Red Flagging Fake News

Suhas Gupta, Kevin Drever, Imran Manji

```r
# load packages
library(data.table)
library(foreign)
library(sandwich)
library(lmtest)
library(stargazer)
library(lfe)
library(car)
library(ggplot2)
library(data.table)
library(knitr)
library(rgeolocate)
library(data.table)
library(knitr)
library(lmtest)
library(ri2)
library(dplyr)
library(forcats)
```

Null & Alternate Hypothesis

- *NULL Hypothesis* : **Make people aware of the prevalence of fake news has no effect on its believabiliy**
- *Alternate Hypothesis* : **General flags about fake news reduce its believability**

Calculating the sample size

In this section, we calculate the minimum required sample size for our experiment.

The statistical power of an experiment is the experiment's abilitiy to reject the NULL hypothesis when a specific alternate hypothesis is true.

$$\alpha = P(\text{reject } H_0|H_0)$$

where α is the significance level. We select a significance level of $\alpha = 0.05$ as a tolerance for Type I errors in our experiment.

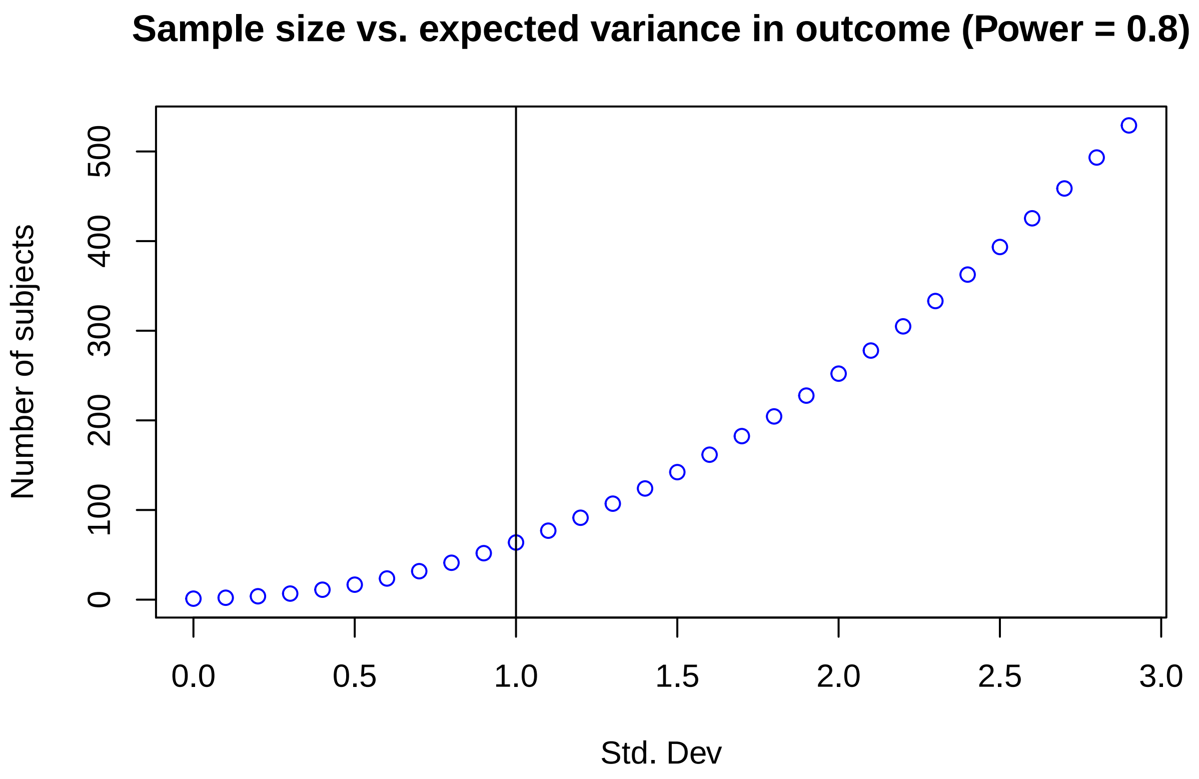
Now that we have chosen our significance level, we would like to minimize the probability of Type II error. i.e. we would like to **maximize** the power of our test against the relevan alternative. Mathematically, power is

$$power = 1 - P_r(\text{Type II Error}) = P_r(\text{reject } H_0|H_1\text{is true})$$

- We would set the required power of our experiment to be **80%** for this study as a reasonable expectation.

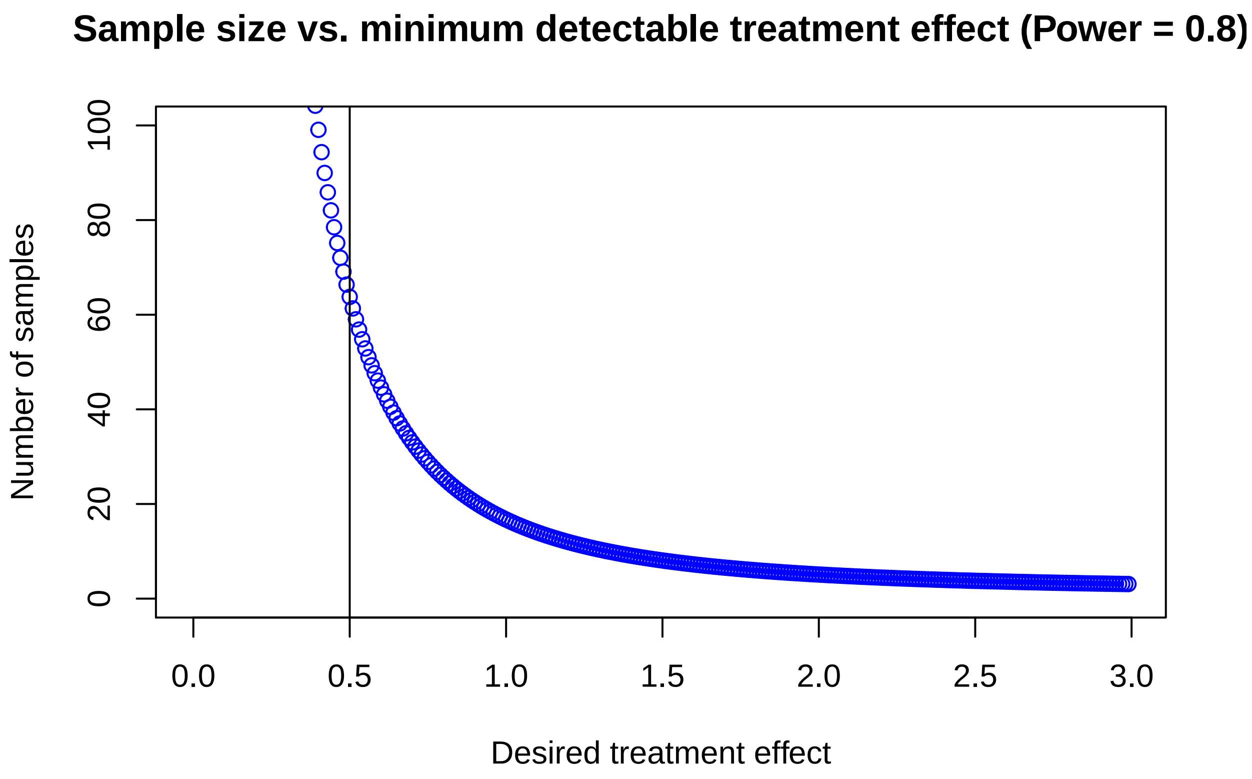
- To calculate the power for the test, we need to conjecture an expected ATE and the standard deviation for the outcome in the experiment.
- The outcome is a rating on a scale of 0-10 on how successfull the red flag was in reducing the believability of the fake/misleading social media post. We would like our experiment to be able detect a difference in means of minimum 2 points on this scale.
- We do expect the measured values for this rating to vary significantly as we poll subjects with different political opinions, life experiences and political affiliations. To be on the conservative side, we would like to have enough power in our experiment to minimize Type II errors when the std. deviation is at least 2.5 times the minimum detectable treatment effect.

```
power_sim <- function(ate,sig_level=0.05,power=0.8,alternate_hyp="two.sided", sd = 1){
    result <- NA
    sims <- seq(1e-5,sd,by=0.1)
    for(i in seq_along(sims)){
        result[i] <- power.t.test(d=ate,
                               sig.level=sig_level,
                               power=power,
                               sd=sims[i],
                               alternative=alternate_hyp)$n
        }
    return(result)
    }
sd <- 3
expected_ate <- 0.5
x <- seq(1e-5,sd, by = 0.1)
samples <- power_sim(ate=expected_ate,sd = sd)
plot(x = x, y=samples,col = 'blue',
     xlab="Std. Dev",
     ylab = 'Number of subjects',
     main = "Sample size vs. expected variance in outcome (Power = 0.8)")
abline(v=1.0,col='black',lwd=1)
```



The above plot shows that we need a minimum sample size of 100 to achieve a power of 0.8 when the outcome variable has a standard deviation 1.0 times the treatment effect. The plot below, validates that the absolute value of minimum treatment effect doesn't change the sample size requirement significantly and that this is determined mostly by the expected variance in the measurement data.

```
power_sim_by_ate <- function(ate_vector,sig_level=0.05,power=0.8,alternate_hyp="two.sided",sd = 1){
    result <- NA
    for(i in seq_along(ate_vector)){
        result[i] <- power.t.test(d=ate_vector[i],
                                  sig.level=sig_level,
                                  power=power,
                                  sd=sd,
                                  alternative=alternate_hyp)$n
        }
    return(result)
    }
sd <- 1
expected_ate <- 3
x <- seq(1e-5,expected_ate,by=0.01)
samples <- power_sim_by_ate(ate=x,sd = sd)

plot(x = x, y=samples,col = 'blue',
     xlab= "Desired treatment effect",
     ylab = 'Number of samples',ylim=c(0,100),
     main = "Sample size vs. minimum detectable treatment effect (Power = 0.8)")
abline(v=0.5,col='black',lwd=1)
```



Covariate questions in the survey

- Age
- Political affiliation
- Registered Voter / non-voter
- race
- are you active on social media?

