Education, Age, and the Liberal Vote: A Stratified Analysis of Canadian Voter Preferences

STA304 - Winter 2025 - Assignment 2

Group 24: You Yu Fu, Eric Guo, Finn Tran, Lina Jin

1 Introduction

Education, in general, has a significant impact on the development of the human brain. Prior research tells us, through various curriculums, education has exposed us to various ideologies and attitudes that ultimately shape our individual opinions (Leś, 2024), which in turn, links education's role in political views and involvement (Hooghe et al., 2015). Studies have shown that an individual's level of education is influential on their political views (Scott, 2022), and that those with a higher level of education tend to gravitate towards left-leaning views of the liberal and New Democratic Party (NDP) views rather than their right-leaning counterparts (Kiss et al., 2023). Though, to what extent does education level affect the voting behaviours of Canadian Liberals, when controlling for age?

In our study, we hope to focus our landscape on the Canadian Liberal Party, to see whether or not the level of education impacts the voting patterns of Canadian liberals, utilizing the 2019 Canadian Federal Election Study, which is a survey intended to gather Canadians' opinions and thoughts before and after the 2019 Canadian Federal Election. The target population is Canadians aged 18 or older. The frame population is Canadians of or over the age of 18 with access to a phone in the database. The sampling population are the Canadians who voluntarily took the survey via phone. This gives us an insight into the patterns of Canadian voters and what thoughts, motivations, interests, accompany these voters during social, economic, and political issues.

If our findings reflect the findings of these papers, what does a more-educated, left-learning voting pattern of citizens say about the educational system of Canada? By examining Liberal voting behaviour based on age and stratified by education level, we aim to assess whether higher education is a strong predictor of Liberal voting patterns in Canada and encourage discussions on how education shapes political engagement in Canadian society.

2 Data

For this project, we will be analyzing the **phone survey data set of the 2019 Canadian Federal Election Study** to investigate if **education level and age** affect voting party preference. To do this, we will be **stratifying our data into based on education level**, using **International Standard Classification of Education (ISCED) levels** (Government of Canada, Statistics Canada, 2023) and **the Government of Canada 2021 Census Education categories** (Government of Canada, Statistics Canada, 2021) as a reference to group our participants based on their response to *q61* in the Phone Data Dictionary which asks "What is the highest level of education that you have completed?".

The ISCED scale offers a standardized framework for comparing education systems across countries (Government of Canada, Statistics Canada, 2023). We utilized this scale in conjunction with variables from the 2021 Canadian Census data to stratify our population accordingly.

By stratifying the sample, we are able to improve the precision of our estimates by accounting for within-group variations of education levels.

Table 1: Our mapping of Phone Data Dictionary Q61 to Education levels

| Response	Corresponding Education Level	Education Level Definition
No schooling	0	Early childhood education
Some elementary school	0	Early childhood education
Completed elementary school	0	Early childhood education
Some secondary/high school	0	Lower secondary education
Completed secondary/high school	1	High school diploma
Some technical, community college	1	High school diploma
Some university	1	High school diploma
Completed technical, community college	2	Short-cycle tertiary education
Bachelor's degree	3	Bachelor or equivalent level
Master's degree	4	Master/PhD or equivalent level
Professional degree or doctorate	5	PhD or equivalent level

Respondents who picked the option ("Don't know", "Refused to answer" or "Skipped") were removed from the data set, with only 11 out of the over 4000 participants choosing this response. The other variables of interest were cleaned, where we removed all NA values and invalid voter ages (age < 18). We used q11 to determine surveyor voting choice, where we grouped them into three categories: "Liberal" for those who said they would vote Liberal, "Not Liberal" for

those who said they would vote for a party that was not the Liberal party, and "Unknown" for those who skipped the question or chose not to answer.

Given that over 25% of phone survey participants chose not to say which party they were going to vote for in the upcoming election, we will generalize this category (q11) into 3 responses, 'Unknown', 'Liberal', and 'Not Liberal'. Removing all these observations could introduce bias into our analysis and decrease the significance of our results.

