysis concludes that there is a statistically significant positive relationship between age and voter preference, yet the level of education and interest in politics also contribute greatly. We conclude that electoral campaigns should target older, educated individuals interested in politics if they wish to capture the votes of an "easy target", or instead focus on younger individuals that are not interested in politics to capture the individuals that would not vote otherwise.

Introduction

Political engagement and voter turnout are critical components of a healthy democracy, and there is a long-standing interest in understanding the factors that influence a person's likelihood to vote. Political parties will often position their campaigns to appeal to people of different ages to maximize the number of votes. Historically, there has been a stark difference in voting patterns and behavior across people of all ages, which a difference in upbringing can typically explain, and values at different stages in a person's life.[1]

Across Europe, political parties have launched widespread campaigns targeting various groups within society, whether it be based on age, education, race, or income, all

in the name of incentivizing voting.[2] This has often come as a method to combat the issue of declining voter turnout within the general European Union – a trend that has been observed since the early 1970s.[3] An important factor that parties and other organizations involved in elections are age and generation one is part of.[4] Age is a reflection of one's experience and will have a significant impact on how one decides to vote and if one votes at all. There are clear intrinsic and extrinsic factors that cause individuals of all ages to vote, oftentimes it is the specific messaging used, but age appears to be a principal and influential factor.

Voter targeting, or the practice of targeting specific voters with specific messages or information, is not a problem in and of itself. It can be a useful and legitimate way for political parties to effectively communicate with voters and promote their platforms. However, voter targeting can become a problem if it is conducted in a way that is manipulative, unethical, or misleading. For example, if a political party or candidate uses voter targeting to spread false or misleading information to influence voter behavior or interest in voting, it can undermine the integrity of the electoral process. If voter targeting is used to exclude or discourage certain groups of people, based on age, race, gender, etc. it can weaken the democratic process.

This serves as the foundation for our **Research Question**: "Is there a statistically significant relationship between age and interest in voting among eligible voters in the European Union?" We furthermore intend to determine if age is more influential than other independent variables, such as education,

[1] Uppal, Sharanjit, and Sébastien LaRochelle-Côté. 2012. "Factors Associated with Voting." Safec.ca. 2012. safec.ca/userfiles/file/%C3%89conomie/EN/04%20Studies/43%20Work-Revenue-Workforce/431%20Employement%20and%20revenue/2012-PerspectivesLabourIncome-Voting.PDF.

[2] "Press Corner." n.d. European Commission - European Commission. Accessed December 18, 2022. ec.europa.eu/commission/presscorner/detail/pt/qanda_21_6212.
[3] Beley, Jules. 2019. "Is Participation in EU Elections Decreasing? | SciencesPo - Dossiers Documentaires." May 2019. dossiers-bibliotheque.sciencespo.fr/une-vie-politique-europeenne-european-political-life/participation-eu-elections-decreasing.
[4] Mijnen, Lisette 2020. "How a festival got young people to vote. Utrecht University". https://www.uu.nl/en/background/how-a-festival-got-young-people-to-vote

happiness, and interest in politics? Identifying what strongly influences one's interest in voting is critical in understanding how political parties are assembled in today's society. Furthermore, there is great value in understanding what factors influence one's interest, as it can inform us on how political parties may position their campaigns to improve voter turnout in the future.

Background

Voting is an important way for citizens to influence the decisions that affect their lives and the direction of their community, city, and country. It allows individuals to express their opinions and preferences on issues and choose leaders to represent them. In a democracy, voting is a key mechanism for holding elected officials accountable and ensuring that they represent the interests of the people. There are several underlying reasons why age has an effect on voting behavior across Europe ranging from voting experiences, education level, and social conformity.

Recently, the EU has placed additional focus on further regulating voter targeting by improving transparency while promoting free and fair elections.[5] Voter targeting is a critical tool for political parties and candidates as they seek to engage with voters and promote their message and policy stances. In particular, voter targeting allows political parties to tailor their messages to specific groups of voters, which is a more effective method than a one-size-fits-all approach. It can also increase voter turnout by reaching groups of registered voters who may be less likely to vote and encouraging them to participate. It also serves to improve campaign strategies by helping political candidates better understand the needs and concerns of specific groups of voters.

