Which Region of Canada and Which Advertising Channel Should the Conservative Party of Canada Focus More On to Boost Up Support?

Group 59: Yujia Zheng 1005329299 & Xi Zhang 1004707784

2020-10-06

Non-Technical Executive Summary

Petit Poll has been asked by the Conservative Party of Canada to survey people from all regions in Canada to provide monthly polling updates. We have designed a survey for the upcoming federal election and included some questions addressing different aspects of interest. Firstly, we asked the participants if they are Canadian citizens, we ask for the regions they come from, and the highest education level the participants have completed. Then we are curious about the ways participants used most frequently to get to know election-related information, and we would ask about what aspects respondents think the governing party should make improvements on. Lastly but most importantly, we asked them which political party they would vote for if they are eligible to vote in the upcoming federal election.

We create a table in R to demonstrate the number of participants from each region and the number of participants within each region who would vote for the Conservative Party. In the table, we also evaluate the percentage of votes to the Conservative Party for different regions. We have found that the percentages of votes to the Conservative Party are highest in British Columbia and Nova Scotia and the lowest percentage is in Saskatchewan. This finding suggests that the Conservative Party members should maintain their votes in British Columbia and Nova Scotia. In the meantime, they are supposed to spend more time on advertising in Saskatchewan to boost up the votes there. Furthermore, we plot the distribution of the most frequently used channels that participants use to known about election-related information. We found that social platforms including Facebook, Twitter are the most popular tool used by the respondents, and television is also adorable. This finding is essential because it implies that the Conservative Party members should be more active on social media and television so that more audiences would get exposure to them.

There are still some existing issues and weaknesses in our sampling method, survey, and analysis. The sampling method we used is Simple Random Sample Without Replacement. This specific method involves the disadvantage that it is extremely difficult to have a full list of phone numbers of our population. In addition, our sample size is 166 and this is small in comparison with the whole population. Thus, the bias may be built into the result and our data may lead to an inaccurate conclusion. Also, the survey includes some issues that should be tackled. Firstly, the provided answers to the survey questions may not cover all the choices the participants are thinking of. Secondly, the questions are designed such that participants can only pick one answer to each of the questions. This makes it convenient to create the dataset and analyze the data, but reduces the diversity of data and causes loss of information. To tackle the issue of small sample size, we may keep collecting more observations and add them to the dataset. This procedure may require some additional sampling techniques. In addition, the designed questions can be developed so that the participants are allowed to pick more than one option and write down their own answers that are not provided.

Introduction

The federal election in Canada has drawn the eyes of people around the world. While people have different attitudes toward the election because they have different characteristics. In this paper, we use R to clean and tidy the data of responses collected from the survey conducted for the federal election, and we are interested in investigating how percentages of votes to Conservative Party differ between regions, and which channels are most frequently used by participants to get to know election-related information.

In this report, we mainly focus on three variables. Firstly, PARTY_OF_INTEREST is a categorical variable and it records which party the participant will vote for if he is eligible to vote in the federal election. Secondly, WAY_TO_GET_INFO is a categorical variable and indicates the ways participants use to get to know election-related information. Furthermore, REGION is a categorical variable and indicates which region of Canada the participant is from. We have found that the percentages of votes to the Conservative Party are highest in British Columbia and Nova Scotia and the lowest happens in Saskatchewan. In addition, the participants use social platforms (Twitter, Facebook, Instagram, etc) most frequently to get to know the information about political parties and elections.

In this report, firstly there is a Research Methodology section, which addresses questions about our population, frame, sample, sampling method, estimation of the cost, issues of non-response, and privacy. There is a section of Data and Results. It includes the source and description of our data, it also includes the plots and a brief explanation of variables. Then, the Discussion section contains the features of the variable and relationship which we are interested in, and their implications, which are all explained in detail. The Weaknesses and Next steps section talk about the existing issues of our survey, data, and analysis, and how we can resolve them in the future. The Appendix includes a link to our survey and screenshots of it. And finally, there are the References for all R packages and other tools and resources we used in this paper.

