Final-explorations

This document is a narrative of explorations of the Kaiser dataset, prior to the final paper.

Note: the data file `31118130.csv` is a symlink to the original file in the "../3 + 4" directory in this repository.

```
raw_data <- read_csv("31118130.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_character(),
##   end_state1 = col_double(),
##   acarot = col_number(),
##   endq1 = col_double(),
##   q2rot = col_logical(),
##   endq2 = col_double(),
##   endq10 = col_double(),
##   q17rot = col_number(),
##   endq22 = col_double(),
##   partyrot = col_number(),
##   length = col_double(),
##   reprostat = col_logical(),
##   qctr = col_double(),
##   xctr = col_double(),
##   wt1 = col_double(),
##   weight = col_double(),
##   standwt = col_double(),
##   weight_ssrs = col_double(),
##   cdc2013 = col_double()
## )
## i Use `spec()` for the full column specifications.
```

```
raw_data
```

```
## # A tibble: 1,676 x 168
##    id    d1    end_state1 hispanic race  racethn nativity racethn2 acarot aca
##    <chr> <chr>      <dbl> <chr>    <chr> <chr>   <chr>    <chr>     <dbl> <chr>
##  1 0000~ Male         789 No       White WHITE,~ <NA>     WHITE, ~     41 Some~
##  2 0000~ Male        1148 No       White WHITE,~ <NA>     WHITE, ~     14 Some~
##  3 0000~ Male         901 Yes      White HISPAN~ U.S.     HISPANI~     14 Very~
##  4 0000~ Male         685 No       White WHITE,~ <NA>     WHITE, ~     41 Some~
##  5 0000~ Male         656 No       White WHITE,~ <NA>     WHITE, ~     41 Some~
##  6 0000~ Male         704 No       Othe~ OTHER,~ <NA>     OTHER, ~     41 Very~
##  7 0000~ Male         749 No       White WHITE,~ <NA>     WHITE, ~     41 Very~
##  8 0000~ Fema~       1001 Yes      Othe~ HISPAN~ Another~ HISPANI~     14 Some~
##  9 0000~ Male         737 No       White WHITE,~ <NA>     WHITE, ~     41 Very~
## 10 0000~ Male         933 No       White WHITE,~ <NA>     WHITE, ~     14 Very~
## # ... with 1,666 more rows, and 158 more variables: q1rot <chr>, q1 <chr>,
## #   endq1 <dbl>, q2rot <lgl>, q2rot2 <chr>, q2a <chr>, q2b <chr>, q2c <chr>,
```

```
## #   q2d <chr>, q2e <chr>, q2f <chr>, q2g <chr>, endq2 <dbl>, q3 <chr>,
## #   q27 <chr>, q4rot <chr>, q4 <chr>, q5rot <chr>, q5 <chr>, q6 <chr>,
## #   q7rot <chr>, q7rot2 <chr>, q7a <chr>, q7b <chr>, q8 <chr>, q9rot <chr>,
## #   q9 <chr>, q10rot <chr>, q10a <chr>, q10b <chr>, q10c <chr>, endq10 <dbl>,
## #   q16 <chr>, q17rot <dbl>, q17 <chr>, q18rot <chr>, q18a <chr>, q18b <chr>,
## #   q18c <chr>, q18e <chr>, q18f <chr>, q18g <chr>, q18h <chr>, q18i <chr>,
## #   q18j <chr>, q11rot <chr>, q11 <chr>, q12rot <chr>, q12a <chr>, q12b <chr>,
## #   q12c <chr>, q12d <chr>, q12e <chr>, q12f <chr>, q12g <chr>, q12h <chr>,
## #   q12i <chr>, q13rot <chr>, q13 <chr>, q14rot <chr>, q14 <chr>, q15 <chr>,
## #   q19_q20rot <chr>, q19 <chr>, q20 <chr>, q21 <chr>, q22 <chr>, endq22 <dbl>,
## #   age <chr>, age2 <chr>, recage <chr>, recage2 <chr>, recage3 <chr>,
## #   recage4 <chr>, recage5 <chr>, child <chr>, marital <chr>, rvote <chr>,
## #   voted <chr>, voted2rot <chr>, voted2 <chr>, voted2ot <chr>,
## #   inclosstotal <chr>, employ <chr>, recemploy <chr>, essential <chr>,
## #   hcworker2 <chr>, hcworker3 <chr>, coverage <chr>, agecov <chr>,
## #   covtype <chr>, agecovtype <chr>, covselfother <chr>, q23 <chr>,
## #   q23ot1 <chr>, q23ot2 <chr>, q23ot3 <chr>, q23ot4 <chr>, rsex <chr>,
## #   gendervar <chr>, ...
```

Opening question: among survey respondents who respond "refuse to answer" to the question of whom they voted for for president, is there a detectable bias? Are liberals or conservatives, Biden or Trump or "other" voters, more likely to do this? Is there a way we can tell?

We can't know for sure, but the question is, can we make a data-supported argument, based on the data in this survey, that supports or rejects this idea?

First, let's get a breakdown of the responses to the question:

```
table(raw_data$voted2)
```

```
##
##   Don't know Donald Trump    Joe Biden      Refused Someone else
##            9          417          733          111           32
```

```
# hat tip to https://stackoverflow.com/a/45386128/13603796
table(raw_data$voted2) %>% prop.table() %>% `*`(100) %>% round(2)
```

```
##
##   Don't know Donald Trump    Joe Biden      Refused Someone else
##         0.69        32.03        56.30         8.53         2.46
```

So 8.5% of respondents refused to answer the question. That's a significant amount, 1 in 12.

First, how do they lean? There are many possible variables we could look at; let's start with `ideology`:

```
table(raw_data$ideology)
```

```
##
## Conservative   Don't Know      Liberal     Moderate      Refused
##          527           68          424          617           40
```

Let's start to break this down.

```
ideology_by_refused_voted2 <- raw_data %>%
  select(ideology, voted2) %>%
  filter(voted2 == "Refused")
table(ideology_by_refused_voted2$ideology)
```

```
##
## Conservative   Don't Know      Liberal     Moderate      Refused
```

```
##          38          6          16          37          14
```