Reproducible R Data Management

2 credits – BIOL 693 (CRN 37704), NRM 693 (CRN 37707)

University of Alaska Fairbanks – Spring Semester 2020

TIME & LOCATION
Lecture: Monday evenings, 1800–1900 (1 hour)
Lab: Monday evenings, 1900–2200 (3 hours)

INSTRUCTOR
Cody Deane

Phone: 907.888.3134   Email: cdeane2@alaska.edu

Office: Suite 113 Murie Building   Office hours: TBD

NO PREREQUESITE COURSES REQUIRED
There is no prerequisite required for this course, but students are recommended to familiarize themselves with R, RStudio, and introductory R materials before the semester begins. Available to undergraduates with permission of instructor.

REQUIRED READINGS, VIDEO, AND MEDIA
Students will need to bring a laptop to class with R and RStudio installed.

– Students are especially encouraged to enroll in this course if they are in the process of preparing their data for the types of analyses statistics courses taught in BIOL/WLF 680.

All course materials will be delivered via Blackboard.

COURSE DESCRIPTION
This course will instruct students in reproducible data management in Program R from processing raw data to managing the output from statistical models. There will be an emphasis on understanding R data structures. This course is not an alternative to BIOL/WLF 680 nor any other statistics course.

INSTRUCTIONAL METHODS
Lectures will review and demonstrate fundamental concepts of weekly learning objectives. Quizzes and lab exercises will focus on completing R tasks, fixing errors in R code, and articulating R task goals.

IN-DEPTH COURSE DESCRIPTION
This course will familiarize students with reproducible data management in Program R, a process generally consisting of 1) processing, proofing, and understanding raw data (~2/5 course material), 2) converting raw data into formats for common statistical analyses and data archiving (~2/5 course material), and 3) storing, using, and presenting model results from common parametric techniques (e.g.; plotting fitted values, <1/5 course material). There will be a strong emphasis on understanding R data structures. This course is not an alternative to BIOL/WLF 680 (or an alternative to any statistics course) nor a Tidyverse course (including ggplot).

COURSE GOALS & STUDENT LEARNING OUTCOMES

1) Introduce reproducible data management practices from raw data to results plotting
- Learning outcome: Participants will learn how to rely on R for any challenge that can be accomplished in Excel and the world of possibilities that exist beyond Excel

2) Understand R data structures so that the source of almost any R error can be traced
- Learning outcome: Students will understand attributes of characters, numeric values, factors, NA values, vectors, matrices, data frames, arrays, lists and how this knowledge combined with command 'str()', 'dim()', 'names()', 'table()' can diagnose almost any R error

3) Articulating project goals and R code to increase trouble-shooting efficiency
- Learning outcome: Few problems in life can be solved if the issue can't be articulated. There will be a strong emphasis on expressing R code and describing R goals with common language.
- Learning outcome: Thorough understanding of bracketing within matrices, data frames, arrays
- Learning outcome: participants will tackle most tasks without Tidyverse packages, not because they aren't useful, but because their language is a caricature of base R.

4) for loops – many new R users have an innate sense for loops are the solution, they're usually not
- Learning outcome: for loops are a powerful tool in R (and many similar languages) but they're poorly understood. Students will learn how they work, when they should be used and avoided
- Learning outcome: The 'apply(),' 'tapply(),' and 'table()' commands, and writing short functions can replace most tasks thought to best be accomplished in Excel or with for loops

5) Mastering 1-to-many relationships with R: row data often include 1-to-many relationships but using these relationships to our advantage can be difficult in R
- Learning outcome: Summarizing and condensing data within factors
- Learning outcome: Reproducible joining of data from different sources using factors

6) Visualizing and understanding data from creative perspectives
- Learning outcome: Students will learn to rely on plotting to accelerate understanding and clear thinking about research objectives; emphasis will be placed on current thoroughness preventing future headaches

STUDENT PROTECTIONS AND SERVICES STATEMENT

Every qualified student is welcome in my classroom. As needed, I am happy to work with you, disability services, veterans' services, rural student services, etc to find reasonable accommodations. Students at this university are protected against sexual harassment and discrimination (Title IX), and minors have additional protections. For more information on your rights as a student and the resources available to you to resolve problems, please go the following site:www.uaf.edu/handbook/

SUPPORT SERVICES

Disabilities: Please contact the instructor and the Office of Disabilities Services (208 WHITAKER BLDG, 474-5655) if you require additional assistance. Students should also contact the UAF Writing Center for additional assistance with the assignments (801 Gruening, 474-5314, uaf-disabilityservices@alaska.edu).