While it is possible to control certain factors that may affect the likelihood of voting at the individual level, such as nationality, gender, and education, it is not possible to control the shared experiences of a segment of individuals brought up under a given political regime or line of thinking (left/right government, for example) that may make a certain age group different from others in a country. These political generational effects are determined by a country's political history. However, when all factors are grouped into a singular analysis, we see an increasing voting trend that grows at decreasing rates. This is the typical shape of a learning curve, and it suggests that the underlying mechanism for this individual aging effect is the habituation of socially conformist behavior.[6]

In a 1968 study titled "Aging, Voting, and Political Interest", a linear correlation between age and voting was discovered. Its data suggests that young adults vote less relative to elderly people, with middle-aged members voting most frequently. They highlighted that the data describing this trend was often considered invalid since factors such as education and sex ratio are subjective to the generation and could be the driving factors of influence. Our study has found evidence for the claim that age does indeed have a strong influence on one's likelihood to vote even with the consideration of generational factors.[7]

It is important to note that there are countless factors that can influence one's interest in voting, and it is an incredibly complex task in assigning the exact measurable effect each factor has. As a result, we do not expect our analysis to provide a definitive, comprehensive overview of all factors that could affect voting preferences,

[5]EDPS. "Online Targeting for Political Advertising: Stricter Rules Are Necessary | European Data Protection Supervisor." Accessed December 18, 2022. edps.europa.eu/press-publications/press-news/press-releases/2022/online-targeting-political-advertising-stricter_en.

[6]Glenn, Norval D., and Michael Grimes. 1968. "Aging, Voting, and Political Interest." American Sociological Review 33 (4): 563. doi.org/10.2307/2092441.
[7]"Press Corner." n.d. European Commission - European Commission. Accessed December 18, 2022. ec.europa.eu/commission/presscorner/detail/pt/qanda_21_6212.

however, we do expect several statistically significant results.

The notion that individual happiness or subjective well-being can influence political behavior, including voting, is a well-established one. Many studies have concluded that subjective well-being has a positive relationship with voting.[8] It is important to note that the relationship between happiness and voting is complex and may be influenced by a variety of factors, including individual characteristics, the political context, and the specific election in question, according to studies done by Dolan, Paul, and Robert Metcalfe.[9]

Political interest is also an important factor influencing voter turnout. Studies show that individuals more interested in politics and who have higher levels of political knowledge are more likely to vote in elections. It is also true that, in many established democracies, women tend to express lower levels of political interest and give fewer correct answers to political knowledge questions than men. This gender gap in political knowledge and interest has been documented in several studies, including those conducted in the United States and Europe. There are several potential reasons for this gender gap. One possibility is that women may be less exposed to political information and less engaged in political discussions than men, leading to lower levels of political knowledge. In addition, social norms and expectations may discourage women from expressing interest in politics or participating in political activities, as discussed in Achim Goerres' paper.[10]

It is well-established that there is a negative relationship between income inequality and voter turnout. One reason for

this relationship may be that those individuals are more likely to feel disconnected from the political process and less likely to believe that their vote will make a difference. In addition, financial barriers to voting, such as the need to take time off work or pay for transportation to polling stations, may be more significant for lower-income individuals in unequal societies. It is important to note that the relationship between income inequality and voter turnout is complex and multifaceted, and other factors, such as the level of political polarization, the strength of civil society, and the accessibility of the voting process, can also affect turnout.[11]

The supporting literature has led us to formulate the **Research Question**: "Is there a statistically significant relationship between age and interest in voting among eligible voters in the European Union?" Through our research, we aim to discover the most important contributors to the voting preferences of Europeans, to establish clear priorities for electoral campaign targeting, as well as the future development of the European Union concerning the smooth running of its democratic processes.

Our **Null Hypothesis** states that there is no relationship between age and voting. On the contrary, the **Alternative Hypothesis** states that there is a positive relationship between age and voting.

H_0: No relationship exists between age and voting.   (H_0: $\beta_1 = 0$)

H_1: A positive relationship between age and voting exists. (H_1: $\beta_1 > 0$)

Similarly, the selection of the used variables has been informed by the academic literature referenced above. We proceed with the description of said variables, along with the rest of the data description.

[8]Daniel Gray, Harry Pickard, Luke Munford, 2021 "Election Outcomes and Individual Subjective Wellbeing in Great Britain". https://onlinelibrary.wiley.com/doi/full/10.1111/ecca.12362
[9]Dolan, Paul, and Robert Metcalfe. 2008. "Electing Happiness: Does Happiness Affect Voting and Do Elections Affect Happiness?" york.ac.uk/media/economics/documents/discussionpapers/2008/0830.pdf.
[10]Goerres, Achim. 2007. "Why Are Older People More Likely to Vote? The Impact of Ageing on Electoral Turnout in Europe." The British Journal of Politics and International Relations 9 (1): 90–121. doi.org/10.1111/j.1467-856x.2006.00243.x.

[11]Gallego, Aina. 2010. "Understanding Unequal Turnout: Education and Voting in Comparative Perspective." Electoral Studies 29 (2): 239–48. doi.org/10.1016/j.electstud.2009.11.002.