- education (< high school, high school , undergrad, grad)

Covariates in regression (not in survey)

- Mturk subject
- location of subject

Experimental Design

2 x 2

- treatment :
  - banner or no banner
  - tweet is false or true
- block by party affiliation and gender

Treatment & control assginment

- how to randomly assign while blocking for above

Regression Models

Outcome: Score on how many headlines were correctly identified by subjects (equally balanced True and False posts)

1. Baseline model

outcome ~ general_flag on survey page

2. Model with co-variates

outcome ~ red_flag * gender + red_flag * political_affiliation + factor(age_group) + factor(education) + red_flag * location + registered_voter + race + social_media_active

3. Model with treatment-covariate interactions

   - Test if fake news red flagging affects democrats and republicans differently
   - Test if fake news red flagging affects different age groups differently
   - Test if fake news red flagging affects voters and non voters differently

Pilot data analysis

```r
pilot_dataset <- fread('./data/pilot/pilot_data_07262020.csv')
head(pilot_dataset)
```

```
##                                                        StartDate
## 1:                                                    Start Date
## 2: {""ImportId"":""startDate"",""timeZone"":""America/Denver""}
## 3:                                           2020-07-17 18:52:45
## 4:                                           2020-07-17 19:46:06
## 5:                                           2020-07-17 19:48:21
## 6:                                           2020-07-17 20:01:07
##                                                      EndDate
## 1:                                                  End Date
## 2: {""ImportId"":""endDate"",""timeZone"":""America/Denver""}
## 3:                                         2020-07-17 18:59:24
## 4:                                         2020-07-17 19:48:33
## 5:                                         2020-07-17 19:59:28
## 6:                                         2020-07-17 20:04:12
##                        Status                    IPAddress
## 1:              Response Type                   IP Address
## 2: {""ImportId"":""status""} {""ImportId"":""ipAddress""}
## 3:                 IP Address                 73.93.90.157
## 4:                 IP Address                98.234.117.52
## 5:                 IP Address               71.244.172.196
## 6:                 IP Address                 173.67.9.152
##                        Progress       Duration (in seconds)
## 1:                     Progress       Duration (in seconds)
## 2: {""ImportId"":""progress""} {""ImportId"":""duration""}
## 3:                          100                         399
## 4:                          100                         147
## 5:                          100                         666
## 6:                          100                         185
##                        Finished
## 1:                     Finished
## 2: {""ImportId"":""finished""}
## 3:                         True
## 4:                        True
## 5:                        True
## 6:                        True
##                                                        RecordedDate
## 1:                                                    Recorded Date
## 2: {""ImportId"":""recordedDate"",""timeZone"":""America/Denver""}
## 3:                                              2020-07-17 18:59:25
## 4:                                              2020-07-17 19:48:34
## 5:                                              2020-07-17 19:59:28
## 6:                                              2020-07-17 20:04:13
##                       ResponseId                    RecipientLastName
## 1:                   Response ID                  Recipient Last Name
## 2: {""ImportId"":""_recordId""} {""ImportId"":""recipientLastName""}
## 3:             R_u8Geu0CykTxNh6h
## 4:             R_2EgXOU7K2nTlgqG
## 5:             R_1NgJqwlFgUpLjT1
## 6:             R_33woiiDhvnhBZZR
##                       RecipientFirstName                     RecipientEmail
## 1:                  Recipient First Name                    Recipient Email
## 2: {""ImportId"":""recipientFirstName""} {""ImportId"":""recipientEmail""}
## 3:
## 4:
## 5:
## 6:
##                           ExternalReference                     LocationLatitude
## 1:                  External Data Reference                    Location Latitude
## 2: {""ImportId"":""externalDataReference""} {""ImportId"":""locationLatitude""}
## 3:                                                               37.777099609375
## 4:                                                              36.5802001953125
## 5:                                                           39.1269073486328125
## 6:                                                           39.1269073486328125
##                       LocationLongitude                    DistributionChannel
## 1:                   Location Longitude                   Distribution Channel
## 2: {""ImportId"":""locationLongitude""} {""ImportId"":""distributionChannel""}
## 3:               -122.40599822998046875                              anonymous
## 4:                  -121.84429931640625                              anonymous
## 5:                     -76.697998046875                              anonymous
## 6:                     -76.697998046875                              anonymous
##                       UserLanguage
## 1:                   User Language
## 2: {""ImportId"":""userLanguage""}
## 3:                              EN
## 4:                              EN
## 5:                              EN
## 6:                              EN
##                                            Q_RecaptchaScore
## 1:                                         Q_RecaptchaScore
## 2:                    {""ImportId"":""Q_RecaptchaScore""}
## 3: 0.90000000000000002220446049250313080847263336181640625
## 4: 0.90000000000000002220446049250313080847263336181640625
## 5: 0.90000000000000002220446049250313080847263336181640625
## 6: 0.90000000000000002220446049250313080847263336181640625
##
## 1: Thank you for participating in this survey - we greatly appreciate it! The survey should take les
## 2:
## 3:
## 4:
## 5:
## 6:
##                              Q2                      Q3
## 1: Are you a registered voter? What is your age group?
## 2:     {""ImportId"":""QID8""} {""ImportId"":""QID9""}
## 3:                         Yes                   21-40
## 4:                         Yes                   21-40
## 5:                         Yes                   21-40
## 6:                         Yes                   21-40
##                                       Q4
## 1: What is your politicial affiliation?
## 2:             {""ImportId"":""QID10""}
## 3:                             Democrat
## 4:                                Other
## 5:
## 6:                             Democrat
##                                            Q5                                 Q6
## 1: What is your highest level of education? What ethnicity do you identify as?
## 2:                  {""ImportId"":""QID19""}           {""ImportId"":""QID20""}
## 3:                           Graduate degree                              Asian
## 4:                           Graduate degree                          Caucasian
## 5:                            College degree                          Caucasian
## 6:                            College degree                  Hispanic / Latinx
##                                                                    Q7
## 1: Do you consider yourself to be active on social media platforms?
## 2:                                          {""ImportId"":""QID18""}
## 3:                                                               Yes
## 4:                                                                No
## 5:                                                                No
## 6:                                                                No
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise
## 2:
## 3:
## 4:
## 5:
## 6:
```

```
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise y
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise y
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise y
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise y
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise y
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise y
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise y
## 2:
## 3:
## 4:
## 5:
## 6:
##                                                          8A
## 1: Do you believe that the content of the above post is true?
## 2:                                {""ImportId"":""QID32""}
## 3:                                                       Yes
## 4:
```

```
## 5:                                                           Yes
## 6:
##                                                               9A
## 1: Do you believe that the content of the above post is true?
## 2:                                    {""ImportId"":""QID33""}
## 3:                                                           Yes
## 4:
## 5:                                                           Yes
## 6:
##                                                              10A
## 1: Do you believe that the content of the above post is true?
## 2:                                    {""ImportId"":""QID34""}
## 3:                                                            No
## 4:
## 5:                                                            No
## 6:
##                                                              11A
## 1: Do you believe that the content of the above post is true?
## 2:                                    {""ImportId"":""QID35""}
## 3:                                                           Yes
## 4:
## 5:                                                           Yes
## 6:
##                                                              12A
## 1: Do you believe that the content of the above post is true?
## 2:                                    {""ImportId"":""QID36""}
## 3:                                                           Yes
## 4:
## 5:                                                            No
## 6:
##                                                              13A
## 1: Do you believe that the content of the above post is true?
## 2:                                    {""ImportId"":""QID37""}
## 3:                                                            No
## 4:
## 5:                                                            No
## 6:
##                                                              14A
## 1: Do you believe that the content of the above post is true?
## 2:                                    {""ImportId"":""QID38""}
## 3:                                                           Yes
## 4:
## 5:                                                            No
## 6:
##                                                              15A
## 1: Do you believe that the content of the above post is true?
## 2:                                    {""ImportId"":""QID39""}
## 3:                                                           Yes
## 4:
## 5:                                                           Yes
## 6:
##                                                              16A
## 1: Do you believe that the content of the above post is true?
## 2:                                    {""ImportId"":""QID40""}
## 3:                                                         No
## 4:
## 5:                                                         No
## 6:
##                                                           17A
## 1: Do you believe that the content of the above post is true?
## 2:                                   {""ImportId"":""QID41""}
## 3:                                                         No
## 4:
## 5:                                                         No
## 6:
```

```r
names(pilot_dataset)
```

```
##  [1] "StartDate"             "EndDate"               "Status"
##  [4] "IPAddress"             "Progress"              "Duration (in seconds)"
##  [7] "Finished"              "RecordedDate"          "ResponseId"
## [10] "RecipientLastName"     "RecipientFirstName"    "RecipientEmail"
## [13] "ExternalReference"     "LocationLatitude"      "LocationLongitude"
## [16] "DistributionChannel"   "UserLanguage"          "Q_RecaptchaScore"
## [19] "Q1"                    "Q2"                    "Q3"
## [22] "Q4"                    "Q5"                    "Q6"
## [25] "Q7"                    "8B"                    "9B"
## [28] "10B"                   "11B"                   "12B"
## [31] "13B"                   "14B"                   "15B"
## [34] "16B"                   "17B"                   "8A"
## [37] "9A"                    "10A"                   "11A"
## [40] "12A"                   "13A"                   "14A"
## [43] "15A"                   "16A"                   "17A"
```