Note: The code to generate this report is saved in "Conservative.Rmd" and is available on my Github page: https://github.com/YujiaZheng/Conservative.git

Survey Methodology

1. Population frame and sample

Population: We are targeting the population who are Canadian citizens (i.e. are eligible to vote), and care about the Candian federal election. Sampling Frame: Our sampling frame is a list of Canadian phone numbers. We develop a survey through the platform "Google Surveys". Then, we send the owners of phone numbers the link to the survey. Sample: 166 Canadian citizens actually observed

2. Sampling method

We choose to use Simple Random Sample Without Replacement (SRSWOR). The reason for our choice: We need to randomly select phone numbers from our sampling frame, i.e. the list of Canadian phone numbers. If we have a large enough list of phone numbers, this SRSWOR method would ensure each of the individuals has an equal chance to be selected. Important statistical properties of SRSWOR: The sample mean (y_bar) is a design-unbiased estimator of the population mean The design-based variance of sample mean (y_bar) is given by: Var(y_bar) = (1-n/N)*(Var(y)^2/n). By using this formula, it is possible for us to find more than one optimum value of the parameter. In this case, the bias of the estimator (y_bar) can also be controlled

3. How to reach our desired respondents

We develop a survey through the platform "Google Surveys". Then, we send the link of the survey by messages to the people selected who own Canadian phone numbers.

4. Estimation of the cost

Taking the issue of non-response into consideration, we plan to send 1000 messages to communicate the details with the selected people (in order to actually get a not very small sample size, and in fact we get 250

responses). This may cost about 10 dollars. Moreover, we may value the time we have devoted to sending the messages at, say 100 dollars. This is a kind of opportunity cost.

5. Steps to deal with non-response and how non-response may affect the survey

Many of the selected people do not respond to our survey. To be more specific, only 150 of the 1000 individuals completed the survey. To deal with the non-response, we resend the link of the survey to 100 people. Finally, we get 250 respondents in total, and we filter 166 Canadian citizens to create a dataset to support analysis. By looking at the responses, we find that these 166 people come from various regions across Canada, complete various education levels, and have different attitudes towards the election.

In terms of how non-response will affect our survey, if it can be assumed that the 166 people who responded have identical characteristics as the rest of the population, then non-response will not affect the result of our survey. However, if people who responded are different from people who did not respond in some way, then the sample formed by those who respond may not represent the entire population, and thus would introduce bias to our results, for example, misleading estimates of the mean of education level.
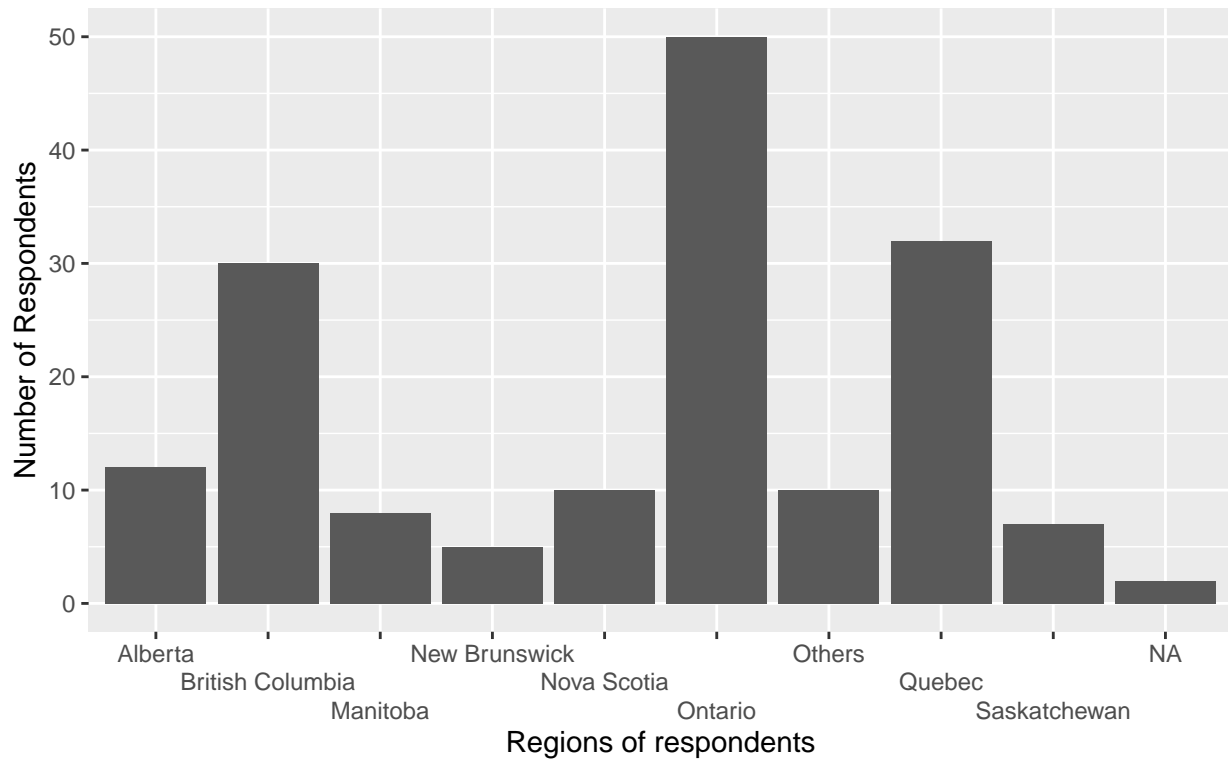
6. Protection of respondent privacy

