# BioSci 1540: Computational Biology

Nathan L. Brouwer

2021-08-25

# Contents

# Chapter 1

# Welcome

This is the syllabus.

```r
install.packages("bookdown")
# or the development version
# devtools::install_github("rstudio/bookdown")
```

# 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 and also an elective for most Biological Sciences majors.

This syllabus contains both key course policies and also is a repository for useful information about 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 research that you may find useful.

Your first assignment in the class will be to read through the syllabus and complete a "Syllabus treasure hunt" assignment on Canvas.

If you have a question about the syllabus (after you complete the syllabus treasure hunt) or even just spot a typo, post a comment on GitHub at https://github.com/brouwern/BIOSCI_1540

You'll be shown how to add comments on GitHub in class. The next page of the syllabus is a page for practicing adding comments.

# Chapter 3

# Github comments practice

Scroll down and locate the heading for the leter your last name begins with. Then locate the bullet point for theletter your first name begins with and add a comment to that line of code.

If you intials do not appear add a comment to the "other" label.

## 3.1 A

- E
- J
- other

## 3.2 B

- A
- H
- L
- M
- O
- S
- other

## 3.3 C

- E
- J
- M
- R

- Z
- other

## 3.4   D

- D
- J
- K
- L
- La
- Le
- Li
- Lo
- R
- other

## 3.5   E

- Z
- other

## 3.6   F

- C
- G
- L
- Z
- other

## 3.7   G

- E
- J
- M
- S
- other

## 3.8   H

- D
- Ja
- Je
- Ji

- Jo
- S
- other

## 3.9 J

- V
- other

## 3.10 K

- M
- other

## 3.11 M

- G
- C
- N
- Na
- Ne
- Ni
- S
- other

## 3.12 N

- M
- other

## 3.13 N

- N
- other

## 3.14 P

- C
- J
- H
- S
- other

## 3.15   P

- J
- other

## 3.16   R

- A
- C
- R
- Ra
- Re
- Ri
- Ro
- V
- Va
- Ve
- Vi
- other

## 3.17   S

- A
- B
- D
- E
- En
- Er
- I
- J
- S
- V
- Va
- Ve
- Vi
- Vo
- other

## 3.18   T

- A
- other

## 3.19   V

- T
- Ta
- Ti
- Te
- To
- other

## 3.20   W

- A
- G
- other

## 3.21   X

- X
- other

## 3.22   Y

- H
- S
- X
- other

## 3.23   Z

- Y
- other

# Chapter 4

# Meeting times

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

**Location:** 1502 Posvar Hall **Days:** Tuesday & Thursday
**Times:** 4:00PM - 5:15AM
Sept 31, 2021-Dec 10, 2021

The first two weeks all lectures will be live-streamed on Zoom: https://pitt.zoom.us/j/99372114800
Password: "BLOSUM"

There is no recitation for this course.

The official Pitt **Academic Calendar** can be found [here](https://catalog.upp.pitt.edu/mime/media/212/3179/Academic+Calendar+2021-2022.pdf: https://catalog.upp.pitt.edu/mime/media/212/3179/Academic+Calendar+2021-2022.pdf

# Chapter 5

# Instructor

**Nathan L. Brouwer**, PhD Office: Langley Hall 215A
(In the former Neuroscience office suite)
Email: nlb24 at pitt.edu
Github: https://github.com/brouwern

# Chapter 6

# 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 7

# Course communication - Overview

## 7.1 Primary modes of communication

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 communcation will occur on Friday's, but important updates may be sent out at other times.

## 7.2 Secondary modes of communication

### 7.2.1 Github "issues"

In this class we will be making extensive use of the code management site GitHub. Most communication about homework and other assignments will occur via GitHub's Issues] feature. In particular, to foster discussion we will use GitHub Issues to create forums around assignments and you will be required to submit comments.

### 7.2.2 Microsoft SharePoint

Files containing code for this course will be distributed using Microsoft Sharepoint. Sharepoint has messaging features which I may pilot test but will not rely on.

# Chapter 8

# 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 reposted (with identifying details removed) to Canvas.

# Chapter 9

# 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 10

# 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 11

# 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.

Importantly, we'll be taking a very novel approach to the textbook for this course: **you will actually be building it yourself!**

Details will be provided in-class, but the gist regarding the textbook is:

1. Most weeks you will be provided 1 or 2 chapters of the book in the form of computer files ("scripts").
2. For homework you will do exercises embedded in the chapters.
3. When the exercises are successfully completed you will be able to convert the files to Web Pages and post them to your own Portfolio of book chapters and other completed activities.

# Chapter 12

# 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!.

## 12.1 Primary software

### 12.1.1 Programming environment: R

We will use the open-source language R)(https://cran.r-project.org/) for almost everything in this course.

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

However, **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/.

### 12.1.2 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

### 12.1.3 Cloud-based R

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/

## 12.2 Code sharing

### 12.2.1 Microsoft SharePoint

R code and other files will be distributed to you via a personal folder on **SharePoint**, accessed via your **OneDrive** account.

You will need to "sync" your OneDrive account to your computer desktop to allow me to help you with your work. No code files will be shared via email or Canvas during the course.

**NOTE**: If you do not have access to a computer you can sync your OneDrive account please contact me so we can workout a system.

### 12.2.2 GitHub

Most code for this course will be available on GitHub and you will need to create an account for some activities.

Links will be distributed directly to you, but you can also see key resources pinned to my profile at https://github.com/brouwern

## 12.3 R packages

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

### 12.3.1 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!.

### 12.3.2 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 13

# 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 14

# 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.

Exam revision assignments are not extra credit and do not add points to your test score.

# Chapter 15

# File names for homework

## 15.1   Files provided by me

You will be provided templates for most files you end up submitting for homework in this class. These file names will contain:

1. Your Pitt user name (e.g. nlb24)
2. Your last name or a concatenated version of it (e.g. brouwer)

If you recieve an incorrect file name you may update it with your user name and last name. Otherwise, Please do not alter these file names unless instructed.

## 15.2   Files created by you

All files your create on your ownwhich 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"

## 15.3   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 16

# 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. Universal face covering is required in all classrooms and in every building on campus, without exceptions, regardless of vaccination status. This means you must wear a face covering that properly covers your nose and mouth when you are in the classroom. If you do not comply, you will be asked to leave class. It is your responsibility to have the required face covering when entering a university building or classroom. For the most up-to-date information and guidance, please visit coronavirus.pitt.edu. and check your Pitt email for updates before each class.

# Chapter 17

# 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.

## 17.1   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 18

# Office hours

## 18.1   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 offered initially only via Zoom.  After the first 2 weeks of the semester I evaluate if I will begin in-person office hours on Wednesdays.

## 18.2   UTA OFFICE HOURS

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

## 18.3   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 19

# 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 20

# 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 21

# Videos of lectures

The first two weeks of class lecture will be recorded using Zoom and posted to Panopto.

After the first lecture I will distribute the Panopto folder where the videos will be saved.

Note: Since this class is not meant to be taken asynchronously I will not actively support use of these videos.

# Chapter 22

# Zoom

Class will be in-person and streamed live via Zoom the first two weeks.

https://pitt.zoom.us/j/99372114800
Password: "BLOSUM"

# Chapter 23

# 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/walterpage carson/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.wordpre ss.com/2015/09/08/rereading-stearns-and-hueys-some-modest-advice-to-graduate-students/

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

# Chapter 24

# 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 25

# 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 26

# 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/