

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 Profile, 2016 Census Canada*, n.d.)). 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. Voters, predominantly Asian and European groups, favour the Liberal party over the others. The liberal supporters are mostly 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 ((*Canadian Federal Election Results in Central Toronto*, n.d.)). Over the past few elections, the Liberal leaders had an absolute majority of all districts. Last year, the Liberal party won all 23 out of 23 districts in the city of Toronto (Meredith Bond (n.d.)). 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.

Part 3. Survey Methodology.

Our questionnaire includes questions that investigate the pandemic's effect on electors' evaluation, both emotionally and rationally, of current policies. 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 convenient sampling will also reduce our budget to below 400

dollars, which is only constructed with the cost of printing and human labor for putting up posters. We will get into details of the drawbacks of our method in the discussion section.

This study stimulated a representative sample of 500 participants; responses to each question are generated separately and then combined into a collective dataset using R (Wickham et al. (2019)). A sample with a population of 6,196,731 (*Toronto Population* 2020), a margin of error of $\pm 5\%$, and a confidence rate of 95%, would have an effective sample size of 385 subjects (*Sample Size Calculator* 2020).

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 $\pm 4\%$. 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 (John Ibbitson (2019)).

Sample Simulation.

method: non-probability sampling simulate population N = 500

```
library(tidyverse)

## -- Attaching packages -----
## 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 -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

# 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"

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"

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"
```

```

household_income <- c("<$19,999", "$20,000-$39,999", "$40,000-$59,999", "$60,000-$79,999", "$80,000+")
d4_income<-sample(household_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"

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"

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"

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"

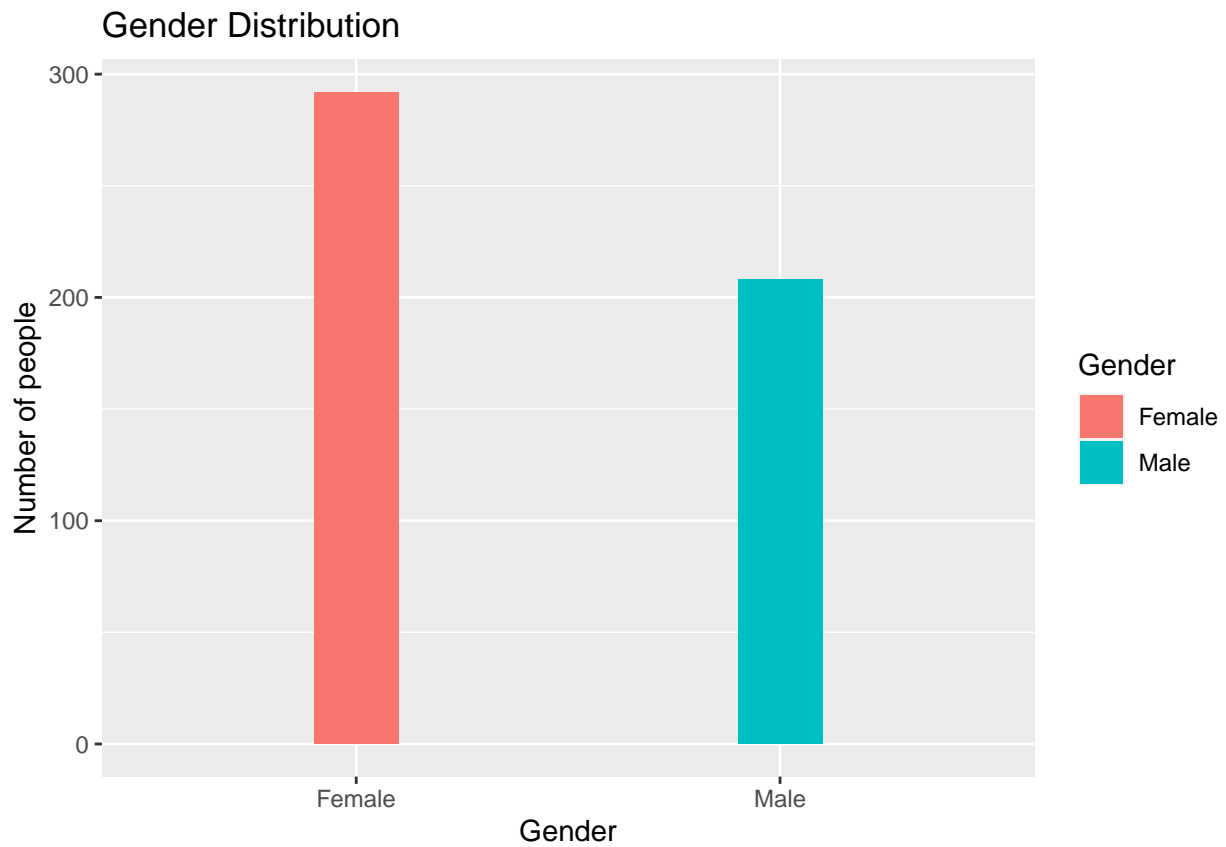
#Form a data set
my_data <- data.frame("Gender"=d1_gender, "Ethnicity"=d2_ethnicity, "Age_Group"=d3_age, "Household_income"=d4_income, "Life_Impact"=d5_effect, "Satisfaction_Lvl"=d6_efficiency, "Vote"=d7_vote)
head(my_data)

##   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

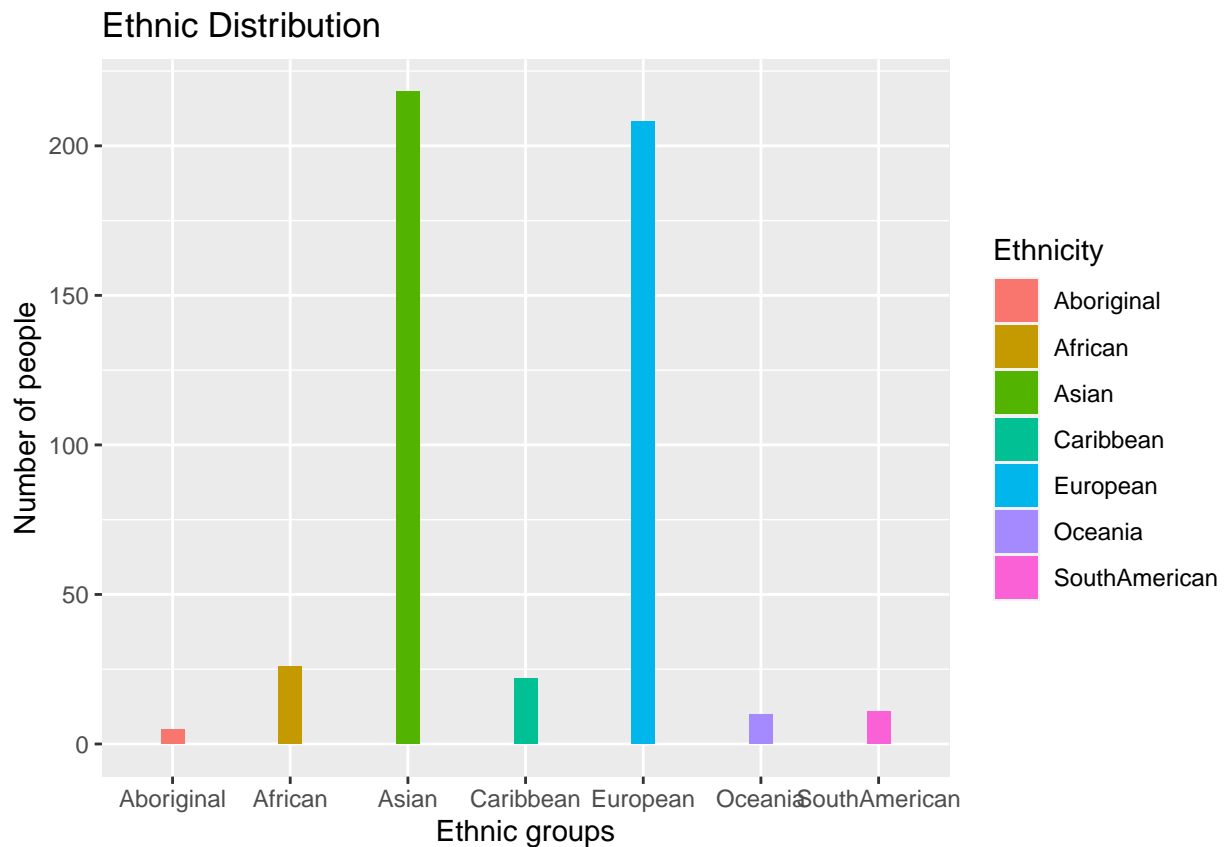
#Note: Each column is not related

#Plot data to get a overview
ggplot(my_data, aes(x = d1_gender, fill = d1_gender)) +
  geom_bar(width=0.2) +
  labs( x = "Gender", y = "Number of people", title ="Gender Distribution") +
  guides(fill = guide_legend(title = "Gender"))

