# POLS 7012 Final Exam 2023 (Answer Key)

Due December 13, 2023 @ Noon (12pm)

I was struck by a rather extraordinary claim that I read in this recent article. According to the latest wave of the American Values Survey, approximately 25% of Americans agree with the following statement:

> The government, media, and financial worlds in the U.S. are controlled by a group of Satan-worshipping pedophiles who run a global child sex trafficking operation.

Perhaps you recognize this as a core tenet of the QAnon conspiracy theory. But do over 80 million Americans actually believe this? For this year's final exam, I'd like us to explore that claim more deeply. Please write and submit a PDF report rendered from an R script (optionally: a Quarto document) that conducts the following analyses:[1]

1. The 2023 AVS survey is not yet publicly available, but the surveys from 2021 and 2022 are, both of which ask questions about QAnon. To download the data, go to this link and search for "American Values Survey" (make sure to select the checkbox "Project Only", otherwise it'll give you summaries rather than the raw data files). Download the 2021 and 2022 datasets and codebooks, and load the data into R.

```
library(tidyverse)
library(here)
library(haven)

# note: my .dta files are in the final-2023 project folder,
# but yours might be in a different place
avs_2021 <- read_dta(here('PRRI AVS 2021 Sept 30 2021.dta'))

avs_2022 <- read_dta(here('AVS 2022_Client Final Weighted.dta'))
```

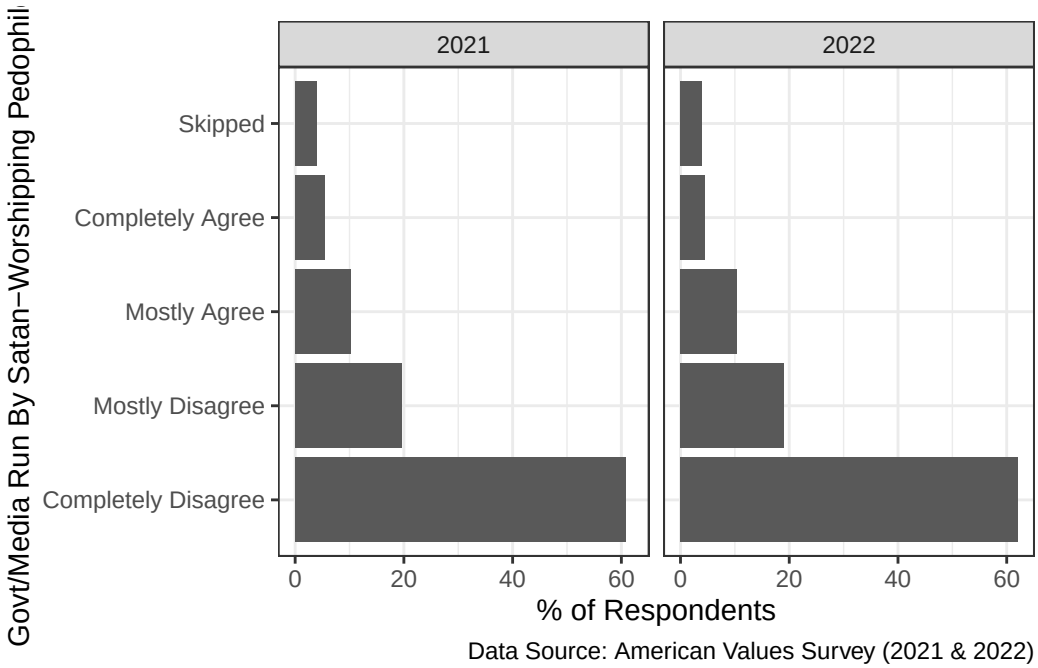
2. What is the distribution of responses for that question about the Satan-worshipping pedophiles? Create two bar charts, one for 2021 and one for 2022, properly labeled and ordered. In all the analyses that follow do *not* drop the respondents who skipped the question. Let's leave them in as a separate category.