Firstly, at the time of data collection, we will inform our potential respondents that 1) our collection has legal authority; 2) the purpose of using personal information in this survey and the following analysis; 3) the title, the business address, and contact phone number of our company (Petit Poll). The survey is completely anonymous. All the phone numbers will not be released to the public. And in the actual dataset we created for analysis, we will assign each respondent a unique ID to identify them.

Data and Results

```
## # A tibble: 6 x 7
##       ID IS_CITIZEN REGION EDU_LEVEL WAY_TO_GET_INFO ASPECT_TO_IMPRO~
##    <int> <chr>      <chr>      <dbl> <chr>           <chr>
## 1     1 Yes        Ontar~         3 Public radio/T~ Education
## 2     2 Yes        Others        NA Social platform Economy/Unemplo~
## 3     3 Yes        Quebec         4 Public radio/T~ Education
## 4     4 Yes        Briti~         3 Social platform Education
## 5     5 Yes        Briti~         1 People around   Healthcare
## 6     6 Yes        Briti~         3 Social platform Healthcare
## # ... with 1 more variable: PARTY_OF_INTEREST <chr>
```

Above is an overview of our survey_result dataset. The data was obtained from our designed survey called "Upcoming Federal Election Survey". Data was collected from 250 participants, including non-citizens, but only the 166 citizens are filtered to be included in the final dataset survey_data. There are in total 7 variables in the dataset. ID is a unique number assigned to the respondent. IS_CITIZEN is to indicate that the respondent is a Canadian citizen. And other variables: REGION, EDU_LEVEL, WAY_TO_GET_INFO, ASPECT_TO_IMPROVE, and PARTY_OF_INTEREST are introduced with a plot and a brief explanation below.

