BioSci 1540: Computational Biology

Nathan L. Brouwer

2022-08-19

Contents

Chapter 1

Welcome

This is the syllabus.

Chapter 2

Introduction

This is the syllabus for BIOSCI 1540: Computational Biology.

This is both the first course in the core sequence of classes of the Computational Biology major, as well an elective for most Biological Sciences majors.

This syllabus contains both key course policies and also is a repository for useful information about the rapidly expanding and evolving field of computational biology.

The first several pages contain key information about the course. Subsequent pages are organized alphabetically by major topic.

Pages marked "**Appendix**" are additional information about computational biology or undergraduate research that you may find useful.

Chapter 3

Meeting times

The course will meeting **in-person** on Tuesdays and Thursday.

Location: 1502 Posvar Hall **Days:** Tuesday & Thursday
Times: 4:00PM - 5:15PM Aug 30, 2021-Dec 8, 2022

There is no recitation for this course.

The official Pitt **Academic Calendar** can be found here

Chapter 4

Instructor

Nathan L. Brouwer, PhD Office: Langley Hall A255
In the "Langely Bridge" where the advisors offices used to be. Email: nlb24 at pitt.edu
Github: https://github.com/brouwern

Chapter 5

Academic integrity

Dr. Brouwer, the Bio Sci department, and the University all take academic integrity very seriously. Cheating includes any form of plagiarism, including copying other students' work or using other resources without proper attribution.

1. If you are caught cheating on a graded assignment, you will receive a zero on the assignment.
2. If you are caught cheating on an exam, you will receive a zero on the exam, an F in the course, and an Academic Integrity Violation Report will be filed.

Below is the University's Policy on Academic Integrity:

"Students in this course are expected to comply with the University of Pittsburgh School of Arts & Sciences Academic Integrity Code located at www.as.pitt.edu/faculty/policy/integrity.html. Any student suspected of failing to meet the student obligations of the code during the semester will be required to participate in the procedures for adjudication, initiated at the instructor level. This may include, but is not limited to, confiscation of the assignment of any individual suspected of violating the code. A minimum sanction of a zero score for the assignment will be imposed. Violation of the Academic Integrity Code requires the instructor to submit an Academic Integrity Violation Report to the Dean."

Chapter 6

Assessments Overview

The primary forms of assessment in the course will i) exams, ii) regular submission of "portfolio" assignments, and iii) individual projects.

6.1 Primary assessments

The key assessments are:

1. **1 R programming basics quiz**: A "take-home" Canvas quiz at the end of the 2nd week of class.

2. **3 Unit Exams**: In-class exams to test you mastery of computational biology concepts and their implementations in R.
3. **1 Capstone Exam**: During the **final exam** period you will take an exam on material from the 4th and final unit of the class, which serves as a capstone unit where themes from throughout the course are brought together. This is a semi-cumulative exam.

4. **1 to 2 weekly portfolio assignments**: Each week you will be given an assignment to test your understanding of key concepts and programming principles. Through these assignments you will build up an **online portfolio** of all your code from throughout the semester.
5. **1 to 2 individual projects**: Over the course of the semester you will begin working on individual-based analyses of unique biological datasets or problems. You will build up your analyses incrementally over the course of the semester by adapting code from your portfolio assignments to your individual datasets.

6.2   Secondary assessments

1. **Software checkpoints** to assure that you have the proper software involved. These are usually completed by submitting a screen grab of your computer console showing successful installation of the software.
2. **Code checkpoints** to assure that you can execute key functions correctly before attempting portfolio assignments. These are usually completed by submitting a screen grab of your computer console showing successful execution of the code.
3. **In-class TopHat** polls, questions and small assignments. These are typically graded for participation.
4. **In-class Worksheets**; you'll usually complete these in-class and submit and image of them online via Canvas. These are typically graded for participation / completion.
5. **Other assignments**: Other small assignments may be given for completion outside of class.

6.3   Pracice problems

Other materials will be provided for optional practice. These will be set to zero points.

1. **Canvas Quiz questions** related to vocab, code, concepts etc.
2. **Optional Portfolio assignments**
3. **Previous exam questions** or full **Practice tests**. These questions may not always align perfectly with current course content but will give you an idea of the types of questions that will appear on an upcoming exam.

6.4   Test-taking policies

Except for the first R quiz, tests will be in-person during the regularly schedule time and administered as **Canvas Quiz.** Components of exam typically include

1. Stand-alone canvas quiz questions
2. Code-based questions which you'll answer via a separate file, submitted via Canvas. Submissions are typically images or code files.

6.5   Key test policies

1. All answers will be submitted by the end of the normal class period.
2. With prior approval you will be able to take 1 exam remotely via Zoom over the course the semester. The final must be taken in-person.
3. Your lowest of the three regular unit exams will be dropped, e.g. if you miss an exam entirely due to illness. The final exam cannot be dropped.
4. Make-up exams and exam re-takes will not be offered.

