CSB1020HF, LEC0142
Introduction to R

Offered by the *Centre for the Analysis of Genome Evolution & Function (CAGEF)*

Instructors:
Dr. David S. Guttman, CSB, CAGEF david.guttman@utoronto.ca
Dr. Calvin Mok, CAGEF Bioinformatics calvin.mok@utoronto.ca

Time:
September 4 – October 16 (7 weeks)
Wed, 1:00 - 4:00pm
ESC3087

Enrollment:
20 graduate students
Auditor spaces based on availability

Course Objectives

This is a beginner's introduction to R and the RStudio environment for individuals with no prior experience or background. Individuals who complete the course will be able to:

- Work with the RStudio and Markdown Notebook environment and navigate the R programming language.
- Understand data structures and data types.
- Import data into R and manipulate data frames.
- Transform 'messy' datasets into 'tidy' datasets.
- Make exploratory plots as well as publication-quality graphics.
- Use string searching and manipulation to clean data.
- Perform basic statistical tests and run a regression model.
- Use flow control and build branching code.

Throughout the course we'll work with a set of data that takes us through the various steps of analysis from importing to data wrangling to statistical analysis and visualization. Each class will consist of a short introductory section followed by 'code-along' hands-on learning that will gradually build up the lecture's topic(s). Students are expected to have access to a computer during class and are encouraged to ask questions while coding along with the instructor. A homework assessment will be assigned after each class to reinforce the skills learned and a final project will test overall knowledge and application. The course materials will be provided through Quercus and lectures will be held in-person using the University of Toronto JupyterHub and its RStudio server.

Course Availability

This course will be held in-person (unless otherwise determined) and will be available to graduate students in CSB and EEB. Auditor spaces will be based upon available space to postdocs, staff, and faculty, although only registered students will be evaluated. The course will count as a single module (0.25 credits) for CSB graduate students. All graduate students interested in taking the course for credit should enroll through ACORN.

Anyone wishing to audit the course should fill out the request form at
https://forms.gle/JAJFK76vKcWA6kKD7

Evaluation

Item	Note	% Mark
Completed R Markdown Notebook	7 lectures x 2% each*	14%
Homework Assignments	6 weekly assignments x 6% each	36%
Term project	Due 2 weeks after the end of the course	50%

* a 3.5% bonus (0.5% per lecture) will be awarded for submitting notebooks within 24 hours of lecture completion.

Pre-requisites: Access to a laptop computer. No prior programming experience required.

Reference Material: *R for Data Science* (http://r4ds.had.co.nz/)

Syllabus

Class	Topic
1	**Introduction to R, RStudio and R Markdown Notebooks**: R and R Markdown basics, best coding practices, functions and syntax, data types and structures, mathematical operations with R objects, installing R packages, getting help.
2	**How to read, write, and manipulate your data:** Importing text and Excel files, the dplyr package and functions to manipulate tabular data.
3	**Introduction to Tidy Data**: Wide versus long data formats, reshaping data with the tidyverse package.
4	**Data visualization with ggplot2**: The grammar of graphics; scatter, line, box, bar, and density plots, among other types of graphics.
5	**Data cleaning with regular expressions (RegEx):** Introduction to RegEx; inspecting, cleansing, and data wrangling using RegEx; classes, quantifiers, operators, pattern-matching, and string manipulation.
6	**Linear regressions**: Simple and multiple linear regressions, ANOVA, ANCOVA, model selection.
7	**Flow control**: for loops, conditional statements (if, while, repeat, next, and break); troubleshooting loops.

Subject to change