title: "How the outbreak of COVID-19 influence popular vote?" author: "Weiwei Song" date: "1/22/2021" output: pdf_document

R Markdown

Back in February of 2020, the world situation is completely different. At that time, the unemployment rate in the United States was the lowest in decades, the stock market repeatedly hit record highs, and the situation of President Trump's re-election was very good. However, the COVID-19 pandemic changed this situation dramatically. Many people say that Trump's response to coronavirus is one of the main reason he failed in the election. Is that true?The count of everyone's ballot is called the popular vote. Under the model "Time For Change" model created by Professor Alan Abramowitz.Presidential elections can be predicted with Presidential approval ratings, economic growth and how long the White House has been controlled by one party instead of the other. The last factor can be ignored.Let's use the survey data to do an analysis whether coronavirus is a important indicator when people make their choice in the future "president".

```r
library(tidyverse)
```

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

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     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
anes_pilot_2020<-read_csv("../data/anes_pilot_2020ets_csv.csv")
```

```
##
## -- Column specification ---------------------------------------------------------
## cols(
##   .default = col_double(),
##   StartDate = col_character(),
##   EndDate = col_character(),
##   RecordedDate = col_character(),
##   ResponseId = col_character(),
##   qmetadata_Browser = col_character(),
##   qmetadata_Version = col_character(),
##   `_v2` = col_character(),
##   qmetadata_Resolution = col_character(),
##   relig1_11_TEXT = col_character(),
##   ethnic1 = col_character(),
##   ethnic2 = col_character(),
##   mauga = col_character(),
##   pk_cjus = col_character(),
##   pk_germ = col_character()
## )
## i Use `spec()` for the full column specifications.
```

```r
#clear all missing data, and rename the variable
trump_data <- anes_pilot_2020%>%
    select(apppres7:healthcarepres7)%>%
    filter(apppres7!=99,frnpres7!=99,immpres7!=99,
           econpres7!=99,covidpres7!=99,healthcarepres7!=99)%>%
```

```
      transmute(Job=apppres7,relation=frnpres7,immgration=immpres7,economy=econpres7,covid=covidpres7,he

trump_data$FORM<- ifelse(trump_data$covid==88,2,1)
trump_data$healthcare<- ifelse(trump_data$healthcare==77,NA,trump_data$healthcare)
trump_data$covid <-ifelse(trump_data$covid==88 ,NA,trump_data$covid)
trump_data<-trump_data%>%
    mutate(degree=coalesce(covid,healthcare))

summary_data<-trump_data%>%
        group_by(FORM, degree)%>%
        count(FORM)%>%
        ungroup()%>%
        group_by(FORM)%>%
        mutate(total=sum(n), prop=n/total)
```

There are two differnt wording in terms of healthcare or Covid-19, let's check if the two different working influence people'answer related to this question. FORM = 1: Do you approve, disapprove, or neither approve nor disapprove of the way Donald Trump is handing the coronavirus(COVID-19)outbreak? FORM = 2: Do you approve, disapprove, or neither approve nor disapprove of the way Donald Trump is handling health care?

we need to test the hypothesis whether for a random people in the survey group, whether the two question actually ask the same topic for them. In another word, the variable covid and healthcare follow the same distribution A good way is to use Chi-Square test.

