Data Exploration: The Political Psychology of COVID-19

Yao Yu

November 30, 2021

In this Data Exploration assignment we will be working with a cleaned version of the replication data from Bartusevicius et al.'s (2021) article, forthcoming in *Psychological Science*, "The psychological burden of the COVID-19 pandemic is associated with anti-systemic attitudes and political violence."

Bartusevicius et al. (2021, 1) investigate "whether the psychological burden of the COVID-19 pandemic relates to anti-systemic attitudes (dissatisfaction with the fundamental social and political order), peaceful political activism, and political violence." They conducted surveys in the United States, Denmark, Italy, and Hungary in order to investigate these questions. These surveys were conducted in two waves; in the version of the data cleaned for the class all independent variables and control variables were collected in the first wave, while all outcome variables were collected in the second wave.

We will introduce one new concept—analysis of a list experiment—in this assignment. The remaining questions will focus on reviewing different ways to visually present your data analysis. Considering how best to illustrate your findings is especially important as you work on your final projects.

If you have a question about any part of this assignment, please ask! Note that the actionable part of each question is **bolded**.

COVID-19, Political Activism, and Political Violence

Data Details:

- File Name: `covid_data.rds`

- Source: This file is a cleaned version of the replication data for Bartusevicius et al. (2021). The full replication data are available here if you are curious.

Variable Name	Variable Description
`id`	Unique respondent ID
`country`	Respondent's country of residence
`age`	Numeric variable indicating one of 15 categories ranging from 18-22 (1) to 90+ (15); VERY IMPORTANT to note that these are not ages in years
`gender`	Binary variable for gender with 0 for female and 1 for male
`race`	Respondent's self-identified race; this variable is only defined for respondents in the US
`education`	Numeric variable indicating one of three categories of educational attainment: low (1), medium (2), or high (3)
`spectrum`	Self-reported political ideology from 1 (the left) to 10 (the right); this is the measure of ideology used in the non-US countries
`US_ideo`	Self-reported political ideology from 1 (very liberal) to 5 (very conservative) for US respondents only

Variable Name	Variable Description
covid_burden	An index averaging ten items designed to measure perceived burden from COVID asking respondents to agree or disagree with statements like "I have felt extremely unwell as a consequence of the coronavirus crisis," and "The extraordinary measures taken by the government in response to the coronavirus crisis make me concerned about my democratic rights." The range of the index is from 1 (lowest COVID burden) to 7 (highest COVID burden)
radical_intention	A radicalism intention scale ranging from 1 (lowest radicalism intention) to 7 (highest radicalism intention) formed as the average of four items asking people to agree or disagree with statements like "I would participate in a public protest against oppression of my group even if I thought the protest might turn violent."
list_control	The number (out of 4) of statements the respondent agreed with in the control version of the list experiment
list_treatment	The number (out of 5) of statements the respondent agreed with in the treatment version of the list experiment (which added the socially undesireable statement to the control list)
support_BLM	For this and all remaining variables respondents were asked "During the past few weeks, demonstrations and protests took place across the United States. In the context of these events, have you, personally, done any of the following during the past few weeks? If not, would you do this if you had the chance?" Answers were on a 0-4 scale where 0 is "no, would never do", 1 is "no, but would do if had the chance," 2 is "yes, once or twice," 3 is "yes, several times," and 4 is "yes, often." This variable refers to the statement, "Went to the streets to protest in support of the Black Lives Matter movement."
protest_rioting	This variable refers to the statement, "Went to the streets to protest against rioting and looting." It uses the same 0-4 scale.
protect_property	This variable refers to the statement, "Went to the streets to protect the property and safety of yourself or others against rioters." It uses the same 0-4 scale.
confront_police	This variable refers to the statement, "Engaged in physical confrontations with the police." It uses the same 0-4 scale.
confront_BLM	This variable refers to the statement, "Engaged in physical confrontations with Black Lives Matter protesters." It uses the same 0-4 scale.
destroy_property	This variable refers to the statement, "Engaged in the destruction or of property." It uses the same 0-4 scale.
confront_citizens	This variable refers to the statement, "Engaged in physical confrontations with other citizens." It uses the same 0-4 scale.

```r
#Load the data
covid_data <- read_rds('covid_data.rds')
```

Question 1

List experiments are one approach to avoiding social desireability bias—the concern that respondents to surveys give the answer that they think the survey taker wants to hear when the truth is embarrassing or not socially acceptable. In a list experiment, respondents are randomized into a control group, which receives a list of innocuous statements that are unlikely to be subject to social desirability bias, or a treatment group, which receives the same list of neutral statement plus one sensitive statement that is socially objectionable.