```

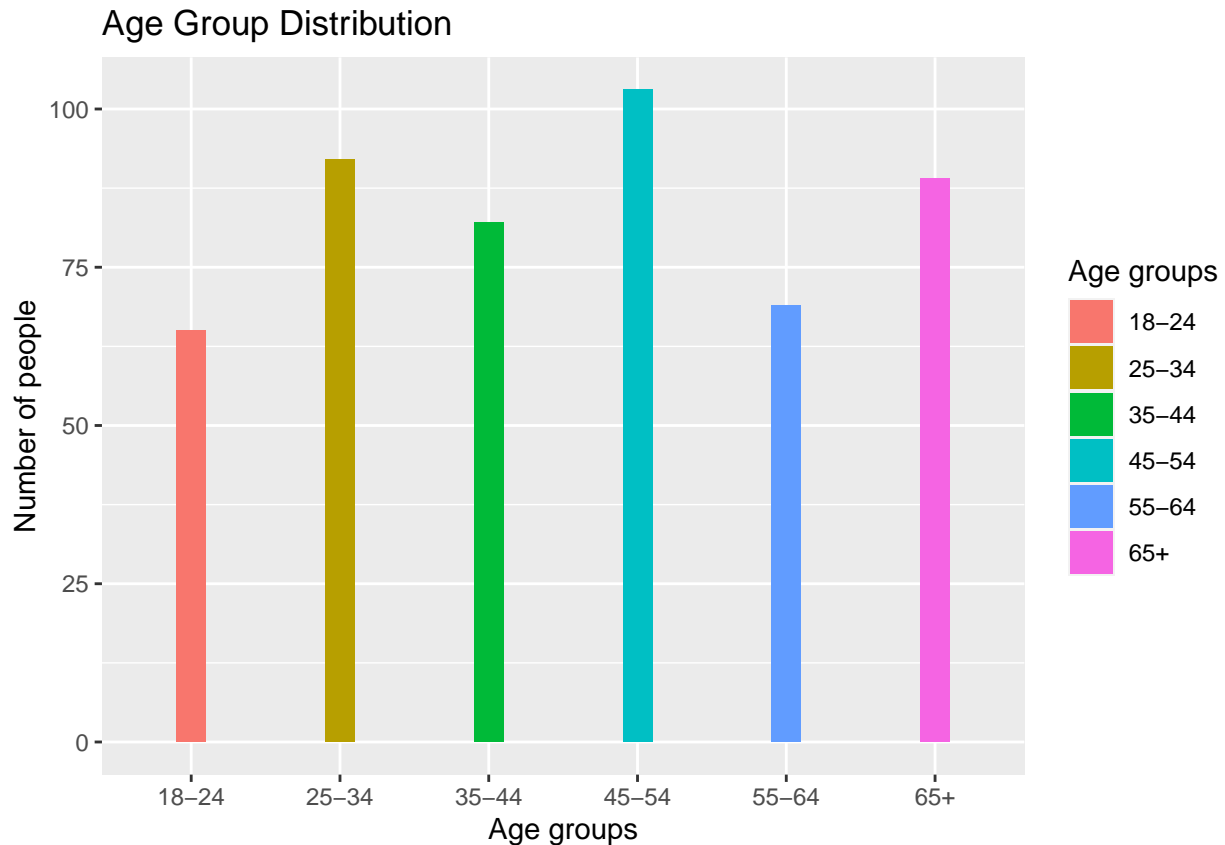


```
#Plot data to get a overview  
ggplot(my_data, aes(x = d2_ethnicity, fill = d2_ethnicity)) +  
  geom_bar(width=0.2) +  
  labs(x = "Ethnic groups", y = "Number of people", title = "Ethnic Distribution") +  
  guides(fill = guide_legend(title = "Ethnicity"))
```