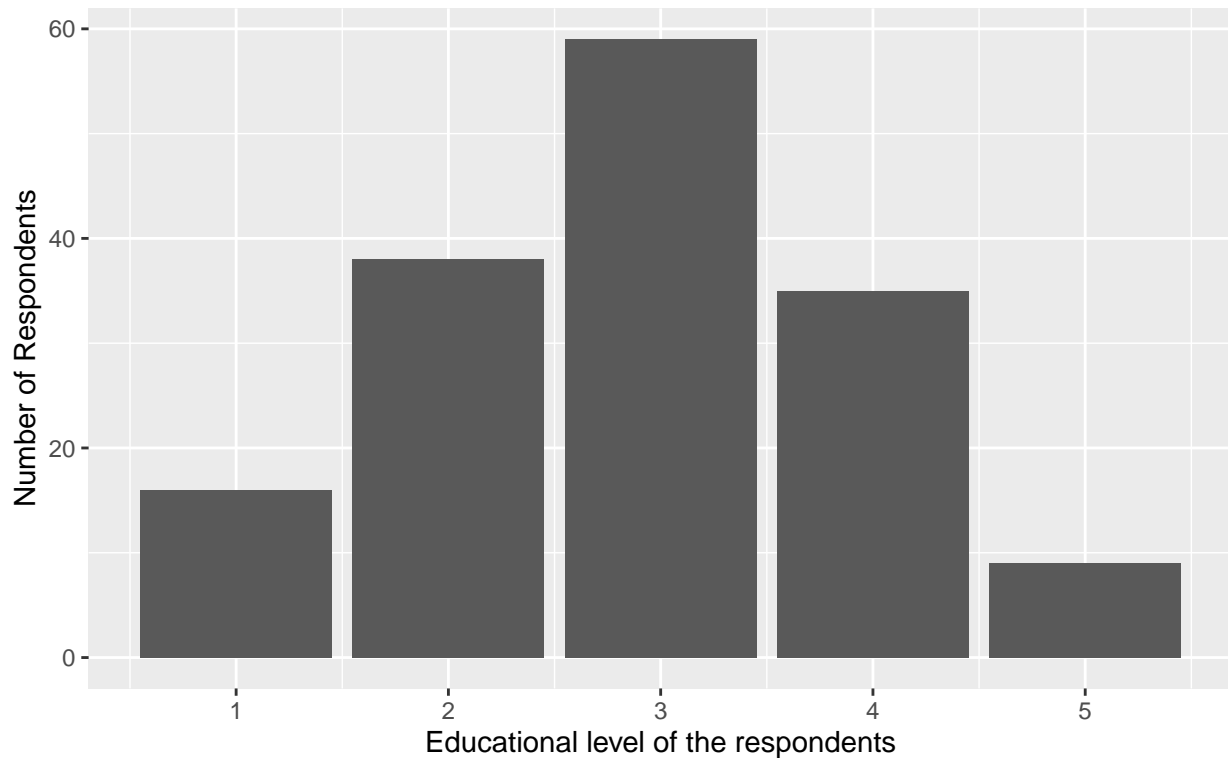Figure 1: Distribution of the regions of respondents

REGION is a categorical variable that indicates which region of Canada the respondent is from. The participants from Ontario occupy the largest proportion of total respondents (50 of them are from Ontario). Participants from British Columbia account for the second-largest and there are 30 participants from this province. The New Brunswick participants account for the smallest percentage of the total and there are 5 participants from this region.

Figure 2: Distribution of the educational level of the respondents

EDU_LEVEL is a numeric variable that takes value from 1 to 5. 1 represents an educational level below high school, 2 represents a high school diploma, 3 represents a Bachelor's degree, 4 represents a Master's degree, and 5 represents a Ph.D. or higher. The respondents are divided into 5 groups by the highest education level they have completed. The group of participants who obtained a Bachelor's degree is the largest group while the group representing a Ph.D. degree is the smallest one. The distribution of the education level of the participants is bell-shaped and roughly looks like a Normal distribution.

