DATA607 - Assignment1

Michael Ippolito

2/4/2021

Overview

In their article, Why Americans Don't Vote, Thomson-Deveaux et al. (2020) explored the reasons why a large number of eligible voters (35 to 60 percent) don't vote in US elections and at the voting habits of voters broken out by various categories (age, level of education, race, gender, and income). The data collected confirmed the well-accepted notion that older, more educated people with higher incomes and stronger party affiliations are more likely to vote. Of voters who never, rarely, or only sometimes vote in US elections, Thomson-Deveaux et al. reported the top reasons why that was the case. My assignment will focus on the specific polling question that dealt with the reason why people often don't vote, categorized by the frequency with which they do vote.

Article citation:
FiveThirtyEight (2020). Why Americans Don't Vote. https://projects.fivethirtyeight.com/non-voters-poll-2020-election/.

Load libraries, retrieve data from github, and parse CSV

```
library(tidyverse)
```

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

```
## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.0.6      v dplyr   1.0.3
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1
```

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

```
library(RCurl)
```

```
##
## Attaching package: 'RCurl'
```

```
## The following object is masked from 'package:tidyr':
##
##     complete
```

```
#original URL on fivethirtyeight,com:
#https://raw.githubusercontent.com/fivethirtyeight/data/master/non-voters/nonvoters_data.csv
nonvoters_csv <- getURL("https://raw.githubusercontent.com/mmippolito/cuny/main/data607/assignment1/non
nonvoters <- read.csv(text = nonvoters_csv)
```

Revelant variables

The data included the responses from all survey questions, tabulated in 119 variables and 5,836 observations.
For this assignment, the subset I chose included the following variables:

voter_category How often the voter voted:

  always
  sporadic
  rarely/never

Q29 Survey question #29, which asked voters to mark which of the following

  ten reason were most important factors in why they chose not to vote.
  (A value of 1 indicates the voter marked it as important; a value of -1 means
  they answered the question, but didn't mark this answer as important; and N/A
  means they didn't answer the question.)

  Q29_1 I didn't like any of the candidates
  Q29_2 Because of where I live, my vote doesn't matter
  Q29_3 No matter who wins, nothing will change for people like me
  Q29_4 Our system is too broken to be fixed by voting
  Q29_5 I wanted to vote, but I didn't have time, couldn't get off work,
      something came up, or I forgot
  Q29_6 I'm not sure if I can vote
  Q29_7 Nobody talks about the issues that are important to me personally
  Q29_8 All the candidates are the same
  Q29_9 I don't believe in voting
  Q29_10 Other

Create Q29 array

```
# Make Q29 into an array so it can be later iterated over in a "for" loop
q29 <- c(
  "I didn't like any of the candidates",
  "Because of where I live, my vote doesn't matter",
  "No matter who wins, nothing will change for people like me",
  "Our system is too broken to be fixed by voting",
  "I wanted to vote, but I didn't have time, couldn't get off work, something came up, or I forgot",
  "I'm not sure if I can vote",
  "Nobody talks about the issues that are important to me personally",
  "All the candidates are the same",
  "I don't believe in voting",
  "Other"
)
```

Create subset from relevant variables

First, filter out voters who "always" vote, and only select voter_category, weight, and the Question 29 responses.

```
# Filter out voters who always vote; only select specific variables
nonvoters_29 <- as_tibble(select(nonvoters, voter_category, weight, Q29_1:Q29_10)) %>%
  filter(voter_category != "always")
nonvoters_29[1:5,]     # Display first 5 observations
```

```
## # A tibble: 5 x 12
##   voter_category weight Q29_1 Q29_2 Q29_3 Q29_4 Q29_5 Q29_6 Q29_7 Q29_8 Q29_9
##   <chr>           <dbl> <int> <int> <int> <int> <int> <int> <int> <int> <int>
## 1 sporadic         1.08    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 2 sporadic        0.682    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 3 rarely/never     1.06    -1    -1    -1    -1    -1    -1    -1    -1     1
## 4 sporadic        0.952    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 5 sporadic         1.37    NA    NA    NA    NA    NA    NA    NA    NA    NA
## # ... with 1 more variable: Q29_10 <int>
```