```
ggplot(summary_data, aes(degree,prop,fill=factor(FORM)))+
  geom_col(position="dodge")+
  labs(title=" Comparing the result of two versions of questions", fill=" FORM", x= " the degree of disa
```
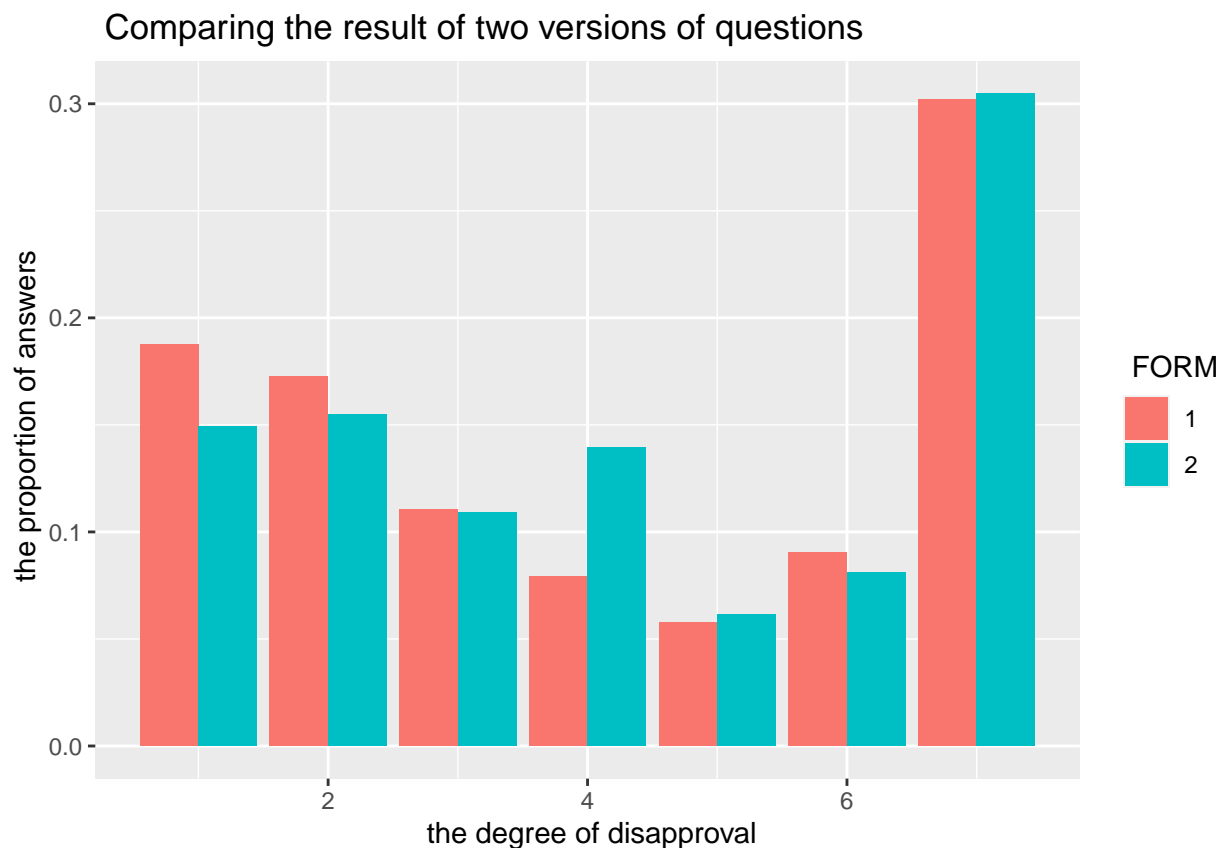
```r
test_data<- as.table(rbind(covid=table(trump_data$covid),
healthcare=table(trump_data$healthcare)))
chisq.test(test_data)
```

```
##
##  Pearson's Chi-squared test
##
## data:  test_data
## X-squared = 34.781, df = 6, p-value = 4.753e-06
```

```r
#make a rough table only includes three categories: approval, netrual,and disapproval

approval_data<-trump_data%>%
  group_by(FORM, degree)%>%
  count(FORM)%>%
  ungroup()%>%
  group_by(FORM)%>%
  mutate(total=sum(n), prop=n/total)%>%
  mutate(rate=if_else(degree<4, "approval",
                       if_else(degree==4,"neutral","disapproval")))%>%
  ungroup()%>%
  group_by(FORM,rate)%>%
  summarize(Total=sum(n))
```

```
## `summarise()` regrouping output by 'FORM' (override with `.groups` argument)
```

