Syllabus

MARB 6360
F 2-4:30 OCNR 240
Fall 2024
Office Telephone: 361-825-6024
Cell: 361-443-5676

Dr. Christopher E. Bird
Office: TH 234, Zoom
Office Hours: W-Th 1:30-4
E-Mail: cbird@tamucc.edu

Computation for 21st Century Biologists

Course Description

Catalog Course Description

This course is designed to prepare and enable students to use computational tools for bioinformatic applications in advanced courses and independent research projects. Students will be introduced to powerful open-source computing tools used in biological research for creation, organization, manipulation, processing, analysis, and archiving of big data. While not a formal requirement, it is assumed that students have a firm command of basic algebra.

Extended Course Description

This is a 3-credit course for undergraduates that introduces the powerful open-source computing tools that are used in biological research for the creation, organization, manipulation, processing, analysis, and archiving of "big data". This course is designed to prepare and enable students to use computational tools for bioinformatic applications in advanced courses and independent research projects. The primary topics covered are: data formats and repositories, command line Linux computing and scripting, regular expressions, super-computing, computer programming with PYTHON and R, data visualization with R, version control and dissemination of scripts and programs with GIT and GitHub, and typesetting with markdown. The usage of large language models (e.g. GPT) will be incorporated throughout the course.

Whether you want to learn basic data handling skills for your research project or you are curious about a career in bioinformatics and "big data", this course will provide you with the proper foundations.

Learning Objectives (Student learning objectives)

Upon the successful completion of this course, students should be able to:

1. ***Recognize, describe, and organize data into "tidy" data structures***
2. ***Locate scientific data repositories and download data***
3. ***Operate UNIX/LINUX (super)computers from command line***
4. ***Construct and modify computer programming/scripting logic structures for processing biological data***
5. ***Describe and use regular expressions to query data***
6. ***Use version control software (git) in coordination with GitHub to organize and manage projects***
7. ***Typeset with LaTeX or MarkDown***
8. ***Use the most popular open-source tools for biological data manipulation***
    a. ***Shell scripting (bash)***
    b. ***Statistical computing (R, tidyverse)***
    c. ***Scientific computing (python)***

Syllabus

9. Use large language models effectively to assist in data processing and teach others how you automated your data processing and analysis.
10. Automate the processing/analysis of a data set and create a GitHub repository that could be submitted as supplemental material for a manuscript to a peer-reviewed journal

Major Course Requirements

Computation for 21st Century Biologists will convene once a week for 2.5 hours. Class periods will involve interactive lectures that require each student to have a computer designed for content creation (Linux, OSX, Windows, not chrome, not iOS, not Android). Homework exercises will embellish upon concepts addressed in lecture. **Participation** involves attending lectures and performance on unannounced quizzes. Weekly **Assignments** will be given to reinforce concepts covered in lectures and encourage students to start using computational tools. **Exams** will be used to evaluate comprehension of the materials covered in lectures and assignments.

Rather than having a final exam, *graduate students* are expected to complete a **Final Project** involving the automation of the manipulation and/or analysis of data, This project, including the code should be archived on GitHub. A report written in Latex or Markdown will be due during the final exam period (this can be a markdown document in your GitHub repo, such as the README.md. The report should be concise in stating what the problem is, describing the strategy used for the solution, and describing how the code works (be sure to include a flow-chart or outline describing what code does). Those taking MARB 6360 will give a 10-15 minute presentation during the Final period on their project.

Project examples: automatically process data from experimental apparatus; image analysis; automated reporting of experimental results; downloading and organizing data from online repositories; etc…

Student learning outcomes will be assessed using in-class exercises, assignments, and exams. Your final grade will be based on the percentage you earn out of the total possible points, extra points may be built into exams or other assignments. It is also possible to lose points by turning in assignments late. Statistical manipulations to adjust grades, *if* used (at the Instructor's discretion), will be performed for each exam individually and all assignments in aggregate. A standard grading scale will be used:

A = 90 - 100 %
B = 80 - 89.9 %
C = 70 - 79.9 %
D = 60 - 69.9 %
F = 0 - 59.9 %

ACTIVITY	% of FINAL GRADE

Syllabus

Participation	10
Assignments	20
Exam 1	10
Exam 2	10
Final Project Report Final Project Presentation	40 10

Required or Recommended Readings

Textbook

R for Data Science. Grolemund and Wickham. (Free)

Recommended or Supplemental Reading

Computing skills for biologists: a toolbook. Allesina & Wilmes 2019.