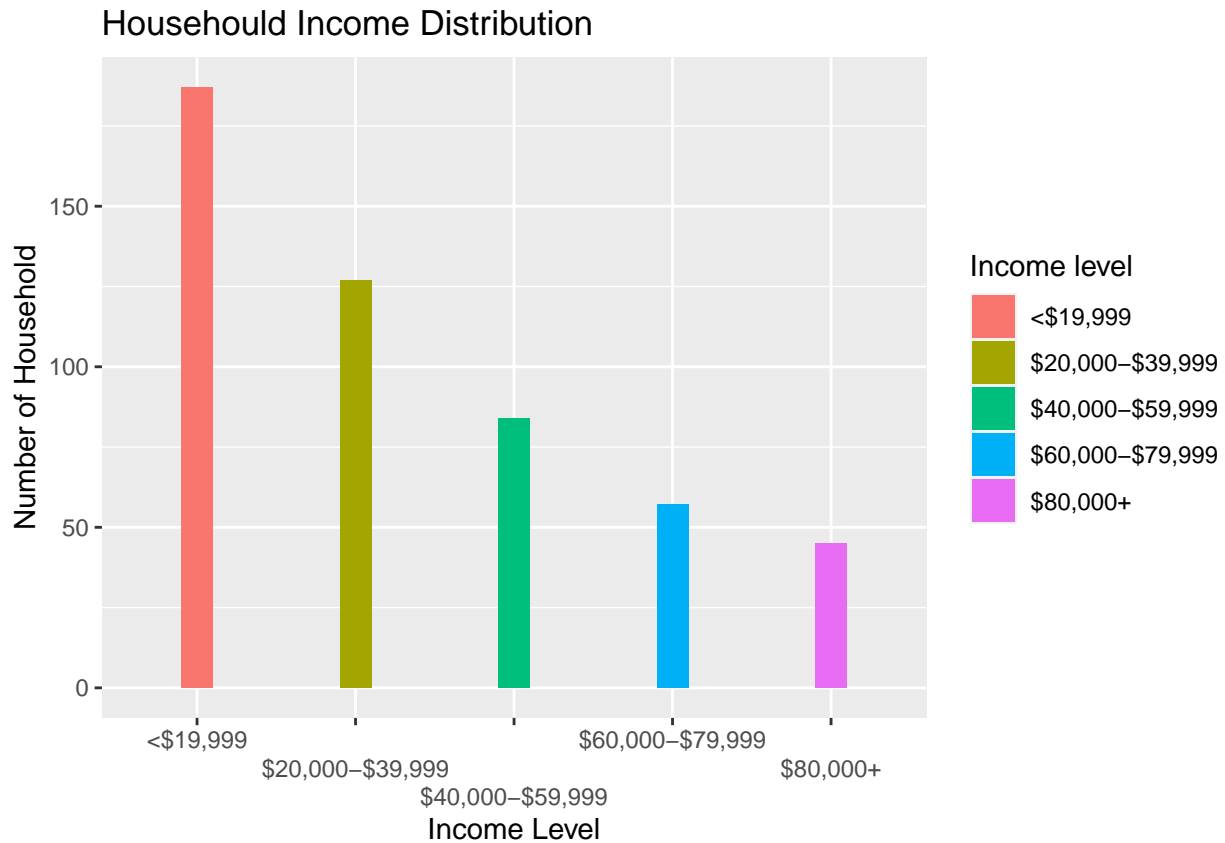
Majority of the survey participants are Asian and Europeans. These two ethnic groups takes up the more percentage of the population than all the others combined.

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



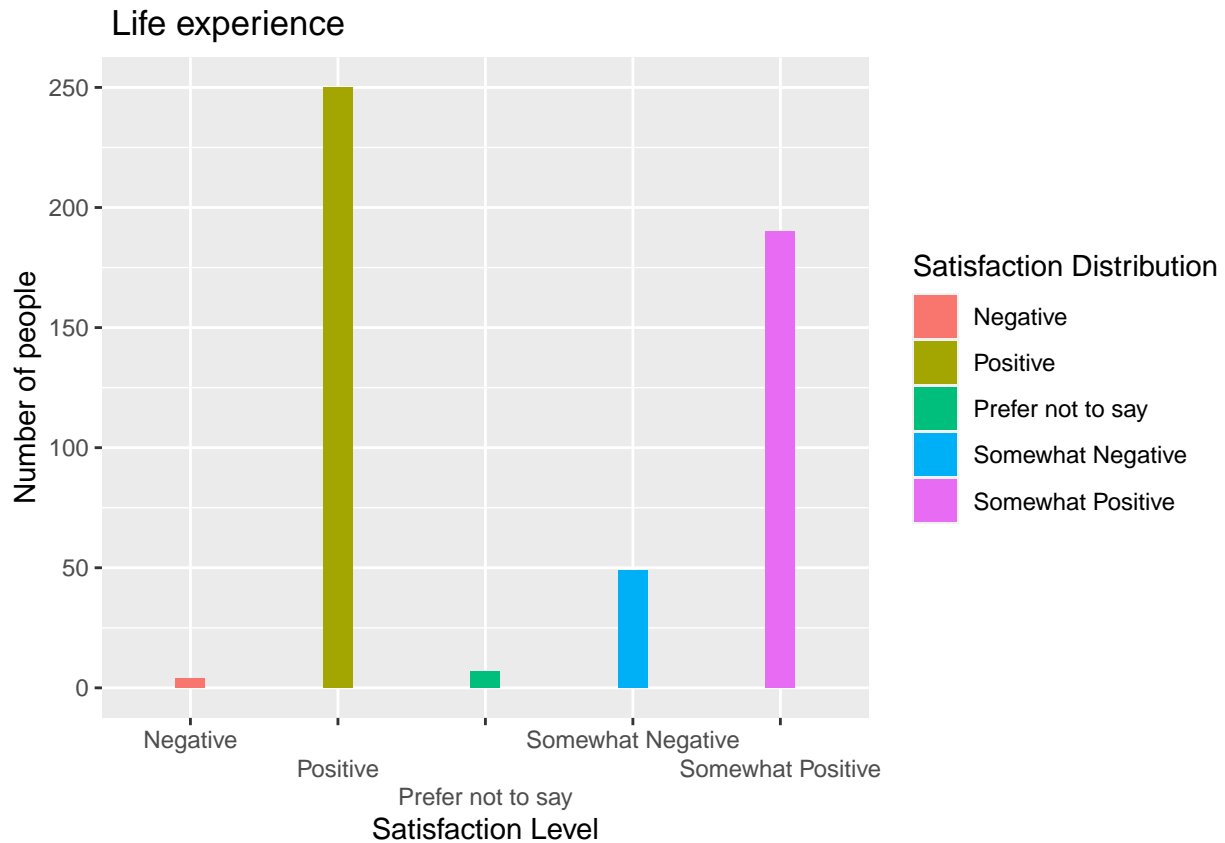
Above, we have an age distribution map toward the survey group, the majority of the participants are aged from 45-54, followed by the group of people who are 25-34 years old, which is almost the same amount compared to people who are over 65 years old, and around 80 people who are 35-44 years old, 72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

```
#Plot data to get a overview
ggplot(my_data, aes(x = d4_income, fill = d4_income)) +
  geom_bar(width=0.2) +
  labs( x = "Income Level", y = "Number of Household", title = "Household Income Distribution") +
  guides(fill = guide_legend(title = "Income level")) +
  scale_x_discrete(guide = guide_axis(n.dodge=3))
```



