2020-toronto-election-prediction

Group 62; Minhui Pan, Sizhen Wu, Haoting Yu

10/8/2020

Part 1. Summary

Significant changes have happened in Canada since last year's federal election. Since the COVID-19 outbreak, the federal government and local leaders have placed different measures and contingency plans to react to the pandemic. We are working with the liberal party to investigate Toronto residences' (excluding the GTA area) opinion on the office's work in the past year and to reflect on the effectiveness of their reactive policies during the pandemic outbreak. The current survey simulated 500 participants' responses as our representative sample. The distribution of the simulation is designed based on demographic layouts according to the 2016 census Canada (Census Canada, 2016). Simulation results were analyzed according to gender, age, racial groups, and income. The results showed that most voters in Toronto are satisfied with the government's reactive measures to COVID-19 and think that they are very efficient. They favour the Liberal party over the others, predominantly Asian and European groups. The liberal supporters are constituted by female Europeans and Asians, whose household income is below 40,000 dollars per year.

Part 2. Introduction

In this opinion polling, we are investigating the political interest of Toronto residences eligible to vote if an election was held in 2020. Toronto has always been the top supporter of the liberal party. Over the past few elections, the liberal leaders had an absolute majority of all districts (wiki). Last year, the liberal party won all 23 out of 23 districts in the city of Toronto. However, as the national polling suggests, the support rate for the two most powerful parties, the liberal and the conservative, in Canada has been gradually declining on a national level. We are interested in investigating whether this trend is present in Toronto and whether the liberal leaders in Toronto should be cautious of people's attitude change after the pandemic.

With that in mind, our survey includes questions that investigate the pandemic's effect on electors' evaluation, both emotionally and rationally, of current policies.

Part 3.Survey Methodology.

Survey questions consist of gender, racial group by global geography, age, income range, personal experience in the pandemic, and government measures evaluation. We are not collecting data involving address, contact, or other personal information. There are seven questions in total, requiring approximately two minutes to finish; participants will have up to one week to complete the survey with a maximum one attempt, until it closes after one month (starting from October 02, 2020). We are using online panels to post survey link and scan code on social media sites, such as Twitter and Facebook via official government account and local district websites. Social media users are predominantly younger adults (We chose Twitter and Facebook because their user average user age is higher than other sites). Therefore, we will also put up posters in local grocery stores, such as Walmart and Shoppers, to get participants from various age groups.

This study stimulated a representative sample of 500 participants; responses to each question are generated separately and then combined into a collective dataset. A sample with a population of 6,196,731 (worldpopulationreview, 2020), a margin of error of ± 5%, and a confidence rate of 95%, would have an effective sample size of 385 subjects ().

Under our best response rate assumption, this study uses a sample size of 500 for simulation. Since we are using non-probability sampling, no margin of error can be associated with this survey. For comparison purposes, a probability sample of 500 participants with this population would have a margin of error ± 4%. Studies typically use weighted samples to control the non-response bias for small sample sizes; however, we are simulating a dataset much larger than the effective sample size. Therefore, to avoid confusion and preserve the data's simplicity, we did not include any weighting method here. On the other hand, to ensure the sample's representativeness, the simulation is designed according to the probability distribution of each variable in the 2019 election results posted on the Global Mail ().

Simulate an appropriate number of responses to your survey. Don't forget to simulate in relation to the survey methodology that you proposed.