Respondents are then asked to report *how many* of the statements on the list they agree with, instead of being asked if they agree with specific items. It is important to note that this method cannot reveal if any particular person agrees with the sensitive item (unless they agree with all of the items when in the treatment group), but it can reveal the proportion of people who agree with the sensitive statement. To find the estimated proportion of people who agreed with the sensitive statement, researchers just subtract the mean number of items agreed with among those who saw the control list from the mean number of items agreed with among those who saw the treatment list (which included the sensitive statement).

In this paper, Bartusevicius et al. run a list experiment in which the control group saw four neutral statements like "I enjoy listening to music," while the treatment group saw those same four statements plus the sensitive statement, "I would use violence against the government or other authorities (for example, during a protest)."

Part a

Estimate the proportion of respondents in the whole sample who agreed with the sensitive statement about being willing to use political violence.

```
mean(covid_data$list_treatment, na.rm = TRUE) - mean(covid_data$list_control, na.rm = TRUE)
```

```
## [1] 0.06705619
```

Part b

It's useful to use this method to see if the proportion of people who agree with the sensitive statement is different among subsets of people who differ on a predictor variable of interest. In this case, we are interested in whether the psychological burden of COVID-19 is associated with increased willingness to use political violence.

Define an indicator variable that is TRUE if the respondent has a perceived COVID burden equal or higher to the median value in the sample and FALSE if it is less than the median value. Estimate the proportion of people who agree with the sensitive statement and have low perceived COVID burden and the proportion of people who agree with the sensitive statement and have high perceived COVID burden. Compare the two values and comment on what you find.

```
median_covid_burden <- median(covid_data$covid_burden)

covid_data_clean <- covid_data %>%
  mutate(has_covid_burden = (covid_burden >= median_covid_burden))

covid_data_clean %>%
  group_by(has_covid_burden) %>%
  summarize(mean_list_control = mean(list_control, na.rm = TRUE),
            mean_list_treatment = mean(list_treatment, na.rm = TRUE),
            diff = mean_list_treatment - mean_list_control)
```

```
## # A tibble: 2 x 4
##   has_covid_burden mean_list_control mean_list_treatment   diff
##   <lgl>                        <dbl>               <dbl>  <dbl>
## 1 FALSE                         2.05                2.08 0.0289
## 2 TRUE                          2.17                2.28 0.112
```

Now let's take some time to review several of the different visualization strategies we have used this semester.

Question 2

Density plots are a good way to show the distribution of a variable of interest in a population. **Make a density plot with perceived COVID-19 burden as the variable of interest. Use different colors to simultaneously show the density of this variable in each country in the sample. Comment on any notable differences between countries.**

```r
covid_data_clean %>%
  ggplot(aes(x = covid_burden, fill = country)) +
  geom_density(alpha = 0.2) +
  theme_light() +
  labs(
    title = "Perceived COVID-19 Burden by Country",
    x = "Perceived COVID-19 Burden",
    y = "Density",
    fill = "Country"
  )
```