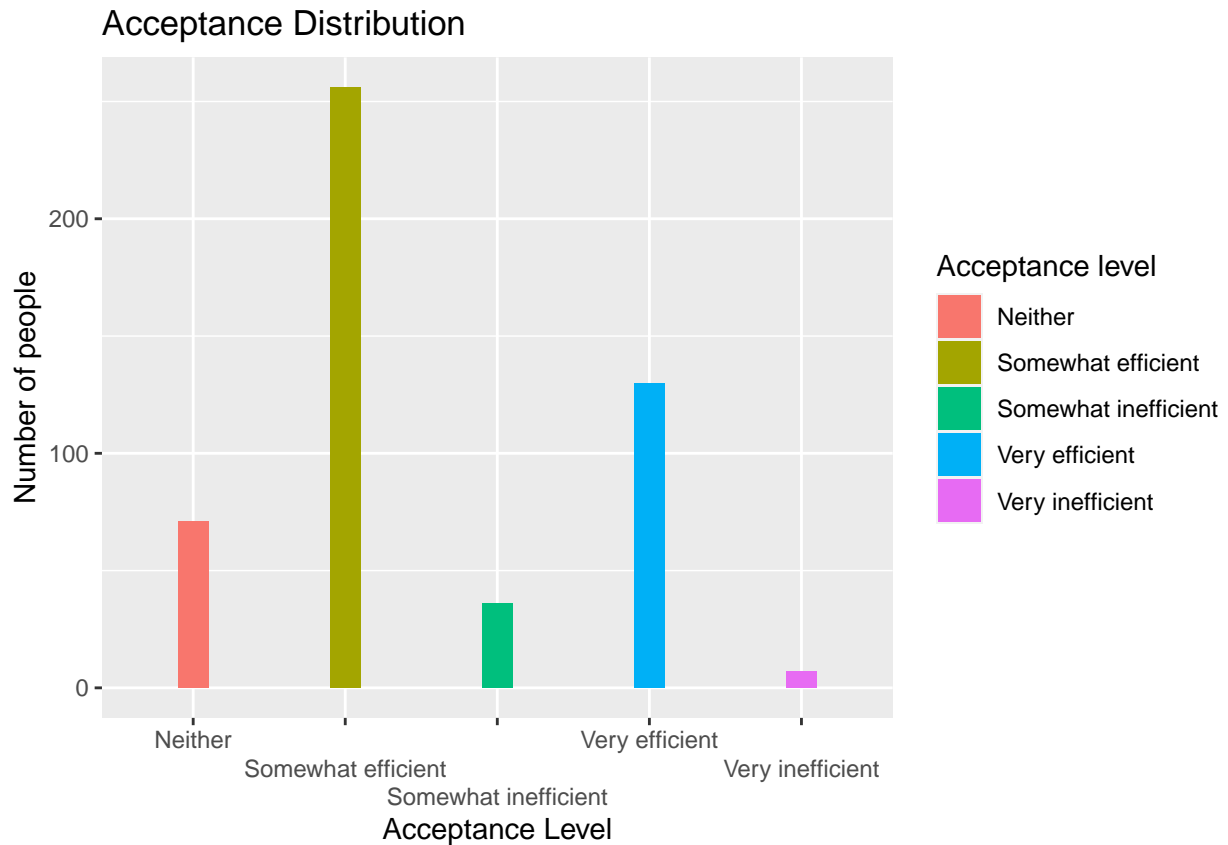
Here is an income distribution bar chart for the 500 participants. The majority of the participants' income per year is lower than \$19999, and amount of people in the graph presents a decreasing trend, with the salary increase.

```
#Plot data to get a overview
ggplot(my_data, aes(x = d5_effect, fill = d5_effect)) +
  geom_bar(width=0.2) +
  labs( x = "Satisfaction Level", y = "Number of people", title = " Life experience") +
  guides(fill = guide_legend(title = "Satisfaction Distribution")) +
  scale_x_discrete(guide = guide_axis(n.dodge=3))
```



