

Reading In Data

Shiro Kuriwaki

Last updated September 13, 2018

By now, you should have:

- Downloaded R or Stata
- Understood what's what in a GUI

WHERE IS YOUR FILE?

Computer files (data, documents, programs) are organized hierarchically, like a branching tree. Folders can contain files, and also other folders. To load a dataset, you need to specify where that file is.

We denote the hierarchy of a folder by the / (slash) symbol. `data/input/2018-08` indicates the `2018-08` folder, which is included in the `input` folder, which is in turn included in the `data` folder.

Files (but not folders) have “file extensions” which you are probably familiar with already: `.docx`, `.pdf`, and `.pdf` already. The file extensions you will see for this course a lot are:

- `.csv`: A comma separated values file
- `.xlsx`
- `.dta`: Stata data
- `.sav`: SPSS data
- `.Rds`: R data single file
- `.Rdata`: R data (could be multiple files)

The goal of the rest of this document is to enable you to write code that will read in data files into your GUI.

READING IN R

Use the haven package. There is a function to read and write certain types of data files.

```
library(haven)
```

```
hh <- readRDS("data/input/2018-08/HHP_August2018_data.Rds")
```

You can call the name you gave it to browse it:

```
hh
```

For a discrete outcome like a poll question, the count function is quick and useful. Enter the name of your dataset and the name of the variable you care about.

To use the count function you need the dplyr package. Add `library(dplyr)` to the beginning of your code. Even easier, add `library(tidyverse)`, that will load dplyr and other related packages we will end up using.

```
library(tidyverse)
```

Now let's use count

```
count(hh, QM3AR1)
```

```
## # A tibble: 4 x 2
##   QM3AR1          n
##   <fct>        <int>
## 1 Strongly approve    414
## 2 Somewhat approve    381
## 3 Somewhat disapprove 244
## 4 Strongly disapprove 388
```

READING IN STATA

The command to read in a Stata file is use

```
use HHP_August2018_data.dta, clear
```

Actually, probably the harder part is getting your computer to find that file. In my case, it looks like

```
cd "~/Dropbox/survey-notes/data/input/2018-08"
```