Perceived COVID-19 Burden by Country
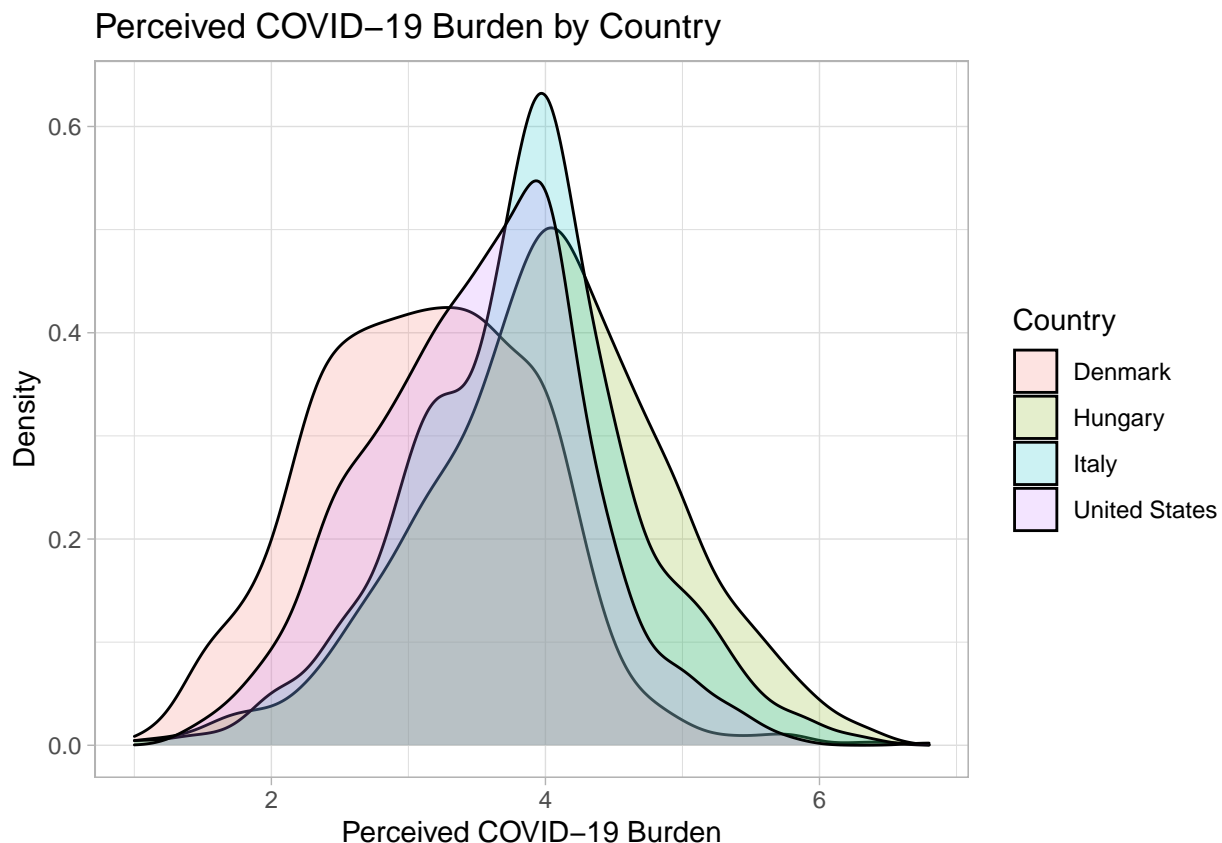
Denmark! They had COVID-19 under control at the time (much better than the other 3 countries).

Question 3

We already analyzed the list experiment. Let's look at a different way to evaluate the link between the psychological burden of COVID and radicalization: linear regression.

We have shown the results of bivariate linear regressions using regresion tables before, but `ggplot` makes it easy to visualize bivariate regression lines. As an example, here is code that shows an OLS regression line of radicalism intention on COVID burden. The key is to set method = 'lm' inside geom_smooth (recall that `lm` is the function in `R` for fitting a regression). The `se` argument controls whether or not a confidence interval is

