Study on Canadian Federal Election 2019

Nuo Chen

27 September 2020

Contents

Introduction

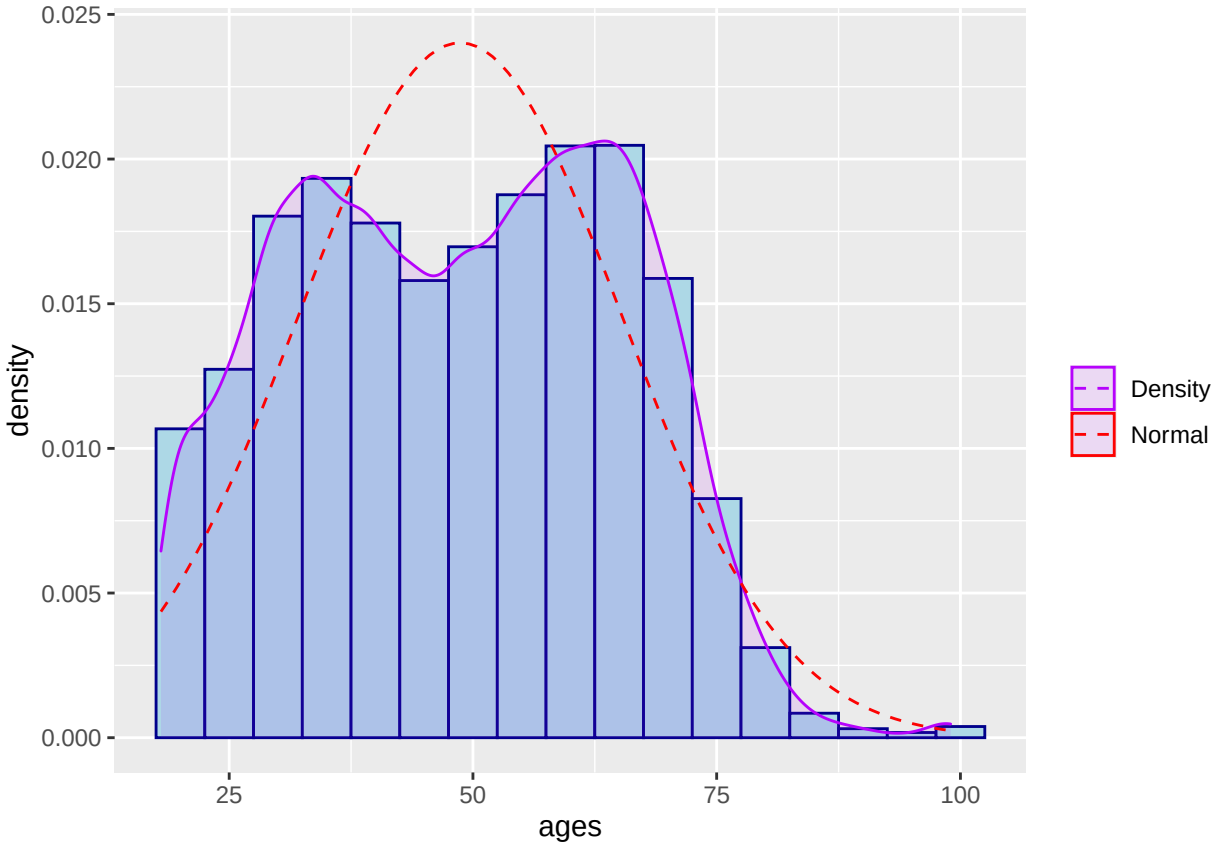
Can we forecast Federal Election's outcome, people's favorite Party using voters personal information as variable? It seems no one has a definite answer on whether the election result is predictable, but it was stated by many scientists that there are correlations between voters' factors and the candidate that he/she votes for. This is a very interesting and widely researched domain. We've carried out a data analysis on the 2019 Canadian Federal Election internet survey trying to see if we can explain people's political interest by the information we gathered from the survey.

