Final Project

Joshua Cai

10/28/2020

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Introduction

For this research project, I will be looking at the question: does distrust in government institutions affect voter turnout? I thought that this question would be interesting to study, given the election that the U.S just went through, and how there is a large amount of polarization pertaining to the state of our current government. I have never seen so much conflict between the people and the government, and I thought that by studying this question, I could gain some insight into the effects of our current political climate. My hypothesis would be that as distrust for a government institution increases, voter turnout would decrease. If someone is to vote, they have to have a certain level of trust for the positions they are voting for and there has to be a belief that the people in these positions will help them. If that trust is missing, then what is there to incentivize someone to take the time to register and go to the polls?

```r
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------------------------------------------------
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ----------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(ggthemes)
data <- read_csv("nonvoters_data.csv")
```

```
##
## -- Column specification ----------------------------------------------------------------------------
```

```
## cols(
##   .default = col_double(),
##   educ = col_character(),
##   race = col_character(),
##   gender = col_character(),
##   income_cat = col_character(),
##   voter_category = col_character()
## )
## i Use `spec()` for the full column specifications.
```

Data

I will be studying a data set from FiveThirtyEight, which contained a large set of data studying the question: Why Many Americans don't vote. This particular data set consisted of answers from an extensive survey, and I will be using two questions from this survey as the basis of my research project:

Question 20. Are you currently registered to vote? 1. Yes 2. No

8. How much would you say you trust each of the following?

- The presidency
- Congress
- The Supreme Court
- The Centers for Disease Control (CDC)
- Election officials
- The intelligence community (e.g. FBI or CIA)
- The news media
- The police
- US Postal Service

[Accordion grid - FLIP 1-4;4-1] 1. A lot 2. Some 3. Not much 4. Not at all

My independent variable would be the question 8 answers, which asked whether a person trusted a certain institution, and asked them to rate their trust of that institution from 1-4. The variable is coded =1 for "A Lot", =2 for "Some", =3 for "Not Much", and =4 for "Not at All". I only looked at the answers pertaining to The presidency, Congress, The Supreme Court, and Election Officials, as these categories seem the most relevant to the election. The dependent variable would be question 20, whether the person actually registered to vote or not. The variable is coded = 1 for "Yes", and = 2 for "No". However, for the regression, I changed the value for "No" to equal "0", as it would have made seeing the relationship of these two variables easier. The research design for this experiment is cross-sectional.

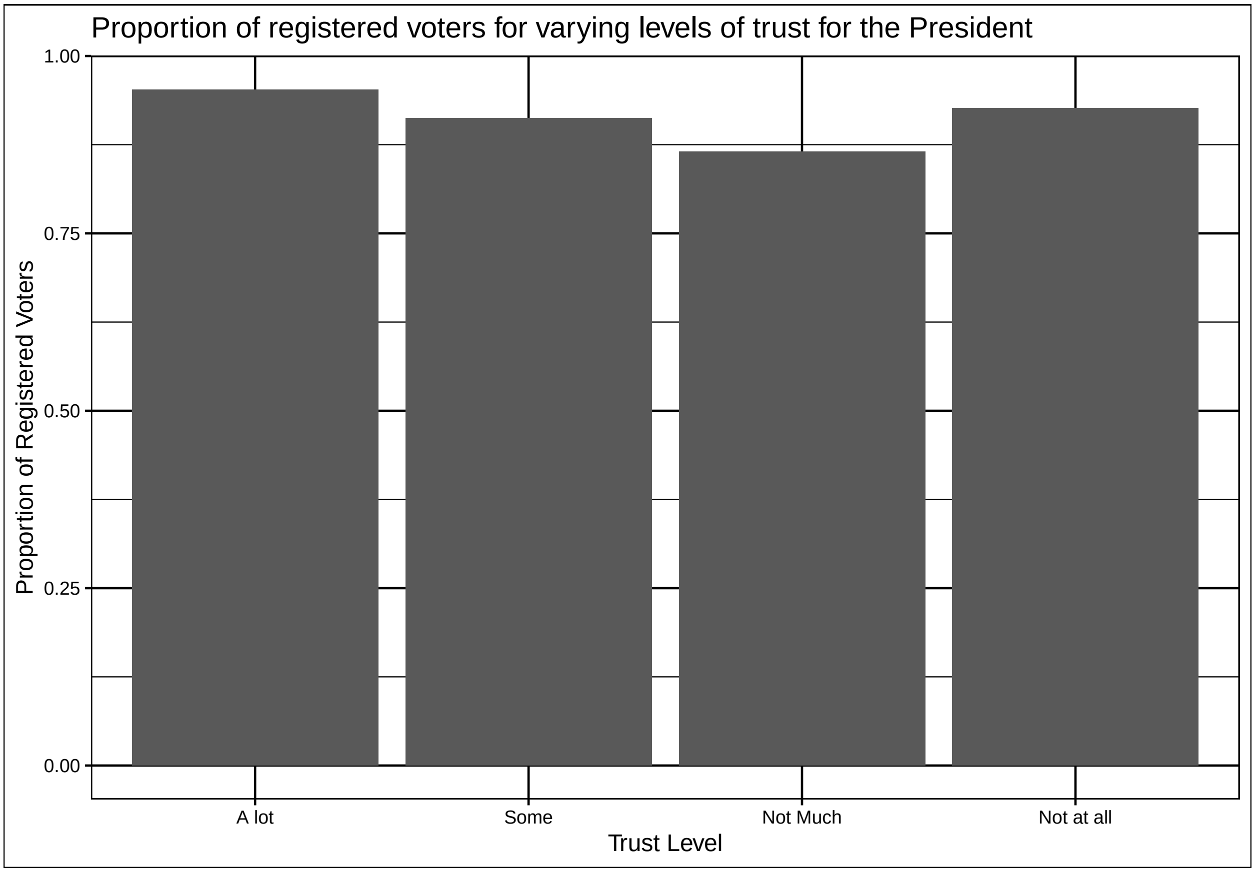
Filtering the data to just include these two questions.

```
filtered <- data %>%
  select(Q8_1, Q8_2, Q8_3, Q8_5, Q20)
```