Pivot on Q29 responses

Now create 10 different tibbles–one for each answer in question 29–and group on voter category.

```
# Iterate over each answer in Question 29
for(i in 1:10) {

  # Concatenate i to create variable name string
  q <- paste("Q29_", i, sep = "")

  # Create tibble with weighted count of voters who answered that this was
  # an important reason why they didn't vote
  categories <- select(nonvoters_29, voter_category, weight, i + 2) %>%
    filter(!is.na(get(q)) & get(q) == 1) %>%
    group_by(voter_category) %>%
    summarize(wt = sum(weight))

  # Create tibble with weighted counts of voters who answered this question at all
  totals <- select(nonvoters_29, voter_category, weight, i + 2) %>%
    filter(!is.na(get(q))) %>%
    group_by(voter_category) %>%
    summarize(wt_total = sum(weight))

  # Merge the two tibbles
  subset <- merge(categories, totals, by = "voter_category")

  # Create a new variable for percentage and print the new tibble
  subset <- mutate(subset, percentage = wt * 100 / wt_total)
  print(subset)

  # Plot the bar chart
  print(ggplot(data = subset, mapping = aes(x = voter_category, y = percentage)) +
    geom_bar(stat = "identity", mapping = aes(color = voter_category, fill = voter_category)) +
```
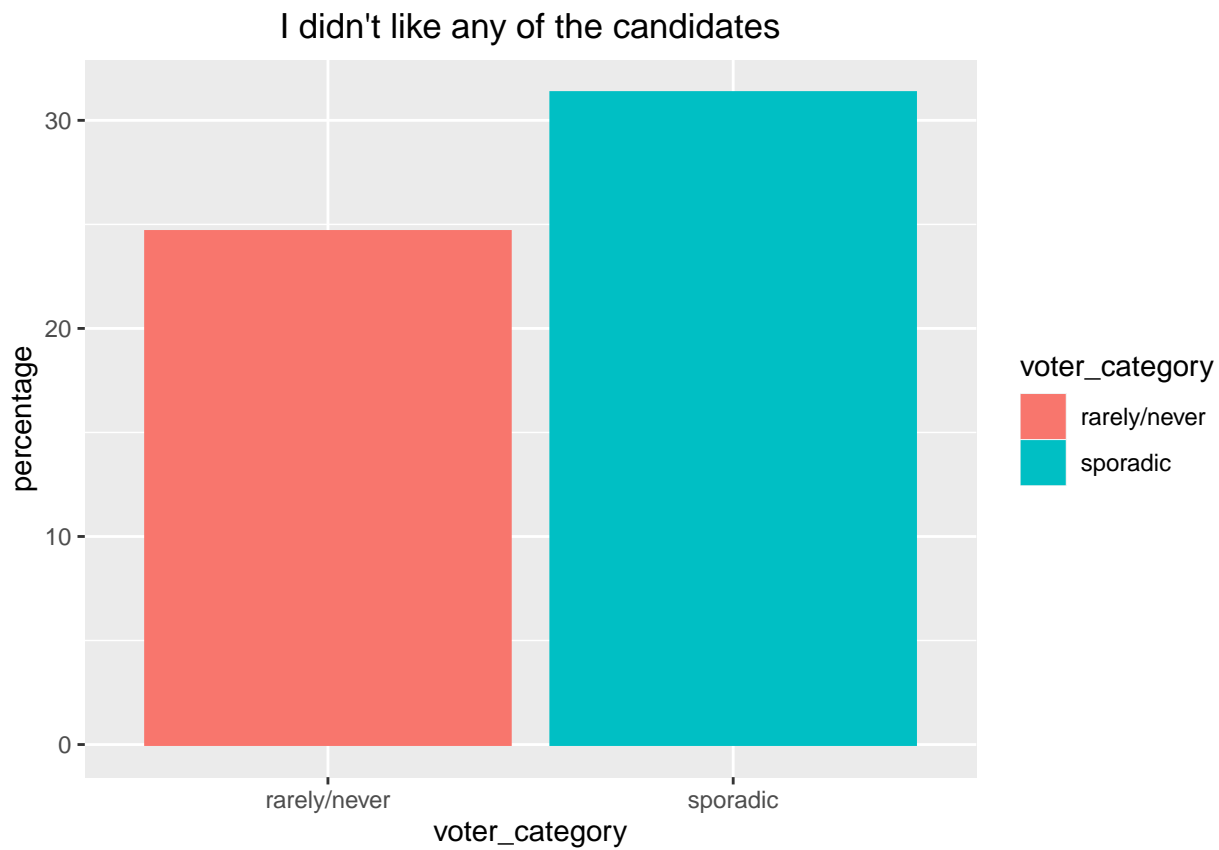
```
    ggtitle(q29[i]) +
    theme(plot.title = element_text(hjust = 0.5)))

}
```
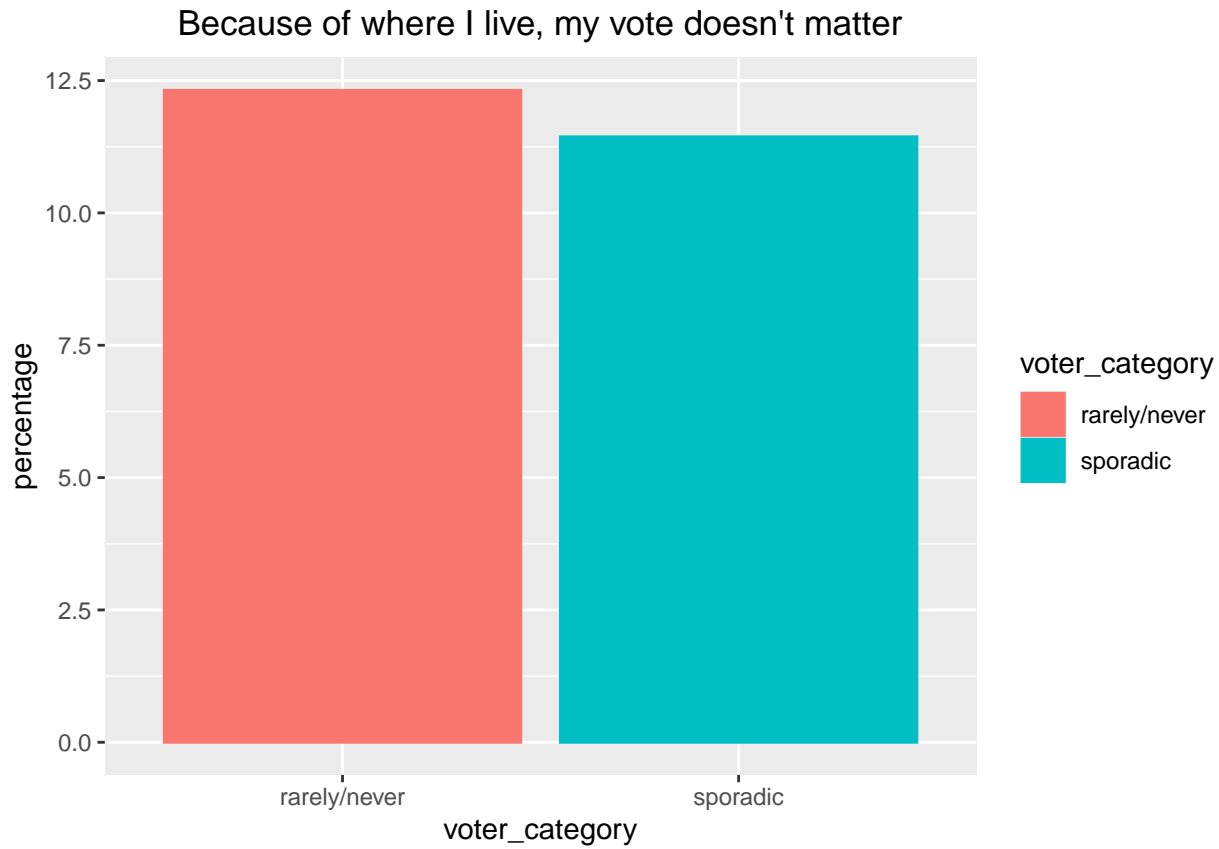
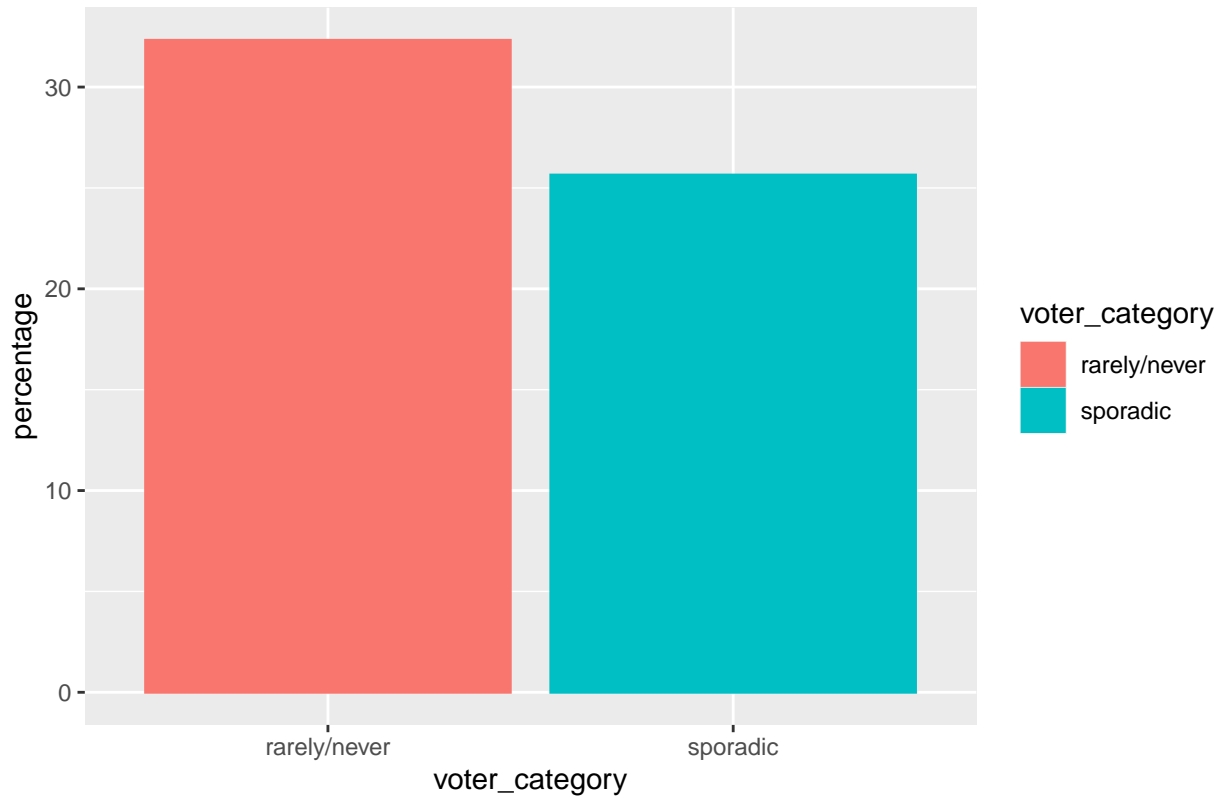voter_category wt wt_total percentage 1 rarely/never 282.2935 1144.7680 24.65945 2 sporadic 99.2918

I didn't like any of the candidates

316.8703 31.33516
voter_category wt wt_total percentage 1 rarely/never 140.9815 1144.7680 12.31529 2 sporadic 36.2433