also plotted. (Some of you made similar plots are part of Data Exercise 7, but we did not ask you to plot the confidence intervals.)

```
# plot the regression line
covid_data %>%
  ggplot(mapping = aes(x = covid_burden, y = radical_intention)) +
  geom_smooth(method = 'lm', se = TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 443 rows containing non-finite values (stat_smooth).
```
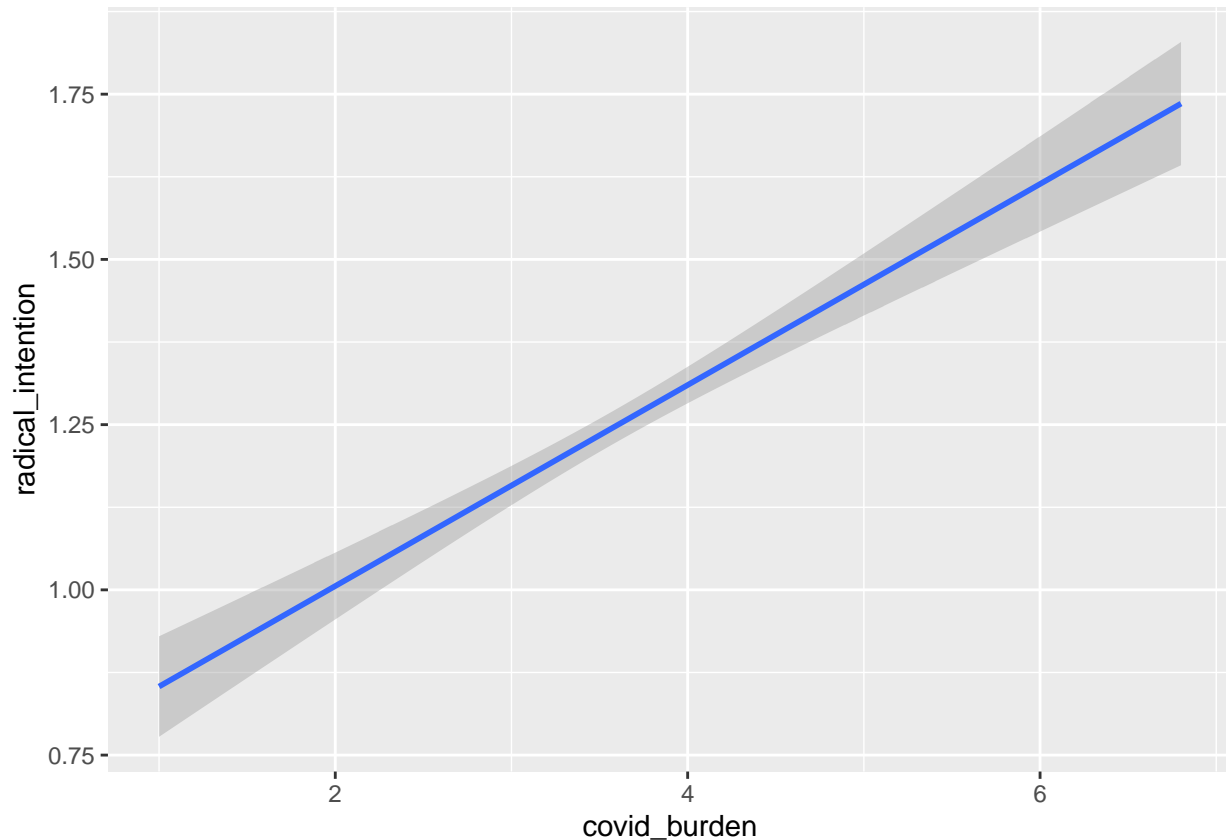

This plot shows a clear positive relationship between COVID burden and radicalism intention. Now you'll try a slightly more complicated version of this task.