Plot Analysis between voter turnout and trust for the Presidency

```
presidential_trust <- filtered %>%
  select(Q8_1, Q20) %>%
    filter(Q8_1 != -1) %>%
    count(Q8_1)
```

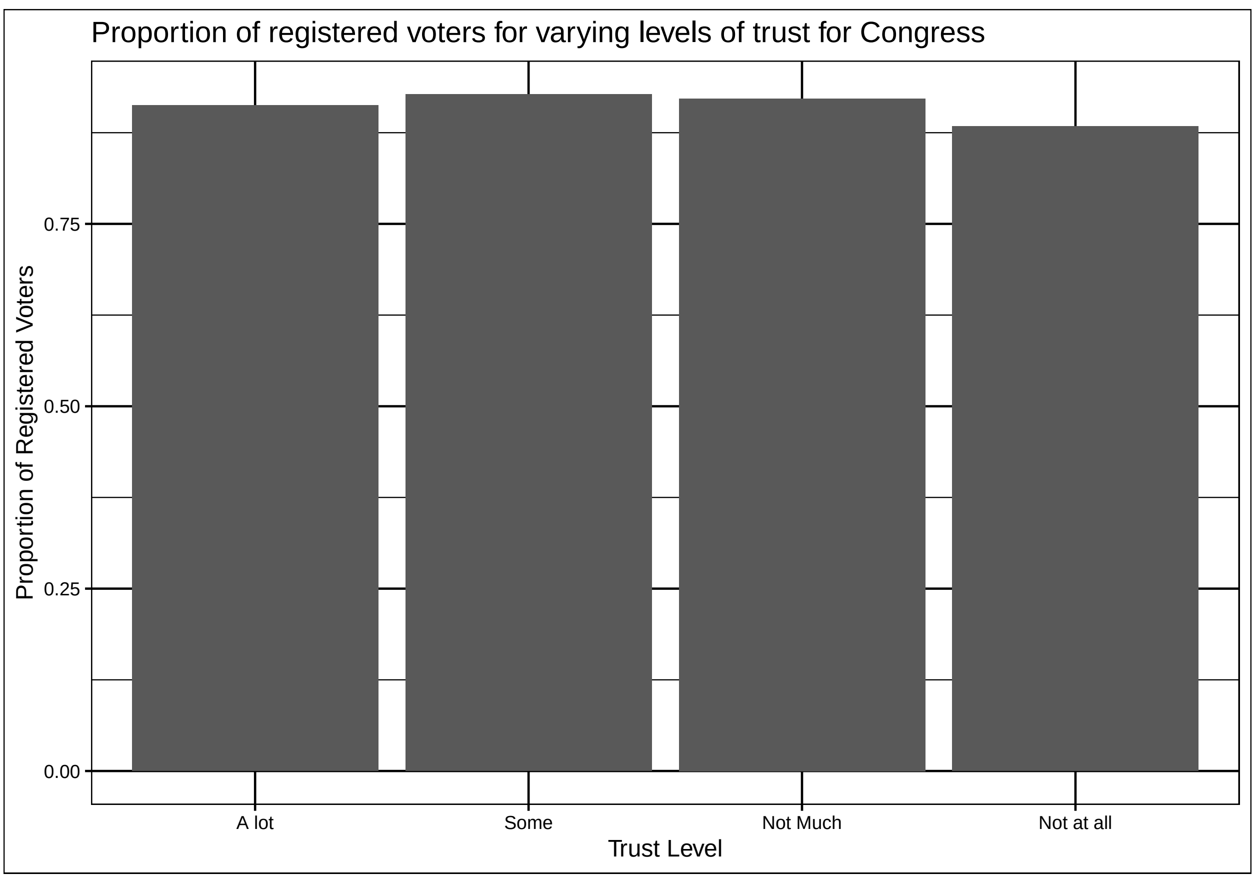
```r
## I found the proportion of people that voted for each level of distrust.
## I wrote out the total number of people that indicated a specific level of distrust
## instead of a command. I could not figure out a command that would take a specific
## value of the the table I made above in presidential_trust, so
## I wrote them all out instead.
presidential_trust_1 <- filtered %>%
  filter(Q8_1 == 1) %>%
  filter(Q20 == 1) %>%
 count(Q8_1) %>%
 mutate(prop = n/948)

presidential_trust_2 <- filtered%>%
  filter(Q8_1 == 2) %>%
  filter(Q20 == 1) %>%
  count(Q8_1) %>%
  mutate(prop = n/1563)


presidential_trust_3 <- filtered %>%
  filter(Q8_1 == 3) %>%
  filter(Q20 == 1) %>%
  count(Q8_1) %>%
  mutate(prop = n/1093)

presidential_trust_4 <- filtered %>%
  filter(Q8_1 == 4) %>%
  filter(Q20 == 1) %>%
  count(Q8_1) %>%
  mutate(prop = n/2185)


bind_rows(presidential_trust_1, presidential_trust_2,
          presidential_trust_3, presidential_trust_4) %>%
    mutate(pres_response = case_when(
    Q8_1 == 1 ~ "A lot",
    Q8_1 == 2 ~ "Some",
    Q8_1 == 3 ~ "Not Much",
    Q8_1 == 4 ~ "Not at all"
  )) %>%
    dplyr::mutate(pres_response = factor(pres_response,
                                   levels = c("A lot", "Some", "Not Much",
                                              "Not at all"))) %>%
  ggplot(aes(x = pres_response, y = prop)) +
  geom_col()+
    labs(title = "Proportion of registered voters for varying levels of trust for the President",
     x = "Trust Level", y = "Proportion of Registered Voters")+
  theme_foundation(base_size = 9)
```



