**EEOB590A: Data Science in R**

**Instructor: Dr. Haldre Rogers**

**231 Bessey Hall**

**Tuesday lecture 2:30-3:30, Thursday coding workshop 2:30-3:30**

**2 credits**

**Course description:**

In this course, students will learn how to follow best practices for conducting reproducible research. Using the open-source software, R, we will practice manipulating, analyzing, and graphing the types of data commonly collected in ecology and experimental evolutionary biology. In this class, the students will work through the problems that inevitably arise when analyzing real, messy datasets plagued by missing values, small sample sizes, over-dispersion etc.

Tuesdays will consist of a lecture to introduce the topic, followed by a coding session to show how the concept is applied in R. On Thursdays, students will be asked to apply the skills to an example dataset. For homework, students will then apply what they have learned to their own dataset or one provided by the instructor. For the final project, each student will analyze their own dataset, and turn in the script, methods & results sections, and manuscript-quality figures/tables. We will finish the semester with a symposium where each student will give a short presentation about his or her research.

**Requirements:** Students should have a basic understanding of R (e.g. how to load data, make a simple graph), or be willing to spend extra time the first few weeks learning. A basic background in statistics (e.g. STAT401(now587)/402, EEOB 590 - Advanced Biostatistics) is helpful but not required. Students should bring a dataset, ready for analysis or request a dataset from their advisor or the instructor.

**Course Objectives:**

By the end of the semester, students will:

1. Be comfortable using R to import, explore, and graph data.
2. Know how to follow best practices for sharing data and code.
3. Be comfortable discussing statistical analyses and sharing code.
4. Have analyzed their own dataset, written the analysis section and the results section of a paper, and have produced manuscript-quality graphics.
5. Become part of an active community of R-users in Ecology & Evolutionary Biology at Iowa State.

**Topics we will cover include:**

* Intro to R, R Studio, and GitHub
* Data management plans
* Designing datasheets and databases
* Data munging/wrangling
* Data exploration
* Data visualization (base graphics and ggplot)
* Intro to linear models
* Other topics to be determined by the course participants & instructor

**Course textbook**

“R for Data Science” by Hadley Wickham and Garrett Grolemund (O’Reilly). Copyright 2017 Garrett Grolemund, Hadley Wickham, 978-1-491-91039-9. <https://r4ds.had.co.nz/>

**Resources and other courses/opportunities for building R skills:**

* ***LunchinatoRs*** - Every Friday from 12-1. Informal brown bag seminar where people take turns sharing code and getting feedback or teaching others new skills. 334 Bessey.
* ***Iowa State Statistics Consulting*** is a fantastic resource. They can help with (from their website) "research design, sample size calculations, choosing statistical methods, use of statistical computing packages (R, SAS, and some JMP) to analyze data, and interpretation of results. Just make an appointment here: http://stat.iastate.edu/statistical-consulting
* **EEOB 590: Spring, Odd Years: Advanced Biostatistics with Dean Adams**
  + Review of the basic univariate and multivariate statistics commonly used in evolutionary and ecological research. The goal of the course is to give students a general idea of what statistical methods are commonly used in evolutionary ecology, which methods are appropriate for which types of data, and to provide a general knowledge of how the methods work.
* R-ladies-Ames - local hub of a global organization for women using R.
  + rladies.org
  + meetup.com/R-Ladies-Ames
  + twitter.com/RLadiesAmes
* **Various Statistics courses useful for ecology/evolutionary biology, with sections that use R:** 
  + STAT579 - Learning how to code in R
  + STAT587- Intro to statistics
  + STAT402- Experimental design
  + STAT406- Spatial data, typically taught by Dr. Phil Dixon
  + STAT407- Multivariate analyses
  + STAT444- Bayesian data analysis
  + STAT457 - Categorical data, linear models

**Reference books**

* Bolker, B. M. 2008. Ecological models and data in R. Princeton University Press, Princeton, N.J.
* Clark, J. S. 2007. Models for ecological data : an introduction. Princeton University Press, Princeton, N.J.
* Crawley- The R Book. Available online at: http://www.kharms.biology.lsu.edu/CrawleyMJ\_TheRBook.pdf
* Fox, G.A., Negrete-Yankelevich, S. and Sosa, V.J. eds., 2015. *Ecological statistics: contemporary theory and application*. Oxford University Press, USA.
* Kery, M. 2010. Introduction to WinBUGS for Ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses. Academic Press.
* Zuur, A. F., E. N. Ieno, and G. M. Smith. 2007. Analysing ecological data. Springer, New York ; London. http://link.springer.com/book/10.1007%2F978-0-387-45972-1 (free download)
* Zuur, Ieno, Walker, Savelieve, and Smith. 2009. Mixed Effects Models and Extensions in Ecology with R. http://link.springer.com/book/10.1007%2F978-0-387-87458-6 (free download)

**Online learning resources**

* Great (free) book to learn R: Hands-on programming with R by Garrett Grolemund. <https://rstudio-education.github.io/hopr/index.html>
* https://www.coursera.org/learn/r-programming
* http://ecology.msu.montana.edu/labdsv/R/labs/R\_ecology.html
* http://swirlstats.com/students.html
* Tips on generalized linear mixed effects models at the GLMM wiki: http://glmm.wikidot.com/faq
* http://ropensci.github.io/reproducibility-guide/sections/introduction/

**Class Schedule**

Note that this is subject to change, depending on the pace of the class and what is working/not working.

