# EPID 799C Syllabus

**Course Title:** R for Epidemiologists

**Prerequisites:** None

**Credit Hours:** 3

**Semester:** Fall, 2018

**Time & Location:** MW 10:10-11:25, Room MC 1303.

**Course Website:** <http://learnr.web.unc.edu/>

**Course Email:** All course communications will take place via the UNC\_EPID799C\_Fall18 google group.

**Instructor of Record:** Steven R. Meshnick, MD, PhD

Professor, Epidemiology

3301 Michael Hooker Research Center CB #7435

**Facilitators:** Mike Fliss, MSW mike.dolan.fliss@gmail.com

Sara Levintow, MSPH levintow@email.unc.edu

Nick Brazeau, nbrazeau@med.unc.edu

**Purpose:** R for Epidemiologists is a 3-credit class intended to be the most effective and efficient way for UNC Epidemiology students to establish a foundation in the R programming language, RStudio IDE, and functional programming modalities. We give special attention to R topics and packages relevant for epidemiological data management, analysis, and visualization. Homework assignments are designed to ease epidemiology students into the language efficiently by building on the UNC Epidemiology core curriculum, using familiar datasets and questions seen in other classes. In addition to homework, each student completes a project of their choosing to incorporate R programming into their existing research or coursework. There are no prerequisites, but students who have taken EPID 716 will benefit from their past assignments, and students who have taken core epidemiology courses (EPID 710, 715, etc.) will benefit from having been introduced to fundamental epidemiology concepts used in this course.

**Learning Objectives:** The objectives of this course are:

1. *Learn Syntax:* To familiarize students with R syntax and programming concepts in preparation for working with R code in advanced epidemiology courses and in real research settings.
2. *Introduce Functional and Iterative Programming:* To gain practice solving complex programming problems by writing new code, adapting existing code, and using pre-built packages in the R framework.

**Lectures:** See course schedule on website.

**Office Hours:** Course facilitators are available by appointment, and a schedule for office hours will be announced early in the semester.

**Texts & Resources:**The course makes extensive use of R, a powerful, free, open-source, and community-supported statistical programming language that connects well with growing data science principles.

Students will find instructions to install R, RStudio, and the course website:

<http://learnr.web.unc.edu/r-for-epi-workshop/>

The following free resources are suggested:

1. Installing [R & Rstudio](https://docs.google.com/document/d/1y42G5uDiUzzDfuFIxsC4BjotiEw9YQSPL_58KATMbv0/edit#heading=h.glvfq4g3ave)
2. *R for Beginners* (<https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf>)

A free, short book on elementary R programming

1. *R for Data Science* (<http://r4ds.had.co.nz/>)

A likewise free book on more advanced programming concepts and packages

1. Online “Live” Introduction to R (free course): (<https://www.datacamp.com/courses>)

NOTE: Interested students can receive free, full access to Datacamp for the duration of the class

**Grading**

10% Attendance

50% Homework assignments

10% Project proposal

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:** There are five homework assignments that cover the basics of epidemiological analysis. The assignments broadly investigate the relationship between early prenatal care and preterm birth, using the NC Births Dataset that epidemiology students have previously worked with in EPID 716. Homework assignments are posted on the course website. To submit homework, email the course facilitators the Microsoft Word assignment document with (1) answers inline and (2) code pasted at the end of the document.

**Project Proposal:** Students will incorporate R programming into their existing research or coursework through a project of their choosing. Halfway through the semester, students will email the course facilitators a brief description of their project topic and proposed use of R programming.

**Final Project:** Students will complete their proposed project over the last month of the course, concluding in a final presentation. Some class time will be dedicated to project work sessions with assistance offered by the course facilitators. Students must have individual project topics, but are encouraged to support each other by sharing progress and collaborating where possible. Final presentations will be brief and provide a summary of the project methods and findings. More guidance will be given later on in the semester.

**Collaboration and the UNC Honor Code:** All academic work in this course, including homework assignments, 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 or homework keys from a previous year) 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 educational purposes, including but not limited to (a) completing homework assignments as a team, (b) learning about programming concepts that underlie problems or solutions, (c) finding examples of similar or simpler problems with ready solutions that can be adapted to the problem at hand, and (d) reviewing solutions to unrelated problems to learn about the process of finding a solution (i.e., troubleshooting programming errors).

**Statement on learning and teaching:** The teaching team acknowledge that not everyone learns the same, and not all classrooms are welcoming to all. The ideal balance of more or less lecture, group work, individual work, homework, practical application, and theory differ not just by individual preference, but learning style and ability. Further, these learning identities are not entirely separate from our own identities, experiences and diverse cultures. A profession or institution’s culture, including that of epidemiology and education institutions like our own, can both explicitly and implicitly preference some cultures and identities over others (e.g. : <http://www.dismantlingracism.org/uploads/4/3/5/7/43579015/whitesupcul13.pdf>). We therefore acknowledge ways that education settings can be gendered and racialized, and aspire to create a space that can not only accommodate, but draw on strengths from these diversities where possible. However, this is non-trivial, deep work - for now, concretely, we aim to balance traditional (note to reader: which tradition?) classroom lecture dynamics with small group activities and coding together, and supporting more introverted and more extroverted learners by balancing talking time in class. We ask for your help and mindfulness to achieve that balance. Beyond this, students are encouraged to bring their full realities into the classroom, as questions, statements, or otherwise, at any time about the applicability and limits of course content. Lastly, there is no hard, underlining binary between the teaching team and students. We are students ourselves, so welcome feedback to improve the course to better meet the needs of the diverse community of public health aligned epidemiologists and statisticians who seek to learn R … ultimately to improve community health.