PS 3 Section Week 3

Zachary Lorico Hertz

09 September 2025

Agenda for today

1. Housekeeping
2. Review key operations
3. Individual practice exercise
4. Group collaboration on practice exercise
5. Recap + share-out
6. Office hours

Housekeeping Announcements

- Slides and answer keys to any in-section exercises will be posted every Thursday on Github.
- You **must attend section**. You get one absence excused no questions asked; further unexcused absences drop your participation by 20% per absence. For excused absences to be considered, you must submit a request through this form.
- Starting next week, to receive full credit for attendance you must also fill out the feedback survey in your email you receive after section - you can fill it with N/As, if you have no feedback.

REMEMBER: Our steps to survive the start

1. What is the shape of my dataset?
2. What are the variables?
3. What is my research question(s)?

Putting this into practice

```
1 dat <- read.csv("export.csv")
```

Putting this into practice

```
1 dat <- read.csv("export.csv")
2 nrow(dat)
```

```
[1] 1153
```

Putting this into practice

```r
dat <- read.csv("export.csv")
nrow(dat)
```

```
[1] 1153
```

```r
names(dat)
```

```
 [1] "treatment"           "nativity"            "asian_ethnicity"
 [4] "zlh_ethnic_lf_merged" "zlh_intra_lf_merged"  "zlh_gcb_merged"
 [7] "ft.mobilization_1"    "FTs_1_delta"          "FTs_2_delta"
[10] "FTs_4_delta"          "FTs_5_delta"
```

Codebook

Variable	Description
treatment	Binary treatment indicator
nativity	Country of birth
asian_ethnicity	Respondents' ethnic background
zlh_ethnic_lf_merged	Ethnic linked fate (5-point scale)
zlh_intra_lf_merged	Racial linked fate (5-point scale)
zlh_gcb_merged	Generic congressional ballot preference
ft.mobilization_1	Likelihood to vote in local elections (0-100)
FTs_1_delta	Post-pre difference in FT ratings for Democrats
FTs_2_delta	Post-pre difference in FT ratings for Republicans
FTs_4_delta	Post-pre difference in FT ratings for Whites
FTs_5_delta	Post-pre difference in FT ratings for Asians

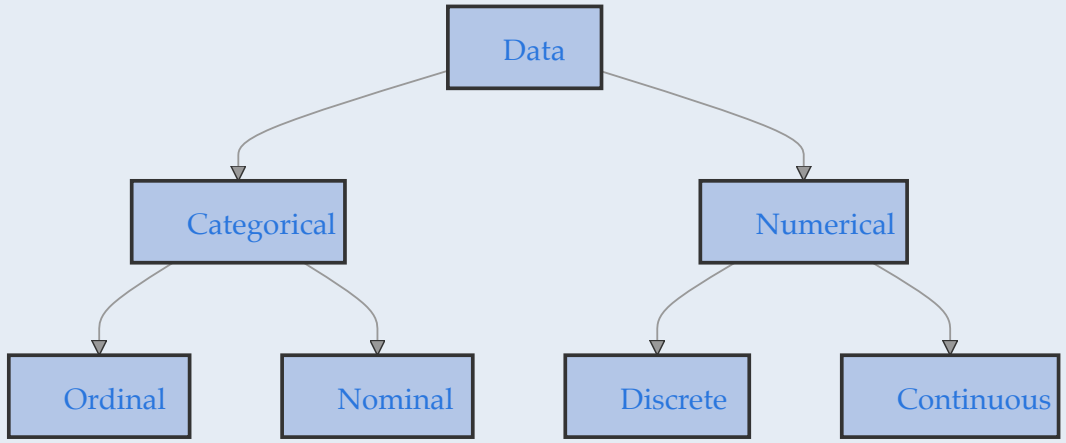
We can take a peek at the data:

```
1 head(dat)
```

```
  treatment                          nativity                 asian_ethnicity
1         1  I was born in another country:                           Chinese
2         0  I was born in another country:                          Filipino
3         0  I was born in another country:                          Filipino
4         0 I was born in the United States                           Chinese
5         1 I was born in the United States                         Taiwanese
6         0 I was born in the United States Indian, Pakistani or South Asian
        zlh_ethnic_lf_merged zlh_intra_lf_merged zlh_gcb_merged
1              Somewhat agree      Somewhat agree   The Democrat
2              Somewhat agree      Somewhat agree The Republican
3           Strongly disagree   Strongly disagree The Republican
4              Somewhat agree      Strongly agree The Republican
5              Somewhat agree      Strongly agree   The Democrat
6 Neither agree nor disagree      Strongly agree   The Democrat
```

mfw students don’t look at the data before proceeding

Data Types


Data
Categorical
Numerical
Ordinal
Nominal
Discrete
Continuous

What variables are categorical?

