Title: R for Epidemiologists

Prerequisites: None

Credit Hours: 2

Semester: Fall, 2016

Time & Location: MW 10:10-11:00, 2303 McGG

Instructor of Record:
M. Alan Brookhart, PhD
Professor, Epidemiology
2105F McGavran-Greenberg CB #7435

Facilitators:
Xiaojuan Li, MSPH xli@live.unc.edu
Nat MacNell, MSPH macnell@unc.edu
Mike Fliss, MS mike.dolan.fliss@gmail.com

Purpose: This course is an introduction to the R language and environment for statistical computing and graphics. The use of *R base* software for data management and analysis is emphasized, along with a few specialized packages.

Lectures: See course schedule

Office Hours: For an appointment, email Nat or Xiaojuan

Learning Objectives: The objectives of this course are:

1. To familiarize students with R syntax and elementary R programming concepts in preparation for working with R code in advanced epidemiology courses and in real research settings.
2. To learn traditional approaches to data management and other methods made possible by the R language.
3. To gain practice solving complex programming problems in the context of epidemiology by writing new code or by identifying, configuring, troubleshooting, evaluating, and adapting user-submitted R code from the Comprehensive R Archive Network (i.e. R packages).

Texts & Resources

Course notes are provided online at sakai.unc.edu.
The following free resources are suggested:

1) *R for Beginners* (https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf)
   A free, short book on elementary R programming
2) Online "Live" Introduction to R (free course):
   (https://www.datacamp.com/courses)

Grading

10% Attendance
40% Homework problems
20% Project milestones (turned in w/ homework)
30% Final project

Final average	Undergraduate Grade	Graduate grade
95 or above	A	H
85-94	B	P
75-84	C	P
70-74	D	L
Below 70	F	F

Undergraduate + and – grades will be decided at the discretion of the instructors, based primarily on the final project.

Homework

In addition to the project, there are five homework assignments that cover the basics of epidemiologic analysis. These assignments will cover the same topic areas as the project: (1) nuts & bolts of R, (2) functions, loops and subsetting, (3) data management, (4) graphics, and (5) modelling. To turn in homework, submit the homework assignment word document with answers inline, with your code pasted in at the end.

Final Project

Students will be asked to complete a data analysis project as part of the course. Students will apply R programming skills to a dataset of the student's choice and follow the steps typically used in analysis projects. Guidance for these steps is found in each homework assignment. These exercises are turned in with the homework assignments but are graded for completion only.

As a capstone, students will also be asked to complete a special analysis of their choice using the dataset. These special analysis techniques (like GIS or advanced modeling) are covered in the second half of the class, but do not have required homework exercises. These analyses should use at least one R package beyond base R, either a package discussed in class or another of the students' choosing.

Suggested datasets will be available for students without their own. Students must have individual project topics, but are encouraged to support each other by sharing progress and collaborating where possible. More information about the project is available on Sakai. At the end of the course, students will combine their work on the project exercises with this special analysis to create a research poster to show their project.

Collaboration and the UNC Honor Code

All academic work in this course, including homework, exams, and projects, is designed to be collaborative - a process that combines a student's new work with existing work created by others. Indeed, the R language itself is a collaboration between many authors. In this context, it is important to correctly attribute credit to sources used in the course of students' work and to accurately identify the student's individual contribution; representing others' work as your own is a violation of the UNC Honor Code. Copying solution code for a specific problem verbatim (e.g. from a fellow student) defeats the educational purpose of the course, as does posting the problem on a community programming forum (e.g. Stack Exchange).
Students are strongly encouraged to consult their peers and any resources they can find (including programming forums) for other educational purposes, including but not limited to (a) learning about programming concepts that underlie problems or solutions, (b) finding examples of similar or simpler problems with ready solutions that can be adapted to the problem at hand, and (c) reviewing solutions to unrelated problems to learn about the process of finding a solution (i.e. troubleshooting programming errors).