2020 September to October 6 Monthly Polling Intentions analysis for Conservative, the party gains the most favor from high-income groups and the old generation

Tianxiao Ma

October 7, 2020

Executive summary

This report was commissioned to analyze Ontario resident's polling intentions through online surveys collected from September 1st, 2020 to October 6th, 2020. The report provides a basic understanding of general voters' characteristics and a more detailed understanding of voters who support our client, the Conservative Party.
The survey divides respondents into various groups associated with their age, income, and favored political party. Conservative and Liberal are the two main parties holding the majority of polls in Ontario. Further analysis reveals people with low income and young age have an obvious tendency to vote for Liberal while voters aged over 50 are more willing to choose the Conservative Party. Moreover, Conservative voters have the highest average after-tax income of 52000CAD. These findings appear to reveal the targeting voters and the direction of developing policy guidelines for the Conservative Party. Furthermore, make appropriate adjustments on policy targeting the group that least favors the Conservative Party and take full advantage of current policy such as lowering the tax rate to retain votes. Online surveys might not be able to provide 100% reliable and accurate results for further analysis due to the lack of engagement, anonymous property of the survey. Other unaccounted hidden factors may also influence the survey results, such as unemployment rate influences income and makes it hard to calculate annual income. In future, it may be worth adding questions associated with current issues such as how COVID-19 influences income. Conducting surveys through phone calls and mails may also increase the reliability of results.

Introduction

The act of 'labeling' is categorizing people and objects by specific attributes; it happens on a daily basis in the new era of big data. People label each other just like supermarket employees organize goods by categories, and sometimes people store and analyze these 'labels' using computers. 'Labelling' is essential for political parties to gain a better insight of targeting voters. Political parties like Conservative, Liberal, Green and NDP all develop different campaign strategies according to the voting population they want to engage. Candidates of each political party have to be prepared for general elections at all times. However, this is quite hard to define who is likely to vote for them. Understanding the needs of different population groups categorized by voter's income and age is crucial for policy making, future direction and winning the election. With the data collected in the survey, Conservative party is able to focus on the group that least favors them and make appropriate adjustments on their policy guidelines, and make full advantages of current policy such as lower the tax rate to retain votes. In the next section, we're going to introduce the survey methodology and visualize results with appropriate explanations. Some Weakness and bias of the survey are included in the discussion section with some future opportunities advice. After that, I'll put discussion appendices and references at the end of this report.

The code is stored in the github file link: https://github.com/rubytianxiaoma/STA304-PS2/blob/main/Monthly%20Poll%20result%20Sep%20to%20Oct%202020.Rmd (https://github.com/rubytianxiaoma/STA304-PS2/blob/main/Monthly%20Poll%20result%20Sep%20to%20Oct%202020.Rmd)

Survey methodology

The objective of this election survey is to analyze the proportion of Ontario residents supporting Conservative Party of Canada for the 2020 Canadian Election. The target population is all the Ontario residents that registered with the government system, and the sample population is 300 Ontario residents that registered with the government system. A complete list of email addresses of Ontario residents that registered with the government system(e.g. CRA) constitutes the sampling frame.
To ensure that the selected sample is representative, Simple random sampling without replacement(SRSWOR) method is used in the survey. A distinct sample size 300 is selected with equal probability to be chosen for a population of size 100000. Under SRSWOR, Let {(i, yi), i ∈ S} be the survey data, consisting of the labels i of the units sampled, and the associated values of y. Let the sample mean described as

$$\bar{y} = \sum i \in S(y_i)$$

Sample Variance:

$$s_y^2 = 1/n - 1 \sum i \in S(y_i - \bar{y})$$

The sample mean

$$\bar{y}$$

is a design-unbiased estimator for the population mean

$$\mu_y$$

$$E(\bar{y}) = \mu_y$$

An unbiased variance estimator for

$$\bar{y}$$

is given by

$$V(\bar{y}) = (1 - n/N)s_y^2/n$$

where the expectation of variance of y_bar equals to variance of y_bar.

All the responses are collected from September 1st to October 6, 2020, and analyzed on October 6th. These responses can closely describe the Ontario electoral district's final voting results. The survey is sent to desired respondents' email address registered with the government system. To avoid nonresponses, the company stimulates desired respondents by sending them 5 dollars to their bank account after finishing the survey, and the 5CAD reward part is highlighted at the beginning of the survey. If there's no response after a week, a reminder email will be sent to respondents again. Non-responses usually mean they're not interested in voting or the email is detected as junk mail so the desired respondent doesn't get a chance to notice it. Estimated cost consists of a reward fee(5 CAD/person, 1500CAD in total) and a labor cost (1000CAD), the total amount is roughly 2500CAD.