Dataset

The dataset has gather the attitudes and options of Canadian during and after the federal election in 2019. Various variables have been collected by individuals such as age, education, citizenship, living place, and political interests. We noticed that most (nearly all) of the variables are categorical other than numerical value. Therefore, the study we carried out will focus on how Canadian political stands are explained by these categorical personal information. For the sake of simplicity we concentrate our analysis on the Party people will vote as first choice, and their level of interest in politics in general. We are aware that numbers of choices have been collected by the survey, however, we believe people weight more on their first choice when it comes to vote for a Party.

Age distribution

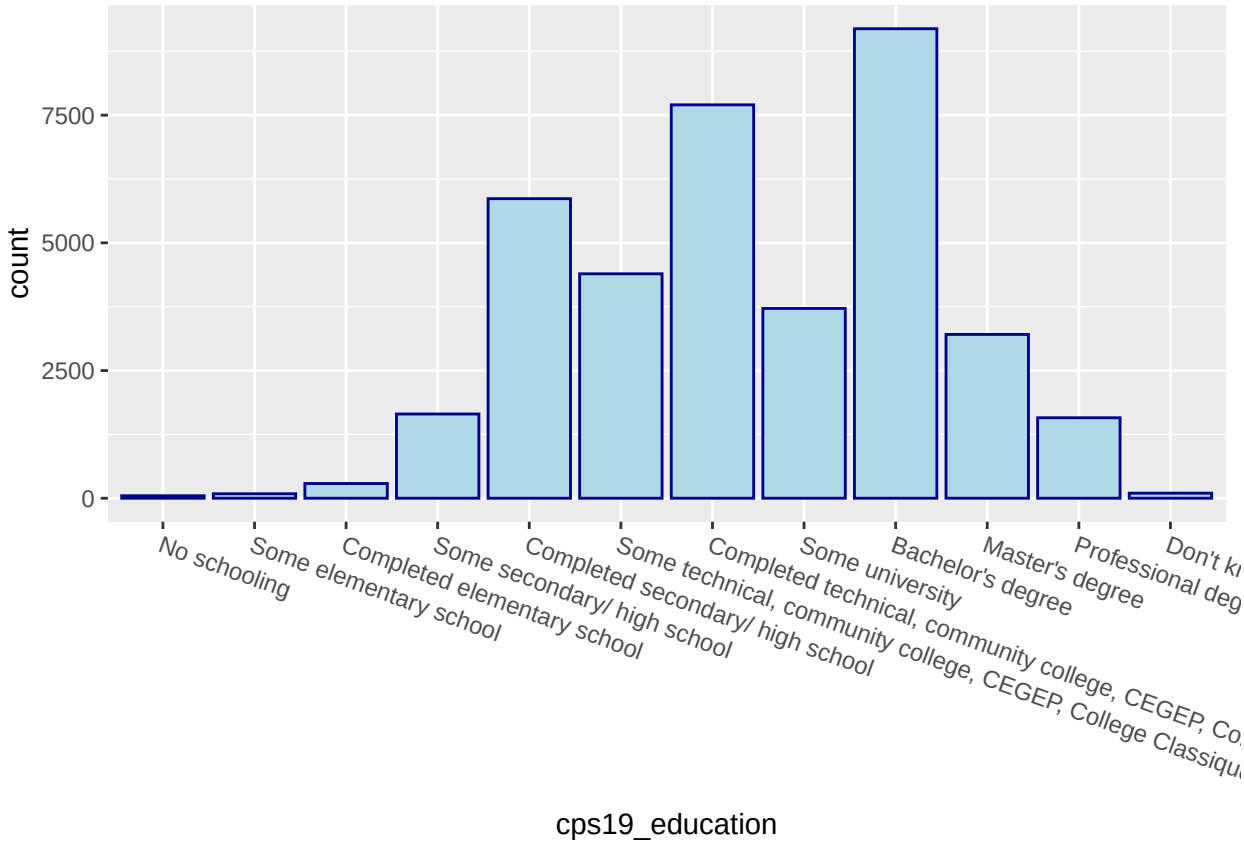
We firstly look at the age distribution of the sample to check if this survey has been distributed randomly and independently to the audience other than to a specific group. As we plot the histogram of the age, along with a fitted (equal sample mean, and standard deviation) normal density function. We observed that the age distribution of the sample is very close to the normal density curve except that we have few samples around the mean. This could be that we have more participants at age around 50-75 than expected, so the mean of the age has been shifted more to the right. But overall, we see that the age distribution is normal.