method: non-probablity sampling simulate population N = 500

```
library(tidyverse)
```

```
## -- 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   1.4.0      v forcats 0.5.0
## -- Conflicts -------------------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
# Simulation result
set.seed(500)

gender <- c("Male","Female")
d1_gender<-sample(gender,500,replace=TRUE,prob=c(0.42,0.58))
head(d1_gender)
```

```
## [1] "Male"   "Male"   "Male"   "Female" "Male"   "Female"
```

```r
ethnicity <- c("Aboriginal","Asian","African","Caribbean","European","SouthAmerican","Oceania")
d2_ethnicity<-sample(ethnicity,500,replace=TRUE,prob=c(0.01,0.43,0.05,0.06,0.4,0.03,0.02))
head(d2_ethnicity)
```

```
## [1] "Asian"         "Asian"         "SouthAmerican" "Asian"
## [5] "Asian"         "Asian"
```

```r
age <- c("18-24","25-34","35-44","45-54","55-64","65+")
d3_age<-sample(age,500,replace=TRUE,prob=c(0.11,0.17,0.14,0.18,0.13,0.15))
head(d3_age)
```

```
## [1] "45-54" "35-44" "65+"   "45-54" "18-24" "35-44"
```

```r
househould_income <- c("<$19,999","$20,000-$39,999","$40,000-$59,999","$60,000-$79,999",">$80,000")
d4_income<-sample(househould_income,500,replace=TRUE,prob = c(0.37,0.26,0.18,0.1,0.09))
head(d4_income)
```

```
## [1] "<$19,999"         "<$19,999"         "<$19,999"         "<$19,999"
## [5] "$20,000-$39,999" "$60,000-$79,999"
```

```r
effect <- c("Positive","Somewhat Positive","Somewhat Negative","Negative","Prefer not to say")
d5_effect<-sample(effect,500,replace=TRUE, prob = c(0.48,0.41,0.09,0.01,0.02))
head(d5_effect)
```

```
## [1] "Positive"          "Somewhat Negative" "Positive"
## [4] "Positive"          "Somewhat Positive" "Positive"
```

```r
efficiency <- c("Very efficient","Somewhat efficient","Neither","Somewhat inefficient","Very inefficient
d6_efficiency<-sample(efficiency,500,replace=TRUE, prob = c(0.27,0.46,0.14,0.09,0.03))
head(d6_efficiency)
```

```
## [1] "Somewhat efficient" "Somewhat efficient" "Somewhat efficient"
## [4] "Somewhat efficient" "Somewhat efficient" "Very efficient"
```

```r