To protect respondent privacy, there's a pop up window stating what data we expect from respondents and how we are going to use collected data before respondents answering the survey. Respondents can choose to keep answering the following questions by clicking the 'continue' button. Besides it, every surveyor and data analyst working in our company has to sign a data protection agreement to avoid data leak. Questions like which party respondents will vote for if there's a general election holding today, their age, income, and the satisfaction level of the current government in different categories such as tax rate, child benefit and minimum wage will be involved in the survey.

Discussion

Dataset

Based on the survey questions, I randomly simulated 100000 responses associated with the respondent's age, income, and favored party with consideration of the relationship between each variable. Without proper adjustments, the simulated data is likely to misjudge the results by assuming everybody has the same probability to vote for a certain party solely based on last year's voting result. There's a 'positive association between income and conservative voting that high-income group' (Lind J. T., 2010), because one of the campaign strategies is to lower income taxes. In addition, the average salary has a positive relationship with people's age. This pattern is analyzed by using dataset 'Income of individuals by age group, sex and income source, Canada, provinces and selected census metropolitan areas' (Statistics Canada). Based on Ontario seat projection issued by Éric Grenier on Oct 6th, we calculated the probability Ontario residents would vote for each party. The voting probability is 61.6% ,28.9%, 0%, 9.5% corresponding to Liberal, Conservative, Green, and New Democratic Party(NDP). These percentages are set as 'default', then we added the correlation between age group and income into the simulation by adjusting the probability of which political party people under different age would vote for.
People are divided into 8 groups based on age. One thing to point out is since it's hard to adjust the probability for people who answered their age as 'not willing to say', I assigned default probability for voting and Ontario average annual salary without adjustment to the group. After the simulation is set, we randomly selected 300 rows of data from the population with the size of 100000 and without replacement. Below are the five figures constructed based on the sample.

```
## -- Attaching packages ---------------------------------------------------------------------
----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------------------------------------------
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Results

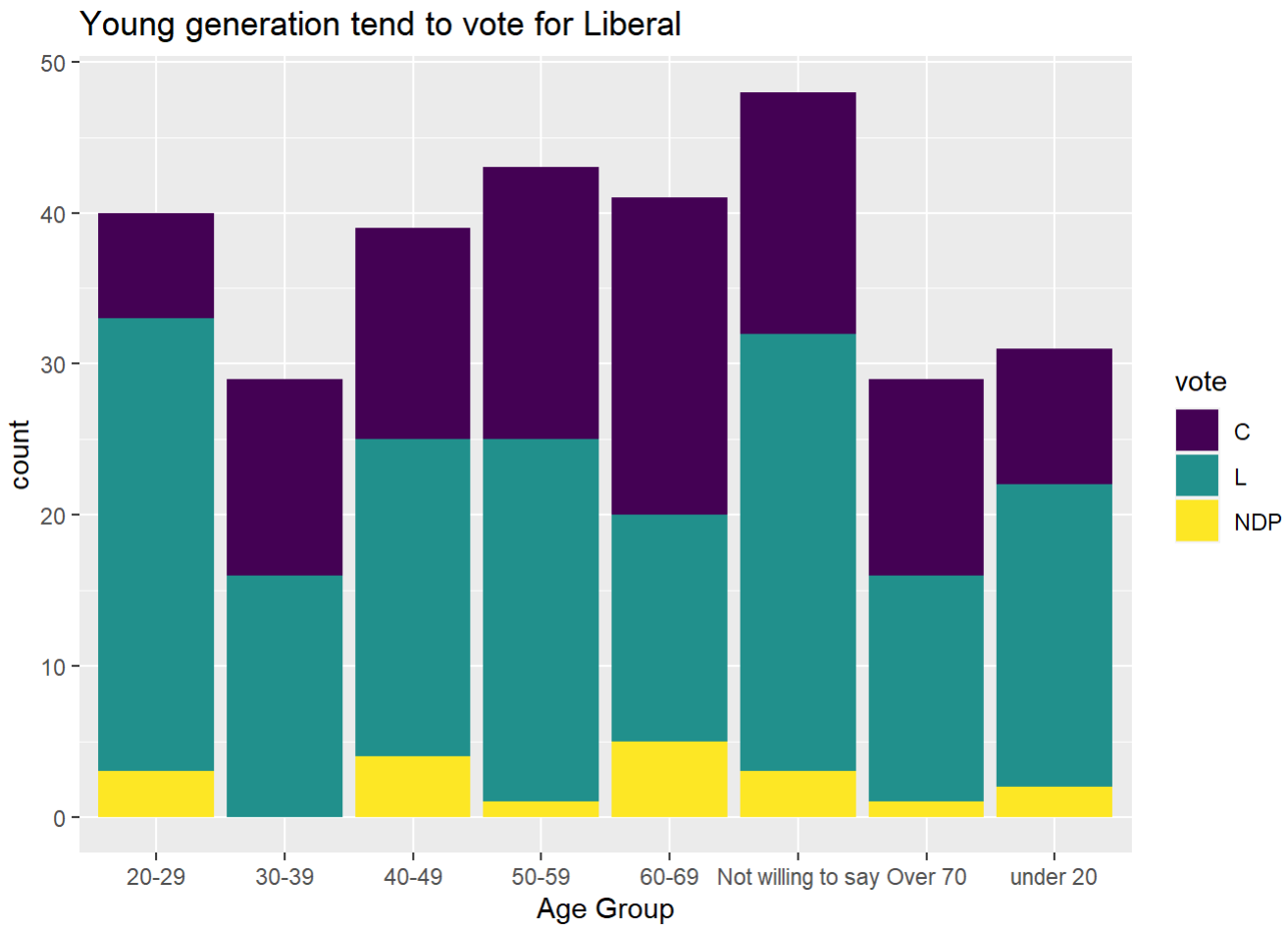


Figure 1:Young generation tend to vote for Liberal