```r
data_pruned <- pilot_dataset[ 3:nrow(pilot_dataset),]
data_pruned[, c(6,7)] <- lapply(data_pruned[, c(6,7)], as.numeric)
```

```
## Warning in lapply(data_pruned[, c(6, 7)], as.numeric): NAs introduced by
## coercion
```

```r
question_col_names <- c('8B','9B','10B','11B','12B','13B','14B','15B','16B','17B',
                        '8A','9A','10A','11A','12A','13A','14A','15A','16A','17A')


# replace all empty strings with NA
for(i in c(26:length(names(data_pruned)))){
    data_pruned[[i]][data_pruned[[i]]==''] <- NA
}

# Check the data
head(data_pruned[, 26:length(names(data_pruned))])
```

```
##      8B    9B   10B   11B   12B   13B   14B   15B   16B   17B    8A    9A   10A   11A   12A
## 1: <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>  Yes  Yes   No  Yes  Yes
## 2:  Yes  Yes   No  Yes  Yes   No  Yes   No   No   No <NA> <NA> <NA> <NA> <NA>
## 3: <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>  Yes  Yes   No  Yes   No
## 4:   No  Yes   No  Yes   No   No  Yes  Yes   No   No <NA> <NA> <NA> <NA> <NA>
## 5: <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>  Yes  Yes   No  Yes   No
## 6: <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>   No  Yes   No  Yes   No
##     13A  14A  15A  16A  17A
## 1:   No  Yes  Yes   No   No
## 2: <NA> <NA> <NA> <NA> <NA>
## 3:   No   No  Yes   No   No
## 4: <NA> <NA> <NA> <NA> <NA>
## 5:   No  Yes  Yes   No   No
## 6:   No   No   No   No   No
```

```r
# Set assignment group variable (treatment = 1 , conrol = 0)
data_pruned[, assignment := ifelse(is.na(data_pruned[,'8B']),0,1)]
```

```
## Warning in `[.data.table`(data_pruned, , `:=`(assignment,
## ifelse(is.na(data_pruned[, : Invalid .internal.selfref detected and fixed by
## taking a (shallow) copy of the data.table so that := can add this new column
## by reference. At an earlier point, this data.table has been copied by R (or
## was created manually using structure() or similar). Avoid names<- and attr<-
## which in R currently (and oddly) may copy the whole data.table. Use set* syntax
## instead to avoid copying: ?set, ?setnames and ?setattr. If this message doesn't
## help, please report your use case to the data.table issue tracker so the root
## cause can be fixed or this message improved.
```

```r
head(data_pruned[, 26:length(names(data_pruned))])
```

```
##      8B   9B  10B  11B  12B  13B  14B  15B  16B  17B   8A   9A  10A  11A  12A
## 1: <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>  Yes  Yes   No  Yes  Yes
## 2:  Yes  Yes   No  Yes  Yes   No  Yes   No   No   No <NA> <NA> <NA> <NA> <NA>
## 3: <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>  Yes  Yes   No  Yes   No
## 4:   No  Yes   No  Yes   No   No  Yes  Yes   No   No <NA> <NA> <NA> <NA> <NA>
## 5: <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>  Yes  Yes   No  Yes   No
## 6: <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>   No  Yes   No  Yes   No
##     13A  14A  15A  16A  17A assignment
## 1:   No  Yes  Yes   No   No          0
## 2: <NA> <NA> <NA> <NA> <NA>          1
## 3:   No   No  Yes   No   No          0
## 4: <NA> <NA> <NA> <NA> <NA>          1
## 5:   No  Yes  Yes   No   No          0
## 6:   No   No   No   No   No          0
```

```r
head(data_pruned)
```

```
##               StartDate             EndDate     Status      IPAddress Progress
## 1: 2020-07-17 18:52:45 2020-07-17 18:59:24 IP Address   73.93.90.157      100
## 2: 2020-07-17 19:46:06 2020-07-17 19:48:33 IP Address  98.234.117.52      100
## 3: 2020-07-17 19:48:21 2020-07-17 19:59:28 IP Address 71.244.172.196      100
## 4: 2020-07-17 20:01:07 2020-07-17 20:04:12 IP Address   173.67.9.152      100
## 5: 2020-07-17 19:58:48 2020-07-17 20:13:25 IP Address 174.195.207.41      100
## 6: 2020-07-17 20:08:21 2020-07-17 20:17:59 IP Address  67.188.128.89      100
##    Duration (in seconds) Finished        RecordedDate        ResponseId
## 1:                   399       NA 2020-07-17 18:59:25 R_u8Geu0CykTxNh6h
## 2:                   147       NA 2020-07-17 19:48:34 R_2EgX0U7K2nTlgqG
## 3:                   666       NA 2020-07-17 19:59:28 R_1NgJqwlFgUpLjT1
## 4:                   185       NA 2020-07-17 20:04:13 R_33woiiDhvnhBZZR
## 5:                   876       NA 2020-07-17 20:13:25 R_3nAiDFypCL0OxYV
## 6:                   578       NA 2020-07-17 20:18:00 R_3QXbtPAqPrxpNRD
##    RecipientLastName RecipientFirstName RecipientEmail ExternalReference
## 1:
## 2:
```

```
## 3:
## 4:
## 5:
## 6:
##       LocationLatitude      LocationLongitude DistributionChannel UserLanguage
## 1:     37.777099609375 -122.40599822998046875           anonymous           EN
## 2:    36.5802001953125    -121.84429931640625           anonymous           EN
## 3: 39.1269073486328125       -76.697998046875           anonymous           EN
## 4: 39.1269073486328125       -76.697998046875           anonymous           EN
## 5:         33.87890625 -117.53530120849609375           anonymous           EN
## 6: 36.6808013916015625 -121.61640167236328125           anonymous           EN
##                                            Q_RecaptchaScore     Q1  Q2    Q3
## 1: 0.90000000000000002220446049250313080847263336181640625 Female Yes 21-40
## 2: 0.90000000000000002220446049250313080847263336181640625 Female Yes 21-40
## 3: 0.90000000000000002220446049250313080847263336181640625 Female Yes 21-40
## 4: 0.90000000000000002220446049250313080847263336181640625 Female Yes 21-40
## 5: 0.90000000000000002220446049250313080847263336181640625 Female Yes 21-40
## 6: 0.90000000000000002220446049250313080847263336181640625   Male Yes   61+
##             Q4              Q5                Q6  Q7   8B   9B  10B  11B  12B
## 1:   Democrat Graduate degree             Asian Yes <NA> <NA> <NA> <NA> <NA>
## 2:      Other Graduate degree         Caucasian  No  Yes  Yes   No  Yes  Yes
## 3:             College degree         Caucasian  No <NA> <NA> <NA> <NA> <NA>
## 4:   Democrat  College degree Hispanic / Latinx  No   No  Yes   No  Yes   No
## 5:      Other  College degree         Caucasian Yes <NA> <NA> <NA> <NA> <NA>
## 6: Republican Graduate degree         Caucasian  No <NA> <NA> <NA> <NA> <NA>
##     13B  14B  15B  16B  17B   8A   9A  10A  11A  12A  13A  14A  15A  16A  17A
## 1: <NA> <NA> <NA> <NA> <NA>  Yes  Yes   No  Yes  Yes   No  Yes  Yes   No   No
## 2:   No  Yes   No   No   No <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 3: <NA> <NA> <NA> <NA> <NA>  Yes  Yes   No  Yes   No   No   No  Yes   No   No
## 4:   No  Yes  Yes   No   No <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 5: <NA> <NA> <NA> <NA> <NA>  Yes  Yes   No  Yes   No   No  Yes  Yes   No   No
## 6: <NA> <NA> <NA> <NA> <NA>   No  Yes   No  Yes   No   No   No   No   No   No
##    assignment
## 1:          0
## 2:          1
## 3:          0
## 4:          1
## 5:          0
## 6:          0
```

```r
# Compute the score against answer guide
answer_guide <- c('Yes','Yes','No','Yes','No','No','Yes','Yes','No','No',
                  'Yes','Yes','No','Yes','No','No','Yes','Yes','No','No')

compute_scores <- function(dataset,answer_guide){
    for(i in 1:nrow(data_pruned)){
    dataset[i,"score"] <- sum(dataset[i,26:45] == answer_guide,na.rm=TRUE)
    }
    return(dataset)
}


data_w_scores <- compute_scores(data_pruned,answer_guide)
head(data_w_scores)
```

```
##            StartDate             EndDate     Status      IPAddress Progress
## 1: 2020-07-17 18:52:45 2020-07-17 18:59:24 IP Address   73.93.90.157      100
## 2: 2020-07-17 19:46:06 2020-07-17 19:48:33 IP Address  98.234.117.52      100
## 3: 2020-07-17 19:48:21 2020-07-17 19:59:28 IP Address 71.244.172.196      100
## 4: 2020-07-17 20:01:07 2020-07-17 20:04:12 IP Address   173.67.9.152      100
## 5: 2020-07-17 19:58:48 2020-07-17 20:13:25 IP Address 174.195.207.41      100
## 6: 2020-07-17 20:08:21 2020-07-17 20:17:59 IP Address  67.188.128.89      100
##    Duration (in seconds) Finished        RecordedDate        ResponseId
## 1:                   399       NA 2020-07-17 18:59:25 R_u8Geu0CykTxNh6h
## 2:                   147       NA 2020-07-17 19:48:34 R_2EgXOU7K2nTlgqG
## 3:                   666       NA 2020-07-17 19:59:28 R_1NgJqwlFgUpLjT1
## 4:                   185       NA 2020-07-17 20:04:13 R_33woiiDhvnhBZZR
## 5:                   876       NA 2020-07-17 20:13:25 R_3nAiDFypCLOOxYV
## 6:                   578       NA 2020-07-17 20:18:00 R_3QXbtPAqPrxpNRD
##    RecipientLastName RecipientFirstName RecipientEmail ExternalReference
## 1:
## 2:
## 3:
## 4:
## 5:
## 6:
##       LocationLatitude      LocationLongitude DistributionChannel UserLanguage
## 1:     37.777099609375 -122.40599822998046875           anonymous           EN
## 2:    36.5802001953125    -121.84429931640625           anonymous           EN
## 3: 39.1269073486328125       -76.697998046875           anonymous           EN
## 4: 39.1269073486328125       -76.697998046875           anonymous           EN
## 5:         33.87890625 -117.53530120849609375           anonymous           EN
## 6: 36.6808013916015625 -121.61640167236328125           anonymous           EN
##                                             Q_RecaptchaScore     Q1  Q2    Q3
## 1: 0.90000000000000002220446049250313080847263336181640625 Female Yes 21-40
## 2: 0.90000000000000002220446049250313080847263336181640625 Female Yes 21-40
## 3: 0.90000000000000002220446049250313080847263336181640625 Female Yes 21-40
## 4: 0.90000000000000002220446049250313080847263336181640625 Female Yes 21-40
## 5: 0.90000000000000002220446049250313080847263336181640625 Female Yes 21-40
## 6: 0.90000000000000002220446049250313080847263336181640625   Male Yes   61+
##            Q4              Q5                Q6  Q7   8B   9B  10B  11B  12B
## 1:   Democrat Graduate degree             Asian Yes <NA> <NA> <NA> <NA> <NA>
## 2:      Other Graduate degree         Caucasian  No  Yes  Yes   No  Yes  Yes
## 3:             College degree         Caucasian  No <NA> <NA> <NA> <NA> <NA>
## 4:   Democrat  College degree Hispanic / Latinx  No   No  Yes   No  Yes   No
## 5:      Other  College degree         Caucasian Yes <NA> <NA> <NA> <NA> <NA>
## 6: Republican Graduate degree         Caucasian  No <NA> <NA> <NA> <NA> <NA>
##     13B  14B  15B  16B  17B   8A   9A  10A  11A  12A  13A  14A  15A  16A  17A
## 1: <NA> <NA> <NA> <NA> <NA>  Yes  Yes   No  Yes  Yes   No  Yes  Yes   No   No
## 2:   No  Yes   No   No   No <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 3: <NA> <NA> <NA> <NA> <NA>  Yes  Yes   No  Yes   No   No   No  Yes   No   No
## 4:   No  Yes  Yes   No   No <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 5: <NA> <NA> <NA> <NA> <NA>  Yes  Yes   No  Yes   No   No  Yes  Yes   No   No
## 6: <NA> <NA> <NA> <NA> <NA>   No  Yes   No  Yes   No   No   No   No   No   No
##    assignment score
## 1:          0     9
## 2:          1     8
## 3:          0     9
## 4:          1     9
## 5:          0    10
## 6:          0     7
```

```r
# Compute the SD and point estimates with pilot data
sd_pilot <- data_w_scores[, sd(score)]
sd_pilot
```

```
## [1] 1.3434
```

```r
d <- data_w_scores[, .(scores=mean(score)), by = assignment]
mod <- lm(score ~ assignment, data_w_scores)
ate <- diff(d$scores)
ate
```

```
## [1] -0.27381
```

```r
stargazer(mod, type="text")
```

```
## 
## ===============================================
##                         Dependent variable:    
##                     ---------------------------
##                                score           
## -----------------------------------------------
## assignment                     -0.274          
##                                (0.536)         
##                                                
## Constant                      7.857***         
##                                (0.364)         
##                                                
## -----------------------------------------------
## Observations                     26            
## R2                              0.011          
## Adjusted R2                     -0.030         
## Residual Std. Error        1.364 (df = 24)     
## F Statistic             0.261 (df = 1; 24)     
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

