BCB/EEOB 546X, Computational Skills for Biological Data, Fall 2017
Wednesday/Friday, 1:10-2:30p.m., 334 Bessey Hall

Instructor: Dr. Matt Hufford   Office Hours: By Appointment
339A Bessey Hall

Instructor: Dr. Tracy Heath   Office Hours: By Appointment
243 Bessey Hall

Instructor: Dr. Dennis Lavrov   Office Hours: By Appointment
343A Bessey Hall

Course Website: https://eeob-biodata.github.io/BCB546X-Fall2017/

Course Description:

With the advent of big data across a number of sub-disciplines within biology, there is an increasing need for biologists to be comfortable and competent with command-line data processing and analysis. This course will develop basic skills necessary for biologists working with big data sets. Topics will include UNIX commands, scripting in R and Python, version control using Git and GitHub, use of high performance computing clusters, and writing effective data-management plans. These topics will be taught using a combination of lectures and computational exercises.

Required Text:

Buffalo, V. Bioinformatics Data Skills. Sebastopol, CA: O'Reilly Media, Inc., 2015. Print and eBook.

This textbook can be purchased on Amazon and directly through O'Reilly Media:

Amazon: http://amzn.to/2k26Vuy
O'Reilly: http://shop.oreilly.com/product/0636920030157.do#

In addition, Iowa State University has purchased 6 seats for O'Reilly eBooks through Safari-ProQuest. This means that only 6 users at a time can access any/all of the O'Reilly books. Please feel free to utilize this resource, but be sure to close your browser window when done reading in order to free up a seat for other users. If you receive a message that the book is unavailable, keep checking back until a seat is available. A direct link to the book through the ISU library is found below:

http://proquest.safaribooksonline.com.proxy.lib.iastate.edu/9781449367480?uicode=iastate

Correspondence:

All correspondence should be conducted via Slack. You have been invited to join the BioDataSkills Slack team. We recommend that you download the desktop application, which is available for Windows, Mac OS X, and Linux (there are also mobile apps). Via Slack, you can send group messages, private messages to any instructor (or all three at once) or class member, and hold discussions regarding lecture topics and assignments. It is also possible to upload files and bits of code on Slack. We have found this platform to be extremely useful for course facilitation and for collaborating on computational projects.

Learning Outcomes:

By the end of this course, students should be able to:
1) Navigate through their computers, create and modify files and directories, and process data using basic Unix commands.
2) Become familiar with basic R syntax and data structures and implement these in data analysis and plotting.
3) Utilize the Python scripting language for more sophisticated data processing.
4) Become familiar with various genomic data types (range, sequence, and alignment data) and learn how to write scripts and analysis pipelines for working with these data.
5) Become familiar with high performance computing resources at Iowa State as well as how and when to employ these resources.
6) Explore additional resources/topics in computational biology including manuscript preparation in LaTeX and Overleaf and creation of NSF-style Data Management Plans.

Hardware and Software Requirements:

In order to complete computational exercises and the group project for this course, students will need access to a Unix operating system. The Mac OS X operating system is based on Unix. Mac users can access a command line interface through the "Terminal" application (located in the /Applications/Utilities folder). For PC users, a Unix environment can be emulated using the "Git Bash" program, by partitioning your hard drive and installing Ubuntu, or by installing Cygwin. These solutions for PC users may not be ideal for all course exercises and we can provide access to Mac computers if you would prefer to go this route. We will also demonstrate how to log into HPC-class, a computer cluster for coursework at ISU. When on HPC-class, you are in a Unix environment.

Links for PC Unix emulation:
Git Bash: https://openhatch.org/missions/windows-setup/install-git-bash
Ubuntu: https://help.ubuntu.com/community/Wubi
Cygwin: https://www.cygwin.com

Software requirements for various portions of the course will be communicated through the GitHub repository found at the following url:

https://eeob-biodata.github.io/BCB546X-Fall2017/install

Grading:

Assignment 1: Unix	15%
Assignment 2: R	15%
Assignment 3: Python	15%
Assignment 4: Data Management Plan	15%
Group Project and Presentation	40%

Late work and plagiarism/other forms of academic misconduct will result in a zero grade for the relevant assignment. Students suspected of academic misconduct will be referred to the Dean of Students Office. For information on ISU's Standards of Academic Integrity, please see the following website:

http://www.dso.iastate.edu/ja/academic/students

Grading Scale:

Grade	Score (=s)
A	s ≥ 92
A-	90 ≤ s <92
B+	88 ≤ s <90
B	82 ≤ s <88
B-	80 ≤ s <82
C+	78 ≤ s <80
C	72 ≤ s <78
C-	70 ≤ s <72
D+	68 ≤ s <70
D	62 ≤ s <68
D-	60 ≤ s <62
F	s <60

Course Schedule:

A link to our course schedule including topics covered by date, reading assignments, and assignment due dates can be found at the link below. This is a *tentative* schedule and we will communicate any changes throughout the semester on Slack and in lecture.

https://docs.google.com/spreadsheets/d/10DN_Qsf7QJaBDkEfhGxnSG1fvaDosrXjAKAFe_PsV3o/edit?usp=sharing