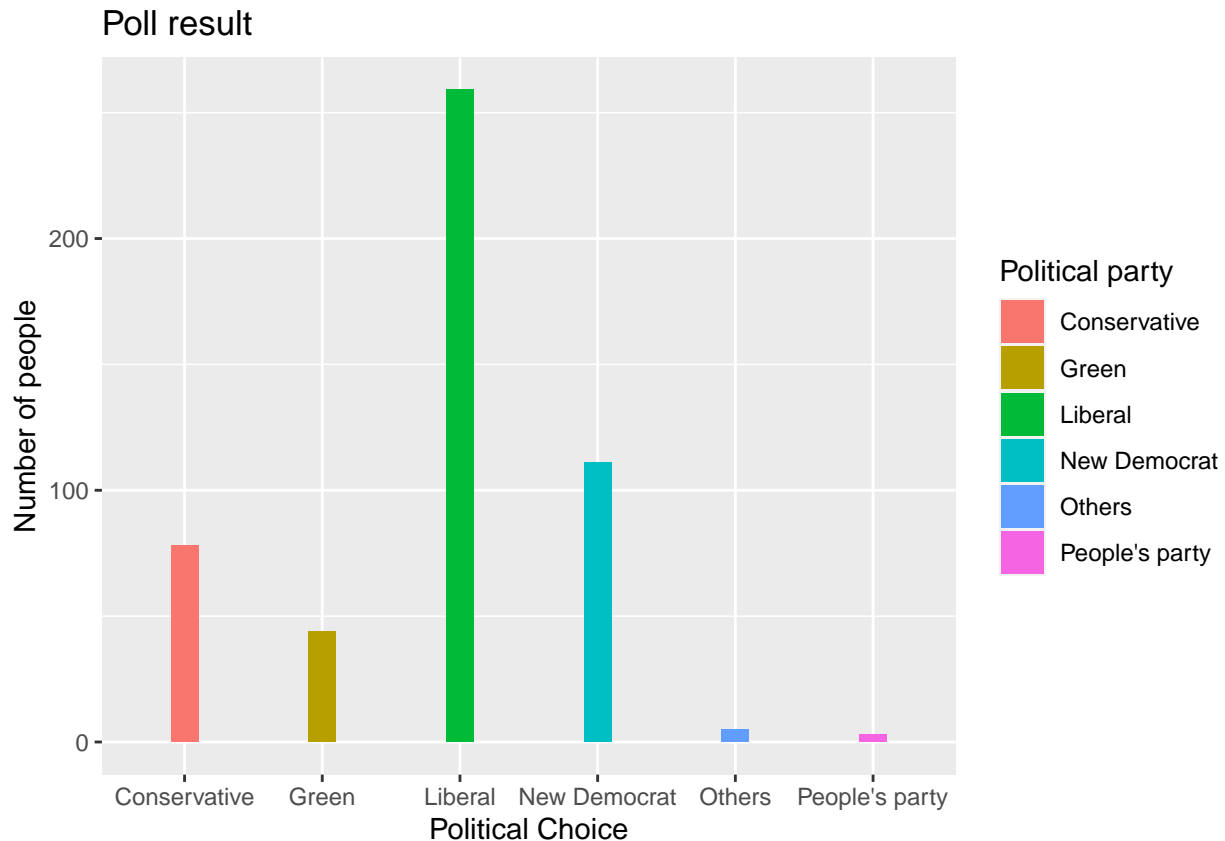
This section shows how people feel about government policies in reaction to COVID-19. Most of the citizens are optimistic about government measures for the pandemic. More than half of the people who had selected positive, strongly agree, others somewhat agrees.

```
ggplot(my_data, aes(x = d6_efficiency, fill = d6_efficiency)) +
  geom_bar(width=0.2) +
  labs( x = "Acceptance Level", y = "Number of people", title = "Acceptance Distribution") +
  guides(fill = guide_legend(title = "Acceptance level")) +
  scale_x_discrete(guide = guide_axis(n.dodge=3))
```

Most participants have a high recognition level for the government's reaction measures during the pandemic, with some benefit of the doubt.

```
#Plot data to get a overview
ggplot(my_data, aes(x = d7_vote, fill = d7_vote)) +
  geom_bar(width=0.2) +
  labs( x = "Political Choice", y = "Number of people", title = "Poll result") +
  guides(fill = guide_legend(title = "Political party"))
```



Support for the Liberal party is still significantly higher than the others. Unlike the national trend, the NDP seems to have a higher support rate than the Conservatives in Toronto.