In general, this bar plot illustrates the voting behaviours among various age groups. It is obvious that Conservative and Liberal are the main parties holding the majority of polls. Liberal won the most polls with little advantage than the Conservative by Oct 6th, 2020. Voters under 29 years old have an obvious tendency towards Liberal. Voters

aged over 50 are more willing to choose Conservative.

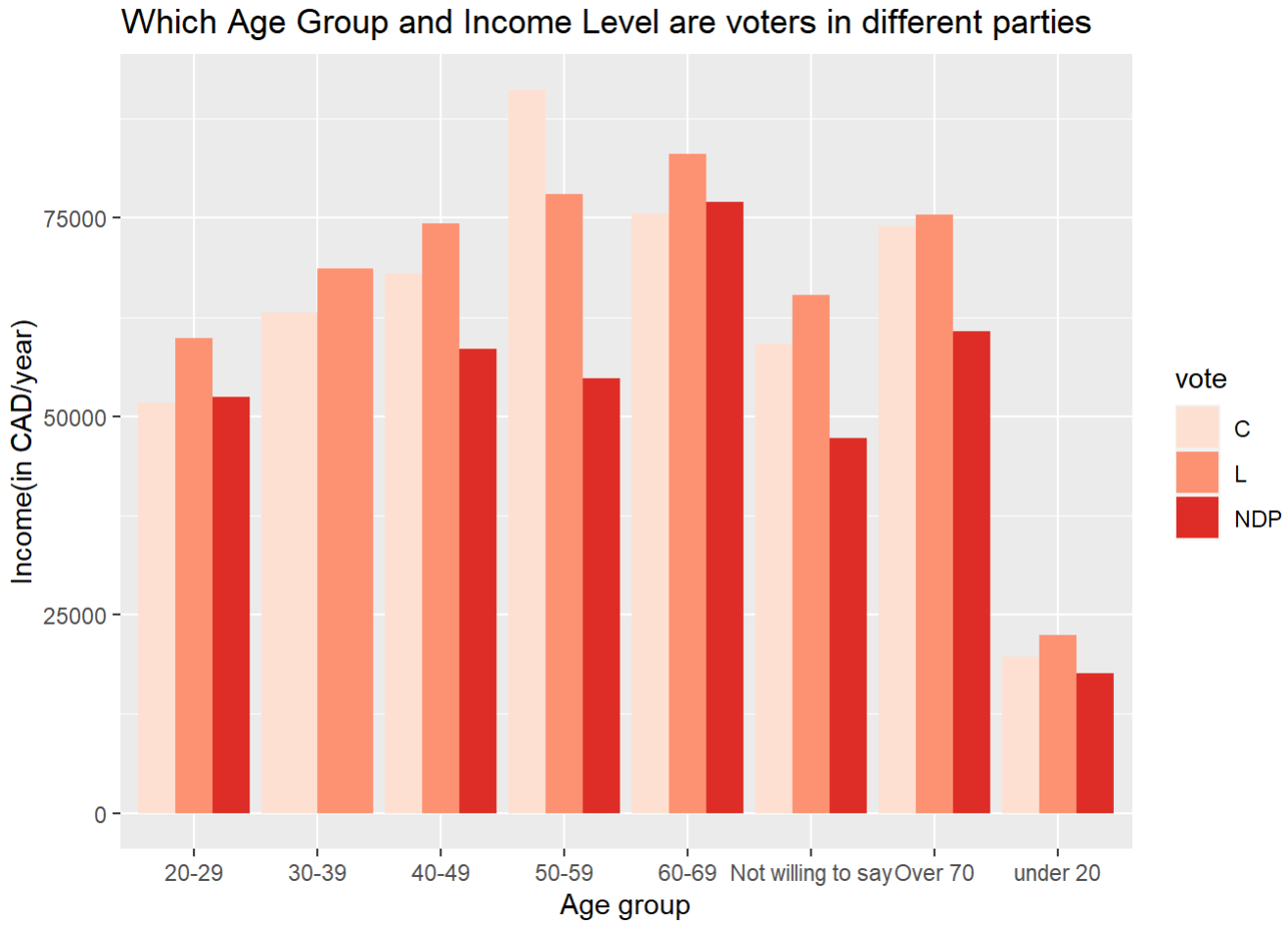


Figure 2: Which Age Group and Income Level are voters in different parties

This is a multiple grouped bar plot attempt to determine which group of voters tend to vote for each party under various age and income circumstances. Political parties respondents voted for are distinguished by color, light pale represents the Conservative, orange shows the Liberal and red represents the New Democratic Party(NDP). From the multiple grouped bar plot, we can conclude that voters who vote for Liberal have the highest income in general, people who vote for NDP have the lowest income and Conservative voters have medium level income. Conservative voters aged from 50-59 years old have higher income than any other party's voter, and the median wage for Conservative voters is above 50000CAD. However, the result may be biased because my sample size is relatively small considering the variance of income distribution among different age groups. Some outliers will significantly influence my results.

