

Tutorial: A Crash Course in R Programming

Olivia Lau

Google Inc.
Olivia.Lau@Post.Harvard.Edu

Keywords: statistical computing, lexical scoping, programming, profiling

Abstract

This course will cover common problems in R code, and offer techniques to make code more efficient and less prone to user error. By the end of the course, participants should be able to write efficient statistical software and validate existing code.

Outline

Topics will include:

- 1) Overview of R (20 minutes)
 - a) Units: Numeric, integer, character, and factor
 - b) Data structures: Vector, matrix, array, list, and data frames
 - c) Manipulating data structures
- 2) Writing efficient R code (40 minutes)
 - a) Writing (and avoiding) loops
 - b) Vectorizing over matrices, arrays, and lists
- 3) Writing functions (20 minutes)
 - a) Environments
 - b) Passing arguments and proper lexical scoping
- 4) Making R code easy-to-use (40 minutes)
 - a) Implementing model functions in your code
 - b) Generic functions
- 5) Validating statistical models (30 minutes)
 - a) Monte Carlo simulation
 - b) Unit tests
- 6) Making R run fast (30 minutes)
 - a) Profiling R code
 - b) Calling compiled languages (C, C++, Fortran) from R

Intended Audience

UseRs (and programmers from other statistical languages) who are interested in learning how to write fast, easy-to-use R code.

Required Knowledge

Familiarity with a command line-based statistical package (e.g., R, SAS, etc) is required. Instructions on installing R and an R editor will be distributed via email prior to the course.

Disclaimer: The views expressed in this presentation are those of the presenter and not necessarily of Google Inc.