Biology 364/664 Syllabus
Advanced Data Analysis in Biology
Spring 2022
Mon 1:30-2:50 or 3:00-4:20 and Wed 1:30-4:20

Prof. Ken Field

January 18, 2022

1 Contact Information

- 208 Biology Building
- kfield@bucknell.edu
- @ProfKenField on Twitter
- @KField-Bucknell on GitHub

2 Course Description

Reproducibility, transparency, and avoiding questionable research practices while discussing how to design experiments and then collect, analyze, explore, and present data. Using "big data" from their own research projects or public transcriptomic datasets students will learn to analyze/visualize complex biological datasets. Includes hands-on work with R. No prior programming experience required.

3 Course Objectives

1. Students will analyze, visualize, and intepret real-world datasets using reproducible data science methods and R, R markdown, and Git.
2. Students will learn to identify and avoid questionable research practices when designing experiments, analyzing data, and presenting results.
3. Working as a team, students will complete novel projects utilizing whole-transcriptome or whole-genome datasets.
4. Students will present their final projects using complex multi-dimensional data visualizations.

4 Grading

- Eight Homeworks, worth 12.5 points each = 100 pts.
- Four Data Projects, worth 25 points each = 100 pts.

- Takehome Midterm xam = 100 pts.
- Takehome Final Exam = 100 pts.
- **Total = 400 pts.**

The Homework assignments and Data Projects will be graded using labor-based grading as described by Asao Inoue. These assignments will utilize a goal-oriented grading system as you develop your skills as a data scientist. When you (or your group) complete all of the goals associated with each project, you will earn 85% of the available points. Failure to do so will result in a lower grade.

To earn more than 85%, you must do the following:

- Propose and perform additional analysis and visualization comparisons (discuss available options with me).
- Routinely assist in the learning and proficiency of your peers.

5 Textbooks and Readings

Required Texts and their Abbreviations, which are used on the syllabus:

Main textbook:

- Applied Statistics with R: A Practical Guide for the Life Sciences by Justin C. Touchon, (ASR)

Leanpub textbooks (pay what you want for PDF or online version):

- R Programming for Data Science, by Roger D. Peng (PRP)
- Exploratory Data Analysis, by Roger D. Peng (EDA)
- Open Intro Statistics 4 (OIS)

Open Access Books, available from online for free:

- R for Data Science (R4DS)
- Statistical Thinking for the 21st Century (21st)
- Additional reading assignments will be distributed via the course git repository in the Readings subdirectory.

6 Policies

6.1 Attendance

Your attendance at all classes and lab is expected, but not a graded part of the course. If you will need to miss lab for any reason, contact Prof. Field before class to make arrangements. However, I want everyone to know that absences due to health concerns will always be accommodated.

6.2 Readings

Readings must be done BEFORE the class where they are assigned. For dates with multiple reading assignments, browse each to determine the sections that are most useful to you. There will be overlap between the various textbooks and you should choose the book that is best for you and your individual background.

6.3 Homework Assignments

Homeworks are due before 11:59pm on Friday on the week they are assigned. Homework assignments will be submitted on GitHub (more instructions to follow). A 25% deduction will be assessed for submissions not received on time. Assignments will not be accepted after noon on Sunday.

6.4 Collaboration and Citation

For all projects and homework assignments working together is acceptable **and encouraged**. It is not ethical to do someone else's work or to have someone do your work. You must cite **all** resources used to work on your homework and projects. Citations should be done at the end of the document. These references can be to books, Stack Overflow and other web resources, and discussions with other students. Working together and discussion is not allowed on takehome exams.

6.5 Academic Integrity Policy

Read Academic Responsibility at Bucknell for policies regarding academic integrity. Any questions concerning academic responsibility or misconduct will be referred to the Board of Review for Academic Responsibility without hesitation. Always cite the source of any information from outside sources, including online sources and classmates. Assignments may be screened using software designed to detect plaigarism. Unless explicitly directed otherwise, all takehome exams are expected to represent individual, not collaborative, work.

6.6 Bucknell University Honor Code

- I will not lie, cheat, or steal in my academic endeavors.
- I will forthrightly oppose each and every instance of academic dishonesty.
- I will let my conscience guide my decision to communicate directly with any person or persons I believe to have been dishonest in academic work.
- I will let my conscience guide my decision on reporting breaches of academic integrity to the appropriate faculty or deans.

6.7 Accommodations

Any student who needs an accommodation based on the impact of a disability should contact the Office of Accessibility Resources at OAR@bucknell.edu, 570-577-1188 or in room 212 Carnegie Building who will coordinate reasonable accommodations for students with documented disabilities.

7 License

8 BIOL 364 Syllabus: Weekly Topics

Week:Date	Topic	Reading	Project
w1:19Jan2022	Using R, Rstudio	**ASR1** PRP4-5	Making Graphs in R
w2:24Jan2022	Simple Data Exploration, Git	**ASR2-3** PRP6-10 EDA4-7 OIS1-4	Data Exploration in R
w3:31Jan2022	Data Visualization	**ASR4-5** EDA15-16 OIS5-7 21st6-7	**R Tutorial Project**
w4:07Feb2022	Advanded Hypothesis Testing	**ASR6** **ASA.pdf** OIS8-9 21st8-9	Exploratory Data Analysis
w5:14Feb2022	Questionable Data Practices	**Fraser.pdf** R4DS22-25 21st10-13,26-27	Multiple Testing Model Fitting
w6:21Feb2022	Questionable Data Practices	**Forstmeier.pdf** PRP12 EDA12-14 21st16-21	**QRP Case Studies Project**
w7:28Feb2022	Reproducibility	**Errington.pdf**	
w8:07Mar2022	Transcriptomics	RNA-seqlopedia	**TAKEHOME MIDTERM**
14Mar2022	**SPRING BREAK**		
w9:21Mar2022	Transcriptomics	Brooks.pdf	Big Data Wrangling
w10:28Mar2022	Transcriptomics	21st32 Williams.pdf SARTools.pdf	Big Data Visualization
w11:4Apr2022	Transcriptomics Project	minitufte.pdf	**Transcriptomics Pipeline Project**
w12:11Apr2022	Transcriptomics Project		Open Science Preregistration
w13-14:18Apr2022	Transcriptomics Project		**Transcriptomics Presentations**
04May2020	**TAKEHOME FINAL**		

Table 1: BIOL364/664 Weekly Syllabus. Touchon, Applied Statistics with R (ASR), Peng R Programming for Data Science (PRP), Peng Exploratory Data Analysis (EDA), Open Intro Statistics (OIS), R for Data Science (R4DS), and A Primer in Biological Data Analysis and Visualization Using R (HART) refers to chapters assigned as reading. **Bold font** indicates required reading.