Website:

https://github.com/tamucc-comp-bio/classroom_repo_2024

List of Supplies

A computer designed for content creation (Linux, Mac, Windows, not chrome, not iOS, not Android).

Course Policies

***Attendance/Tardiness*:** Attendance is expected. If you are late, don't make a disturbance and you will be responsible for catching yourself up to where we are.

***Late work and Make-up Exams*:** 10% of total possible score is deducted per day late. Inform professor as soon as you find out that you will miss and exam. Make arrangements with professor for make-up.

Extra Credit: No

Cell Phone/Electronic Device Usage: Required

Preferred methods of scholarly citations

Academic Integrity/Plagiarism

University students are expected to conduct themselves in accordance with the highest standards of academic honesty. Academic misconduct for which a student is subject to penalty includes all

forms of cheating, such as illicit possession of examinations or examination materials, falsification, forgery, complicity, or plagiarism. (Plagiarism is the presentation of the work of another as one's own work.) In this class, academic misconduct or complicity in an act of academic misconduct on an assignment or test will result in a failing grade.

AI Requirement Statement

Students are expected to use "artificial intelligence" (e.g. large language models) in this course. When you submit work created using artificial intelligence, you are responsible for the content, not the ai. If you used ai in a novel fashion, you should describe how it was used to accomplish the task.

Dropping a Class

I hope that you never find it necessary to drop this or any other class. However, events can sometimes occur that make dropping a course necessary or wise. ***Please consult with your academic advisor, the Financial Aid Office, and me, before you decide to drop this course.*** Should dropping the course be the best course of action, you must initiate the process to drop the course by going to the Student Services Center and filling out a course drop form. Just stopping attendance and participation WILL NOT automatically result in your being dropped from the class. (12/3) is the last day to drop a class with an automatic grade of "W" this term.

Classroom/Professional Behavior

Texas A&M University-Corpus Christi, as an academic community, requires that each individual respect the needs of others to study and learn in a peaceful atmosphere. Under Article III of the Student Code of Conduct, classroom behavior that interferes with either (a) the instructor's ability to conduct the class or (b) the ability of other students to profit from the instructional program may be considered a breach of the peace and is subject to disciplinary sanction outlined in article VII of the Student Code of Conduct. Students engaging in unacceptable behavior may be instructed to leave the classroom. This prohibition applies to all instructional forums, including classrooms, electronic classrooms, labs, discussion groups, field trips, etc.

Statement of Civility

Texas A&M University-Corpus Christi has a diverse student population that represents the population of the state. Our goal is to provide you with a high-quality educational experience that is free from repression. You are responsible for following the rules of the University, city, state and federal government. We expect that you will behave in a manner that is dignified, respectful and courteous to all people, regardless of sex, ethnic/racial origin, religious background, sexual orientation or disability. Behaviors that infringe on the rights of another individual will not be tolerated.

Statement of Academic Continuity

In the event of an unforeseen adverse event, such as a major hurricane and classes could not be held on the campus of Texas A&M University–Corpus Christi; this course would continue through the use of Canvas and/or email. In addition, the syllabus and class activities may be modified to allow continuation of the course. Ideally, University facilities (i.e., emails, web sites, and Canvas) will be operational within two days of the closing of the physical campus.

Syllabus

However, students need to make certain that the course instructor has a primary and a secondary means of contacting each student.

Mental Health and Well-Being
The university aims to provide students with essential knowledge and tools to understand and support mental health. As part of our commitment to your well-being, we offer access to Telus Health, a service available 24/7/365 via chat, phone, or webinar. Scan the QR code to download the app and explore the resources available to you for guidance and support whenever you need it.

