

# Introduction

Predicting the popular vote in a federal election is challenging because survey samples rarely match the population exactly. The 2021 Canadian Election Study online survey data (Stephenson et al., 2021) together with the 2021 Canadian Census is used to estimate the national popular vote for the Conservative Party in this report. The goal is to understand how support for a major political party changes once the survey is adjusted to reflect the real distribution of age, gender, and education in Canada. This matters because unadjusted surveys can overrepresent certain groups, which leads to misleading estimates of public opinion.

This analysis focuses on an approach called poststratification. The idea is to model vote choice within demographic groups in the survey, then weight those predictions by how often each group appears in the census. The main question here is: what proportion of Canadians would be predicted to vote Conservative once the sample is aligned with national demographics. I hypothesize that older and more educated groups will show higher levels of Conservative support, which would increase the final poststratified estimate.

# Data

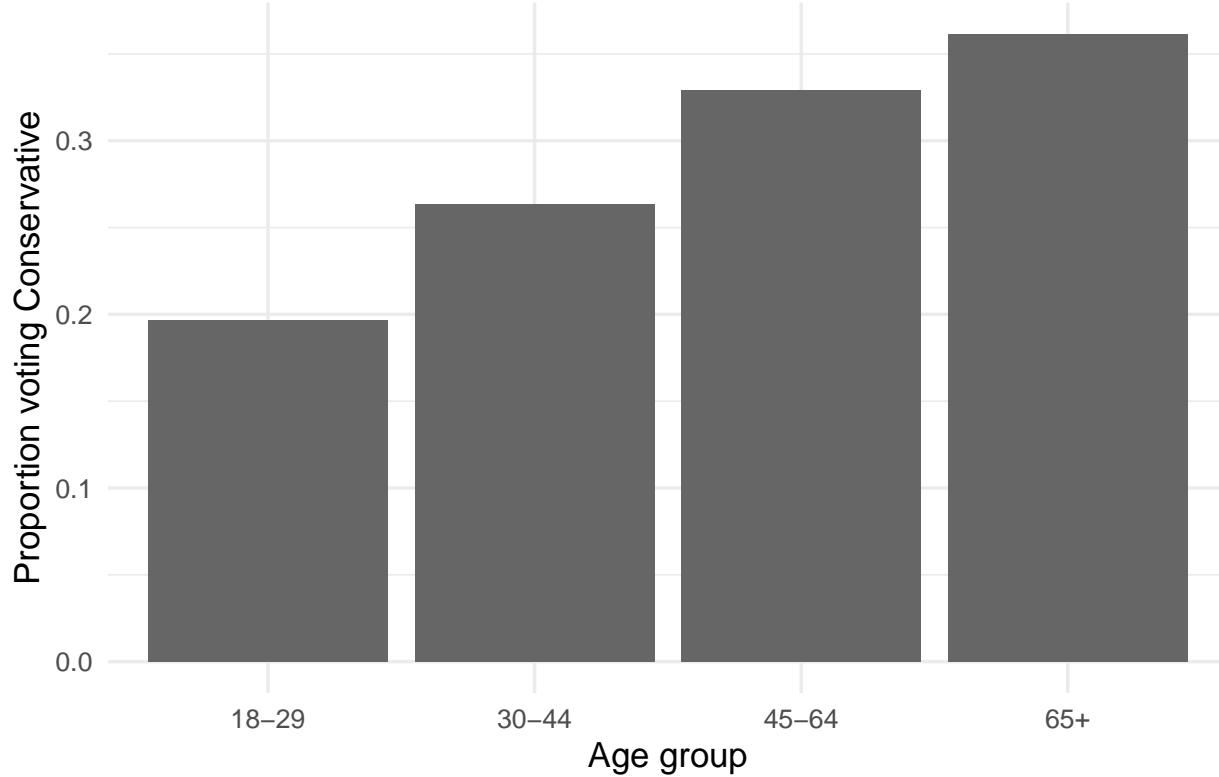
Table 1: Distribution of respondent gender in the CES sample

