The Changing Effect of Class and Religion on Voting Behavior in the Canadian Federal Election

Jessica Long

2023-03-21

Abstract

Using data from the Canada Election Study (CES) survey for the 2004 and 2021 Canadian federal elections, I analyzed the influence of class (through union membership and education level) and religion on voter preference between the Liberal, Conservative, and NDP parties. In particular, I investigated if there was a shift in the party preference of the different groups under consideration between 2004 and 2021. I used a weighted multinomial logistic regression model with interaction terms to look at the effects of membership in each group, and the difference in effect between the years. I found that union members prefer the Liberals to the Conservatives, and the NDP to the Liberals. There is some preference for the Conservatives over the Liberals among non-university educated voters. Catholics generally prefer the Liberal party, but there was a notable increase in support for the Conservatives in 2021. Both mainline and non-mainline Protestants prefer the Conservative party over the Liberals, with the non-mainline Protestants showing stronger support.

Introducation

My project explores the impact of personal background characteristics on vote choice in the 2004 and 2021 Canadian federal elections, and if the effect of particular characteristics have changed over time. In particular, I looked at education level (as a proxy for economic class), union membership and religion. Understanding how these factors influence people's votes and how the influence changed with time may be useful for predicting election results or trying to gain support for a party. More generally, it would give insight into Canadian society and how different groups of people behave.

In Canada, voters vote for a Member of Parliament for their riding, and the party with the most MPs takes power. I will be focusing my analysis on the three largest parties: Conservative Party (center right), Liberal Party (center left), and New Democratic Party/NDP (left-wing). The other parties were excluded for receiving too few votes, and additionally some were created after 2004. I have also excluded the province of Quebec from the analysis as politics in that province are notably different from the rest of the country. The issue of Quebec sovereignty is a key factor in elections, and the local sovereignist party, the Bloc Quebecois, generally wins a fair share of the MP seats in the province.

I considered several hypotheses regarding the influence of education level, union membership, and religion on vote choice between the three major parties. From the analysis, there is a preference for the Liberal Party amongst Catholic voters, though Catholics have become more likely to vote for the Conservatives as of 2021. Protestant voters are more likely to vote Conservative compared to Liberal, and non-mainline Protestants have a stronger preference for Conservatives than the mainline denominations. Union members prefer the NDP to the Liberals, and the Liberals to the Conservatives. There is some indication that non-university educated voters are more likely to vote Conservative compared to Liberal. However, several of the effects of the considered features were not significant at the 0.05 level.

Hypotheses

The first area I focused on is voting preferences by class. Traditionally, the NDP has relied on working-class voters as its primary base, and has had close connections with unions. However, in recent years the Conservative Party has been gaining support from working-class voters concerned with traditional values and immigration. Meanwhile, working class support for the Liberal party has declined (Polacko, Kiss, and Graefe 2022). Additionally, the relationship between the NDP and unions has declined, due to a federal ban on union (and corporate) donations to parties in 2003 (in favor of more state funding), and some unions embracing strategic voting by supporting Liberal candidates in order to keep out the Conservatives (Savage 2010). However, the NDP still enjoys support from its traditional voting base. Based on these observations, I expected some increase in the probability of respondents without university education voting for Conservatives compared to for the Liberal party. I also expected an increase in the probability of union members voting for Liberals, and some decrease in the probability of voting NDP.

The second area of interest was religion. Traditionally, there is a prominent Protestant/Catholic divide in Canadian politics and culture, which often overlaps with the English/French divide. Catholic voters have traditionally supported the Liberal party, while Protestants tended to vote for the Conservatives more (Stephenson 2009). However, the Catholic-Liberal relationship has been weakening in more recent years. One possible explanation is that Catholic voters are now concerned more about traditional values vs secular ones, as opposed to the old Catholic vs Protestant conflict (Rayside, Sabin, and Thomas 2017). So I expected a decrease in the probability of Catholics voting for Liberals, and an increase in the probability of voting Conservatives. From the Protestant side, there has been a decline in the preference for Conservatives from mainline Protestants (Anglican, Lutheran, Presbyterian, and United), but not from other Protestant groups. One possible reason is that levels of religiosity have declined for mainline Protestants, resulting in them prioritizing traditional values less when voting (Wilkins-Laflamme 2016). They may also have moved away from the historic sectarian conflict, as with the Catholics. So I expected a decrease in the probability of mainline Protestants voting Conservative.

Data and Methodology

The data I used was from the Canada Election Study (CES), for the years 2021 (**ces_2021?**) and 2004 (Blais et al. 2005). The dataset includes information gathered from campaign-period and post-election surveys. The 2004 surveys were conducted via telephone, while the 2021 survey was conducted online. Some survey responses were grouped together as I was more interested in the effect of the more general category. The education levels below secondary level were combined into one category. Note that in Canada, universities are academic institutions that grant degrees (e.g. Bachelor's, Master's), while colleges offer technical training and grant diplomas. Responses for religion were grouped together into Mainline_Protestant, Other_Protestant, Other_Christian, and Other_religion as needed. Along with the responses for union membership, education level, and religion, I also included responses for province of residence, gender, native language (either English or French), income category, marital status, and age to be used in data exploration. I then combined the datasets and added an indicator variable for the year.

I decided to use a multinomial logistic regression model, with the response variable being the party voted for between the three major parties. I chose this model as it is able to estimate probabilities while providing a way to assess the influence of particular features. I investigated the demographic distribution in the two datasets to see if weights should be used for this model, in order to have the data be more in line with the voting population. Using census data from 2006 and 2021 (Canada 2022), I computed the proportions in the Canadian population (minus Quebec) for gender, age group, education level, and combinations of the three. For age group, I looked at 20-24, 25-65 and 65+ as that was what was convenient to use based on the census data. In the survey data, I included those less than 24 years old in the first category (the Canadian voting age is 18). Performing t-tests on various categories, I determined that there was a significant difference in the survey and census data for gender and at least some of the age groups and education levels. I decided to not account for union membership and religion in the weighting as there wasn't census data accounting for those categories and education together, making it hard to determine the proper proportions in the population.

