Replication of

Dan Xu

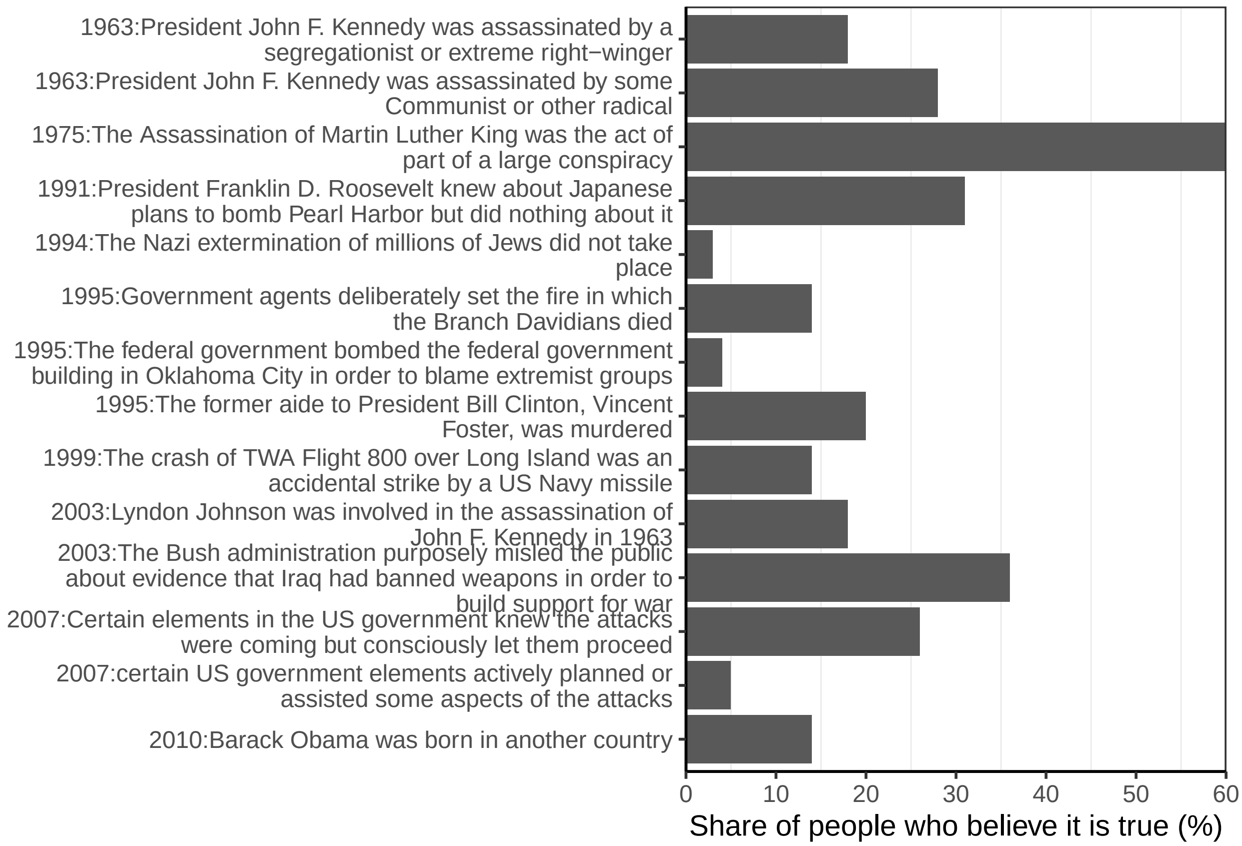
3/26/2021

Abstract

"Fake news" has become a recurring theme in American politics since the 2016 U.S. election. Many have raised the question of whether social media plays a role in spreading political misinformation and disinformation. This paper is a replication study of two economists Allcott and Gentzkow on fake news propogation in the 2016 U.S. election. Both of the authors examined how deliberately fabricated, as opposed to erroneous or biased stories from well-known or little-known sources spread on social media platforms, such as Facebook.By conducting a survey with 1200 participants a few months after the 2016 U.S. election, the authors found that fake news did not swing the election as people barely remembered it.Therefore, people's voting behaviors in the 2016 election had not been affected by fake news in social media.

Introduction

Written by Stanford University professor Matthew Gentzkow and NYU professor Hunt Allcott, the original paper aims to address the following questions: Has social media become a major platform for people to receive political information? How do American adults trust information they see on social media? How would such information affect their voting behaviors and their candidate preferences? Fake news is no news. In fact, it has been permeated in American politics for decades. The following graphs illustrate how the political conspiracy theories spread between 1963 and 2010. We can see that the mostly widely circulated one is that the assassination of Martin Luther King was the act of part of a large conspiracy in 1975. 60 percent of the American population believe that it is true, followed by the second mostly widely spread conspiracy of 2003, where the Bush administration was believed to purposely misled the publi about evidencne that Iraq banned weapons in order to build support for war.If we pay close attention to the conspiracies, it seems that there is no correlation between media and the spread of such conspiracy theories. The graphs above illustrate the political conspiracy theories spread between 1963 and 2010. We can see that the mostly widely circulated one is that the assacination of Martin Luther King was the act of part of a large conspiracy in 1975. 60 percent of the American population believe that it is true, followed by the second mostly widely spread consiracy of 2003, where the Bush administration was believed to purposely misled the publi about evidencne that Iraq banned weapons in order to build support for war.If we pay close attention to the conspiracies, it seems that there is no correlation between media and the spread of such conspiracy theories.


1963:President John F. Kennedy was assassinated by a segregationist or extreme right−winger
1963:President John F. Kennedy was assassinated by some Communist or other radical
1975:The Assassination of Martin Luther King was the act of part of a large conspiracy
1991:President Franklin D. Roosevelt knew about Japanese plans to bomb Pearl Harbor but did nothing about it
1994:The Nazi extermination of millions of Jews did not take place
1995:Government agents deliberately set the fire in which the Branch Davidians died
1995:The federal government bombed the federal government building in Oklahoma City in order to blame extremist groups
1995:The former aide to President Bill Clinton, Vincent Foster, was murdered
1999:The crash of TWA Flight 800 over Long Island was an accidental strike by a US Navy missile
2003:Lyndon Johnson was involved in the assassination of John F. Kennedy in 1963
2003:The Bush administration purposely misled the public about evidence that Iraq had banned weapons in order to build support for war
2007:Certain elements in the US government knew the attacks were coming but consciously let them proceed
2007:certain US government elements actively planned or assisted some aspects of the attacks
2010:Barack Obama was born in another country
0
10
20
30
40
50
60
Share of people who believe it is true (%)