Data Description

voted: Voted in the last national election (0 = No, 1 = Yes)
revpolintr: How interested in politics? (1 = Not interested at all, 4 = Very interested)
agea: Age of the respondent
hinctnta: Household's total net income (1 = 1st decile, 10 = 10th decile)
eisced: Highest level of education (1 = less than lower secondary, 7 = higher tertiary education, >= MA level)
happy: Respondent's self-reported happiness (0 = Extremely unhappy, 10 = Extremely happy)
minority: Part of same race as most (0 = No, 1 = Yes)
male: Respondent is a male (0 = No, 1 = Yes)
atchctr: Emotional attachment to country (0 = Not at all emotionally attached, 10 = Very emotionally attached)
pplhlp: Most of the time people try to be helpful (0 = People mostly look out for themselves, 10 = People mostly try to be helpful)
stfgov: Satisfied with national government (0 = Extremely dissatisfied, 10 = Extremely satisfied)

The data has been acquired from round 10 of the ESS, using data from 32 countries. Conducted every two years among various European countries, the ESS uses face-to-face interviews to collect data on changing values, attitudes, and behaviors in Europe.

Fig. 1. Summary of Variables Used

Variable	Obs	Mean	Std. dev.	Min	Max
voted	16,721	.7198732	.4490744	0	1
agea	17,940	50.88645	18.4512	15	90
eisced	17,973	4.267624	1.636931	1	7
hinctnta	14,695	5.448452	2.771929	1	10
revpolintr	18,018	2.220335	.8817298	1	4
atchctr	17,951	8.127235	2.111797	0	10
stfgov	17,640	4.289452	2.710664	0	10
happy	18,014	7.139003	2.053278	0	10
minority	17,919	.0800826	.2714286	0	1
male	18,060	.448505	.497355	0	1
pplhlp	17,980	4.792436	2.48935	0	10

Preliminary Screening

Before conducting our analysis, we examined the data to identify and address potential issues that could bias results. First, we tabulated and visualized the variables, allowing us to recognize missing/erroneous values. Such data points were imputed to remove their influence. No significant outliers were found in any of the variables, most likely due to the restricted range of accepted inputs for the majority of the questions (such as anchored scales or yes/no questions).

We also created three dummy variables – **voted**, **minority,** and **male** – to facilitate our analysis. **voted** is derived from **vote** in the

ESS, asking respondents if they have voted in the last elections in their country. It is designed to account for either "yes" and "no" values, expressed as 1 or 0 respectively. Only respondents who are eligible to vote have been considered in our analysis.

Generated from **feethngr**, **minority** asks respondents if they feel part of the same race/ethnic group as most people in their country, and is processed to only include "yes" and "no" values. While not the ideal representation of being a minority, it is the very perception that interests us for the analysis, as the perception of being a minority is what may impact one's actions. This is due to a sense of belonging to a marginalized group, which can motivate members of minority populations to exercise their right to vote.[12]

Additionally, **male** is generated from **gndr**, and turned into a dummy so its impact can be easily observable.

Lastly, **polintr** has been reversed. Since its original scale went from 1 to 4, with the highest number representing the lowest interest in politics, we found it to be counterintuitive. The newly reversed variable **revpolintr** maintains all the properties of **polintr**, yet makes more sense with the highest value of 4 representing the highest interest in politics.

---

[12]Gee, L. (2018, December 14). How Identity Politics Drives Voter Turnout. Harvard Business Review. hbr.org/2018/12/how-identity-politics-drives-voter-turnout

Descriptive Statistics

As seen in Fig. 1, the sample size is sufficiently large, reducing the effect of potential bias and making it less likely that the sample will be skewed by outlier values and so more representative of the population.

Our main dependent variable **voted** has a mean of 0.72, thus most respondents had voted in the last election. Yet, the election type is not specified, nor is the typical voting behavior. Our assumption is that if one voted in the last elections at any level, he also votes in all other elections. The binary nature of **voted** may also cause issues with the model, as it is not a continuous variable and does not subject itself well to linear regression.

The mean value of the key independent variable **agea** is 50.89, resulting in an approximate average respondent age of 51.

The most common response for **polintr** is 2 (the mean is 2.22), signifying a mild interest in politics in most respondents. Less than 10% are "very interested" in politics. However, narrow anchored scales as seen here can lead to bias by limiting the range of responses and the amount of information collected, resulting in simplified models which are unable to capture the nuance of data.

The average value of **minority** is 0.08, thus the vast majority of respondents do not perceive themselves as racial or ethnic minorities in their respective countries.

The variable **eisced** describes the highest education level on the ES-ISCED scale. The mean is 4.27, so an average respondent has between an upper secondary (value of 3), an advanced vocational (4), or a bachelor's degree (5). We assume proportionality here, thus a one-unit increase in the level of education has the same effect on **votes** no matter the starting point or the level. This is not the most realistic assumption, but it is necessary for our analysis.

Representing a self-reported measure of happiness, **happiness** introduces bias, as true happiness is impossible to measure and few

people would wish to report being unhappy. Hence, there is a skewness of the distribution towards higher happiness scores (Fig. 2.)