Part 4: Data Analysis

```
# 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

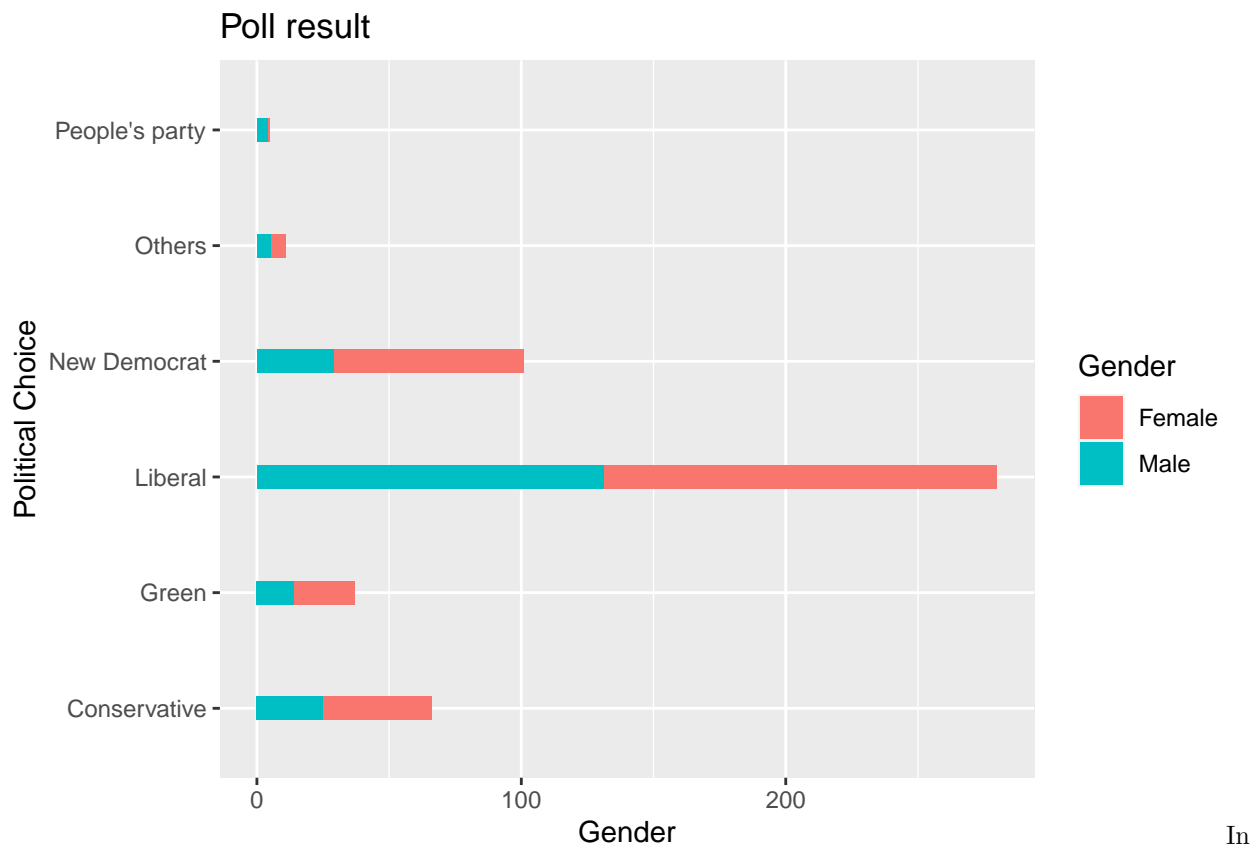
D.m

## [1] 208

#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 = "Political Choice", y = "Gender", title = "Poll result") +
coord_flip()
```



In general, almost all parties have slightly more female supporters than male supporters, this is consistent with the gender distribution in our sample, meaning that there is little gender differences in elector's political view. However, we do not have enough evidence to judge the significance of this difference.

#Income lvl VS Election choices

```
set.seed(500)
```

```
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
```

```
D.b
```

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

```
D.c
```

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

```
D.d
```

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

```
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
```

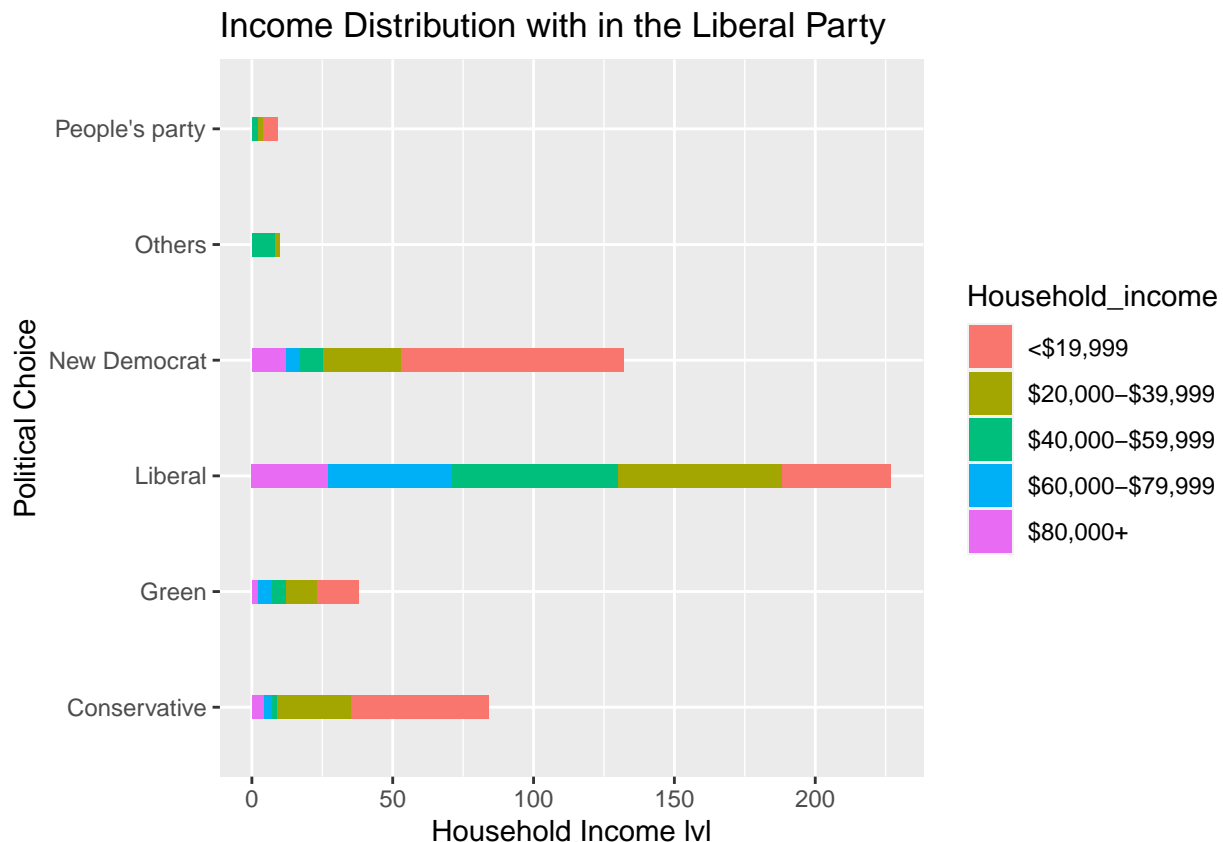
```
#Overview of household income distribution in each category
```

```
ggplot(my_data, aes(x = Party2,fill=Household_income)) +
```

```
  geom_bar(width=0.2) +
```

```
  labs( x = "Political Choice", y = "Household Income lvl", title ="Income Distribution with in the Lib
```

```
  coord_flip()
```