Part a

Graph bivariate linear regressions of radicalism intention on COVID burden for each country. Try to do this on a single plot instead of four different plots. Comment on what you find. Note that you may not want to plot confidence intervals to avoid cluttering the plot.

```
# plot the regression line
covid_data %>%
  ggplot(mapping = aes(x = covid_burden, y = radical_intention, color = country)) +
  geom_smooth(method = 'lm', se = FALSE) +
  theme_light() +
  labs(
    title = "Radicalism Intention on COVID Burden for each Country",
    x = "COVID Burden",
```
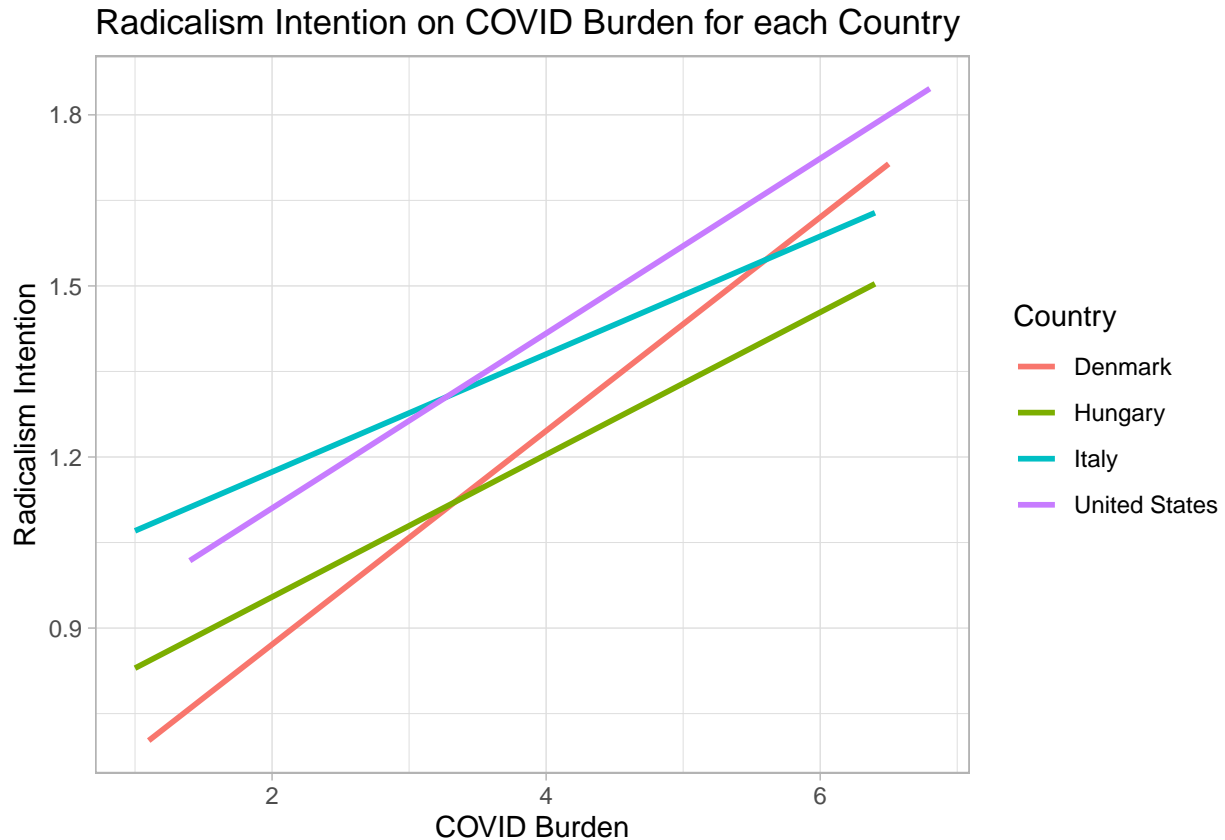
```
    y = "Radicalism Intention",
    color = "Country"
  )
```

`geom_smooth()` using formula 'y ~ x'

Warning: Removed 443 rows containing non-finite values (stat_smooth).

Radicalism Intention on COVID Burden for each Country

Part b

Fit a multivariate linear regression of radicalism intention on COVID burden with education, age, and gender as control variables. Use the complete sample; don't subset by country. Present the results in a regression table and comment on your findings. Make sure to comment on the main explanatory variable (COVID burden) but also comment on the control variables in the context of theories we have learned about in class. Hint: The stargazer package makes good regression tables; set type = 'text'.

```
fit_1 <- lm(radical_intention ~ covid_burden + education + age + gender, data = covid_data)

stargazer(fit_1, header = FALSE)
```

Part c

Are these results robust to the inclusion of a control variable for political ideology? These data have ideology variables, but use different ones for the US and non-US samples. **Using either a US or non-US subset of the data fit the same regression as above, but add the relevant ideology variable. Comment**

Table 2:

	Dependent variable:
	radical_intention
covid_burden	0.142***
	(0.014)
education	0.043***
	(0.016)
age	−0.045***
	(0.004)
gender	0.190***
	(0.025)
Constant	0.881***
	(0.068)
Observations	4,125
R^2	0.073
Adjusted R^2	0.072
Residual Std. Error	0.800 (df = 4120)
F Statistic	81.279*** (df = 4; 4120)
Note:	*p<0.1; **p<0.05; ***p<0.01

on if previous estimates were stable, why they might be different, and on what you find out about the effect of ideology. What do the results say about the power of worldview in shaping political behavior?

```
covid_data_us <- covid_data %>%
  filter(country == "United States")

fit_2 <- lm(radical_intention ~ covid_burden + education + age + gender + US_ideo, data = covid_data_us)

stargazer(fit_2, header = FALSE)
```

Table 3:

	Dependent variable:
	radical_intention
covid_burden	0.083**
	(0.034)
education	0.011
	(0.032)
age	−0.076***
	(0.008)
gender	0.104*
	(0.053)
US_ideo	−0.148***
	(0.022)
Constant	1.974***
	(0.170)
Observations	971
R^2	0.151
Adjusted R^2	0.147
Residual Std. Error	0.821 (df = 965)
F Statistic	34.404*** (df = 5; 965)
Note:	*p<0.1; **p<0.05; ***p<0.01

Question 4

There are seven different outcome variables associated with respondents' activities related to the summer 2020 BLM protests. (Note that these variables were only collected for US respondents.) **Fit seven different multivariate linear regressions, one for each outcome variable. Use COVID burden, education, age, and gender as independent variables. Make a single coefficient plot with the point estimate and 95% confidence intervals for the coefficient on COVID burden for each outcome variable.** Hint: look back at the solutions to Data Exercise 8 for ideas about how to code this.

```
fit_support_BLM <- lm(support_BLM ~ covid_burden + education + age + gender + US_ideo,
                  data = covid_data_us)
fit_protest_rioting <- lm(protest_rioting ~ covid_burden + education + age + gender + US_ideo,
```

```
                             data = covid_data_us)
fit_protect_property <- lm(protect_property ~ covid_burden + education + age + gender + US_ideo,
                             data = covid_data_us)
fit_confront_police <- lm(confront_police ~ covid_burden + education + age + gender + US_ideo,
             data = covid_data_us)
fit_confront_BLM <- lm(confront_BLM ~ covid_burden + education + age + gender + US_ideo,
             data = covid_data_us)
fit_destroy_property <- lm(destroy_property ~ covid_burden + education + age + gender + US_ideo,
             data = covid_data_us)
fit_confront_citizens <- lm(confront_citizens ~ covid_burden + education + age + gender + US_ideo,
             data = covid_data_us)

plot_summs(fit_support_BLM, fit_protest_rioting, fit_protect_property,
           fit_confront_police, fit_confront_BLM, fit_destroy_property,
           fit_confront_citizens, scale = TRUE, plot.distributions = TRUE)
```