For the final weights, I used the proportion in the census data divided by the proportion in the survey data for most of the education levels. These proportions were accounting for year, age group, and gender. The exceptions were Completed_secondary, Some_college, and Some_university. This was because the census did not account for people who had some post-secondary education but not all, so I thought those people would've been included under the completed secondary category in the census. For these, I found the ratio of their proportion in the survey data to the sum of their proportions. I then multiplied this ratio by the leftover proportion in the census data not accounted for by the other education levels to estimate their proportion in the population, then computed the weight as with the other categories. I also further filtered the data to remove other gender responses (in 2021) and no-answer responses for education as they were not accounted for in this weighting scheme.

Results

I fitted using the multinom function in the nnet R package, with the weights described previously. For the features, I used the dummy variables for union membership, education level, religion, and year, as well as the interaction of year with the other variables. Liberal was used as the base response as they were the winning party in both years.

The coefficients, p-values, and significance (at 0.05 level) are as follows:

```
print(results, nrow = 100)
```

```
##                                                Conservative          NDP
## (Intercept)                                     0.264818402 -0.462191939
## year                                           -0.299086537  0.011298062
## union                                          -0.402776140  0.435122528
## education_Bachelors_degree                     -0.200609093 -0.335448003
## education_Completed_college                     0.232392690 -0.046299400
## education_Completed_secondary                   0.158176878 -0.188333998
## education_Less_than_secondary                   0.122334510  0.186310181
## education_Masters_degree                       -0.349737719 -0.225657897
## education_Professional_degree_or_doctorate      0.172864637  0.288184629
## education_Some_college                          0.439935843  0.136643793
## education_Some_university                      -0.310539344 -0.277591244
## religion_Catholic                              -0.676071632 -0.498796132
## religion_Mainline_Protestant                    0.009038011 -0.233663702
## religion_No_answer                              0.438933327 -0.232229443
## religion_Non_religious                         -0.129504237  0.520512959
## religion_Other_Christian                        0.755020194  0.355415416
## religion_Other_Protestant                       0.694418800  0.117125428
## religion_Other_religion                        -0.827016061 -0.490556464
## year:union                                      0.129387592 -0.204245397
## year:education_Bachelors_degree                 0.073922611 -0.182894702
## year:education_Completed_college                0.021126212  0.106934149
## year:education_Completed_secondary              0.137759900  0.386492373
## year:education_Less_than_secondary             -0.165769417  0.261032599
## year:education_Masters_degree                   0.009540797 -0.209145726
## year:education_Professional_degree_or_doctorate -0.424898853 -0.775816234
## year:education_Some_college                    -0.286474940 -0.055041220
## year:education_Some_university                  0.335707153  0.479736822
## year:religion_Catholic                          0.514701115 -0.017539985
## year:religion_Mainline_Protestant               0.165555096  0.001174235
## year:religion_No_answer                        -0.670698285 -0.009747111
```

```
## year:religion_Non_religious                      -0.056750347  0.062988045
## year:religion_Other_Christian                    -0.560684833 -0.454907315
## year:religion_Other_Protestant                    0.029807681 -0.140742617
## year:religion_Other_religion                      0.278983036  0.570072810
##                                                 Conservative_pval     NDP_pval
## (Intercept)                                          2.423805e-03 4.241013e-05
## year                                                 1.481820e-03 9.251432e-01
## union                                                7.740433e-03 6.077646e-03
## education_Bachelors_degree                           1.883953e-01 5.672478e-02
## education_Completed_college                          4.362886e-02 7.345067e-01
## education_Completed_secondary                        2.401370e-01 2.532375e-01
## education_Less_than_secondary                        3.540465e-01 2.181775e-01
## education_Masters_degree                             2.034314e-01 4.466504e-01
## education_Professional_degree_or_doctorate           5.521216e-01 3.571465e-01
## education_Some_college                               3.653126e-02 5.845787e-01
## education_Some_university                            1.105558e-01 2.182029e-01
## religion_Catholic                                    1.432879e-07 1.889148e-03
## religion_Mainline_Protestant                         9.428409e-01 1.555900e-01
## religion_No_answer                                   3.484039e-01 7.225577e-01
## religion_Non_religious                               3.549308e-01 1.147143e-03
## religion_Other_Christian                             4.305961e-04 1.892139e-01
## religion_Other_Protestant                            3.681439e-05 5.970674e-01
## religion_Other_religion                              6.521039e-04 8.312294e-02
## year:union                                           4.463833e-01 2.486977e-01
## year:education_Bachelors_degree                      6.502979e-01 3.316450e-01
## year:education_Completed_college                     8.673794e-01 4.710094e-01
## year:education_Completed_secondary                   3.708687e-01 3.636825e-02
## year:education_Less_than_secondary                   2.736958e-01 1.188983e-01
## year:education_Masters_degree                        9.738570e-01 5.054863e-01
## year:education_Professional_degree_or_doctorate      1.786805e-01 2.320213e-02
## year:education_Some_college                          2.101391e-01 8.372600e-01
## year:education_Some_university                       1.111078e-01 4.540807e-02
## year:religion_Catholic                               2.696716e-04 9.207930e-01
## year:religion_Mainline_Protestant                    2.423696e-01 9.948856e-01
## year:religion_No_answer                              1.632274e-01 9.883218e-01
## year:religion_Non_religious                          7.055827e-01 7.107142e-01
## year:religion_Other_Christian                        4.272024e-02 1.902084e-01
## year:religion_Other_Protestant                       8.709456e-01 5.583576e-01
## year:religion_Other_religion                         2.762613e-01 5.290578e-02
##                                                 Conservative_sig NDP_sig
## (Intercept)                                                 TRUE    TRUE
## year                                                        TRUE   FALSE
## union                                                       TRUE    TRUE
## education_Bachelors_degree                                 FALSE   FALSE
## education_Completed_college                                 TRUE   FALSE
## education_Completed_secondary                              FALSE   FALSE
## education_Less_than_secondary                              FALSE   FALSE
## education_Masters_degree                                   FALSE   FALSE
## education_Professional_degree_or_doctorate                 FALSE   FALSE
## education_Some_college                                      TRUE   FALSE
## education_Some_university                                  FALSE   FALSE
## religion_Catholic                                           TRUE    TRUE
## religion_Mainline_Protestant                               FALSE   FALSE
## religion_No_answer                                         FALSE   FALSE
```

```
## religion_Non_religious                                  FALSE    TRUE
## religion_Other_Christian                                 TRUE   FALSE
## religion_Other_Protestant                                TRUE   FALSE
## religion_Other_religion                                  TRUE   FALSE
## year:union                                              FALSE   FALSE
## year:education_Bachelors_degree                         FALSE   FALSE
## year:education_Completed_college                        FALSE   FALSE
## year:education_Completed_secondary                      FALSE    TRUE
## year:education_Less_than_secondary                      FALSE   FALSE
## year:education_Masters_degree                           FALSE   FALSE
## year:education_Professional_degree_or_doctorate         FALSE    TRUE
## year:education_Some_college                             FALSE   FALSE
## year:education_Some_university                          FALSE    TRUE
## year:religion_Catholic                                   TRUE   FALSE
## year:religion_Mainline_Protestant                       FALSE   FALSE
## year:religion_No_answer                                 FALSE   FALSE
## year:religion_Non_religious                             FALSE   FALSE
## year:religion_Other_Christian                            TRUE   FALSE
## year:religion_Other_Protestant                          FALSE   FALSE
## year:religion_Other_religion                            FALSE   FALSE
```

