Male......Voters Prefer to Vote for Conservative

Group 9 - Yijie Zhao, Yuze Kang, Shuyu Duan & Rachel Oh

October 7, 2020

Executive Summary

Petit Poll, the polling company that has been cooperating with Conservative, updates the status of polling monthly, in Ontario, Canada. By sending online surveys to emails of vast voting age population in Ontario, not only opinions but also characteristics of these eligible voters are acquired and analyzed, for further estimation and reference. From voter's age, gender, preference of political party, to their income and education level, we have discovered that...

The surveys are randomly sent to a considerable amount of people, in order to obtain unbiased results that can represent the voting population in Ontario, still we've missed out a lot of responses which is reasonable as not everyone can be reached in time on the Internet even in this time.

1) the task (including specifying the client and the focus);
2) what was done;
3) what was found;
4) why this matters;
5) some weaknesses; and
6) areas for future work.

Introduction

In Canadian federal election, Ontario has always been an important battleground. As the most populous province of Canada, Ontario counts 121 federal electoral districts (i.e., ridings in Canadian English) holding voting rights and has 121 seats, more than one third in the House of Commons. Therefore, polling of Ontario voters has a strategic significance for all ambiguous political parties, especially for Conservative. However, does it definitely lead Conservative to winning federal government by leading the poll before federal election? Absolutely not. Even a higher voting rate can not ensure a final winning, as a fact shows in 2019 Canadian federal election that Conservative lost government while gained more votes than Liberal. So both votes and seats matter.

There is a variety of interior or exterior factors that affects each voter's decision on the election date and causes the consequence that polling is not as much reliable as we think. In the following context, we decide to focus on some common factors of voters themselves, such as age, sex, education and income level, to find the relation to their preferences of political parties. In addition to polling updates in October, we try to figure out if there is any common features of voters who prefer to vote Conservative in the federal election, for Conservative Party's reference. Based on that, no matter the voters are rational or irrational on that big date, we might use this observation to help Conservative stabilize their current voters and win over more voters of Ontario, to win more seats in such an important province. Besides, we discuss the weakness and limitation contained in our whole analyzing process for readers' reference and for improvement next time.

All the code and data supporting this analysis is available at: https://github.com/JessieZ32/Polling-for-Conservative.

Survey Methodology

To achieve the aims of our report, we need to collect data first. Our target population is all the eligible voters of Ontario, i.e., any at least 18-year-old Canadian citizen without deprived of civil right whose residential address is located in Ontario. And our sampling population is whom we can possibly reach out among the target population. For data collection, we first designed an appropriate anonymous survey for voters of Ontario, including all information, or say, variables we are interested in, through SurveyMonkey (*Survey Monkey Inc.*, n.d.). The link and a screenshot of our survey is available at the appendix of this report.

We collaborated with some big Canadian communications and media companies, Rogers, Bell and Telus, as they own lists of customers' email addresses which may cover most of our target population (i.e., our frame). Through their email system, we massively sent emails with our survey URL link in. Both in content of email and survey, we promised to keep all the collected data residential, with respect to the protection of participants' privacy, and asked for ignorance if people who received this email weren't not eligible voters. Then we collected all the data from 7958 respondents as samples. Through the whole process, the sampling method that we used is simple random sampling without replacement (SRSWOR). We hope sample can be chosen with an equal probability, making the collected sample could represent target population with as less bias as possible.

As for the costs of sampling process, we only need to calculate the wage we pay for employees who did the job of sending emails. Doing a survey through some kind of modern media can cost less that survey face-to-face. However, there exists some errors due to our sampling design. For example, delivering the survey to voters online can definitely cause non-response error. There is no way to ensure them checking email regularly and doing survey willingly. But we can increase the rate of response by setting up some rewards, for example, we can set a raffle page and all the respondents shall do a lucky raffle once they submit the survey. Hence, we can reduce the influence of non-response error as much as possible.