Education

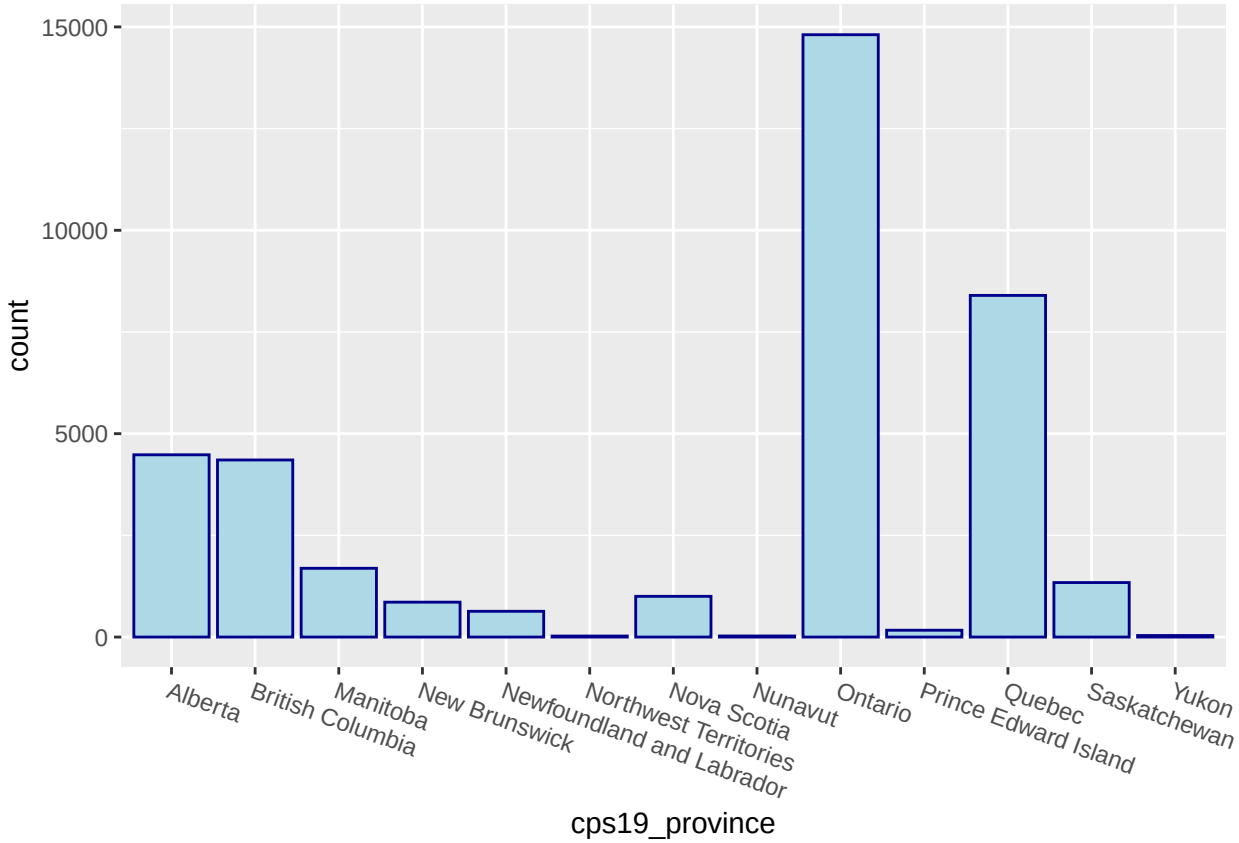
The second plot is the histogram of highest education obtained by the samples. We noticed most of the samples have completed bachelor's degree, followed by colleges, and secondary school. Very few people have

completed professional degree or doctorate. We consider the sample is representative of the true population. This is very important when taking large sample for statistical analysis. We always want to draw a dataset to represent the population. If we have bias in our sample, we will get bias in our analysis using this sample.