By adding up the individual coefficient and the corresponding interaction, we also get the impact of each variable on the log odds in the year 2021:

```
print(effects_2021, nrow = 100)
```

```
## # A tibble: 16 x 3
##    variable                                Conservative     NDP
##    <chr>                                          <dbl>   <dbl>
##  1 union                                         -0.273   0.231
##  2 education_Bachelors_degree                    -0.127  -0.518
##  3 education_Completed_college                    0.254   0.0606
##  4 education_Completed_secondary                  0.296   0.198
##  5 education_Less_than_secondary                 -0.0434  0.447
##  6 education_Masters_degree                      -0.340  -0.435
##  7 education_Professional_degree_or_doctorate    -0.252  -0.488
##  8 education_Some_college                         0.153   0.0816
##  9 education_Some_university                      0.0252  0.202
## 10 religion_Catholic                             -0.161  -0.516
## 11 religion_Mainline_Protestant                   0.175  -0.232
## 12 religion_No_answer                            -0.232  -0.242
## 13 religion_Non_religious                        -0.186   0.584
## 14 religion_Other_Christian                       0.194  -0.0995
## 15 religion_Other_Protestant                      0.724  -0.0236
## 16 religion_Other_religion                       -0.548   0.0795
```

First we look at the individual coefficients, which correspond to the effect on the log odds in 2004. Union membership has a negative effect on log odds for voting Conservative, and a positive effect for voting NDP, which both these effects being significant. These match up with the expectations for voting behavior of union members from before. Most of the effects for education level are not significant, aside from the positive effect of having college education on the log odds of voting Conservative. For the non-significant effects, the coefficients for having less than secondary education are positive for Conservative. For NDP, the effect is positive for some college and less than secondary, but negative for completed college and secondary. This suggests

working class voters are more likely to vote Conservative than Liberal in 2004. The relative preference for NDP is not as strong as expected. Lastly, we look at the religious effects. The coefficients for both parties are negative, and this effect is significant, which agrees with the expected Catholic preference for the Liberal Party. The effect on log odds for Conservative for other Protestants is positive and significant. The corresponding effect for mainline Protestants is negative and not significant. This matches our assumption that non-mainline Protestants have a stronger preference for the Conservative Party than mainline Protestants.

Next we will look at the interaction terms, which will indicate how the effect of a feature has changed from 2004 to 2021, as well as the overall effect on log odds in 2021. Both coefficients for union membership are not significant. We see a negative effect for the interaction term for NDP, but the overall effect on the log odds for voting NDP relative to Liberal in 2021 is positive, as predicted. From the relevant education coefficients, the only significant interaction is the term for completed secondary and voting NDP, which is positive. For the other interaction coefficients for NDP, completed college and less than secondary are positive while some college is negative. For the Conservative interaction coefficients, completed college and secondary are positive, while less than secondary and some college are negative. Overall, the log odds for these four education levels are positive, except for less than secondary and Conservative. This suggests voting class workers are more likely to vote for these two parties compared to the Liberals in 2021, although the relative change in preference from 2004 is varied. For religion, the Catholic interaction term for voting Conservative is positive and significant, as expected. However, the overall effect on log odds is still negative for the non-Liberal parties, suggesting that Catholic voters overall still prefer the Liberal party. The corresponding terms for mainline and other Protestants is positive, but not significant. The interaction coefficient for mainline Protestant is larger, but the overall effect of other Protestant on voting Conservative is larger.

Conclusion

In 2004, union members were more likely to vote Liberal than Conservative, and more likely to vote NDP than Liberal. In 2021, the relative probability of union members voting NDP over Liberal has dropped, while the relative probability of union members voting Conservative over Liberal has increased. However, neither of these effects was significant at the 0.05 level. While I had hypothesized the change in probability for NDP vs Liberal, I did not anticipate the effect for Conservative vs Liberal, as many unions oppose the Conservative Party. However, it could be that individual union members do not always vote in accordance with the unions' strategies, and have other reasons for voting Conservative.