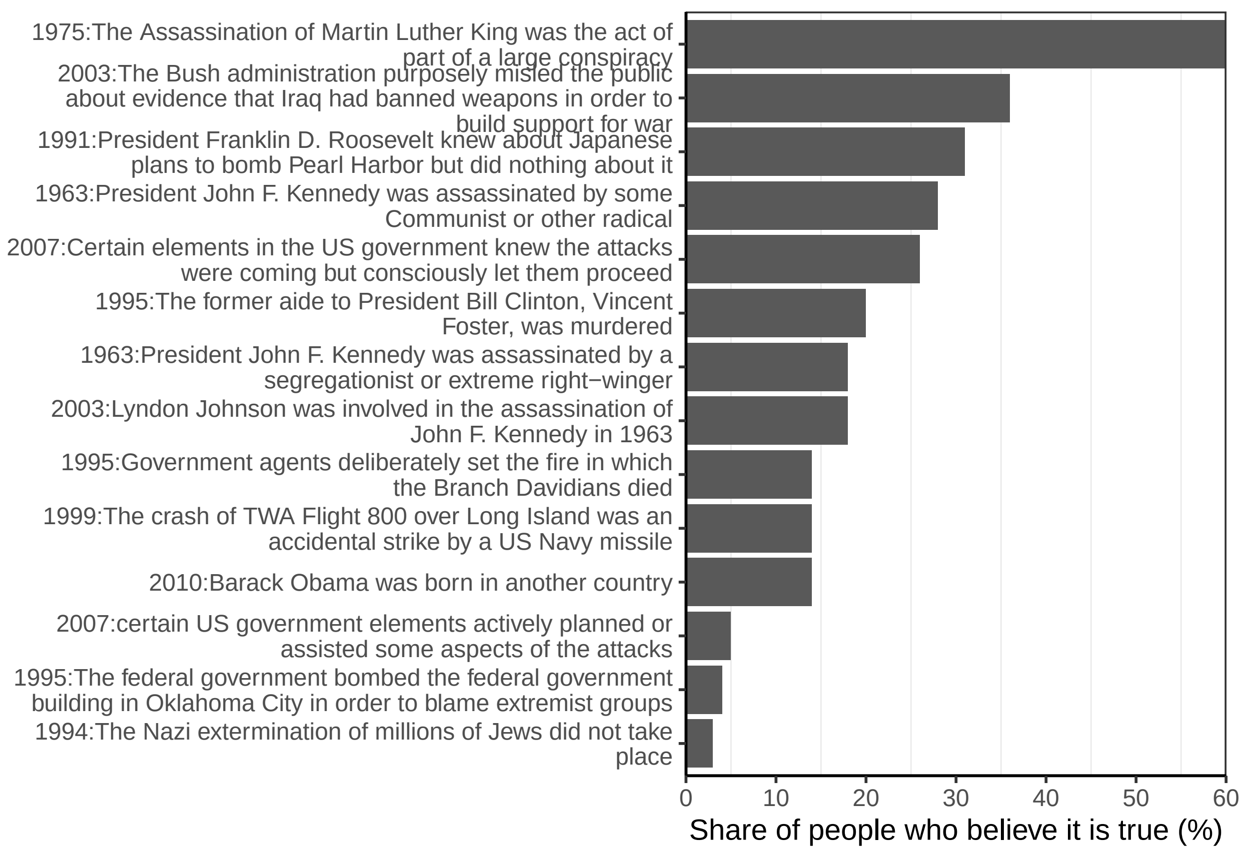

1975:The Assassination of Martin Luther King was the act of part of a large conspiracy
2003:The Bush administration purposely misled the public about evidence that Iraq had banned weapons in order to build support for war
1991:President Franklin D. Roosevelt knew about Japanese plans to bomb Pearl Harbor but did nothing about it
1963:President John F. Kennedy was assassinated by some Communist or other radical
2007:Certain elements in the US government knew the attacks were coming but consciously let them proceed
1995:The former aide to President Bill Clinton, Vincent Foster, was murdered
1963:President John F. Kennedy was assassinated by a segregationist or extreme right−winger
2003:Lyndon Johnson was involved in the assassination of John F. Kennedy in 1963
1995:Government agents deliberately set the fire in which the Branch Davidians died
1999:The crash of TWA Flight 800 over Long Island was an accidental strike by a US Navy missile
2010:Barack Obama was born in another country
2007:certain US government elements actively planned or assisted some aspects of the attacks
1995:The federal government bombed the federal government building in Oklahoma City in order to blame extremist groups
1994:The Nazi extermination of millions of Jews did not take place
0
10
20
30
40
50
60
Share of people who believe it is true (%)

Table 1: Results

Pro	title
Clinton	41
Trump	115

Table 2: Results

Pro	total
Clinton	5518.6
Trump	26156.3

The authors conducted a survey after the presidential election of 2016 to gather data. Specifically, Gentzkow and Allcott ran an online survey of 1,208 US adults recruited from MonkeyMonday in late November, a few weeks after the election. Given that the nature of the recruitment method, the sample size for the survey was not randomly selected. It it indeed biased and the data may not be representative of the population. The main goal of the survey was designed to obtain information about how people rely on social media as an information source to get informed; how exposure to fake news affects people's political behaviors; who are more likely to believes fake News.

Data

Dataset

As the authors put it, there were far more pro-Trump fake news shares — 30.3 million shares for false stories that favored Trump, versus 7.6 million that favored Clinton.My replicated results confirm that there were 41 fake new stories that had been shared on Facebook regarding Clinton, and 115 regarding Trump. However, the total number of fake news stories shared on Facebook was 31674.9 combined; and it was 5518.6 and 26156.3 for Clinton and Trump respecitively. The auhtors would need to specify how they calculated fake news shares on Facebook.

Post - election survey The study was conducted through a post - election survey and a database to understand how people produce, circulate, and receive political fake news. Alongside questions about the respondents' demographics and political affiliations, the survey asked participants for their "most important source of news and information" about the 2016 election. Then, the respondents were presented with 15 news headlines, equally split between pro-Clinton and pro-Trump news, including some delibrately fabricated false stories, fake stories that really existed, and some true ones as well, and placebo headlines — which were untrue and made-up claims that were neither real nor fake. The respondents then were asked to indicate if they "recall seeing this reported or discussed prior to the election?' Following this quesiton, the respondents were asked to evalute the authenticity of the news stories. The question was"At the time of the election, would your best guess have been that this statement was true?"

In addition, to understand how political fake news was produced, circulated and received amongst American adults, after aquiring consent from the survey participants to participate and a promise to provide their best answers, the survey asked the respondents to provide their demographic information by answering question including their party identificaetion/affliation, who and when they voted for in the 2016 presidential election, and their gender, educational background, race/ethnicity, zipcode. The survey also asked about 2016 election news consumption, including how much time they spent on reading, watching, or listening to election news in general and on social media in particular, and asked them to identify the most important source of news and information about the 2016 election.

Accoriding to the authors, the survey was distributed to 1,208 participants.However, there are 1437 observatiions because some respondents have not completed the survey or have withdrawn from the survey.

Descriptive statistics with missing values

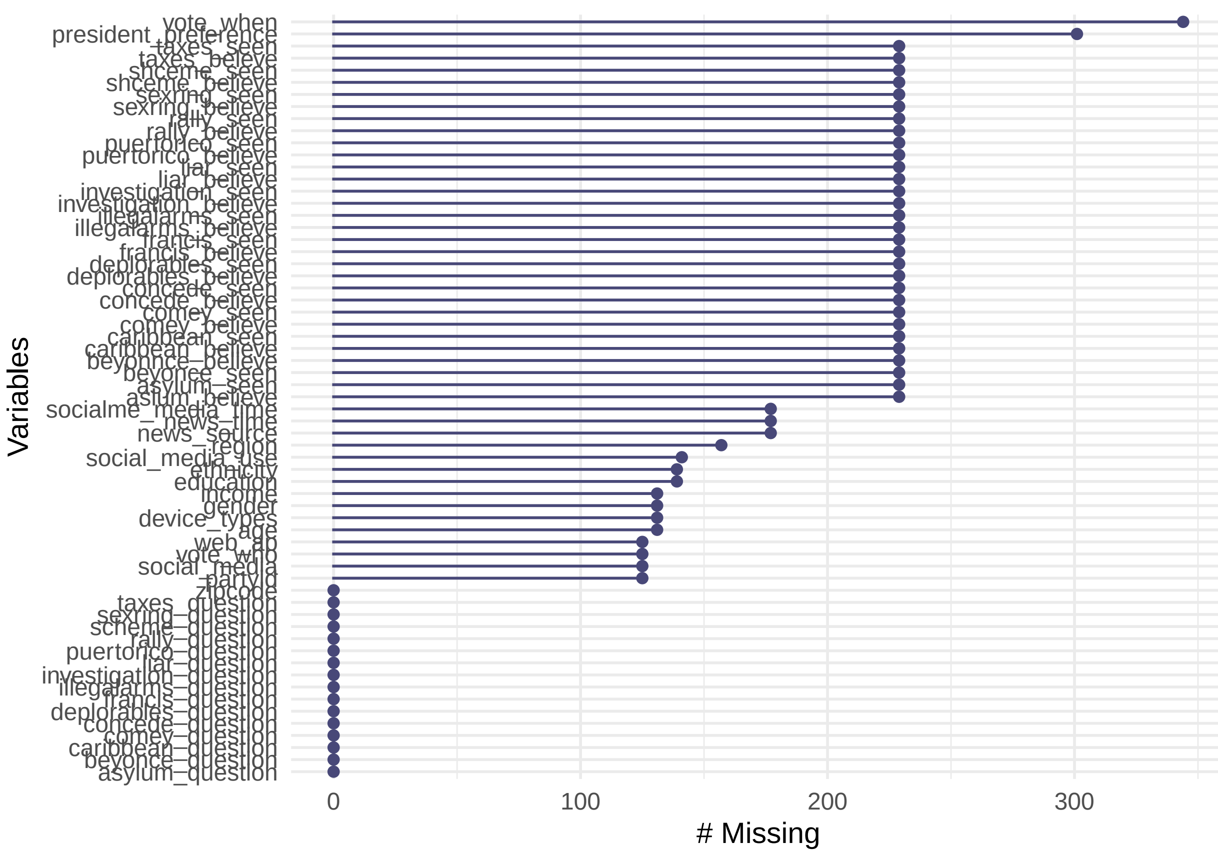
Before we delve into the analysis, let us take a look at the features of the data.

