

Section 6: Your replication data

Gov 50 Section

October 14, 2026

Today's goal: By the end of section, you calculate the first red square in Figure 1 of Barber and Pope yourself.

Your TF runs each **Watch** chunk. Run it on your laptop at the same time. Then do the **Try it** with a partner.

Block 0: Set up and knit

Do each step with your TF.

1. Make a folder with the name `gov50-replication` on your Desktop or in Documents. Keep all of your project files in this folder.
2. Go to ascamara.github.io/gov50. Under Section 6, download `gov50-section6-lab.Rmd` and `Ideology_Trump.dta`.
3. Move both files into `gov50-replication`.
4. In RStudio, click **File > Open File**. Open the `.Rmd` that is in `gov50-replication`.
5. In the Console, run `install.packages(c("haven", "dplyr", "ggplot2"))`. Do not type any other code in the Console.
6. Click **Knit**.

The chunk below reads the data. Then `glimpse()` prints a short view of the data. `glimpse()` comes from the `dplyr` package. If your PDF shows **Rows: 13,000** and **Columns: 36**, your setup works.

```
library(haven)
library(dplyr)
bp <- read_dta("Ideology_Trump.dta")
glimpse(bp)
```

If the error says	Do this
there is no package called does not exist in current working directory	Do step 5. Put the <code>.Rmd</code> and the <code>.dta</code> in the same folder. Then do step 4 again.

Block 1: Look at the data

Watch

Figure 1 needs three facts about each answer:

1. Who gave the answer? The column **republican** tells us.
2. Which cue did the person see? The column **condition** tells us.
3. What was the answer? The column **Support** tells us.

The data has 13,000 rows. But only 1,300 people took the survey. Look at the first 12 rows to see why.

```
bp[1:12, c("caseid", "Question", "republican", "condition", "Support")]
```

One row is one answer to one question. Each person answered 10 questions.

Column	Values
republican, democrat	1 = yes. 0 = no. Leaners count with their party.
condition	1 = Trump takes the liberal side. 2 = Trump takes the conservative side. 3 = a cue from Republicans in Congress. 5 = no cue. There is no 4.
Support	1 = liberal answer. 0 = conservative answer. NA = "don't know."

`table()` counts how many times each value appears in a column. `useNA = "ifany"` also counts the NAs.

```
table(bp$Support, useNA = "ifany")
```

The mean of a column of 0s and 1s is the share of 1s. Thus the mean of **Support** is the share of liberal answers. `na.rm = TRUE` skips the "don't know" answers.

```
mean(bp$Support, na.rm = TRUE)
```

68% of all answers are liberal.

Now keep only the Republicans. `subset()` keeps the rows where **republican** is 1. `nrow()` counts the rows. Divide by 10 to count the people.

```
reps <- subset(bp, republican == 1)
nrow(reps) / 10
mean(reps$Support, na.rm = TRUE)
```

442 Republicans took the survey. 42% of their answers are liberal.

Try it

```
# 1. Make a data set with the name dems. Keep the rows where democrat is 1.
# 2. How many Democrats took the survey? Use nrow(), then divide by 10.
# 3. What share of Democrat answers is liberal? Use mean().
# 4. Compare this share with the 42% for Republicans. Write your answer as a comment.
```

Block 2: Calculate the first red square

Watch

Figure 1 compares two groups of Republicans. One group saw no cue. The other group saw the liberal Trump cue. Make a data set for each group.

```
no_cue <- subset(reps, condition == 5)
lib_cue <- subset(reps, condition == 1)
```

Calculate the share of liberal answers in each group. Then subtract.

```
mean(no_cue$Support, na.rm = TRUE)
mean(lib_cue$Support, na.rm = TRUE)
mean(lib_cue$Support, na.rm = TRUE) - mean(no_cue$Support, na.rm = TRUE)
```

With no cue, 40% of Republican answers are liberal. With the liberal Trump cue, 56% are liberal. The difference is 0.16, or 16 percentage points. This difference is the effect of the cue. Figure 1 shows it as the first red square.