Fig. 2. Distributions of **happy** and **atchctr**
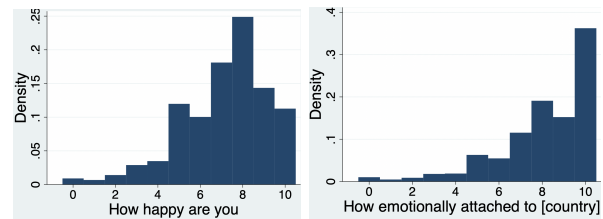
The variable **atchctr** shows ~35% of the respondents feeling very emotionally attached to their country (score of 10), and ~80% of respondents are attached to their country (score of 7+). Such a skewed distribution, as also seen in **happy**, may introduce bias, potentially threatening the linear nature of the model (if the effects are severe).

Fig. 3. Box-Plot **voted** vs **agea**
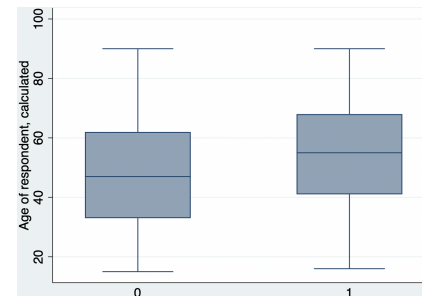
The box-and-whiskers plot alludes to the existence of the investigated relationship. The average age is higher for those who vote, as is the 3rd quartile. These observations contradict our null hypothesis, serving as grounds to continue our study by conducting multivariate analysis to conclusively prove or disprove H_0.

Before beginning our analysis, we ran a correlation matrix on our variables to determine if Factor Analysis (FA) is a suitable method and if any endogenous variables have been included. For FA, a correlation of >0.5 is considered to indicate a strong correlation, while <0.3 indicates a weak correlation between a pair of variables. Fig. 4 shows that among the resulting values, all but two values in the matrix fell in the "weak" or "no correlation" range. Therefore the correlation

is not strong enough, indicating that they should not be grouped as factors. Hence, FA is not the most appropriate technique to use. Additionally, variables such as **revpolintr**, which we were wary of including at first, proved to be valid for this model, showing that while there is some correlation with **voted**, it is not too strong.

Fig. 4. Correlation Matrix

	voted	revpol~r	agea	hinctnta	eisced
voted	1.0000				
revpolintr	0.2880	1.0000			
agea	0.1397	0.1529	1.0000		
hinctnta	0.0901	0.1156	-0.3582	1.0000	
eisced	0.1523	0.2270	-0.2206	0.3768	1.0000
happy	0.1132	0.0859	-0.1608	0.2867	0.1606
minority	-0.0551	-0.0324	-0.0235	-0.0337	-0.0710
atchctr	0.1724	0.1428	0.2258	-0.0374	-0.0074
pplhlp	0.0904	0.1026	-0.0129	0.1134	0.0930
stfgov	0.1135	0.1395	0.0422	0.1069	0.0710

	happy	minority	atchctr	pplhlp	stfgov
happy	1.0000				
minority	-0.0479	1.0000			
atchctr	0.1648	-0.0777	1.0000		
pplhlp	0.2480	-0.0137	0.0807	1.0000	
stfgov	0.2418	-0.0098	0.1580	0.2889	1.0000

Multivariate Analysis

Therefore, the analysis of the impact of our independent variables on the exercise of voting rights (**voted**) in the last national election was done through multivariate linear regression. Country weights were used to more accurately represent the countries, since samples for each country were roughly equal, despite population differences. Seven models have been fitted and displayed to best identify the most impactful variables. Furthermore, only adjusted R-squared figures are shown to prevent overfitting and allow for an appropriate comparison of models and variables after the analysis.

The overall results are satisfactory. As seen in Fig. 5 below, an initial regression was run on our main dependent variable **voted** against the key independent variable **agea**. A statistically significant relationship has emerged with an R-squared of 0.024 and a p-value lower than 0.01.

This signifies that 2.4% of the variability in voting behavior is explained by the variability in **agea**. This may appear a low, insignificant value, but it is once again worth considering that voting, or any preferences of an individual for that matter, are an incredibly complex phenomenon that cannot be determined by any single set of influencing factors, never mind a single specific one. In other words, a regression on only one explanatory variable lacks completeness and is likely to suffer from omitted variable bias.

As a result, we added more control variables one at a time to discern the contribution of each one to the model and visualized the model's rising strength in explaining the data.

The second model includes **eisced**. Here too we see a statistically significant, positive relationship, that brings the total model adj. R-squared value to 0.091. We included this variable as individuals with a higher level of education are intuitively more likely to be informed about the electoral process, understand the issues at stake and have the resources to participate in an election.

Similarly, **hinctnta**, **polintr**, and **stfgov** show a statistically significant positive relationship to **voted**. Household income is significant, raising our adj. R-squared to 0.103. This is in line with our expectations and theory as many studies, including Solt (2008), and Anderson and Beramendi (2008) find that individuals living in more unequal societies are less likely to vote. This relationship is particularly strong among lower-income individuals, who may feel less invested in the political process or less able to participate due to financial constraints.[13]

Political interest is an intuitive inclusion within the variables – politically engaged people are more likely to vote than someone

[13]Gallego, Aina. 2010. "Understanding Unequal Turnout: Education and Voting in Comparative Perspective." Electoral Studies 29 (2): 239–48. doi.org/10.1016/j.electstud.2009.11.002.