Because of where I live, my vote doesn't matter

316.8703 11.43790

voter_category wt wt_total percentage 1 rarely/never 369.9487 1144.7680 32.31648 2 sporadic 81.2421

No matter who wins, nothing will change for people like me

316.8703 25.63891
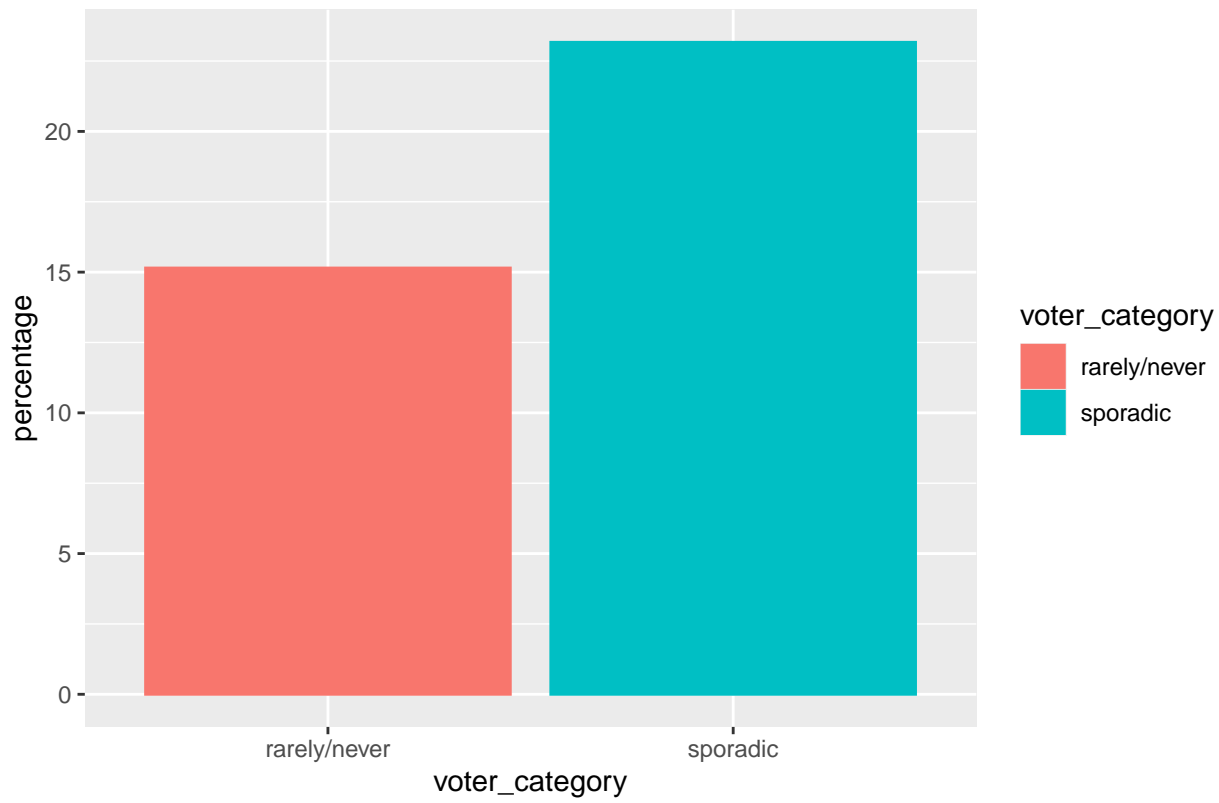voter_category wt wt_total percentage 1 rarely/never 263.0228 1144.7680 22.97608 2 sporadic 44.3060

## Our system is too broken to be fixed by voting
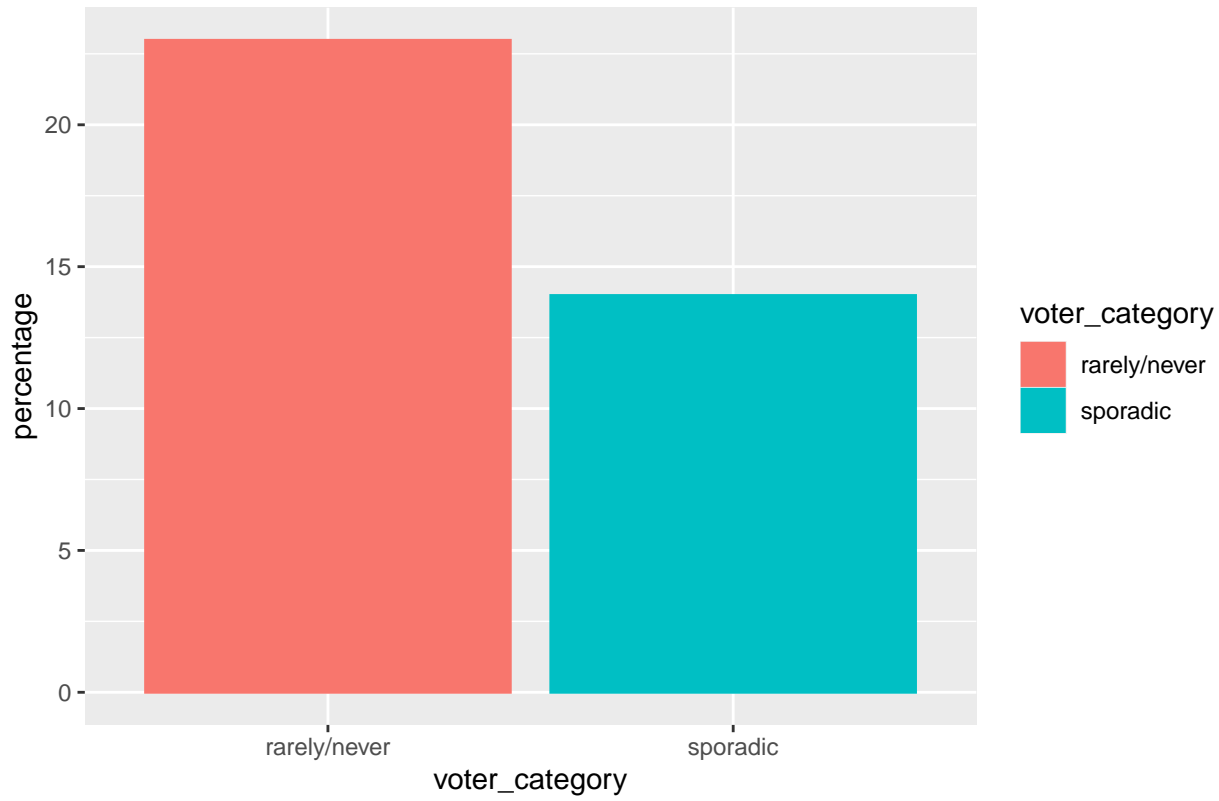
316.8703 13.98238

voter_category wt wt_total percentage 1 rarely/never 173.3809 1144.7680 15.14551 2 sporadic 73.4010