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Topic** | **Homework (due following Tuesday)** | **Chap.** |
| 27-Aug-19 | Introduction to course, GitHub, RStudio | Get setup with GitHub |  |
| 29-Aug-19 | Coding Thursday: Getting comfortable with GitHub & RStudio | Set up your own repository on GitHub with raw data; Fill out Analysis\_Outline |  |
| 3-Sep-19 | Introduction to R |  | 2,4,6,8 |
| 5-Sep-19 | Coding Thursday: getting comfortable with R | Script that reads in data and libraries |  |
| 10-Sep-19 | Introduction to R, continued |  | 3 |
| 12-Sep-19 | Coding Thursday: more coding in R | Read tidy data paper by Hadley Wickham |  |
| 17-Sep-19 | Introduction to Reproducible Research |  | NA |
| 19-Sep-19 | Coding Thursday: Data management plan | Data Management Plan |  |
| 24-Sep-19 | Introduction to Data Wrangling, tidyverse, piping |  | 9,10,14 |
| 26-Sep-19 | Coding Thursday: Data wrangling | Finish practice script from class today |  |
| 1-Oct-19 | Data Wrangling continued |  | 12,13 |
| 3-Oct-19 | Coding Thursday: Data wrangling | Finish practice script from class today |  |
| 8-Oct-19 | Visualizing datasets, intro to ggplot2 |  | B1-ggplot, B5-ExDatAn |
| 10-Oct-19 | Coding Thursday: Exploring and visualizing datasets | Data wrangling script and exploratory graphics from your own dataset |  |
| 15-Oct-19 | Data simulation |  | NA |
| 17-Oct-19 | Coding Thursday: Data simulation | Simulated data set & accompanying script |  |
| 22-Oct-19 | Choosing your statistical approach |  | NA |
| 24-Oct-19 | Coding Thursday: Choosing your statistical approach | Proposed statistical approach for your dataset |  |
| 29-Oct-19 | Data analysis part 1 |  | 18 |
| 31-Oct-19 | Coding Thursday: Data analysis | First draft of data analysis |  |
| 5-Nov-19 | Data analysis part 2 |  | Fox6,13 |
| 7-Nov-19 | Coding Thursday: Data analysis | Final data analysis |  |
| 12-Nov-19 | Graphing part 1 |  | 22 |
| 14-Nov-19 | Coding Thursday: Graphing | First draft of graphs |  |
| 19-Nov-19 | Graphing part 2 |  | NA |
| 21-Nov-19 | Coding Thursday: Graphing | Final draft of graphs |  |
| 26-Nov-19 | No class- Thanksgiving |  |  |
| 28-Nov-19 | No class- Thanksgiving |  |  |
| 3-Dec-19 | What to include in a paper |  | NA |
| 5-Dec-19 | Coding Thursday: Preparation for presentations | Written methods & results section |  |
| 10-Dec-19 | Final presentations & peer review |  |  |
| 12-Dec-19 | No class |  |  |

**Academic Dishonesty**

The class will follow Iowa State University’s policy on academic dishonesty.  Anyone suspected of academic dishonesty will be reported to the [Dean of Students Office](http://www.dso.iastate.edu/ja/academic/misconduct.html).

**Accessibility Statement**

Iowa State University is committed to assuring that all educational activities are free from discrimination and harassment based on disability status. Students requesting accommodations for a documented disability are required to meet with staff in Student Accessibility Services (SAS) to establish eligibility and learn about related processes.  Eligible students will be provided with a Notification Letter for each course and reasonable accommodations will be arranged after timely delivery of the Notification Letter to the instructor.  Students are encouraged to deliver Notification Letters as early in the semester as possible.  SAS, a unit in the Dean of Students Office, is located in room 1076, Student Services Building or online at [www.sas.dso.iastate.edu](http://www.sas.dso.iastate.edu/).  Contact SAS by email at [accessibility@iastate.edu](mailto:accessibility@iastate.edu) or by phone at 515-294-7220 for additional information.

**Dead Week**

This class follows the Iowa State University Dead Week policy as noted in section 10.6.4 of the [Faculty Handbook](http://www.provost.iastate.edu/resources/faculty-handbook).

**Discrimination and Harassment**

Iowa State University does not discriminate on the basis of race, color, age, ethnicity, religion, national origin, pregnancy, sexual orientation, gender identity, genetic information, sex, marital status, disability, or status as a U.S. Veteran. Inquiries regarding non-discrimination policies may be directed to Office of Equal Opportunity, 3410 Beardshear Hall, 515 Morrill Road, Ames, Iowa 50011, Tel. 515-294-7612,  Hotline 515-294-1222, email [eooffice@iastate.edu](mailto:eooffice@mail.iastate.edu)

**Religious Accommodation**

Iowa State University welcomes diversity of religious beliefs and practices, recognizing the contributions differing experiences and viewpoints can bring to the community. There may be times when an academic requirement conflicts with religious observances and practices. If that happens, students may request reasonable accommodation for religious practices. In all cases, you must put your request in writing. The instructor will review the situation in an effort to provide a reasonable accommodation when possible to do so without fundamentally altering a course. For students, you should first discuss the conflict and your requested accommodation with your professor at the earliest possible time. You or your instructor may also seek assistance from the [Dean of Students Office](http://www.dso.iastate.edu/sa/) at 515-294-1020 or the [Office of Equal Opportunity](http://www.eoc.iastate.edu/) at 515-294-7612.

**Contact Information**

If you are experiencing, or have experienced, a problem with any of the above issues,  email [academicissues@iastate.edu](mailto:academicissues@iastate.edu)