Biology 364/664 Syllabus
Advanced Data Analysis in Biology
Spring 2020
Mon Wed 2:00pm to 4:52pm

Prof. Ken Field

January 10, 2020

1 Contact Information

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

2 Course Description

Data exploration and visualization using state-of-the-art computational techniques. Using "big data" from their own research projects or public transcriptomic datasets, students will learn to analyze/visualize complex biological datasets. Lab includes hands-on work with R/virtual reality. 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

$$\begin{aligned} Eight\,Homeworks,\ worth\,12.5\,points\,each &= 100pts. \\ Four\,Data\,Projects,\ worth\,25\,points\,each &= 100pts. \\ Takehome\,Midterm\,Exam &= 100pts. \\ Final\,Project &= 200pts. \\ Takehome\,Final\,Exam &= 100pts. \\ \textbf{Total} &= \textbf{600pts}. \end{aligned}$$

The Data Projects and Final Project will be graded using labor-based contract 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 21/25 or 170/200 pts (85%). 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:

- A Primer in Biological Data Analysis and Visualization Using R, by Gregg Hartvigsen, ISBN: 9780231166997 (Hart)

  This is the only hardcopy textbook for the course.

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

  These are Leanpub books, available from LeanPub for a "pay what you want" price.

- R for Data Science (R4DS)
- 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.

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 noon 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 by noon on Friday. 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

1. I will not lie, cheat, or steal in my academic endeavors.
2. I will forthrightly oppose each and every instance of academic dishonesty.
3. 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.
4. 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 Heather Fowler, Director of the Office of Accessibility Resources at hf007@bucknell.edu, 570-577-1188 or in room 212 Carnegie Building who will coordinate reasonable accommodations for students with documented disabilities. The college will make reasonable accommodations for persons with disabilities.

7 License

8 BIOL 364 Syllabus: Weekly Topics

Week:Date	Topic	Reading	Project
w1:13Jan2020	Using R, Rstudio, Git	HART1-3 PRP4-5	Making Graphs in R
w2:20Jan2020	Simple Data Exploration	HART4 PRP6-10 EDA4-7 OIS1-4	Data Exploration in R
w3:27Jan2020	Data Visualization	HART5,11 EDA15-16 OIS5-7	**R Tutorial Project**
w4:03Feb2020	Advanded Hypothesis Testing	HART6-8 OIS8-9	Exploratory Data Analysis
w5:10Feb2020	Questionable Data Practices	R4DS22-25 Fraser.pdf	Multiple Testing Model Fitting
w6:17Feb2020	Questionable Data Practices	PRP12 EDA12-14 Forstmeier.pdf	**QRP Case Studies Project**
w7:24Feb2020	Transcriptomics	RNA-seqlopedia Brooks.pdf	QC & Mapping Statistics
w8:02Mar2020	Transcriptomics	SARTools.pdf Williams.pdf	**Transcriptomics Pipeline Project I**
09Mar2020	**SPRING BREAK**		
w9:16Mar2020	Interactive Data	minitufte.pdf	**Transcriptomics Pipeline Project II**
w10:23Mar2020	Final Project	reproducible-code.pdf	**TAKEHOME MIDTERM** Data Archives
w11:30Mar2020	Final Project		Data Wrangling
w12:06Apr2020	Final Project		Preregistration **Data Visualization Project**
w13:13Apr2020	Final Project		**Final Project Presentations**
w14:20Apr2020	Final Project		**Final Project Presentations**
06May2020	**TAKEHOME FINAL**		

Table 1: BIOL364/664 Weekly Syllabus. 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.