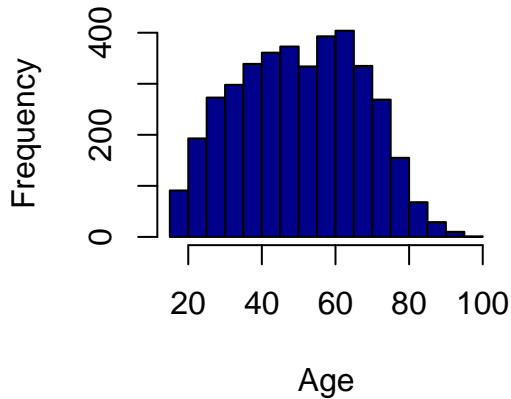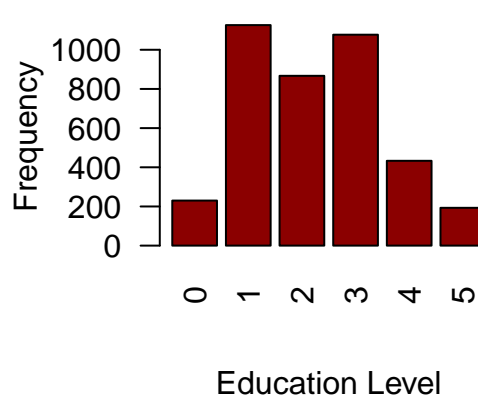Plotting Education Level (based on q61), and Voting preference (based on q11):
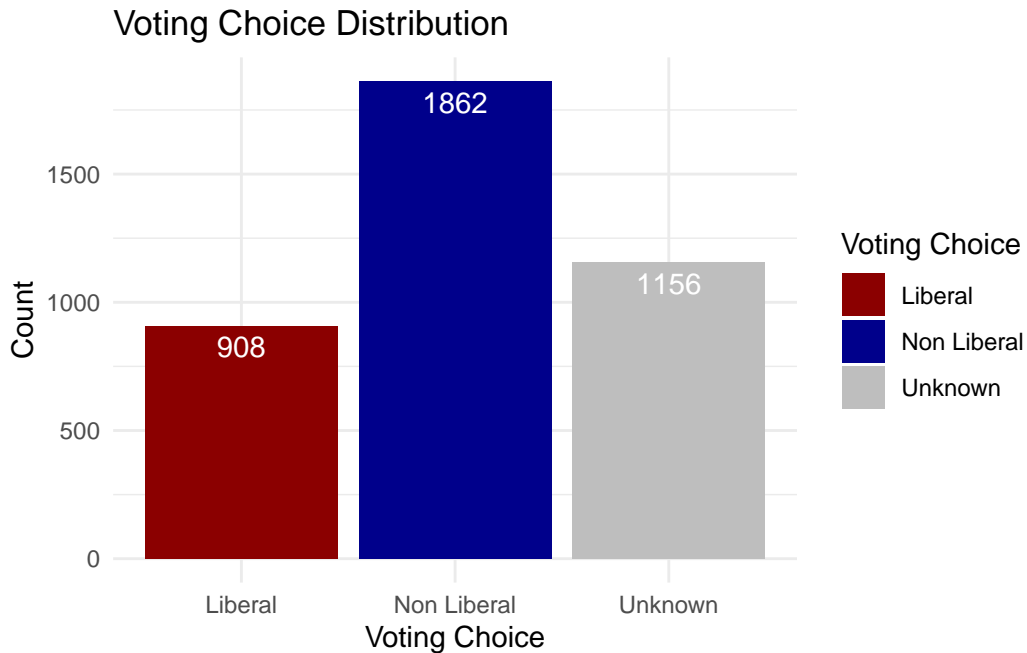
Age Boxplot
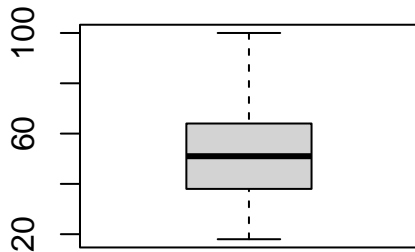
Education Boxplot
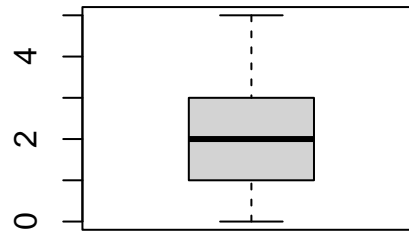
Age Distribution

