DSCI 210: Final Exam Review Sheet

Comments

- This exam will cover only content from R
- The in-class portion will be worth 50 points, the take-home portion is worth 100 points
- The goal of this exam is to measure your understanding of R – not necessarily "how to use" R.

Topics

Basics

- Understand the difference between R and R Studio
- Know how to read in data using the Import > Dataset, understand the basic options that can be specified when reading in data using this method, i.e. Heading, Separator, NA, etc.
- Understand the basic data structures in R, i.e. vectors, data.frames, and lists
- Understand best practices in using the help() function

Manage Data

- Know how to create a basic vector, data.frame, or list
- Know to refer to particular elements of a vector or data.frame, e.g. what does x[2] do? Or xy[2,3], or xy$variable1
- Know to subset a vector or data.frame, e.g. xy[,3:5] or xy[c(1,3,5), 3:5]
- Understand how to understand the structure of a data.frame, e.g. can you understand output from the str() function.
- Understand the usage of the functions like as.numeric(), as.character(), or as.date()

Summarize / Plotting Data

- Know how to use basic summary functions, e.g. mean(), median(), length(), etc.
- Know how to use more advanced summary functions, e.g. table(), summary(), aggregate(), apply(), by()
- Understand the basics of plotting in R, e.g. graphics window, plotting functions, adding additional points or lines to existing plots
- Understand how to modify basic graphical parameters, e.g. adding a title, axes labels, edit range of axes, change color or type of plotting character, etc.
- Understand the basics of the lattice() package

Function Writing

- Know how to create a basic function in R
- Understand how functions use their local variables, default options, etc.
- Understand how objects are returned by functions, i.e. by using the return() function

Jackknife & Cross-Validation

- Understand why one might use the jackknife procedure
- Understand how to implement a simple jackknife procedure in R
- Understand the purpose of cross-validation
- Understand how to implement a simple cross-validation procedure in R

Remaining Topics

- Understand the purpose of the Twitter API
- Understand what types of summaries can be used for Twitter data, e.g. plotting over time, plotting by ScreenName, word frequency analysis, wordclouds, etc.
- Understand the basics of web scraping via the rvest() package
- Understand the purpose of the SelectorGadget Chrome add-in that was used in class
- Understand the basics of direct file download, i.e. download.file()