```r
## Power calculation
power.t.test(d=ate,sig.level=0.95,power=0.8,sd=sd,alternative="two.sided")
```

```
## 
##      Two-sample t test power calculation 
## 
##               n = 21.818
##           delta = 0.27381
##              sd = 1
##       sig.level = 0.95
##           power = 0.8
##     alternative = two.sided
## 
## NOTE: n is number in *each* group
```

```r
# Modify the column names for better readability
data_mod <- rename(data_w_scores,
       Gender = Q1,
       Reg_Voter = Q2,
       Age_bin = Q3,
       Party = Q4,
       Education = Q5,
       Ethnicity = Q6,
       Soc_Med_Active = Q7
        )
head(data_mod[,19:27])
```

```
##    Gender Reg_Voter Age_bin      Party       Education         Ethnicity
## 1: Female       Yes   21-40   Democrat Graduate degree             Asian
## 2: Female       Yes   21-40      Other Graduate degree         Caucasian
## 3: Female       Yes   21-40             College degree         Caucasian
## 4: Female       Yes   21-40   Democrat  College degree Hispanic / Latinx
## 5: Female       Yes   21-40      Other  College degree         Caucasian
## 6:   Male       Yes     61+ Republican Graduate degree         Caucasian
##    Soc_Med_Active   8B   9B
## 1:            Yes <NA> <NA>
## 2:             No  Yes  Yes
## 3:             No <NA> <NA>
## 4:             No   No  Yes
## 5:            Yes <NA> <NA>
## 6:             No <NA> <NA>
```

MTurk data

Mturk survey 1 was done with a higher reward and no check for BOTs. The survey subject count was 100 and responses were obtained within 24 hours due to the high reward (5-8 min task paid $1.5).

```
mturk1_dataset <- fread('./data/Mturk_1/Mturk_1_data.csv')
head(mturk1_dataset)
```

```
##        StartDate       EndDate        Status       IPAddress Progress
## 1:    Start Date      End Date Response Type      IP Address Progress
## 2: 7/24/20 22:43 7/24/20 22:45    IP Address    99.75.53.174      100
## 3: 7/24/20 22:43 7/24/20 22:45    IP Address   68.33.126.140      100
## 4: 7/24/20 22:43 7/24/20 22:45    IP Address   98.19.217.229      100
## 5: 7/24/20 22:43 7/24/20 22:45    IP Address  174.85.199.139      100
## 6: 7/24/20 22:43 7/24/20 22:45    IP Address 209.159.199.248      100
##    Duration (in seconds) Finished  RecordedDate        ResponseId
## 1: Duration (in seconds) Finished Recorded Date       Response ID
## 2:                   124     True 7/24/20 22:45 R_3wSF9NPrJQkHDjj
## 3:                   112     True 7/24/20 22:45 R_2wHbTKc7249gZQY
## 4:                    99     True 7/24/20 22:45 R_2Et95GjQbJ9BgaR
## 5:                   142     True 7/24/20 22:45 R_3LimuwbiSdyN053
## 6:                   152     True 7/24/20 22:45 R_1rwQr9otPszxd5D
##      RecipientLastName   RecipientFirstName  RecipientEmail
## 1: Recipient Last Name Recipient First Name Recipient Email
## 2:
## 3:
## 4:
## 5:
## 6:
##          ExternalReference  LocationLatitude  LocationLongitude
## 1: External Data Reference Location Latitude Location Longitude
## 2:                               42.02209473       -88.17050171
## 3:                               38.86700439       -76.81729889
## 4:                               34.45120239       -84.15299988
## 5:                               34.34539795       -86.27400208
## 6:                               44.14149475       -103.2052002
##     DistributionChannel  UserLanguage Q_RecaptchaScore
## 1: Distribution Channel User Language Q_RecaptchaScore
## 2:            anonymous            EN              0.9
## 3:            anonymous            EN              0.9
## 4:            anonymous            EN              0.7
## 5:            anonymous            EN              0.9
## 6:            anonymous            EN              0.9
##
## 1: Thank you for participating in this survey - we greatly appreciate it! The survey should take les
## 2:
## 3:
## 4:
## 5:
## 6:
##                             Q2                                Q38
## 1: Are you a registered voter? Did you vote in the 2012 election?
## 2:                          No                                 No
## 3:                         Yes                                Yes
## 4:                         Yes                                Yes
## 5:                         Yes                                Yes
## 6:                         Yes                                Yes
##                                   Q39                           Q37
## 1: Did you vote in the 2016 election? What is your marital status ?
## 2:                                 No                        Single
## 3:                                Yes                       Married
## 4:                                Yes                       Married
## 5:                                Yes                       Married
## 6:                                Yes                        Single
##                                                Q40
## 1: What is your primary language of communication?
## 2:                                         English
## 3:                                         English
## 4:                                         English
## 5:                                         English
## 6:                                         English
##                                      Q36                      Q3
## 1: What is your annual household income? What is your age group?
## 2:                              < $60000                   21-40
## 3:                     $150000 - $250000                   21-40
## 4:                      $60000 - $150000                   21-40
## 5:                              < $60000                   21-40
## 6:                              < $60000                   21-40
##                                      Q4
## 1: What is your politicial affiliation?
## 2:                                Other
## 3:                             Democrat
## 4:                           Republican
## 5:                                Other
## 6:                                Other
##                                          Q5                                 Q6
## 1: What is your highest level of education? What ethnicity do you identify as?
## 2:                             Some college                              Asian
## 3:                          Graduate degree                          Caucasian
## 4:                           College degree                  Hispanic / Latinx
## 5:                             Some college                          Caucasian
## 6:                     High school graduate                          Caucasian
##                                                                      Q7
## 1: Do you consider yourself to be active on social media platforms?
## 2:                                                              Yes
## 3:                                                              Yes
## 4:                                                              Yes
## 5:                                                              Yes
## 6:                                                              Yes
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise
## 2:
## 3:
## 4:
## 5:
```

```
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise y
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise y
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise y
## 2:
## 3:
## 4:
## 5:
## 6:
##
## 1: Disclaimer: Content on social media may contain false or misleading information. Please exercise y
## 2:
## 3:
## 4:
## 5:
## 6:
##                                                      8A
## 1: Do you believe that the content of the above post is true?
## 2:
## 3:                                                   Yes
## 4:                                                   Yes
## 5:                                                    No
## 6:
##                                                      9A
## 1: Do you believe that the content of the above post is true?
## 2:
## 3:                                                    No
## 4:                                                   Yes
## 5:                                                   Yes
## 6:
##                                                     10A
## 1: Do you believe that the content of the above post is true?
## 2:
## 3:                                                   Yes
## 4:                                                   Yes
## 5:                                                    No
## 6:
##                                                     11A
## 1: Do you believe that the content of the above post is true?
## 2:
## 3:                                                   Yes
## 4:                                                         No
## 5:                                                        Yes
## 6:
##                                                           12A
## 1: Do you believe that the content of the above post is true?
## 2:
## 3:                                                        Yes
## 4:                                                         No
## 5:                                                        Yes
## 6:
##                                                           13A
## 1: Do you believe that the content of the above post is true?
## 2:
## 3:                                                        Yes
## 4:                                                        Yes
## 5:                                                         No
## 6:
##                                                           14A
## 1: Do you believe that the content of the above post is true?
## 2:
## 3:                                                         No
## 4:                                                         No
## 5:                                                        Yes
## 6:
##                                                           15A
## 1: Do you believe that the content of the above post is true?
## 2:
## 3:                                                        Yes
## 4:                                                        Yes
## 5:                                                        Yes
## 6:
##                                                           16A
## 1: Do you believe that the content of the above post is true?
## 2:
## 3:                                                        Yes
## 4:                                                         No
## 5:                                                         No
## 6:
##                                                           17A   SCO
## 1: Do you believe that the content of the above post is true? Score
## 2:                                                                 9
## 3:                                                        Yes     3
## 4:                                                        Yes     5
## 5:                                                         No     8
## 6:                                                                 7
```