5. It is your responsibility to assure that you have adequate power in your devices and internet access to complete exams. Please take your power cord to class!

6. Adjustments will be made as necessary for students with DRS accommodations.

7. You may not communicate with other students while taking a test.

6.6 Test dropping policy

Your lowest regular unit test will be dropped. The final exam CANNOT be dropped.

6.7 Final Exam Policies

The final will cover conceptual and programming material covered after the last regular unit exam. You will also need to draw on R skills from throughout the semester.

The final exam will …

1. Take place during finals week on the day scheduled by the University
2. Is required
3. Follow the general format of other tests.

The final exam is required and will be scheduled as per the University calendar.

As noted elsewhere your lowest test score will be automatically dropped and your score for tests calculated based on your 2 best tests.

6.8 Exam "buffer question" policy

The unit tests and the final exam will have 1 to 2 "buffer point." The way this work is that there will be, for example, 26 questions but the test will be worth a maximum of 25 points. The maximum score on the test will be capped at 25 points. If you miss 1 question you will get 25/25 points instead of 25/26. If you miss 2 questions you will get 24/25 and so forth. Scores greater than 100% will *not* be given.

Chapter 7

Assessments - Homework policies

Homework for outside of class will be assigned via Canvas.

ß **Late assignments:** Canvas will automatically deduct 0.5% of your score on all non-test assignments per hour, rounding to the nearest hour (12%/day). 29 minutes = 0 hours = full credit, 31 minutes = 1 hour.

Chapter 8

Course communication - Overview

All important details will be sent via Canvas messages and mentioned/discussed in class. It is your responsibility to regularly check Canvas messages to stay abreast of the course. Most communication will occur on Fridays, but important updates may be sent out at other times if necessary.

Chapter 9

Course communication - Email and Canvas Messages

Announcements to the whole course via Canvas will occur typically on **Friday** afternoons. Other communication related to course administration will also occur via Canvas messages.

Your can message me on Canvas or email me at nlb24 at pitt.edu.

Please include "Comp Bio" as the 1st thing in your email subject line with an informative bit of information as the "...". e.g. "Comp Bio 2: problem accessing TopHat". (Questions related to code are best submitted via GitHub).

If you use another email service as your primary email (eg GMail) please set your Pitt email to forward there.

For info on forwarding your Pitt email to your personal account follow this link: https://bit.ly/2Riz7dx

I try to answer all emails received on weekdays before 5 pm within 24-36 hrs. Emails received after 5 pm will be answered at the earliest the following morning. Emails received on the weekend will be answered Monday.

Please consult this syllabus before asking questions about course policies and the schedule, and refer to relevant information such as URLs, subject headings or dates. Screengrabs are super helpful. If the entire answer to your question can be found in the syllabus I will likely respond by saying "*This is in the syllabus, Cheers, Dr. B.*".

Questions relevant to the whole class may be re-posted (with identifying details removed) to Canvas.

Chapter 10

Communication - University email policy

Each student is issued a University e-mail address (username at pitt.edu). This e-mail address may be used by the University for official communication. Students are expected to read e-mail sent to this account on a regular basis. Failure to read and react to University communications in a timely manner does not absolve the student from knowing and complying with the content of the communications.

The University provides an e-mail forwarding service that allows students to read their e-mail via other service providers. Students that choose to forward their e-mail from their pitt.edu address to another address do so at their own risk. If e-mail is lost as a result of forwarding, it does not absolve the student from responding to official communications sent to their University e-mail address.

To forward e-mail sent to your University account, go to http://accounts.pitt.edu, log into your account, click on Edit Forwarding Addresses, and follow the instructions on the page. Be sure to log out of your account when you have finished.

For the full E-mail Communication Policy, go to bc.pitt.edu/policies/policy/09/09-10-01.html.

Chapter 11

Catalog description

The catalog description of this course is:

> "This course is designed to give students a broad understanding of how computational approaches can be used to solve problems in biology. Both the biological and computational underpinnings of the methods will be addressed."

Chapter 12

Course materials overview

Thanks to an Open Educational Resources (OER) grant from The Office of the Provost, there are no books to buy for this course - all course materials are free and will be provided digitally through Canvas.

Reading will take several forms:

1. files containing computer code (scripts)
2. articles and blog posts,
3. textbook readings

Chapter 13

Course materials - Software

Installation and account creation for all software will be covered in detail during class or via assignments and accompanying videos. If you want to get prepared early for the course you can explore these on your own.

For information on getting started with R see Get R Done!.

13.1 Primary software

13.1.1 Programming environment: R

We will use the open-source language R for almost everything in this course.

NOTE: If you have used R before that's awesome!

However, **PLEAE PLEASE PLEASE update your installation when the class begins** - R goes through regular updates and works best with the newest version.

