GEOG 5680/6680 Introduction to R – Summer 2021 (online edition)

Professor Simon Brewer
simon.brewer@geog.utah.edu

Course description: Over the past few years, the R programming language has become increasingly widespread in academia and industry for statistical analysis, bioinformatics, data mining and graphics. This increase in its use means that many people have encountered it through journals, websites, coursework and colleagues, but may have little training in the basics of the language and are unaware of the full set of possible applications.

This course is designed to give an intensive introduction to R as a programming language and as a graphical tool. We will also cover some basic analysis, although the focus will largely be on the language itself. It is aimed at people with little to no prior experience with R or programming, although some basic knowledge of statistics is assumed.

Course goals: Students completing the course will have experience in:

- One of the most commonly used and generally applicable software packages for statistical analysis
- Manipulating data in R
- Programming in R including functions and control loops
- The wide range of graphical tools for making figures
- Basic statistical model building in R
- Extending base R with add-on packages
- Report generation and website building

Class policies:

- **Any assignment, including the project report, turned in after the due date will only be worth half the earned points**
- Materials (exercises, project report) must be turned in electronically through Canvas
- Collaboration between students is encouraged, but **final products must show evidence of individual effort**. If not, no credit will be given.
- The material covered in class will be cumulative, and later classes will depend on earlier classes. Following the schedule of modules is encouraged

Online changes: This class was originally designed as an intensive 5-day workshop with 3 modules per day. With the transition to online teaching, this removes the need to have this taught in an intensive fashion, but it will follow the same modular format as before. The class consists of 15 modules (see table below), each designed to last around 90 minutes. Note that the modules will be somewhat progressive, so it is strongly suggested that you take them in the order they are posted.

Each module will focus on a particular aspect of working with R, and will consist of a short introductory lecture, a series of computer exercises and a short graded exercise designed to reinforce the lesson.

One difficulty with running an online class that involves a lot of computer lab work is obtaining help if you are stuck. There will be several channels that you can use to contact the instructor if you encounter problems

- Class Slack channel
- Canvas message
- Email

It is highly recommended to use the Slack channel where possible, as this will permit other students to follow any discussion basd on your question. There will also be a series of weekly meeting slots available through Canvas. These are 30 minutes one-on-one sessions where you can go over problems with the instructor, or discuss any issues related to the class.

Note that you are free to take all the modules within a week or space them out over the first summer session. The deadline for all exercises and work to be submitted is the end of the first summer session, June 25th.

IMPORTANT: We will be using RStudio as a front-end to R. These will both be available through the CSBS Citrix server, but it is strongly recommended that you download and install these on your own computer if possible. Installation instructions will be posted prior to the start of the class.

GEOG 5680 Preliminary Class Schedule

Module	Topic	Notes
1	Class introduction	Introduction to R and RStudio; getting data in and out of R
2	Data manipulation I	Vectors, matrices, data frames, lists
3	Basic plotting	Simple plots and plot parameters
4	Control statements	Looping, if-else
5	R functions	Function content, help. Writing your own functions
6	Writing reports in R	RMarkdown
7	Data manipulation II	tidyverse, dplyr
8	Extending basic plots	ggplot2, reshape2
9	Simple inference tests	*t*-test
10	Introduction to statistical modeling in R	lm and glm
11	Data manipulation III	data.table
12	Making maps in R	tmap; intro spatial data
13	R and Github	Using github to store projects

14	Web applications with Shiny	Simple UI/Server framework
15	Further applications	Spatial data, multi-level, foreach

Grading
Letter grades will be assigned following the scheme provided below, using .5 as the break point:

A	95+	C+	70~74
A-	90~94	C	65~69
B+	85~89	C-	60~64
B	80~84	D	50~59
B-	75~79	E	~49

Grades are based on two sets of assignments, in-class exercises and a class project.

In–class exercises (15x5pts):
Each lab contains one or more exercises, designed to provide further experience in the techniques used. The exercises generally ask to repeat part of the lab, but with different datasets, or slightly different questions. The exercises for each module will be worth a total of 3pts. Answers to the exercises should be provided in Word or other electronic format, and will usually require code and/or figures to be saved from R.

Guidelines for final project (25pts):
Students are required to carry out a short data analysis study, including exploration and modeling. Students may either a) use a predefined data set and set of questions; b) use their own data set. In the second case, please check with the instructor to make sure that your planned analysis is suitable for the course.

The **project report** should be written using R Markdown and submitted as an html page. We will cover the steps to generate these reports during the class. The report should include the code you have used, the results you obtained and a brief discussion of the steps you took and the results you obtained. There should be an emphasis on exploration of the original dataset, in addition to any more advanced analysis.

ACADEMIC MISCONDUCT SYLLABUS STATEMENT

- Academic misconduct will not be tolerated. Penalties may include failure of an assignment, the entire course, and/or the filing of formal charges with appropriate university authorities. Academic misconduct includes, but is not limited to, cheating, misrepresenting one's work, and plagiarism:
- Cheating involves the unauthorized possession or use of information in an academic exercise, including unauthorized communication with another person during an exercise such as an examination.

- Misrepresenting one's work includes, but is not limited to, representing material prepared by another as one's own work or submitting the same work in more than one course without prior permission of all instructors.
- Plagiarism means the intentional unacknowledged use or incorporation of any other person's work in one's own work offered for academic consideration or public presentation.

As the only institution in the state classified in the highest research category (R1), at the University of Utah you will have access to state-of-the-art research facilities and be able to be part of the knowledge creation process. You will have the opportunity to do research of your own with faculty who are leading experts in their field, engaging in programs that match your research interests. Further, you will interact with and often take classes with graduate students that provide an advanced understanding of the knowledge in your field.

The University of Utah seeks to provide equal access to its programs, services and activities for people with disabilities. If you will need accommodations in this class, reasonable prior notice needs to be given to the instructor and to the Center for Disability Services, 162 Olpin Union Building, 581-5020 (V/TDD) to make arrangements for accommodations. All written information in this course can be made available in alternative format with prior notification to the Center for Disability Services.