vote <- c("Liberal","New Democrat","Green","Others","Conservative","People's party")
d7_vote<-sample(vote,500,replace=TRUE,prob = c(0.56,0.22,0.07,0.02,0.12,0.01))
head(d7_vote)
```

```
## [1] "Liberal"      "New Democrat" "Liberal"      "Liberal"      "Liberal"
## [6] "Liberal"
```

```r
#Form a data set
my_data <- data.frame("Gender"=d1_gender,"Ethnicity"=d2_ethnicity, "Age_Group"=d3_age, "Household_incom
head(my_data)
```

```
##   Gender      Ethnicity Age_Group Household_income       Life_Impact
## 1   Male          Asian     45-54         <$19,999          Positive
## 2   Male          Asian     35-44         <$19,999 Somewhat Negative
## 3   Male  SouthAmerican       65+         <$19,999          Positive
## 4 Female          Asian     45-54         <$19,999          Positive
## 5   Male          Asian     18-24 $20,000-$39,999 Somewhat Positive
## 6 Female          Asian     35-44 $60,000-$79,999          Positive
##      Satisfaction_Lvl         Vote
## 1 Somewhat efficient      Liberal
## 2 Somewhat efficient New Democrat
## 3 Somewhat efficient      Liberal
## 4 Somewhat efficient      Liberal
## 5 Somewhat efficient      Liberal
## 6     Very efficient      Liberal
```

```r
#Note: Each column is not related
```
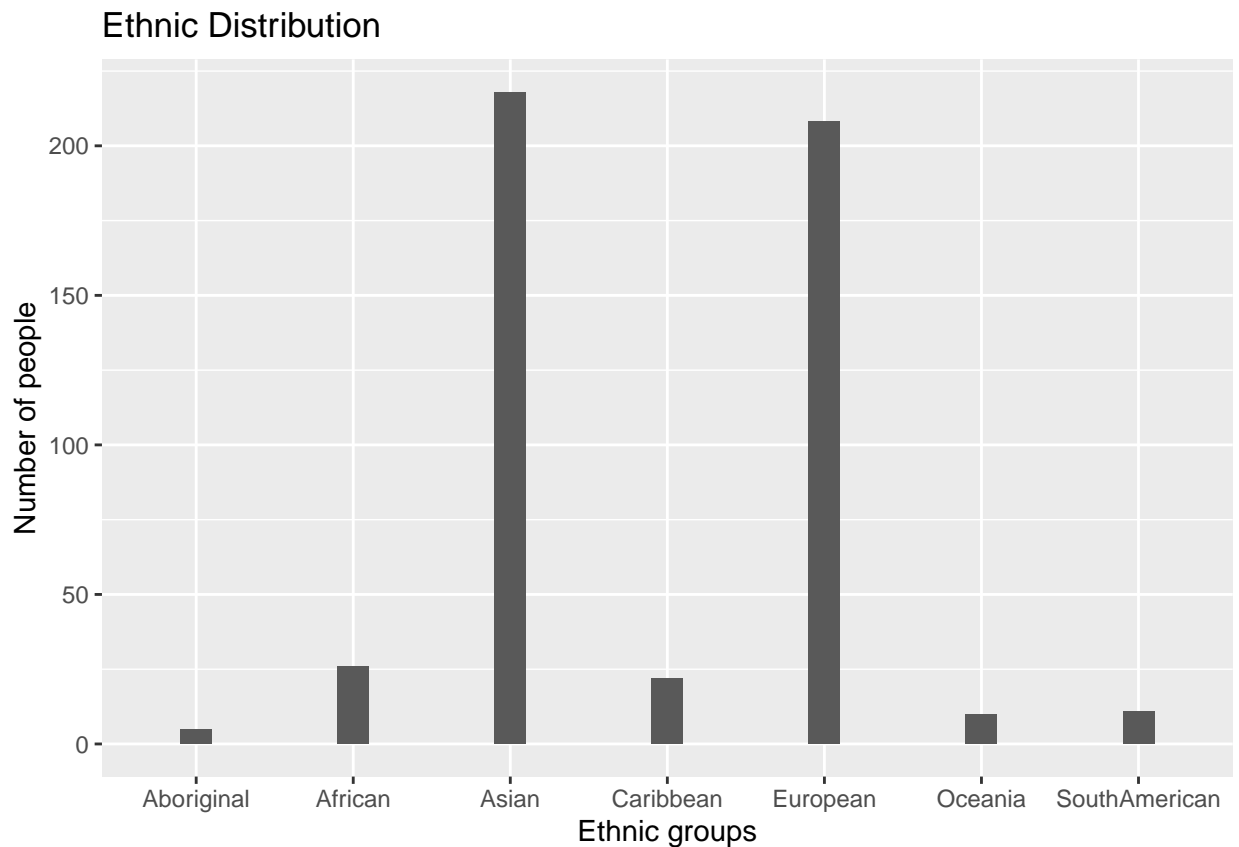
```r
#Plot data to get a overview
ggplot(my_data, aes(x = d1_gender)) + geom_bar(width=0.2)+labs( x = "Gender", y = "Number of people",
                                                title ="Gender Distribution")
```
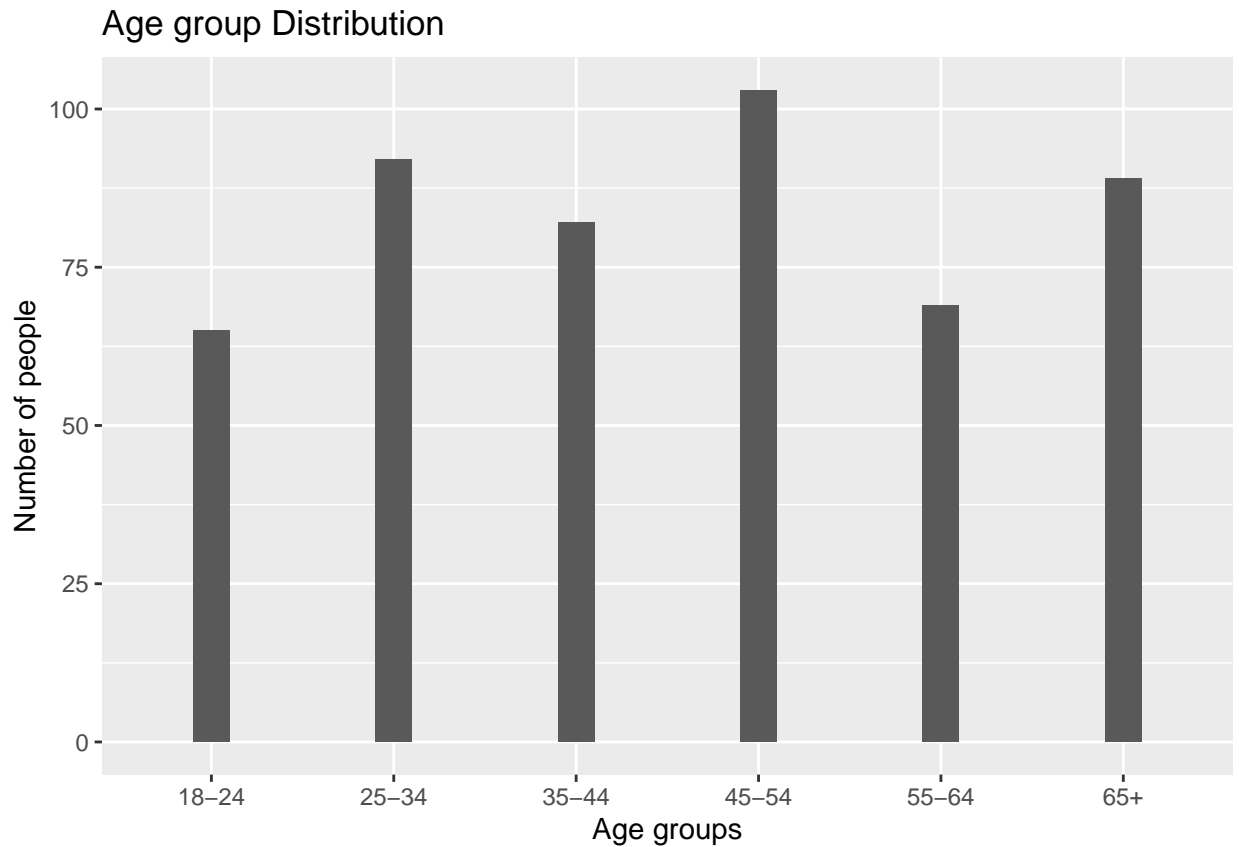
Gender Distribution
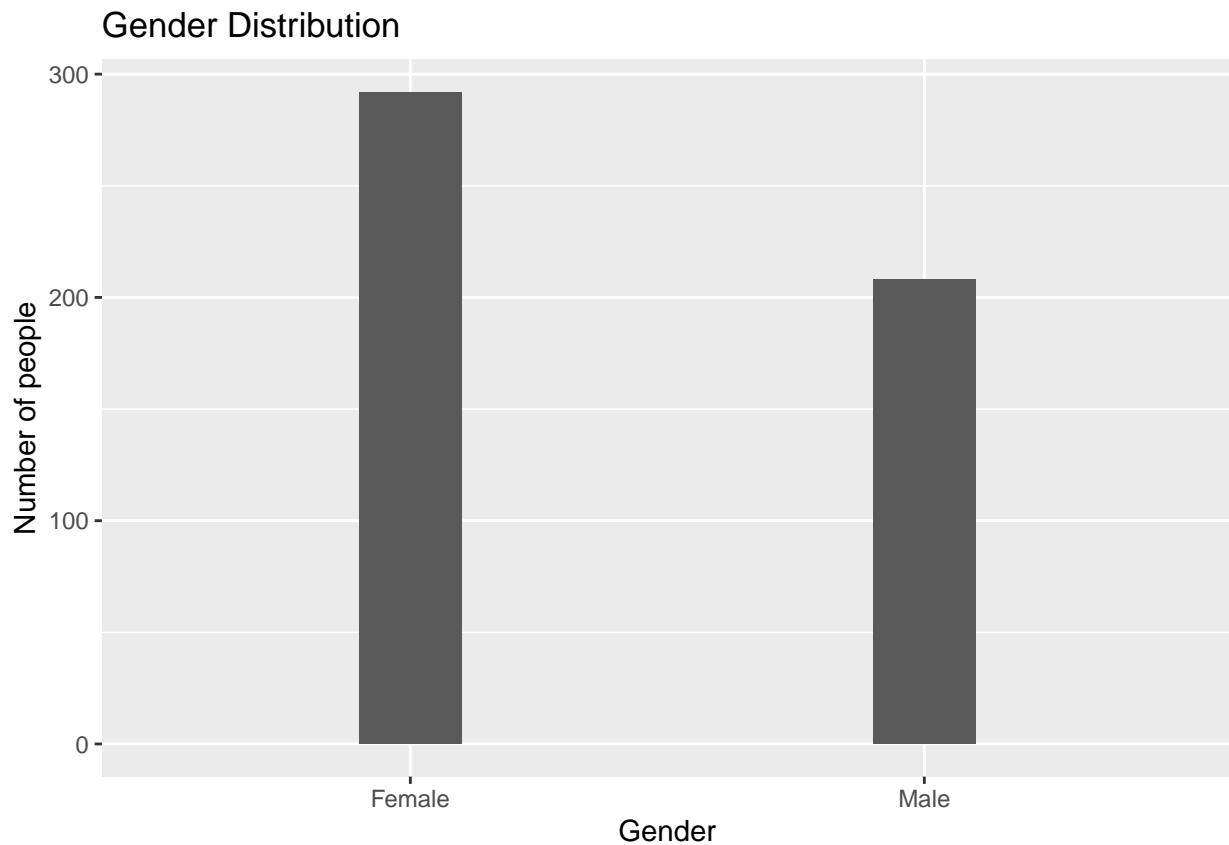
We have made a gender distribution graph on the 500 people who have taken the survey from the first bar chat above. The majority of them are female, which is about 20% more than males.

```
#Plot data to get a overview
ggplot(my_data, aes(x = d2_ethnicity)) + geom_bar(width=0.2)+labs( x = "Ethnic groups", y = "Number of p
                                                    title ="Ethnic Distribution")
```
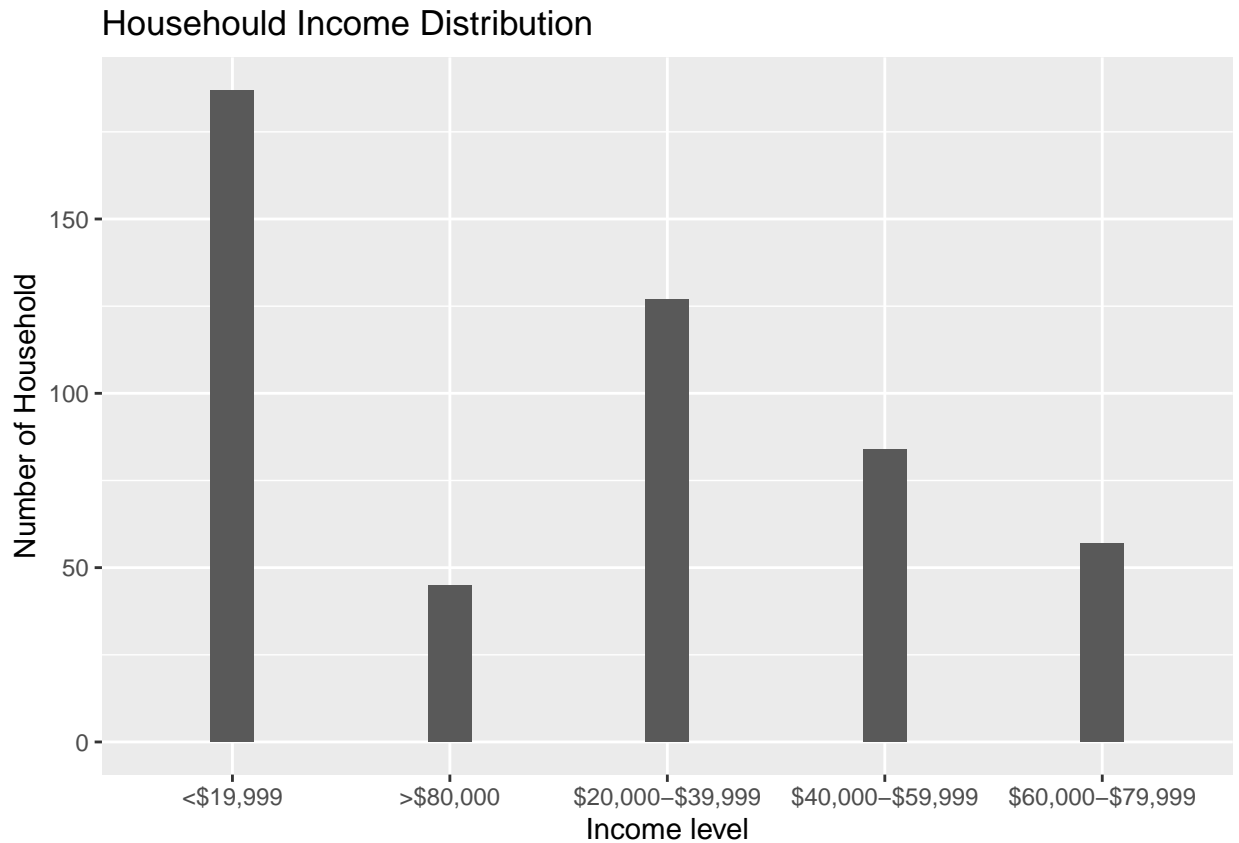
Ethnic Distribution