```
data_pruned <- mturk1_dataset[ 3:nrow(mturk1_dataset),]
data_pruned[, c(6,7)] <- lapply(data_pruned[, c(6,7)], as.numeric)
```

```
## Warning in lapply(data_pruned[, c(6, 7)], as.numeric): NAs introduced by
## coercion
```

```
question_col_names <- c('8B','9B','10B','11B','12B','13B','14B','15B','16B','17B',
                        '8A','9A','10A','11A','12A','13A','14A','15A','16A','17A')
for(i in c(26:length(names(data_pruned)))){
    data_pruned[[i]][data_pruned[[i]]==''] <- NA
}

# Check the data
# head(data_pruned[, 26:length(names(data_pruned))])

# Set assignment group variable (treatment = 1 , control = 0)
data_pruned[, assignment := ifelse(is.na(data_pruned[,'8B']),0,1)]
```

```
## Warning in `[.data.table`(data_pruned, , `:=`(assignment,
## ifelse(is.na(data_pruned[, : Invalid .internal.selfref detected and fixed by
## taking a (shallow) copy of the data.table so that := can add this new column
## by reference. At an earlier point, this data.table has been copied by R (or
## was created manually using structure() or similar). Avoid names<- and attr<-
## which in R currently (and oddly) may copy the whole data.table. Use set* syntax
## instead to avoid copying: ?set, ?setnames and ?setattr. If this message doesn't
## help, please report your use case to the data.table issue tracker so the root
## cause can be fixed or this message improved.
```

```
head(data_pruned[, 26:length(names(data_pruned))])
```

```
##       Q3         Q4                   Q5                Q6  Q7   8B   9B  10B
## 1: 21-40   Democrat      Graduate degree         Caucasian Yes <NA> <NA> <NA>
## 2: 21-40 Republican       College degree Hispanic / Latinx Yes <NA> <NA> <NA>
## 3: 21-40      Other         Some college         Caucasian Yes <NA> <NA> <NA>
## 4: 21-40      Other High school graduate         Caucasian Yes   No  Yes   No
## 5: 21-40   Democrat       College degree         Caucasian Yes <NA> <NA> <NA>
## 6: 41-60 Republican       College degree         Caucasian Yes   No  Yes   No
##     11B  12B  13B  14B  15B  16B  17B   8A   9A  10A  11A  12A  13A  14A  15A
## 1: <NA> <NA> <NA> <NA> <NA> <NA> <NA>  Yes   No  Yes  Yes  Yes  Yes   No  Yes
## 2: <NA> <NA> <NA> <NA> <NA> <NA> <NA>  Yes  Yes  Yes   No   No  Yes   No  Yes
## 3: <NA> <NA> <NA> <NA> <NA> <NA> <NA>   No  Yes   No  Yes  Yes   No  Yes  Yes
## 4:  Yes  Yes   No  Yes   No   No   No <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 5: <NA> <NA> <NA> <NA> <NA> <NA> <NA>   No  Yes   No   No   No   No   No   No
## 6:  Yes  Yes   No   No  Yes   No  Yes <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
##     16A  17A SC0 assignment
## 1:  Yes  Yes   3          0
## 2:   No  Yes   5          0
## 3:   No   No   8          0
## 4: <NA> <NA>   7          1
## 5:   No   No   6          0
## 6: <NA> <NA>   6          1
```

```
answer_guide <- c('Yes','Yes','No','Yes','No','No','Yes','Yes','No','No',
                  'Yes','Yes','No','Yes','No','No','Yes','Yes','No','No')

compute_scores <- function(dataset,answer_guide){
    for(i in 1:nrow(data_pruned)){
    dataset[i,"score"] <- sum(dataset[i,26:length(names(dataset))] == answer_guide,na.rm=TRUE)
    }
    return(dataset)
}

data_w_scores <- compute_scores(data_pruned,answer_guide)
head(data_w_scores[,31:length(names(data_w_scores))])
```

```
##      8B   9B  10B  11B  12B  13B  14B  15B  16B  17B   8A   9A  10A  11A  12A
## 1: <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>  Yes   No  Yes  Yes  Yes
## 2: <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>  Yes  Yes  Yes   No   No
## 3: <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>   No  Yes   No  Yes  Yes
## 4:   No  Yes   No  Yes  Yes   No  Yes   No   No   No <NA> <NA> <NA> <NA> <NA>
## 5: <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>   No  Yes   No   No   No
## 6:   No  Yes   No  Yes  Yes   No   No  Yes   No  Yes <NA> <NA> <NA> <NA> <NA>
##     13A  14A  15A  16A  17A SCO assignment score
## 1:  Yes   No  Yes  Yes  Yes   3          0     3
## 2:  Yes   No  Yes   No  Yes   5          0     5
## 3:   No  Yes  Yes   No   No   8          0     4
## 4: <NA> <NA> <NA> <NA> <NA>   7          1     5
## 5:   No   No   No   No   No   6          0     6
## 6: <NA> <NA> <NA> <NA> <NA>   6          1     2
```

```
sd <- data_w_scores[, sd(score)]
sd
```

```
## [1] 1.1527
```

```
d <- data_w_scores[, .(scores=mean(score)), by = assignment]
mod <- lm(score ~ assignment, data_w_scores)
ate <- diff(d$scores)
ate
```

```
## [1] 0.40902
```

```
stargazer(mod, type="text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                                score
## -----------------------------------------------
## assignment                    0.409*
##                              (0.227)
##
## Constant                     4.451***
##                              (0.160)
##
## -----------------------------------------------
## Observations                   101
## R2                            0.032
## Adjusted R2                   0.022
## Residual Std. Error       1.140 (df = 99)
## F Statistic             3.250* (df = 1; 99)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

```
## Power calculation
power.t.test(d=ate,sig.level=0.95,n=nrow(data_w_scores),sd=sd,alternative="two.sided")
```

```
##
##      Two-sample t test power calculation
##
##               n = 101
##             delta = 0.40902
##                sd = 1.1527
##         sig.level = 0.95
##             power = 0.99303
##       alternative = two.sided
##
## NOTE: n is number in *each* group
```

EDA with Mturk data set 1

```
data_w_scores[, .(score_mean = mean(score)), by=assignment]
```

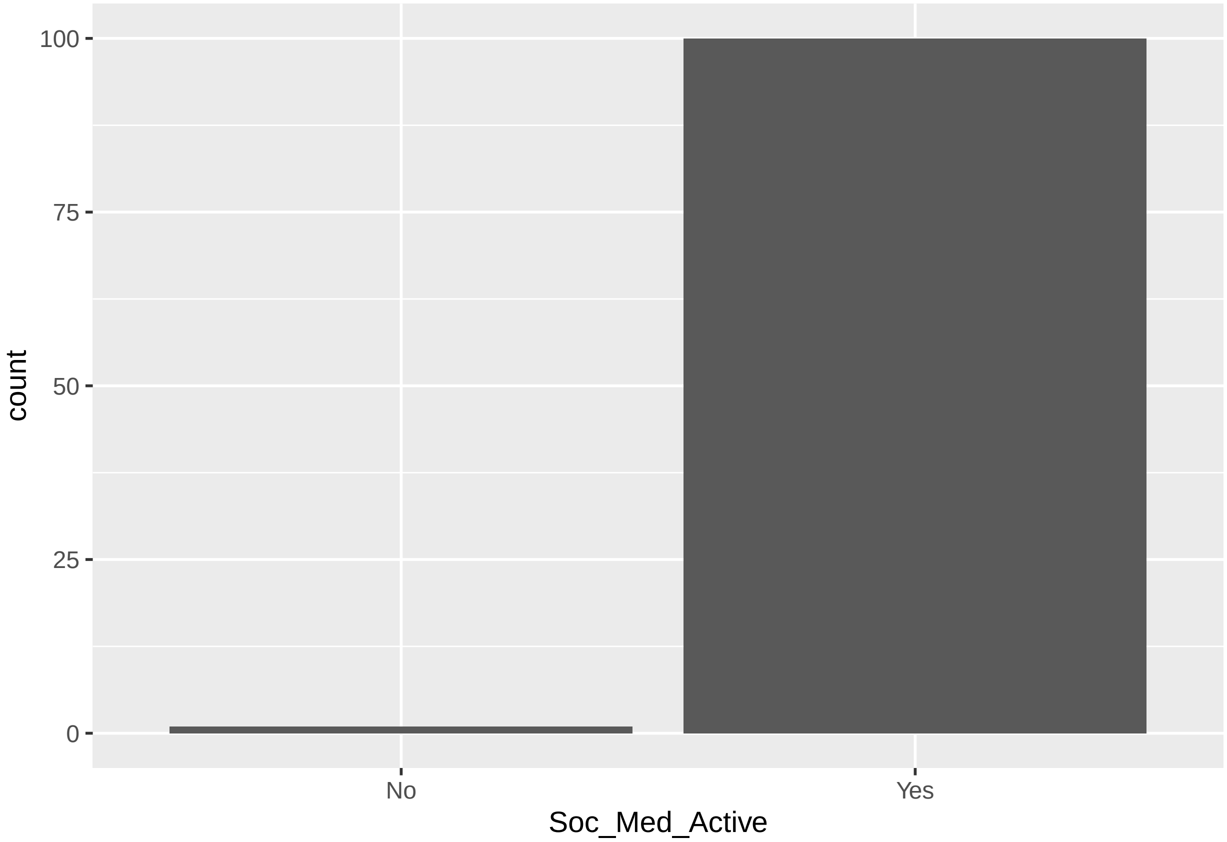
```
##    assignment score_mean
## 1:          0      4.451
## 2:          1      4.860
```

```
data_mod <- rename(data_w_scores,
       Gender = Q1,
       Reg_Voter = Q2,
       Age_bin = Q3,
       Party = Q4,
       Education = Q5,
       Ethnicity = Q6,
       Soc_Med_Active = Q7
        )

data_mod$Gender[data_mod$Gender==''] <- 'Unanswered'

names(data_mod)
```

```
##  [1] "StartDate"             "EndDate"               "Status"
##  [4] "IPAddress"             "Progress"              "Duration (in seconds)"
##  [7] "Finished"              "RecordedDate"          "ResponseId"
## [10] "RecipientLastName"     "RecipientFirstName"    "RecipientEmail"
## [13] "ExternalReference"     "LocationLatitude"      "LocationLongitude"
## [16] "DistributionChannel"   "UserLanguage"          "Q_RecaptchaScore"
## [19] "Gender"                "Reg_Voter"             "Q38"
## [22] "Q39"                   "Q37"                   "Q40"
## [25] "Q36"                   "Age_bin"               "Party"
## [28] "Education"             "Ethnicity"             "Soc_Med_Active"
## [31] "8B"                    "9B"                    "10B"
## [34] "11B"                   "12B"                   "13B"
## [37] "14B"                   "15B"                   "16B"
## [40] "17B"                   "8A"                    "9A"
## [43] "10A"                   "11A"                   "12A"
## [46] "13A"                   "14A"                   "15A"
## [49] "16A"                   "17A"                   "SC0"
## [52] "assignment"            "score"
```