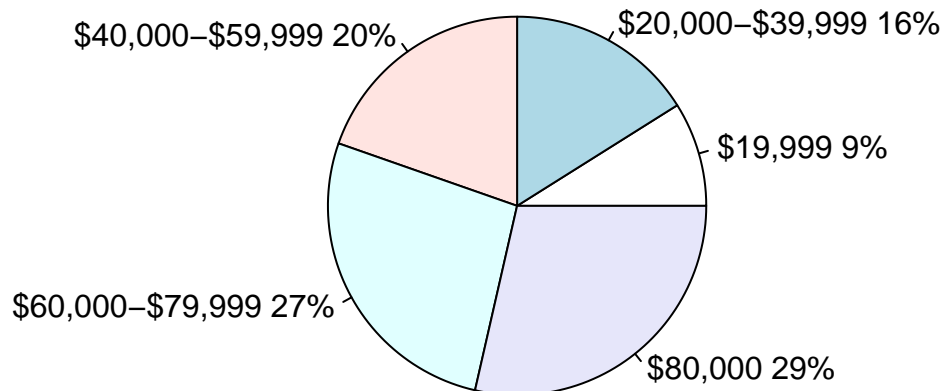


```
#Distribution of people who have voted for Liberal
```

```
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 within the Liberal Party")
```

Voter Demographics within the Liberal Party



The majority of the people who voted for the Liberal party have a higher than average SES status. According to the pie chart, over 50% of the participants have a household income over 60,000 dollars per year.

```
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
```

A.b

```
## [1] 218
```

A.c

```
## [1] 26
```

A.d

```
## [1] 22
```

A.e

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

```
A.f
```

```
## [1] 11
```

```
A.g
```

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

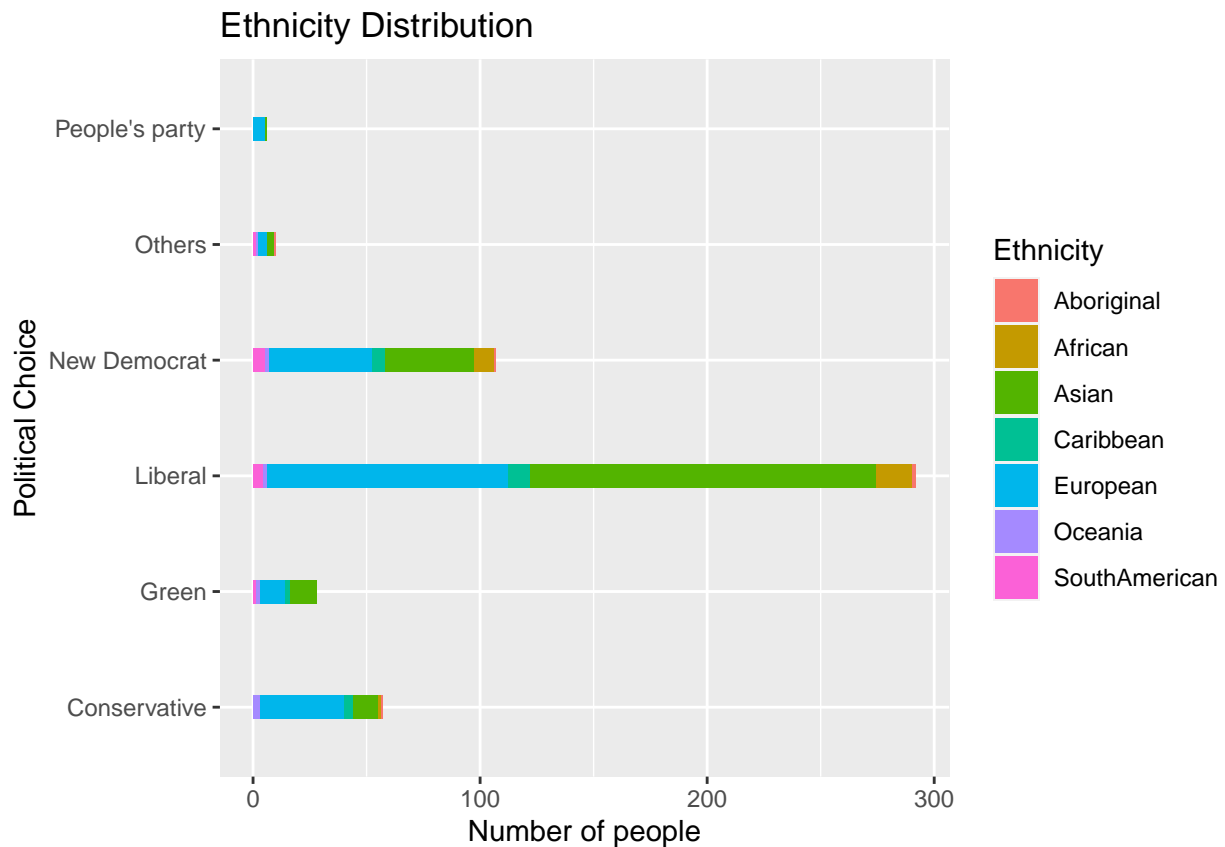
```
#Connect ethnicity with vote
```

```
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
```

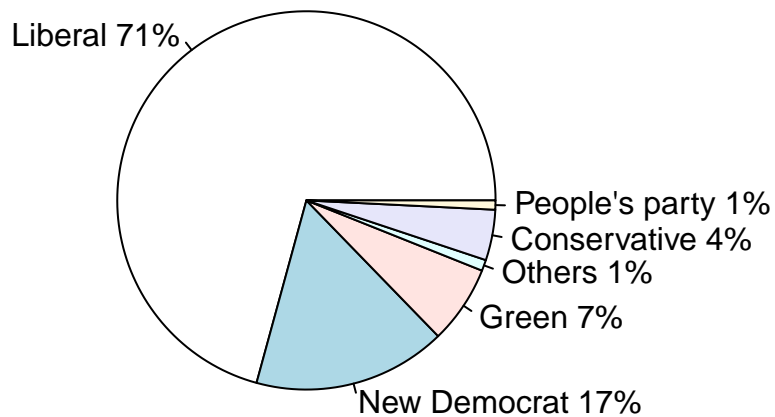
```
# Get a overview of different ethnicity's election preference
```

```
ggplot(my_data, aes(x = Party3,fill=Ethnicity)) +
  geom_bar(width=0.2) +
  labs( x = "Political Choice", y = "Number of people", title ="Ethnicity Distribution") +
  coord_flip()
```