This bar chart that tells us how the 500 people are divided by ethnicity. The result is divided into Aboriginal, African, Asian, Caribbean, Central & Southern American, European, and Oceania. As the graph tells, people who have taken this survey are the majority of Asians and Europeans.

```
#Plot data to get a overview
ggplot(my_data, aes(x = d3_age)) + geom_bar(width=0.2)+labs( x = "Age groups", y = "Number of people",
```

Age group Distribution

Above, we have an age distribution map toward the survey group, the majority of the participants are aged from 45-54, followed by the group of people who are 25-34 years old, which is almost the same amount compared to people who are over 65 years old, and around 80 people who are 35-44 years old, 72 from 55-64, and 63 from 18-24.

```r
#Plot data to get a overview
ggplot(my_data, aes(x = d4_income)) + geom_bar(width=0.2)+labs( x = "Income level", y ="Number of Housel
```

Househould Income Distribution

Here is an income distribution bar chart for the 500 participants. The majority of the participants' income per year is lower than $19999, and amount of people in the graph presents a decreasing trend, with the salary increase.

```r
#Plot data to get a overview
ggplot(my_data, aes(x = d5_effect)) + geom_bar(width=0.2)+labs( x = "Level", y = "Number of people", ti
```

Life experience
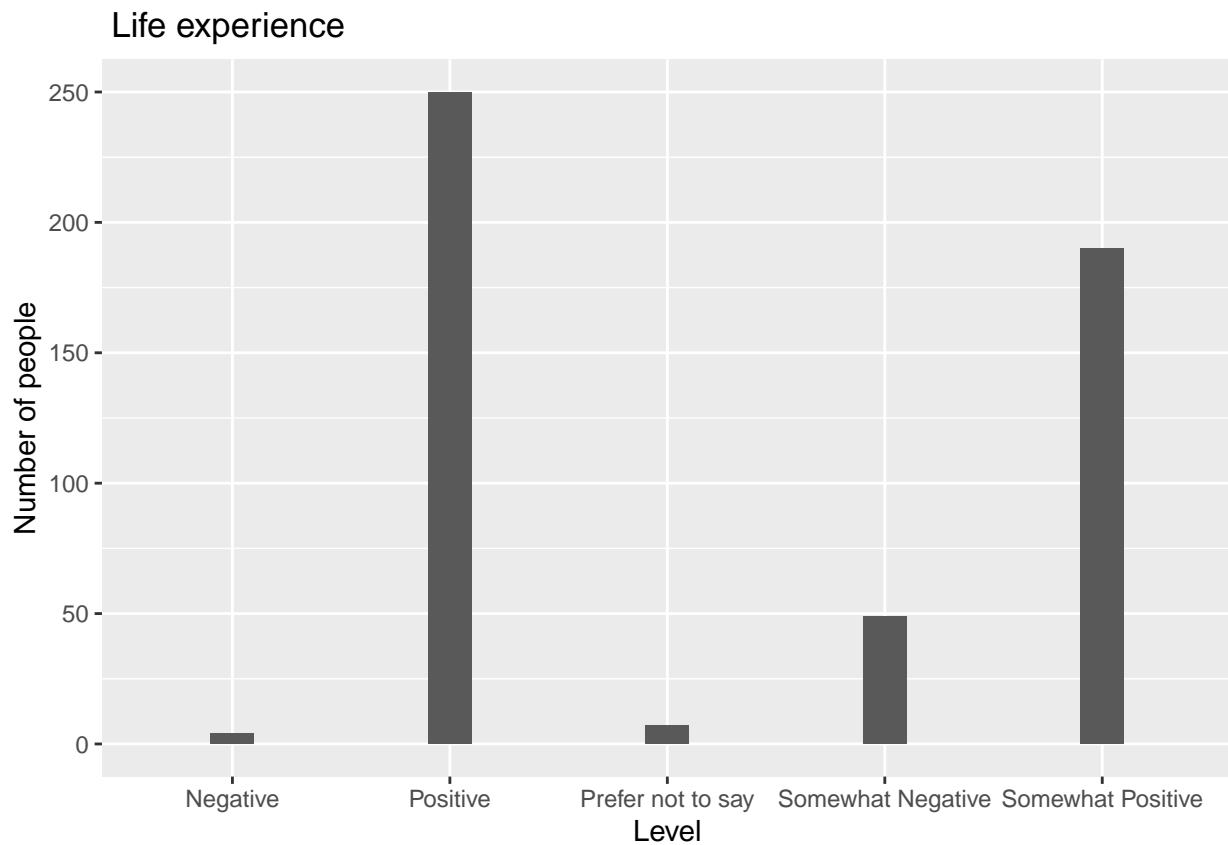
This section shows how people feel about government policies in reaction to COVID-19. Most of the citizens are optimistic about government measures for the pandemic. More than half of the people who had selected positive, strongly agree, others somewhat agrees.

```
ggplot(my_data, aes(x = d6_efficiency)) + geom_bar(width=0.2)+labs( x = "Degree of Satisfaction", y = "
```

Satisfaction Distribution
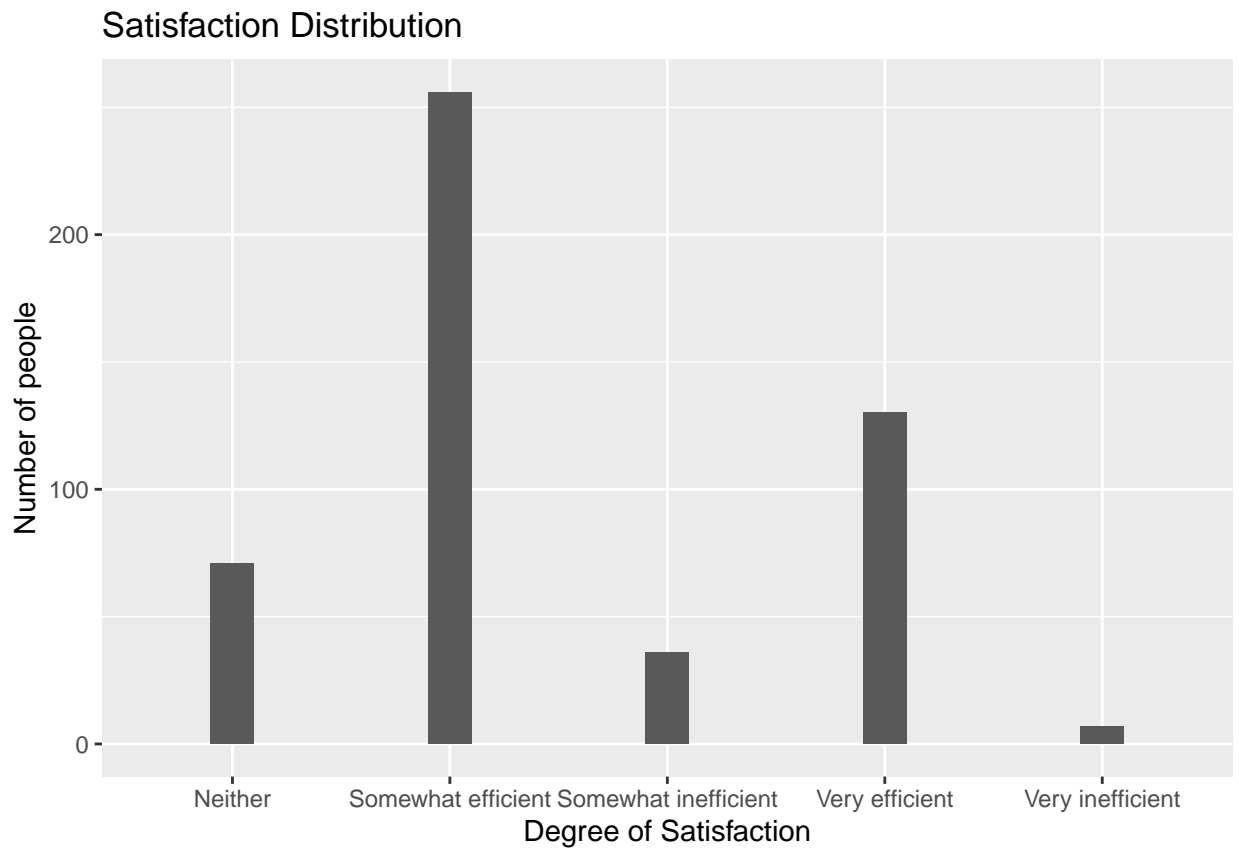

```r
#Plot data to get a overview
ggplot(my_data, aes(x = d7_vote)) + geom_bar(width=0.2)+labs( x = "Types of Party", y = "Distribution",
```