Loading required namespace: broom.mixed
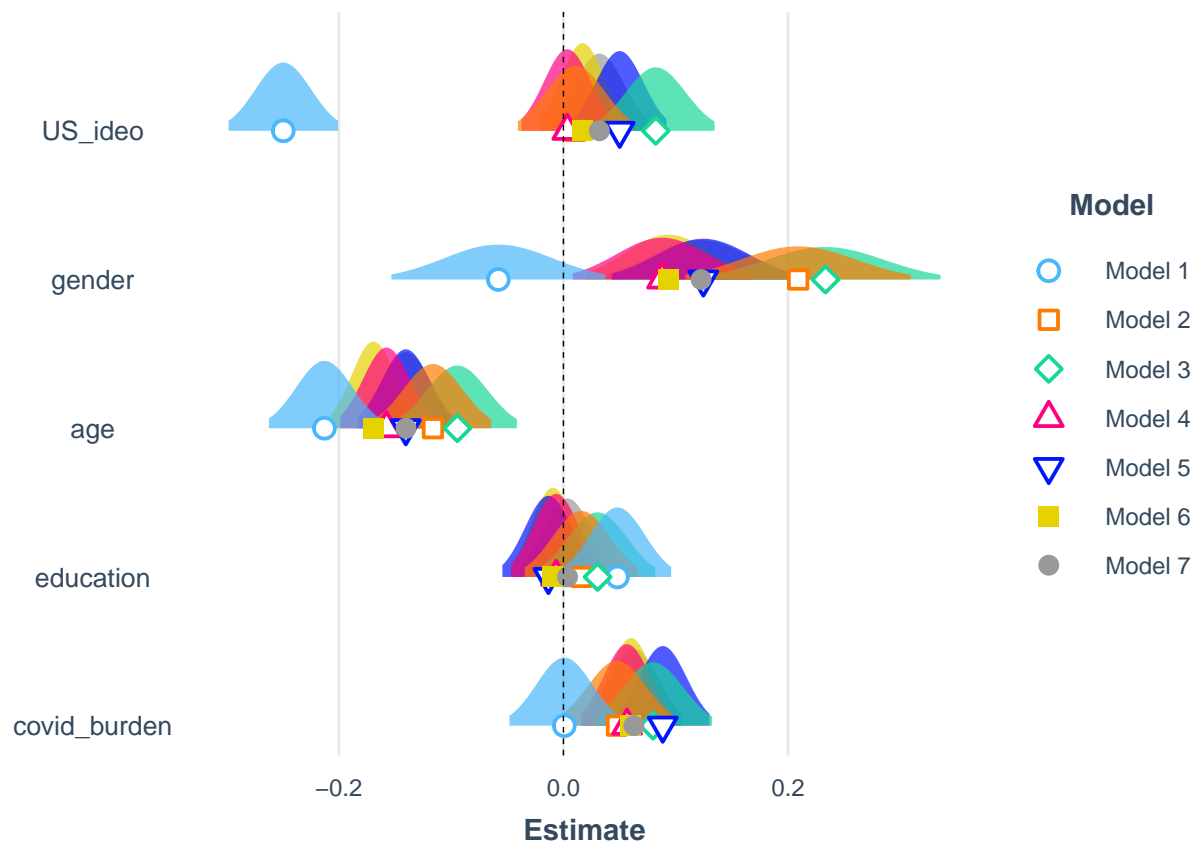

```
#          model.names = c("Model 1"),
#          coefs = c("Age:Treated" = "age:treatedTRUE",
#                    "Treated" = "treatedTRUE",
#                    "Age" = "age",
#                    "Japanese Friend" = "jpfriend")) +
# labs(title = "Coefficient Plot")
```