Fig. 5. Regression Results

VARIABLES	(1) voted	(2) voted	(3) voted	(4) voted	(5) voted	(6) voted	(7) voted
agea	0.00400***	0.00763***	0.00827***	0.00693***	0.00640***	0.00643***	0.00647***
	(0.000198)	(0.000361)	(0.000404)	(0.000423)	(0.000434)	(0.000439)	(0.000442)
eiscd		0.0570***	0.0433***	0.0279***	0.0286***	0.0287***	0.0256***
		(0.00380)	(0.00455)	(0.00464)	(0.00466)	(0.00471)	(0.00465)
hinctnta			0.0195***	0.0148***	0.0140***	0.0132***	0.0120***
			(0.00273)	(0.00272)	(0.00272)	(0.00275)	(0.00282)
revpolintr				0.101***	0.0948***	0.0926***	0.0975***
				(0.00830)	(0.00834)	(0.00846)	(0.00850)
atchctr					0.0241***	0.0233***	0.0202***
					(0.00372)	(0.00376)	(0.00387)
stfgov						0.00225	-0.00114
						(0.00263)	(0.00276)
happy							0.0119***
							(0.00392)
minority							-0.0624**
							(0.0255)
male							-0.0224
							(0.0142)
pplhlp							0.00746**
							(0.00305)
Constant	0.509***	0.0331	-0.0584*	-0.132***	-0.285***	-0.278***	-0.335***
	(0.0110)	(0.0281)	(0.0333)	(0.0328)	(0.0405)	(0.0410)	(0.0463)
Observations	16,604	16,533	13,729	13,706	13,655	13,481	13,369
Adj. R-squared	0.024	0.091	0.103	0.138	0.147	0.145	0.153

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

who is not, since interest impacts their passion for the issues and candidates at stake in elections. Hence, **revpolintr** increases the adjusted R-squared value to 0.138.

Attachment to one's country similarly shows a positive correlation, passing the 1% significance test. This is in line with what one could expect regarding patriotism. Voters who have a strong feeling of attachment and pride towards their country will be more inclined to vote and impact the direction of their country.[14]

Interestingly, **stfgov** does not have a strong impact on our model's ability to explain the data, as seen by the lack of asterisks next to the value. This signifies a high p-value, which means that the positive relationship present is not statistically significant at any relevant level (including 10%). This is surprising, as we expected there to be a significant negative relationship, showing that as satisfaction with the government decreases, people tend to vote more to change the government. For example, according to "Emotions and voting in EU referendums", voters use referendums as a vehicle to express their dissatisfaction or judgment on a government by deciding to vote or not to vote.[15] The inclusion of this variable has reduced our model's ability to explain the data by 0.2%. While this is not a significant deviation, the decrease in R-squared itself tells us this variable should not be included.

Lastly, the most comprehensive model was also tested to control for any other variables that could contribute to one's interest in politics. These were grouped due to two main reasons. First, the underlying literature suggests that we have already analyzed the most important contributors to voting behavior separately and that these last remaining factors are not as crucial. Secondly, they reflect attitudes and opinions more than facts (apart from male), and therefore it is not illogical to group them to be able to see their impact, as any individual variable's

[14]Lubbers, Marcel and Coenders, Marcel 2017. "Nationalistic attitudes and voting for the radical right in Europe" ... https://journals.sagepub.com/doi/10.1177/1465116516678932

[15]Garry, John. 2013. "Emotions and Voting in EU Referendums." European Union Politics 15 (2): 235–54. doi.org/10.1177/1465116513514780.

contribution may not have noticeably changed the R-squared. In fact, we see that the combined impact of the last four variables only increases the R-squared by .008, up to 0.153.

Interestingly, this very group has also shown the most deviations from both our expectations and the literature. While **happy** is in line with the expectations, **minority** and **male** show a trend opposite to what is suggested by academics.

In fact, both of the latter display a negative direction, implying that the respondents who have identified as minorities and/or males vote less than their counterparts. There are countless reasons why minority groups may not vote more than their counterparts. For example, it could be related to a lack of attachment to the country, differences in cultural and religious values and customs, or simply a lack of interest in a country that is not considered "home".

Furthermore, **male** is not statistically significant relative to other explanatory variables, failing to pass even the 10% significance level. While it is unclear as to what is causing the misalignment of **male** to previous studies, it is evident that further research would need to be completed to reassess this relationship. Nevertheless, given it is not statistically significant it should not be considered in this model.

Similarly to **happy**, **pplhlp** demonstrates a significant positive relationship, in-line with the suggestions of academics. Various studies, including "Electing Happiness: Does Happiness Affect Voting," by Paul Dolan, Robert Metcalfe, and Nattavudh Powdthavee have found that individuals reporting higher levels of happiness are more likely to participate in the political process, including voting in elections. The authors also found that the relationship between happiness and voting was stronger for individuals who were more politically engaged, such as those who

had a higher level of political knowledge or were more interested in politics.