Province

We then plot the histogram of the province of people live in. It is obvious that most people did the survey are from Ontario and Quebec which are the two largest province in Canada. Alberta, British Columbia, and Saskatchewan also represent a large potion of the participants. These are actually the groups that we should focus as the voters from these regions often has more impact on the final election result.



Gender

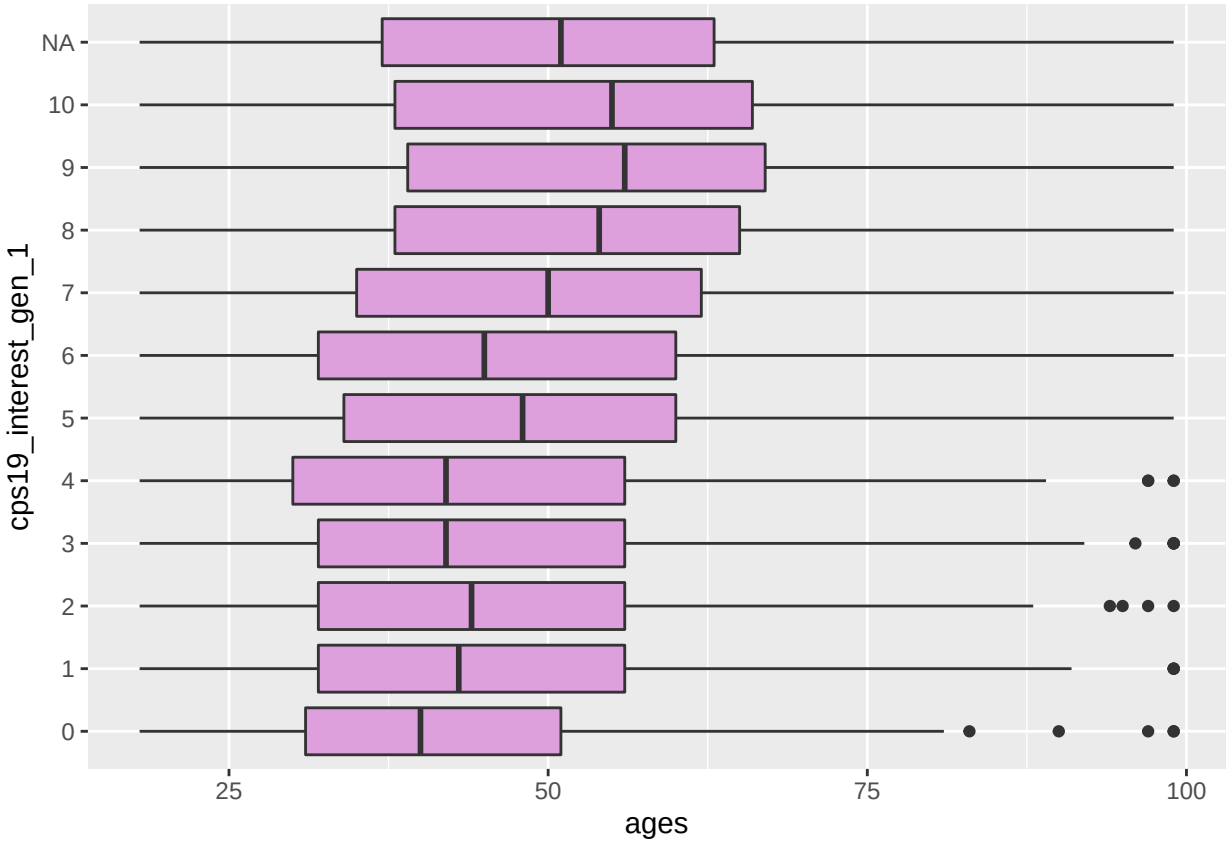
Looking at the gender/province histogram, we noticed that we have more female participants than male participants. This is somehow counter intuitive as we first believe we should have almost equal male and female participants. However we did an quick research and found out that the percentage of Canadian female is 50.4%. This explains the ratio of female is higher then male in this survey.