Poll result
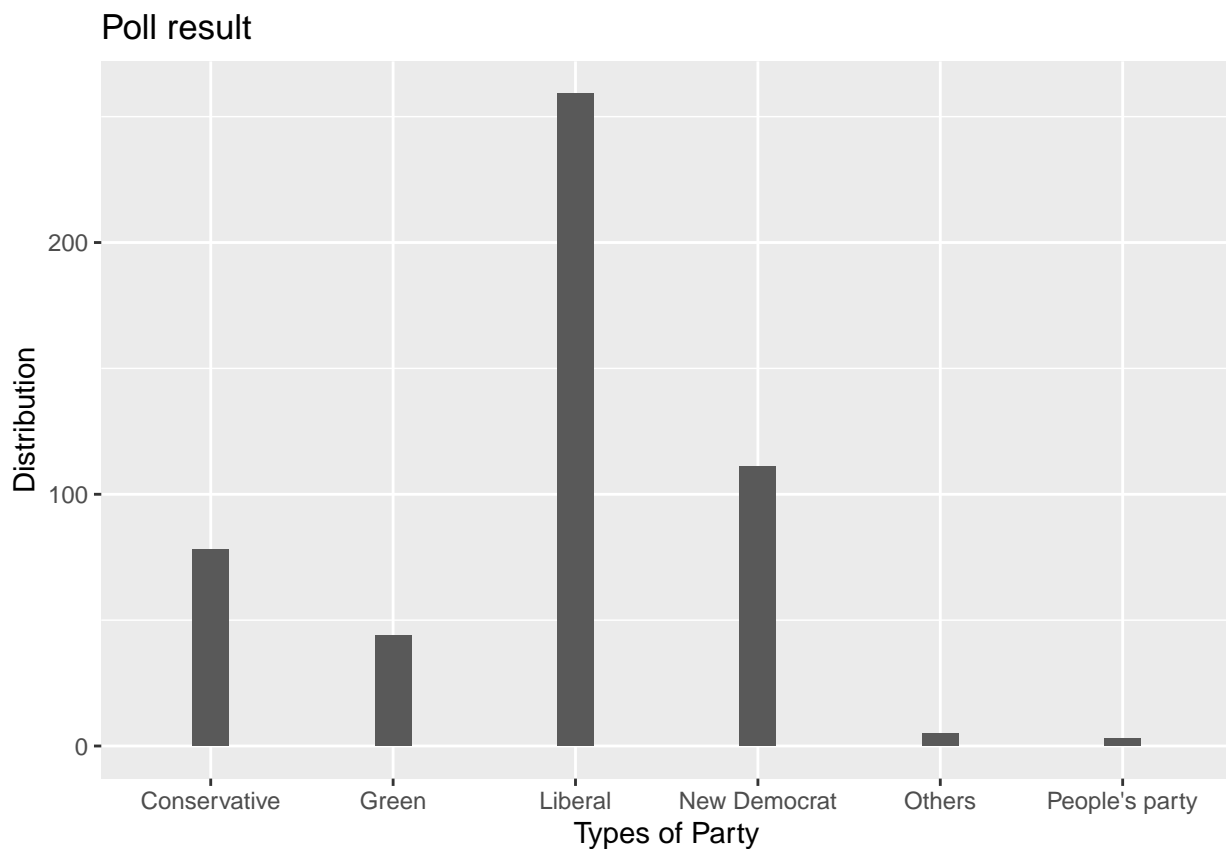

```r
# Gender VS Election choice
set.seed(500)
my_data$Party<-NA
#Candidates demographic(how many male and female in survey population)
D.f<-nrow(my_data[my_data$Gender=='Female',])
D.m<-nrow(my_data[my_data$Gender=='Male',])
D.f
```

```
## [1] 292
```

```r
D.m
```

```
## [1] 208
```

```r
#Connect gender with election preference
Party<- c("Liberal","New Democrat","Green","Others","Conservative","People's party")
Party.data.f<-sample(Party,D.f,replace=TRUE,prob=c(0.32,0.15,0.04,0.01,0.07,0.002))
Party.data.m<-sample(Party,D.m,replace=TRUE,prob=c(0.24,0.07,0.03,0.01,0.05,0.008))
my_data[my_data$Gender=='Female',"Party"]<-Party.data.f
my_data[my_data$Gender=='Male',"Party"]<-Party.data.m

ggplot(my_data, aes(x = Party,fill=Gender)) + geom_bar(width=0.2)+labs( x = "Type of Party", y = "Gender
                                               title ="Poll result")
```
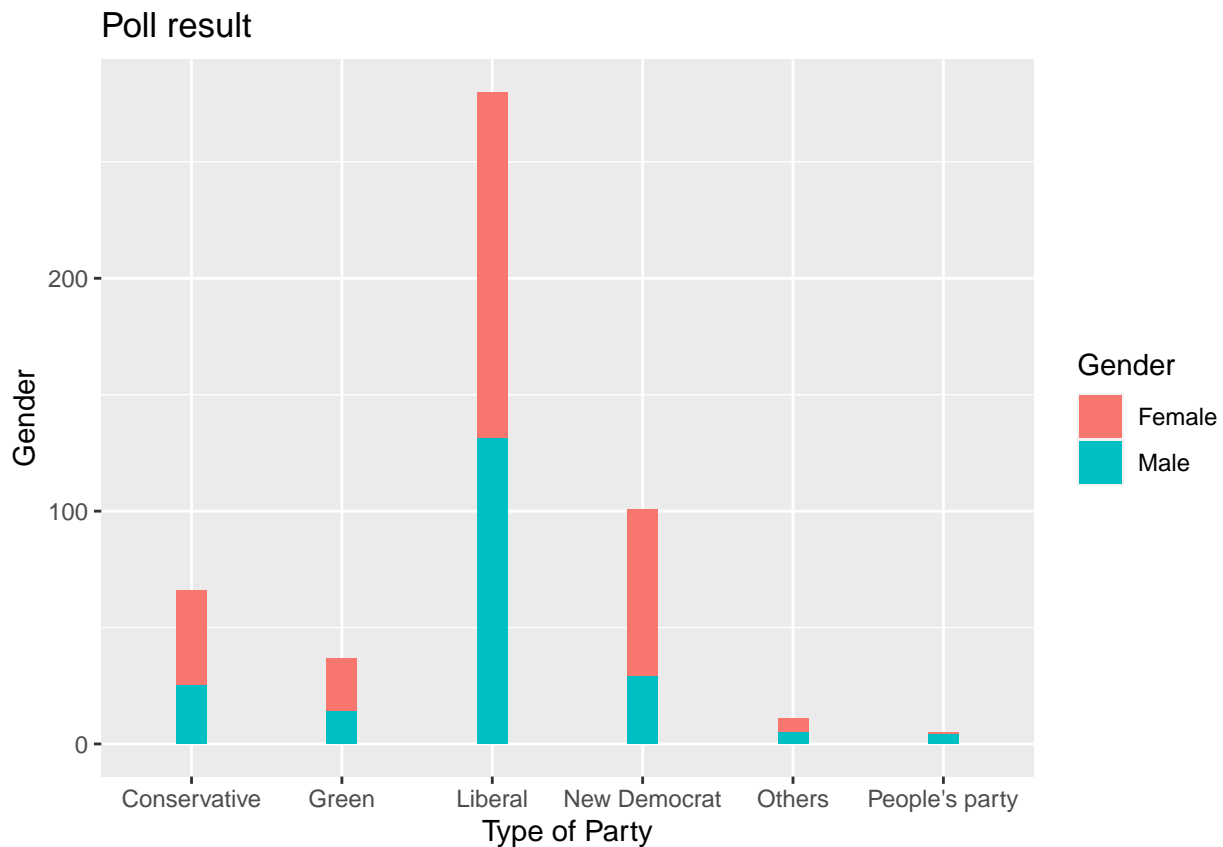
Poll result

