

Section 7: ggplot and your descriptives

Gov 50 Section

October 21, 2026

Today's goal: By the end of section, you make the descriptive table and the descriptive plot for your replication.

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

The project

The project has two parts: a replication and an extension.

Section	Replication	Extension
Introduction	The paper's big question, research question, theory, and hypotheses	The same, for your extension
Descriptives	1 table and 1 plot	The same, for your subgroup
Main analysis	1 regression table and 1 effect plot, each with an interpretation	The same, for your subgroup
Conclusion	How do the findings inform the big question? What are the limitations?	The same, for your extension

For most extensions, you repeat the analysis for one subgroup. A subgroup can be a group of people, such as the people with high political knowledge. It can also be one question, such as abortion. If your group wants to do a different type of extension, talk to Antonio.

Section	What we do
7 (today)	Make the descriptive table and plot
8	Run a regression. Read and write a regression table.
9	Plot the regression coefficients in an effect plot
10	Extension workshop, and time to finish earlier labs
11	Presentation practice

Block 0: Knit

1. Download `gov50-section7-lab.Rmd` from ascamara.github.io/gov50.
2. Move it into your `gov50-replication` folder, next to `Ideology_Trump.dta`.
3. Open the `.Rmd`. Click **Knit**.

The chunk adds a `party` column and a `cue` column. It also makes `people`, with one row for each person. If your PDF shows 1300, your setup works.

```
library(haven)
library(ggplot2)
bp <- read_dta("Ideology_Trump.dta")
bp$party <- "Independent"
bp$party[bp$republican == 1] <- "Republican"
bp$party[bp$democrat == 1] <- "Democrat"
bp$cue <- factor(bp$condition, levels = c(5, 1, 2, 3),
                 labels = c("No cue", "Trump liberal", "Trump conservative", "Congress"))
people <- subset(bp, Question == "Abortion")
nrow(people)
```

If the error says	Do this
there is no package called	In the Console, run
does not exist in current working directory	<code>install.packages(c("haven", "ggplot2"))</code> . Put the <code>.Rmd</code> and the <code>.dta</code> in the same folder. Then open the <code>.Rmd</code> again.

Block 1: The descriptive table

Watch. `table()` counts the rows for each value of a column. Each row of `people` is one person, so this table gives the number of people in each party.

```
table(people$party)
```

Try it 1. Make the descriptive table for your replication.

```
# 1. Count the number of people who got each cue.
# 2. Count the number of people for each pair of a cue and a party.
```

Which group has the fewest people?

Block 2: The grammar of graphics

The “gg” in `ggplot` means the **grammar of graphics**. You make a plot in layers. Each `+` adds one layer. Every `ggplot` has the same template:

```
ggplot(DATA, aes(MAPPINGS)) +
  GEOM() +
  ...one more layer for each +
```

- `DATA` is the data frame.
- `aes()` tells R which column goes on which axis.
- The geom tells R how to draw the data.

Watch. We make a bar chart of the number of people in each party, one layer at a time.

Step 1. The data only. R does not know which column to use. It draws an empty plot.

```
ggplot(people)
```

Step 2. Add the mapping. R puts `party` on the x-axis. It has no geom, so it draws no data.

```
ggplot(people, aes(x = party))
```

Step 3. Add a geom. `geom_bar()` counts the rows for each party. The bars show the numbers of your first table.

```
ggplot(people, aes(x = party)) +  
  geom_bar()
```

Try it 2. Make a bar chart of the number of people who got each cue.

```
# Use people.
```

Which cue did the most people get? How many people got it?

Block 3: Types of geoms

The type of the column tells you which geom to use. For more geoms, see the [ggplot cheat sheet](#).

Plot	Geom	Use it for
Bar chart that counts rows	<code>geom_bar()</code>	A column with categories, such as <code>party</code>
Histogram	<code>geom_histogram()</code>	A column with numbers, such as <code>knowledge</code>
Bar chart with heights that you calculated	<code>geom_col()</code>	A share or a mean for each category
Points	<code>geom_point()</code>	A share for each category, or two columns with numbers (a scatter plot), such as <code>age</code> and <code>knowledge</code>
Line plot	<code>geom_line()</code>	A number that changes with another number, such as the share of liberal answers at each level of <code>knowledge</code>

Watch. Do people who know more about politics give more liberal answers? Or fewer?