Political interest

Now we will look at the more interesting part of this study, political interest, and how it is affected by other factors that have been collected in the data set.

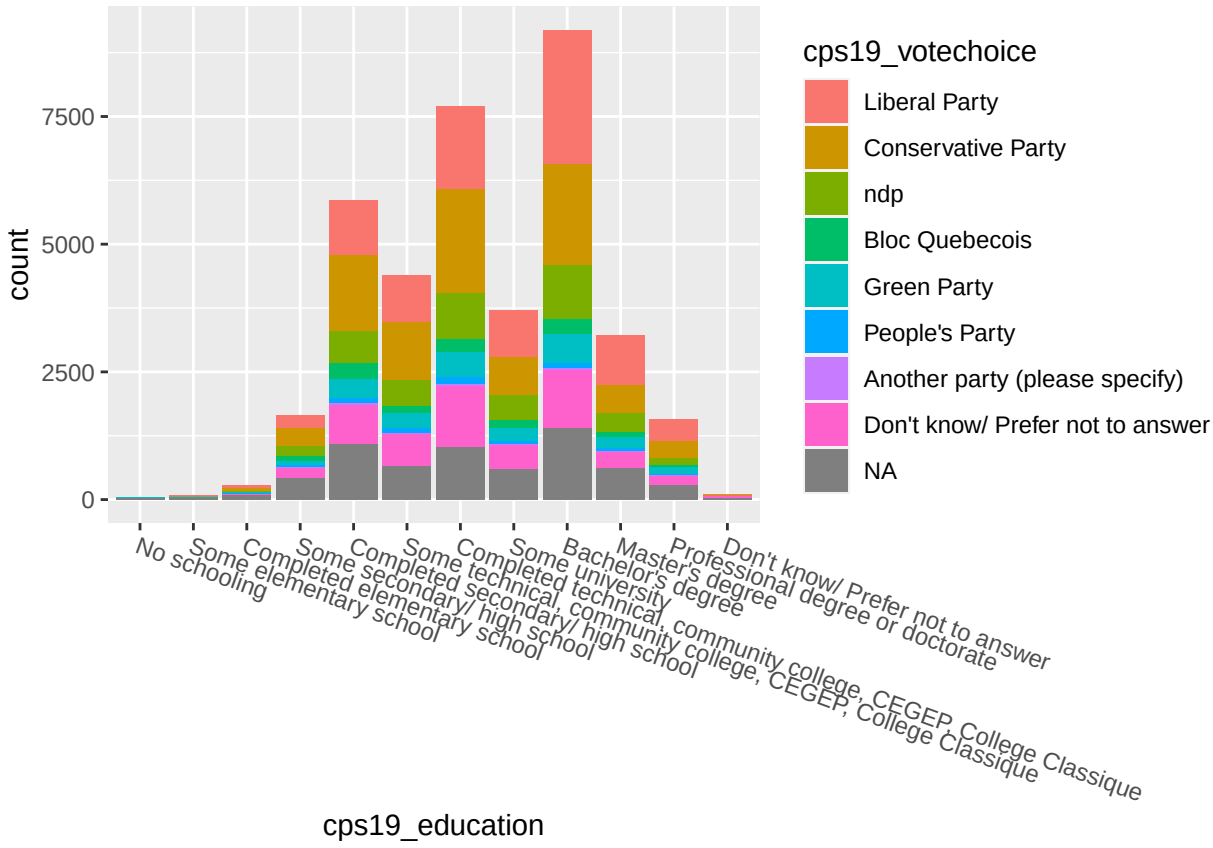
Are you interested in politics?