Poll result between gender and election: The percentage distribution between all parties are almost evenly distributed, about a slightly more than a half of participants are female, and less than a half are male, applied to all parties supportive.

```r
set.seed(500)
#Income lvl VS Election choices
my_data$Party2<-NA
#Candidates demographic(household income lvl)
D.a<-nrow(my_data[my_data$Household_income=='<$19,999',])
D.b<-nrow(my_data[my_data$Household_income=='$20,000-$39,999',])
D.c<-nrow(my_data[my_data$Household_income=='$40,000-$59,999',])
D.d<-nrow(my_data[my_data$Household_income=='$60,000-$79,999',])
D.e<-nrow(my_data[my_data$Household_income=='>$80,000',])
D.a
```

```
## [1] 187
```

```r
D.b
```

```
## [1] 127
```

```r
D.c
```

```
## [1] 84
```

```r
D.d
```

```
## [1] 57
```

```r
D.e
```

```
## [1] 45
```

```
#Connect Household income lvl with election action
Party2<- c("Liberal","New Democrat","Green","Others","Conservative","People's party")
Party2.data.a<-sample(Party2,D.a,replace=TRUE,prob=c(0.05,0.07,0.012,0.0014,0.045,0.0024))
Party2.data.b<-sample(Party2,D.b,replace=TRUE,prob=c(0.09,0.04,0.014,0.0018,0.03,0.0033))
Party2.data.c<-sample(Party2,D.c,replace=TRUE,prob=c(0.11,0.02,0.02,0.014,0.005,0.00369))
Party2.data.d<-sample(Party2,D.d,replace=TRUE,prob=c(0.15,0.02,0.016,0.0016,0.012,0.00035))
Party2.data.e<-sample(Party2,D.e,replace=TRUE,prob=c(0.16,0.05,0.008,0.0012,0.028,0.00026))
my_data[my_data$Household_income=='<$19,999',"Party2"]<-Party2.data.a
my_data[my_data$Household_income=='$20,000-$39,999',"Party2"]<-Party2.data.b
my_data[my_data$Household_income=='$40,000-$59,999',"Party2"]<-Party2.data.c
my_data[my_data$Household_income=='$60,000-$79,999',"Party2"]<-Party2.data.d
my_data[my_data$Household_income=='>$80,000',"Party2"]<-Party2.data.e

#Overview of household income distribution in each category
ggplot(my_data, aes(x = Party2,fill=Household_income)) + geom_bar(width=0.2)+labs( x = "Type of Party",
                                                                      title ="Poll result")
```
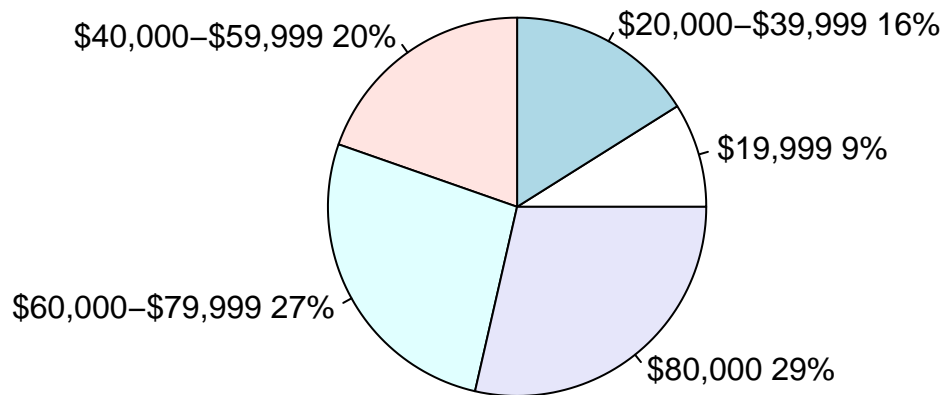