Statements on Student Grade Appeals, Disabilities Accommodations, Civil Rights Reporting, and Campus Emergencies can be accessed at https://www.tamucc.edu/faculty/faculty-affairs/assets/syllabi-statements.pdf

Schedule
(course outline)

Date	Lecture Topic	HW Due
	Theme I: Welcome to the Matrix	
Wk 0	1. Course overview 2. Biological Data Repositories, Structures, Formats 3. Computer set up	
Wk 1	Linux Boot Camp I 1. UNIX philosophy 2. Navigating/creating/manipulating directories & files 3. How to get help: man pages 4. Basic commands: cd, ls, cp, mv, mkdir, rm, tr, cut, cat, head, tail, 5. Commands useful for manipulating data files in text streams	Assignment 0
Wk 2	Linux Boot Camp II 1. Wildcards, substituting characters, permissions, sudo 2. Pattern matching with grep & regex 3. Intro to Computer Programming: Shebang!, Scripting, For Loops	Assignment 1
Wk 3	Linux Boot Camp III 1. More Computer Programming with bash 2. Logic: if-then-else; looping with while, and GNU parallel 3. Functions: diy commands 4. Advanced text stream manipulation: sed, paste, …	Assignment 2 SuperComputer Acct
Wk 4	Version Control & Supercomputing 1. Linux repositories and tools for biologists	Assignment 3

Syllabus

	2. Version control with git 3. Super computing, ssh	
	Theme II: Wrangling and Visualizing Data With R	
Wk 5	R Boot Camp I 1. R Philosophy 2. Command line R 3. R data types & structures 4. Math, equalities, logic 5. Basic statistical functions 6. Reading and writing data	**Exam 1** Install R & R Studio
Wk 6	R Boot Camp II 1. Scripting & writing good code 2. Loops & if-then decision logic 3. R Studio 4. Functions 5. Libraries 6. Random numbers 7. Vectorized loops 8. Debugging 9. More basic stats 10. Base R plots	Assignment 5
Wk 7	R Boot Camp III 1. tidyverse 2. Basic reading & manipulating data 3. Basic computing statistics 4. Visualization of data w ggplot2	Assignment 6
Wk 8	R Boot Camp IV 1. Advanced tidyverse, pipelines 2. Manipulating & wrangling data	Assignment 7
	Theme III: Programming the Matrix	
Wk 9	Python Boot Camp I 1. Intro to Python 2. Data structures 3. Functions 4. Decision logic and loops 5. Reading and writing files	**Exam 2** Install Anaconda
Wk 10	Python Boot Camp II 1. Writing code 2. Modules & Program Structure 3. Errors and exceptions 4. Debugging 5. Testing & Profiling	Assignment 9

Syllabus

Wk 11	Scientific Computing w/ Python 1. NumPy and SciPy 2. Pandas 3. Biopython 4. Other modules	Assignment 10
Wk 12	Scientific Typesetting w/ Latex 1. Latex document structure 2. Typsetting 3. Latex packages for biologists	Assignment 11
Wk 13	Putting It All Together	Assignment 12
Final	**Final Exam: Becoming THE ONE**	

Date	Graduate Students: Final Project Schedule
	Theme I: Welcome to the Matrix
Wk 0	
Wk 1	
Wk 2	Submit Project Idea (2.5%)
Wk 3	
Wk 4	Submit Project Plan/Outline (2.5%)
	Theme II: Programming the Matrix
Wk 5	Link to GitHub repository for project w/ readme (2.5%)
Wk 6	Commit at least 1 working function to GitHub (2.5%)
Wk 7	Commit at least 2 working functions with data I/O to GitHub (2.5%)
Wk 8	
	Theme III: Becoming THE ONE

Syllabus

Wk 9	
Wk 10	Latex or Markdown draft/ progress report (see syllabus section G for description of report); include a description of tasks left to achieve. Code and data committed to GitHub (2.5%)
Wk 11	
Wk 12	
Wk 13	Final Report in Latex or Markdown; Working code and data committed to GitHub. (25%)
Final	MARB 6360 Final Presentations