Results & Discussion

Among 7958 respondents who are eligible voters in Ontario, we collect all their information from the survey and combine them into a data set file. We set up several variables of which we think matter polling or voting behavior, such as age, sex, personal annual income level, education level and federal electoral districts voters live. In the following context, we use software R and R packages like tidyverse and ggplot2 to analyze the data of those variables we collected from survey respondents and draw some tables and graphs to visualize the data (R Core Team 2020; Wickham et al. 2019; Wickham 2016).

I. October Polling Updates for Canadian Political Parties

· CON represents Conservative Party of Canada; · NDP represents New Democratic Party; · GRN represents Green Party of Canada; · IND represents Independent; · PP represents People's Party; · OTH represents Bloc Québécois and other parties or individual candidates.

The table below shows the ratio of polling in Ontario for Canadian Political Parties updated in October from the sample data we collect.

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 7 x 2
##   PoliticalP_supported Polling
##   <chr>                  <dbl>
## 1 CON                    0.320
## 2 GRN                    0.0608
## 3 IND                    0.00503
## 4 LIB                    0.424
```

```
## 5 NDP                    0.171
## 6 OTH                    0.00377
## 7 PP                     0.0160
```

From the table, we can see that Conservative has a polling ratio of over 30% and Liberal is still in a moderately dominant position among polling in Ontario as usual. Except Liberal, around 17% polling ratio of New Democratic should also be cared about. The rest parties' polling accounts for less than 0.1%.

## II. Ontarian Voters in Different Age pro Conservative

We draw a Kernel density plot for voters in different age who support Conservative in Ontario.

Age Distribution of Conservative Supporters
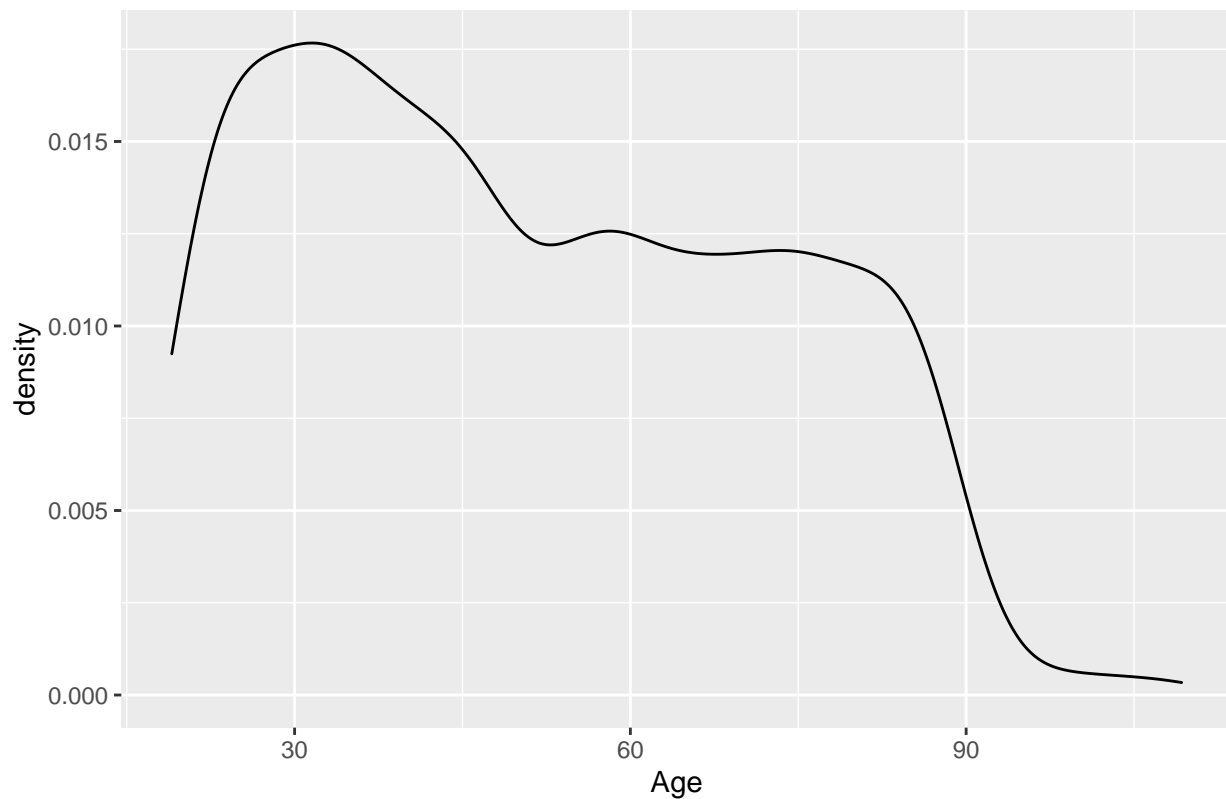
This density plot indicates the distribution of respondents who prefer to vote Conservative. It is showed that this distribution is a multi-modal mode, with an obvious peak at the age of 30 and two other inapparent peaks at the age of near 60 and 75 respectively. The overall distribution is right-skewed. The maximum is the first peak of age 30. The density has two sudden falls: the first is after age 35 to 55 and the second is after age 85. From age 18 to 30, the density rises rapidly. From age 55 to 85, the density is pretty stable.

## III. Male or Female's Preferences for Political Parties

For male and female voters, we draw a horizontal bar plot to present the preferences for parties of different sex.