In this analysis, I used education level as a proxy for class, with the non-university levels representing the working class. This is imperfect as colleges offer diplomas for more white-collar/professional fields also. This may be why the education effects are varied and often not significant at the 0.05 level. In 2004, there is some indication that non-university educated voters are more likely to vote Conservative than Liberal. This lines up with the result from Polacko, Kiss, and Graefe (2022) that the working class shift to the Conservative party was underway in 2004. In 2021, this continued with the working class generally preferring both the Conservative and NDP parties to the Liberal party. However, the change in relative probability from 2004 differed between the various non-university education levels.

In 2004, the Catholic preference for the Liberal party over the other two is significant at the 0.05 level. The increase in the probability of Catholics voting for Conservatives is also significant at that level, though overall the Catholics still prefer the Liberal party. For Protestants, both mainline and other denominations prefer the Conservative party over the Liberals in 2004, with non-mainline Protestants having a stronger, and significant preference. In 2021, both types of Protestants increased their preference for the Conservative party, the mainline denominations more so, although this effect is not significant. This runs counter to my hypotheses that the non-mainline denominations would be demonstrating increased preference for the Conservatives while the mainline preference would decrease. It could be that combining several denominations into these two categories has obscured some of the effects of the specific Protestant denominations. Wilkins-Laflamme (2016) also accounted for religiosity/degree of importance religion had in the respondent's life. This additional variable may capture more information about the effect of religion on voting.

A future analysis could include data collected for other federal elections, which would give a more detailed view of trends throughout time. I could look at the effects on voting of the university education levels, or

other religions aside from Catholic and Protestant Christianity. It may be useful to look at trends at a provincial or regional level, since Canada traditionally has had regional preferences for different parties. I could also do a separate analysis on Quebec, given that Catholicism is traditionally a key part of Quebec society, and that unions in Quebec tend to support the sovereignist Bloc Quebecois (2). Further analysis could be used to make a simple model for predicting voting trends in a future federal election.

References

Blais, André, Joanna Everitt, Patrick Fournier, and Gidengil. 2005. "2004 Canadian Election Study (CES)." Institute for Social Research.

Canada, Statistics. 2022. "Table 98-10-0384-01 Highest Level of Education by Census Year: Canada, Provinces and Territories, Census Metropolitan Areas and Census Agglomerations." https://doi.org/https://doi.org/10.25318/9810038401-eng.

Polacko, Matthew, Simon Kiss, and Peter Graefe. 2022. "The Changing Nature of Class Voting in Canada, 1965–2019." *Canadian Journal of Political Science/Revue Canadienne de Science Politique* 55 (3): 663–86. https://doi.org/10.1017/S0008423922000439.

Rayside, David, Jerald Sabin, and Paul E. J. Thomas. 2017. *Religion and Canadian Party Politics*. UBC Press.

Savage, Larry. 2010. "Contemporary Party-Union Relations in Canada." *Labor Studies Journal* 35 (March): 8–26. https://doi.org/10.1177/0160449X09353028.

Stephenson, Laura. 2009. "The Catholic-Liberal Connection: Untangling the Specifics." *APSA 2009 Toronto Meeting Paper*.

Wilkins-Laflamme, Sarah. 2016. "The Changing Religious Cleavage in Canadians' Voting Behaviour." *Canadian Journal of Political Science / Revue Canadienne de Science Politique* 49 (3): 499–518. http://www.jstor.org/stable/26291381.