vote, but I didn't have time, couldn't get off work, something came up, or I forgot
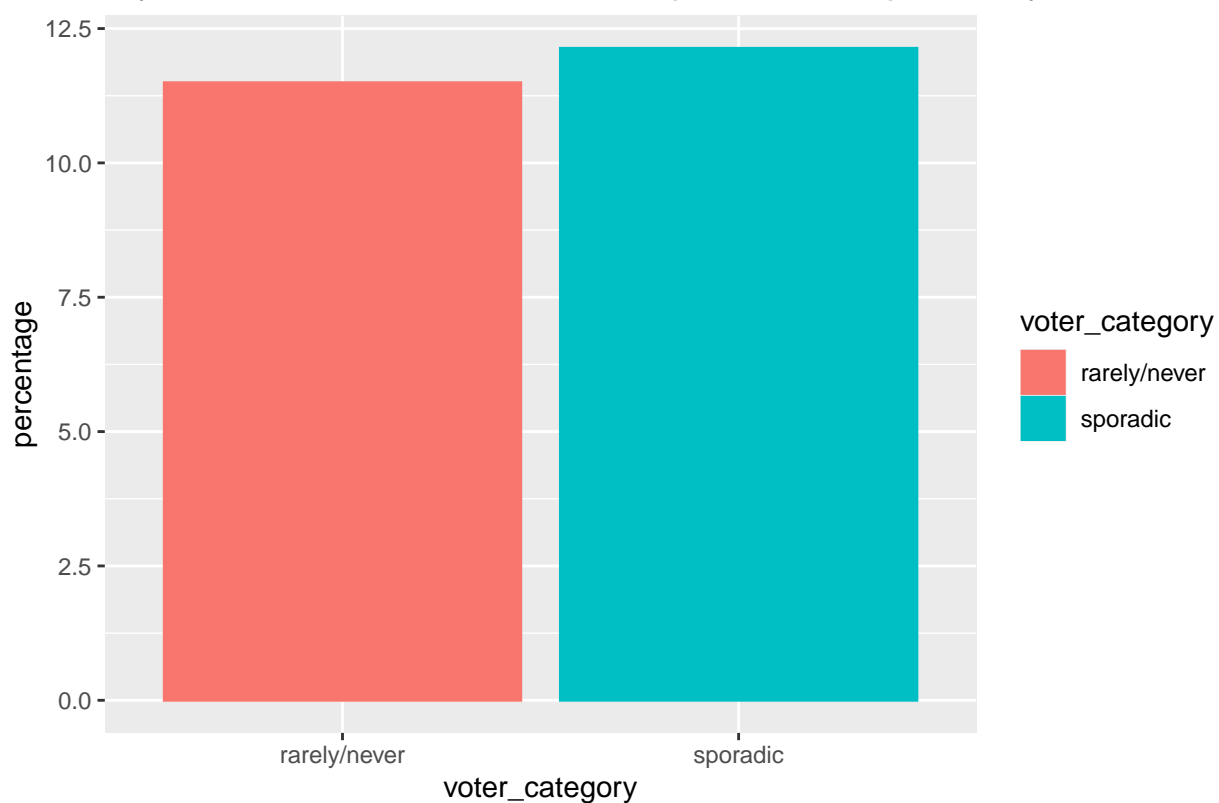

316.8703 23.16437
voter_category wt wt_total percentage 1 rarely/never 56.1102 1144.7680 4.901447 2 sporadic 5.0034 316.8703

I'm not sure if I can vote
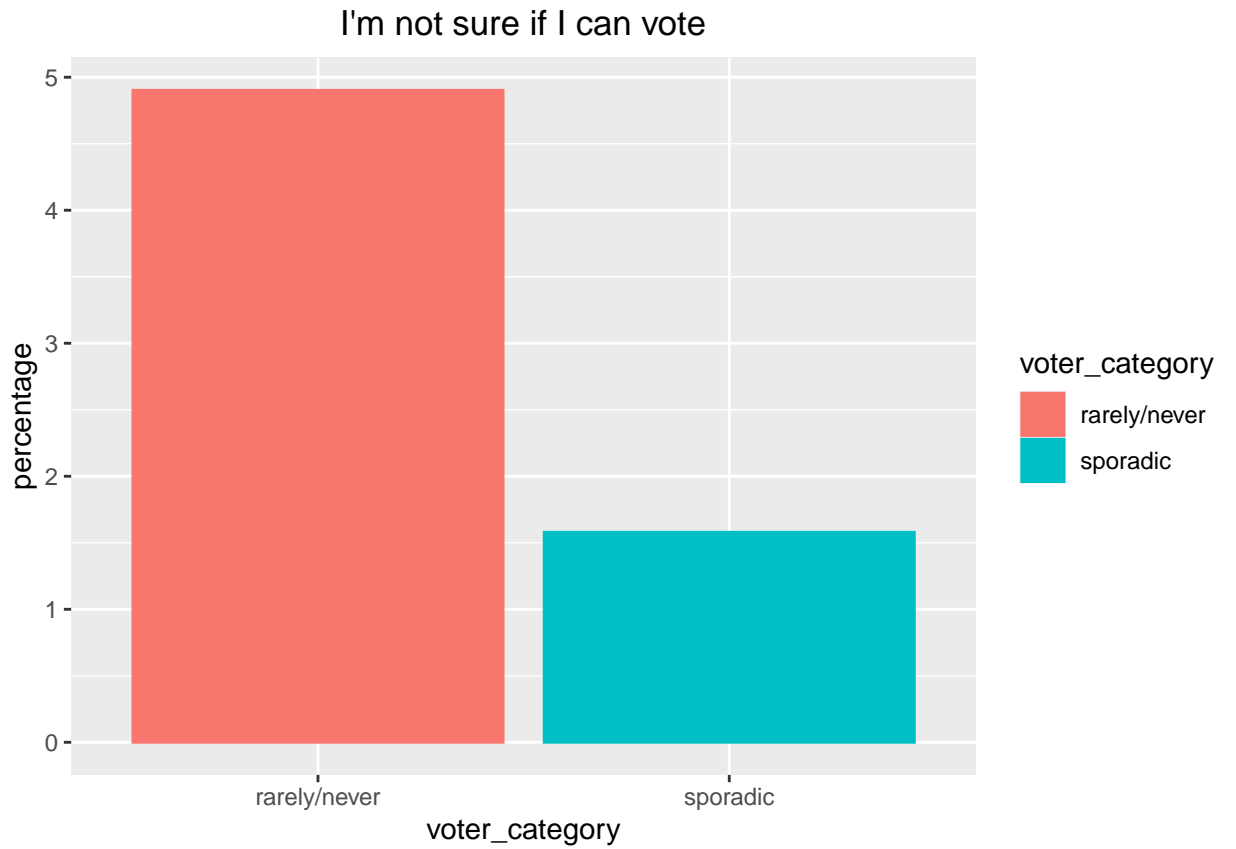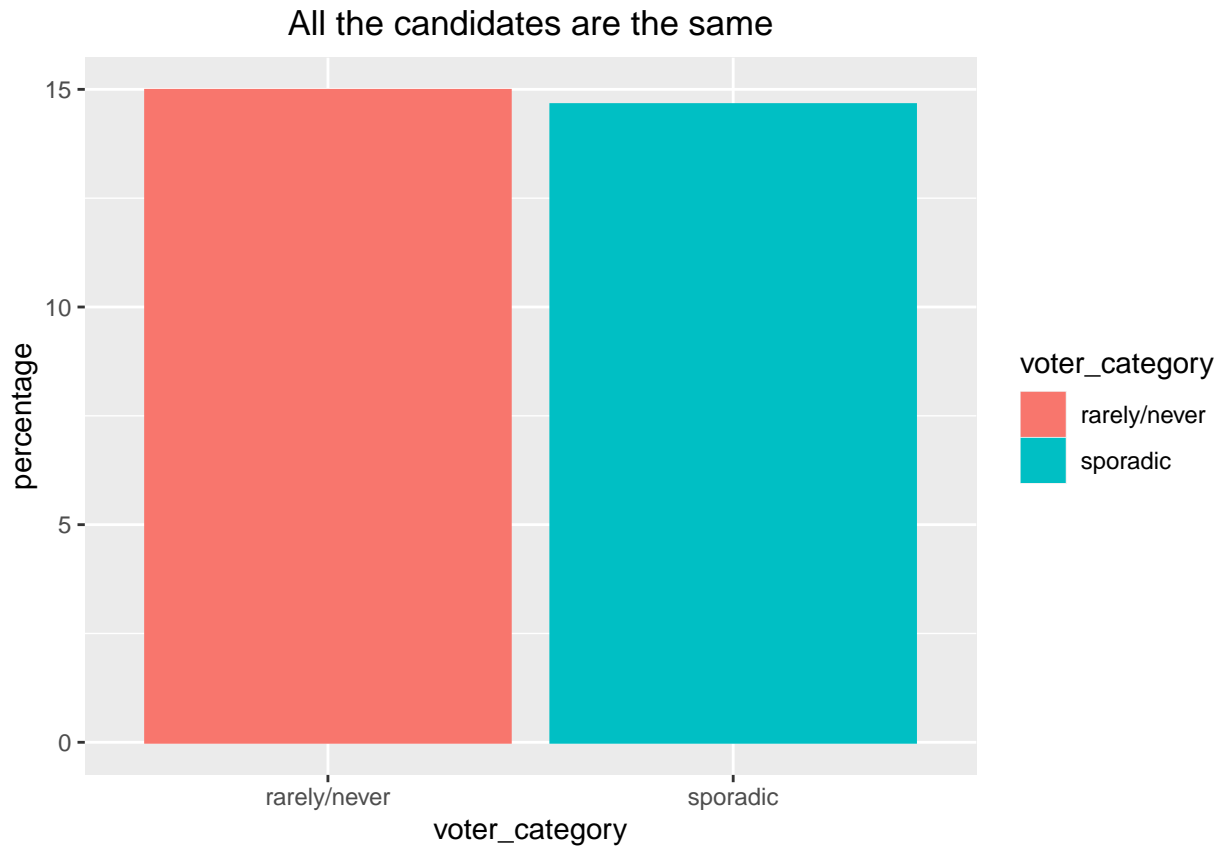