Plot Analysis between voter turnout and trust for Congress

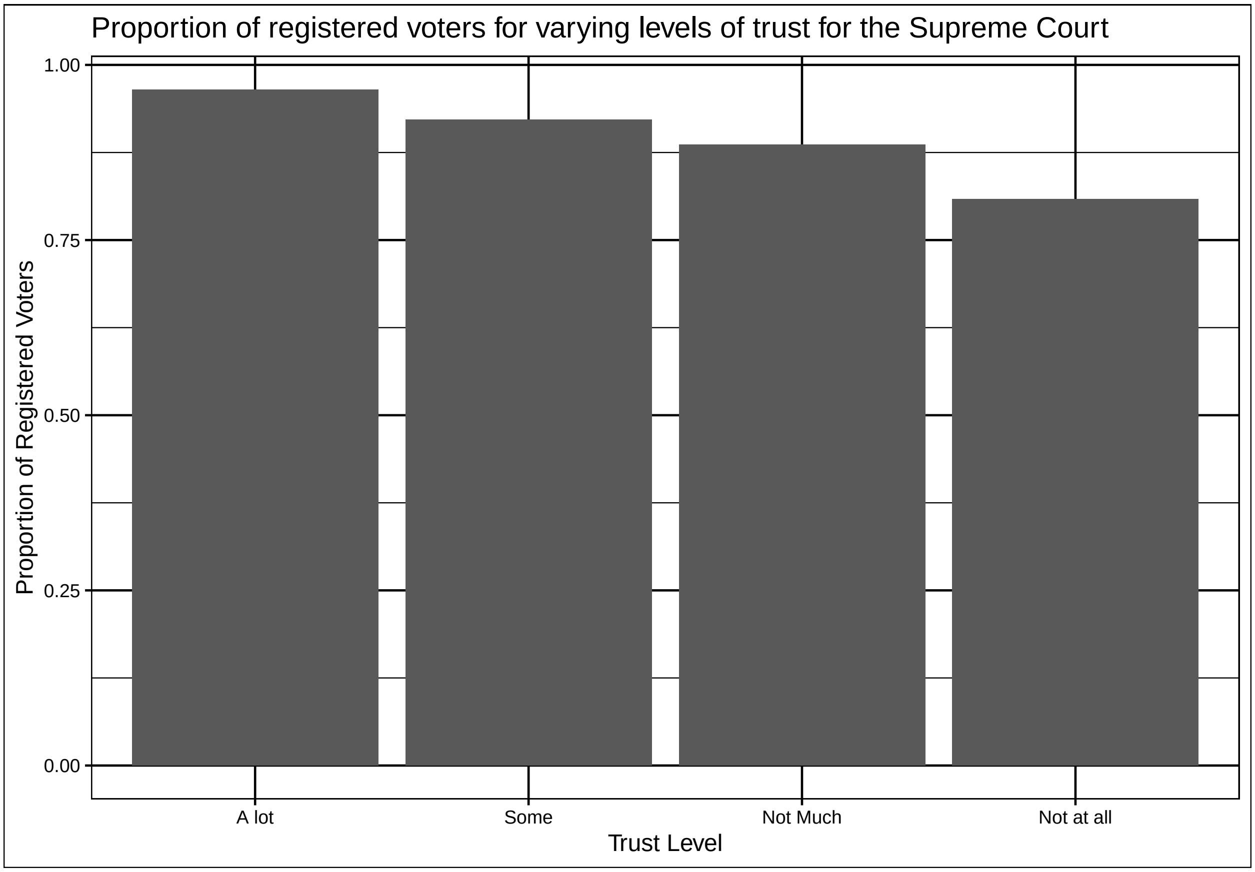
```
congress_trust <- filtered %>%
  select(Q8_2, Q20) %>%
   filter(Q8_2 != -1) %>%
    count(Q8_2)


congress_trust_1 <- filtered %>%
  filter(Q8_2 == 1) %>%
  filter(Q20 == 1) %>%
  count(Q8_2) %>%
  mutate(prop = n/228)

congress_trust_2 <- filtered%>%
  filter(Q8_2 == 2) %>%
  filter(Q20 == 1) %>%
  count(Q8_2) %>%
  mutate(prop = n/2000)


congress_trust_3 <- filtered %>%
  filter(Q8_2 == 3) %>%
  filter(Q20 == 1) %>%
  count(Q8_2) %>%
  mutate(prop = n/2354)
```

```
congress_trust_4 <- filtered %>%
  filter(Q8_2 == 4) %>%
  filter(Q20 == 1) %>%
  count(Q8_2) %>%
  mutate(prop = n/1205)


bind_rows(congress_trust_1, congress_trust_2,
          congress_trust_3, congress_trust_4) %>%
    mutate(congress_response = case_when(
    Q8_2 == 1 ~ "A lot",
    Q8_2 == 2 ~ "Some",
    Q8_2 == 3 ~ "Not Much",
    Q8_2 == 4 ~ "Not at all"
  )) %>%
    dplyr::mutate(congress_response = factor(congress_response,
                                levels = c("A lot", "Some", "Not Much",
                                           "Not at all"))) %>%
  ggplot(aes(x = congress_response, y = prop)) +
  geom_col()+
    labs(title = "Proportion of registered voters for varying levels of trust for Congress",
     x = "Trust Level", y = "Proportion of Registered Voters")+
  theme_foundation(base_size = 9)
```