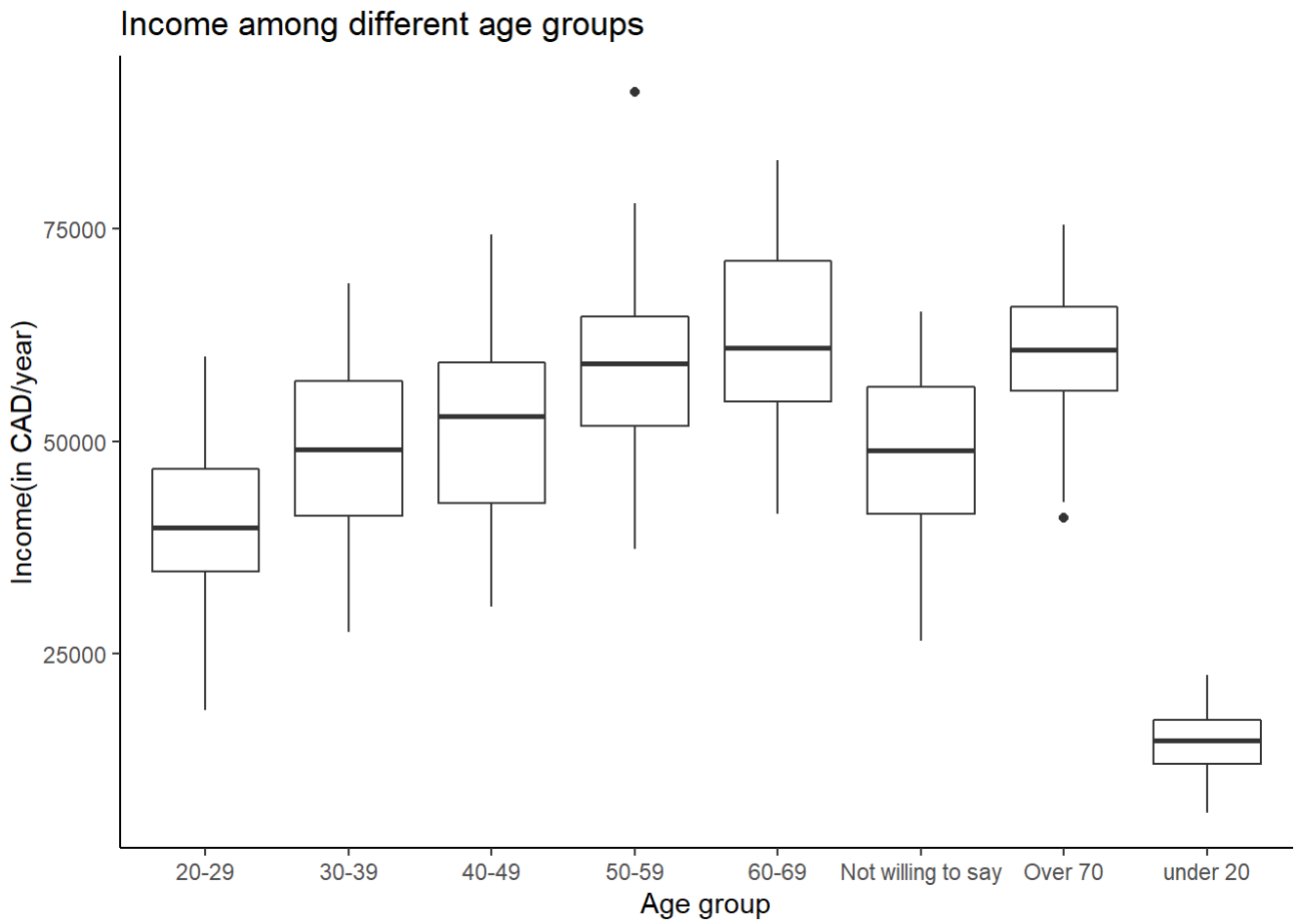
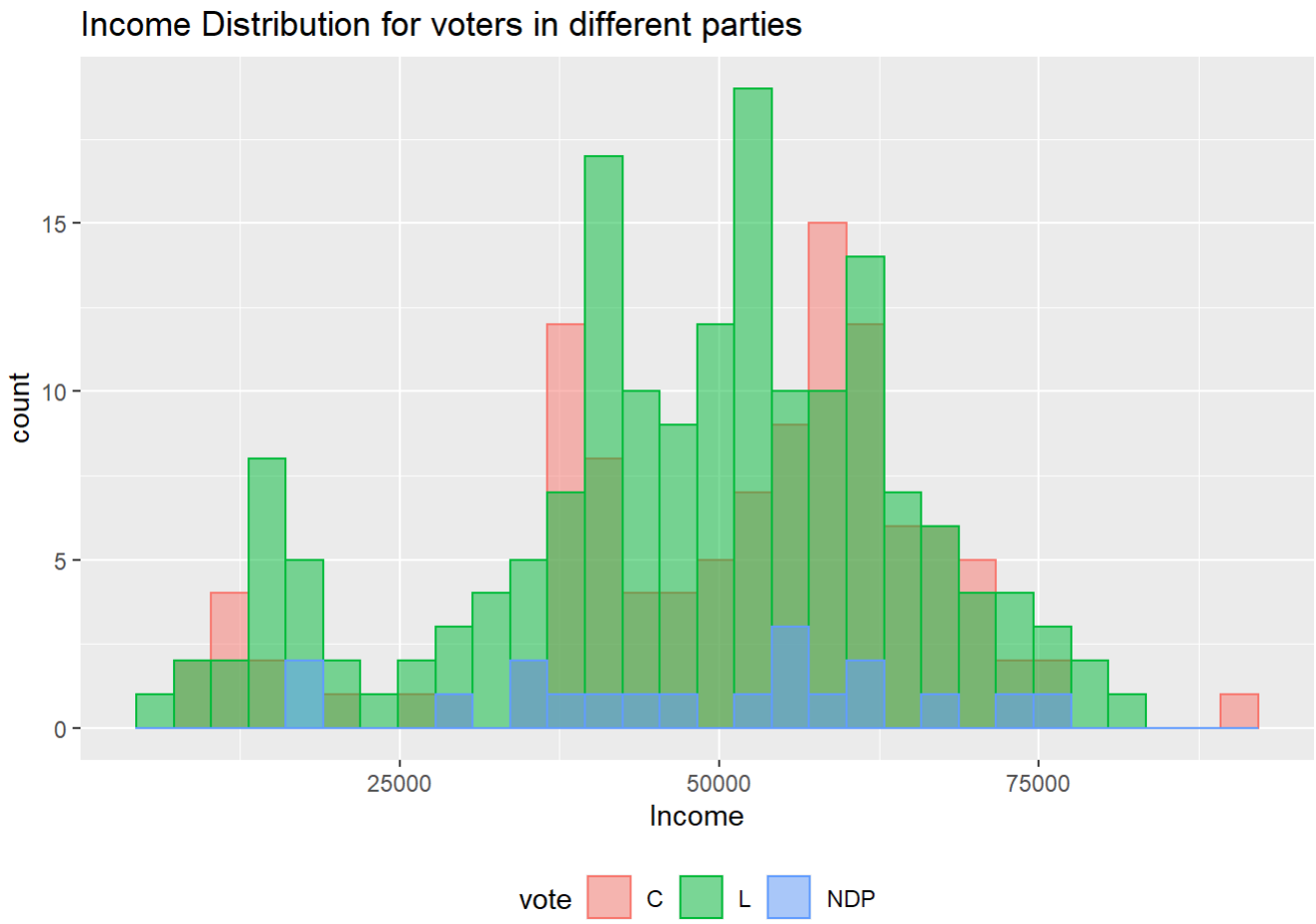


Figure 3: Income among different age groups

The grouped box plot illustrating the relationship between different age groups and their average after-tax income. Respondents whoever answered ‘not willing to say’ are set to have Ontario average income. There is a positive relationship between age and income. People who are under 20 tend to have a mean income at around 15000CAD, which is far away from Ontario residents average level 49000CAD. There is an outlier for people aged from 50-59 who have extremely high income, and there is an outlier for people aged over 70 with slightly low income.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




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

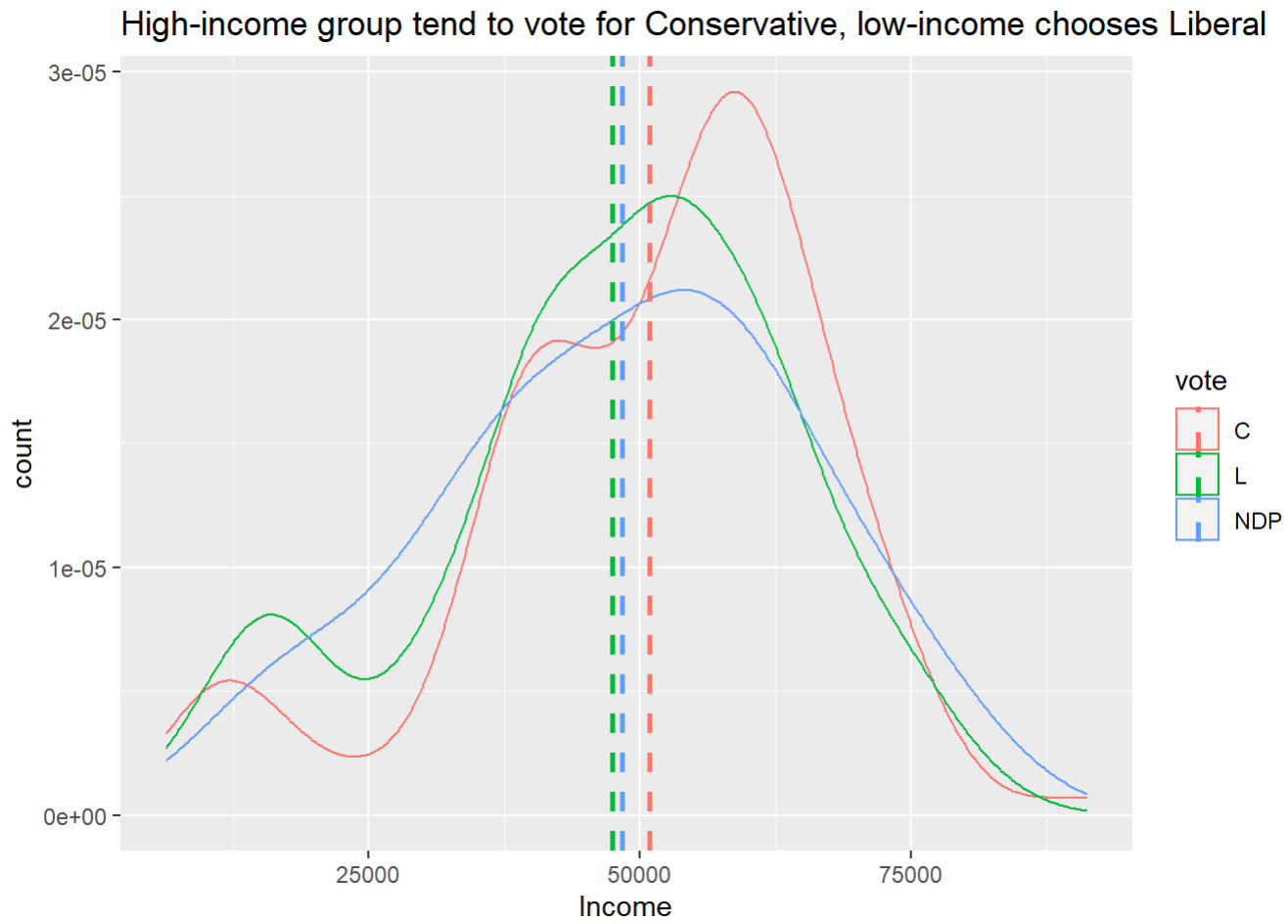


Figure 4: Income distribution for voters in different parties. Figure 5: Low-income group tend to vote for Liberal Party and high-income group is more likely to vote for Conservative. Figure 4 and 5 try to determine the income distribution for voters in different political parties. We can clearly see that Conservative voters have the highest mean income and Liberal voters have the lowest mean income. There's a left-skewed distribution for people who vote for Liberal and the peak is at around 53000CAD. There's a normal distribution for both voters choosing Conservative and NDP. The average income for conservative voters is around 52000 CAD, the value for Liberal and NDP is around 46500 CAD and 49800 CAD respectively.

I summarized the main findings as followed:

Conservative and Liberal are the two main parties holding the majority of polls in Ontario.

People with low income and the young have an obvious tendency to vote for Liberal.

Voters for Conservative have the highest average after-tax income 52000CAD.

Voters aged over 50 are more willing to choose Conservative.

Next Step: How can Conservative attract more voters?

To gain more votes, Conservative could develop campaign strategies that appeal to the young generation and people with low income. For the young, conservative could try to become more contemporary with issues related to young people, and pay more attention to develop the education system. Increasing the Student Grants, loose

repayment period, lower student loan interest rate, and providing more job opportunities for new graduates are potential effective ways to appeal to young voters. Lowering the interest rate and offering more benefits to low-income groups are also useful to gain votes from them. To retain existing voters, Conservative Party has to keep showing up in public and emphasize existing campaign strategies such as lowering the tax rate, increasing the minimum wage to keep the votes from people with high income and the old generation.

Weakness and potential opportunities