1.579006
voter_category wt wt_total percentage 1 rarely/never 131.5600 1144.7680 11.49228 2 sporadic 38.4431

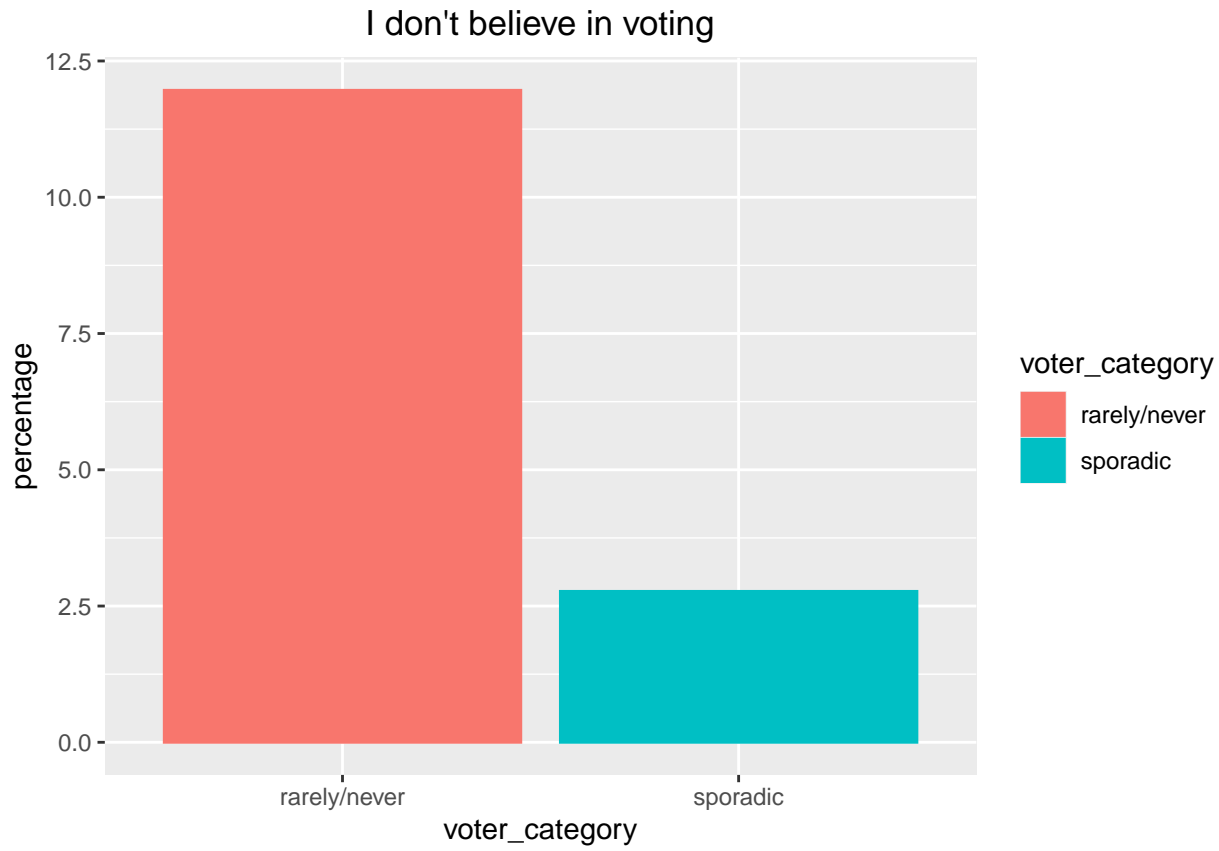## Nobody talks about the issues that are important to me personally

316.8703 12.13212

voter_category wt wt_total percentage 1 rarely/never 171.4582 1144.7680 14.97755 2 sporadic 46.4219
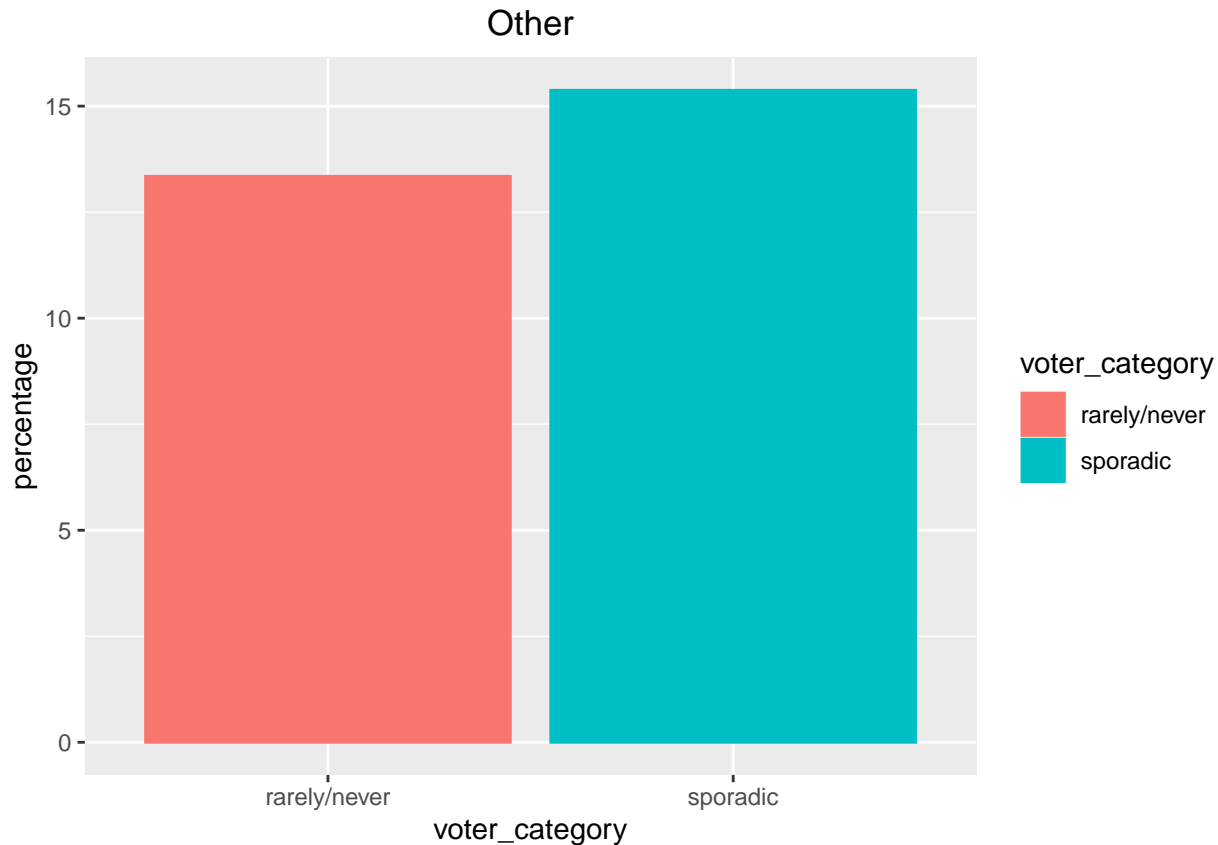
All the candidates are the same