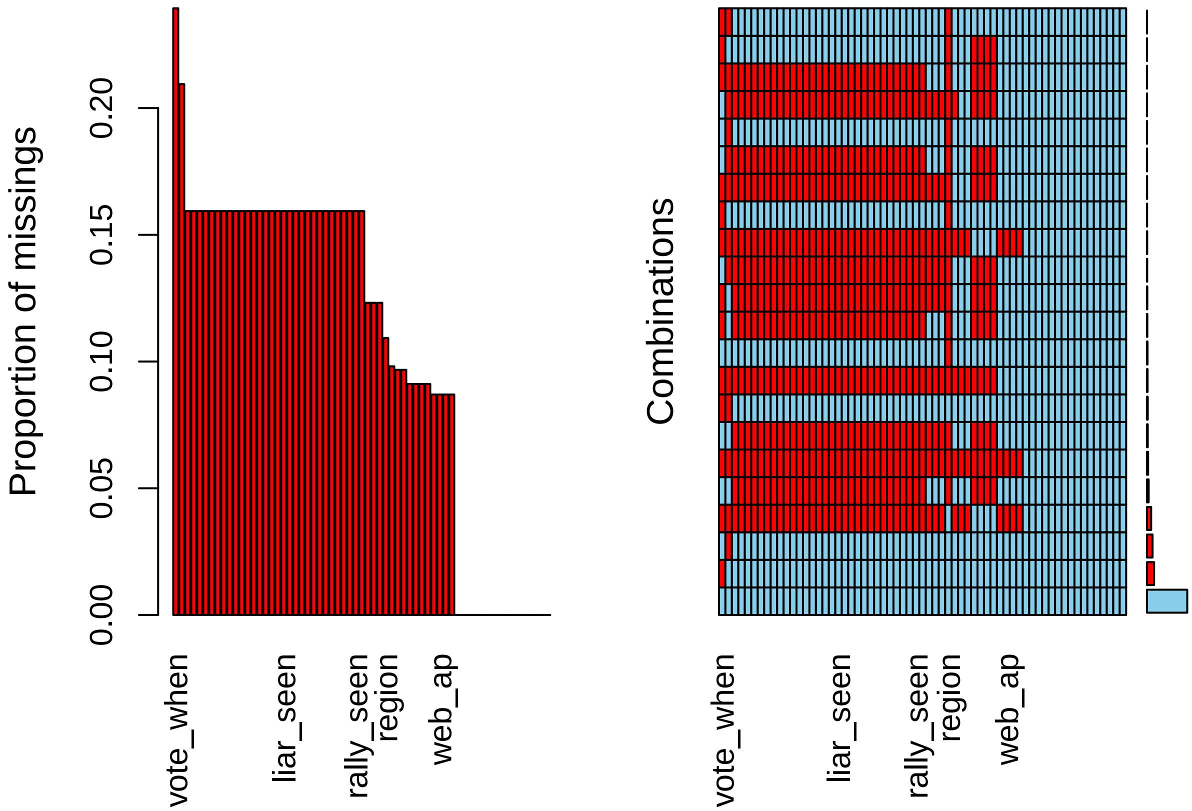
Descriptions	Value
Sample size (nrow)	1437
No. of variables (ncol)	63
No. of numeric/interger variables	47
No. of factor variables	0
No. of text variables	16
No. of logical variables	0
No. of identifier variables	0
No. of date variables	0
No. of zero variance variables (uniform)	0
%. of variables having complete cases	0% (0)
%. of variables having >0% and <50% missing cases	100% (63)
%. of variables having >=50% and <90% missing cases	0% (0)
%. of variables having >=90% missing cases	0% (0)

As the table above shows, there are 63 variables with 1437 observations in the dataset. However, there are a lot of missing values contained in this dataset. # Descriptive analysis of the survey sample

x
887
63

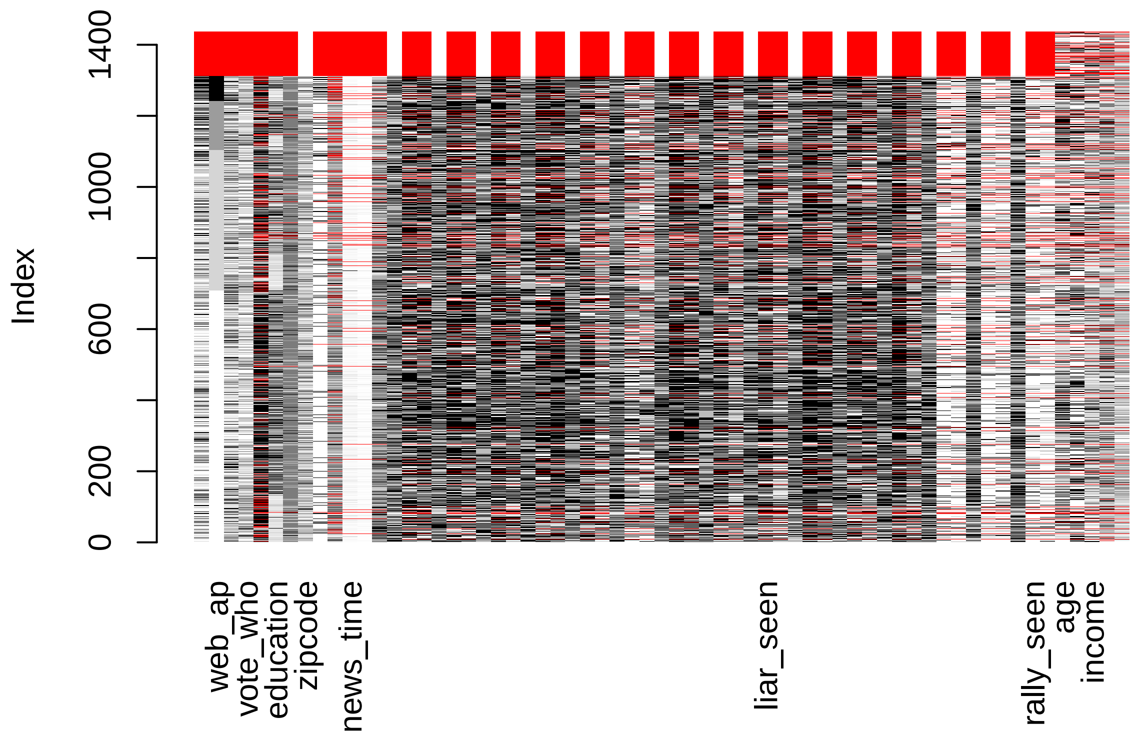
As shown above, there are 887 complete cases.






```
##
##  Variables sorted by number of missings:
##                Variable      Count
##               vote_when 0.23938761
##    president_preference 0.20946416
##             francis_seen 0.15935978
##          francis_believe 0.15935978
##         illegalarms_seen 0.15935978
##      illegalarms_believe 0.15935978
##              asylum_seen 0.15935978
##            aslum_believe 0.15935978
##             sexring_seen 0.15935978
##          sexring_believe 0.15935978
##               comey_seen 0.15935978
##            comey_believe 0.15935978
##               taxes_seen 0.15935978
##             taxes_beleve 0.15935978
##             beyonce_seen 0.15935978
##         beyonnce_believe 0.15935978
##          puertorico_seen 0.15935978
##       puertorico_believe 0.15935978
##                liar_seen 0.15935978
##             liar_believe 0.15935978
##              shceme_seen 0.15935978
##           shceme_believe 0.15935978
##       investigation_seen 0.15935978
##    investigation_believe 0.15935978
##            caribbean_seen 0.15935978
```

```
##        caribbean_believe 0.15935978
##         deplorables_seen 0.15935978
##      deplorables_believe 0.15935978
##              concede_seen 0.15935978
##          concede_believe 0.15935978
##                rally_seen 0.15935978
##             rally_believe 0.15935978
##                 news_time 0.12317328
##      socialme_media_time 0.12317328
##               news_source 0.12317328
##                    region 0.10925539
##          social_media_use 0.09812109
##                 education 0.09672930
##                 ethnicity 0.09672930
##                       age 0.09116214
##                    gender 0.09116214
##                    income 0.09116214
##              device_types 0.09116214
##              social_media 0.08698678
##                    web_ap 0.08698678
##                   partyid 0.08698678
##                  vote_who 0.08698678
##                   zipcode 0.00000000
##          francis_question 0.00000000
##     illegalarms_question 0.00000000
##           asylum_question 0.00000000
##          sexring_question 0.00000000
##            comey_question 0.00000000
##            taxes_question 0.00000000
##          beyonce_question 0.00000000
##      puertorico_question 0.00000000
##             liar_question 0.00000000
##           scheme_question 0.00000000
##   investigation_question 0.00000000
##        caribbean_question 0.00000000
##     deplorables_question 0.00000000
##          concede_question 0.00000000
##            rally_question 0.00000000
```




```r
pct_miss(Survey_clean) # percentage of missing value in the data.
```