```r
ggplot(approval_data, aes(rate,Total,fill=factor(FORM)))+
  geom_col(position="dodge")+
  labs(title="The overall rating of the results",x="rate",y="count", fill="form")
```

The overall rating of the results
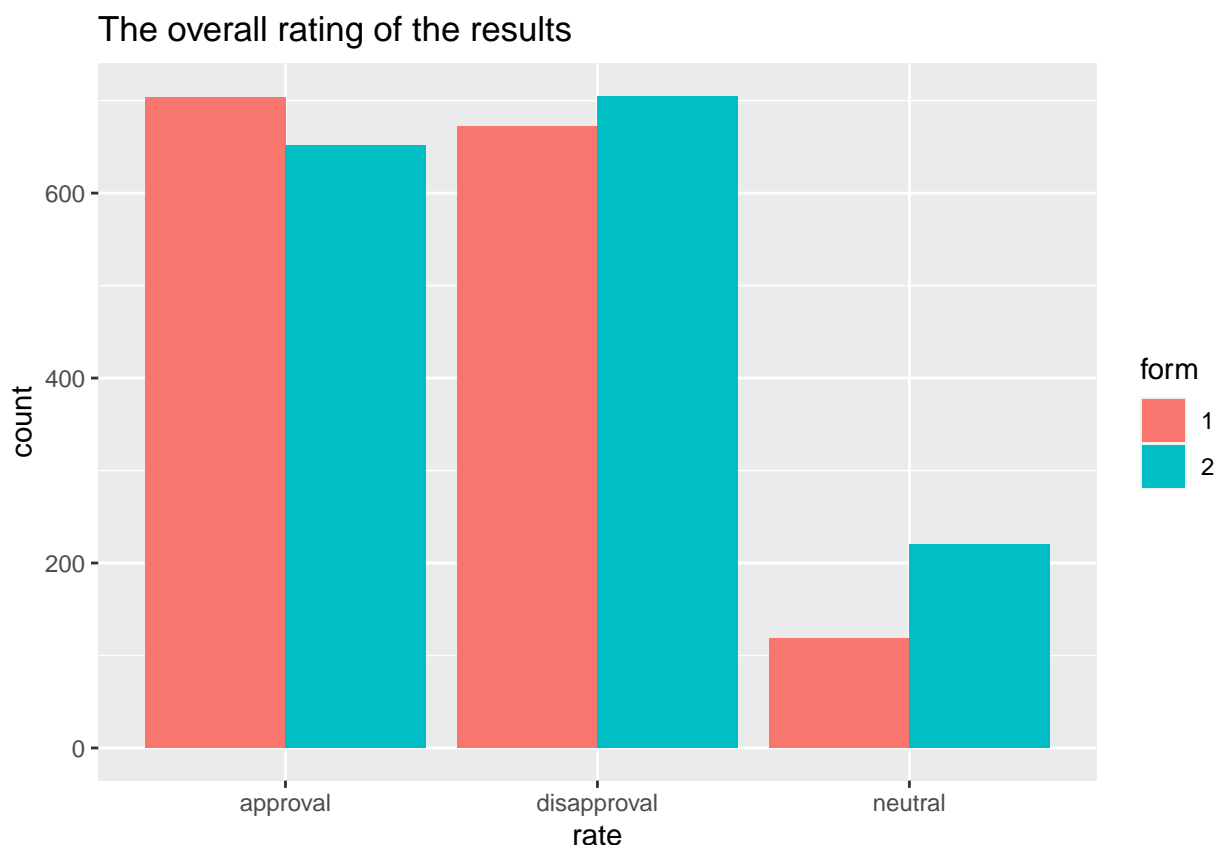
The p value is essentially zero, which doesn't give us enough evidence that the two wording techniques doesn't influence people's answer. People who answered question form 1 have less people hold neutral attitude.This makes sense because for majority of people, Trump's job performance regarding to coronavirus's outbreak in USA may not equal to what he did with heath care system. People may complains about the Trump's error in dealing with COVID-19 but they may find that his policy about health care system is acceptable. Overall, large percentage of citizens do disapprove extremely strongly about what Trump did to health care and COVID-19.

Is there any possibilities that people's concern about coronavirus and healthcare system really influence their choice?

```
cl_data<- anes_pilot_2020%>%
  select(hlthcare1:covid_elect)
```

Is people's understanding or the information media gives to them affect their option about the coronavitus?

```
mis_data<- anes_pilot_2020%>%
  select(mis_covid1:conf_covid2)
```