[1] Unlike the problem sets, this exam must be completed individually, without help from others.

```r
# let's start by cleaning up the data and creating a single dataframe
# with responses from both waves

avs_2021 <- avs_2021 |>
  # label
  mutate(qanon = case_when(Q34_b == -1 ~ 'Skipped',
                           Q34_b == 1 ~ 'Completely Agree',
                           Q34_b == 2 ~ 'Mostly Agree',
                           Q34_b == 3 ~ 'Mostly Disagree',
                           Q34_b == 4 ~ 'Completely Disagree')) |>
  # Make it an ordinal variable
  mutate(qanon = factor(qanon,
                        levels = c('Completely Disagree', 'Mostly Disagree',
                                   'Mostly Agree', 'Completely Agree',
                                   'Skipped')))


avs_2022 <- avs_2022 |>
  # label
  mutate(qanon = case_when(Q20_3 == -1 ~ 'Skipped',
                           Q20_3 == 1 ~ 'Completely Agree',
                           Q20_3 == 2 ~ 'Mostly Agree',
                           Q20_3 == 3 ~ 'Mostly Disagree',
                           Q20_3 == 4 ~ 'Completely Disagree')) |>
  # Make it an ordinal variable
  mutate(qanon = factor(qanon,
                        levels = c('Completely Disagree', 'Mostly Disagree',
                                   'Mostly Agree', 'Completely Agree',
                                   'Skipped')))


d <- bind_rows(
  # add a year variable and keep selected variables
  # that we need for the analysis
  avs_2021 |>
    mutate(year = 2021) |>
    select(year, Respid, kp_weight, qanon,
           ppgender, xparty4, ppethm, ppage, ppeducat),

  avs_2022 |>
    mutate(year = 2022) |>
```

```
    select(year, Respid, kp_weight, qanon,
           ppgender, xparty4, ppethm, ppage, ppeducat)
)
```

```
d |>
  group_by(year) |>
  count(qanon) |>
  mutate(pct = n / sum(n) * 100) |>
  ggplot(mapping = aes(x=pct, y=qanon)) +
  geom_col() +
  labs(
    x = '% of Respondents',
    y = 'Govt/Media Run By Satan-Worshipping Pedophiles',
    caption = 'Data Source: American Values Survey (2021 & 2022)'
  ) +
  theme_bw() +
  facet_wrap(~year)
```




3. I'm particularly interested in the group of respondents who "Completely Agree" with the statement. Let's call this group the **QAnon Believers**. What percent of survey respondents are QAnon Believers in each wave?

```r
d <- d |>
  mutate(qanon_believer = as.numeric(qanon == 'Completely Agree'))

d |>
  group_by(year) |>
  summarize(pct_qanon_believers = mean(qanon_believer) * 100,
            num = sum(qanon_believer))
```

```
# A tibble: 2 x 3
   year pct_qanon_believers   num
  <dbl>               <dbl> <dbl>
1  2021                5.34   134
2  2022                4.56   115
```

*It looks like the share of respondents who "Completely Agree" with the statement is small, approximately 5% of each sample. Most of the work is happening in the "Mostly Agree" camp. I am genuinely curious which part of the statement they "mostly" agree with, and which part they disagree with. But I wouldn't characterize anyone who says "Mostly Agree" to that statement as a QAnon believer.*

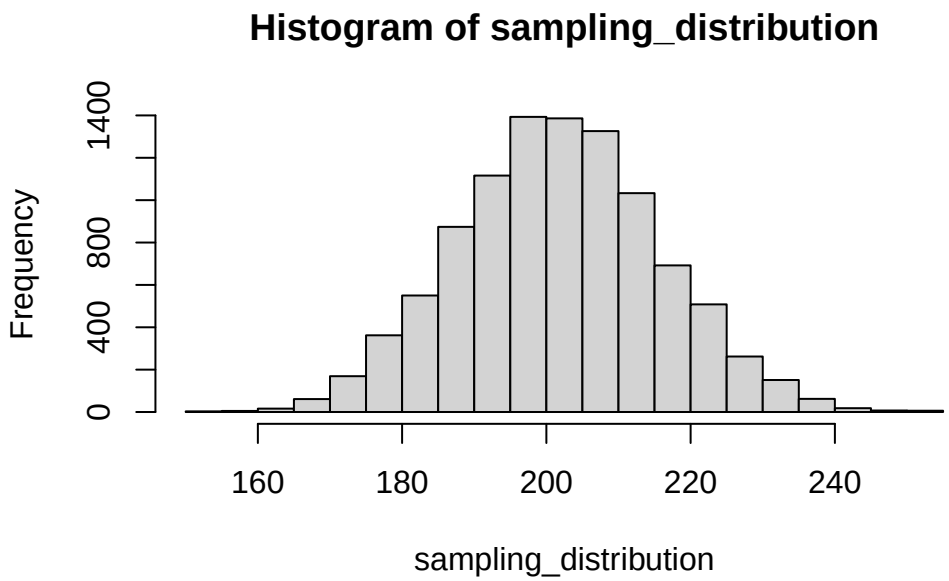
4. Scott Alexander argues that 4% of Americans will say pretty much anything on a survey. That's the percent of survey respondents who report believing that "shape-shifting reptilian people control our world by taking on human form and gaining power" (source). So my null hypothesis is that roughly 4% of Americans, through some combination of joking around and/or accidentally pressing the wrong button, will express complete agreement with any conspiracy theory. Is our result here consistent with that null hypothesis? Simulate a sampling distribution and compute a p-value.

```r
# create a population where the null hypothesis is true
null_america <- sample(c(0,1),
                       size = 1e6,
                       prob = c(0.96, 0.04),
                       replace = TRUE)

# here's a function to draw a sample from Null America and
# count how many respondents are QAnon Believers
get_sample_statistic <- function(n){

  this_sample <- sample(null_america, size = n, replace = FALSE)

  return(sum(this_sample))
```

```
}

get_sample_statistic(n = nrow(d))
```

```
[1] 204
```

```
# now generate 10,000 random surveys from Null America
num_samples <- 1e4

sampling_distribution <- replicate(num_samples,
                                   get_sample_statistic(n = nrow(d)))


hist(sampling_distribution)
```