Post-Regression Estimation

The following is the description of the analyses we conducted to ensure the statistical integrity and validity of our findings.

Multicollinearity refers to a situation in which two or more predictor variables in a regression model are highly correlated. This can pose a problem as it can affect the interpretation of the regression coefficients and make it difficult to determine the unique effects of each predictor on the outcome variable. Multicollinearity can also make the model less stable and more sensitive to small changes in the data.

Fig. 6. VIF Analysis

| Variable	VIF	1/VIF
hinctnta	1.35	0.739330
eisced	1.35	0.739974
happy	1.22	0.818981
agea	1.22	0.820371
revpolintr	1.21	0.828097
stfgov	1.16	0.862699
atchctr	1.14	0.880482
pplhlp	1.11	0.900965
male	1.05	0.953225
minority	1.01	0.990497
Mean VIF	1.18	

To detect multicollinearity, the Variance Inflation Factor (VIF) has been computed. VIF greater than >4 represents the presence of multicollinearity, while a VIF >10 signifies very strong multicollinearity. On the other hand, a VIF score of 1 indicates that no multicollinearity is found. Having undergone this analysis, our model has shown a mean VIF score of 1.18, with the highest VIF score received by the variable **hinctnta** at 1.35 – all of which are reasonable figures demonstrating that multicollinearity is not present.

Heteroscedasticity refers to the situation in which the variance of a variable is non-constant across observations. This means that the data exhibits unequal dispersion, with some observations showing a larger variation

than others. This may have implications for the validity of our statistical test. Since the data used in our analysis is cross-sectional (as opposed to time series), heteroscedasticity is not as important, yet it remains a concern.

As seen in Fig. 7, heteroscedasticity is certainly present and significant in the case of this regression. To further confirm this, we continued our analysis by running the Breusch-Pagan test for heteroscedasticity, which resulted in a high chi-squared value of 797.78, as well as Prob>chi-squared = 0.0. This further validates our concerns regarding heteroscedasticity.

Fig. 7. Visualization of Heteroscedasticity
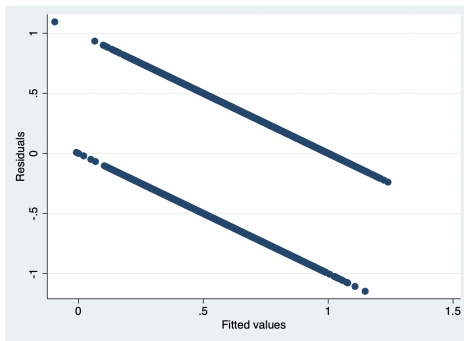
As a result, we repeated our regression model, this time integrating robust standard errors as opposed to previously used standard errors. This allows us to see if the level of heteroscedasticity present in our model is sufficient to reject the null hypothesis.

Robust standard errors adjust for heteroscedasticity by using a different formula to calculate the standard errors of estimators. It also adjusts for the unequal variance of errors, which can make the estimates more accurate and reduce the bias of the estimates.

As seen in Fig. 8, the new regression with robust errors has proven satisfactory. Its results are in-line with the outcome of the previous regression, with the same R-squared and p-values throughout. Most importantly, the p-values and significance of our main independent variable, **agea**, remain consistent. Interestingly, the R-squared of the standalone model of **age** vs **voted** is higher with robust errors, at 0.049.

Endogeneity is the degree to which the independent variable of a regression model is correlated with the error term. The implementation of robust standard errors aims to minimize the potential effects of endogeneity.

Fig. 8. Regression Results (robust errors)

VARIABLES	(1) voted	(2) voted	(3) voted	(4) voted	(5) voted	(6) voted	(7) voted
agea	0.00611***	0.00763***	0.00827***	0.00693***	0.00640***	0.00643***	0.00647***
	(0.000376)	(0.000361)	(0.000404)	(0.000423)	(0.000434)	(0.000439)	(0.000442)
eisced		0.0570***	0.0433***	0.0279***	0.0286***	0.0287***	0.0256***
		(0.00380)	(0.00455)	(0.00464)	(0.00466)	(0.00471)	(0.00465)
hinctnta			0.0195***	0.0148***	0.0140***	0.0132***	0.0120***
			(0.00273)	(0.00272)	(0.00272)	(0.00275)	(0.00282)
revpolintr				0.101***	0.0948***	0.0926***	0.0975***
				(0.00830)	(0.00834)	(0.00846)	(0.00850)
atchctr					0.0241***	0.0233***	0.0202***
					(0.00372)	(0.00376)	(0.00387)
stfgov						0.00225	-0.00114
						(0.00263)	(0.00276)
happy							0.0119***
							(0.00392)
minority							-0.0624**
							(0.0255)
male							-0.0224
							(0.0142)
pplhlp							0.00746**
							(0.00305)
Constant	0.355***	0.0331	-0.0584*	-0.132***	-0.285***	-0.278***	-0.335***
	(0.0218)	(0.0281)	(0.0333)	(0.0328)	(0.0405)	(0.0410)	(0.0463)
Observations	16,604	16,533	13,729	13,706	13,655	13,481	13,369
Adj. R-squared	0.049	0.091	0.103	0.138	0.147	0.145	0.153