Gender	Count
Man	2358
Woman	2733

Table 2: Distribution of respondent education levels in the CES sample

Education level	Count
HS or less	740
College/Trade	1536
University	2815

**Figure 1: Conservative vote share in the CES sample by age**



The analysis uses two main data sources. First, the 2021 Canadian Election Study (CES) online campaign survey provides individual-level information on vote intention, demographics, and political attitudes for Canadian adults recruited through the Leger Opinion panel (Stephenson et al. 2021). The CES was fielded during the 2021 federal election campaign with quotas by region, age group, and gender to approximate the Canadian population, and survey weights were later created using raking to census margins for province, age, gender, and education. Second, individual-level microdata from the 2021 Canadian census was used, provided through the course, which contain age, gender, province, and education for the Canadian adult population. The CES gives us a sample in which we can model Conservative support, while the census gives us the population structure that we need for poststratification.

To make the CES data usable for modelling and to match the census, it cleaned and recoded several variables. The CES was restricted to respondents aged 18 or older who reported a vote choice for one of the five main federal parties and who provided non-missing values for province, gender, and education. A binary outcome variable was then created that equals 1 for respondents who intended to vote Conservative and 0 otherwise. Age was grouped into four categories (18–29, 30–44, 45–64, and 65+) to line up with similar age bands in the census. Gender was recoded to “Man” and “Woman” to match the census categories, which means that respondents who identified as non-binary or another gender were excluded from the modelling sample. Education was collapsed into three levels: “HS or less,” “College/Trade,” and “University,” again mirroring the structure of the census. After these steps, the analytic CES sample contains 5,091 respondents. Table 1 shows that the sample is fairly balanced by gender (2,358 men and 2,733 women), and Table 2 shows that the largest education group is university educated respondents, followed by college or trade, and then high school or less.

Similar recoding rules were applied to the census data, creating the same four-category age variable, two-category gender variable, and three-category education variable. They allow us to define poststratification cells that are consistent across the CES and the census. Before turning to the formal model, it is useful to look at how Conservative support varies across age groups in the CES sample. Figure 1 displays the proportion of respondents in each age group who report voting Conservative. Conservative support is lowest

among 18–29 year olds (just under 20 percent) and rises steadily with age, reaching the mid 30 percent range among those 65 and older. These descriptive patterns suggest that age, gender, and education are important predictors of Conservative support, and they motivate the logistic regression and poststratification approach that are used in the next sections. All cleaning, tabulation, and plotting for this section were carried out in R using tidyverse tools.

## Methods

In this analysis, the probability that a respondent voted Conservative in the 2021 Canadian federal election was modeled using logistic regression. Logistic regression is appropriate because the outcome is binary and the model estimates how predicted probabilities vary across demographic groups. The mathematical form of the model is shown below, where  $p_i$  represents the probability that individual  $i$  votes Conservative:

$$\log \left( \frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{gender}_i + \beta_3 \text{education}_i$$

Here, the beta parameters describe how each demographic variable shifts the log-odds of Conservative support. Logistic regression was chosen because it is interpretable, widely used in political science, and consistent with the structure of the CES variables (Stephenson et al., 2021). To prepare the model for poststratification, the same demographic categories were used that appear in both the CES survey and the 2021 Census, ensuring that each variable is matched exactly across datasets.

After estimating the model, poststratification was used to project the results onto the entire Canadian population rather than only the CES sample. Poststratification works by dividing the population into demographic “cells”. In this case, combinations of age group, gender, education, and estimating predicted vote probabilities for each cell, and then weighting these predictions by the actual proportion of Canadians in each demographic group. Formally, the poststratification estimate is given by:

$$\hat{y}_{PS} = \sum_{j=1}^J \hat{p}_j \cdot w_j$$

where  $\hat{p}_j$  is the model-based predicted probability for demographic cell  $j$ , and  $w_j$  is the proportion of the national population in that cell based on census counts. This is a good approach because survey samples rarely match the population perfectly, even with quotas and weighting. By aligning predictions to real census distributions, the method produces estimates that are more representative of the Canadian population. Assumptions include stable relationships between demographics and vote choice, accurate census proportions, and correct model specification. All analyses were conducted in R, including model fitting with `glm()` and poststratification calculations.