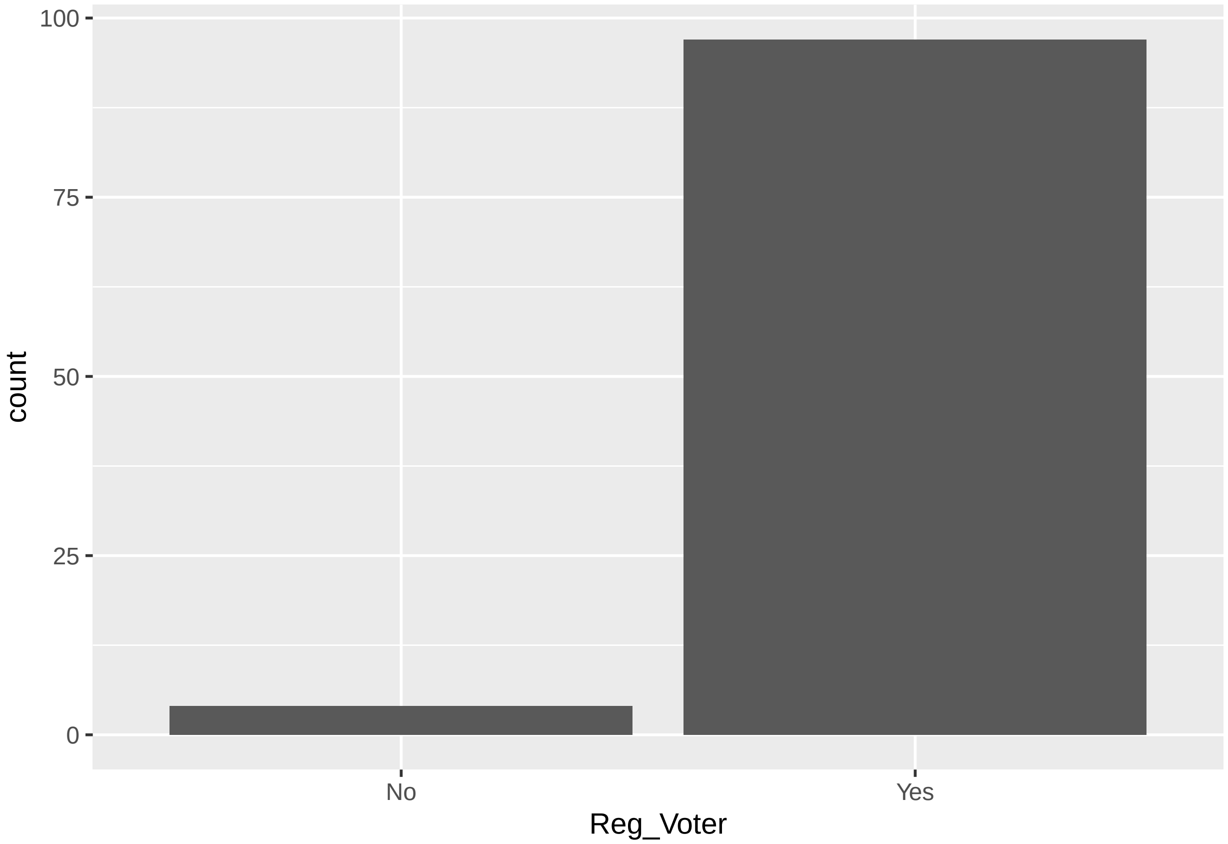
Nearly everyone in the Mturk survey considered themselved active on social media

```
ggplot(data_mod) + geom_bar(aes(x = Soc_Med_Active))
```



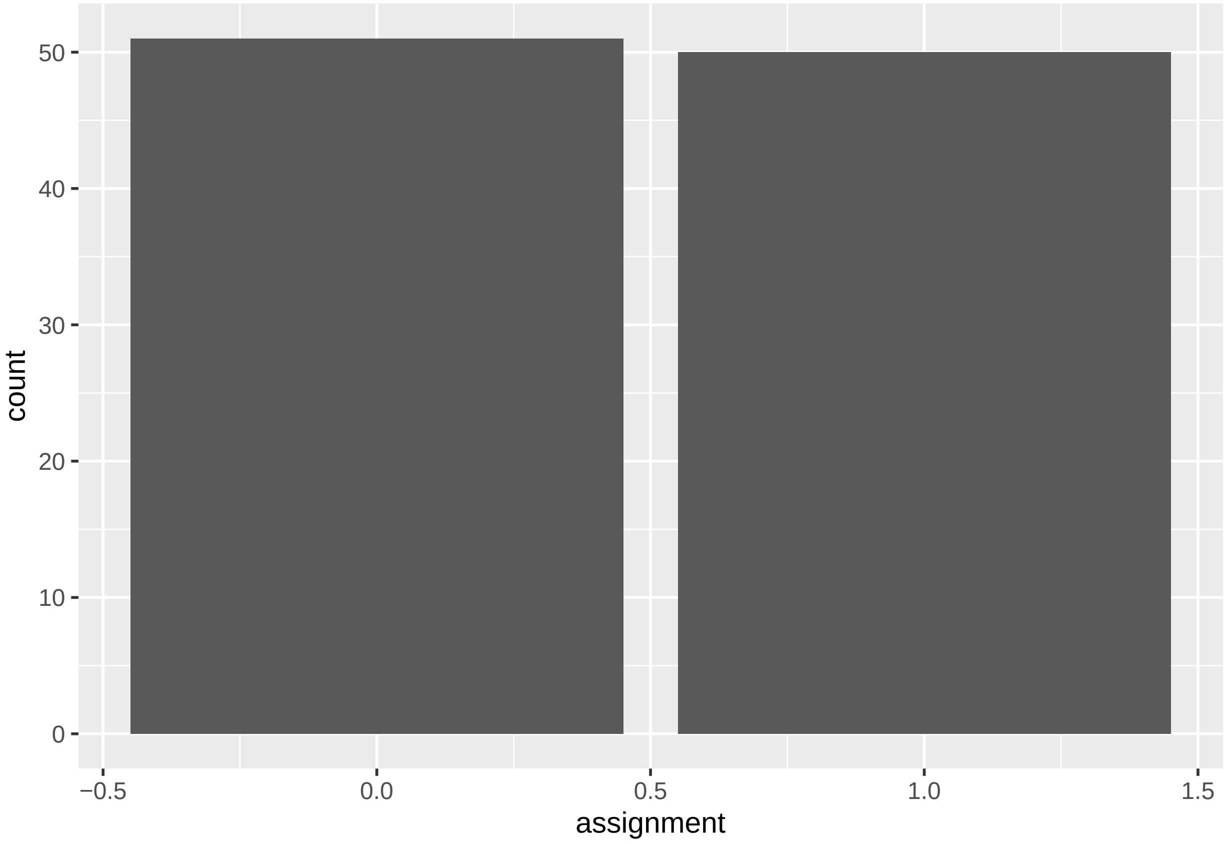
Also nearly everyone said that they were registered as a voter currently

```
ggplot(data_mod) + geom_bar(aes(x = Reg_Voter))
```



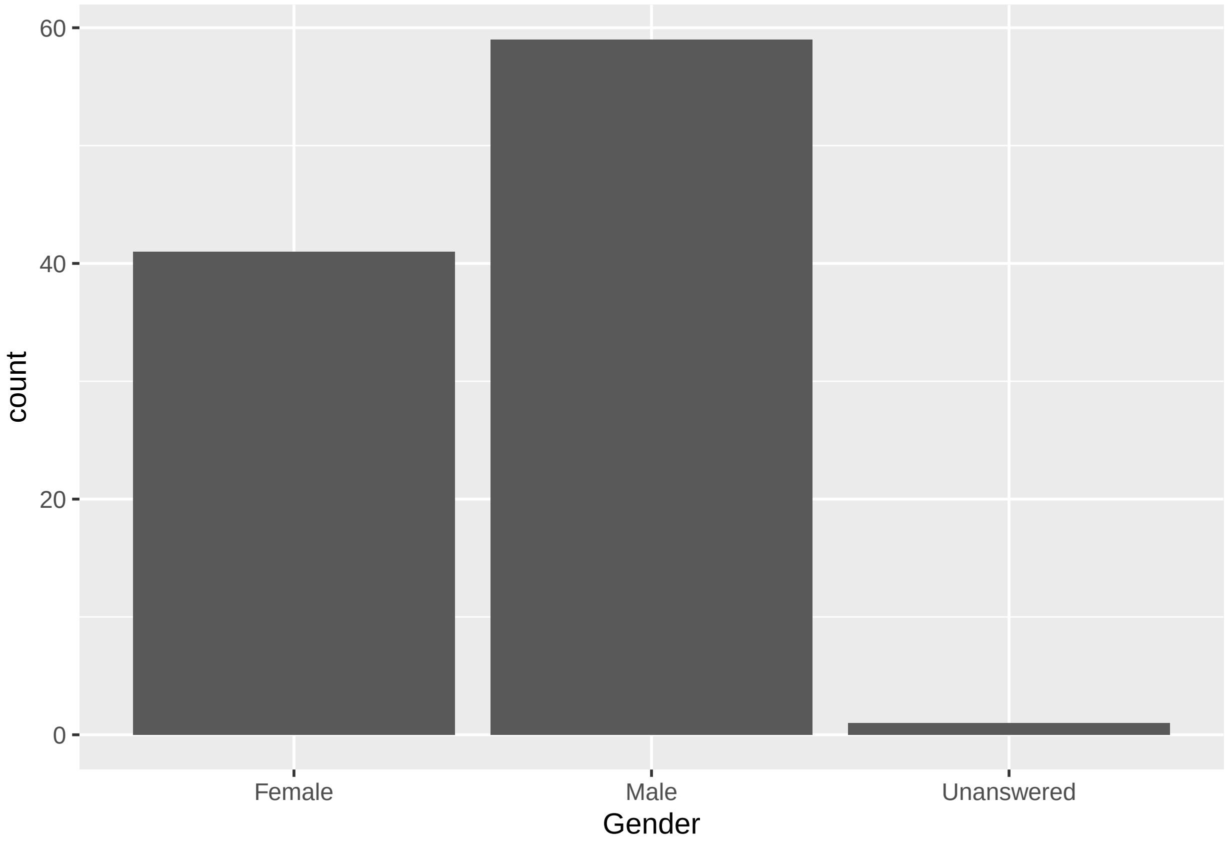
The randomization worked well in the survey software and we had an equal allocation to treatment and control groups in the experiment

```
ggplot(data_mod) + geom_bar(aes(x = assignment))
```



