Lecture 6. Tidyverse vs. Base R

Marcy Shieh

October 15, 2020

Final Project

- Start brainstorming your seminar papers
- Some ways to do this:
    - What question(s) am I interested in?
    - What are my independent and dependent variables?
    - How am I going to measure those variables?
    - What dataset(s) will I need?
    - What kind of methods will I need to use?
- Next few weeks:
    - You will be working on the data, methods, results, and discussion sections of your papers in this class
    - Assignments will force you to think about it *now* instead of the last two weeks of the semesters

Some tips for seminar papers

- Write an observational study with real data
  - It's tempting to just write a lit review or a research design if the class allows it. . . and never touch the paper again.
  - But you want to use your time wisely!
  - Do NOT opt for the long lit review
  - If your thinking is, "I'm going to write a research design but it's too ambitious for me to executive in the next six weeks. . . " That is an excuse. Do a proof of concept! If you want to do a survey but you don't have money to pay survey respondents right now, get a list of 100 undergrads and ask them to take it for free! If you want to interview people once covid is over, set up a few Skype interviews!

Ways I can help

- ▶ If there is a particular method you are interested in using but have no idea where to start, let me know!

- ▶ If there is an R package that you want to use but need some encouragement, let me know!

- ▶ I am not an expert in all methods and packages, but I love figuring things out. If I can help other people as I indulge in one of my favorite hobbies, it will make the rabbit hole much more satisfying!

But for now, we're going to talk about base R and tidyverse...

- ▶ There's a whole discourse
- ▶ People are passionate about this

What is base R?

- base R refers to basic functions in R, i.e., functions you don't have to install any packages for
- functions like plot(), table(), sort(), if()...all base R!

What is Tidyverse?

- a collection of packages developed by Hadley Wickham, the Chief Scientist of RStudio

- creates a layer of code on top of base R, changing its structure (just a bit, though)

- you *pipe* when you use tidyverse: %>%
    - I think of it as a power strip

- Why is Tidyverse useful?
    - it is great for data manipulation!

Do I have to pick one or the other?

- No, I use both...
- Some things are easier in base R, others easier in Tidyverse
- When you start using R more, you'll develop an intuition for what you think is easier
- But...this upsets me some people...

Replication materials

- ▶ What you do in the privacy of your R console is none of my business...

- ▶ BUT if you're not just "playing around" with data and you plan to make it replicable someday, you should choose one or the other

- ▶ I will show you base R and Tidyverse equivalents so you will be versed in both

- ▶ For the final project, PLEASE use Tidyverse

- ▶ I am not a dogmatic person but...

  - ▶ Tidyverse is just easier for *most* data cleaning tasks...

  - ▶ And it is good to be consistent with your coding, especially if the project can turn into something you will keep working on, add collaborators on, etc.