`aggregate()` calculates the share for all four conditions at the same time. It has three parts.

- `Support ~ condition` is a formula. The column on the left of `~` is the column to summarize. The column on the right of `~` splits the rows into groups. Read it as “Support, split by condition.”
- `data = reps` tells R which data set to use.
- `FUN = mean` tells R what to calculate for each group. Write `mean` without parentheses.

`aggregate()` skips the NAs, so you do not need `na.rm = TRUE`. The result is a small data set with one row for each condition.

```
rep_shares <- aggregate(Support ~ condition, data = reps, FUN = mean)
rep_shares
```

Now calculate the effect of every cue at the same time. Subtract the no-cue share from each share. `rep_shares$Support[rep_shares$condition == 5]` is the share in the row where `condition` is 5.

```
rep_shares$effect <- rep_shares$Support - rep_shares$Support[rep_shares$condition == 5]
rep_shares
```

The `effect` column is close to the red squares in Figure 1. The numbers are not exactly the same, because the paper also controls for race. The effect for condition 5 is 0, because no cue is the baseline.

Try it

```
# 1. Make dem_shares. Use aggregate() on dems, the data set from Block 1.
# 2. Add an effect column to dem_shares. Subtract the no-cue share.
# 3. Which cue changes the answers of Democrats the most? Write your answer
#    as a comment.
```

Block 3: Read Figure 1

This block has no code. Look at Figure 1 on the screen. Talk with a partner first. Then we talk as a class.

- A red square is for Republicans. A blue circle is for Democrats. A green triangle is for independents.
- The height of a marker is the effect of the cue. The effect is the change in the share of liberal answers, compared with no cue.
- The line through a marker is the 95% confidence interval. If the line crosses 0, we cannot say that the cue had an effect.

When ____ see the ____ cue, the share of liberal answers goes up (or down) by ____ percentage points.

1. Find the red square above “Liberal Trump.” Find about the same number in the **effect** column of **rep_shares**. Say it in words with the sentence above.
2. Find the three blue circles. Compare them with the **effect** column in **dem_shares**. Do their lines cross 0? What does this tell you about Democrats?

Block 4: Groups

This block has no code. Sit with your project group. Answer these two questions with your group.

1. What outcome do you want to study? For example, you can study the answers on one policy, such as the minimum wage.
2. Which groups of people do you want to compare? For example, you can compare people with low and high political knowledge.

For ideas, look at the figures and tables in Barber and Pope and in Tappin (2022). Tappin did the same type of experiment with 34 policies. His data is at osf.io/t2bpj.

Block 5 (Optional): Plot the red squares

Watch

This plot shows the **effect** column from Block 2. It is Figure 1 for Republicans, without the confidence intervals.

First, give each condition in **rep_shares** a name. The names go in the same order as the rows.

```
rep_shares$cue <- c("Trump liberal", "Trump conservative", "Congress", "No cue")
```

You build a ggplot in layers. Each + adds one layer. Put each + at the end of a line.

```
library(ggplot2)
ggplot(rep_shares, aes(x = cue, y = effect))      # Layer 1: the axes.
```

```
ggplot(rep_shares, aes(x = cue, y = effect)) +
  geom_point()                                  # Layer 2: a dot for each row.
```

```
ggplot(rep_shares, aes(x = cue, y = effect)) +  
  geom_point() +  
  geom_hline(yintercept = 0, linetype = "dashed") + # Layer 3: a dashed line at 0.  
  labs(x = "Cue", y = "Effect", title = "Republicans") # Layer 4: labels.
```

Try it

```
# 1. Give each condition in dem_shares a name. Use the same names as above.  
# 2. Make the same plot for Democrats. Use the title "Democrats".  
# 3. Compare your two plots with Figure 1. Write one difference as a comment.
```

Before you leave

Knit to PDF. Upload the PDF to Canvas. If the knit fails, the error gives the line number. Fix that line. If you cannot fix it, ask your TF.