316.8703 14.65013

voter_category wt wt_total percentage 1 rarely/never 136.9328 1144.7680 11.961620 2 sporadic 8.7735

I don't believe in voting

316.8703 2.768798

voter_category wt wt_total percentage 1 rarely/never 152.7566 1144.7680 13.34389 2 sporadic 48.7104

316.8703 15.37235

Conclusions

As evidenced by the data, voters who never, rarely, or only sporadically vote tend to feel as if nothing will change, regardless of the outcome of an election. Almost as often, they report not feeling any affinity toward any particular candidate. Further, many of them claim that the system is in disrepair and can't be fixed by voting.

While the above results are interesting (albeit disheartening!), I'd find it even more telling to further break down the most significant response by gender, race, income, or level of education; this might indicate which voters feel disenfranchised and why they feel that way, thereby guiding public policy decisions on possibly mitigation efforts.

People who feel as if nothing will change

Break the results of response #3 of question 29 out by age, education, race, gender, and income.

```
# Filter out voters who always vote; only select response #3 (people who feel as if nothing will change)
nonvoters_29_3 <-
  as_tibble(select(nonvoters, voter_category, weight, ppage, educ, race, gender, income_cat, Q29_3)) %>%
  filter(voter_category != "always" & !is.na(Q29_3))

#cut the age variable into categories
nonvoters_29_3 <- mutate(nonvoters_29_3, age_category = cut(ppage, c(18, 30, 40, 50, 60, 70, 80, 120)))
nonvoters_29_3[1:5,]     # Display first 5 observations
```

```
## # A tibble: 5 x 9
##   voter_category weight ppage educ     race  gender income_cat Q29_3 age_category
##   <chr>           <dbl> <int> <chr>    <chr> <chr>  <chr>      <int> <fct>
## 1 rarely/never    1.06     61 High s~  White Female $40-75k       -1 (60,70]
## 2 rarely/never    0.844    48 College  White Male   $40-75k       -1 (40,50]
## 3 rarely/never    1.02     46 Some c~  Othe~ Male   Less than~     1 (40,50]
## 4 rarely/never    0.984    50 High s~  Black Female Less than~     1 (40,50]
## 5 rarely/never    0.464    47 Some c~  Hisp~ Male   $40-75k       -1 (40,50]
```

```r