Figure 3: Distribution of the ways to get information about political parties and elections
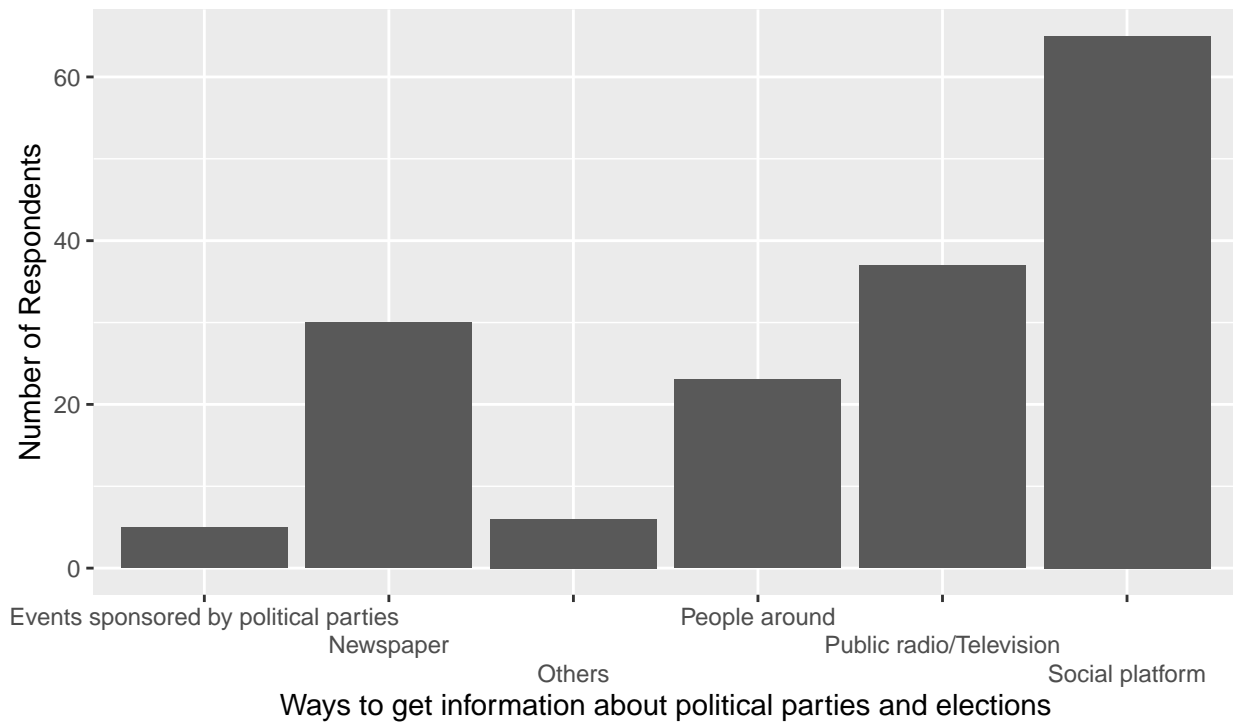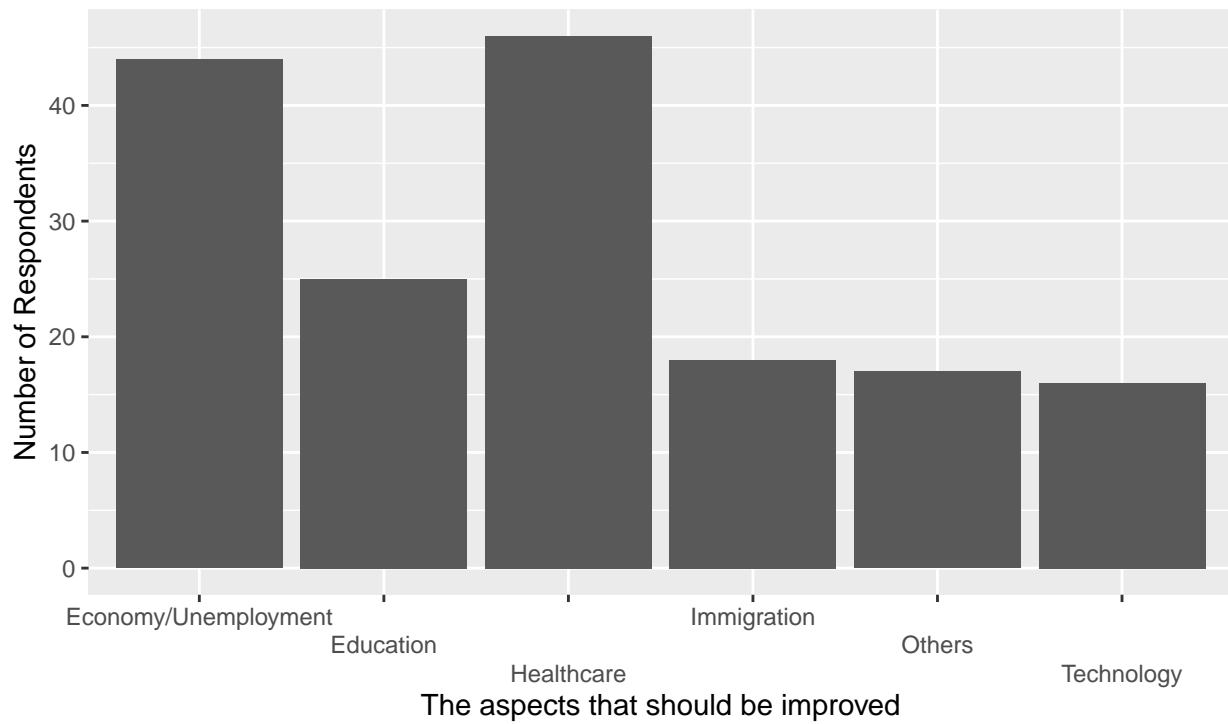
Figure 3: Distribution of the ways to get information about political parties and elections

WAY_TO_GET_INFO is a categorical variable and indicates the ways participants use to get election-related information. The participants use social platforms (Twitter, Facebook, and Instagram, etc) most frequently to learn about politics and election issues. Public radio or television is popular as well. While the least people prefer to get information about elections through the events sponsored by political parties.