Summary with missing values

```
## [1] 10.65491
```

```r
n_miss(Survey_clean) # number of missing values
```

```
## [1] 9646
```

```r
n_complete(Survey_clean) # without missing value
```

```
## [1] 80885
```

```r
as_shadow(Survey_clean) # A matrix with missing and non missing:
```

```
## # A tibble: 1,437 x 63
##    social_media_NA web_ap_NA partyid_NA vote_who_NA vote_when_NA education_NA
##    <fct>           <fct>     <fct>      <fct>       <fct>        <fct>
##  1 !NA             !NA       !NA        !NA         !NA          !NA
##  2 !NA             !NA       !NA        !NA         !NA          !NA
##  3 !NA             !NA       !NA        !NA         !NA          !NA
##  4 !NA             !NA       !NA        !NA         !NA          !NA
##  5 !NA             !NA       !NA        !NA         !NA          !NA
##  6 !NA             !NA       !NA        !NA         NA           !NA
```

```
##  7 !NA             !NA       !NA        !NA         !NA          !NA
##  8 !NA             !NA       !NA        !NA         NA           !NA
##  9 !NA             !NA       !NA        !NA         !NA          !NA
## 10 NA              NA        NA         NA          NA           NA
## # ... with 1,427 more rows, and 57 more variables: ethnicity_NA <fct>,
## #   zipcode_NA <fct>, social_media_use_NA <fct>, president_preference_NA <fct>,
## #   news_time_NA <fct>, socialme_media_time_NA <fct>, news_source_NA <fct>,
## #   francis_question_NA <fct>, francis_seen_NA <fct>, francis_believe_NA <fct>,
## #   illegalarms_question_NA <fct>, illegalarms_seen_NA <fct>,
## #   illegalarms_believe_NA <fct>, asylum_question_NA <fct>,
## #   asylum_seen_NA <fct>, aslum_believe_NA <fct>, sexring_question_NA <fct>,
## #   sexring_seen_NA <fct>, sexring_believe_NA <fct>, comey_question_NA <fct>,
## #   comey_seen_NA <fct>, comey_believe_NA <fct>, taxes_question_NA <fct>,
## #   taxes_seen_NA <fct>, taxes_beleve_NA <fct>, beyonce_question_NA <fct>,
## #   beyonce_seen_NA <fct>, beyonnce_believe_NA <fct>,
## #   puertorico_question_NA <fct>, puertorico_seen_NA <fct>,
## #   puertorico_believe_NA <fct>, liar_question_NA <fct>, liar_seen_NA <fct>,
## #   liar_believe_NA <fct>, scheme_question_NA <fct>, shceme_seen_NA <fct>,
## #   shceme_believe_NA <fct>, investigation_question_NA <fct>,
## #   investigation_seen_NA <fct>, investigation_believe_NA <fct>,
## #   caribbean_question_NA <fct>, caribbean_seen_NA <fct>,
## #   caribbean_believe_NA <fct>, deplorables_question_NA <fct>,
## #   deplorables_seen_NA <fct>, deplorables_believe_NA <fct>,
## #   concede_question_NA <fct>, concede_seen_NA <fct>, concede_believe_NA <fct>,
## #   rally_question_NA <fct>, rally_seen_NA <fct>, rally_believe_NA <fct>,
## #   age_NA <fct>, gender_NA <fct>, income_NA <fct>, region_NA <fct>,
## #   device_types_NA <fct>
```

```
bind_shadow(Survey_clean) #The initial matrix concatenated with the matrix with missing and non missing
```

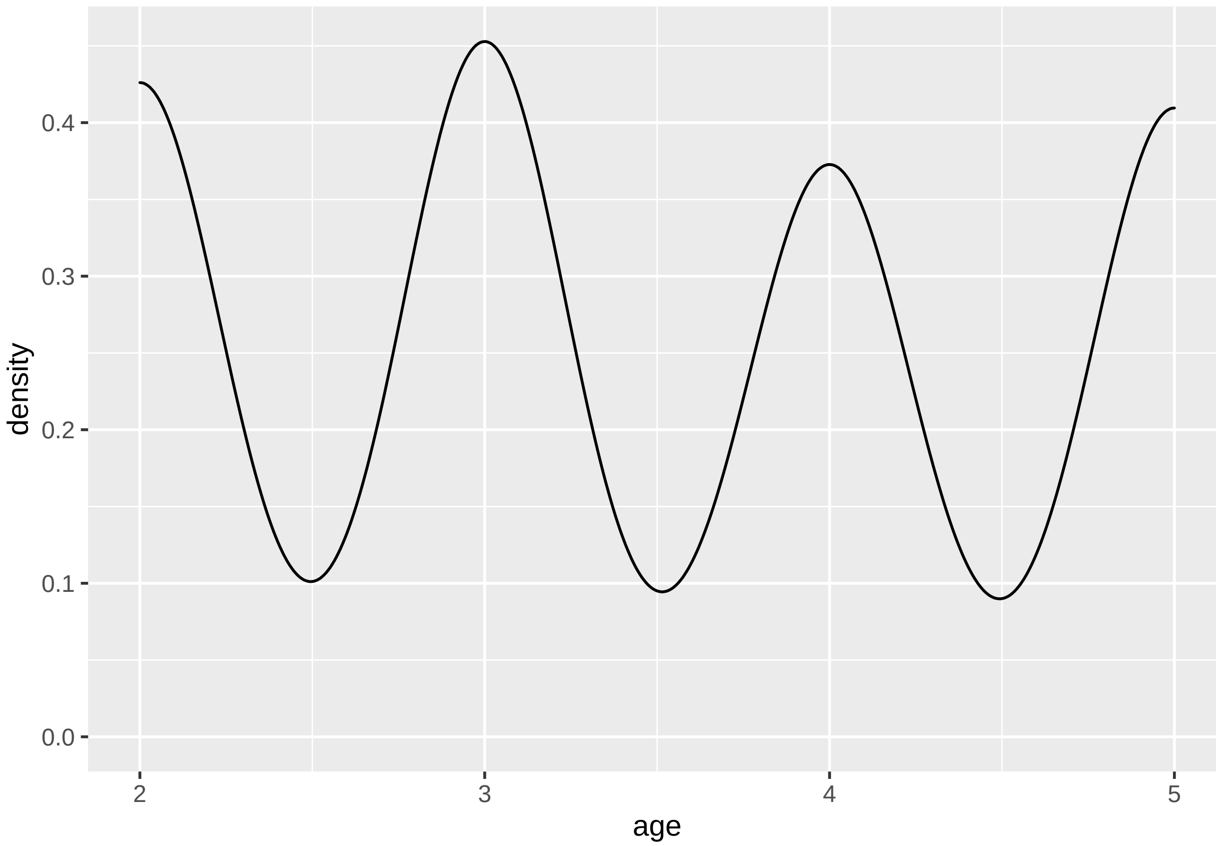
```
## # A tibble: 1,437 x 126
##    social_media web_ap partyid vote_who vote_when education ethnicity zipcode
##           <int>  <int>   <int>    <int>     <int>     <int>     <int> <chr>
##  1            1      1       1        1         4         2         5 "55417"
##  2            3      3       2        1         4         2         5 "91913"
##  3            1      4       7        4         4         1         1 "00000"
##  4            1      1       1        1         4         2         5 "15205 "
##  5            2      2       6        2         3         2         2 "47404"
##  6            2      2       4        6        NA         2         5 "55712"
##  7            1      2       3        1         4         2         6 "02536"
##  8            2      2       4        6        NA         3         5 "34287"
##  9            2      3       3        1         2         2         2 "53703"
## 10           NA     NA      NA       NA        NA        NA        NA ""
## # ... with 1,427 more rows, and 118 more variables: social_media_use <int>,
## #   president_preference <int>, news_time <int>, socialme_media_time <int>,
## #   news_source <int>, francis_question <chr>, francis_seen <int>,
## #   francis_believe <int>, illegalarms_question <chr>, illegalarms_seen <int>,
## #   illegalarms_believe <int>, asylum_question <chr>, asylum_seen <int>,
## #   aslum_believe <int>, sexring_question <chr>, sexring_seen <int>,
## #   sexring_believe <int>, comey_question <chr>, comey_seen <int>,
## #   comey_believe <int>, taxes_question <chr>, taxes_seen <int>,
## #   taxes_beleve <int>, beyonce_question <chr>, beyonce_seen <int>,
## #   beyonnce_believe <int>, puertorico_question <chr>, puertorico_seen <int>,
## #   puertorico_believe <int>, liar_question <chr>, liar_seen <int>,
```

```
## #   liar_believe <int>, scheme_question <chr>, shceme_seen <int>,
## #   shceme_believe <int>, investigation_question <chr>,
## #   investigation_seen <int>, investigation_believe <int>,
## #   caribbean_question <chr>, caribbean_seen <int>, caribbean_believe <int>,
## #   deplorables_question <chr>, deplorables_seen <int>,
## #   deplorables_believe <int>, concede_question <chr>, concede_seen <int>,
## #   concede_believe <int>, rally_question <chr>, rally_seen <int>,
## #   rally_believe <int>, age <int>, gender <int>, income <int>, region <int>,
## #   device_types <int>, social_media_NA <fct>, web_ap_NA <fct>,
## #   partyid_NA <fct>, vote_who_NA <fct>, vote_when_NA <fct>,
## #   education_NA <fct>, ethnicity_NA <fct>, zipcode_NA <fct>,
## #   social_media_use_NA <fct>, president_preference_NA <fct>,
## #   news_time_NA <fct>, socialme_media_time_NA <fct>, news_source_NA <fct>,
## #   francis_question_NA <fct>, francis_seen_NA <fct>, francis_believe_NA <fct>,
## #   illegalarms_question_NA <fct>, illegalarms_seen_NA <fct>,
## #   illegalarms_believe_NA <fct>, asylum_question_NA <fct>,
## #   asylum_seen_NA <fct>, aslum_believe_NA <fct>, sexring_question_NA <fct>,
## #   sexring_seen_NA <fct>, sexring_believe_NA <fct>, comey_question_NA <fct>,
## #   comey_seen_NA <fct>, comey_believe_NA <fct>, taxes_question_NA <fct>,
## #   taxes_seen_NA <fct>, taxes_beleve_NA <fct>, beyonce_question_NA <fct>,
## #   beyonce_seen_NA <fct>, beyonnce_believe_NA <fct>,
## #   puertorico_question_NA <fct>, puertorico_seen_NA <fct>,
## #   puertorico_believe_NA <fct>, liar_question_NA <fct>, liar_seen_NA <fct>,
## #   liar_believe_NA <fct>, scheme_question_NA <fct>, shceme_seen_NA <fct>,
## #   shceme_believe_NA <fct>, investigation_question_NA <fct>,
## #   investigation_seen_NA <fct>, ...
```