```
As our company issued surveys through email, there might be a bias that people who check their e
mail on time are mostly aged less than 55 years old. We may not be able to collect as much data
from people over 55 years old as we expected. Additionally, it is hard to distinguish if the res
pondents won’t lie in the survey. The survey only consists of some basic questions asking desire
d respondents’ age, income, and favored party. Moreover, there might be some lurking variables i
nfluencing the result analysis such as the unemployment rate.
```

In the future, it may be worth adding questions associated with current issues such as how COVID-19 influences income. Conducting surveys through phone calls and mails may increase the reliability of the results as well. Using SRSWOR method with sample size 300 and population size 100000 might be too small because we’re dividing the 300 rows into many groups, thus outliers have a significant influence when analyzing each group’s distribution. In the future, we can increase the sample size to have a more representative sampling.

Discussion appendices that detail the survey

Link to the survey: https://rubytianxiaoma.typeform.com/to/C7oiRv8A (https://rubytianxiaoma.typeform.com/to/C7oiRv8A)

Screenshots of the survey Note: Available answers for question 2: what age category best describes you?

Under 20, 20-29, 30-39, 40-49, 50-59, 60-69,70 and older, I prefer not to say

Available answers for question 3: what’s your after-tax income?

Under $15000, $15000-34999, $35000-54999, $55000-74999, $75000-94999, $95000-114999, Over $115000, I do not work

Canadian Voting Intentions

$5 Giveaway! We'd like to know about how you might vote in a general election.

1 min to complete

Ready press Enter ↵

“

Your Privacy means everything to us! We may require some personal information from you for analyzing purposes. Your data will be protected by us. Please continue if you consent that you are willing to answer the questions in this survey.

Continue press Enter ↵

1→ Let's get started! Are you an Ontario resident? *

A Yes

B No

2 → What age category best describes you? *

Type or select an option

3 → What's your after-tax income? *

Type or select an option

