R Tutorial

Andie Creel

January, 2023

Special thanks to Sam Maher who wrote an R Tutorial in 2018 that influenced this.

Goal

This is R Studio. The goal of this lecture is to see some basics about R so that we can dive into more exciting things tomorrow.

Overview of R Studio

Layout of RStudio:

- Script
  - Where you will be writing your own programs
- Environment/../Git
  - Mostly just environment and Git
  - Show which data objects you have loaded in memory
  - Eventually where you'll do version control with GitHub (last day)
- Files/Plots/Packages/Help/Viewer
  - Helps you load packages and other files to load
  - Default window when you're trying to get with a function from a package (we will get to this)
- Console/Terminal
  - where the code actually run
  - Scripts executes in console
  - Code disappear in the console, whereas a script saves your code
  - To run something in the console, type it in and hit "enter"
  - Important to know there is a "terminal" in R studio. Again, just know it's there.

Console

- We can type code straight into the console and run it.
- The console in R Studio knows your running R code.

```
# Type the following in the console, then press enter
2+3
```

```
## [1] 5
```

```r
# variable assignment happens with an arrow (on a mac can do option + -)
a <- 2
b <- 3

a + b
```

```
## [1] 5
```

How to write and execute a script

Writing Scripts and Running Code

- New Script: File > New File > R Script (or R Markdown, which is what I'm using now), or just the New Document script in the upper left hand corner of the screen > R Script (or R Markdown)
- Keyboard shortcut to run a section of code: highlight or put your cursor on that line and hit Ctrl + Enter (Windows) or Command + Enter (Mac)
- Executing code: the run button at the top right corner of the script
- For R Markdowns, we also have the Knit button at the top of the script.
  - That will run all code and "knit" it together into a pdf or HTML or doc
  - File type is specified at the top of the R Markdown file

R Markdown

- I wrote this pdf using an R Markdown (**show them quickly**).
- I love these because I'm able to write lots of notes to myself while I'm coding
  - it produces a nice shareable file
  - can share my notes, code and results with others.
- You can do the minipsets with a script or r markdown.

Mini psets

- If you have no experience with R, start with a script.
- If you want to do a R Markdown and am used to them, that's fine.

You can "clean up" the Environment after you've executed code by clicking the broom icon. **This will delete everything in your environment**.

Basic Data Types and COMMENTS

```r
# This is a comment, you can use '#' to write notes to yourself in your code
# - Comments are what make or break good coders, and coders who can collaborate with others.
# - If you ever thing you writing "too" many comments, you almost always not.
# - The things you think are obvious in your code wont be to others (nor yourself in a year when you ge

# Numeric -- integer: no decimal points
```

```r
myInt <- 1

# Numeric -- double: decimal points
myNum <- 2.4

# logical (Boolean): a true/false statement
myBool_1 <- (3 < 4)
myBool_2 <- (3 > 4)

# character (string)
myChar_a <- "a"
myChar_b <- 'b'
```

Ways to store datatypes

```r
# vector: can only be a vector of one data type (numeric, logical, string)
myVec_n <- c(1, 2, 3, 4, 5)
myVec_s <- c(str, "b", "c")

# matrix: can only be a matrix of one data type
myMat_n <- matrix(c(myVec_n,
                    6, 7, 8, 9, 10),
                  nrow = 2,
                  ncol = 5)

# Lists: Very powerful, but somewhat confusing. For now, just know they exist
myList <- list(2, "c", myMat_n)
myList[[1]] # returns numeric
```

```
## [1] 2
```

```r
myList[[2]] # returns string
```

```
## [1] "c"
```

```r
myList [[3]] # returns matrix
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1    3    5    7    9
## [2,]    2    4    6    8   10
```

```r
# data frame: can have multiple data types
myDF <- as.data.frame(myMat_n)
colnames(myDF) # these don't mean anything to me
```

```
## [1] "V1" "V2" "V3" "V4" "V5"
```

```
colnames(myDF) <- c("age_yr", "weight_lb", "income_$", "height_ft", "height_in")
```

Dataframes:

- Like matrices
- Can have different data types in each column
- Reference specific columns using the "$" operator, followed by the name of the column

```
# investigate one column
myDF$age_yr
```

```
## [1] 1 2
```

```
#create a new column
myDF$nonsense <- myDF$age_yr + myDF$weight_lb
```

- For the most part, you'll be loading new data by reading a csv
- You might have to create one at some point.
- By looking at how they're created we can get a better sense of what goes into them

```
# Create the data frame
BMI <-  data.frame(
   gender = c("Male", "non-binary","Female"),
   male = c(T, F, F),
   height = c(152, 171.5, 165),
   weight = c(81, 93, 78),
   Age = c(42,38,26)
)

# Try referencing one column
BMI$male # version 1
```

```
## [1]  TRUE FALSE FALSE
```

```
BMI[,2] #version 2
```

```
## [1]  TRUE FALSE FALSE
```

```
# Try referencing one row
BMI[1,]
```

```
##   gender male height weight Age
## 1   Male TRUE    152     81  42
```

```
# Try referencing one cell
BMI$height[1] # version 1
```

```
## [1] 152
```

```r
BMI[1,3] # version 2
```

```
## [1] 152
```

A word of caution

- Make sure you don't over write your variables by accident.

```r
# assigning new value to same variable (something to do carefully)
a <- 5
a <- a + 1 # If you run this line more than one, you will NOT get six
a
```

```
## [1] 6
```

```r
# assigning new value to new variable
a <- 5
a_new <- a + 1 # If you run this line more than one, you WILL get six
a_new
```

```
## [1] 6
```

Functions

Functions: once you have initialized them, they take in an input, perform a set of operations on them, and then give you some return value.

Example on board

- consider the function: myF(x) { y <- x + 3; return(y)}
- what does myF(3) return? 6

Points:

- These are helpful when you have something that you do often
- Rule of thumb: if you're copying and pasting code 3 times or more, make function
- (i say if you are going to copy past ever, because even if you think it'll only be twice it'll probably be more)
- Recent example for me:
  - wrote a function to take a date and return the season
  - Wrote a function to get kelvin and return Fahrenheit

```r
myF <- function(x){
  y <- x - x^2
  return(y)
}

myF(.5)
```

```
## [1] 0.25
```

```
myF(.25)
```

```
## [1] 0.1875
```

```
myF(.7)
```

```
## [1] 0.21
```

Loops

- for loops: iterates through a task for a set number of times
- Consider these loops (psuedo code):
  - For (i in 1 through 4) { print i }
  - For (i in 1 through 4) { print i / 4}
- Can be helpful when
  - Iterate through a column of data and do something to each row
  - Construct a new column and want to construct each row by scratch

```
#simple
for (i in 1:4){
  print(i)
}
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
```

```
# more involved
for (i in 1:4){
  print(i/4)
}
```

```
## [1] 0.25
## [1] 0.5
## [1] 0.75
## [1] 1
```

```
# combining loop and function
for (i in 1:4){

  y = myF(i/4)

  print(y)
}
```

```
## [1] 0.1875
## [1] 0.25
## [1] 0.1875
## [1] 0
```

```r
# manipulating a column (from above: version one of referencing a cell)
for (i in 1:length(BMI$Age)) {
  BMI$Age[i] <- BMI$Age[i] + 1 # everyone aged on year
}

# Making a new column
for (i in 1:length(BMI$Age)) {
  BMI$Age[i] <- BMI$Age[i] + 1 # everyone aged one year
}
```

If statements

- sometimes you want to execute a task ONLY if a certain condition is met.
- Open the BMI df:
  - Our RA did not record men's ages right
  - All men are actually 3 years younger than what's recorded
  - What would the correct DF look like?
- If statements let you fix a mistake like this
- Also demonstrates why the Boolean (true/false or indicator) variable is so powerful

```r
# goes through each row and changes age if someone is male
for (i in 1:length(BMI$male)) {

   if (BMI$male[i] == TRUE) {
    BMI$Age_new[i] <- BMI$Age[i] - 3
  }else{
    BMI$Age_new[i] <- BMI$Age[i]
  }
}
```

Other R Tutorials

Princeton Getting Started with R

UCLA Getting Started with R

Some specific packages

We haven't covered packages yet, but a few good resources for the tomorrow.

ggplot

dplyr and tidyr