Phone Data Distribution

Voting Choice Distribution

Education LEVEL

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	1.000	2.000	2.238	3.000	5.000

AGE

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
18.00	38.00	51.00	50.89	64.00	100.00

From the summary statistics and graph, we can see that the survey population is decently well educated, with a mean and median ISCED score of around 2. This means that the average survey respondent has at least completed community college. The ISCED_Level distribution appears to be right skewed, with the interquartile range from 1 to 3 from the box plot and histogram. The age seems slightly right-skewed, with the interquartile range from late 30's to early 60's from the box plot, but has a higher range of ages from 18 to late 90's. Among the web survey respondents, it shows a slight bi-modal distribution on the histogram, with the interquartile range from from late 30's to early 60's.

We can make further comparisons on phone survey participants in the 'Unknown' category with those who provided an answer.

Bar Plot for 'Unknown'   **Bar Plot for 'Known'**

ISCED Level   ISCED Level
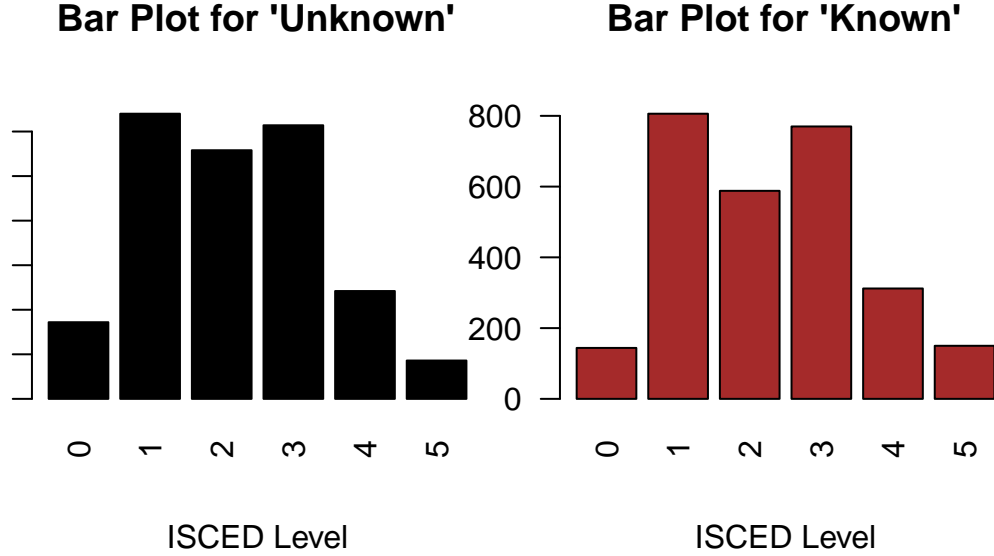
These distributions are very similar, therefore, they could be removed without introducing signficant bias into our data. Since we are considering those who specifically voted Liberal in the 2019 Canadian Federal Election, those who inputted Unknown will be insignificant to our study anyway.

3 Methods

To calculate the Confidence Interval for a proportion and stratified sampling, the following formula will be used:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\sum_{h=1}^{H} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_h^2}{n_h}}$$

Where $\hat{p} = \sum_{h=1}^{H} W_h \hat{p}_h$ and $W_h = \frac{N_h}{N}$

In the formula,

- H is the number of Education Level strata.

- h denotes a specific stratum.

- $\hat{p}_h$ specifies the sample proportion of Liberal voters from the strata h.

- N_h is the population size of the stratum h, with numbers from Statistics Canada (Statistics Canada, 2021).

5

- n_h is the sample size of the stratum h.

- s_h^2 indicates the sample variance of the stratum h.

