Programming day three: debugging, functions, and functional programming

Princeton Sociology Methods Camp

Angela Li[1]

Princeton University

August 28, 2023

[1] These slides are the collective effort of everyone who has contributed to the Princeton Sociology Methods Camp.

Feedback from last session

See Coding Feedback slides

Data we'll be working with

In-class lecture example: same data as day 1 (AddHealth data)

Data we'll be working with

In-class lecture example: same data as day 1 (AddHealth data)

Homework example: CDC data where parents report that a vaccine caused autism in their child

Data we'll be working with

Source: National Public Radio

When do we write our own functions?

Times when user-written functions, often in conjunction with the apply family, come in handy, are:

- **Repeat a process**: on day 1, we reviewed how to use a for loop to avoid having to copy/paste code to repeat some process; functions can provide a more efficient and flexible way to avoid this repetition
- **Transparency about what the code is doing**: R has a plethora of packages that have built-in functions for many things we might want to do- for instance, the dplyr functions we reviewed yesterday that help with summary statistics; functions that facilitate "bootstrapped" quantities of interest. But what these functions are doing internally can be a black box, so to make sure we understand what's going on, we may want to write the function ourselves

How to approach writing a function

- Ask yourself: "what problem am I trying to use this function to solve?"
- Once you've identified the problem, try writing code *outside of the function* to deal with a few *simple cases* of the problem
- Then, see what you can do to generalize the code from step two so that it can handle a variety of versions of the problem

Basic structure of a function

```
functionname <- function(argument1, argument2, argument3...)
{
what to do
return(what to return)
}
```

For each example, we'll discuss as a class:

1. What problem is the function trying to solve?
2. What are the function's arguments in this case? (we may also for shorthand refer to these as a function's inputs)
3. What is the function doing with those arguments?
4. What does the function return? What class is it? (feel free to check using R)

Live coding

Let's jump to our .qmd file to live code together!