Below is a box plot of the level of interest in politics on scale of 0 to 10 by age. The middle bar of a box represents the mean of sample in the group, the left and right edge of the box represent the maximum deviation from the mean. We clearly observe that elder participants show more interest in politics than the younger ones. This is also very representative of the true population. Younger generations are more attracted by social media, technology, entertainment than politics while adults are more concerned by politics, culture, public policy, etc. It is also an hint that the politic stands can vary from age group to age group.



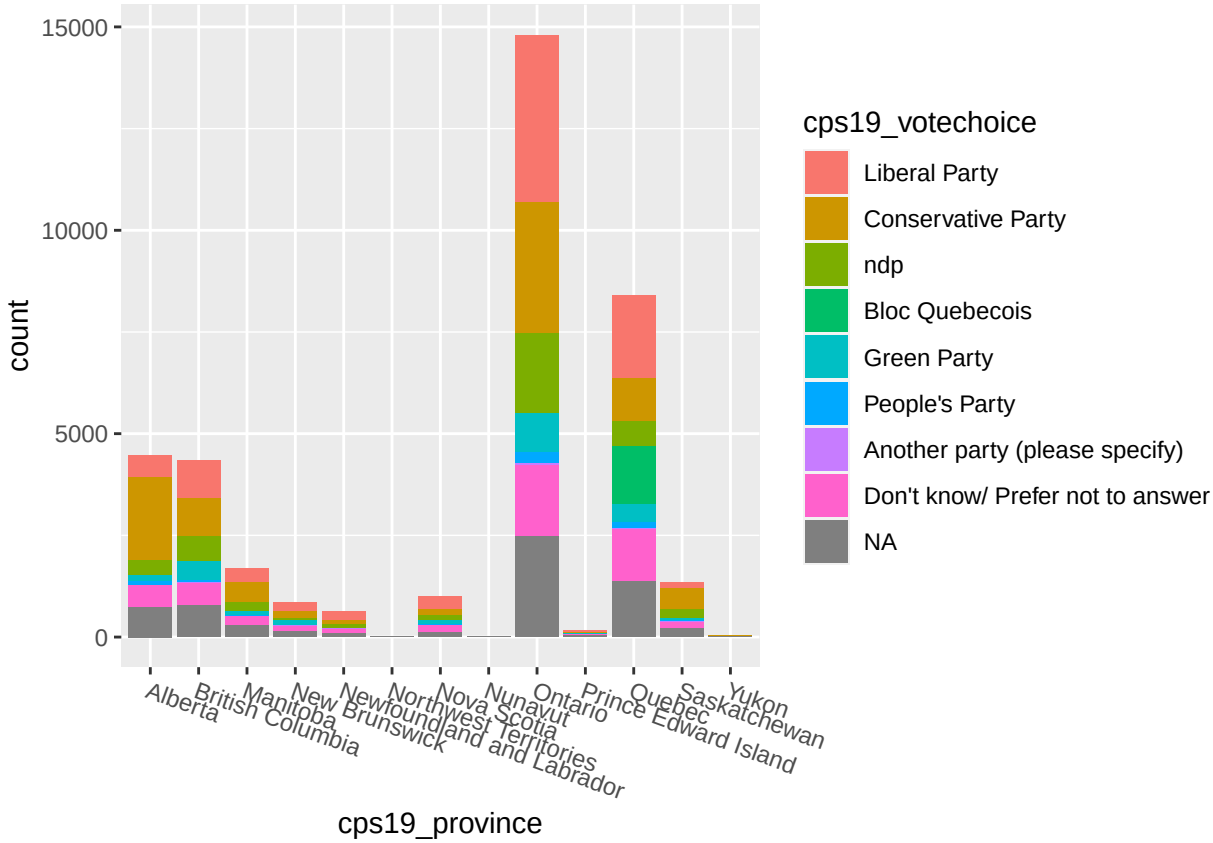
Your school tells you which Party to vote?