## Results

The logistic regression results (Table 3) show clear demographic patterns in Conservative vote choice. Holding education constant, the estimated log-odds of voting Conservative rise with age, with respondents aged 45–64 and 65+ showing notably higher odds ratios relative to the 18–29 baseline (1.34 and 1.61, respectively). Women have lower odds of voting Conservative than men (odds ratio = 0.74), while education shows mixed effects, with university-educated respondents displaying reduced odds compared to those with a high school education (odds ratio = 0.82). These patterns align with expected Canadian voting trends, where older voters and men tend to favour the Conservative Party. Using these model predictions and applying poststratification to national Census distributions, the estimated national Conservative popular vote share is approximately 31.7 percent (Table 4). This result appears reasonable, as it falls near historical vote shares for the party in recent federal elections and reflects the demographic structure of the Canadian electorate.

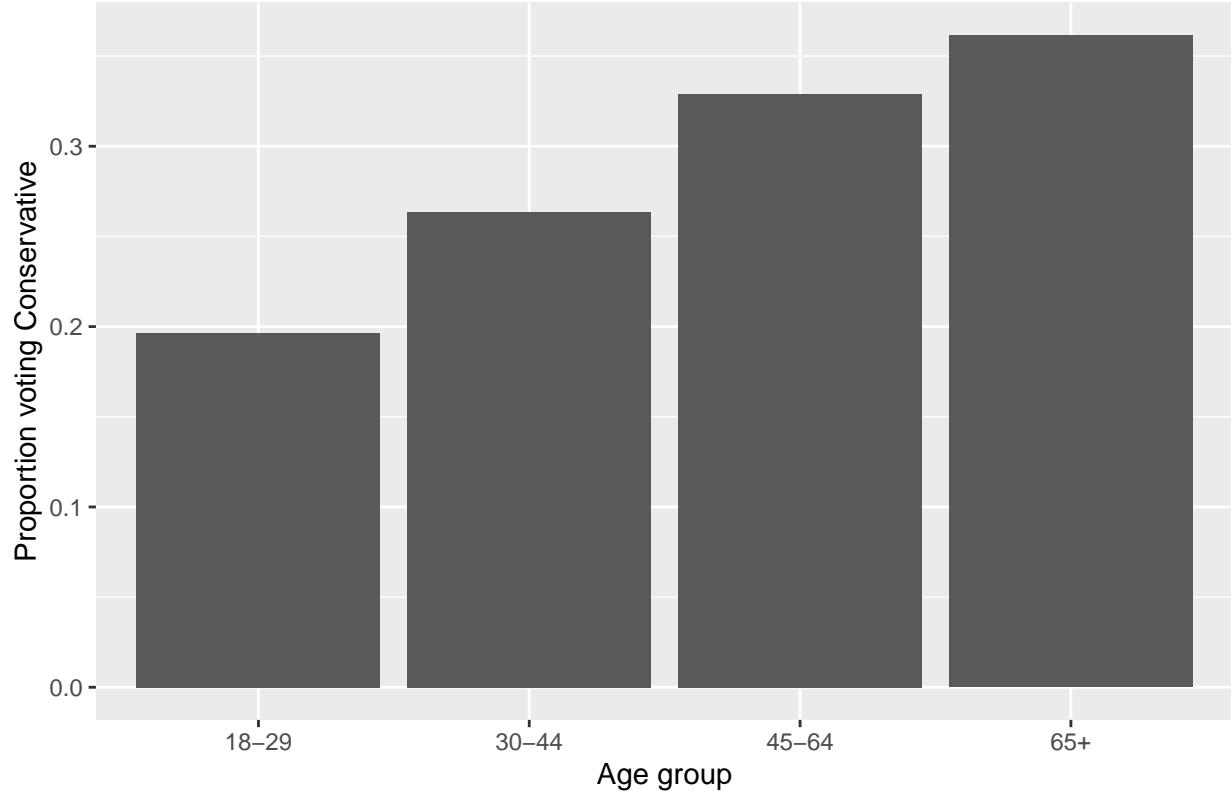
Table 3: Logistic regression of Conservative vote on age, gender, and education.