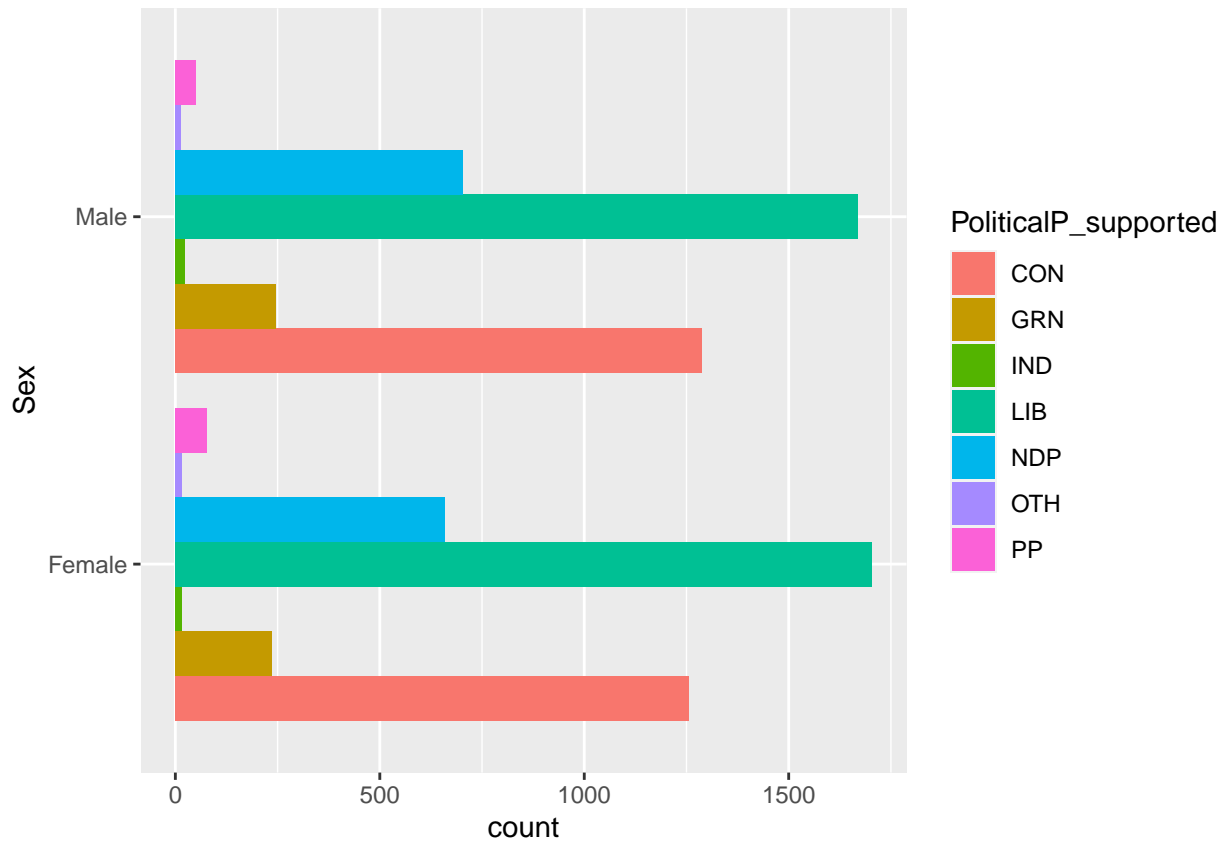
The ratios of male and female to the total sample we collect are almost equal. From the bar plot above, we can find that male and female show similar preference for each party. There are slight differences in supporting Liberal and People's Party between male and female. Both male and female prefer the most to vote Liberal and then Conservative.

IV. Polling Associated to Education Level of Respondents

We draw a vertical bar plot to represent the polling for different parties of groups with different education level
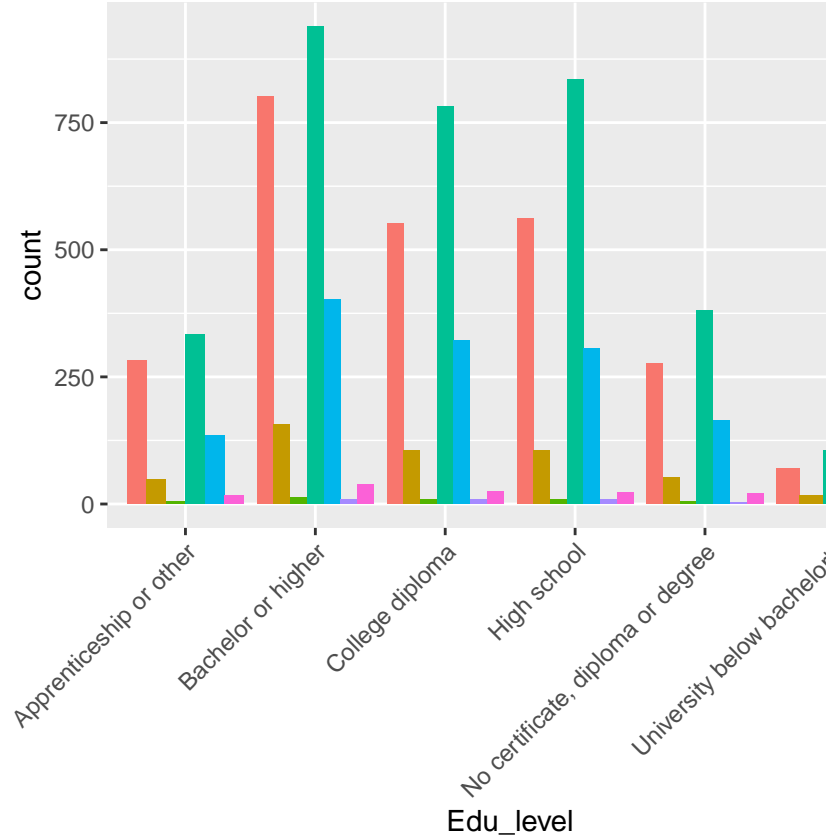

and we concern about the data for Conservative mostly.

From the graph above, we can see that among all the supporters for Conservative, the largest proportion is respondents with the highest level of education, like holding a bachelor degree or higher level degree.

V. The Association Between Average Income and Polling for Conservative

We collect the annual after-tax incomes from our respondents. We also question them which federal electoral districts of Ontario they are living in. After we calculate the following two variables: · the average after-tax annual incomes of different federal electoral districts; · the polling for Conservative in different federal electoral districts; we draw a scatter plot for these two variables.