Variable	Description
treatment	Binary treatment indicator
nativity	Country of birth
asian_ethnicity	Respondents' ethnic background
zlh_ethnic_lf_merged	Ethnic linked fate (5-point scale)
zlh_intra_lf_merged	Racial linked fate (5-point scale)
zlh_gcb_merged	Generic congressional ballot preference
ft.mobilization_1	Likelihood to vote in local elections (0-100)
FTs_1_delta	Post-pre difference in FT ratings for Democrats
FTs_2_delta	Post-pre difference in FT ratings for Republicans
FTs_4_delta	Post-pre difference in FT ratings for Whites
FTs_5_delta	Post-pre difference in FT ratings for Asians

R variables and data types

R has a specific set of variable classes when working with different types of data. Common errors can arise from attempting operations on a variable with the wrong class.

Category	Subtype	R Variable Class	Type	Example
Numerical	Discrete	integer	Whole numbers	1, 100, -9
Numerical	Continuous	numeric	Decimals	0.1, -0.09, 234.567
Categorical	Nominal	character	Text	“Black”, “White”
Categorical	Nominal/Ordinal	factor	Categorical	“Support”, “Oppose”, “Not sure”

Class matters

`summary()` behaves differently, depending on the variable type. For example:

```
1 summary(dat$asian_ethnicity)
```

```
   Length     Class      Mode
     1153 character character
```

```
1 summary(dat$FTs_5_delta)
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
-79.0000  -2.0000   0.0000   0.3601   3.0000  50.0000        6
```

Why does it help to check?

Sometimes, R considers numbers weirdly.

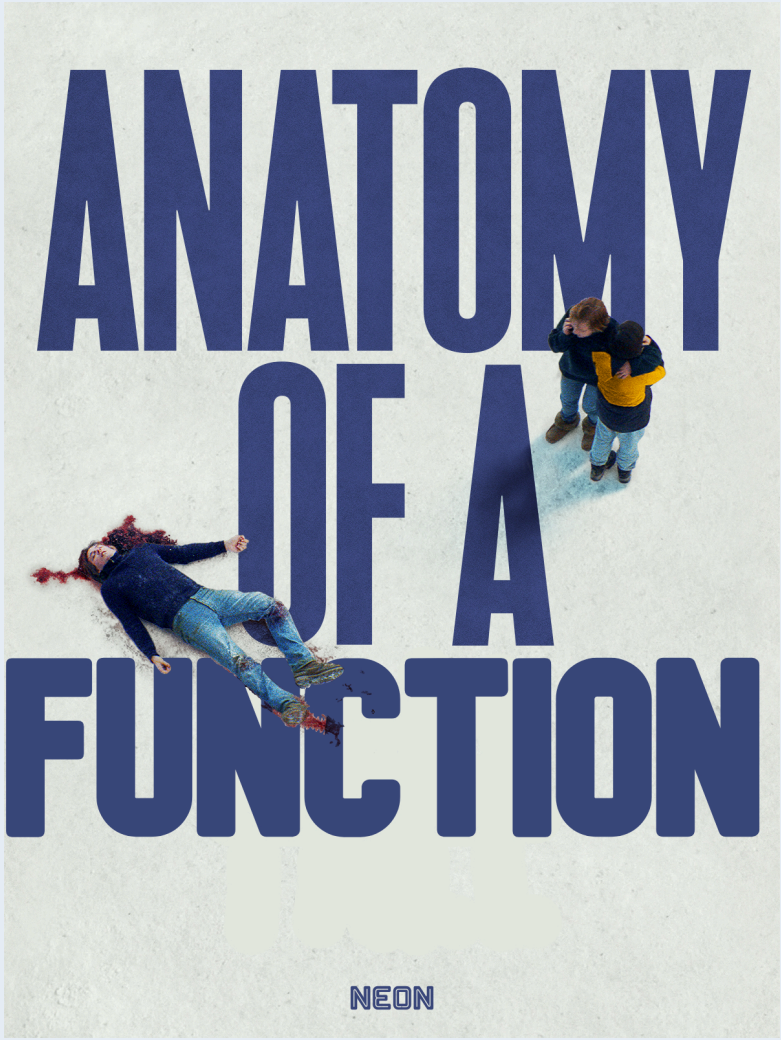
```
1 head(dat$treatment)
```

```
[1] "1" "0" "0" "0" "1" "0"
```

```
1 class(dat$treatment)
```

```
[1] "character"
```