There does seem to be a slight skew in the distribution of participant's gender towards the Male gender (One subject did not answer the gender question)

```
ggplot(data_mod) + geom_bar(aes(x = Gender))
```

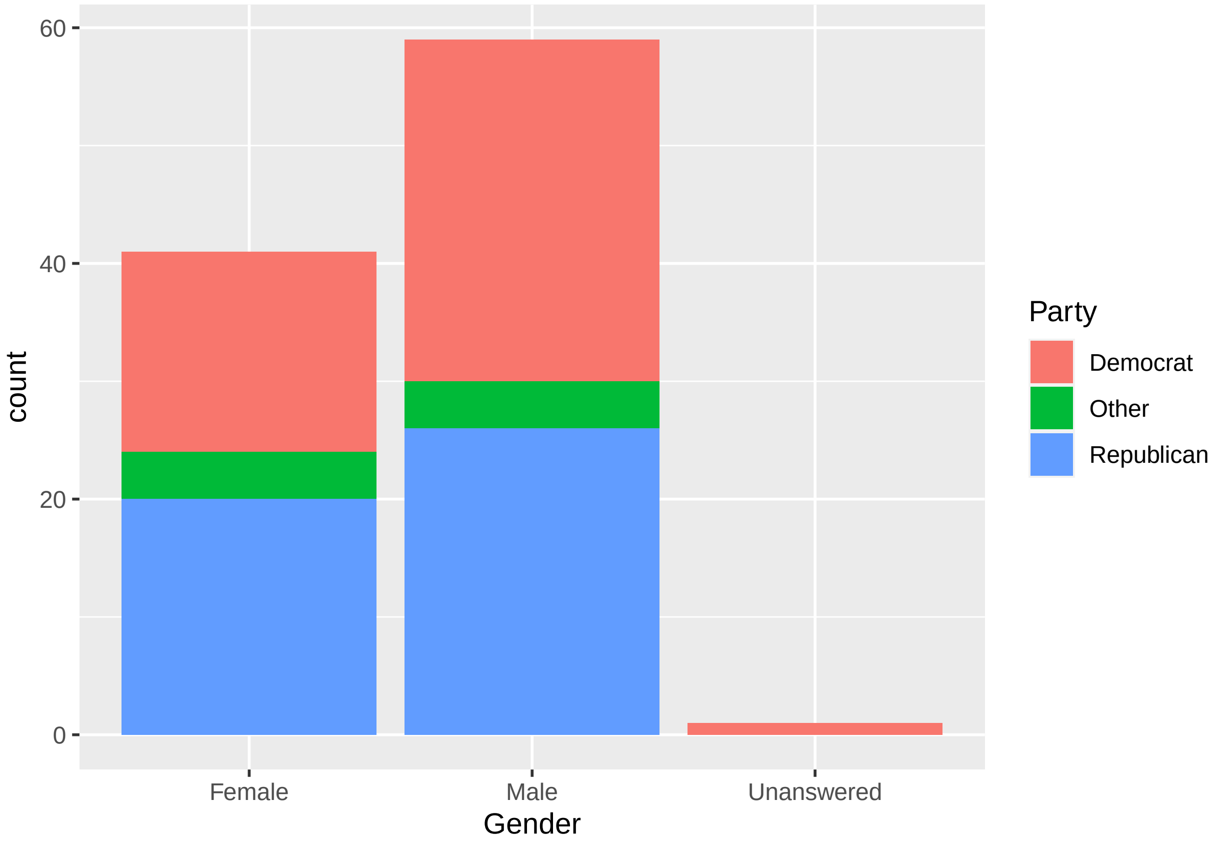


Within each gender category, we see that the party affiliation is approximately evenly distributed.

```
data_mod[, .N, by=.(Gender,Party)]
```

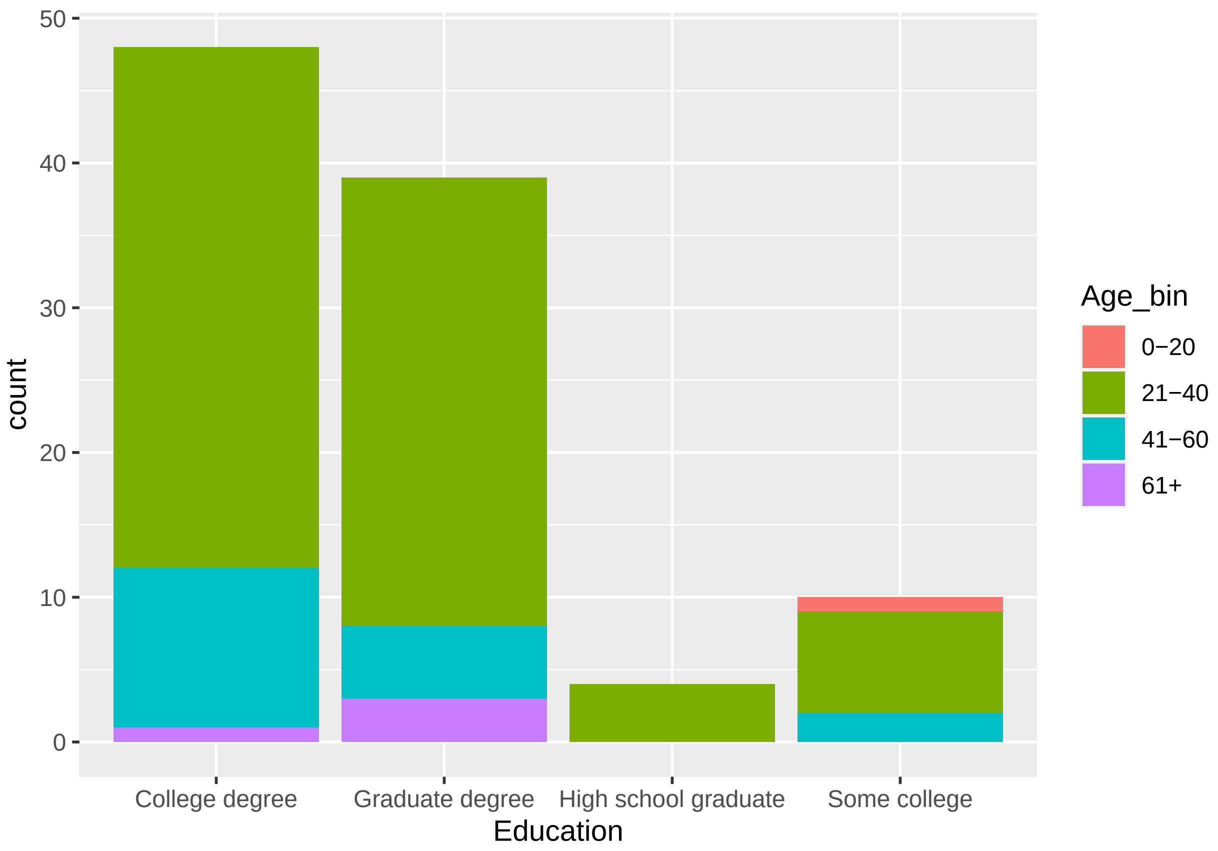
```
##          Gender      Party  N
## 1:      Female   Democrat 17
## 2:      Female Republican 20
## 3:      Female      Other  4
## 4:        Male      Other  4
## 5:        Male   Democrat 29
## 6:        Male Republican 26
## 7: Unanswered   Democrat  1
```

```
ggplot(data_mod) + geom_bar(aes(x = Gender,fill=Party))
```

