Biology 364-664 Syllabus Spring 2023
Advanced Data Analysis in Biology

Prof. Ken Field

January 15, 2023

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 transcriptomic datasets students will learn to analyze/visualize complex biological networks. Includes hands-on work with R. No prior programming experience required.

Table of contents

1 Contact Information

- Pronouns: He/Him
- Office: 203C Biology Building
- kfield@bucknell.edu
- @ProfKenField on Twitter
- @ProfKen@fediscience on Fediverse
- @KField-Bucknell on GitHub
- Computer Lab: Rooke Chemistry 009

2 Meeting Times

Section	Tuesday	Thursday
01	1 - 4pm	1 - 2pm
Office hours		2 - 3pm
02	1 - 4pm	3 - 4pm
Office hours		4 - 5pm

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

Assignment	Points each	Total
Eight Homeworks	12.5	100
Four Data Projects	25	100
Takehome Midterm Exam	100	100
Takehome Final Exam	100	100

Assignment	Points each	Total
Total		400

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 weekly schedule:

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 is 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. Classes will be taught in hybrid format as necessary. Please inform Prof. Field at least one hour before class starts if you will be attending remotely.

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 3:00pm 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, Artificial Intelligence, 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 (or something) do your work. You must cite **all** resources used to work on your homework and projects. This includes any use of ChatGPT, OpenAI, or other machine-learning language models. Citations should be done at the end of the document in a section called **Acknowledgements**. These references can be to books, Stack Overflow and other web resources (including a link), ChatGPT (including the prompt used), and discussions with other students. Working together, use of ChapGPT, and online discussions are 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 plagiarism or the use of AI. 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 or illness 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

Creative Commons plays an important role in open science, open data, open source efforts. This class is covered by a Creative Commons license. The license we'll use for class materials, code, and presentations is covered by the "Attribution-ShareAlike 4.0 International" license, which is commonly called the CC BY-SA 4.0 license. Some of the materials for this course, including portions of the Syllabus, are derived from work by Roger H. French @frenchrh Kyocera Professor, Materials Science, Case Western Reserve University.

8 Weekly Schedule

The weekly schedule will be updated regularly and the current version can be found at: https://github.com/Bucknell-Biol364/Biology364-2023/tree/main/00-Syllabus