To loop or not to loop?
A whistle-stop tour of iterating in R

Fiona Christie

R Glasgow Presentation

9th April 2019

My expe  ience (🙄)

- Data scientist for two years.
- Almost all analysis in R (with a smattering of Python).
- Worked with production code.
- Tested code using *testthat* package.
- Version control in git with GitHub.

Iterating

- Execute a function over a series of elements.
- Many ways to do this in R.
- Key aim is to avoid code duplication.
  - Secondary aim to avoid making your eyeballs sad.

for()

- Often what people reach for first, but often not the best way.
- Not as slow as they were historically.
- My main objection: they’re clunky and ugly.
- Hadley Wickham suggests they’re *too* flexible.
  - Not always clear what is going on.

Wait, do I even need to write a loop?

- No (usually).
- R is a functional style language.
  - R pushes you to build and use functions
- A functional is a function that takes a function as an input.

Apply family of functions

- What I typically reach for.
- apply, lapply, sapply, vapply, mapply, rapply and tapply
- I'm going to focus on the first three.
- In base R
  - Will briefly touch on a tidyverse equivalent later on.

apply(X, MARGIN, FUN, ...)

- X is an array or matrix
- Margin is 1 if you want to parse by row, 2 to parse by column.
- If you want to do something other than just print to the screen you'll have to save the output to a variable.
- This is a functional.

lapply(X, FUN, …)

- Takes almost any R object- data frames, matrices, lists, vectors, ….
- Returns objects as a list
- Useful if you have a collection of objects (e.g. a list of data frames) and you'd like to apply a function to each of them
  - For example, sanity check dimensions of each item in a list.

sapply(X, MARGIN, FUN, ...)

- Similar to lapply, but tries to simplify the returned object as much as possible- actually a wrapper function.
  - Not always predictable.
  - Best just to use it interactively.
- Worth noting that you can use the functions to execute rather than store.
  - E.g. draw a series of plots, read in files, print to the screen.

Tidyverse – map (in purr)

- Different functions depending on the type output you want.
- Idea: apply a function to each element in a vector.
  - Data frames are lists containing vectors of the same length.

Thank you!