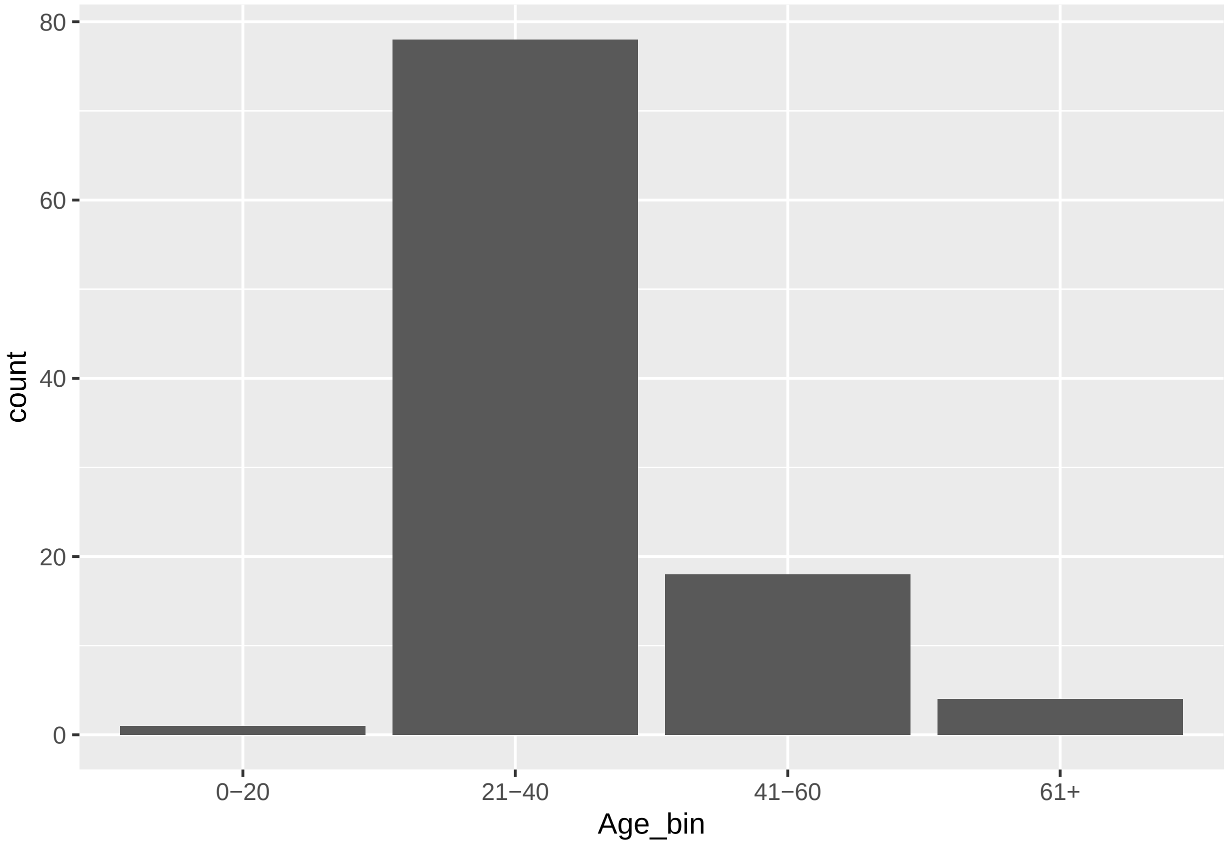


Our dataset does appear to consist mostly of people with atleast a college degree or higher and the participants mostly belong to the 21-40 age bucket.

```
ggplot(data_mod) + geom_bar(aes(x = Education,fill=Age_bin))
```



```
ggplot(mutate(data_mod, Age = fct_infreq(Age_bin))) + geom_bar(aes(x = Age_bin))
```



```
data_mod[, .N, by=.(Party,Age_bin)]
```

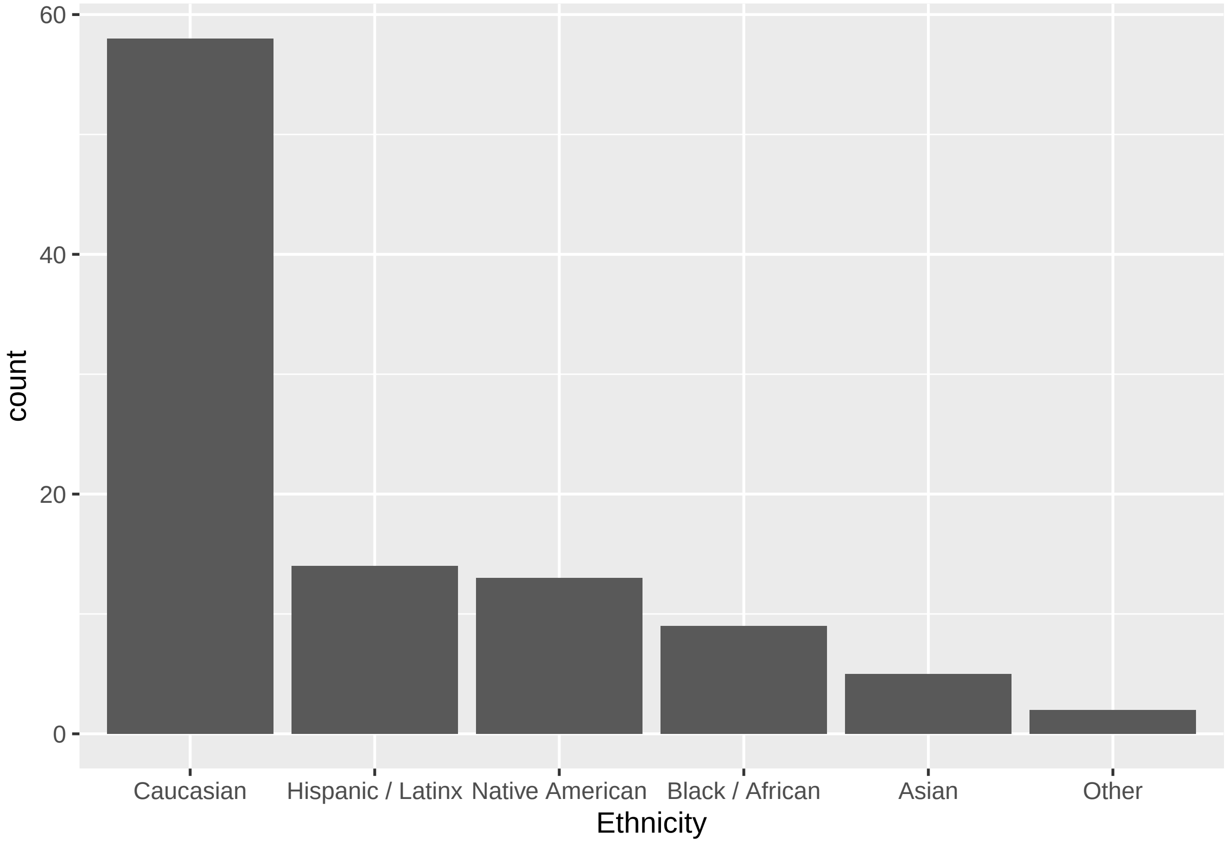
```
##         Party Age_bin  N
## 1:   Democrat   21-40 40
## 2: Republican   21-40 33
## 3:      Other   21-40  5
## 4: Republican   41-60 10
## 5:      Other   41-60  3
## 6:   Democrat   41-60  5
## 7:   Democrat     61+  1
## 8: Republican     61+  3
## 9:   Democrat    0-20  1
```

In terms of ethinicity of the randomly sampled subjects, the majority were Caucasian followed by approximately equal counts of Hispanic and Native americans, followed by african americans and asians.

```
data_mod[, .N, by=Ethnicity]
```

```
##            Ethnicity  N
## 1:         Caucasian 58
## 2: Hispanic / Latinx 14
## 3:   Native American 13
## 4:   Black / African  9
## 5:             Asian  5
## 6:             Other  2
```

```
ggplot(mutate(data_mod, Ethnicity = fct_infreq(Ethnicity))) + geom_bar(aes(x = Ethnicity))
```



```
compute_robust_ci<- function(mod,clustering = FALSE,data=NA) {
  coefs <- names(mod$coefficients)
  if (clustering){
    # calculate robust clustered standard errors
    robust_se <- sqrt(diag(vcovCL(mod,cluster = data)))
  }
  else{
    # calculate robust standard errors without clustering
    robust_se <- sqrt(diag(vcovHC(mod)))
  }
  ci_ll <- NA
  ci_ul <- NA
  for(i in 1:length(coefs)){
    ci_ll[i] <- mod$coefficients[[coefs[i]]] - 1.96 * robust_se[i]
    ci_ul[i] <- mod$coefficients[[coefs[i]]] + 1.96 * robust_se[i]
  }
    ci_custom <- matrix(c(ci_ll,ci_ul), nrow = length(coefs), byrow = FALSE)
    return(ci_custom)
}

compute_robust_se<- function(mod,clustering = FALSE,data=NA) {
  coefs <- names(mod$coefficients)
  if (clustering){
    # calculate robust clustered standard errors
    robust_se <- sqrt(diag(vcovCL(mod,cluster = data)))
  }
  else{
    # calculate robust standard errors without clustering
    robust_se <- sqrt(diag(vcovHC(mod)))
  }

    return(robust_se)
}
```

```
mod <- lm(score ~ assignment+factor(Gender)+factor(Party)+factor(Age_bin)+ factor(Ethnicity)+factor(Edu
se_custom <- compute_robust_se(mod)
stargazer(mod,type="text")
```

```
##
## =================================================================
##                                           Dependent variable:
##                                       ---------------------------
##                                                  score
## -----------------------------------------------------------------
## assignment                                      0.507**
##                                                 (0.240)
##
## factor(Gender)Male                               0.002
##                                                 (0.246)
##
## factor(Gender)Unanswered                       -2.848**
##                                                 (1.165)
##
## factor(Party)Other                               0.202
##                                                 (0.495)
##
## factor(Party)Republican                          0.015
##                                                 (0.261)
##
## factor(Age_bin)21-40                            -0.010
##                                                 (1.275)
##
## factor(Age_bin)41-60                            -0.348
##                                                 (1.305)
##
## factor(Age_bin)61+                              -0.309
##                                                 (1.443)
##
## factor(Ethnicity)Black / African                 0.483
##                                                 (0.658)
##
## factor(Ethnicity)Caucasian                       0.046
##                                                 (0.569)
##
## factor(Ethnicity)Hispanic / Latinx               0.349
##                                                 (0.615)
##
## factor(Ethnicity)Native American                 0.683
##                                                 (0.627)
```

```
##
## factor(Ethnicity)Other                           0.518
##                                                 (1.038)
##
## factor(Education)Graduate degree                 0.186
##                                                 (0.264)
##
## factor(Education)High school graduate           -0.641
##                                                 (0.639)
##
## factor(Education)Some college                   -0.789*
##                                                 (0.456)
##
## Constant                                       4.304***
##                                                 (1.429)
##
## -----------------------------------------------------------------
## Observations                                      101
## R2                                               0.209
## Adjusted R2                                      0.058
## Residual Std. Error                         1.119 (df = 84)
## F Statistic                               1.384 (df = 16; 84)
## =================================================================
## Note:                                 *p<0.1; **p<0.05; ***p<0.01
```