**Instructor:** Robert Harbert, Ph.D.

**Contact Info:** Email: [rharbert@stonehill.edu](mailto:rharbert@stonehill.edu)

Office: TBA

**Office Hours:** Mondays and Thursdays 1:00 – 3:00 PM or by appointment.

**Lecture Schedule:** BIO 200 A, MW 4:00-5:15PM, SSC142

**Required Lecture Material:**

* **Course textbook:**

*Required:* Buffalo, V., 2015. *Bioinformatics Data Skills: Reproducible and robust research with open source tools*. " O'Reilly Media, Inc.". (ISBN: 978-1449367374) \*\*Available new and used on Amazon\*\*

*Recommended:* Lesk, A., 2014. *Introduction to bioinformatics*. Oxford University Press. (ISBN: 9780199651566)

* **Laptop.** Access to a computer, preferably running a Mac or Unix/Linux operating system. Common bioinformatics software installation (including the R and Python programming languages) will be covered in class.

**BIO 200 Course Description:**

This course introduces common concepts and tools in the field of Bioinformatics with a focus on developing a basic skill set for working with large biological data sets. The digital age has resulted in a period of rapid growth of data, and in biology this is revolutionizing how we look at the world. Understanding how the field uses computational tools to manage and study these massive datasets is a crucial skill set for the modern Biology student. This course will cover the major sources of data in biology and an overview of the myriad of computational tools available.

**Course Goals & Objectives**:

After having completed Introduction to Bioinformatics you will be able to:

* Discuss Biology as a subdiscipline in “Data Science”
* Understand the major sources of ‘big-data’ in Biology and the scale and nature of the data being produced.
* Perform fundamental operations (data input/output, statistics, data visualization) in both R and Python programming environments.
* Use common bioinformatics tools using the Unix command line, R, and Python.
* Run bioinformatics programs from the Unix command line to perform analysis of DNA sequence data.
* Understand the importance of reproducibility and open access for data and computer code in bioinformatics.

**BIO 101 Tentative Lecture Schedule**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **Day** | **Lecture** | **Topic** | **Reading** | **Assignments** |
|  |  |  |  |  |  |
| 8/28 | T | 1 | Course Expectations & Intro |  |  |
| 9/3 | M |  | *Labor Day – No Class* |  |  |
| 9/5 | W | 2 | L: Biology as a “Data Science” | BDS: Preface, 1-17 |  |
| 9/10 | M | 3 | L: Data Collection Day |  |  |
| 9/12 | W | 4 | P: Intro to R | BDS pgs. 175-198 |  |
| 9/17 | M | 5 | L: High-Performance Computing |  |  |
| 9/19 | W | 6 | P: Data Wrangling in R | BDS pgs. 199-260 |  |
| 9/24 | M | 7 | L: Modern DNA sequencing |  |  |
| 9/26 | W | 8 | P: Introduction to the Unix command line and common DNA sequence file formats | BDS pgs. 125-173, 339-352 |  |
| 10/1 | M | 9 | L: Alignment and DNA string comparison |  |  |
| 10/3 | W | 10 | P: Pairwise Alignment, BLAST, kmers | BDS pgs. 355-378 |  |
| 10/8 | M | 11 | *Columbus Day – No Class* |  |  |
| 10/10 | W |  | L: Multiple Sequence Alignment – *mafft, Muscle* |  |  |
| 10/15 | M | 12 | L: Phylogenetics |  |  |
| 10/17 | W | 13 | P: Evolutionary tree building – *phylotaR,* *RAxML, TNT* | TBA |  |
| 10/22 | M | 14 | --------------------------------- |  | **EXAM 1** |
| 10/24 | W | 15 | L: The importance of open-science | BDS pgs. 68-69 |  |
| 10/29 | M | 16 | L: Microbiomes/Metagenomics |  | Project Proposals DUE (Tue. 11PM) |
| 10/31 | W | 17 | P: Taxonomic classification of mixed samples – *Kraken/Centrifuge* | TBA |  |
| 11/5 | M | 18 | L: Molecular Medicine |  |  |
| 11/7 | W | 19 | P: Cancer Sequencing | TBA |  |
| 11/12 | M | 20 | L: Geospatial Bioinformatics |  |  |
| 11/14 | W | 21 | P: Biodiversity and Ecological Niche Modeling | TBA |  |
| 11/19 | M | 22 | P: Ecological Forecasting |  |  |
| 11/21 | W | 23 | *Thanksgiving break – No Class* |  |  |
| 11/26 | M | 24 | Python 1 – Libraries, loops, and functions |  |  |
| 11/28 | W | 25 | Python 2 – Biopython |  |  |
| 12/3 | M | 26 | Peer project consultation day |  |  |
| 12/5 | W | 27 | Term Project Presentation Day |  |  |
| 12/10 | M | 28 | Catch-up or Exam Review |  |  |
| TBA | -- |  | FINAL EXAM |  | Comprehensive Exam |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

\*These topics will be presented as pre-recorded lectures available on eLearn Blackboard

**EXAMS:**

One mid-term and one final exam will be given in this class. The date and location for the cumulative final exam will be posted on Stonehill’s website and announced in class.

**Assignments:**

**Homeworks:** Will be assigned on a weekly basis. These will typically be readings and reflection writing but will also include out-of-class data analysis.

**Practical Session Reports:**  Approximately every other class period in this course will consist of hands-on, practical learning. In these sessions we will be learning a new bit of bioinformatics software and applying it to some practice data. After each of these sessions you will be required to submit a “lab-report” like write-up of what went on during the practical session and what we found.

**Term paper/project/presentation:**  After the mid-term exam we will begin working on the term project for this course. Each student will identify an area of bioinformatics not covered in this course and do an exploratory project to investigate the kinds of analyses being done, the input data required, and the kinds of insites that can be derived from their chosen tool or method. Ideally it will be possible to develop a brief teaching module where the chosen tool or method is demonstrated for the class. Unlike a typical term paper, this assignment will consist of a brief project report (1-2 pages single spaced) and a short (10 minute) presentation to be given in front of the class at the end of the term. We will also devote a class period to peer review your developing projects.