- $Z_{\alpha/2}$ is the critical value from the standard normal distribution corresponding to a 95% confidence level ($\alpha = 0.05$)

Below is the constructed logistic regression model:

$$\log \left(\frac{1 - P(\text{Vote Liberal})}{P(\text{Vote Liberal})} \right) = \beta_0 + \beta_1 (\text{Education Level}) + \beta_2 (\text{Age})$$

- Education Level, the ordinal variable that the population is stratified by, defining the respondent's education level

- Age, discrete variable for the respondent's age

- $P(VoteLiberal)$ describes the probability of the person voting for the Liberal party. The intercept value β_0 references the expected log-odds of a person with Education Level = 0 and age 0 to vote for the Liberal party. The interpretation is not relevant as a person of age 0 cannot vote.

- The coefficient β_1 identifies the expected effect of Education Level in log-odds of voting Liberal compared to the baseline education level (ISCED Level = 0) when Age is kept constant.

- β_2 is the expected effect of an unit change in Age on the log-odds of voting Liberal, when Education Level is kept constant.

4 Results

Present a table showing the estimated proportion of votes for the selected party along with the 95% confidence interval, and include text describing this table and the key takeaways.

In Table 2, we present the confidence intervals of phone survey:

Table 4: Proportion of Outcome Variable and 95% Confidence Interval of Outcome Variable

	Proportion of Outcome Variable	95% Confidence Interval of Outcome Variable
Phone Survey	0.20914	(0.1950576, 0.2232315)

The proportions and 95% confidence intervals of outcome variable of interest calculated for the Canadian Election Study 2019 phone survey data. The 95% Confidence Interval (CI) for the proportion of participants who voted Liberal was calculated using the stratified phone

survey data. The estimated proportion of Liberal voters in the sample is 20.91%. Based on the confidence interval, we can say that we are 95% confident that the true proportion of Liberal voters in the population of Canadian voters falls between 19.51% and 22.32%. This interval accounts for the variability in our sample and suggests that, if we were to repeat this sampling process many times, approximately 95% of the resulting confidence intervals would contain the true population proportion of Liberal voters.

Below is the estimate regression model:

Table 5: Estimated Logistic Regression Model

	Estimate (β)	Standard Error	p-value
Intercept	-2.373643	0.163359	< 2e-16 ***
ISCED Level	0.208224	0.038918	9.28e-08 ***
Age	0.012804	0.002728	2.77e-06 ***

Accounting for stratification, -2.373643 is the log-odds intercept for a person with no school on the ISCED scale (ISCED level = 0) and zero years old. This is just a baseline variable, since anyone who is 0 years old cannot vote.

β_1 is positive and p-value is < 0.05, meaning that higher education is associated with higher odds of voting Liberal at significance level. The log odds is 0.208224, meaning that for each one-unit increase in education level, the log-odds of voting Liberal increase by around 20.8%.

β_2 is positive with a log-odds value of 0.012804 and p-value is < 0.05, meaning that older age is slightly associated with higher log-odds of voting Liberal.

5 Discussion

Based on our findings, we found that both education and age have a positive effect on the likelihood of voting Liberal, with education level increasing the log-odds of voting for the Liberal party by 17.5% and age increasing the log-odds of voting for the Liberal party by 0.8%. While both are statistically significant, we can suggest that education level has a larger effect on the likelihood for voting Liberal party than age, drawing parallels from previous studies that have suggested that Canadians with a higher level of education tend to vote Liberal (citation). Age has some impact, which also may be correlated with the fact that those with a higher level of education tend to be older (citation).

Though, since Liberal party is only 20.91% of our population, we cannot say that education level is exclusively associated with the Liberal party, or that education is the only thing that is increasing the likelihood of voting Liberal.