Some time on syntax



The $ operator

The $ operator lets you access specific columns (variables) in your dataset.

Syntax: `dataframe$column_name`

```
1 # Access the treatment variable
2 head(dat$treatment)
```

```
[1] "1" "0" "0" "0" "1" "0"
```

```
1 # Access the nativity variable
2 head(dat$nativity)
```

```
[1] "I was born in another country:"  "I was born in another country:"
[3] "I was born in another country:"  "I was born in the United States"
[5] "I was born in the United States" "I was born in the United States"
```

Think of it as: "Hey R, from the `dat` dataset, give me the `treatment` column"

Anatomy of `subset()`

The `subset()` function filters your data based on conditions you specify.

Syntax: `subset(dataframe, condition)`

Anatomy of `subset()`

```r
# Keep only rows where treatment equals 1
head(subset(dat, treatment == 1))
```

```
   treatment                          nativity                 asian_ethnicity
1          1  I was born in another country:                         Chinese
5          1 I was born in the United States                       Taiwanese
7          1  I was born in another country:                         Chinese
8          1 I was born in the United States                        Japanese
9          1 I was born in the United States                         Chinese
10         1  I was born in another country: Indian, Pakistani or South Asian
         zlh_ethnic_lf_merged zlh_intra_lf_merged zlh_gcb_merged
1               Somewhat agree      Somewhat agree   The Democrat
5               Somewhat agree      Strongly agree   The Democrat
7               Somewhat agree      Somewhat agree The Republican
8            Somewhat disagree   Somewhat disagree   The Democrat
9  Neither agree nor disagree      Somewhat agree   The Democrat
10              Strongly agree      Strongly agree   The Democrat
```

Anatomy of `subset()`

```r
# Keep only rows where FTs_1_delta is greater than 0
head(subset(dat, FTs_1_delta > 0))
```

```
   treatment                          nativity                 asian_ethnicity
2          0  I was born in another country:                          Filipino
4          0 I was born in the United States                           Chinese
6          0 I was born in the United States Indian, Pakistani or South Asian
7          1  I was born in another country:                           Chinese
8          1 I was born in the United States                          Japanese
13         1 I was born in the United States                           Chinese
         zlh_ethnic_lf_merged        zlh_intra_lf_merged zlh_gcb_merged
2             Somewhat agree           Somewhat agree The Republican
4             Somewhat agree           Strongly agree The Republican
6  Neither agree nor disagree           Strongly agree   The Democrat
7             Somewhat agree           Somewhat agree The Republican
8           Somewhat disagree         Somewhat disagree   The Democrat
13 Neither agree nor disagree Neither agree nor disagree   The Democrat
```

Anatomy of `subset()`

```r
# Multiple conditions with &
head(subset(dat, treatment == 1 & FTs_1_delta > 0))
```

```
   treatment                          nativity                 asian_ethnicity
7          1  I was born in another country:                           Chinese
8          1 I was born in the United States                         Japanese
13         1 I was born in the United States                          Chinese
14         1 I was born in the United States                         Japanese
16         1 I was born in the United States                           Korean
17         1  I was born in another country: Indian, Pakistani or South Asian
         zlh_ethnic_lf_merged       zlh_intra_lf_merged zlh_gcb_merged
7             Somewhat agree            Somewhat agree The Republican
8          Somewhat disagree         Somewhat disagree   The Democrat
13 Neither agree nor disagree Neither agree nor disagree   The Democrat
14 Neither agree nor disagree Neither agree nor disagree The Republican
16            Somewhat agree            Somewhat agree   The Democrat
17            Somewhat agree            Somewhat agree   The Democrat
```

Think of it as: "Give me only the rows where this condition is TRUE"

The assignment operator <–

The assignment operator `<–` saves your work by creating new objects.

