Biology 364/664 Syllabus
Analysis and Visualization of Biological Data
Spring 2019
Tues Thurs 11:00am to 11:52am and Tues 1:00pm to 4:52pm

Prof. Ken Field

December 28, 2018

1 Contact Information

Prof. Ken Field

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

2 Course Description

Introduction to the process of data exploration and visualization using state-of-the-art computational techniques. Using "big data" from public archives or their own research projects, students will learn how to rigorously analyze and visualize complex biological datasets. Lab will include hands-on work with R and virtual reality. No 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

$$\begin{aligned} Ten\ Homeworks,\ worth\ 10\ points\ each &= 100pts. \\ Four\ Data\ Projects,\ worth\ 25\ points\ each &= 100pts. \\ Takehome\ Midterm\ Exam &= 100pts. \\ Final\ Project &= 100pts. \\ Takehome\ Final\ Exam &= 100pts. \\ \textbf{Total} &= \textbf{500pts}. \end{aligned}$$

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 completes all of the goals associated with each project, you will earn 43/50 or 85/100 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 Abbreviation, which is used in the class:

- A Primer in Biological Data Analysis and Visualization Using R, by Gregg Hartvigsen, ISBN: 9780231166997 (Hart)
- R Programming for Data Science, by Roger D. Peng (PRP)
- Exploratory Data Analysis, by Roger D. Peng (EDA)

  These are Leanpub books, available from LeanPub for a "pay what you want" price.
- R for Data Science (R4DS) is a book teaching R for Tidy Data Science, and is available as a bookdown book on the web R for Data Science, you can buy an ebook for $18 from Google Play Store page for this book.
- Open Intro Statistics version 3 (OIS) is an open source text book on Inferential Statistics, published under a Creative Commons license, for free distribution as a pdf. In addition a copy can be purchased from Amazon for $9.
- Additional reading assignments will be distributed via the course git repository in the readings subdirectory.

6 BIOL 364 Syllabus: Weekly Topics

Day:Date	Foundation	Practicum	Reading	Due
w1a:Tu:8/28/18	ODS Tool Chain	R, Rstudio, Git		
w1b:Th:8/30/18	Setup ODS Tool Chain	Bash, Git, Twitter	PRP4-33	HW1
w2a:Tu:9/4/18	What is Data Science	OIS:Intro2R	PRP35-64	**HW1 Due**
w2b:Th:9/6/18	Data Analytic Style, Git	451SempProj, Git	PRP65-93, OI1-1.9	HW2
w3a:Tu:9/11/18*	Struct. of Data Analysis	ISLR:Intro2R, Loops	PRP94-116, OIS3	**HW2 Due**
w3b:Th:9/13/18*	OIS3 Intro to Data	GapMinder, Dplyr, Magrittr		
w4a:Tu:9/18/18	OIS3, Intro2Data part 2, Data	EDA: PET Degr.	EDA1-31	Proj1
w4b:Th:9/20/18	Hypothesis Testing	GGPlot2 Tutorial	EDA32-58	HW3
w5a:Tu:9/25/18	Distributions	SemProj RepOut1	R4DS1-3	**HW3 Due**
w5b:Th:9/27/18	Wickham DSCI in Tidyverse	SemProj RepOut1	R4DS4-6	**SemProj1**,
w6a:Tu:10/2/18	OIS Found. of Inference	Inference	R4DS7-8	**Proj1 Due**
w6b:Th:10/4/18		Midterm Review	R4DS9-16 Wrangle	
w7a:Tu:10/9/18*	Summ. Stats & Vis.	Data Wrangling		
w7b:Th:10/11/18*	**MIDTERM EXAM**			HW4
w8a:Tu:10/16/18	Numerical Inference	Tidy Check Explore	OIS4	**HW4 Due**
w8b:Th:10/18/18	Algorithms, Models	Pairwise Corr. Plots	OIS5.1-4	Proj 2, HW5
Tu:10/23	**CWRU FALL BREAK**		R4DS17-21 Program	
w9b:Th:10/25/18	Categorical Infer	Predictive Analytics	OIS6.1,2	
w10a:Tu:10/30/18	SemProj	SemProj	OIS7	**SemProj2 HW5 Due**
w10b:Th:11/1/18	Lin. Regr.	Lin. Regr.	OIS8	**Proj.2 due**
w11a:Tu:11/6/18	Inf. for Regression	Curse of Dim.	OIS8	Proj 3
w11b:Th:11/8/18	Model Accuracy	Training Testing	ISLR3	HW6
w12a:Tu:11/13/18	Multiple Regr.	Mul. Regr. & Pred.	ISLR4	**HW6 due**
w12b:Th:11/15/18	Classification		ISLR6	
w13a:Tu:11/20/18	Classification	Clustering	ISLR5	**Proj 3 due**
Th:11/22/18	**THANKSGIVING**			Proj 4
w14a:Tu:11/27/18	Big Data	Hadoop		
w14b:Th:11/29/18	InfoSec	VerisDB		**SemProj3**
w15a:Tu:12/4/18	SemProj ReportOut3			
w15b:Th:12/6/18	SemProj ReportOut3			**Proj4**
	FINAL EXAM	**Monday12/17, 12:00-3:00pm**	Olin 313	**SemProj4 due**

Table 1: DSCI351-451 Weekly Syllabus. Peng R Programming (PRPx.y), Peng Exploratory Data Analysis (EDAx.y), R for Data Science (R4DSx.y), Open Intro Statistics (OISx.y) and Introduction to Statistical Learning with R (ISLRx.y) refers to chapters and sections assigned as reading.

7 Policies

7.1 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.

7.2 Readings

Readings must be done, BEFORE the class, where they are assigned. The reading assignment, is for the class with which it is listed.

7.3 Homework Assignments

Homeworks are due before noon on Friday on the week they are assigned. Homework assignments will be submitted on GitHub after week 1 (more instructions to follow). A 25% deduction will be assessed for submissions not received by noon on Friday. Assignments will not be accepated after noon on Monday.

7.4 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, Wikipedia and other web resources, and discussions with other students. Working together and discussion is not allowed on takehome exams.

7.5 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.

7.6 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.

7.7 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.

8 License