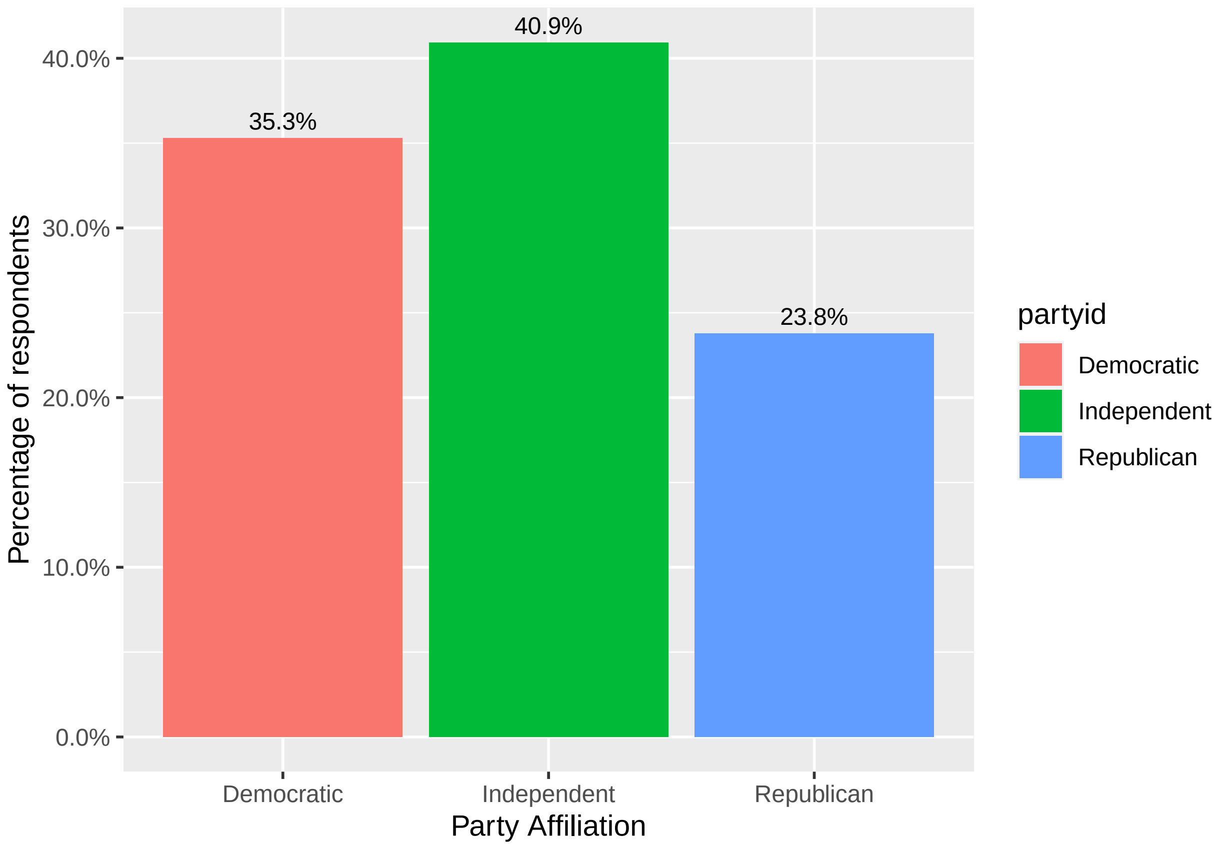
```
ggplot( bind_shadow(Survey_clean),
       aes(x = age,
           fill = partyid)) +
  geom_density(alpha=0.5)
```

Plot with missing values

```
## Warning: Removed 131 rows containing non-finite values (stat_density).
```