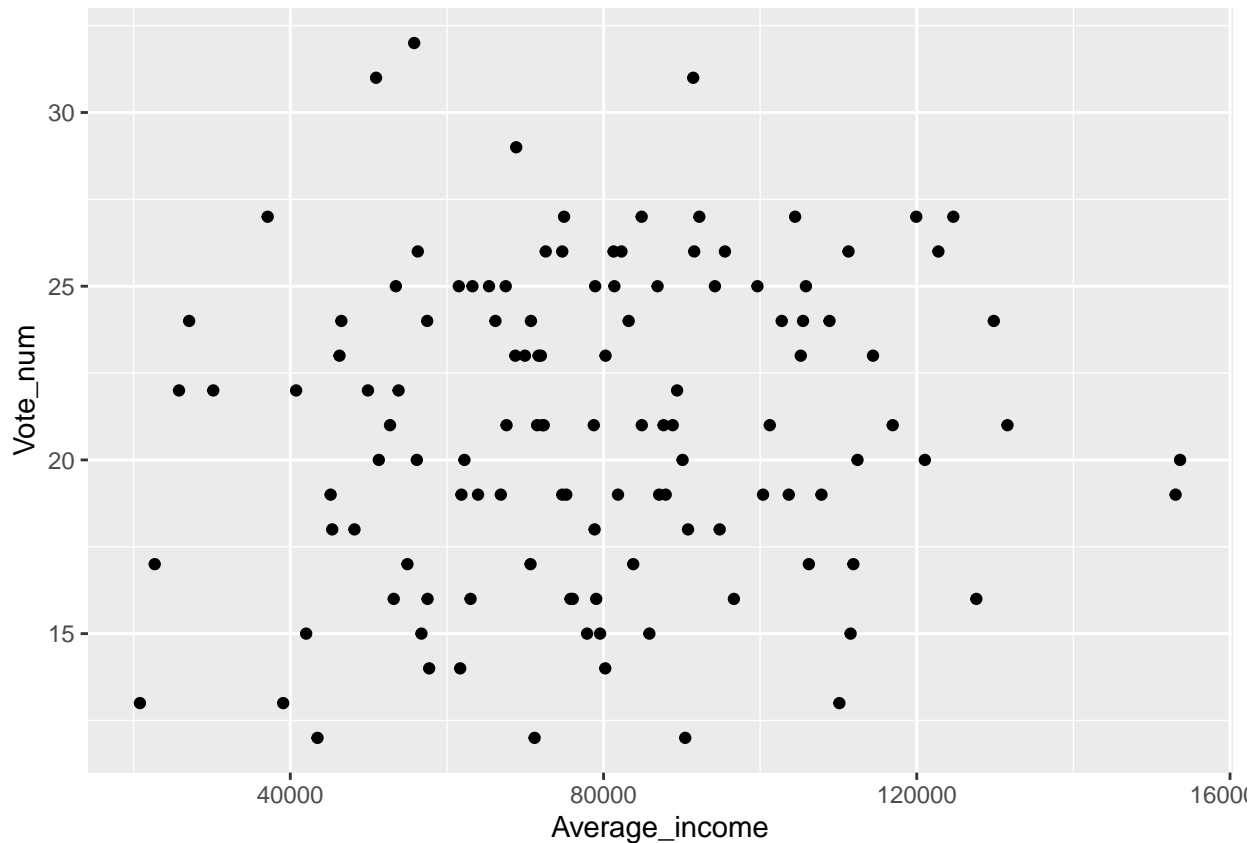
Actually, we could not find a association between the average after-tax annual incomes and the polling for Conservative in different federal electoral districts. These 121 points are almost uniformly distributed near the center part of the whole graph.

VI. Discussion

Online surveys can actually miss out a considerable proportion of people who do not often use computers or the internet, for example the elderly, or people with heavy financial burden. Maybe we can reach the respondents by both online surveys and offline interviews, or in another combined way, to receive responses from all groups of people in the population. Simple Random Sampling might be biased when the sample size is not large enough. In this case the number of responses we collected is not ideally large enough to display all characteristics of the population in Ontario. Even if a person's personal information is secured, he or she may not be fully honest when filling the questionnaire (For example, some people are very sensitive when it comes to their income). Intentionally or Consciously, this might always leads to some influences to our results of survey.

Appendices

reference

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Survey Monkey Inc. n.d. San Mateo, California, USA. www.surveymonkey.com.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.