Statistics 260: R for Data Sci

Lecture 1: Introduction to R and Getting Started

Brad Mcneney

Introducing R

What is R?

- R is an open-source environment for statistical computing and graphics.
  - An implementation of the S language (https://en.wikipedia.org/wiki/S_(programming_language))
- Started in the mid-1990's by Ross Ihaka and Robert Gentlemen at Auckland University
- Now maintained by a team of experts called the R Development Core Team
- A "packages" system allows any user to bundle R code, data and examples together.
- R and R packages are distributed through the Comprehensive R Archive Network (CRAN).
- SFU has a CRAN mirror at http://cran.stat.sfu.ca

What does "environment" mean?

- R is a fully-functioning programming environment with all the usual constructs, such as
  - conditionals (if-then-else),
  - loops
  - user-defined functions.
- In addition there are built-in facilities for
  - data input, storage, manipulation, and output
  - optimization, matrix computation, etc.,
  - random number generation,
  - data analysis and graphics.
- "Base" R is good, but it is the package system that makes R great.

R packages

- The R package system is the key to R's success
- It has allowed statisticians and other data scientists to implement and distribute their work to be used by others.
- The R package system enforces some rules about how packages are structured, but differences in programming styles of package authors mean different interfaces
  - Users need to be aware of different data structures for input, output and of different styles of graphics

Course objectives

Course objectives

- Learn how to use R for common tasks in data science, with a focus on tools from the "tidyverse".
  - https://www.tidyverse.org/

Getting started

Getting started with R, RStudio and git

- A brief set of instructions are on the canvas page https://canvas.sfu.ca/courses/48263/pages/getting-started-with-r-rstudio-and-git
- Please try to get R and RStudio installed and create an RStudio project linked to the class GitHub repository as soon as possible.
- Those still having trouble after the weekend can ask for help in their Stat 261 class next week.

R reference cards

- RStudio has created several useful reference cards, called cheat sheets, and have collected cheat sheets from other R users.
- See https://www.rstudio.com/resources/cheatsheets/
- For getting started with RStudio you might find the following helpful: https://github.com/rstudio/cheatsheets/raw/master/rstudio-ide.pdf

Starting R

- Start R by starting RStudio.
- The "Console" window is where you can type your commands.
- However, it is good practice to open an R script, type your commands in the script, and then submit the commands to the R console.
  - `Session -> Set Working Directory` to set the working directory
  - `File -> New File -> R Script` to open a new R script
  - type your commands into the script
  - put your cursor on the line you want to submit and hit `Ctrl-enter`
- Save your script for later use.
- More on the RStudio interface at https://support.rstudio.com/hc/en-us/sections/200107586-Using-RStudio

Reading

- Chapter 1 of the text:
  https://r4ds.had.co.nz/introduction.html