Appendix

Data Cleaning

```
library(tidyverse)
library(haven)
library(fastDummies)
library(nnet)
library(DescTools)
library(janitor)
library(sjmisc)
library(cesR)

data_raw <- read_dta('2021 Canadian Election Study v1.0.dta')

data_selected <- data_raw %>%
  select(cps21_genderid,
         cps21_province,
         cps21_education,
         cps21_religion,
         cps21_language_1, # English as native language
         cps21_language_2, # French as native language
         cps21_union,
         cps21_income_cat,
         cps21_marital,
         cps21_yob, # Year of birth
```

```
        pes21_votechoice2021)

# Remove Quebec, keep only votes for Liberal, Conservative, NDP
data_filtered <- data_selected %>%
  filter(cps21_province != 11, pes21_votechoice2021 %in% c(1, 2, 3))

data_cleaned <- data_filtered %>%
  mutate(province = case_when(
    cps21_province == 1 ~ "Alberta",
    cps21_province == 2 ~ "British_Columbia",
    cps21_province == 3 ~ "Manitoba",
    cps21_province == 4 ~ "New_Brunswick",
    cps21_province == 5 ~ "Newfoundland_and_Labrador",
    cps21_province == 6 ~ "Northwest_Territories",
    cps21_province == 7 ~ "Nova_Scotia",
    cps21_province == 8 ~ "Nunavut",
    cps21_province == 9 ~ "Ontario",
    cps21_province == 10 ~ "Prince_Edward_Island",
    cps21_province == 12 ~ "Saskatchewan",
    cps21_province == 13 ~ "Yukon")) %>%
  select(-cps21_province) %>%
  mutate(gender = case_when(
    cps21_genderid == 1 ~ "male",
    cps21_genderid == 2 ~ "female",
    !(cps21_genderid %in% c(1:2)) ~ "other")) %>%
  filter(gender != "other") %>%
  select(-cps21_genderid) %>%
  mutate(education = case_when(
    cps21_education %in% c(1:4) ~ "Less_than_secondary",
    cps21_education == 5 ~ "Completed_secondary",
    cps21_education == 6 ~ "Some_college",
    cps21_education == 7 ~ "Completed_college",
    cps21_education == 8 ~ "Some_university",
    cps21_education == 9 ~ "Bachelors_degree",
    cps21_education == 10 ~ "Masters_degree",
    cps21_education == 11 ~ "Professional_degree_or_doctorate",
    !(cps21_education %in% c(1:11)) ~ "No_answer")) %>%
  filter(education != "No_answer") %>%
  select(-cps21_education) %>%
  mutate(religion = case_when(
    cps21_religion %in% c(1:2) ~ "Non_religious",
    cps21_religion %in% c(8, 13, 16, 18) ~ "Mainline_Protestant",
    cps21_religion %in% c(9, 15, 17, 19, 20, 21) ~ "Other_Protestant",
    cps21_religion == 10 ~ "Catholic",
    cps21_religion %in% c(11, 12, 14) ~ "Other_Christian",
    cps21_religion %in% c(3, 4, 5, 6, 7, 22) ~ "Other_religion",
    !(cps21_religion %in% c(1:22)) ~ "No_answer")) %>%
  select(-cps21_religion) %>%
  rename(native_anglophone = cps21_language_1) %>%
  mutate(native_anglophone = case_when(
    native_anglophone == 1 ~ 1,
    native_anglophone != 1 ~ 0)) %>%
  rename(native_francophone = cps21_language_2) %>%
```

```
  mutate(native_francophone = case_when(
    native_francophone == 1 ~ 1,
    native_francophone != 1 ~ 0)) %>%
  mutate(union = case_when(
    cps21_union == 1 ~ 1,
    cps21_union != 1 ~ 0)) %>%
  select(-cps21_union) %>%
  # New income categories to match with 2021
  mutate(income_cat = case_when(
    cps21_income_cat %in% c(1, 2) ~ "less_than_30k",
    cps21_income_cat == 3 ~ "30_to_60k",
    cps21_income_cat == 4 ~ "60_to_90k",
    cps21_income_cat %in% c(5:8) ~ "more_than_90k",
    !(cps21_income_cat %in% c(1:8)) ~ "No_answer")) %>%
  select(-cps21_income_cat) %>%
  mutate(marital = case_when(
    cps21_marital == 1 ~ "Married",
    cps21_marital == 2 ~ "Living_w_partner",
    cps21_marital == 3 ~ "Divorced",
    cps21_marital == 4 ~ "Seperated",
    cps21_marital == 5 ~ "Widowed",
    cps21_marital == 6 ~ "Never_married",
    !(cps21_marital %in% c(1:6)) ~ "No_answer")) %>%
  select(-cps21_marital) %>%
  rename(age = cps21_yob) %>%
  mutate(age_group = case_when(
    age < 25 ~ 1,
    age %in% c(25:64) ~ 2,
    age >= 65 ~ 3)) %>%
  mutate(vote_choice = case_when(
    pes21_votechoice2021 == 1 ~ "Liberal",
    pes21_votechoice2021 == 2 ~ "Conservative",
    pes21_votechoice2021 == 3 ~ "NDP")) %>%
  select(-pes21_votechoice2021) %>%
  mutate(year = 2021) %>%
  replace(is.na(.), 0)

data_final <- data_cleaned %>%
  select("vote_choice", "union", "education", "religion", "year") %>%
  # Create dummy variables and remove original columns
 dummy_cols(select_columns = c("education", "religion")) %>%
  select(-c("education", "religion"))

# Create dummy variables for all features, for additional analysis
data_final_full <- data_cleaned %>%
  dummy_cols(select_columns = c("gender", "education", "religion", "province", "marital",
                                "income_cat", "age_group")) %>%
  select(-c("gender", "education", "religion", "province", "marital",
            "income_cat", "age_group"))

# Load 2004 CES data
get_ces("ces2004")
```

```r
data_selected_2004 <- ces2004 %>%
  select(cps_rgen, # Gender
         province,
         cps_s3, # Education level
         cps_s9, # Religion
         cps_s17, # First language
         cps_s6a, # Belonging to a union
         cps_s18, # Total household income
         cps_s2, # Marital status
         cps_s1, # Year of birth
         pes_a3_3) # Which party voted for
colnames(data_selected_2004) <-  c('gender', 'province', 'education', 'religion',
                                   'first_language', 'union','income', 'marital',
                                   'birth_year', 'vote_choice')

# Remove Quebec, keep only votes for Liberal, Conservative, NDP
data_filtered_2004 <- data_selected_2004 %>%
  filter(province != 24, vote_choice %in% c(1, 2, 3))

data_cleaned_2004 <- data_filtered_2004 %>%
  mutate(province = case_when(
    province == 48 ~ "Alberta",
    province == 59 ~ "British_Columbia",
    province == 46 ~ "Manitoba",
    province == 13 ~ "New_Brunswick",
    province == 10 ~ "Newfoundland_and_Labrador",
    province == 61 ~ "Northwest Territories",
    province == 12 ~ "Nova_Scotia",
    province == 62 ~ "Nunavut",
    province == 35 ~ "Ontario",
    province == 11 ~ "Prince_Edward_Island",
    province == 47 ~ "Saskatchewan",
    province == 60 ~ "Yukon")) %>%
  # Replace numerical value with label (male/female)
  mutate(gender=as.character(as_factor(gender))) %>%
  mutate(education = case_when(
    education %in% c(1:4) ~ "Less_than_secondary",
    education == 5 ~ "Completed_secondary",
    education == 6 ~ "Some_college",
    education == 7 ~ "Completed_college",
    education == 8 ~ "Some_university",
    education == 9 ~ "Bachelors_degree",
    education == 10 ~ "Masters_degree",
    education == 11 ~ "Professional_degree_or_doctorate",
    !(education %in% c(1:11)) ~ "No_answer")) %>%
  filter(education != "No_answer") %>%
  mutate(religion = case_when(
    religion %in% c(0, 98) ~ "Non_religious",
    religion %in% c(1, 9, 13, 16) ~ "Mainline_Protestant",
    religion %in% c(2, 12, 14, 18, 19, 20) ~ "Other_Protestant",
    religion == 4 ~ "Catholic",
    religion %in% c(5, 7, 10, 17) ~ "Other_Christian",
    religion %in% c(3, 6, 7, 8, 11, 15, 97) ~ "Other_religion",
```

```
    !(religion %in% c(0:20, 97, 98)) ~ "No_answer")) %>%
  mutate(native_anglophone = case_when(
    first_language == 1 ~ 1,
    first_language != 1 ~ 0)) %>%
  mutate(native_francophone = case_when(
    first_language == 5 ~ 1,
    first_language != 5 ~ 0)) %>%
  select(-first_language) %>%
  mutate(union = case_when(
    union == 1 ~ 1,
    union != 1 ~ 0)) %>%
  # New income categories to match with 2021
  mutate(income_cat = case_when(
    income %in% c(1, 2) ~ "less_than_30k",
    income %in% c(3:5) ~ "30_to_60k",
    income %in% c(6:8) ~ "60_to_90k",
    income %in% c(9, 10) ~ "more_than_90k",
    !(income %in% c(1:10)) ~ "No_answer")) %>%
  select(-income) %>%
  mutate(marital = case_when(
    marital == 1 ~ "Married",
    marital == 2 ~ "Living_w_partner",
    marital == 3 ~ "Divorced",
    marital == 4 ~ "Seperated",
    marital == 5 ~ "Widowed",
    marital == 6 ~ "Never_married",
    !(marital %in% c(1:6)) ~ "No_answer")) %>%
  mutate(age = 2004 - birth_year) %>%
  select(-birth_year) %>%
  mutate(age_group = case_when(
    age < 25 ~ 1,
    age %in% c(25:64) ~ 2,
    age >= 65 ~ 3)) %>%
  filter(!is.na(age_group)) %>%
  mutate(vote_choice = case_when(
    vote_choice == 1 ~ "Liberal",
    vote_choice == 2 ~ "Conservative",
    vote_choice == 3 ~ "NDP")) %>%
  mutate(year = 2004) %>%
  replace(is.na(.), 0)

data_final_2004 <- data_cleaned_2004 %>%
  select("vote_choice", "union", "education", "religion", "year") %>%
  # Create dummy variables and remove original columns
  dummy_cols(select_columns = c("education", "religion")) %>%
  select(-c("education", "religion"))

# Create dummy variables for all features, for additional analysis
data_final_full_2004 <- data_cleaned_2004 %>%
  dummy_cols(select_columns = c("gender", "education", "religion", "province", "marital",
                                "income_cat", "age_group")) %>%
  select(-c("gender", "education", "religion", "province", "marital",
            "income_cat", "age_group"))
```