# Create array of fields we're interestedin
fields <- c("age_category", "educ", "race", "gender", "income_cat")

# Iterate over each answer in Question 29
for(i in 1:5) {

  # Create tibble with total weighted counts of voters
  totals <- select(nonvoters_29_3, weight, fields[i]) %>%
    filter(!is.na(fields[i])) %>%
    group_by_at(fields[i]) %>%
    summarize(wt_total = sum(weight))

  # Create tibble with weighted count of voters who answered that this was
  # an important reason why they didn't vote
  grouped <- select(nonvoters_29_3, weight, Q29_3, fields[i]) %>%
    filter(!is.na(fields[i]) & Q29_3 == 1) %>%
    group_by_at(fields[i]) %>%
    summarize(wt = sum(weight))

  # Merge the two tibbles
  subset <- merge(grouped, totals, by = fields[i])

  # Create a new variable for percentage and print the new tibble
  subset <- mutate(subset, percentage = wt * 100 / wt_total)
  print(subset)

  # Plot the bar chart
  print(ggplot(data = subset, mapping = aes(x = fields[i], y = percentage)) +
    geom_bar(position = "dodge", stat = "identity", mapping = aes(color = get(fields[i]), fill = get(fi
    ggtitle(fields[i]) +
    theme(plot.title = element_text(hjust = 0.5)))

}
```
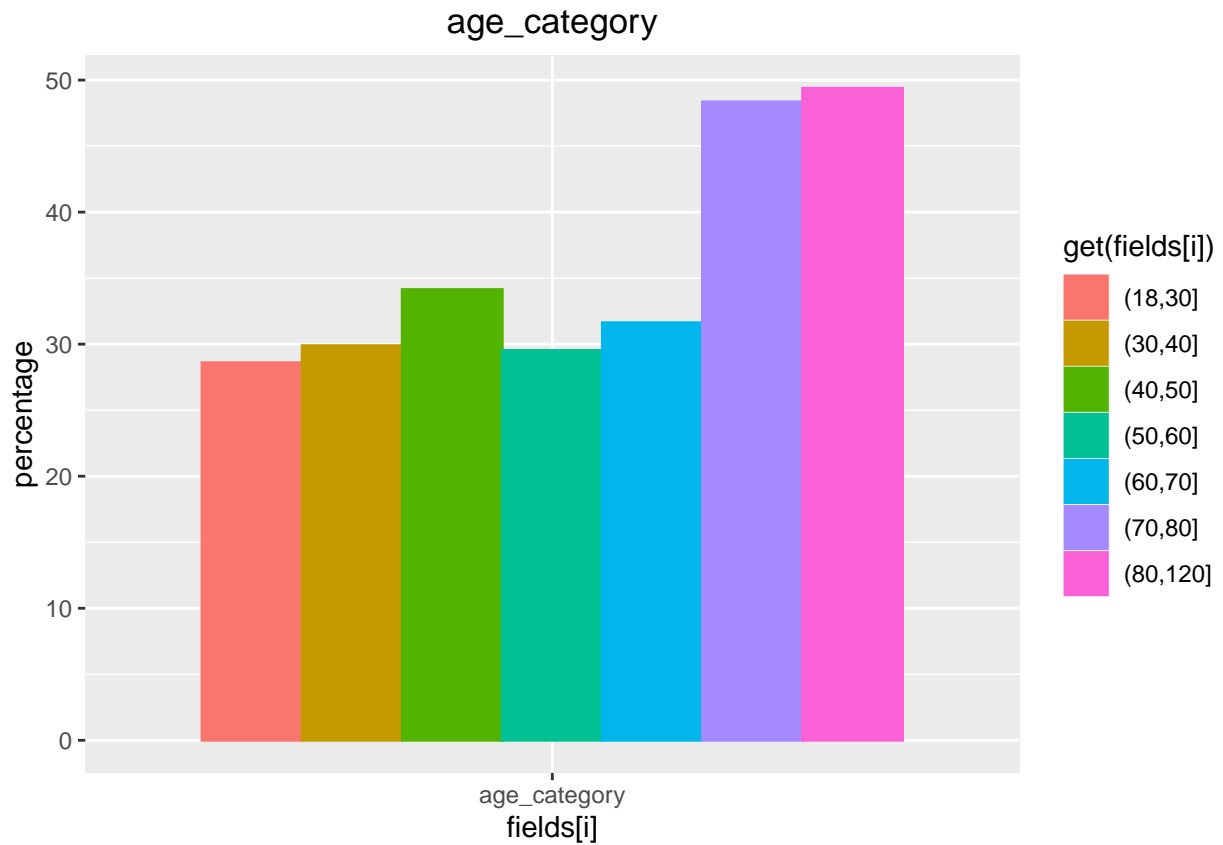
```
##   age_category       wt wt_total percentage
## 1      (18,30] 118.0589 412.8643   28.59509
## 2      (30,40] 114.7357 383.8970   29.88711
## 3      (40,50]  79.4737 232.7961   34.13876
## 4      (50,60]  77.5314 262.5861   29.52609
## 5      (60,70]  39.0753 123.5671   31.62274
## 6      (70,80]  17.0409  35.2443   48.35080
## 7     (80,120]   5.2749  10.6834   49.37473
```

```
##                        educ        wt wt_total percentage
## 1               College  58.8725 221.6404   26.56217
## 2 High school or less 276.7591 877.9354   31.52386
## 3         Some college 115.5592 362.0625   31.91692
```
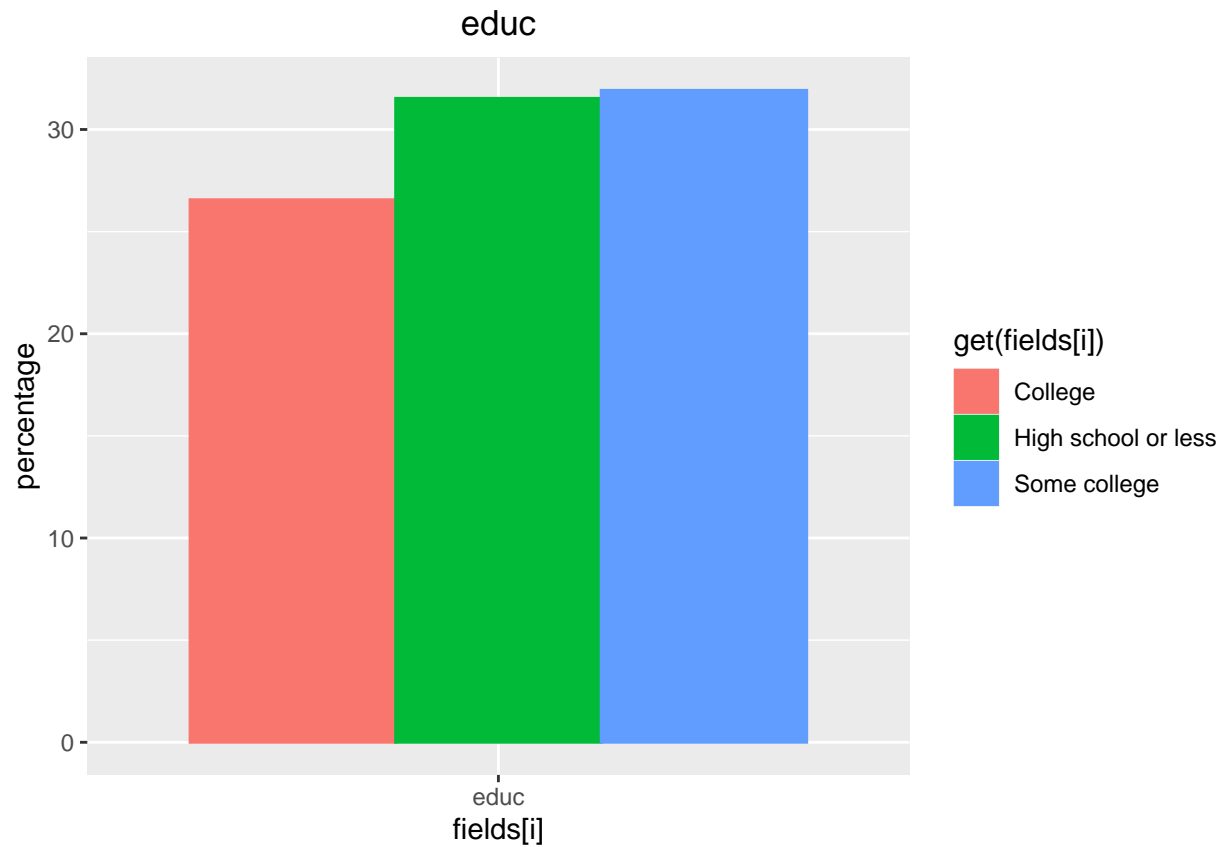
```
##          race        wt wt_total percentage
## 1       Black  68.8955 213.7141   32.23723
## 2    Hispanic  66.5566 309.7612   21.48642
## 3 Other/Mixed  51.5447 124.0881   41.53879
## 4       White 264.1940 814.0749   32.45328
```
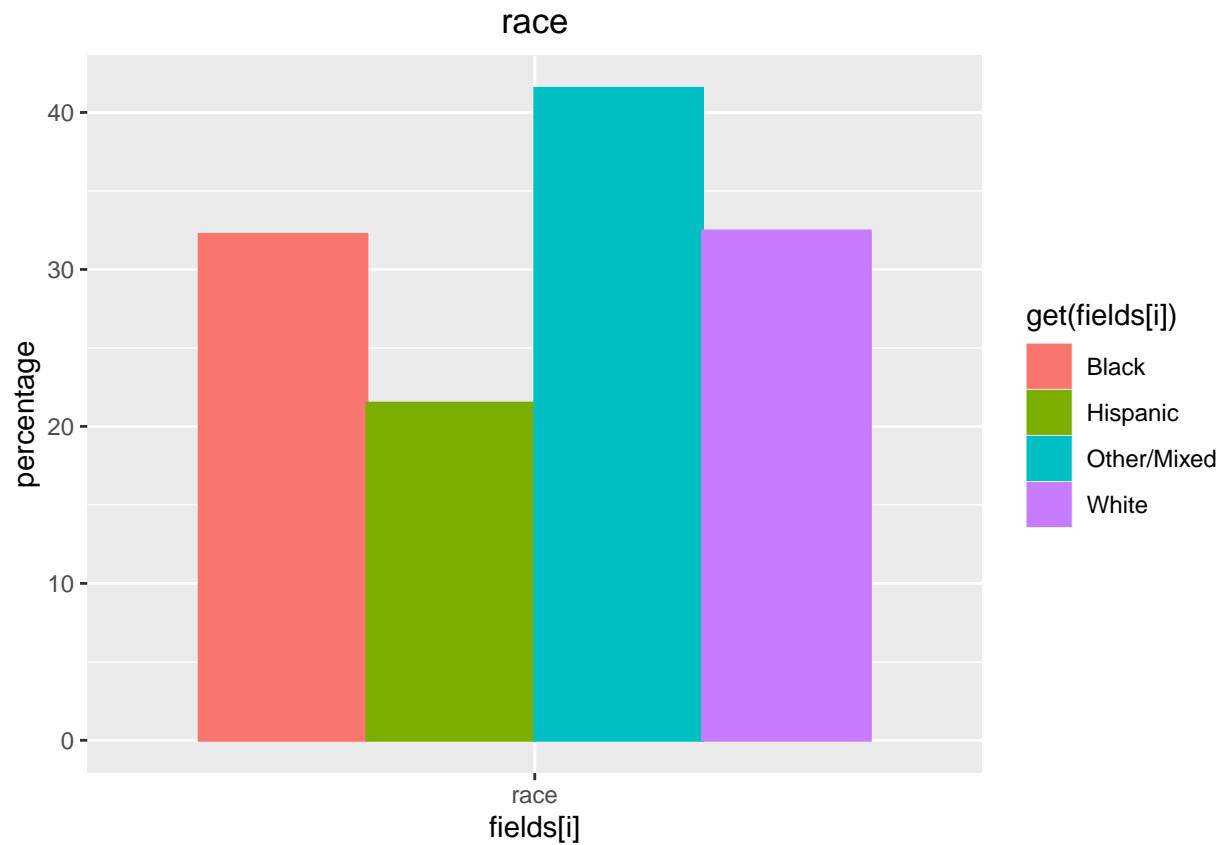
```
##    gender        wt wt_total percentage
## 1 Female 239.2747 787.7183   30.37567
## 2   Male 211.9161 673.9200   31.44529
```