Plot Analysis between voter turnout and trust for the Supreme Court

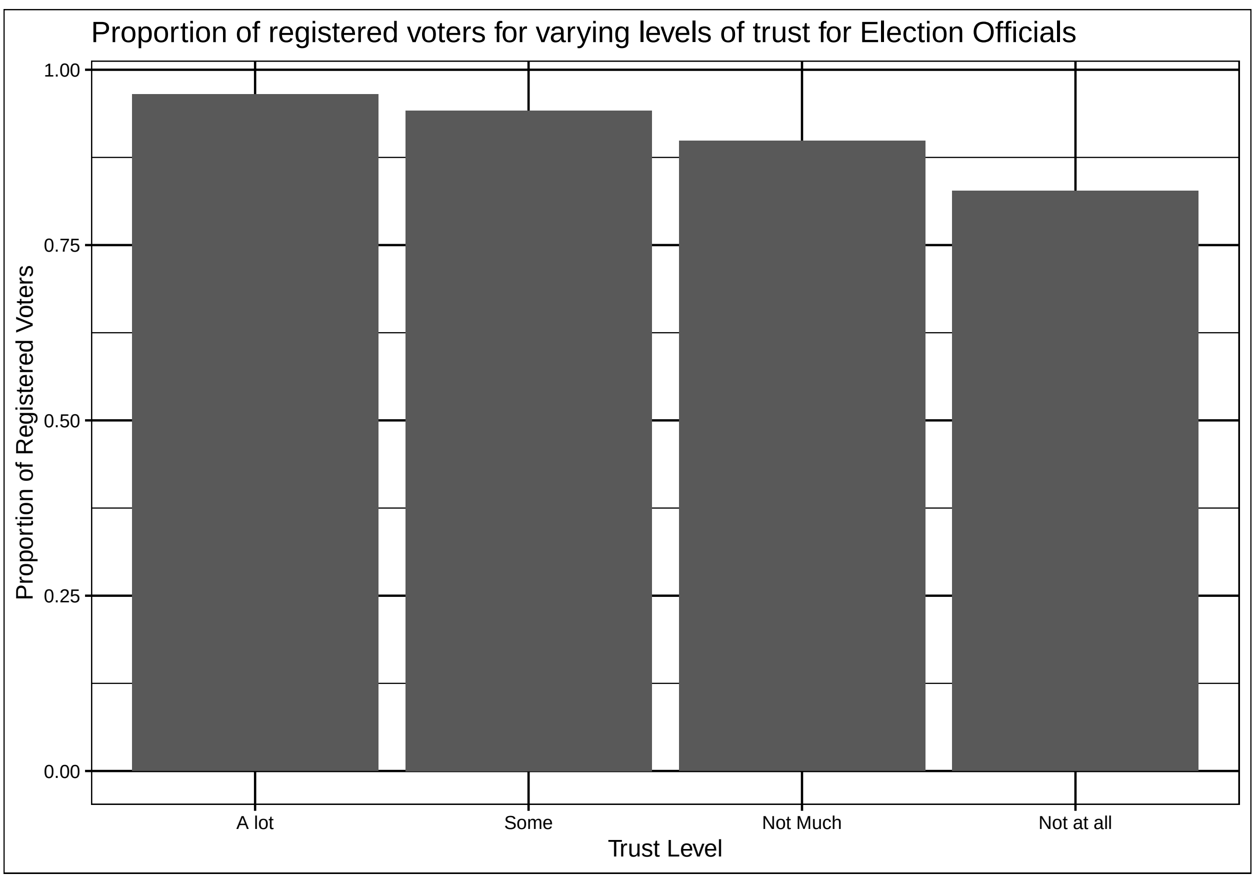
```r
court_trust <- filtered %>%
  select(Q8_3, Q20) %>%
    filter(Q8_3 != -1) %>%
    count(Q8_3)


court_trust_1 <- filtered %>%
  filter(Q8_3 == 1) %>%
  filter(Q20 == 1) %>%
  count(Q8_3) %>%
  mutate(prop = n/1075)

court_trust_2 <- filtered%>%
  filter(Q8_3 == 2) %>%
  filter(Q20 == 1) %>%
  count(Q8_3) %>%
  mutate(prop = n/3187)


court_trust_3 <- filtered %>%
  filter(Q8_3 == 3) %>%
  filter(Q20 == 1) %>%
  count(Q8_3) %>%
  mutate(prop = n/1140)

court_trust_4 <- filtered %>%
  filter(Q8_3 == 4) %>%
  filter(Q20 == 1) %>%
  count(Q8_3) %>%
  mutate(prop = n/392)


bind_rows(court_trust_1, court_trust_2,
          court_trust_3, court_trust_4) %>%
    mutate(court_response = case_when(
    Q8_3 == 1 ~ "A lot",
    Q8_3 == 2 ~ "Some",
    Q8_3 == 3 ~ "Not Much",
    Q8_3 == 4 ~ "Not at all"
  )) %>%
    dplyr::mutate(court_response = factor(court_response,
                                  levels = c("A lot", "Some", "Not Much",
                                             "Not at all"))) %>%
  ggplot(aes(x = court_response, y = prop)) +
  geom_col()+
    labs(title = "Proportion of registered voters for varying levels of trust for the Supreme Court",
     x = "Trust Level", y = "Proportion of Registered Voters")+
  theme_foundation(base_size = 9)
```