4→ Great!
If a general election is held today, which party would you like to vote for? *

Choose as many as you like

- A Conservative
- B Liberal
- C Green party
- D Ontario New Democratic Party
- E I don't want to vote

5→ Would you mind telling us how you think the current government doing on these issues?

From 1, Not good to me, to 5, I'm satisfied with it...

Continue press Enter ↵

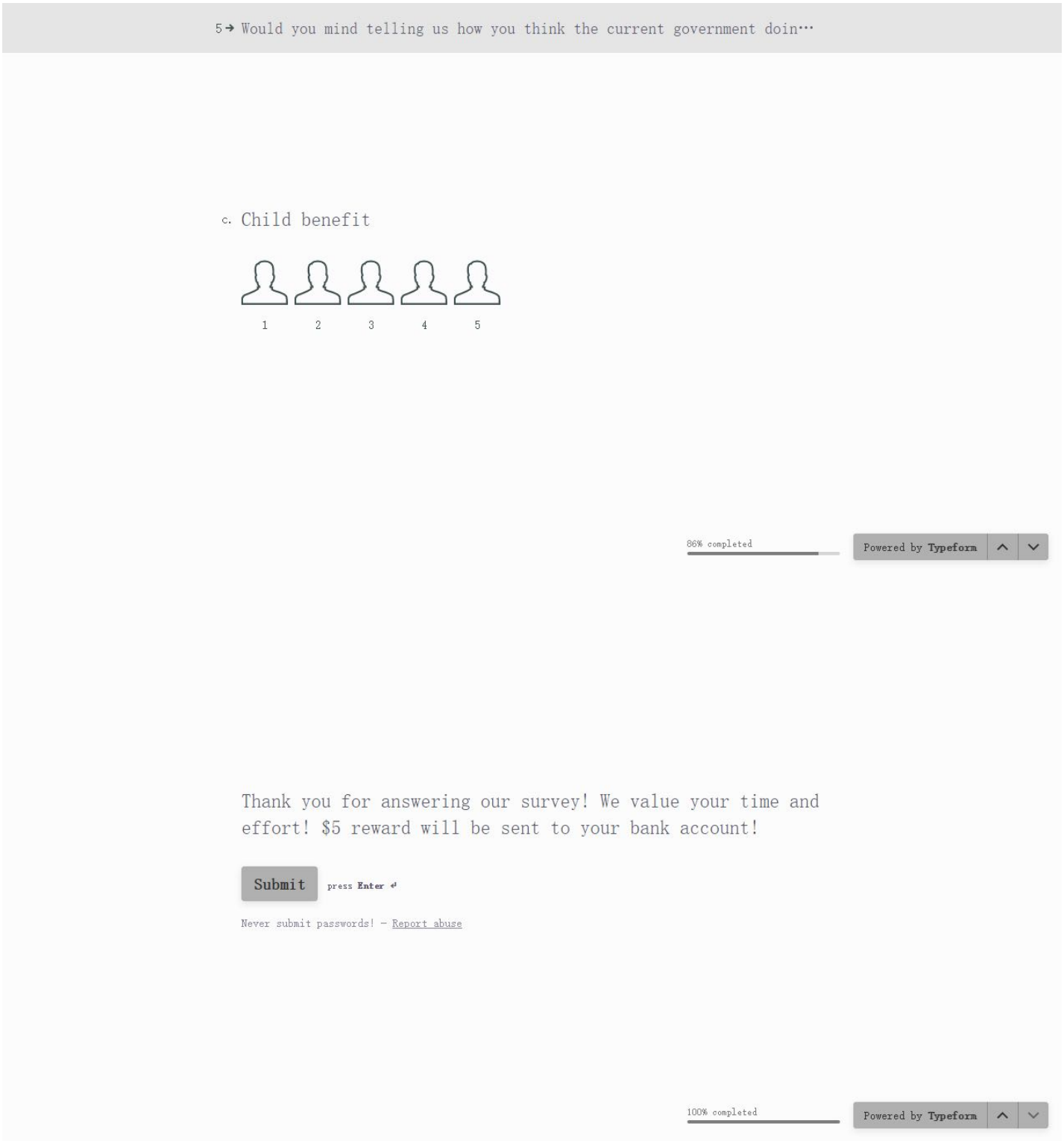
a. Tax rate

1 2 3 4 5

5 → Would you mind telling us how you think the current government doin…

b. Min wage

1 2 3 4 5



References

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ (https://www.R-project.org/).

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686 (https://doi.org/10.21105/joss.01686)

Elin Waring, Michael Quinn, Amelia McNamara, Eduardo Arino de la Rubia, Hao Zhu and Shannon Ellis (2020). skimr: Compact and Flexible Summaries of Data. https://docs.ropensci.org/skimr (https://docs.ropensci.org/skimr) (website), https://github.com/ropensci/skimr (https://github.com/ropensci/skimr).

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Hadley Wickham, Jim Hester and Winston Chang (2020). devtools: Tools to Make Developing R Packages Easier. https://devtools.r-lib.org/ (https://devtools.r-lib.org/), https://github.com/r-lib/devtools (https://github.com/r-lib/devtools).

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.2. https://CRAN.R-project.org/package=dplyr (https://CRAN.R-project.org/package=dplyr)

Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.30.

Statistics Canada. Table 11-10-0239-01 Income of individuals by age group, sex and income source, Canada, provinces and selected census metropolitan areas

Barwell G. (2007) Conservative Party Strategy. In: Wring D., Green J., Mortimore R., Atkinson S. (eds) Political Communications. Palgrave Macmillan, London. https://doi.org/10.1057/9780230286306_3 (https://doi.org/10.1057/9780230286306_3)

Éric Grenier. (2020) “CBC News Canada Poll Tracker.” CBCnews, CBC/Radio Canada,newsinteractives.cbc.ca/elections/poll-tracker/canada/

Éric , Grenier.(2020) “Federal Election 2019 Live Results.” CBCnews, CBC/Radio Canada, newsinteractives.cbc.ca/elections/federal/2019/results/.