

Introduction to R for the Novice Programmer
New Mexico State University
May 16 to 27, 2016

Description: Application of computer programming language R to analyze genomics and transcriptomics data sets. R data types. File input/output. Iterative programming. Fast search algorithms. Visualization. Six 2-hour lectures and six 1.5 office hours.

Learning objectives

- Write R script to manipulate table data
- Perform basic file input and output
- Write efficient programs to access data sets in gigabytes with millions of records
- Visualize genomic data

Instructor: Dr. Joe Song, joemsong@cs.nmsu.edu, Phone: 575-646-4299, Office: SH 141.

Office hours: 1:30-3:00pm Mondays, Wednesdays, and Fridays or by appointment.

Class schedule: 10am-12pm Mondays, Wednesdays, and Fridays in TBA.

Textbook

Matloff, N. (2011). *The Art of R Programming: A Tour of Statistical Software Design*. No Starch Press.

Topics

Day	Programming concepts	Application	Reading	Project
1	Data types (vector, list and data frame) and functions	Introduction to human genome annotation file	Ch1.1, 1.2, 1.3, 1.4, 1.6, 1.7; Ch5	Part 1
2	File input/output and string parsing by regular expression	Reading GENCODE file	Ch10.2.1; Ch11.1, 11.2	Part 2
3	Iterative programming: for-loops	Search in GENCODE	Ch7.1, 7.2	Part 3
4	Binary search and run time comparison with linear search	Fast search in GENCODE	Wikipedia: binary search algorithm	Part 4
5	Visualization	GENCODE and FANTOM5	Ch12	Part 5
6	Student project presentation			