gender
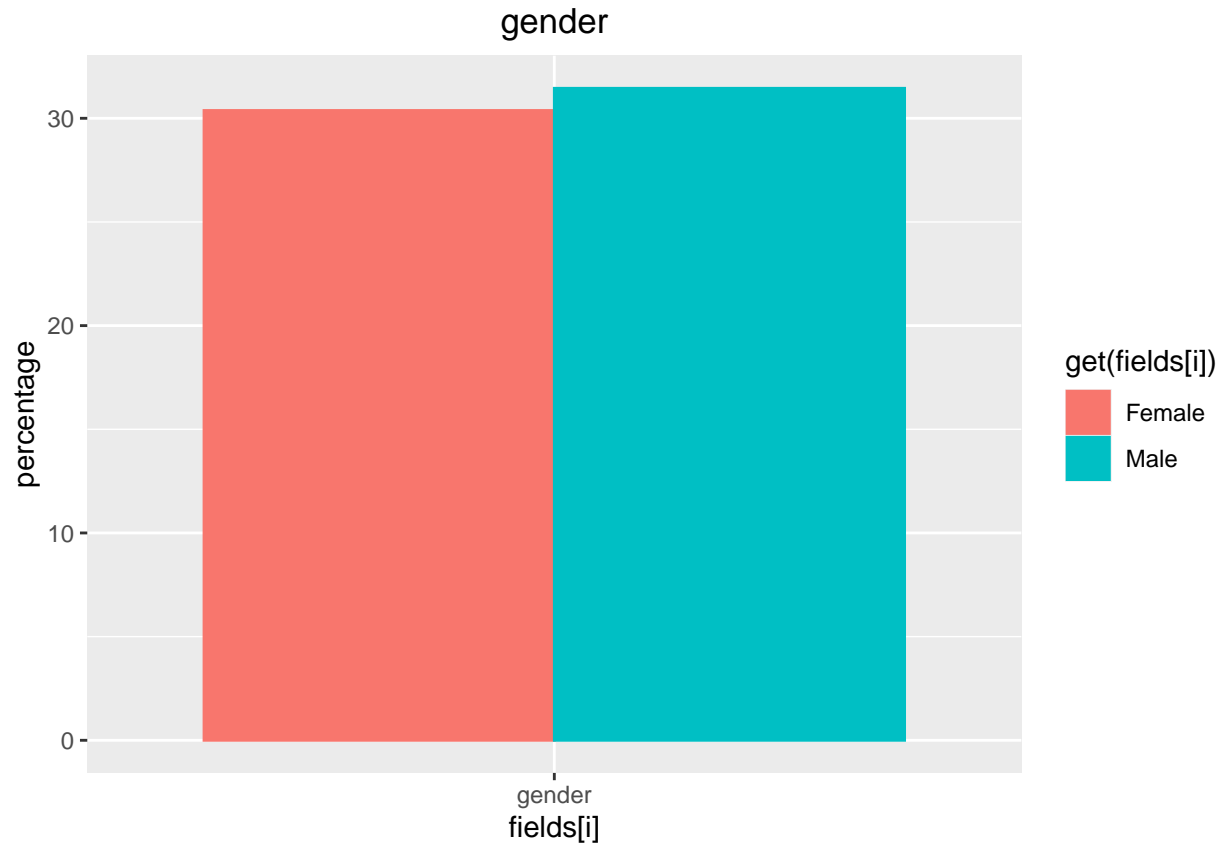
```
##       income_cat       wt wt_total percentage
## 1  $125k or more  50.2482 207.8991   24.16951
## 2        $40-75k 106.0588 347.7087   30.50220
## 3        $75-125k  85.4134 278.3564   30.68491
## 4 Less than $40k 209.4704 627.6741   33.37248
```

income_cat
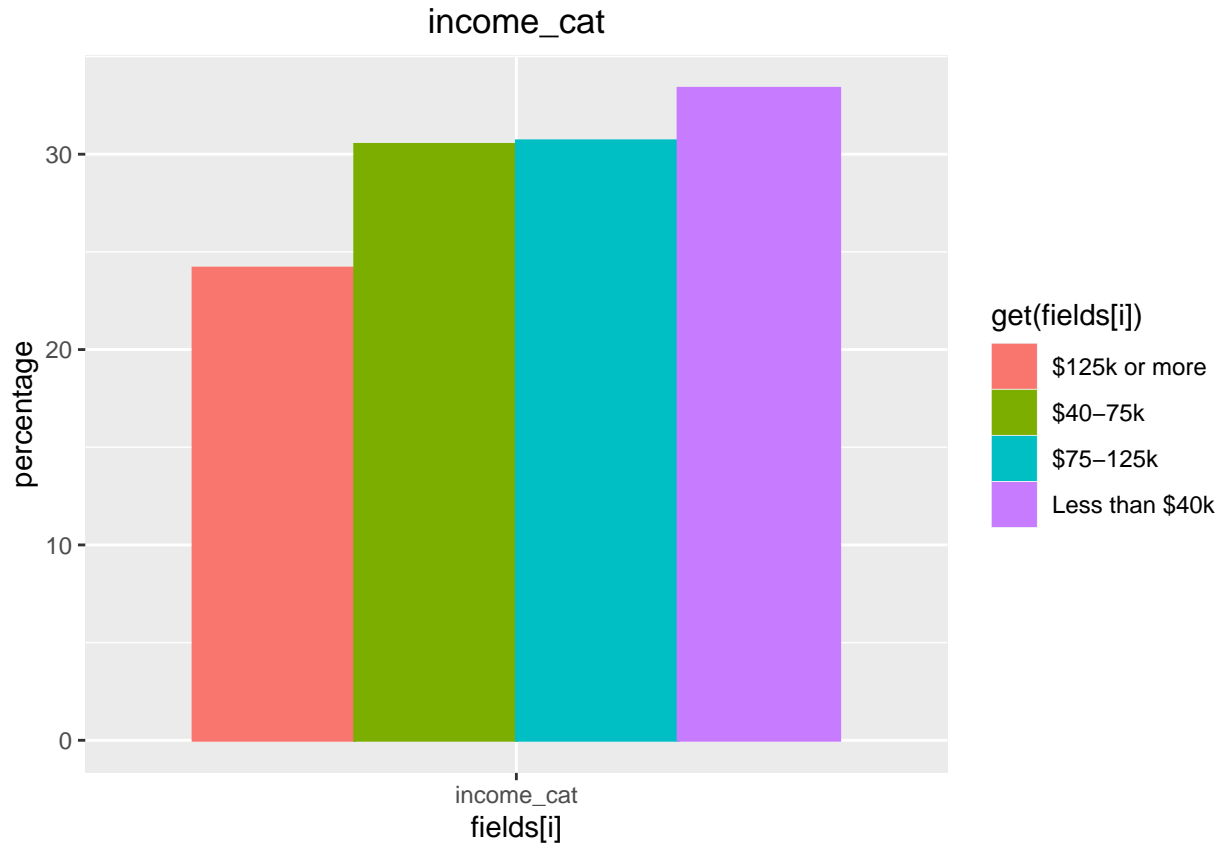
Further Conclusions

The additional analysis for Response #3 illustrates which demographics feel their vote doesn't matter. Generally speaking, there was little variation across demographics, with one notable exception: While older voters (70 and older) tend to vote more often than younger ones, they are also the ones who feel their vote doesn't matter. While this inverse relationship is somewhat surprising, the fact that there was little variation among the other demographics is perhaps even more surprising. For example, my expectation was that minorities and females would comprise a greater percentage of voters who feel disenfranchised, given the demographic of most elected politicians. Instead, there was little variation among races and genders, which leads me to believe that people, in general, feel their vote doesn't matter, regardless of demographic. In a time when there seems to be little to unify us, perhaps our general sense of disillusionment is one way (albeit a depressing one!) in which we can consider ourselves united.