Plot Analysis between voter turnout and trust for Election Officials

```
election_trust <- filtered %>%
  select(Q8_5, Q20) %>%
    filter(Q8_5 != -1) %>%
    count(Q8_5)


election_trust_1 <- filtered %>%
  filter(Q8_5 == 1) %>%
  filter(Q20 == 1) %>%
  count(Q8_5) %>%
  mutate(prop = n/541)

election_trust_2 <- filtered%>%
  filter(Q8_5 == 2) %>%
  filter(Q20 == 1) %>%
  count(Q8_5) %>%
  mutate(prop = n/2634)


election_trust_3 <- filtered %>%
  filter(Q8_5 == 3) %>%
  filter(Q20 == 1) %>%
  count(Q8_5) %>%
  mutate(prop = n/1864)
```

```r
election_trust_4 <- filtered %>%
  filter(Q8_5 == 4) %>%
  filter(Q20 == 1) %>%
  count(Q8_5) %>%
  mutate(prop = n/748)


bind_rows(election_trust_1, election_trust_2,
          election_trust_3, election_trust_4) %>%
    mutate(election_response = case_when(
    Q8_5 == 1 ~ "A lot",
    Q8_5 == 2 ~ "Some",
    Q8_5 == 3 ~ "Not Much",
    Q8_5 == 4 ~ "Not at all"
  )) %>%
    dplyr::mutate(election_response = factor(election_response,
                                  levels = c("A lot", "Some", "Not Much",
                                             "Not at all"))) %>%
  ggplot(aes(x = election_response, y = prop)) +
  geom_col()+
     labs(title = "Proportion of registered voters for varying levels of trust for Election Officials",
     x = "Trust Level", y = "Proportion of Registered Voters")+
  theme_foundation(base_size = 9)
```



