

Stats/CME 195 -- Spring 2016
Summary of topics covered

1. R -- history, popularity, applications, script languages generally (first class)
2. language fundamentals (3/29--3/31)
 - 2.1. data structures (3/29)
 - 2.1.1. constants, variables, functions, types
 - 2.1.2. R's type system
 - 2.1.3. atomic types (numeric, integer, character, factor, etc.)
 - 2.1.4. complex types (vector, matrix, dataframe, list, etc.)
 - 2.1.5. Side topics: basic graphics (plot, histogram, barplot, boxplot etc.) and basic data analysis (table, statistical functions like mean, var, etc.)
 - 2.2. Imperatives
 - 2.2.1. flow control -- branching and iteration (3/31)
 - 2.2.1.1. vectorized operations
 - 2.2.2. user-defined functions and abstraction (3/31, 4/5). scoping rules, argument passing.
 - 2.2.2.1. Ex: Numerical integration. Side topics: random variable series of functions, matplot, saving and loading data to and from disk (4/5)
 - 2.2.2.2. Ex: Clinton emails (4/5, 4/7)
 - 2.3. getting help: manual pages and online resources
3. loose ends (4/7) -- factors and lists
4. At this point in the course I began organizing lectures around programming tasks. I introduced R concepts as needed and in the context of completing a task.
 - 4.1. Manipulating a dataset (4/7). Topics along the way: handling CSV format, random sampling, sort routines and sorting a dataframe by a column, split and lapply, row and column names, vectorizing functions
 - 4.2. The last task led to a diversion on text processing, in order to explore the Clinton emails and Ontario baby names data set further (4/7, 4/12)
 - 4.2.1. basic string manipulation -- nchar, paste, print, cat, strsplit
 - 4.2.2. more complicated tools -- grep, sub/gsub, substr
 - 4.2.3. regular expressions
 - 4.3. Estimating the # of videos on Youtube, exploring the distribution of view counts of Youtube videos. Topics along the way: data scraping (html, browser tools), more regex's, scan, anonymous functions, data transformations (4/14)
 - 4.4. Plotting thematic maps. Topics along the way: S4 classes, DB joins, cut, color palettes (4/19)
5. Final class--ggplot, dplyr intros (4/21)

text manipulation -- stringr

database manipulation -- plyr, dplyr

graphics -- ggplot2

Aside: my programming language learning analogy. Like learning a natural language, it's much easier to learn when you're immersed in it, copying patterns and acquiring new words that seem useful. You can only go so far using a grammar and lexicon. On the other hand if you're immersed too early it's no good. An adult foreigner can spend forever in a new place without ever learning the language. The goal up to the 10/6 class was to give you the rudiments of the language so you could become productive and start learning things in a context in which they're motivated, and you can start interfacing to or adapting others' code as needed.