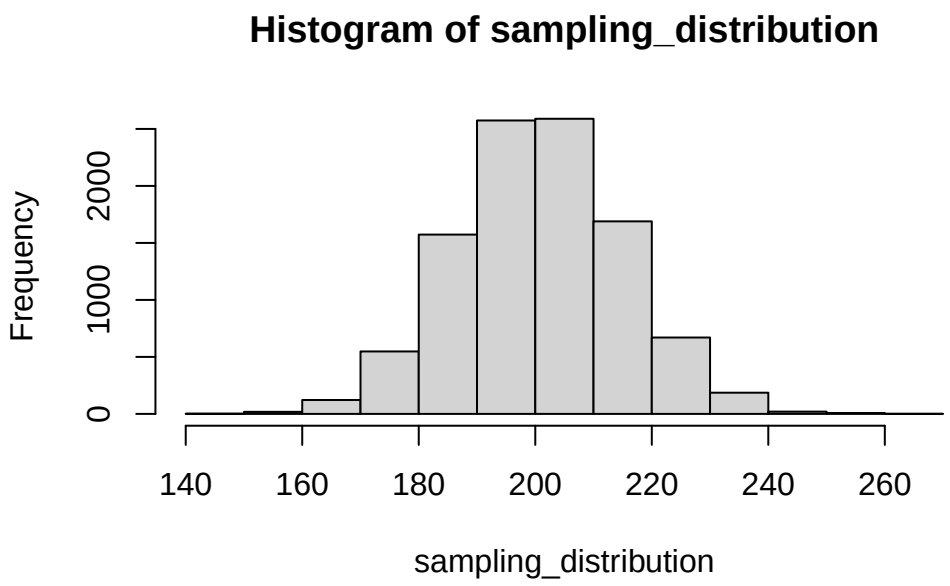



```
sum(sampling_distribution >= 249) / num_samples
```

```
[1] 7e-04
```

*There is more support for QAnon than we would expect to observe due to sampling error alone from that null hypothesis. The Satan-worshipping pedophiles conspiracy theory has a statistically significantly larger number of adherents than the Lizardpeople conspiracy, though not by a large margin (it's a difference of roughly 1 percentage point).*

```
# note that the binomial distribution function does the same thing,
# but faster
sampling_distribution <- rbinom(n = num_samples, size = nrow(d), prob = 0.04)

hist(sampling_distribution)
```




```
sum(sampling_distribution >= 249) / num_samples
```

```
[1] 9e-04
```

5. Since QAnon is an Internet-based conspiracy theory, maybe it is more widespread among younger respondents. What is the average age of QAnon Believers? Is that higher or lower than the average age of everyone else in the survey? Perform a null hypothesis test and report the associated p-value.

```r
d |>
  group_by(qanon_believer) |>
  summarize(mean_age = mean(ppage))
```

```
# A tibble: 2 x 2
  qanon_believer mean_age
           <dbl>    <dbl>
1              0     52.7
2              1     51.6
```