Since our analysis was drawn from a phone survey data, we need to account for selection bias: the survey was voluntary, based on those who had access to a technological device, and those who followed up in the survey. The phone data age distribution demonstrated a slight right-skew, possibly suggesting that an older demographic responded to the phone survey. This may introduce some bias since those who are older are likely to have completed a full education compared to those who are younger, closer to 18 years old. Further analyses could be done focused on certain age groups rather than all the ages to have a more focused analysis on education level on voting patterns. We also need to consider that the ISCED is a universal classification of education-level, meaning that the scale is generalized to fit the standard of all education levels across the globe, where there is much variation in how education is offered. To consider Canadian-specific education standards, we may get more reliable, specific findings from a Canadian measure of education. Furthermore, this scale considers higher levels of education dependent on primary, secondary, and postsecondary levels of education. This means that it may omit those who did not complete a certain level of formal education, limiting our analysis to those who have been formally enrolled in education. Further studies could be done on those who have not completed a level of education (i.e. did not go to university/college, etc.) to see how this affects voting patterns. There was also missing values that were removed from the data, impacting the analyses on our sampling population. Some studies have also stated that higher education can also result in economically right-wing views (Scott, 2022). This suggest that there needs to be a consideration of other factors that influences political opinions.

Overall, our research generated valuable insight between voting trends in Canadian voters and demographic variables such as education and age. Future research and studies could be made to explore the structure of the Canadian educational system that may explain the trend of Liberal voting, or study other variables that may add to the likelihood of voting Liberal.

6 Generative AI Statement

- Generative AI was used to clarify function usage, for formatting inputs in R functions and graphs and creating tables in Markdown (ChatGPT 4o).

- We utilized generative AI to reformulate sentences and improve the clarity of the report (ChatGPT 4o)

  We used generative AI solely as a supplemental tool to assist with code refinement and report writing. While AI was employed for low-level tasks such as grammar checking in code and rewording sentences, the conceptual insights and analyses in both the report discussions and the code were entirely our own.

7 Ethics Statement

The reproducibility of this study has been ensured through comprehensive documentation of the methodologies employed. This includes providing the formula for the confidence interval, describing the logistic regression model, and ensuring that all data sources are thoroughly referenced. Additionally, any data manipulations, such as cleaning and preprocessing, are well documented. Proper statistical methods like T-tests and confidence intervals have been applied to maintain the integrity of the results.

Since the CES 2019 survey data is publicly available, this study does not require Research Ethics Board approval, as it falls under the exemption for research using publicly available data protected by law. The privacy of human participants is safeguarded, as the CES 2019 dataset is anonymized. Furthermore, steps have been taken to ensure that no individual participants can be re-identified, including avoiding manipulations that could expose small sample groups to potential disclosure risks.

8 Bibliography

1. Allaire, J. J., et al. (n.d.). *Introduction to R Markdown.* RStudio. Retrieved April 4, 1991, from https://rmarkdown.rstudio.com/docs/.

2. Dekking, F. M., et al. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

3. Government of Canada, Statistics Canada. (2023, June 29). *International Standard Classification of Education (ISCED).* Government of Canada. Retrieved from www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=1275152.

4. Government of Canada, Statistics Canada. (2024, August 2). *Census profile, 2021 Census of Population profile table.* Retrieved from www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/page.cfm?LANG=E&GENDERlist=1%2C2%2C3&STATISTIClist=1%2C4&DGUIDlist=2021A000011124&HEADERlist=37&SearchText=Canada.

5. Grolemund, G. (2014, July 16). *Introduction to R Markdown.* RStudio. Retrieved April 4, 1991, from https://rmarkdown.rstudio.com/articles_intro.html.

6. Hooghe, M., Dassonneville, R., & Marien, S. (2014). The impact of education on the development of political trust: Results from a five-year panel study among late adolescents and young adults in Belgium. *Political Studies, 63*(1), 123–141. https://doi.org/10.1111/1467-9248.12102.

7. Kiss, S., Polacko, M., & Graefe, P. (2023). The education and income voting divides in Canada and their consequences for redistributive politics. *Electoral Studies, 85*, 102648. doi:10.1016/j.electstud.2023.102648.

8. Leś, Tomasz. (2024). Education, Ideology, and Critical Thinking. 10.4324/9781003386100-5.

9. Scott, R. (2022). Does university make you more liberal? estimating the within-individual effects of higher education on political values. *Electoral Studies, 77*, 102471. doi:10.1016/j.electstud.2022.102471.

9 Appendix

Any additional notes/derivations that are supplementary to the report can be added in an appendix. This section will not be directly graded, but may be included for completion-sake.