Lecture and lab schedule

WEEK	DATE	LECTURE
1	01/13/20	R data structures, importing csv files, 'str()', 'dim()', 'table()'; *bracketing & plotting*
2	01/20/20	*Alaska Civil Rights Day (no class)*
2	01/27/20	Factors, characters, numeric values, and NAs; *bracketing & plotting*
3	02/03/20	Identifying and changing data structures; plotting pointers; *bracketing & plotting*
4	02/10/20	Reproducible data proofing and fixing data errors without Excel; *bracketing & plotting*
5	02/17/20	When and why histograms, density plots, boxplots, 'plot(table())'; *bracketing & plotting*
6	02/24/20	Creating factors for building 1-to-many relationships; *bracketing & plotting*
7	03/02/20	Establishing and connecting different data 1-to-many relationships; *bracketing & plotting*
8	03/09/20	Advanced use of for loops and how they contrast to bracketing, random parsing of data using 'sample()', strategies for missing values (NAs); *bracketing & plotting*; **midterm exam**
9	03/16/20	*Spring break (no class)*
10	03/23/20	Approaches for understanding and accounting for repeated measures and pseudo replication without statistics; exporting data summary tables; *bracketing & plotting*
11	03/30/20	Plotting spatial data and date formatting *OR open week as needed*; *bracketing & plotting*
12	04/06/20	Converting rows with 1-to-many relationships to row×column data in place of crosstab queries or pivot tables (e.g.; capture-recapture, occupancy data formats)
13	04/13/20	Summarizing row×column data (i.e.; capture-recapture and occupancy data) after its been formatted; advanced manipulation of row×column data; *bracketing & plotting*
14	04/20/20	Data structures within glm(), lm(), RMark, unmarked results output: *bracketing & plotting*
15	04/27/20	Fitted values and continuous covariates; the Delta method; removing regression' intercepts for presenting results; *bracketing & plotting*
NA	TBD	*Final exam*
[1]Readings and materials for lectures will be made available as necessary		
[2]The schedule may be amended to accommodate progress or in response to new opportunities		

EVALUATION

Attendance:
- You are expected to attend and to participate in all classes.
- You MUST contact the instructor in advance to request leave for a planned absence or to document an absence due to illness or emergency.
- Missed quizzes and lab assignments will be assessed as *zero* points unless you notified the instructor in advance of your absence via email. If the instructor is notified prior, alternative assignments or arrangements can be made to make-up for the quiz or exam.

Assignments:
- Weekly quizzes will occur at the beginning of lectures.
- Weekly in-class assignments will occur during labs

Exams:
- Midterm exam will closely reflect weekly quizzes, date TBD
- Final exam (comprehensive), date TBD
- Exams include material covered in lecture

Labs:
You are required to attend labs. Lecture quizzes and lab assignments will be a substantial portion of the grade. Labs will require approximately three hours of meeting time each week.

Grading:

Item	Description	Percentage of Total
Quizzes	Lecture quizzes	40%
Exams	Lab assignments	40%
Labs	Midterm and final exam	20%
Total		100%

Letter grades will be assigned using these criteria

Grade	A	B	C	D	F
%	>90.0	89.9-80.0	79.9-70.0	69.9-60.0	<60.0

This course will follows the University of Alaska Fairbanks Incomplete Grade Policy: "The letter "I" (Incomplete) is a temporary grade used to indicate that the student has satisfactorily completed (C or better) the majority of work in a course but for personal reasons beyond the student's control, such as sickness, has not been able to complete the course during the regular semester. Negligence or indifference are not acceptable reasons for an "I" grade."
- *"C"* indicates a satisfactory level of knowledge and performance. This is the minimum acceptable grade that undergraduates may receive for courses to count toward major or minor degree requirements, or as a prerequisite for another course.
- **Academic Integrity:** University of Alaska students are expected to conduct themselves with academic integrity. There is a zero-tolerance policy for plagiarism or cheating https://www.bw.uaf.edu/graduates/academic_honesty.php. Please review the Student Code of Conduct to help you understand what is expected and what measures are taken to address misconduct: http://www.uaf.edu/catalog/current/academics/regs3.html#Student_Conduct