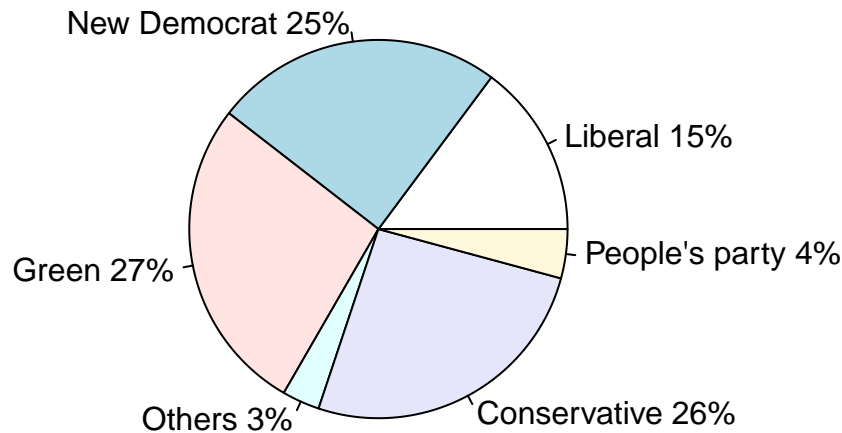
```
#Asian election preference
slices <- c(0.18,0.042,0.017,0.0024,0.011,0.002)
lbls <- c("Liberal","New Democrat","Green","Others","Conservative","People's party")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct) # add percents to labels
lbls <- paste(lbls,"%",sep="") # ad % to labels
p1 <- 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
p2 <- pie(slices, labels = lbls, main="Aboriginal Election Preference")
```

Aboriginal Election Preference



Within each political party, ethnic distribution of their supporters varies. For the Liberals, majority of the supporters are Asian and European, with the Asian population only slightly larger than the European population. We can also find similar first pattern in other parties as well, possibly because these two groups takes up most of Toronto's population makeup. However, in parties like the Conservative and the NDP, European supporters seem to take a larger portion, which is possibly because the Liberal party tend to have policies that benefit the immigrants and the minority groups.

This idea can be justified with the two pie charts shown above. Within ethnic groups, an overly dominant Asian population favors the Liberal party, whereas the Aboriginals tend to favor the NDP and the Conservative. It would be interesting for future studies to further investigate this matter.

Part 4: Discussion

How to keep everyone happy?

In general, the Liberal party has been doing a great job on designing policies that favors the majorities and controlling the pandemic in the city of Toronto. According to our simulation, the Liberal party has a advantage among the two dominant population, Asians and Europeans, who are also likely to have a household income of over 60,000 dollars per year. On the other hand, for minority groups, the support Liberal is much lower, only takes up to 15% among the Aboriginals. These results suggest, although the Liberal party wins most of the support in the city of Toronto, the reason for this advantage is because the city's population is made up of higher social economic status(SES) population. However, these findings also question participants' satisfaction level for pandemic measures. One potential problem is that, people with high SES are more likely to have access to high quality health care and social supports, where the situation might be different for minority groups and those who have lower SES. As one of the major cities in multicultural country, political parties should always be aware that people with different income levels tend

to have different political interests. Our survey is limited in finding the psychology behind the participants' choices, nonetheless, the liberals and the others should be more attentive for different needs of people from all groups.

What matters in term of our survey?

There are many limits to a simulation study. Our sample follow the population distribution from Census Canada (*Census Profile, 2016 Census Canada*, n.d.) during the data generation process, however, there is a high chance that we will not get sample distribution with a perfect match to the population. With respect to this issue, we increased the sample size to 500, which includes more than 100 participants than the effective sample size (385) with Toronto's population.

So here is what we should do next:

We will need to pay more attention to 'Affordability and cost of living', and be aware that most of our supports comes from the 'Rich.' As we pay more attention to the lower-income class, not only can we win more support, but also provide a better society. We can do so by increasing the minimum wage and decreasing the tax overall. Although it may seem to cause a economic problem for the government, this is critical for improving the well-being of all groups in the society. Once, the willingness to spend increases, it will create more job opportunities.

How we can make a bigger stand in the society?

In the main time, it is also important to study why do people choose the other parties, so that we can learn from their progress and mistakes. The Liberal party's primary political focusese includes: 'climate change', 'health care', 'general economy', 'affordable and living costs', 'Tax/senior issues'. Whereas 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 (*Compare the Party Platforms on These Top Election Issues* 2019). Going into depth about participants' attitude on specific political policies is another angle future studies can take to improve support for the Liberals as well as improve citizens' well-being in general.

codes used for this study

<https://github.com/joliw9282/2020-toronto-election-prediction>

Reference

Canadian Federal Election Results in Central Toronto. n.d. https://en.wikipedia.org/wiki/Canadian_federal_election_results_in_Central_Toronto.

Census Profile, 2016 Census Canada. n.d. Statistics Canada, Government of Canada. <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=POPC&Code1=0944&Geo2=PR&Code2=35&SearchText=toronto&SearchType=Begins&SearchPR=01&B1=Ethnic%20origin&TABID=1&type=1>.

Compare the Party Platforms on These Top Election Issues. 2019. CBC News. <https://newsinteractives.cbc.ca/elections/federal/2019/party-platforms/>.

John Ibbitson. 2019. *Election 2019 Results.* Global Mail. <https://www.theglobeandmail.com/politics/election2019/results/>.

Meredith Bond. n.d. *Federal Election Results: Who Won in Toronto and the Gta?* City News. <https://toronto.citynews.ca/2019/10/21/toronto-gta-federal-election-results/>.

Sample Size Calculator. 2020. calculator.net. <https://www.calculator.net/sample-size-calculator.html>.

Toronto Population. 2020. World Population Review. <https://worldpopulationreview.com/world-cities/toronto-population>.

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>.