Results

Regressions

```
regressions <- data %>%
  filter(Q8_1 != -1) %>%
   filter(Q8_2 != -1) %>%
   filter(Q8_3 != -1) %>%
   filter(Q8_5 != -1) %>%
    mutate(Q20 =
           ifelse(Q20 == "2", "0", "1"))



## regression for presidential trust
  pres_regres_test <- lm(Q20 ~ Q8_1, data = regressions)
  pres_regres_test
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_1, data = regressions)
##
## Coefficients:
## (Intercept)         Q8_1
##    0.933752    -0.005693
```

```
  summary(pres_regres_test)
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_1, data = regressions)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.92806  0.07763  0.08333  0.08902  0.08902
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.933752   0.009714  96.122   <2e-16 ***
## Q8_1        -0.005693   0.003240  -1.757    0.079 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2745 on 5724 degrees of freedom
## Multiple R-squared:  0.0005391,  Adjusted R-squared:  0.0003645
## F-statistic: 3.087 on 1 and 5724 DF,  p-value: 0.07896
```

```
  # regression with age as a control
  pres_regres_age <- lm(Q20 ~ Q8_1 * ppage, data = regressions)
 pres_regres_age
```

```
##
```

```
## Call:
## lm(formula = Q20 ~ Q8_1 * ppage, data = regressions)
##
## Coefficients:
## (Intercept)         Q8_1        ppage   Q8_1:ppage
##   0.7710080    0.0102568    0.0028809   -0.0002153
```

```r
summary(pres_regres_age)
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_1 * ppage, data = regressions)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.99984  0.04255  0.07083  0.11122  0.16009
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.7710080  0.0332672  23.176  < 2e-16 ***
## Q8_1         0.0102568  0.0108081   0.949    0.343
## ppage        0.0028809  0.0005798   4.969 6.92e-07 ***
## Q8_1:ppage  -0.0002153  0.0001918  -1.123    0.262
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2718 on 5722 degrees of freedom
## Multiple R-squared:  0.02046,    Adjusted R-squared:  0.01994
## F-statistic: 39.83 on 3 and 5722 DF,  p-value: < 2.2e-16
```

```r
# regression with gender as a control

 pres_regres_gender <- lm(Q20 ~ Q8_1 * gender, data = regressions)
 pres_regres_gender
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_1 * gender, data = regressions)
##
## Coefficients:
##     (Intercept)             Q8_1       genderMale  Q8_1:genderMale
##        0.924399        -0.001654         0.018548        -0.007980
```

```r
summary(pres_regres_gender)
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_1 * gender, data = regressions)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.93331  0.07632  0.08056  0.08595  0.09559
```

```
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.924399   0.013761  67.174   <2e-16 ***
## Q8_1            -0.001654   0.004603  -0.359    0.719
## genderMale       0.018548   0.019429   0.955    0.340
## Q8_1:genderMale -0.007980   0.006481  -1.231    0.218
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2745 on 5722 degrees of freedom
## Multiple R-squared:  0.0008479,  Adjusted R-squared:  0.000324
## F-statistic: 1.619 on 3 and 5722 DF,  p-value: 0.1828
```

```
  ## regressions for congressional trust

  congress_regres_test <- lm(Q20 ~ Q8_2, data = regressions)
congress_regres_test
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_2, data = regressions)
##
## Coefficients:
## (Intercept)         Q8_2
##     0.96028     -0.01521
```

```
summary(congress_regres_test)
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_2, data = regressions)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.94507  0.07014  0.08535  0.08535  0.10056
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.960285   0.012903  74.422  < 2e-16 ***
## Q8_2        -0.015210   0.004446  -3.421 0.000628 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2743 on 5724 degrees of freedom
## Multiple R-squared:  0.002041,   Adjusted R-squared:  0.001866
## F-statistic:  11.7 on 1 and 5724 DF,  p-value: 0.000628
```

```
  # regression with age as a control
 congress_regres_age <- lm(Q20 ~ Q8_2 * ppage, data = regressions)
 congress_regres_age
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_2 * ppage, data = regressions)
##
## Coefficients:
## (Intercept)         Q8_2        ppage   Q8_2:ppage
##   0.9475389   -0.0524729    0.0001332    0.0007654
```

```r
summary(congress_regres_age)
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_2 * ppage, data = regressions)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.01241  0.04225  0.06755  0.10749  0.19206
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.9475389  0.0411682  23.016  < 2e-16 ***
## Q8_2        -0.0524729  0.0141304  -3.713 0.000206 ***
## ppage        0.0001332  0.0007560   0.176 0.860131
## Q8_2:ppage   0.0007654  0.0002615   2.927 0.003431 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2714 on 5722 degrees of freedom
## Multiple R-squared:  0.0232, Adjusted R-squared:  0.02269
## F-statistic:  45.3 on 3 and 5722 DF,  p-value: < 2.2e-16
```

```r
# regression with gender as a control

 congress_test_gender <- lm(Q20 ~ Q8_2 * gender, data = regressions)
 congress_test_gender
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_2 * gender, data = regressions)
##
## Coefficients:
##     (Intercept)              Q8_2        genderMale  Q8_2:genderMale
##        0.951089         -0.011562          0.017843        -0.006896
```

```r
summary(congress_test_gender)
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_2 * gender, data = regressions)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.95047  0.06799  0.08360  0.08644  0.10490
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.951089   0.018142  52.425   <2e-16 ***
## Q8_2             -0.011562   0.006431  -1.798   0.0722 .
## genderMale        0.017843   0.025917   0.688   0.4912
## Q8_2:genderMale -0.006896   0.008941  -0.771   0.4406
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2743 on 5722 degrees of freedom