Question 5: DATA SCIENCE QUESTION

Pick one of the outcomes associated with the BLM protests. Define a new binary variable that takes 1 if the respondent actually engaged in the behavior and 0 if they didn't engage in it or merely hypothetically would engage in it. Fit a regression model of your choosing with your new variable as the outcome, COVID burden as the independent variable, and control variables of your choosing. Plot the predicted probability of engaging in your chosen behavior as a function of COVID burden, holding other variables at their mean (or, for gender, making separate plots for each gender). Make sure to also plot 95% confidence intervals. Hint: look back at the solutions to Data Exercise 6 for ideas about how to code this.

```r
covid_data_protect_property_yes <- covid_data_us %>%
  mutate(protect_property_yes = (protect_property > 1))

fit_yes <- lm(protect_property_yes ~ covid_burden + education + age + gender + US_ideo,
    data = covid_data_protect_property_yes)

fake_data <- tibble(covid_burden = rep(seq(1, 7, by = 1), 2),
      education = mean(covid_data_us$education, na.rm = TRUE),
      age = mean(covid_data_us$age, na.rm = TRUE),
      US_ideo = mean(covid_data_us$US_ideo, na.rm = TRUE),
      gender = c(rep(0, 7), rep(1, 7)))

preds <- predict(fit_yes, newdata = fake_data, se.fit = TRUE)

fake_data <- fake_data %>%
  mutate(pred = preds$fit,
        se = preds$se.fit,
        se_lower = pred - 1.96*se,
        se_upper = pred + 1.96*se)

fake_data %>%
  ggplot(aes(x = covid_burden, y = pred)) +
  geom_line() +
  geom_line(aes(x = covid_burden, se_lower), linetype = "dashed", color = "grey") +
  geom_line(aes(x = covid_burden, se_upper), linetype = "dashed", color = "grey") +
  facet_wrap(~gender) +
  theme_light() +
  labs(
    title = "Predicted Probability of Protecting Propoerty as a function of COVID Burden",
    x = "COVID Burden",
    y = "Predicted Probability of Protecting Propoerty"
  )
```
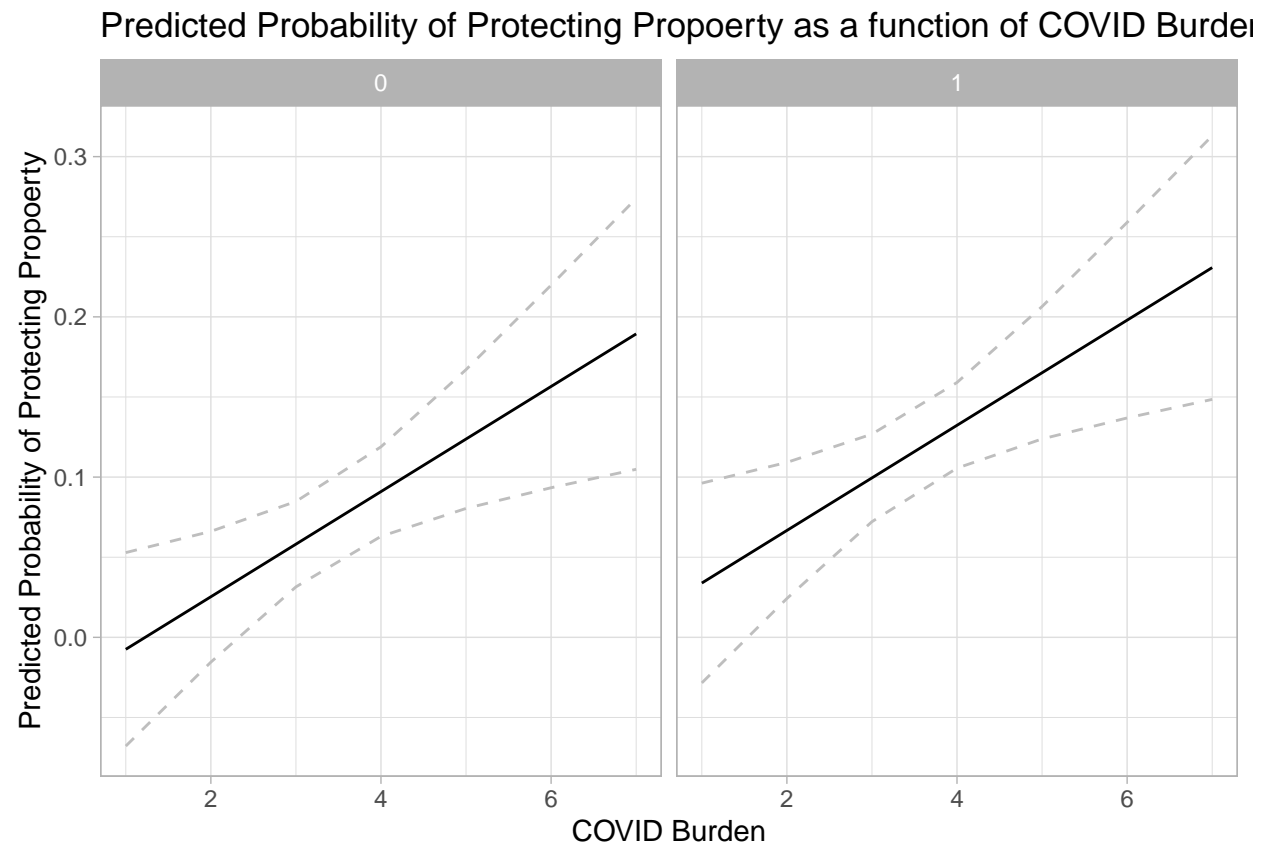
Predicted Probability of Protecting Propoerty as a function of COVID Burden

Question 6

Identify another question you have about the link between COVID burden and political action/violence and use the data to answer it.