Step 1. A histogram. `knowledge` is the number of correct answers on 8 political questions. A histogram puts the numbers into bins and counts the people in each bin. `binwidth = 1` makes one bin for each number of correct answers.

```
ggplot(people, aes(x = knowledge)) +  
  geom_histogram(binwidth = 1)
```

Step 2. A bar chart of shares. `aggregate()` calculates the share of liberal answers at each level of knowledge, as in section 6. We have the heights already, so we use `geom_col()`, not `geom_bar()`.

```
know <- aggregate(Support ~ knowledge, data = bp, FUN = mean)
ggplot(know, aes(x = knowledge, y = Support)) +
  geom_col()
```

Step 3. A line plot. Change only the geom, and you get a different plot. `geom_line()` connects the same shares from left to right.

```
ggplot(know, aes(x = knowledge, y = Support)) +
  geom_line()
```

Try it 3. Make the line plot for Democrats. Then make it for Republicans.

```
# 1. Keep the Democrat rows of bp. Make the line plot for them.
# 2. Do the same for Republicans.
```

Does knowledge move Democrats and Republicans in the same direction? Give the share of liberal answers at knowledge 0 and at knowledge 8 for each party.

Block 4: Axes, labels, and panels

ggplot labels each axis with the column name. A reader needs real labels.

- `labs()` sets the labels. A good title says the finding.
- `ylim(0, 1)` makes the y-axis go from 0 to 1. Each share is between 0 and 1.
- `geom_hline()` adds a horizontal line.
- `facet_wrap()` makes one small plot, a **panel**, for each group.

Watch.

Step 1. Labels, limits, and a line. First, we calculate the share of liberal answers for each cue and party. In section 6, `aggregate()` used one group. Here it uses two groups: `cue + party`. The result has one row for each pair of a cue and a party, so it has $4 \times 3 = 12$ rows. `Support` is the share of liberal answers for that pair. Then we keep the Republican rows.

```
shares <- aggregate(Support ~ cue + party, data = bp, FUN = mean)
rep_shares <- subset(shares, party == "Republican")
rep_shares
```

We put a dashed line at the Republican “No cue” share. `rep_no_cue` keeps that one row of `rep_shares`. `geom_hline()` takes the height of the line from the `Support` column of `rep_no_cue`. We do not type the number, so the line is always at the correct share.

```
rep_no_cue <- subset(rep_shares, cue == "No cue")
ggplot(rep_shares, aes(x = cue, y = Support)) +
  geom_col() +
  geom_hline(data = rep_no_cue, aes(yintercept = Support), linetype = "dashed") +
  ylim(0, 1) +
  labs(x = "Cue", y = "Share of liberal answers",
       title = "The liberal Trump cue moves Republicans to the left")
```

The distance from the top of a bar to the dashed line is the difference in means. This is the effect of the cue.

Step 2. Panels. Now we plot all of `shares`, with `party` on the x-axis. Each party has four rows in `shares`, one for each cue. Without panels, the bars of the four cues stack on top of each other. Each bar is the sum of four shares, so it goes above 1.

```
ggplot(shares, aes(x = party, y = Support)) +  
  geom_col()
```

`facet_wrap()` fixes this. It makes one small plot, a **panel**, for each group.

- `facet_wrap(~ cue)` makes one panel for each cue. The `~` means “by”, as in `aggregate()`.
- Each panel has one bar for each party.
- `ncol = 2` puts the panels in two columns.

```
ggplot(shares, aes(x = party, y = Support)) +  
  geom_col() +  
  facet_wrap(~ cue, ncol = 2)
```

Try it 4. Make your project plot: the shares of all three parties.

```
# Your plot needs these parts:  
# - the bars of shares, with one panel for each party  
# - a dashed line at the "No cue" share of each party  
# - a y-axis from 0 to 1  
# - labels and a title
```

In the Republican panel, how far is each cue bar from the dashed line: “Trump liberal”, “Trump conservative”, and “Congress”? What does each distance tell you about the effect of the cue?

Block 5: Groups (optional)

Sit with your project group. Answer these questions together.

1. What is your research question? What is your theory? What is your hypothesis?
2. Which data and which method will you use to test your hypothesis? For example, you can use a subgroup of the Barber and Pope data, and repeat the replication for it.

Before you leave, tell your TF your answers.

Before you leave

Knit to PDF. Upload the PDF to Canvas. Keep this `.Rmd` in your `gov50-replication` folder. Your project uses this code again.