Syntax: `new_object <– your_code`

Think of it as: “Take the result of this code and save it with this name”

The assignment operator <-

```r
# Save the subset as a new dataset
chinese_subsample <- subset(dat, asian_ethnicity == "Chinese")

# You can now use these new objects
head(chinese_subsample)
```

```
   treatment                          nativity asian_ethnicity
1          1  I was born in another country:         Chinese
4          0 I was born in the United States         Chinese
7          1  I was born in another country:         Chinese
9          1 I was born in the United States         Chinese
11         0  I was born in another country:         Chinese
12         0 I was born in the United States         Chinese
         zlh_ethnic_lf_merged zlh_intra_lf_merged zlh_gcb_merged
1               Somewhat agree      Somewhat agree   The Democrat
4               Somewhat agree      Strongly agree The Republican
7               Somewhat agree      Somewhat agree The Republican
9  Neither agree nor disagree      Somewhat agree   The Democrat
11              Somewhat agree      Somewhat agree The Republican
12              Somewhat agree      Somewhat agree The Republican
```

The `table()` command

The `table()` function counts how many observations fall into each category.

Syntax: `table(dataframe$variable)`

The `table()` command

```
# One-way table: count categories
table(dat$treatment)
```

```
  0   1
564 589
```

```
table(dat$asian_ethnicity)
```

```
                          Chinese                          Filipino
                              294                               253
Indian, Pakistani or South Asian                         Japanese
                              207                               158
                           Korean                     None of these
                               88                                 2
                            Other                         Taiwanese
                               18                                36
                       Vietnamese
                               97
```

Think of it as: "Count how many times each value appears"

The `table()` command

The `table()` function counts how many observations fall into each category.

Syntax: `table(dataframe$variable)`

```r
# Two-way table: relationships between variables
table(dat$treatment, dat$asian_ethnicity)
```

```
  Chinese Filipino Indian, Pakistani or South Asian Japanese Korean
0     138      127                                99       79     44
1     156      126                               108       79     44

  None of these Other Taiwanese Vietnamese
0             0    12        22         43
1             2     6        14         54
```

Think of it as: “Count how many times each value appears”

Individual notebook time

Go to: https://tinyurl.com/ps3-week3-zh

Share-out on the activity

AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

ARTICLE

Do Female Officers Police Differently? Evidence from Traffic Stops

Kelsey Shoub, Katelyn E. Stauffer, Miyeon Song

First published: 24 May 2021 | https://doi.org/10.1111/ajps.12618 | Citations: 34

Kelsey Shoub is Assistant Professor, Department of Political Science, University of South Carolina, Gambrell Hall 350, 817 Henderson Street, Columbia, SC 29208 (kshoub@mailbox.sc.edu). Katelyn E. Stauffer is Assistant Professor, Department of Political Science, University of South Carolina, Gambrell Hall 350, 817 Henderson Street, Columbia, SC 29208 (kstauffer@sc.edu). Miyeon Song is Assistant Professor, Department of Political Science, University of South Carolina, Gambrell Hall 350, 817 Henderson Street, Columbia, SC 29208 (misong@sc.edu).

Read the full text > PDF TOOLS SHARE

Abstract

Political scientists have increasingly begun to study how citizen characteristics shape whether—and how—they interact with the police. Less is known about how officer characteristics shape these interactions. In this article, we examine how one officer characteristic—officer sex—shapes the nature of police-initiated contact with citizens. Drawing on literature from multiple fields, we develop and test a set of competing expectations. Using over four million traffic stops made by the Florida State Highway Patrol and Charlotte (North Carolina) Police Department, we find that female officers are less likely to search drivers than men on the force. Despite these lower search rates, when female officers do conduct a search, they are more likely to find contraband and they confiscate the same net amount of contraband as male officers. These results indicate that female officers are able to minimize the number of negative interactions with citizens without losses in effectiveness.

Debugging tips

- Did I spell everything right?
- Am I referring to the right object?
- Did I run the code in the right order?
- Did I accidentally leave something (un)commented?

After today you should:

- Remember the three key questions to ask yourself when you open a new dataset, and know how to use them
- Feel comfortable using the key operations on Week 2 of the R Cheat Sheet
- Recall common mistakes when debugging your code
- Please fill out the feedback survey in your email(!!!)

Office hours!