Poll result
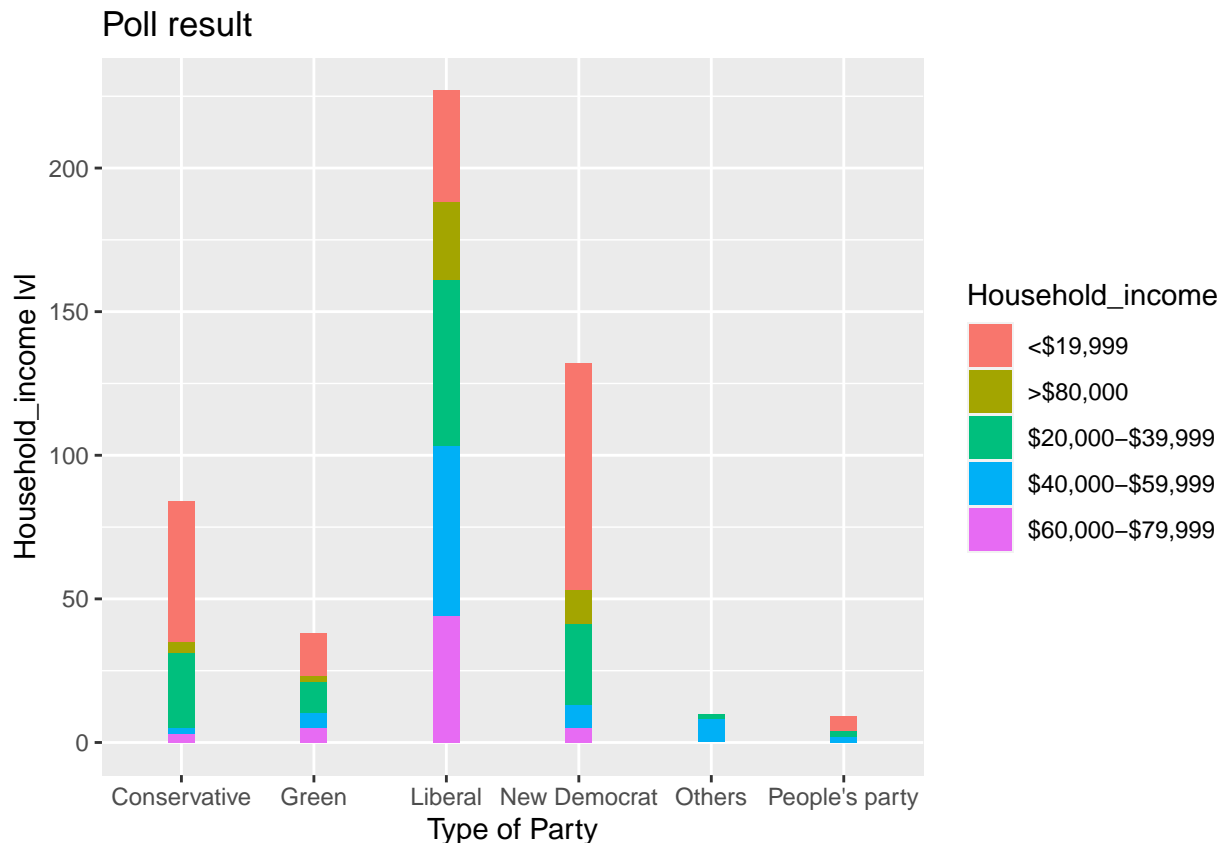
```
#Distribution of people who have voted for Liberal
slices <- c(0.05,0.09,0.11,0.15,0.16)
lbls <- c("$19,999", "$20,000-$39,999", "$40,000-$59,999", "$60,000-$79,999", "$80,000")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct) # add percents to labels
lbls <- paste(lbls,"%",sep="") # ad % to labels
pie(slices, labels = lbls, main="Voter demographics")
```

Voter demographics

Poll result between income and election: The majority of people would select Conservatives in the next election, followed by the Liberal party. In general, almost the same number of middle-class participants would choose liberal and conservative, and people with lower incomes would prefer the Conservative.

```r
set.seed(500)
#Ethnicity VS Election choices
my_data$Party3<-NA
#Candidates demographic(ethnicity)
A.a<-nrow(my_data[my_data$Ethnicity=='Aboriginal',])
A.b<-nrow(my_data[my_data$Ethnicity=='Asian',])
A.c<-nrow(my_data[my_data$Ethnicity=='African',])
A.d<-nrow(my_data[my_data$Ethnicity=='Caribbean',])
A.e<-nrow(my_data[my_data$Ethnicity=='European',])
A.f<-nrow(my_data[my_data$Ethnicity=='SouthAmerican',])
A.g<-nrow(my_data[my_data$Ethnicity=='Oceania',])


A.a
```

```
## [1] 5
```
```r
A.b
```

```
## [1] 218
```
```r
A.c
```

```
## [1] 26
```
```r
A.d
```

```
## [1] 22
```
```r
A.e
```

```
## [1] 208
```
```r
A.f
```

```
[1] 11
```
```r
A.g
```

```
## [1] 10
```

```
#Connect ethnicity with vote
Party3<-c("Liberal","New Democrat","Green","Others","Conservative","People's party")
Party3.data.a<-sample(Party3,A.a,replace=TRUE,prob=c(0.012,0.02,0.022,0.0026,0.021,0.0034))
Party3.data.b<-sample(Party3,A.b,replace=TRUE,prob=c(0.18,0.042,0.017,0.0024,0.011,0.002))
Party3.data.c<-sample(Party3,A.c,replace=TRUE,prob=c(0.17,0.0561,0.0052,0.0014,0.012,0.00004))
Party3.data.d<-sample(Party3,A.d,replace=TRUE,prob=c(0.048,0.012,0.0055,0.0026,0.019,0.00048))
Party3.data.e<-sample(Party3,A.e,replace=TRUE,prob=c(0.052,0.0217,0.0051,0.0033,0.019,0.0019))
Party3.data.f<-sample(Party3,A.f,replace=TRUE,prob=c(0.058,0.0352,0.0048,0.0038,0.019,0.0017))
Party3.data.g<-sample(Party3,A.g,replace=TRUE,prob=c(0.04,0.033,0.0104,0.0039,0.019,0.00048))

my_data[my_data$Ethnicity=='Aboriginal',"Party3"]<-Party3.data.a
my_data[my_data$Ethnicity=='Asian',"Party3"]<-Party3.data.b
my_data[my_data$Ethnicity=='African',"Party3"]<-Party3.data.c
my_data[my_data$Ethnicity=='Caribbean',"Party3"]<-Party3.data.d
my_data[my_data$Ethnicity=='European',"Party3"]<-Party3.data.e
my_data[my_data$Ethnicity=='SouthAmerican',"Party3"]<-Party3.data.f
my_data[my_data$Ethnicity=='Oceania',"Party3"]<-Party3.data.g

# Get a overview of different ethnicity's election preference
ggplot(my_data, aes(x = Party3,fill=Ethnicity)) + geom_bar(width=0.2)+labs( x = "Type of Party", y = "E
                                                      title ="Poll result")
```
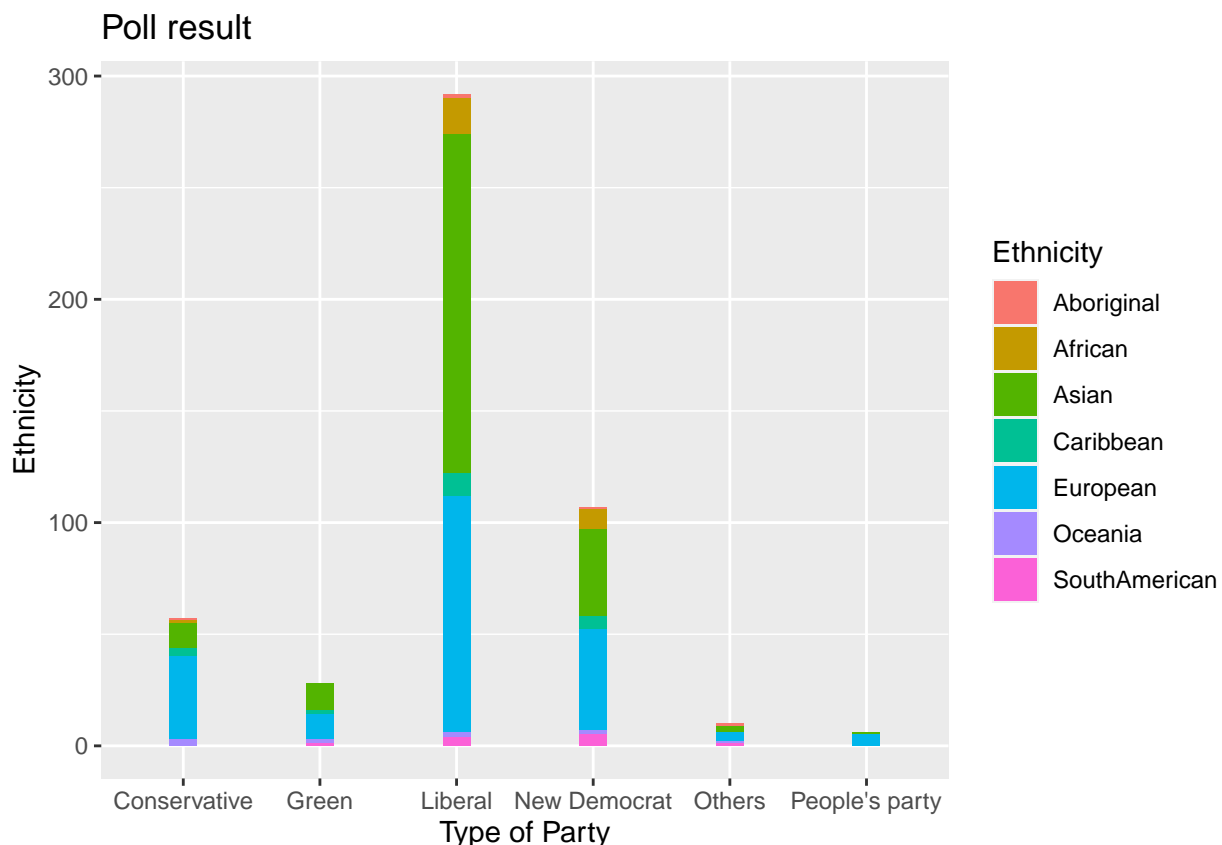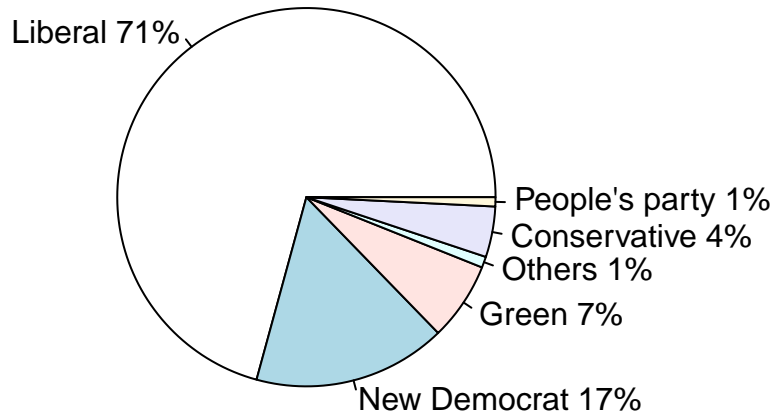


```
#Asian election preference
slices <- c(0.18,0.042,0.017,0.0024,0.011,0.002)
lbls <- c("Liberal","New Democrat","Green","Others","Conservative","People's party")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct) # add percents to labels
```

```
lbls <- paste(lbls,"%",sep="") # ad % to labels
pie(slices, labels = lbls, main="Asian election preference")
```