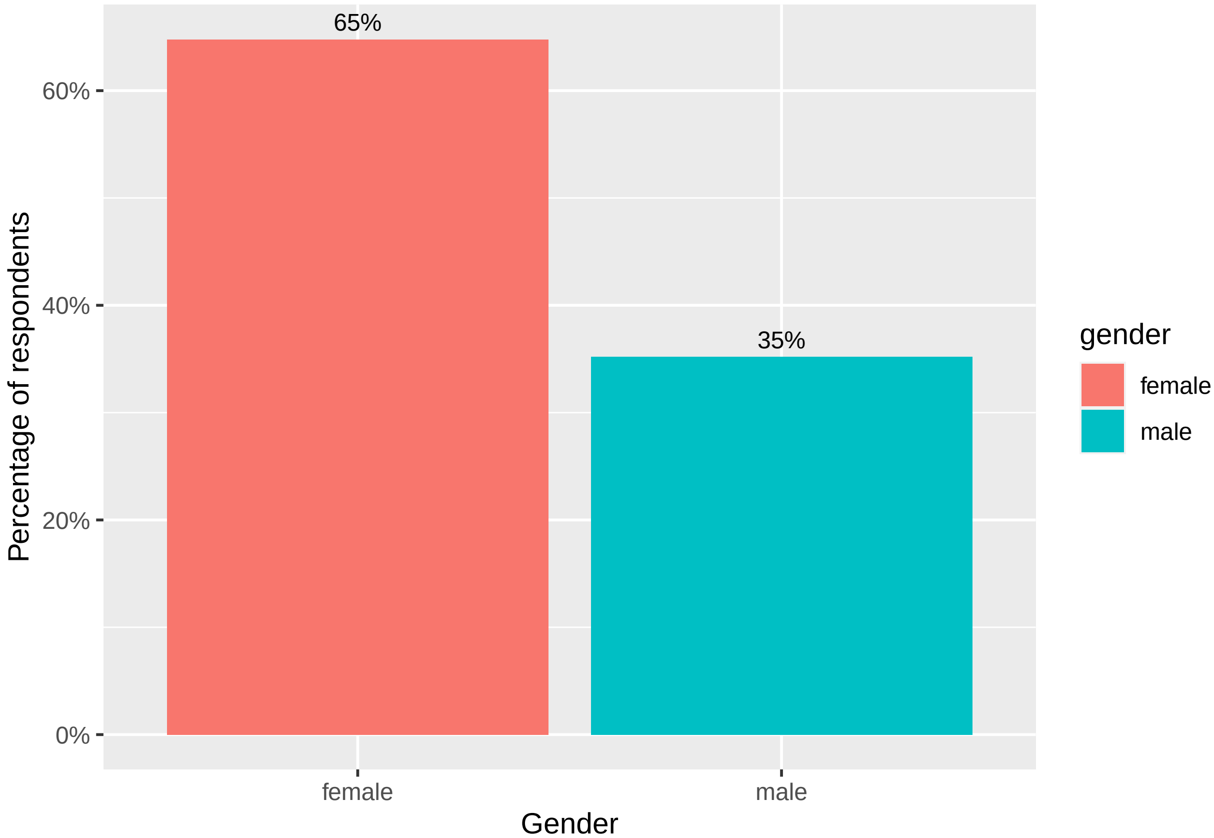

density
0.0
0.1
0.2
0.3
0.4
2
3
4
5
age

Party affliation



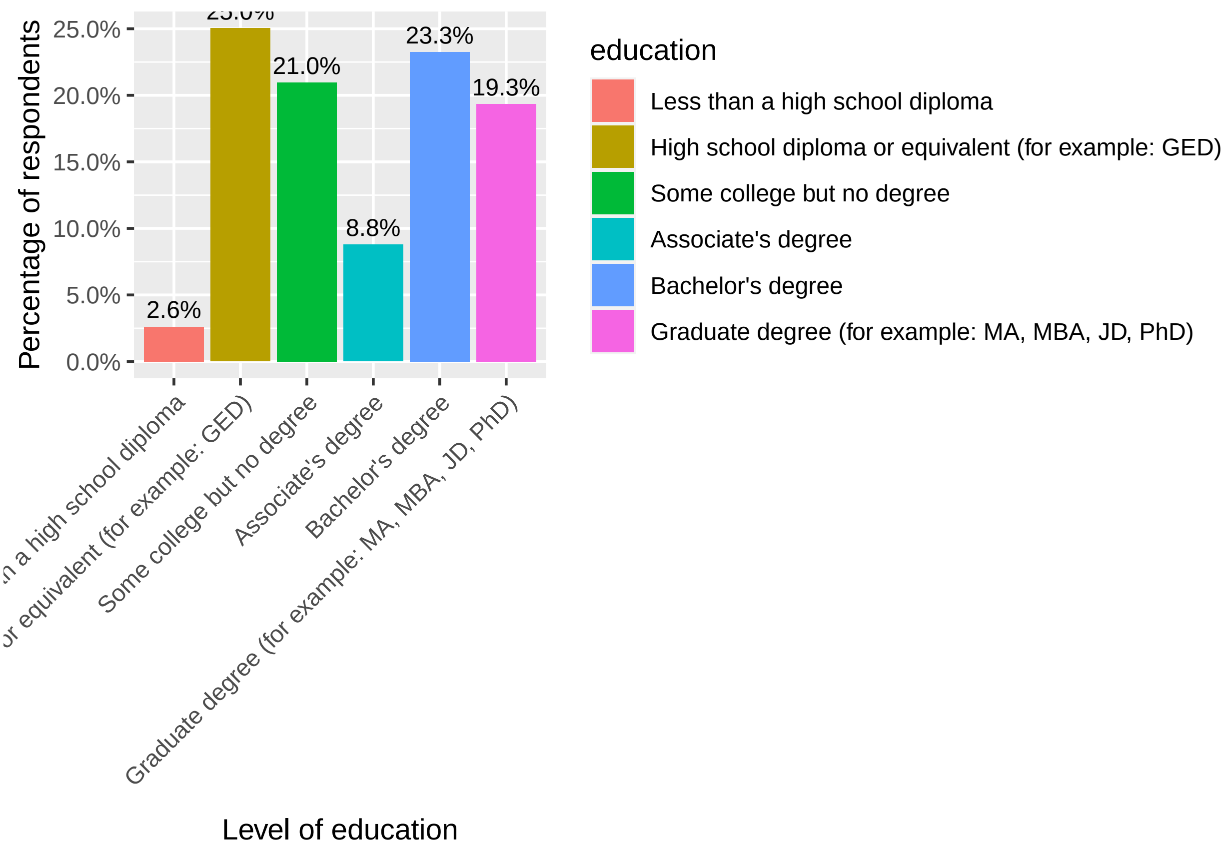
In terms of party arlliation, the sample size contains 49 percent of respondents who identified as Republicans, and 34 percent of respondents who identified as Democrats. However, the proportion is seemingly inconsitent with the polling results from major U.S. polling agencies. According to Gallup, "37% of registered voters identified as independents, 33% as Democrats and 26% as Republicans" in the year of 2017. The inconsistency is potentially a result of different coding strategies. In the survey, respondents were asked to self identified themselves with one of the following categories, namely, "Strongly Republican","Weakly Republican","Lean toward Republican party","Independent","Lean toward Democratic party","Weakly Democratic", "Strongly Democratic", where the "Lean toward Republican party" and "Lean toward Democratic party" categories can also be lumpped into the "Independent" category. If we lumped the respondents who lean toward either party, then the results would be different as the independents would be nearly 41% as the Republicans and Democrats being 35.3 percent and 23.8 percent. In this case, we can conclude that the sample size is skewed toward Republicans.

Gender



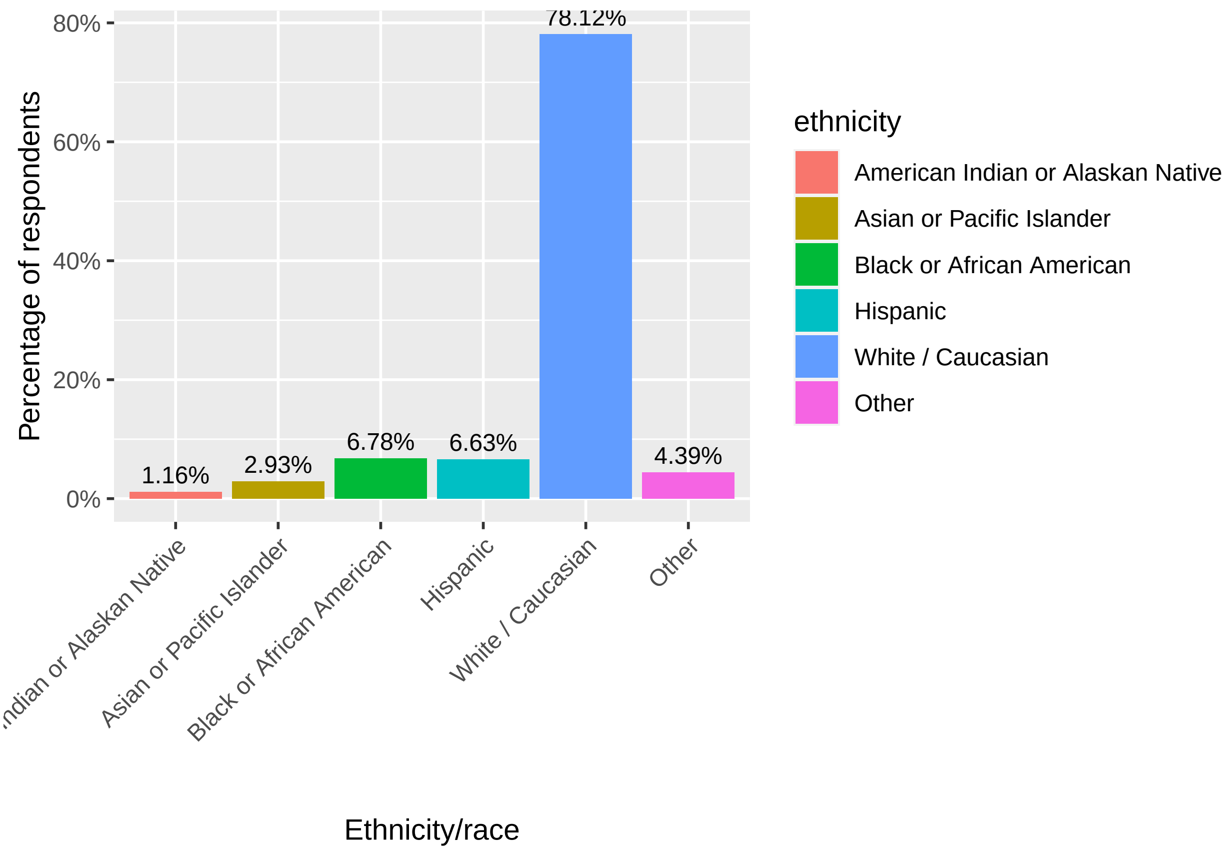
As the graph above shows, the sample is disproportionately female with nearly 2/3 of the survey respondents were women, and only 1/3 of the respondents were men.

education



The sample is also skewed toward highly educated people. As the plot shows, only 27.6 percent of the survey respondents received a high school diploma or less leducation; while more than 70 percent of the survey participants reported receiving at least some college degree level education, more than 40 percent of them reported holding a bachelor's or higher degree.

Ethnicity



As the graph shows, more than 70% of the study participants were white or Caucasian, suggesting that the sample was not drawn using random sampling techniques.

In summary, the sample was disproportionaly white, female, andd highy education, which lacks representativeness and needs to be adjusted with weights.

survey weights

Survey weights

Raking

For public opinion surveys, researchers use multiple methods to adjust unweighted samples. One of the most prevant method for weighting is iterative proportional fitting, which is also referred to as raking. With raking, researchers select “a set of variables where the population distribution is known, and the procedure iteratively adjusts the weight for each case until the sample distribution aligns with the population for those variables (Pew[cite]).” In this study, to adjust the weights, I will select a few variables, such as party identification, gender,education to match the desired population distribution. Then the weights obtained will be adjusted so that the sample in the correct proportion.As mentioned earlier, Raking is relatively simple to implement as it only requires knowing the marginal proportions for each variable used in weighting.

Table 3: Average weights table

Variable	Mean	Notes
Educ_BachelorsOrMore	0.2673650	Weighted average of 18-24 years and 25+ years from 2014 ACS 5-year estimates
Educ_HSOrLess	0.4197737	Weighted average of 18-24 years and 25+ years from 2014 ACS 5-year estimates
Democrat	0.3684771	Weighted results from 2012 ANES, only considering people who were D, R, or I
Independent	0.3429180	Weighted results from 2012 ANES, only considering people who were D, R, or I
Republican	0.2886049	Weighted results from 2012 ANES, only considering people who were D, R, or I
Age	47.1549515	Average age of people 18 or older in the US in 2015
Race_White	0.6160000	https://www.census.gov/quickfacts/table/PST0452 js-a
Male	0.4867939	Share male of pople 18 or older in the US in 2015
Income	76.1625160	Mean income (this is in 2014 inflation-adjusted dollars), times 1.021 for 2.1% inflation (from http://www.usinflationcalculator.com/)
NewsSource_Social	1.2400000	3 is Often, 0 is Never (numbering changed from original Pew coding)
NewsSource_Web	1.5800000	3 is Often, 0 is Never (numbering changed from original Pew coding)

```r
# clean data without NAs
survey_cleanse <- as.data.frame(na.omit(Survey_clean))
```

```r
#recode ethnicity to two groups, white vs. non white
survey_cleanse$ethnicity<-
  fct_collapse(survey_cleanse$ethnicity,
               White = c("White / Caucasian"),
               Non.White = c("American Indian or Alaskan Native","Asian or Pacific Islander","Black or A
```

```r
# recode education
# Educ_BachelorsOrMore 0.2673650
# Educ_HSOrLess 0.4197737
# Educ_Other 0.3128613

survey_cleanse$education<-
  fct_collapse(survey_cleanse$education,
               Educ_BachelorsOrMore = c("Bachelor's degree","Graduate degree (for example: MA, MBA, JD,
               Educ_HSOrLess = c("Less than a high school diploma","High school diploma or equivalent (i
               Educ_Other = c("Some college but no degree","Associate's degree"))
```