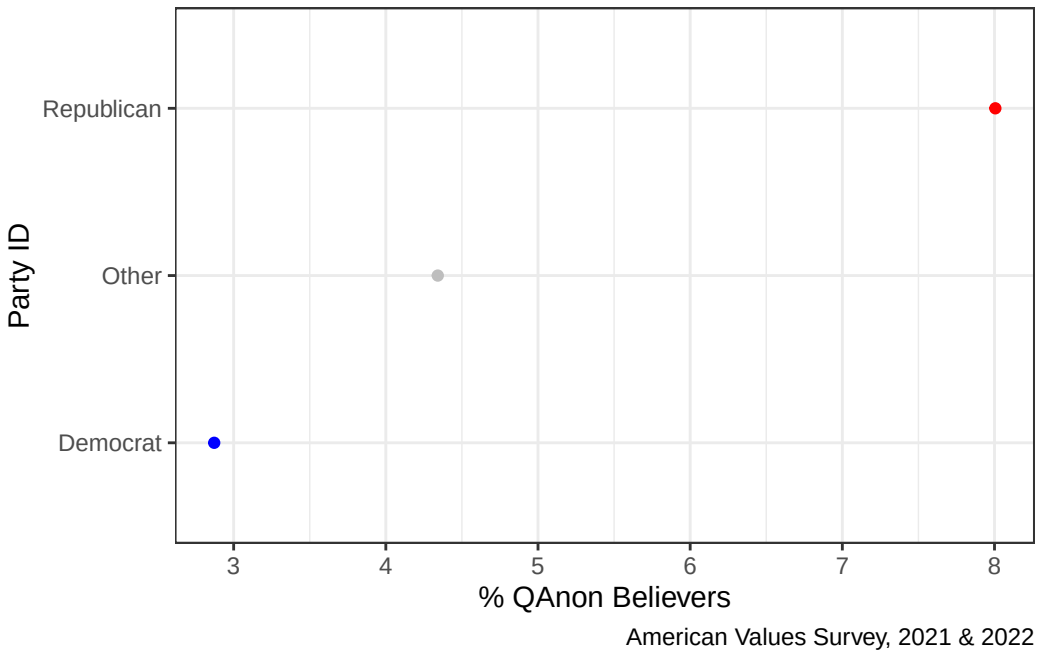
```r
age_model <- lm(ppage ~ qanon_believer, data = d)

summary(age_model)
```

```
Call:
lm(formula = ppage ~ qanon_believer, data = d)

Residuals:
    Min      1Q  Median      3Q     Max
-34.735 -13.735   2.265  14.265  42.402

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     52.7348     0.2527 208.693   <2e-16 ***
qanon_believer  -1.1364     1.1358  -1.001    0.317
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.47 on 5029 degrees of freedom
Multiple R-squared:  0.000199,  Adjusted R-squared:  2.117e-07
F-statistic: 1.001 on 1 and 5029 DF,  p-value: 0.3171
```

*There is no significant difference in average age between the QAnon believers and the rest of the survey respondents.*

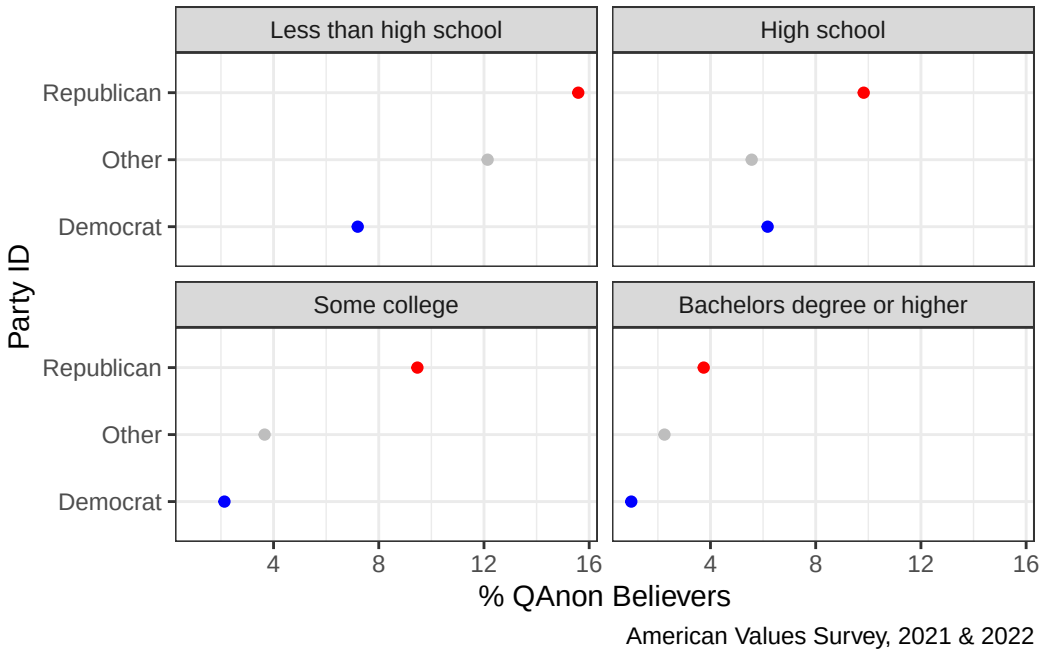
5. Next, let's look at the relationship between political party identification and QAnon Believers. Replicate the following chart, using respondents from both 2021 and 2022 waves of the survey.

```r
d |>
  # relabel
  mutate(partyid = case_when(xparty4 == 1 ~ 'Republican',
                             xparty4 == 2 ~ 'Democrat',
                             TRUE ~ 'Other')) |>
  # convert to ordinal variable
  mutate(partyid = factor(partyid, levels = c('Democrat',
                                              'Other', 'Republican'))) |>
  mutate(qanon_believer = if_else(qanon_believer == 1, 'QAnon Believer',
                                  'Not QAnon Believer')) |>
  group_by(partyid) |>
  count(qanon_believer) |>
  mutate(pct = n / sum(n) * 100) |>
  filter(qanon_believer == 'QAnon Believer') |>
  ggplot(mapping = aes(y=partyid, x=pct, color=partyid)) +
  geom_point() +
  scale_color_manual(values = c('blue', 'gray', 'red')) +
  theme_bw() +
  labs(y = 'Party ID', x = '% QAnon Believers', fill = 'Party ID',
       caption = 'American Values Survey, 2021 & 2022') +
  theme(legend.position = 'none')
```