## Multiple R-squared:  0.00215,    Adjusted R-squared:  0.001627
## F-statistic:  4.11 on 3 and 5722 DF,  p-value: 0.00637
```

```
  ## regressions for court trust

court_regres_test <- lm(Q20 ~ Q8_3, data = regressions)
court_regres_test
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_3, data = regressions)
##
## Coefficients:
## (Intercept)         Q8_3
##     1.01567     -0.04551
```

```
summary(court_regres_test)
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_3, data = regressions)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.97017  0.02983  0.07534  0.07534  0.16636
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.015674   0.010370   97.94   <2e-16 ***
## Q8_3        -0.045508   0.004528  -10.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2721 on 5724 degrees of freedom
## Multiple R-squared:  0.01734,    Adjusted R-squared:  0.01717
## F-statistic:   101 on 1 and 5724 DF,  p-value: < 2.2e-16
```

```
  # regression with age as a control
  court_test_age <- lm(Q20 ~ Q8_3 * ppage, data = regressions)
   court_test_age
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_3 * ppage, data = regressions)
##
## Coefficients:
## (Intercept)         Q8_3        ppage   Q8_3:ppage
##   1.0469357   -0.1080752   -0.0009039    0.0013726
```

```r
summary(court_test_age)
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_3 * ppage, data = regressions)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.00907  0.03489  0.05874  0.10293  0.28446
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.0469357  0.0331547  31.577  < 2e-16 ***
## Q8_3        -0.1080752  0.0140784  -7.677 1.91e-14 ***
## ppage       -0.0009039  0.0006117  -1.478     0.14
## Q8_3:ppage   0.0013726  0.0002677   5.128 3.02e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2694 on 5722 degrees of freedom
## Multiple R-squared:  0.0376, Adjusted R-squared:  0.0371
## F-statistic: 74.52 on 3 and 5722 DF,  p-value: < 2.2e-16
```

```r
# regression with gender as a control

 court_test_gender <- lm(Q20 ~ Q8_3 * gender, data = regressions)
 court_test_gender
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_3 * gender, data = regressions)
##
## Coefficients:
##     (Intercept)             Q8_3       genderMale  Q8_3:genderMale
##         1.00267         -0.03847          0.02386         -0.01310
```

```r
summary(court_test_gender)
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_3 * gender, data = regressions)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.97497  0.03579  0.07426  0.07660  0.17973
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.002672   0.015250  65.751  < 2e-16 ***
## Q8_3             -0.038466   0.006670  -5.767 8.47e-09 ***
## genderMale        0.023860   0.020801   1.147    0.251
## Q8_3:genderMale -0.013100   0.009084  -1.442    0.149
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2721 on 5722 degrees of freedom
## Multiple R-squared:  0.01776,    Adjusted R-squared:  0.01724
## F-statistic: 34.48 on 3 and 5722 DF,  p-value: < 2.2e-16
```

```
  ## regressions for election official trust

elect_regres_test <- lm(Q20 ~ Q8_5, data = regressions)
elect_regres_test
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_5, data = regressions)
##
## Coefficients:
## (Intercept)         Q8_5
##     1.03054     -0.04527
```

```
summary(elect_regres_test)
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_5, data = regressions)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9853  0.0600  0.0600  0.1053  0.1505
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.030536   0.011322   91.02   <2e-16 ***
## Q8_5        -0.045269   0.004316  -10.49   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2719 on 5724 degrees of freedom
## Multiple R-squared:  0.01886,    Adjusted R-squared:  0.01869
## F-statistic:   110 on 1 and 5724 DF,  p-value: < 2.2e-16
```

```
  # regression with age as a control
  elect_test_age <- lm(Q20 ~ Q8_5 * ppage, data = regressions)
  elect_test_age
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_5 * ppage, data = regressions)
##
## Coefficients:
## (Intercept)         Q8_5        ppage   Q8_5:ppage
##    1.081219    -0.108022    -0.001247     0.001340
```

```r
summary(elect_test_age)
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_5 * ppage, data = regressions)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.00289  0.03021  0.06174  0.09613  0.26037
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.0812194  0.0359570  30.070  < 2e-16 ***
## Q8_5        -0.1080222  0.0135233  -7.988 1.65e-15 ***
## ppage       -0.0012472  0.0006559  -1.902   0.0573 .
## Q8_5:ppage   0.0013402  0.0002530   5.298 1.22e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2691 on 5722 degrees of freedom
## Multiple R-squared:  0.0395, Adjusted R-squared:  0.039
## F-statistic: 78.44 on 3 and 5722 DF,  p-value: < 2.2e-16
```

```r
# regression with gender as a control

 elect_test_gender <- lm(Q20 ~ Q8_5 * gender, data = regressions)
 elect_test_gender
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_5 * gender, data = regressions)
##
## Coefficients:
##     (Intercept)              Q8_5        genderMale   Q8_5:genderMale
##         1.00736          -0.03538           0.04275          -0.01821
```

```r
summary(elect_test_gender)
```

```
##
## Call:
## lm(formula = Q20 ~ Q8_5 * gender, data = regressions)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.99651  0.05708  0.06341  0.11067  0.16426
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.007358   0.016575  60.777  < 2e-16 ***
## Q8_5            -0.035384   0.006373  -5.552 2.95e-08 ***
## genderMale       0.042746   0.022702   1.883   0.0598 .
## Q8_5:genderMale -0.018207   0.008660  -2.102   0.0356 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2719 on 5722 degrees of freedom
## Multiple R-squared:  0.01964,    Adjusted R-squared:  0.01912
## F-statistic: 38.21 on 3 and 5722 DF,  p-value: < 2.2e-16
```