In this study, the weighting variables were raked according to their marginal distributions for each of the demographic variables. For this survey, I used the raking method to adjust the data according to population parameters.As the above table shows, I used six variables to define raking weights (excluding income and news sources), namely education, gender, partyid, age, race. I got relative frequencies of these variables using the package weights.

```r
library(weights)
```

```
## Loading required package: Hmisc

## Loading required package: lattice

## Loading required package: Formula

##
## Attaching package: 'Hmisc'

## The following object is masked from 'package:survey':
##
##     deff

## The following objects are masked from 'package:dplyr':
##
##     src, summarize

## The following objects are masked from 'package:base':
##
##     format.pval, units

## Loading required package: gdata
```

```
## gdata: read.xls support for 'XLS' (Excel 97-2004) files ENABLED.

##

## gdata: read.xls support for 'XLSX' (Excel 2007+) files ENABLED.

##
## Attaching package: 'gdata'

## The following objects are masked from 'package:dplyr':
##
##     combine, first, last

## The following object is masked from 'package:purrr':
##
##     keep

## The following object is masked from 'package:stats':
##
##     nobs

## The following object is masked from 'package:utils':
##
##     object.size

## The following object is masked from 'package:base':
##
##     startsWith

## Loading required package: mice

##
## Attaching package: 'mice'

## The following object is masked from 'package:stats':
##
##     filter

## The following objects are masked from 'package:base':
##
##     cbind, rbind
```

```
wpct(survey_cleanse$gender)
```

```
##    female      male
## 0.6561443 0.3438557
```

```
# gender
#  male    female
# 0.3438557 0.6561443

wpct(survey_cleanse$partyid)
```

```
##  Democratic Independent  Republican
##   0.3788050   0.3664036   0.2547914
```

```
# partyid
 # Democratic Independent  Republican
 #  0.3788050   0.3664036   0.2547914

wpct(survey_cleanse$age)
```

```
##         2         3         4         5
## 0.2311161 0.2818489 0.2423901 0.2446449
```

```
#        2         3         4         5
# 0.2311161 0.2818489 0.2423901 0.2446449
wpct(survey_cleanse$education)
```

```
##       Educ_HSOrLess           Educ_Other Educ_BachelorsOrMore
##           0.2175874            0.2998873            0.4825254
```

```
       # Educ_HSOrLess           Educ_Other Educ_BachelorsOrMore
       #     0.2175874            0.2998873            0.4825254

wpct(survey_cleanse$ethnicity)
```

```
## Non.White     White
## 0.1871477 0.8128523
```

```
# Non.White     White
# 0.1871477 0.8128523
```

The next step was to specify the population distribution of the selected variables in a target list. I used the survey mean tables provided by the authors.

```
gender <- c(0.4867939,0.5132061)
education <- c(0.4197737,0.3128613,0.2673650)
ethnicity <- c(0.384,0.6160000) # white vs. non white
partyid <- c(0.3684771,0.3429180,0.2886049)
# the Chilean Census 2002
# age  <- c(.163, .203, .195, .187, .253)
```

Turn them into numeric variables

```
survey_cleanse$gender <- as.numeric(survey_cleanse$gender)
```

```
## Warning: NAs introduced by coercion
```

```
survey_cleanse$education <- as.numeric(survey_cleanse$education)
survey_cleanse$ethnicity <- as.numeric(survey_cleanse$ethnicity)
survey_cleanse$partyid <- as.numeric(survey_cleanse$partyid)
```

```r
# definitions of target list
targets <- list(gender,education,ethnicity,partyid)
# important: to use the same variable names of the dataset
names(targets) <- c("gender","education","ethnicity","partyid")
# id variable
survey_cleanse$caseid <- 1:length(survey_cleanse$gender)
```

```r
library(anesrake)
anesrakefinder(targets, survey_cleanse, choosemethod = "total")
```

```
##     gender  education  ethnicity    partyid
## 0.00000000 0.43032073 0.39370462 0.06762694
```

```r
# https://community.rstudio.com/t/r-anesrake-error-error-in-x-weights-non-numeric-argument-for-binary-o
```

```r
# I apply the anesrake function as follows:
# The maximum weight value is five, weights greater than five will be truncated (cap = 5).
# The total differences between population and sample have to be greater than 0.05 so that to include a
# The maximum number of variables included in the raking procedure is five (nlim = 5).

outsave <- anesrake(targets, survey_cleanse, caseid = survey_cleanse$caseid,
  verbose= FALSE, cap = 5, choosemethod = "total",
  type = "pctlim", pctlim = .05 , nlim = 5,
  iterate = TRUE , force1 = TRUE)
```

```
## [1] "Raking converged in 20 iterations"
```

```r
# summary(outsave)
```

```r
#add weights to dataset
survey_cleanse$weightvec  <- unlist(outsave[1])
n  <- length(survey_cleanse$gender)
```

```r
# weighting loss
((sum(survey_cleanse$weightvec ^ 2) / (sum(survey_cleanse$weightvec)) ^ 2) * n) - 1
```

```
## [1] 0.525237
```

The observations were weighted using weighted means abstracted from national samples for national representativeness. The average weight for education was obtained from the weighted average of American adults who aged 18 and above from 2014 ACS 5-year estimates from United States Census Bureau. Accordingly, respondents who received a bachelor's degree or higher were assigned the weight average of 0.2673650; while other respondents at other levels of education were assigned the average mean of 0.4197737 as the weight. Likewise, for the party identification variable, Respondents who identified with Democrats were assigned the weighted mean of 0.3684771; those who identified themselves as Independents, including Republicans and Democrats who indicated that they leaned toward Independents, were assigned the weighted mean of 0.3429180; For Republican survey participants, they were assigned the average weight of 0.2886049. The weighted means for the party identification were drawns from the survey results of American National Election Studies 2012. For the average of the sample size, 47.1549515 was assigned to the sample as the average age of American adults. In terms of race, the white/Caucasion survey respondnents were assigned with a weighted average of 0.6160000, which represents their proportion of the American population. For sex ratio, Men were assigned with the weighted average of 0.4867939.

To sum up, in terms of the survey design, there are a few issues: the survey does not include attention checks. Attenton checks are a useful tool to to make sure survey respondents are really paying attention, and not simply randomly answer the questions for the sake of completing the survey as quickly as possible.The authors should include such attention check questions to insure the reliability of the survey responses. The income variable is missing in the original survey.

News headlines

T test

```
### non-parametric tests
t.test((survey_cleanse$francis_believe[survey_cleanse$francis_seen == 1]),
       survey_cleanse$francis_believe[survey_cleanse$francis_seen == 2])
```

```
##
##  Welch Two Sample t-test
##
## data:  (survey_cleanse$francis_believe[survey_cleanse$francis_seen == 1]) and survey_cleanse$francis_
## t = 0.98791, df = 348.7, p-value = 0.3239
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.07757464  0.23415658
## sample estimates:
## mean of x mean of y
##  2.224490  2.146199
```

small fake: Fake fake: Big true:

The survey participants were presented with 15 headlines, which were equally split between pro-Clinton and pro-Trump news stories. The authors categorized the stories into “BigFake”, “SmallFake”,FakeFake“,”BigTrue“, referring to ambiguously and unambiguously false stories, and unambiguously true ones, and placebo headlines — which were made-up claims that neither true nor fake news outlets had reported. The parcipants were then asked to report if they had seen it “reported or discussed” during the 2016 U.S. campaign, and how likely they would have been to believe it. False claims included a rumor that the Pope had endorsed Trump for presidency, and that the FBI found evidence of the Clinton Foundation running a pedophile ring. True ones included Trump refusing to confirm that he would concede the election if he lost, as well as Clinton’s comment about Trump supporters belonging in a “basket of deplorables.” Placebos included claims that either Trump or Clinton campaign staff had diverted funds to buy alcohol for expensive parties. Since the authors have not clearly explained the differences amongst these therms, especially between “BigFake” and “FakeFake”, it seems challenging to differentiate them. According to the authors, there were 30 questions ready to be presented to the respondents, but each respondent would only receive a randomly selected 15 of these stories.