Asian election preference


```
#Aboriginal election preference
slices <- c(0.012,0.02,0.022,0.0026,0.021,0.0034)
lbls <- c("Liberal","New Democrat","Green","Others","Conservative","People's party")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct) # add percents to labels
lbls <- paste(lbls,"%",sep="") # ad % to labels
pie(slices, labels = lbls, main="Aboriginal election preference")
```
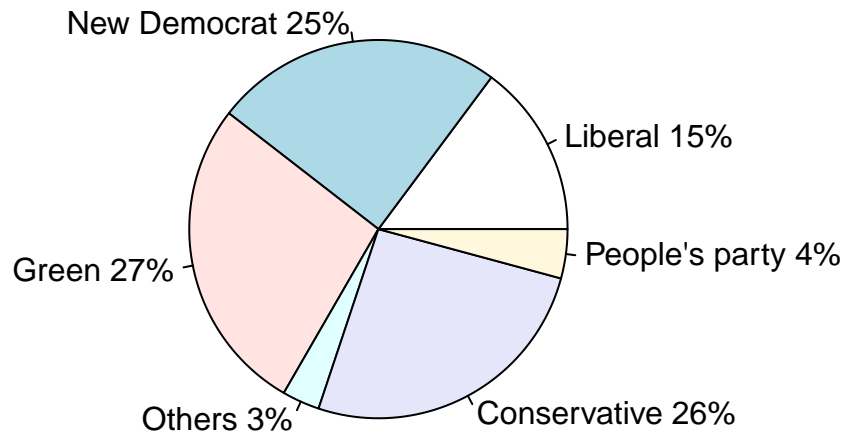
Aboriginal election preference

Pie chart of the distribution of participants categorized by income in liberal parties. Majority of the supportive are high income.

Part 4: Discussion

How to keep everyone happy?

We have done a fantastic job on racial equalizing, particularly on Asian (71%) and other minority groups, since us and conservation party are both immigration friendly. There are more than 85% of the chance that a minority racial group will vote for us, therefore, as long as we keep on paying attention on taking care of social inequality problems, then we are good to go. The only part that we need to pay more attention is how we can make the aboriginal people feel more comfortable. Majority of the Aboriginal people have voted for the People's party, and we need to focus more on the 'corruption and ethics in government'.

What matters in term of our survey?

As results have shown, a political party needs to take care of several factors to receive supports: gender, age, ethnicity, income. Gender equalization is an essential factor in this particular kind of survey because the government can have policies towards all genders, but feel unfair for others because of gender issues. So it can tell by the liberal party's policies preferences by gender. Ethnicity is a huge problem for politics, as we live in a multicultural society. As we know that racism is not accepted no matter what happens globally, politics need to be race/colour-friendly. This survey does have the power to show how minorities feel about current politics. Doing an income background check is necessary for the focus of our political party. We need to focus on making the rich happy and the poor comfortable. People with different levels of income tend to have other interests in political factors. Lastly, our society has people of all ages, just like the income background check. We must know what people think of different ages to make a better environment for all people.

So here is what we should do next:

We have found out that the Liberal party has equal supports by males and females, which means that we are doing a great job on gender equalization. We will need to pay more attention to 'Affordability and cost of living'; most of our supports comes from the 'rich.' As we pay more attention to the lower-income class, we can have a more successful run. To do that, we can do so by increasing the minimum wage and decreasing the tax overall. Although t might sound like the government will go 'broke' afterwards, people are starting to have a better life and better income. The willingness to spend has increased, which can create more job opportunities and more government income. Besides raising people's economic life, we need to balance out and make some benefits towards housing, food, rents, anything that is rational and potentially can help our citizens. By the time we finish working on these problems, we can have supports from all income classes, and people will be happier.

How we can make a bigger stand in the society?

In the main time, a more important thing is that people who have selected conservative party is more than the supports towards liberal party. Our big focus as the liberal party includes: 'climate change', 'health care', 'general economy', 'affordable and living costs', 'Tax/senior issues'. The conservative party once made a point of our increase of carbon tax is 'ineffective'. Beside that, the conservative party's goal of our government deficit is '0', as we only promise to decrease to under $10 billion. Maybe people are not happy with that, we need to spend more time figuring out about raising our local GDP and GNP, increase the net export in order to balance out the deficit. Once we have sorted out the general economy problem, we are not far from getting more supports.

Reference

codes used for this study is linked below. https://github.com/joliw9282/2020-toronto-election-prediction