Figure 4: Distribution of the aspects that respondents think should be improved

ASPECT_TO_IMPROVE is a categorical variable that represents the aspects respondents think the governing party should make improvements on. The two most important aspects the participants think should be improved are Economy/Unemployment and Healthcare. While the numbers of respondents who think the technology, immigration, education, and other aspects should be improved are approximately equal.

Figure 5: Distribution of the parties respondents would vote for

PARTY_OF_INTEREST is a categorical variable and it records which party the participant will vote for if he is eligible to vote in the federal election. The survey result reflects that the respondents who would like to vote for the Liberal Party of Canada occupy the largest proportion of all the respondents. While the respondents who tend to vote for the Conservative Party are the second largest group. This implies that the largest competitor the Conservative Party is facing is the Liberal Party. Only a very small proportion of the respondents support the Green Party or other parties. Note that this result is almost consistent with the results of the 2019 Canadian Federal Election (see link to results in Reference).

Discussion

Discussion of the survey questions:

1. Are you a Canadian citizen?

This question is just to help us filter all the citizens to create our dataset.

2. Which region of Canada are you from?

This question is helping to investigate the geographical locations of the supporters of each political party. The result would give implications to the Conservative Party members to figure out which regions they should make more efforts on.

3. What is the highest level of education you have obtained?

This question aims to understand the educational levels associated with the participants who prefer to vote for the Conservative Party. The result would give implications for members of the Conservative Party to target their supporters.

4. In which way do you usually get information about political parties and elections? Pick the one you get the most information from.

This question is going to help understand how people in Canada get exposure to the issues regarding political parties. The result would guide the members of the Conservative party to choose proper ways to advertise, give speeches, and win over the polls.

5. From your perspective, what is the most important problem facing Canada today?

This question is helping to understand participants' opinions towards the most important problem facing Canada. The result would give implications for Conservative Party members to learn about which type of problem they should resolve in order to win the majority of polls in the election.

6. If you are eligible to vote in the upcoming federal election, which party are you going to vote for?

This question is to help with understanding which political party the participants would vote for in the upcoming federal election. The result gives implications for Conservative Party members to learn about their largest competitors and then react properly.

Discussion of the variable and relationship we are interested in:

The discussion of the general survey data is in the Result section, and the variable and relationship will be discussed in detail here.

The variable that we are interested in is WAY_TO_GET_INFO.

In Figure 3 (see Data and Results Section), we plot the distribution of ways to get information about political parties and elections. Generally speaking, there are five main ways for the respondents to get exposure to political issues. These include events sponsored by political parties, newspapers, people around (family members, friends, etc.), public radio/television, and social platforms (Twitter, Facebook, etc.). Among them, social platforms are the most popular tool used by the respondents. To be more specific, there are almost 70 of the 166 selected respondents using social platforms to get to know election-related information. Furthermore, many respondents prefer to get to know the election and political parties through television programs as well. While the least respondents prefer attending the events sponsored by political parties, only 5 of 166 respondents tend to learn about the election-related information through the events. This result also gives important implications to the Conservative Party about their advertising channels. That is, the Conservative Party members are supposed to spend more money and time on advertising through social platforms and television programs. While sponsoring a public event is not recommended since it is expensive and not attractive enough to potential voters.

```
##            Region No_Respondents No_Respondents_Conserv   Percent
## 1         Alberta             12                      4 0.3333333
## 2  British Columbia           30                     15 0.5000000
## 3        Manitoba              8                      3 0.3750000
## 4     New Brunswick            5                      2 0.4000000
## 5      Nova Scotia            10                      5 0.5000000
## 6          Ontario            50                     11 0.2200000
## 7           Others            10                      2 0.2000000
## 8           Quebec            32                      8 0.2500000
## 9     Saskatchewan             7                      1 0.1428571
## 10            <NA>             2                      1 0.5000000
```

Table 1: The percentage of respondents who would vote for Conservative party in each region

We are also interested in the relationship between the regions of the respondents and which party they would like to vote for.

In Table 1, we used our survey results to summarize the number of respondents from each region, and the number of respondents in each province who would like to vote for the Conservative Party of Canada. Then, we calculated and showed the percentage of the votes to the Conservative Party in each region of Canada.

