

**Introduction to R for the Novice Programmer**  
**New Mexico State University**  
**June 15 to 26, 2015**

**Description:** Application of computer programming language R to manipulate large data set containing digitized information from modern molecular biology. R data types. File input/output. Iterative programming. Fast search algorithms. Visualization. Application to access human genome annotation. Six 2-hour lectures and six 1.5-hour office hours.

**Learning objectives**

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

**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 Skeen Hall Room TBD.

**Textbook**

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

**Topics**

<b>Day</b>	<b>Programming concepts</b>	<b>Application</b>	<b>Project</b>
1	Data types: vector, list and data frame	Introduction to human genome annotation file	Part 1
2	File input/output and string parsing by regular expression	Reading GENCODE file	Part 2
3	Iterative programming: for-loops	Search in GENCODE	Part 3
4	Binary search and compare run time with linear search	Fast search in GENCODE	Part 4
5	Visualization by plots	Summarizing information from GENCODE	Part 5
6	Student project presentation	GENCODE	