Data Exploration

```
# Want to find if data is not representative of overall voting population
# proportionally
# Proportions in population taken from census data in 2006 and 2021
# Proportions were calculated for the population of Canada minus Quebec
# Computations were done in Excel

n_2021 <- nrow(data_cleaned)
n_2004 <- nrow(data_cleaned_2004)

table(data_final$union)/n_2021
```

```
##
##         0         1
## 0.8330898 0.1669102
```

```
table(data_final_2004$union)/n_2004
```

```
##
##         0         1
## 0.8126424 0.1873576
```

```
union_prop_2004 <- 4005.4*1000/23384735 # Number of union workers/eligible voters, in Canada
union_prop_2021 <- 4696.7*1000/28332715
t.test(data_final$union, mu = union_prop_2004)
```

```
##
##  One Sample t-test
##
## data:  data_final$union
## t = -1.0627, df = 8213, p-value = 0.288
## alternative hypothesis: true mean is not equal to 0.1712827
## 95 percent confidence interval:
##  0.1588443 0.1749760
## sample estimates:
## mean of x
## 0.1669102
```

```
t.test(data_final_2004$union, mu = union_prop_2021)
```

```
##
##  One Sample t-test
##
## data:  data_final_2004$union
## t = 2.3178, df = 1755, p-value = 0.02058
## alternative hypothesis: true mean is not equal to 0.1657695
## 95 percent confidence interval:
##  0.1690895 0.2056258
## sample estimates:
## mean of x
## 0.1873576
```

```r
# Difference not significant at 0.05

table(data_cleaned$education)/n_2021
```

```
##
##                 Bachelors_degree                Completed_college
##                       0.31531532                       0.21999026
##              Completed_secondary              Less_than_secondary
##                       0.11711712                       0.01960068
##                   Masters_degree Professional_degree_or_doctorate
##                       0.10445581                       0.04821037
##                     Some_college                   Some_university
##                       0.07755052                       0.09775992
```

```r
table(data_cleaned_2004$education)/n_2004
```

```
##
##                 Bachelors_degree                Completed_college
##                       0.21355353                       0.17653759
##              Completed_secondary              Less_than_secondary
##                       0.20899772                       0.15091116
##                   Masters_degree Professional_degree_or_doctorate
##                       0.05865604                       0.02619590
##                     Some_college                   Some_university
##                       0.07687927                       0.08826879
```

```r
# Higher education seems to be overrepresented
# Test Bachelor's degree and completed secondary to check
t.test(data_final_full$education_Bachelors_degree, mu = 0.130595)
```

```
##
##  One Sample t-test
##
## data:  data_final_full$education_Bachelors_degree
## t = 36.029, df = 8213, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0.130595
## 95 percent confidence interval:
##  0.3052650 0.3253656
## sample estimates:
## mean of x
## 0.3153153
```

```r
t.test(data_final_full_2004$education_Bachelors_degree, mu = 0.196026)
```

```
##
##  One Sample t-test
##
## data:  data_final_full_2004$education_Bachelors_degree
## t = 1.7917, df = 1755, p-value = 0.07335
## alternative hypothesis: true mean is not equal to 0.196026
## 95 percent confidence interval:
```

```
##  0.1943670 0.2327401
## sample estimates:
## mean of x
## 0.2135535
```

```
# Difference is significant for 2021
t.test(data_final_full$education_Completed_secondary, mu = 0.263338)
```

```
##
##  One Sample t-test
##
## data:  data_final_full$education_Completed_secondary
## t = -41.21, df = 8213, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0.263338
## 95 percent confidence interval:
##  0.1101617 0.1240725
## sample estimates:
## mean of x
## 0.1171171
```

```
t.test(data_final_full_2004$education_Completed_secondary, mu = 0.276679)
```

```
##
##  One Sample t-test
##
## data:  data_final_full_2004$education_Completed_secondary
## t = -6.9734, df = 1755, p-value = 4.364e-12
## alternative hypothesis: true mean is not equal to 0.276679
## 95 percent confidence interval:
##  0.1899620 0.2280334
## sample estimates:
## mean of x
## 0.2089977
```

```
# Difference is significant

table(data_cleaned$gender)/n_2021
```

```
##
##    female      male
## 0.5530801 0.4469199
```

```
table(data_cleaned_2004$gender)/n_2004
```

```
##
##    female      male
## 0.5381549 0.4618451
```

```
male_prop_2004 <- 0.482694191
male_prop_2021 <- 0.486704231
t.test(data_final_full$gender_male, mu = male_prop_2004)
```

```
##
##  One Sample t-test
##
## data:  data_final_full$gender_male
## t = -6.521, df = 8213, p-value = 7.397e-11
## alternative hypothesis: true mean is not equal to 0.4826942
## 95 percent confidence interval:
##  0.4361659 0.4576739
## sample estimates:
## mean of x
## 0.4469199
```

```
t.test(data_final_full_2004$gender_male, mu = male_prop_2021)
```

```
##
##  One Sample t-test
##
## data:  data_final_full_2004$gender_male
## t = -2.0889, df = 1755, p-value = 0.03686
## alternative hypothesis: true mean is not equal to 0.4867042
## 95 percent confidence interval:
##  0.4385046 0.4851856
## sample estimates:
## mean of x
## 0.4618451
```

```
# Difference is significant

table(data_cleaned$age_group)/n_2021
```

```
##
##          1          2          3
## 0.05880205 0.73459946 0.20659849
```

```
table(data_cleaned_2004$age_group)/n_2004
```

```
##
##          1          2          3
## 0.05523918 0.72722096 0.21753986
```

```
t.test(data_final_full$age_group_2, mu = 0.738841507)
```

```
##
##  One Sample t-test
##
## data:  data_final_full$age_group_2
## t = -0.87066, df = 8213, p-value = 0.384
## alternative hypothesis: true mean is not equal to 0.7388415
## 95 percent confidence interval:
##  0.7250487 0.7441502
## sample estimates:
## mean of x
## 0.7345995
```

```r
t.test(data_final_full_2004$age_group_2, mu = 0.692875989)
```

```
##
##  One Sample t-test
##
## data:  data_final_full_2004$age_group_2
## t = 3.2304, df = 1755, p-value = 0.001259
## alternative hypothesis: true mean is not equal to 0.692876
## 95 percent confidence interval:
##  0.706369 0.748073
## sample estimates:
## mean of x
##  0.727221
```

```r
t.test(data_final_full$age_group_3, mu = 0.171310728)
```

```
##
##  One Sample t-test
##
## data:  data_final_full$age_group_3
## t = 7.8989, df = 8213, p-value = 3.179e-15
## alternative hypothesis: true mean is not equal to 0.1713107
## 95 percent confidence interval:
##  0.1978412 0.2153558
## sample estimates:
## mean of x
## 0.2065985
```

```r
t.test(data_final_full_2004$age_group_3, mu = 0.228304896)
```

```
##
##  One Sample t-test
##
## data:  data_final_full_2004$age_group_3
## t = -1.0931, df = 1755, p-value = 0.2745
## alternative hypothesis: true mean is not equal to 0.2283049
## 95 percent confidence interval:
##  0.1982242 0.2368555
## sample estimates:
## mean of x
## 0.2175399
```

```r
# Some of the age groups have a significant difference from the population
```