According to the table, we find that 50% of the participants from British Columbia and Nova Scotia would like to vote for the Conservative Party of Canada. This attains the highest percentage among all the regions in Canada. Also, 40% of the participants from New Brunswick tend to vote for the Conservative Party. Such a percentage is also adorable. Furthermore, respondents from Saskatchewan gave the lowest percentage of votes to the Conservative Party, which is only 14.3%.

The summary and analysis of the result give some important implications to the policymakers of the Conservative Party. The support from British Columbia and New Brunswick are already pretty optimistic, so the Conservative Party should keep up in these regions. In addition, the Conservative Party should make more efforts in the regions where they gained less support, for example, Ontario, Saskatchewan, and Prince Edward Island. They may allocate more resources and do more advertising in these regions.

Weaknesses and Future Steps

Some weaknesses of our sampling method, survey, and analysis:

- The sampling method we used is Simple Random Sample Without Replacement (SRSWOR). This specific method entails some disadvantages: It is extremely difficult to have a full list of phone numbers of our population, and it is time-consuming to create a list that is as large as possible
- There could be bias built into our result if our sample size is not large enough, or our sample cannot represent our full population
- Some of the potential answers may not cover all the choices that respondents are thinking of
- The questions are designed such that respondents can pick only one option, in order to make the creation of the dataset and the analysis more convenient, but this may result in the loss of information
- The sample size 166 is probably not large enough, thus our data may not be sufficient, the result and analysis based on the dataset may not be accurate and bias could have occurred.

Future steps we could take to resolve the issues:

- If the sample is not representative enough, we should keep collecting observations to get a larger sample, and some additional sampling techniques may be required
- We could allow respondents to choose more than one options for some questions, in order to get more complete information

Appendices

The link to our survey: https://docs.google.com/forms/d/1XQXlaHRRknBqXOYidlZOn_zdcrYUU2LQliT8r5pMzSg/edit

Screenshots of our survey:

Upcoming Federal Election Survey

In this survey, we will ask about some of your personal information, the channel you get to know the election-related information, your suggestion to the governing party, and the political party you will vote for. Note: we will not share any of your personal information with people other than our survey team.

1. Are you a Canadian citizen?

○ Yes

○ No

2. Which region of Canada are you from?

○ Ontario

○ British Columbia

○ Quebec

○ Alberta

○ Manitoba

○ Nova Scotia

○ New Brunswick

○ Saskatchewan

○ Others (Northwest Territories, Yukon, Nunavut, Prince Edward Island, Newfoundland and Labrador)

○ Prefer not to Answer

3. What is the highest level of education you have obtained?

○ 1: Below high school

○ 2: High school diploma

○ 3: Bachelor's degree

○ 4: Master's degree

○ 5: PH.D. or higher

○ Prefer not to answer

4. In which way do you usually get information about political parties and elections? Pick the one you use most frequently

○ Newspapers

○ Public radio/Television

○ Events sponsored by political parties

○ People around: family, friends, etc.

○ Social platforms: Twitter, Facebook, Instagram, etc.

○ Others

::::

5. From your perspective, which of the following aspects do you think the governing party should make improvements on?

○ Immigration

○ Education

○ Economy/Unemployment

○ Technology

○ Healthcare

○ Others

::::

6. If you are eligible to vote in the upcoming federal election, which party are you going to vote for?

○ Liberal Party of Canada

○ Conservative Party of Canada

○ Bloc Quebecois

○ New Democratic Party

○ Green Party of Canada

○ Others (Independent, Communist, Nationalist)

○ Prefer not to Answer

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/.
- Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
- Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.30.
- Yihui Xie (2015) Dynamic Documents with R and knitr. 2nd edition. Chapman and Hall/CRC. ISBN 978-1498716963

- Yihui Xie (2014) knitr: A Comprehensive Tool for Reproducible Research in R. In Victoria Stodden, Friedrich Leisch and Roger D. Peng, editors, Implementing Reproducible Computational Research. Chapman and Hall/CRC. ISBN 978-1466561595
- Results of the 2019 Canadian federal election. (2020, September 28). Retrieved October 07, 2020, from https://en.wikipedia.org/wiki/Results_of_the_2019_Canadian_federal_election