Conclusion

The results of this experiment showed that as distrust in government institutions increased, voter turnout decreased. This basic relationship is shown by the regressions that I conducted on each of the government institutions that I laid out in the introduction. The barplots that I made also show this relationship, as the proportion of people voting decreases for the increasing levels of distrust. For my regressions, the main coefficient of interest were the slopes of each of the main regressions, so for example, coef(pres_regres_test)["Q8_1"], which was -0.0056933. Since these slopes were negative, they show that for every incremental increase in distrust, the amount of people that registered to vote went down, which proves my hypothesis. I also created other regressions with controls for gender and age, and the basic relationship I laid out in my hypothesis still applied, as voter turnout still decreased when distrust increased, holding variables such as gender and age constant. The main coefficients that I am looking at are statisically significant, as the p-values associated with them are below 0.05, with the exception of the regression for Q8_1. I think there are limitations to my analysis, as distrust of government institutions is certainly not the only consideration people have when deciding not to vote, and for the people and the answers I looked at, there could have been any number of other factors that influenced their decision to go to the polls or not. These could be confounders for my study. Because of this, I don't think the coefficients from my regressions can be interpreted as causal. I also think recording the respondents' geographic location might have been useful for this study, as political views and sentiments about our government could also be affected by the environment that people live and where they come from.

'