If you are having trouble installing R an excellent resource is: https://learning statisticswithr.com/book/introR.html

REFERENCE
R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/.

13.1.2 Integrated developement environment (IDE)

We'll use RStudio as the front-end for working with R.

NOTE: If you've use RStudio before **please update it** when the course begins.

[RStudio](https://www.rstudio.com/: www.rstudio.com/products/rstudio/download/#download

13.1.3   Cloud-based R

We will use several web-based services to run R. First, we will use the online machine-learning platform Kaggle and the Notebooks feature to run code in a user-friendly environment. You can make a Kaggle account with your University or personal email address.

Additionally, as a back-up to running R on your desktop I'll also introduce you to [RStudio Cloud](https://rstudio.cloud/. There are also several excellent primers which we'll use.

RStudio cloud: https://rstudio.cloud/

13.2   R packages

Many R **packages** (aka libraries) will be used in this course. Below are several key ones.

13.2.1   swirl

swirl is an R package that sets up interactive tutorials in your R console. It works both on desktop computers and RStudio cloud and is an excellent tool for learning and reviewing R.

Details for installing swirl can be found here in Get R Done!.

13.2.2   Bioconductor

Bioconductor is a clearing house and repository for high-grade bioinformatics and computational biology software using R.

Details for installing Bioconductor can be found here.

Chapter 14

Disability Resource & Services

If you have a disability for which you are or may be requesting an accommodation, you are encouraged to contact both your instructor and Disabilities Resources and Services. DRS will verify your disability and determine reasonable accommodations.

216 William Pitt Union
(412) 648-7890
(412) 383-7355 (TYY)

Chapter 15

Extra credit policies

No extra credit will be offered in this course.

"Buffer questions" on the mini-tests and the final are NOT extra credit.

Chapter 16

File names for homework

16.1   Files created by you

All files your create on your own which get submitted for homework need to be in lower case and **strictly** adhere to this format:

"pittemail_lastname_assignmenttitle"

Where I will provide the assignment title.

For example:

"nlb24_brouwer_assign_1.rmd"

16.2   Incorrectly named files

Files will be downloaded and processed automatically so it is essential that your files are named correctly; errors may result in your assignment receiving a zero. There will be a grace period of 2 weeks where I will work with everyone to understand and use the file naming system.

It is your responsibility to make sure you file names are correct. If I spot an incorrectly named file I may have time to inform you so you can correct it, but this may not be until after the assignment is due and will result in points being lost as per the late assignment policy.

Chapter 17

Grading scale

Final.Percentage	Grade	GPA
98–100%	A+	4
92–97	A	4
90–91	A-	3.75
88–89	B+	3.25
82–87	B	3
80–81	B-	2.75
78–79	C+	2.25
72–77	C	2
70–71	C-	1.75
68–69	D+	1.25
62–67	D	1
60–61	D-	0.75
59 and below	F	0

Note: Students planning to major in Biological Sciences must pass this course with a C (not C- !) or better. Rounding is not done until final grades are computed and is done by computer to 1 decimal place.

Final letter grades are assigned after rounding and is done automatically by Canvas including the decimal value. For example, a score of 91.99% rounds to 92.0% and is an A, but a score of 91.94% rounds to 91.9% and is an A-.

Note: Students planning to major in Computational Biology must pass this course with a C (not C- !) or better.

Chapter 18

Masks and COVID-19

During this pandemic, it is extremely important that you abide by the public health regulations, the University of Pittsburgh's health standards and guidelines, and Pitt's Health Rules.

These rules have been developed to protect the health and safety of all of us. Please check your Pitt email for updates to the current masking policy.

Chapter 19

Metnal health & wellness

School is hard - please take care of yourself as best you can given the many demands on your time. Diminished mental health, including significant stress, mood changes, excessive anxiety, or problems with sleeping can interfere with your academic performance. You have a support network here at Pitt to help you through challenging times. Acknowledging that you need help, and getting that help, is smart and courageous.

- If you are in an EMERGENCY situation, call 911 or Pitt Police at 412-624-2121.
- If your symptoms are due to FINANCIAL strain, please visit pitt.libguides.com/assistanceresources to see all available University resources.
- If your symptoms are due to strained RELATIONSHIPS, families, or personal crises, please visit the University Counseling Center at www.studentaffairs.pitt.edu/cc/ for free confidential services.
- If your symptoms are strictly related to your COURSE WORK and performance in this course, please contact me.