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Nevertheless, the large heteroscedasticity is a concern, especially as we found the error mean to be non-zero, but instead 0.053. We decided to validate if we have achieved BLUE (best, linear, unbiased, efficient) Ordinary Least square estimators, we must check if the following conditions have been met by our data.

Fig. 9. QQ Plot
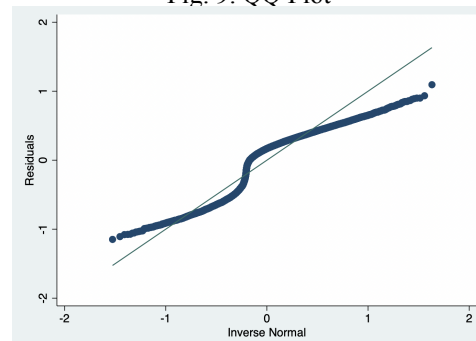
The model will be a poor fit if we attempt to fit a linear model to non-linear data. A QQ plot assesses the normality of residuals. If these are normally distributed, this suggests that the assumptions of the linear model are reasonable. Fig. 9 hints at the non-linear nature of our regression. The deviation is not particularly large, yet the shape is not uniform. This is likely since our dependent variable **voted** is binary. As a remedy, we

decided to run a non-linear, logistic, regression that is usually used for binary outcomes.

Fig. 10. Logistic Regression Results

```
Logistic regression                          Number of obs =    13,369
                                             LR chi2(10)   =   1853.72
                                             Prob > chi2   =    0.0000
                                             Pseudo R2     =    0.1192
```

voted	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
agea	.0201919	.0013856	14.57	0.000	.0174761	.0229077
eisced	.1583915	.0150686	10.51	0.000	.1288575	.1879255
hinctnta	.0559828	.0089477	6.26	0.000	.0384456	.0735199
revpolintr	.6513965	.027268	23.89	0.000	.5979522	.7048409
atchctr	.1065206	.0101717	10.47	0.000	.0865845	.1264567
stfgov	.02787	.0084371	3.30	0.001	.0113335	.0444065
happy	.0678919	.0109554	6.20	0.000	.0464196	.0893641
minority	−.2004855	.0782868	−2.56	0.010	−.3539248	−.0470463
male	−.1371578	.0432916	−3.17	0.002	−.2220078	−.0523077
pplhlp	.024974	.0089867	2.78	0.005	.0073604	.0425875
_cons	−3.919435	.1503629	−26.07	0.000	−4.214141	−3.62473

Fig. 10 shows the outcome of the logistic regression. The Pseudo R-squared of this model is 0.12, showing that a similar proportion of the variance of data is explained by this model as by its linear counterpart. Furthermore, the z-values of most of the variables (apart from **minority** and **pplhlp**) are under the 5% significance level, with most equalling 0.

Fig. 11. ROC Curve

Area under ROC curve = 0.7333
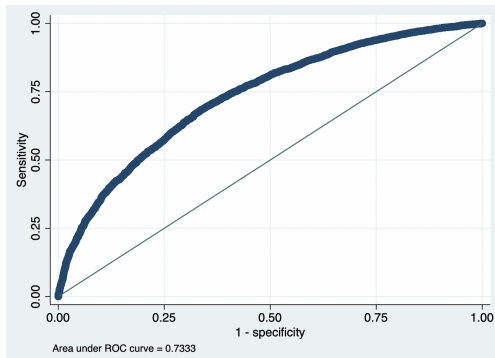
The ROC curve visualizes the predictive power of our model. Fig. 11 shows the model is certainly better than a random guess, shown by the straight diagonal line, but is not perfect by any means, as a "perfect" model would outline the top left corner of the graph. Similarly, the area under the ROC curve of 0.73 means our model holds predictive power.

At last satisfied with the models, we proceed with the conclusion of our results. It is important to note that we consider the results of both models, carefully interpreting the linear model despite its apparent weaknesses, and supporting it with the results of the logistic regression. Further analysis of the validity of the logistic model is not performed, as this is too far outside the scope of this course.

Conclusions

This study aimed to investigate the relationship between age and interest in voting among eligible voters in Europe and to determine whether age is more influential than other variables such as education, happiness, and interest in politics. The results of the statistical analysis directly **answered our Research Question**, indicating that **there is a positive, statistically significant relationship between age and voting preferences**, with older individuals reporting higher levels of interest in voting. In addition, **age was found to be the most influential variable** out of the ones considered (based on its contribution to the adjusted R-squared). However, education and interest in politics are also important in predicting interest in voting.