Computing Weights

```r
proportions_pop <- read_csv("population_proportions.csv") %>%
  # Want to deal with completed secondary, some college, some uni separately
  filter(education != "Completed_secondary")
```

```r
# Compute the proportions in the data
data_counts <- data_cleaned %>% group_by(gender, year, age_group) %>%
  summarise(total_count=n(), .groups = 'drop')
proportions_data <- data_cleaned %>%
  group_by(gender, year, age_group, education) %>%
  summarise(count=n(), .groups = 'drop') %>%
  left_join(data_counts) %>%
  mutate(proportions_data = count/total_count)
# Address the portion of the population that is not accounted for
# Want to assign this portion to Completed_secondary, Some_college, and Some_university
proportions_data_leftover <- proportions_data %>%
  # Calculate the leftover portion that these three make up in the data
  filter(education %in% c("Completed_secondary", "Some_college", "Some_university")) %>%
  group_by(gender, year, age_group) %>%
  mutate(proportions_data_leftover_total=sum(proportions_data)) %>%
  ungroup() %>%
  # Find the ratio between these three relative to the total leftover in the data
  mutate(proportions_data_leftover_ratio = proportions_data/proportions_data_leftover_total)
proportions_data <- proportions_data %>%
  left_join(proportions_pop) %>%
  left_join(proportions_data_leftover) %>%
  replace(is.na(.), 0) %>%
  # Calculate the leftover portion in the population
  group_by(gender, year, age_group) %>%
  mutate(proportions_total=sum(proportion)) %>%
  ungroup() %>%
  mutate(proportions_leftover=1-proportions_total) %>%
  # Estimate the proportions of Completed_secondary, Some_college, and Some_university
  # in the total population using the ratios in the data
  mutate(proportion = ifelse(proportion==0,
                             proportions_data_leftover_ratio*proportions_leftover,
                             proportion))
proportions_data <- proportions_data %>%
# Weight is proportion in population/proportion in sample
  mutate(weight = proportion/proportions_data) %>%
  select(gender, year, age_group, education, weight)

data_final <- data_cleaned %>%
  left_join(proportions_data) %>%
  select("vote_choice", "union", "education", "religion", "weight", "year") %>%
  # Create dummy variables
 dummy_cols(select_columns = c("education", "religion")) %>%
  # Remove original columns
  select(-c("education", "religion"))

# Create dummy variables for all features, for additional analysis
data_final_full <- data_cleaned %>%
  dummy_cols(select_columns = c("gender", "education", "religion", "province", "marital",
                                "income_cat", "age_group")) %>%
  select(-c("gender", "education", "religion", "province", "marital",
            "income_cat", "age_group"))

proportions_pop <- read_csv("population_proportions.csv") %>%
```

```
  # Want to deal with completed secondary, some college, some uni separately
  filter(education != "Completed_secondary")
# Compute the proportions in the data
data_counts <- data_cleaned_2004 %>% group_by(gender, year, age_group) %>%
  summarise(total_count=n(), .groups = 'drop')
proportions_data <- data_cleaned_2004 %>%
  group_by(gender, year, age_group, education) %>%
  summarise(count=n(), .groups = 'drop') %>%
  left_join(data_counts) %>%
  mutate(proportions_data = count/total_count)
# Address the portion of the population that is not accounted for
# Want to assign this portion to Completed_secondary, Some_college, and Some_university
proportions_data_leftover <- proportions_data %>%
  # Calculate the leftover portion that these three make up in the data
  filter(education %in% c("Completed_secondary", "Some_college", "Some_university")) %>%
  group_by(gender, year, age_group) %>%
  mutate(proportions_data_leftover_total=sum(proportions_data)) %>%
  ungroup() %>%
  # Find the ratio between these three relative to the total leftover in the data
  mutate(proportions_data_leftover_ratio = proportions_data/proportions_data_leftover_total)
proportions_data <- proportions_data %>%
  left_join(proportions_pop) %>%
  left_join(proportions_data_leftover) %>%
  replace(is.na(.), 0) %>%
  # Calculate the leftover portion in the population
  group_by(gender, year, age_group) %>%
  mutate(proportions_total=sum(proportion)) %>%
  ungroup() %>%
  mutate(proportions_leftover=1-proportions_total) %>%
  # Estimate the proportions of Completed_secondary, Some_college, and Some_university
  # in the total population using the ratios in the data
  mutate(proportion = ifelse(proportion==0,
                             proportions_data_leftover_ratio*proportions_leftover,
                             proportion))
proportions_data_2004 <- proportions_data %>%
  # Weight is proportion in population/proportion in sample
  mutate(weight = proportion/proportions_data) %>%
  select(gender, year, age_group, education, weight)

data_final_2004 <- data_cleaned_2004 %>%
  left_join(proportions_data_2004) %>%
  select("vote_choice", "union", "education", "religion", "weight", "year") %>%
  # Create dummy variables
  dummy_cols(select_columns = c("education", "religion")) %>%
  # Remove original columns
  select(-c("education", "religion"))

# Create dummy variables for all features, for additional analysis
data_final_full_2004 <- data_cleaned_2004 %>%
  dummy_cols(select_columns = c("gender", "education", "religion", "province", "marital",
                                "income_cat", "age_group")) %>%
  select(-c("gender", "education", "religion", "province", "marital",
            "income_cat", "age_group"))
```