6. Recreate the chart from question 5, but condition on education category. Is there more partisan polarization on this statement for some education groups than others?

```
d |>
  # relabel
  mutate(partyid = case_when(xparty4 == 1 ~ 'Republican',
                             xparty4 == 2 ~ 'Democrat',
                             TRUE ~ 'Other')) |>
  mutate(ppeducat = haven::as_factor(ppeducat)) |>
  # convert to ordinal variable
  mutate(partyid = factor(partyid, levels = c('Democrat', 'Other', 'Republican'))) |>
  mutate(qanon_believer = if_else(qanon_believer == 1, 'QAnon Believer', 'Not QAnon Believ
  group_by(partyid, ppeducat) |>
  count(qanon_believer) |>
  mutate(pct = n / sum(n) * 100) |>
  filter(qanon_believer == 'QAnon Believer') |>
  ggplot(mapping = aes(y=partyid, x=pct, color=partyid)) +
  geom_point() +
  scale_color_manual(values = c('blue', 'gray', 'red')) +
  theme_bw() +
  labs(y = 'Party ID', x = '% QAnon Believers', fill = 'Party ID',
       caption = 'American Values Survey, 2021 & 2022') +
  theme(legend.position = 'none') +
  facet_wrap(~ppeducat)
```




*It appears that the least politically polarized group is the group with Bachelor's Degrees or Above; there is very little difference by party in this group. The most polarized group is the Less Than High School Group, where respondents who identify as Republicans are twice as likely to be QAnon Believers than those who identify as Democrats.*

7. **OPTIONAL BONUS FUN.** The results we got in Question 2 don't exactly match the results reported in the codebooks, do they? That's because the codebook results are **poststratified** (also known as survey-weighted estimates). Basically, the survey respondents are not a perfectly representative sample of the United States population; there are some demographics that are underrepresented among the survey respondents and others that are overrepresented. To make inferences about the American population as a whole, we need to take a weighted average of the survey responses, giving more weight to the underrepresented groups and less weight to the overrepresented groups. The `kp_weight` variable contains those weights. (A respondent with `kp_weight = 3` is from a demographic group that's three times more common in the US population than in the survey. A respondent with `kp_weight = 0.5` is from a demographic group that's half as common in the US population as it is in the survey). Use those weights to replicate the estimates from the 2021 and 2022 codebooks.

```
avs_2021 |>
  count(qanon, wt = kp_weight) |>
  mutate(pct = n/sum(n)*100)
```

```
# A tibble: 5 x 3
  qanon                    n   pct
  <fct>                <dbl> <dbl>
1 Completely Disagree 1433. 57.1
2 Mostly Disagree      515. 20.5
3 Mostly Agree         289. 11.5
4 Completely Agree     156.  6.21
5 Skipped              115.  4.60
```

```
avs_2022 |>
  count(qanon, wt = kp_weight) |>
  mutate(pct = n/sum(n)*100)
```

```
# A tibble: 5 x 3
  qanon                    n   pct
  <fct>                <dbl> <dbl>
1 Completely Disagree 1384. 54.9
2 Mostly Disagree      554. 21.9
3 Mostly Agree         315. 12.5
4 Completely Agree     136.  5.41
5 Skipped              134.  5.29
```