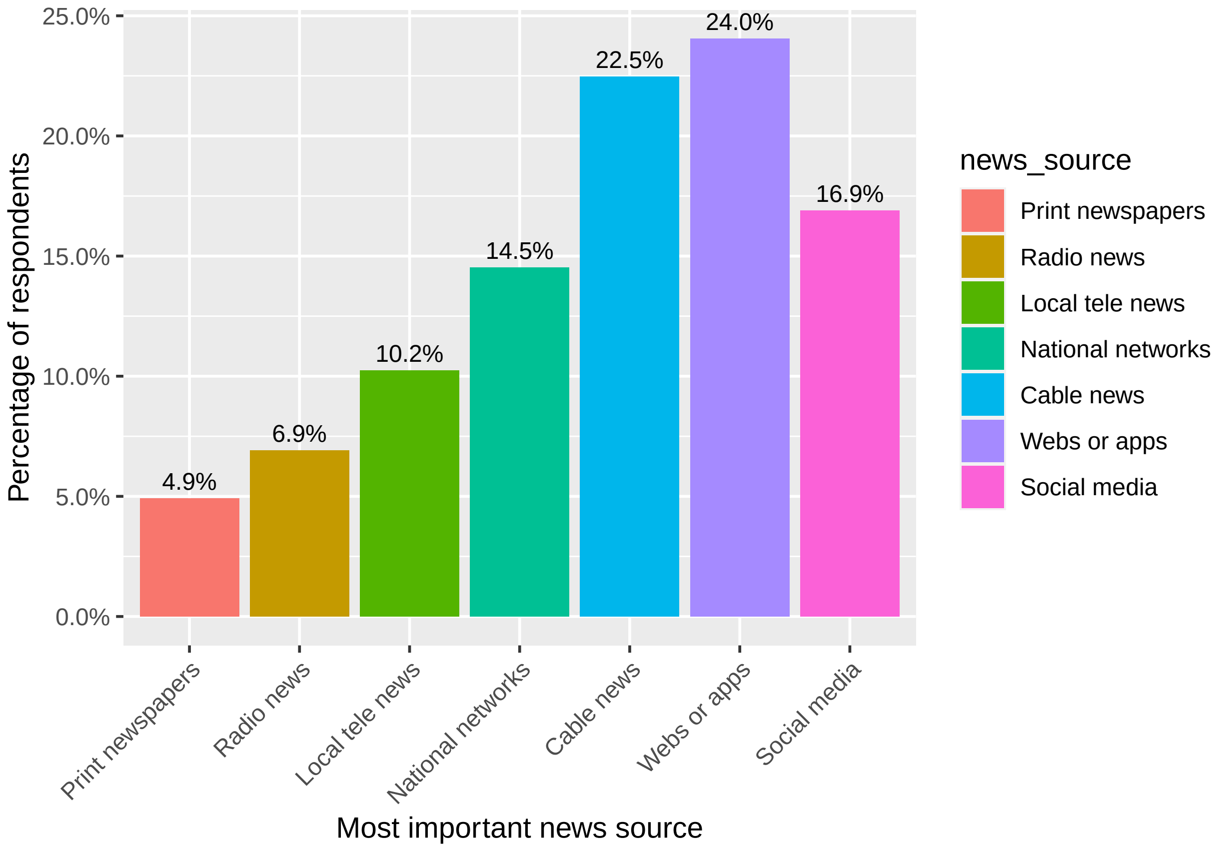
Most important news source to get politically informed

As social media has permeated every aspect of our lives, people have started to heavily rely social media to obtain information. This study tests whether social media has become an important information source

Table 4: Average weights table

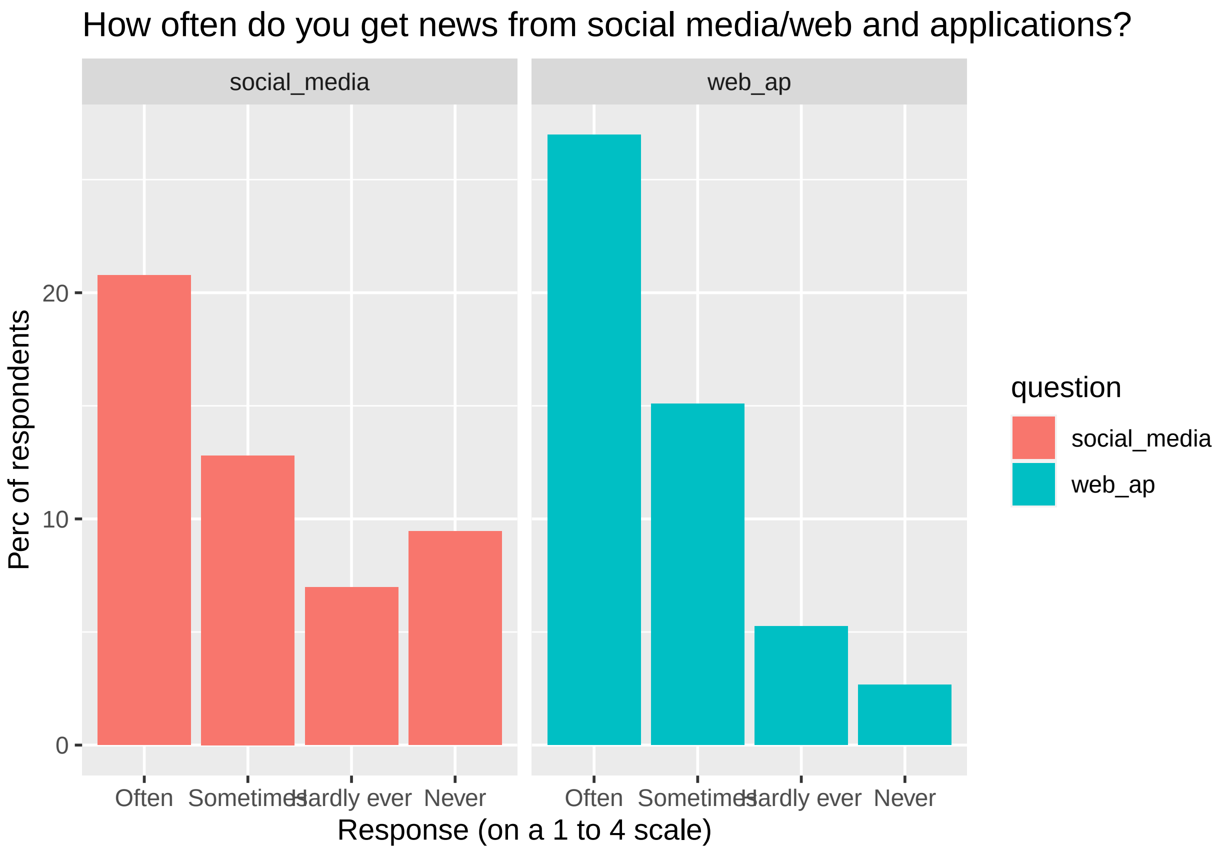
Type	Article Text	Article True/False	Pro-Clinton or pro-Trump?
BigFake	Pope Francis endorsed Donald Trump.	FALSE	Trump
BigFake	An FBI agent connected to Hillary ClintonÍs email disclosures murdered his wife and shot himself.	FALSE	Trump
BigFake	The Clinton Foundation bought $137 million in illegal arms.	FALSE	Trump
BigFake	Mike Pence said that "Michelle Obama is the most vulgar First Lady we've ever had."	FALSE	Clinton
BigFake	In May 2016, Ireland announced that it was officially accepting Americans requesting political asylum from a Donald Trump presidency.	FALSE	Clinton
BigFake	Celebrity RuPaul said that Donald Trump mistook him for a woman and groped him at a party in 1995.	FALSE	Clinton
SmallFake	At the beginning of November, the FBI uncovered evidence of a pedophile sex ring run under the guise of the Clinton Foundation.	FALSE	Trump
SmallFake	Under Donald Trump's tax plan, it is projected that 51% of single parents would see their taxes go up.	TRUE	Clinton
SmallFake	At a rally a few days before the election, President Obama screamed at a protester who supported Donald Trump.	FALSE	Trump
SmallFake	FBI Director James Comey's October 28th letter about new developments in the investigation of Hillary Clinton's emails went only to Republican members of Congress, and not to Democrats.	FALSE	Clinton
SmallFake	A Republican congressman helped broker a deal for Donald Trump to buy a taxpayer-owned building in order to	FALSE	Clinton

from which American adults to get news. In this survey, respondents were asked to report what the most important news sources are for them. Surprisingly, the results suggest show that only 17 percent of the respondents clearly indicated that they used social media as their primary information source. This finding is consistent with thee authors findings. In 2016, people still relied on other communication channels to get informed, especially news websites and cable television.



How often do people use social media to get news?

Therefore, people do not use social media as the primary source to get political information. However, it does not mean that social media does not play any role in information consumption. To the contrary, survey results suggest that respondents heavily rely on social media ton follow political news. The results show that people heavily rely on social media to get informed. More than 20 percent of the respondents indicated that they often use social media to get news and more than 25 percent of the respondents often turned to websites and apps to get informed.



Who believed in political fake news?

One of the goals of this study was to test if the share of respondents that was truly exposed and believed the average fake news article. STo exam what contributed to the respondents' belivability of such political fake news, the authors established a simple linear model as follows to characterize the potential relationship, In this model, B_{ij} takes 1 when people believe, takes 0 when people do not believe in the story; and takes 0.5 when people are not sure about the veracity of the story.For example, if headline a is true, then B takes value 1 if person i responded "Yes" to "would your best guess have been that this statement was true?". B_{ij} is the the dependent variable and a vector Xi in a linear regression.

Model

In this study, the authors established a model to predict who would believe in fake news. The model is specified as follows,

$$y_{i,j} = \alpha + \beta x_n + \epsilon_n$$

where $\epsilon \sim normal(0, \sigma)$.

The model is therefore equivalent to the following sampling involving the residual,

$y_n - (\alpha + \beta X_n) \sim normal(0, \sigma)$

and reducing still further, to $y_n \sim normal(\alpha + \beta X_n, \sigma)$

This model is coded in Stan and R programming languages. The model includes N observations, each with predictor x_n and outcome y_n. The intercept and slope parameters are alpha and beta. The model assumes a normally distributed noise term with scale sigma. This model has improper priors for the two regression coefficients.