One limitation of this study is that it only considers the variables of the ESS, and other not-included variables might significantly affect interest in voting. Furthermore, our model revealed some non-linear tendencies along with heteroscedasticity, therefore our findings should be interpreted with a grain of salt. Nevertheless, the logistic regression results support our findings.

Our findings imply that electoral campaigns should target older, educated individuals interested in politics if they want an "easy target" to capture, or instead younger individuals that are not interested in politics to capture the individuals that would not vote otherwise.

Future research could consider a wider range of variables to more fully understand the factors that influence interest in voting.

Bibliography

Ansolabehere, Stephen, and Eitan Hersh. 2011. "Gender, Race, Age, and Voting: A Research Note." Papers.ssrn.com. Rochester, NY. 2011. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1901654.

Beley, Jules. 2019. "Is Participation in EU Elections Decreasing? | SciencesPo - Dossiers Documentaires." Dossiers-Bibliotheque.sciencespo.fr. May 2019. http://dossiers-bibliotheque.sciencespo.fr/une-vie-politique-europeenne-european-political-life/participation-eu-elections-decreasing.

Dolan, Paul, and Robert Metcalfe. 2008. "Electing Happiness: Does Happiness Affect Voting and Do Elections Affect Happiness?" http://york.ac.uk/media/economics/documents/discussionpapers/2008/0830.pdf.

EDPS. "Online Targeting for Political Advertising: Stricter Rules Are Necessary | European Data Protection Supervisor." Edps.europa.eu. Accessed December 18, 2022. http://edps.europa.eu/press-publications/press-news/press-releases/2022/online-targeting-political-advertising-stricter_en.

Gallego, Aina. 2010. "Understanding Unequal Turnout: Education and Voting in Comparative Perspective." *Electoral Studies* 29 (2): 239–48. https://doi.org/10.1016/j.electstud.2009.11.002.

Garry, John. 2013. "Emotions and Voting in EU Referendums." European Union Politics 15 (2): 235–54. https://doi.org/10.1177/1465116513514780.

Gee, L. (2018, December 14). How Identity Politics Drives Voter Turnout. Harvard Business Review. https://hbr.org/2018/12/how-identity-politics-drives-voter-turnout

Glenn, Norval D., and Michael Grimes. 1968. "Aging, Voting, and Political Interest." *American Sociological Review* 33 (4): 563. https://doi.org/10.2307/2092441.

Goerres, Achim. 2007. "Why Are Older People More Likely to Vote? The Impact of Ageing on Electoral Turnout in Europe." *The British Journal of Politics and International Relations* 9 (1): 90–121. https://doi.org/10.1111/j.1467-856x.2006.00243.x.

Greene, Steven. 1999. "Understanding Party Identification: A Social Identity Approach." *Political Psychology* 20 (2): 393–403. https://doi.org/10.1111/0162-895x.00150.

Kostelka, Filip, André Blais, and Elisabeth Gidengil. 2018. "Has the Gender Gap in Voter Turnout Really Disappeared?" *West European Politics* 42 (3): 437–63. https://doi.org/10.1080/01402382.2018.1504486.

"Press Corner." n.d. European Commission - European Commission. Accessed December 18, 2022. http://ec.europa.eu/commission/presscorner/detail/pt/qanda_21_6212.

Rauh, Christopher. 2017. "Voting, Education, and the Great Gatsby Curve." *Journal of Public Economics* 146 (February): 1–14. https://doi.org/10.1016/j.jpubeco.2016.12.005.

Stockemer, Daniel, and Lyle Scruggs. 2012. "Income Inequality, Development and Electoral Turnout – New Evidence on a Burgeoning Debate." *Electoral Studies* 31 (4): 764–73. https://doi.org/10.1016/j.electstud.2012.06.006.

Uppal, Sharanjit, and Sébastien LaRochelle-Côté. 2012. "Factors Associated with Voting." Safec.ca. 2012. safec.ca/userfiles/file/%C3%89conomie/EN/04%20Studies/43%20Work-Revenue-Workforce/431%20Employement%20and%20revenue/2012-PerspectivesLabourIncome-Voting.PDF.

Wolfinger, Raymond E. 1965. "The Development and Persistence of Ethnic Voting." *The American Political Science Review* 59 (4): 896–908. https://doi.org/10.2307/1953212.

Mijnen, Lisette 2020. "How a festival got young people to vote. Utrecht University". https://www.uu.nl/en/background/how-a-festival-got-young-people-to-vote

Daniel Gray, Harry Pickard, Luke Munford, 2021 *"Election Outcomes and Individual Subjective Wellbeing in Great Britain". https://onlinelibrary.wiley.com/doi/full/10.1111/ecca.12362*

Lubbers, Marcel and Coenders, Marcel 2017. "Nationalistic attitudes and voting for the radical right in Europe". https://journals.sagepub.com/doi/10.1177/1465116516678932