Title: Basics of R Programming: Part 1
Date/Time/Location: [[***]]
Instructor: Haben Michael, PhD Candidate, Dept. of Statistics (haben.michael@stanford.edu)
Prerequisites: If you would like to follow along on your own computer, please have R installed. Detailed instructions for installing R for Apple, PC, or Linux are available here. Please feel free to contact me at the above email address with any questions. No experience with R or programming will be assumed, although some exposure to programming concepts may be helpful.

Description: We will start with the basics: interacting with R, R's data types, and built-in functions. Next we will survey exploratory data analysis, graphics, and getting data in and out of R. The goal will be to tour enough features of R that attendees can later pursue topics in greater depth on their own.

Course Outline:

- introductory remarks about R -- history, popularity, applications, competing languages
- interacting with R -- REPL and scripts
- representing and manipulating data ("nouns" of the language)
    - constants, variables, functions, types
    - R's type system
    - atomic types (numeric, integer, character, factor, etc.)
    - complex types (vector, matrix, dataframe, list, etc.)
- built-in functions ("verbs" of the language)
    - calling functions, parameters
    - descriptive statistics (table, mean, var, etc.)
    - regression and classical statistical tests
    - random routines
- basic graphics
    - scatterplots (aside: polymorphism)
    - histogram, barplot, boxplot
    - graphics options (color, legend, mfrow, etc.), model formulas
- getting data into and out of R
    - importing/exporting locally/remotely
    - memory dumps

—

Title: Basics of R Programming: Part 2
Date/Time/Location: [[***]]
Instructor: Haben Michael, PhD Candidate, Dept. of Statistics (haben.michael@stanford.edu)
Prerequisites: Minimal exposure to R, at the level of the first workshop. However, this workshop won't rely heavily on the first and I will try to review concepts when I do refer back.

Description:R is a powerful platform for statistics and graphics. Part 2 will have 3 parts. First, we will look at how R can be used as a programming language. Second, we will look at basic string manipulation and text processing, which is needed for such varied tasks as data scraping and

manipulating genome data. Finally, we will tie everything together by using what we've learned to answer realistic, useful questions.

Course Outline:

- Programming concepts
    - Flow control -- branching and iteration
    - user-defined functions
    - vectorized operations
- text processing
    - basic string manipulation -- nchar, paste, print, cat, strsplit
    - more complicated tools -- grep, sub/gsub, substr
- applications: perform specific tasks involving the following tools: 1) getting data into R, 2) manipulating data structures, 3) vectorized operations (or loops) 4) text processing, 5) plotting
    - investigate frequency of a given name over time
    - investigate frequency of given names ending in "a" for boys vs. girls
    - investigate relationship between word length and word count

References

- The Quick-R website is a good mix of concise tutorial pages and reference material. http://www.statmethods.net/.
- Modern Applied Statistics with S-Plus by Venables & Ripley (available on-line through Stanford) is an authoritative but introductory textbook. The book gives some background on statistical methods, besides showing how to implement them in R.
- For specific questions, the goto resource used to be the official R mailing list, which has searchable archives covering many years of questions. If you have an R question, almost certainly someone has asked it on the list. Nowadays, mainly due to stackexchange, I would simply use Google, which ranks stackexchange sites and the R mailing list highly. An added benefit of stackexchange is that oftentimes questions about R are closely related to questions about data analysis or statistics, and these are all appropriate topics for stackexchange.