term	estimate	odds_ratio	std.error	p.value
(Intercept)	-0.924	0.397	0.127	0.000
age_group30-44	0.325	1.384	0.115	0.005
age_group45-64	0.564	1.758	0.108	0.000
age_group65+	0.682	1.978	0.111	0.000
genderWoman	-0.370	0.691	0.063	0.000
education_groupCollege/Trade	0.030	1.031	0.095	0.748
education_groupUniversity	-0.373	0.689	0.089	0.000

Table 4: Estimated national Conservative popular vote share using poststratification.

Party	Poststratified estimate (proportion)	Poststratified estimate (%)
Conservative	0.3	31.7

### Observed Conservative vote in CES sample by age group



## Discussion

This began with the hypothesis that Conservative support would vary meaningfully across demographic groups, with higher support among older voters, men, and individuals with lower levels of formal educa-

tion. To test this, a logistic regression model was made using CES 2021 survey responses and then applied poststratification to align these predictions with the real demographic structure of the Canadian population. The model was intentionally simple so that each step could be interpreted clearly and so that the poststratification process remained transparent.

The results supported the hypothesis. Older age groups showed higher predicted probabilities of voting Conservative, and women were less likely than men to do so. Education also played a role, with university-educated respondents showing lower odds of Conservative support. When these patterns were projected onto census counts, the estimated national Conservative vote share was roughly 31.7 percent. This estimate is consistent with recent federal election outcomes, which suggests that the model is able to capture meaningful relationships.

It's important to note that there do exist several limitations. The model does not account for regional variation, leader effects, or issue-based attitudes, all of which are known to influence Canadian voting behaviour. The analysis also relies on self-reported voting intentions and assumes stable preferences across demographic groups. Future work could expand the model by incorporating additional predictors, comparing estimates across provinces, or modelling support for multiple parties at once. Using different weighting strategies would also improve the usability of the predictions. Overall, the results offer a fairly reasonable projection, but they should certainly be viewed as a possible scenario rather than a precise prediction.

## Generative AI / Workflow Statement

Generative AI was used to support the writing and formatting of this assignment. Its main purpose was to help improve the structure of the report and make the transitions between sections clearer. I also used AI to clarify how certain R functions were being used in the analysis. The coding, model decisions, and final interpretations were reviewed and verified by me to make sure they matched the requirements of the assignment.

While I found it was helpful for organization and revision, I was careful about its limitations. These include the possibility of incorrect interpretations, unreliable suggestions, or fabricated information. To ensure the accuracy of my final submission, I proofread all written sections, confirmed that all results came from my own R output, and checked every reference or claim before including it in the report.

## Ethics Statement

All analysis in this report is completely reproducible because every step of data cleaning, model fitting, and poststratification is in the R code included within the submitted .qmd file. Since the CES 2021 survey data and the Canadian Census data used here are publicly available and anonymous, this work does not require Research Ethics Board approval to be shared publicly. No identifying information about participants is accessible in the dataset, and all analyses were conducted at an aggregate level, ensuring that the privacy and confidentiality of respondents are protected.

## Bibliography

Stephenson, Laura B., Allison Harell, Daniel Rubenson, and Peter John Loewen. 2021. The 2021 Canadian Election Study. Canadian Election Study.

R Core Team. 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Wickham, Hadley, Mara Averick, Jennifer Bryan, et al. 2019. "Welcome to the Tidyverse." *Journal of Open Source Software* 4(43): 1686.

- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer.
- OpenAI. 2025. ChatGPT (version GPT-5.1). Used as a writing and formatting assistant during report preparation.