Interestingly, when we plot a histogram again for education but this time grouped by the Party that participant intent to vote. We observed clearly that Liberal Party has gained its support mostly in people have completed higher educations such as Bachelor, Masters, doctorate, etc. In the contrast, Conservative Party gathers its supporters from People with less higher but more technical education background (technical community collect, secondary high school . . .). This is due to the fact that different Party have different focus, strategies, politic stands that may apply to different professions, and academic path. People will vote for the Party they know well and the Party that addresses their issues.



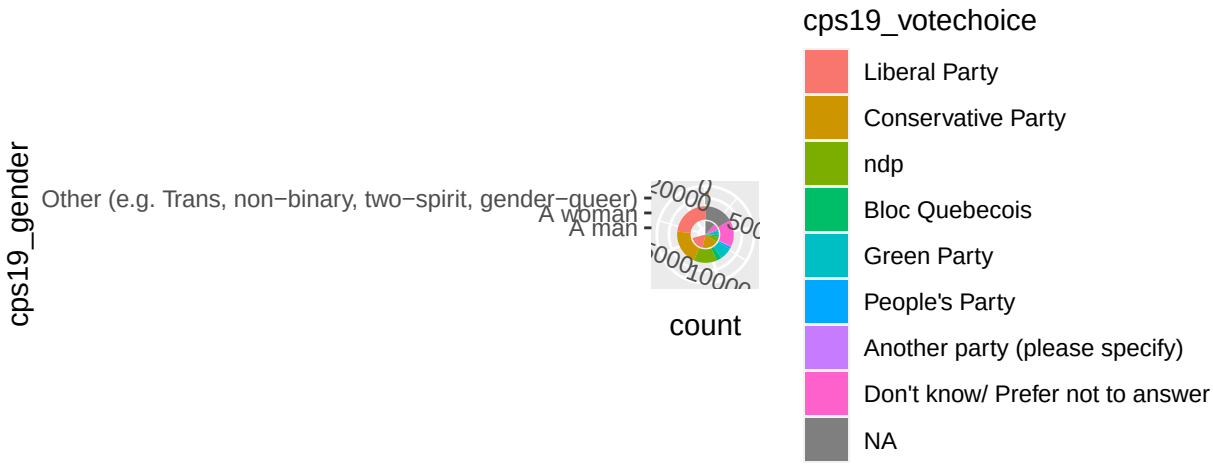
Living with the ones voted the same?

If we plot the region histogram grouped by Party. Liberal Party is obviously outstanding in the 2 largest provinces in Canada. However, people living in Alberta, Saskatchewan, and other small provinces seem to prefer Conservative Party. Parties can have their political strategies and politic stands base on geography. In fact, there is a branch of study called "Electoral Studies in Political Geography" which is the analysis of the methods, the behaviors, and the results in the context of geographic space.



Women prefer the Liberal Party while men are indifferent?

We finally plot a pie chart by gender and group the count by Party choice. Looking at the male participants, we don't see a clear favorite between Liberal Party and Conservative Party. However, female voters obviously prefer the Liberal than the Conservative. This is also a very natural finding that represents the consensus. The effect of candidate gender on election poll is a mixture of psychological and historical factors.



Shortcoming of dataset and analysis

One shortcoming of the dataset is large number of missing values. This is probably due to the natural of online survey - participants have the option to not answer all the questions. Another shortcoming of the analysis is we are just raising questions and stating observations other than testing a hypothesis. For that we need to write null hypothesis and perform statistical testing which is out of the scope of this study.

References

(Stephenson 2020) (Stephenson and Forthcoming 2019)

Stephenson, Allison Harrel, Laura, and Peter Loewen. Forthcoming. 2019. *Measuring Preferences and Behaviour in the 2019 Canadian Election Study*. V1 ed. Canadian Journal of Political Science. https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DUS88V.

Stephenson, Allison; Rubenson, Laura B; Harell. 2020. *2019 Canadian Election Study - Online Survey*. V1 ed. Harvard Dataverse. https://doi.org/10.7910/DVN/DUS88V.