19.1 Resources:

University Counseling Center: 412-648-7930
Sexual Assault Response: 412-648-7856
RE:SOLVE crisis network: 888-796-8226
Pitt Police: 412-624-2121
Pitt Pantry food bank: https://www.studentaffairs.pitt.edu/pittserves/the-pitt-pantry/

Chapter 20

Office hours

20.1 INSTRUCTOR OFFICE HOURS

MONDAY: Noon-1 pm. Zoom: https://pitt.zoom.us/j/92563415506 Password: "BLOSUM" (all caps)

WEDNESDAY: Noon-1 pm Zoom: https://pitt.zoom.us/j/92563415506 Password: "BLOSUM" (all caps)

Other days: By appointment: calendly.com/brouwern

Office hours will be via Zoom and in-person.

20.2 UTA OFFICE HOURS

UTA office hours will be posted by the beginning of the 2nd week of lectures.

20.3 Individual appointments

I'm available for individual 15 minute appointments 10 am to noon on Tuesdays, Thursdays and Fridays. Schedule these appointments via Calendely:https://calendly.com/brouwern/15min

If office hours or available times for appointments do not work for you, please email me to coordinate a different meeting time.

Chapter 21

Point breakdown

You can view the point breakdown and a summary of related policies here. Commenting is enabled; feel free to post questions.

Chapter 22

Semester schedule

The schedule of tests is available here.

A calendar for the semester will be posted with key University dates and tests.

Chapter 23

Weekly schedule

A summary of weekly events is available here. Commenting is enabled; feel free to post questions.

Each Friday I will send out an message via Canvas that will layout the topics and tasks for the following week.

Chapter 24

TopHat

We will use TopHat in-class and for homework. Please bring a charged TopHat compatible device to all lectures and recitations.

TopHat
Join Code: **209015**
www.tophat.com

When logging into TopHat it will ask you for your university. Typing in "University of Pittsburgh" brings up 3 options: **select the first one** that just says "University of Pittsburgh".

Chapter 25

Updates to schedule & syllabus

I reserve the right to update the syllabus, schedule, point allocation and all other components of the course as necessary.

If changes occur after the first day of class, they will be clearly communicated in class and via email, and a revised syllabus and schedule distributed with major changes flagged.

Chapter 26

Videos of lectures

This is an in-person course. Lectures will not be live-streamed. Recordings of at least the audio of each lecture will be posted for reference. I can briefly review key parts of lectures during office hours but will not fill in all the gaps in your notes if you missed class. It is your responsibility to watch the class recordings and obtain notes you missed from classmates.

Chapter 27

Zoom

I will use Zoom for recording my lectures and posting them to Panopto but will not release the link for watching the lecture off-site.

Each student can take 1 regular unit exam remotely during the scheduled exam period. You must contact me ahead of time to do this.

The final exam must be taken in-person.

Chapter 28

Appendix: graduate school

If you are considering graduate school you may wish to check out these resource:

So you want to go to graduate school? O'Connel, Tim. "A conversation about grad school" https://eatmorecookies.wordpress.com/2018/07/17/a-conversation-about-grad-school/

Carson, W.P. 1999. A primer on how to apply and get admitted to graduate school in ecology and evolutionary biology. Bulletin of the Ecological Society of America 80(4):246-250. (PDF here: https://sites.google.com/site/walterpagecarson/publications)

Stearns, S. Some modest advices for graduate students Bulletin of the Ecological Society of America, 1987 - JSTOR https://stearnslab.yale.edu/some-modest-advice-graduate-students

Thoughts on Stearns and Huey Fox, J. 2015. Re-reading Stearns and Hueys some modest advice to gradate students. https://dynamicecology.wordpress.com/2015/09/08/rereading-stearns-and-hueys-some-modest-advice-to-graduate-students/

On the publishing process: a tweet by dsquintana https://twitter.com/dsquintana/status/1154820610750595072?s=20

Chapter 29

Cheat sheets

RStudio produces a number of cheatsheets; the most useful ones for beginners are below.

Basic R code
http://github.com/rstudio/cheatsheets/raw/master/base-r.pdf

RStudio https://github.com/rstudio/cheatsheets/raw/master/rstudio-ide.pdf

RMarkdown https://github.com/rstudio/cheatsheets/raw/master/rmarkdown-2.0.pdf
https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf

Translations of some of these cheat sheets can be found at the bottom of the RStudio page.

Chapter 30

Appendix: Getting involved in (computational) research

Approaching a PI - A guide for undergraduates

https://thefemalescientist.com/guide/madeleine-hann/2053/approaching-a-pi-a-guide-for-undergraduates-in-stem/

Chapter 31

Shortcuts in R

Programmers are often obsessed with using keyboard shortcuts. You'll see me use some, and you might want to learn them.

The back of this cheatsheet is a 1-page summary of shortcuts https://github.c om/rstudio/cheatsheets/raw/master/rstudio-ide.pdf

Another summary sheet of Shortcuts in RStudio is here: https://support.rstu dio.com/hc/en-us/articles/200711853-Keyboard-Shortcuts

A great blog post on shortcuts is here https://appsilon.com/r-studio-shortcuts-and-tips/