Week 1 Lab: Introduction to R

STAT 3021

Class information

TA: Ganghua Wang Email: wang9019@umn.edu

Website: https://keywgh.github.io/teach/STAT3021_LAB

Office Hour: TBD

Location: 495 Ford Hall

Requirements

- Please refer to Syllabus.
- To receive full credits, no late submission. All homework must be turned into Canvas no later than 11:59 pm.
- If you have any questions about homework grade, please contact instructor/grader.

What's R

R is a free statistical programming language used in statistics, machine learning, and data science. R studio is a user-friendly environment for R.

Students in STAT 3021 will learn R from lectures and lab. Students should use R to complete their miscellaneous assignment 2 and practical assignment (more information will be available on the first day of lecture). Students should bring their own laptops to lab sessions.

Installation

Please click on the link below to learn how to install R and R studio if you haven't.

https://youtu.be/d-u_7vdag-0

Getting started

After installing both R and R studio, you can start R studio. There should be three windows; *Console* (left window), *Environment* (top right window), *Files* (bottom right window).

Console is where any work is processed and shows outputs. *If there is any errors in your R command, the error message will appear in the console.*

Environment shows datasets and objects created. You can click on 'History' tab to view previous R commands.

Files show any files in your working directory by default. Notice other tabs such as Plots, Packages, Help, etc.

R as a calculator

Copy and paste or retype each line of the following R commands in your console and hit [Enter] and predict what R will produce.

```
9+7
sqrt(16)
abs(-3)
3^2
3**2
3%2
```

Question: Did you see any error message? Fix the command with the error so it calculates 3 divided by 2.

Creating objects

You can use either '<-' or '=' to create objects.

```
x<-1
y=3
xy<-x*y
```

Two commands above create two objects 'x' and 'y' and 'x' is defined as 1 and 'y' is defined as 3. 'xy' is defined as x multiplied by y. **You can check 'Environment' Window (top right) to see all three objects.**

Try each line of commands below.

```
print(x)
x
y<-10 #now y is 10
y
```

Note that both print(x) and x return the value of x.
When you type 'y<-10', R console does not print the value of y.

Note that R is case-sensitive; hence each of the following commands will produce an error message *Error: object 'X' not found.*

```
print(X)
X
X+3
```

Question: If you type 'x+y' in the Console, what will you see?

```
x+y
```

Question: What is the value of z below?

```
z<-x/y
z
```

Question: Copy and paste the command below into your Console. What output, if any, do you see? Why?

```
X+Y
```

To remove objects ..

```
rm(x)
rm(y, z)
```

Different types of objects

When you want to store more than one number/character to an object, you can create a vector. Suppose you have four students' names and their Exam 1 scores.

```
names<-c("Adam", "Ben", "Chuck", "Dan")
exam1<-c(78, 51, 60, 90)
```

Whenever you want to create a vector, you need 'c()'. The letter 'c' stands for 'combine' or 'column'.

Built-in Functions in R

Suppose you want to find the average (mean) of those four students scores. You can use built-in function 'mean()'.

```
mean(exam1)
```

There are many built-in functions in R. We will learn common ones in class. Certain functions take only a particular type of input. For instance, mean() only works when the input is a number or a vector of numbers.

Question: What will happen if you run each of the following commands? What error messages do you get? Why?

```
mean(names)
mean(A)
```

To find the minimum exam score, use the following command.

```
min(exam1)
```

'min()' is the function that finds the minimum. Similarly 'max()' finds the maximum.

Try it yourself: What is the 'minimum' name of the fours students (Adam, Ben, Chuck, Dan).

Other useful and important tips

Use 'R Stript (Ctrl+Shift+N)' and SAVE your work.

Commands that you typed in Console cannot be saved as a file.

By writing commands in R Script, you can save your work and open the strip file later and continue working on your project. **You can run your commands by highlighting the command and clicking "Run" button, or by hitting [Ctrl]+[Enter] in Window PC or by hitting [Command]+[Enter] in Mac**.

Use Comments

By using '#', you write comments.

```
# This is a comment line

# Homework 1 Problem 1
mean(c(12, 4, 25, 0))
```

Anything you write after '#' is a comment line and R will not run that line. A comment line is used to describe what you are doing.

Help

In R, using 'help(**function**)' or '?', you can open help document of the function.

```
help(hist)
?hist
```

When you are not sure the exact function name that you need, you may use '??'.

```
??histogram
```

Practice Using R

Want more practice? (i) You can try data camp! (More information available on Canvas) or (ii) you can also download the "swirl" package.

Swirl is an interactive coures that teaches you how to program in R.

```
## Install package
install.packages("swirl")
```

First try the "R Programming" lesson

List of lessons:

- R Programming
- Data Analysis
- Regression Models
- Getting and Cleaning Data
- Statistical Inference

```
## Call package
library(swirl)
```

Once the lessons are installed...

```
# Activate interactive course
swirl()
```

Follow the prompts and have fun!

Questions regarding Swirl: swirlstats.com

When you are done with swirl, hit '[Esc]' to exit.

Summary

What you need to know:

- Successfully download and run R and RStudio which are compatible with your operating system.
- Basic operations in R: addition, etc; create objects.
- Bulit-in functions.

- (Most important) How to get help: Google, help(), ?.

Good to know:

- R script.
- Home/working directory.