Model and Results

```
data_final <- data_final %>% rbind(data_final_2004) %>%
  mutate(year = case_when(
    year == 2021 ~ 1,
    year == 2004 ~ 0)) %>%
  relocate(union, .after = year)

# Use Liberal as base, as they are the winners in both years
data_final$vote_choice <- relevel(as.factor(data_final$vote_choice), ref = "Liberal")
fit <- multinom(vote_choice ~ . - weight + year:(. - weight),
                data = data_final, weights=weight, maxit=100)

# Returns dataframe of coefficients, p-values, and significance
get_fit_results <- function(fit){
  coeffs <- as.data.frame(t(summary(fit)$coefficients))
  # Z-score
  z <- t(summary(fit)$coefficients/summary(fit)$standard.errors)
  # p-value
  p <- as.data.frame(1 - pnorm(abs(z), 0, 1)) * 2
  colnames(p) <- c("Conservative_pval", "NDP_pval")
  results <- cbind(coeffs, p) %>%
    mutate(Conservative_sig = (Conservative_pval < 0.05)) %>%
    mutate(NDP_sig = (NDP_pval < 0.05))
  return(results)
}
results <- get_fit_results(fit)

# Adds single and interaction term to give total impact on log odds
get_2021_effects <- function(i, results, k){
  C <- results$Conservative
  C_odds <- C[i] + C[i+k]
  N <- results$NDP
  N_odds <- N[i] + N[i+k]
  return(c(C_odds, N_odds))
}
# k is the number of rows apart a variable is from its interaction with year
k <- 16
variable_names <- fit$coefnames[c(3:18)]
effects_2021 <- sapply(c(3:18), get_2021_effects, results, k)
effects_2021 <- data_frame(variable=variable_names,
                           Conservative = effects_2021[1, ], NDP = effects_2021[2, ])
```