

Computational Tools for Research in Biology (Fall 2021)

BSC 4452— Class Number: 21520

BSC 6451— Class Number: 27027

Meeting: Period 3 (9:35-10:25) in MAEB 0234 (or zoom for online section)

Instructor Information:

Prof. Edward L. Braun

Office Hours: By appt. (514 Carr)

(352) 846-1124

ebraun68@ufl.edu

Office hours:

I will be available in my office Monday's after class – just walk back with me or pop in. You can also make an appointment for a face-to-face or zoom meeting. If you do not wish to wear a face covering in my office, we will move the meeting to zoom.

Course Description:

Information technology has dramatically transformed how research across many disciplines is conducted. This is particularly true in the biological sciences where researchers frequently find themselves faced with massive amounts of diverse data to analyze. As data types and volumes continue to grow, knowledge of scripting, database management, and advanced computing skills are critical for researchers.

Additional Course Information:

Topics covered in the course will address a gap in how research has advanced—and become increasingly computational— while student training in the use of computational tools has lagged.

The course assumes no prior coding or command line skill and covers concepts that will provide the ability for students to apply new technologies to a wide array of research questions. A foundation in data management and analysis concepts opens doors for well-trained scientists and allows them to work in multidisciplinary research.

This course will survey areas where high performance computing, large-scale data access and integration, informatics tools and software, and multi-disciplinary collaboration have had high impact on research as a foundation for computationally enabled research.

Textbooks:

The main texts for the course are:

- *The Linux Command Line*: <http://linuxcommand.org> (my notes and referenced page numbers will be based on the 19.01A PDF, Fifth internet edition from Jan 28, 2019)
- *Python For Everyone*: <https://www.py4e.com/book.php>

Each of these is available as a free PDF download (or for purchase in print).

I will also make some google docs available and may assign readings from the peer-reviewed scientific literature.

Course Assignments and Grading:

Course component	Proportion of grade
Participation	35%
Class project	35%
Take home exercises	30%

Class participation

I expect regular attendance. If you are sick, have a family emergency, or plan to participate in a university-sanctioned event just let me know; those absences are excused. Likewise, I will consider excusing you for any reasonable professional development activities; just discuss them with me. But otherwise I expect to see you in class and you will lose 1% of this portion of the grade for unexcused absences. Although some classes will be lectures many will include discussions; I expect everybody to participate. We will also be learning how to use specific computer programs during some classes and participating in those activities is important.

Take home exercises

There will be no tests, but I will expect you to produce short (1-2 paragraph) summaries of assigned readings.

Grading scale

The following grading scale will be used (values are percentages):

91-100 A	88-90.9 A-	
85-87.9 B+	81-84.9 B	78-80.9 B-
75-77.9 C+	68-74.9 C	65-67.9 C-
62-64.9 D+	58-61.9 D	

Below 58 is failing (E).

Course policies

* Academic dishonesty will not be tolerated. If cheating or plagiarism is suspected, all persons involved will receive a zero on the affected assignment and it will be reported to the Dean of Students Office.

* If you notice another student engaging in activities to you believe to constitute academic dishonesty please report the potential violation. I take all allegations seriously. However, remember that you are bound by the UF honor code, which states that “*in reporting an alleged Student Honor Code violation, a student shall not intentionally or in bad faith make a false or misleading statement.*” In other words, I view intentional false allegations as a type of academic dishonesty.

*Attendance in class is required. However, if you cannot attend on specific days just email with the subject line “UNABLE TO ATTEND CLASS ON <DATE>” and provide a brief explanation. If you have concerns at any time just discuss them with me. My goal is to make the course enjoyable and informative so you will want to attend!

University support services

Resources are available on campus to help students meet academic goals and solve personal problems, which interfere with their academic performance. Resources include:

1. [UF Counseling and Wellness Center](#), 301 Peabody Hall and Radio Rd Facility, 392-1575, personal and career counseling.
2. [Career Resource Center](#), Reitz Union, 392-1601, career development assistance and counseling.
3. [CLAS Academic Advising Center](#), Farrior Hall, 100 Fletcher Drive, 392-1521, provides advise on course selection and course planning to meet graduation requirements

Disability Notice

Students with disabilities enrolled in this course and who may need disability-related classroom accommodations are encouraged to make an appointment to see me before the end of the second week of the term. All discussions will remain confidential, although the [Student Accessibility Services](#) office may be consulted to discuss appropriate implementation of any accommodation requested.

Schedule

NOTE: this schedule is may (and probably will) change – I’ll keep everybody updated)

<i>Week</i>	<i>Reading/Assignment</i>	<i>Topic</i>
1	<i>Download Software</i>	<i>Introduction and course objectives</i>
1	<i>Read TLCL Introduction & Ch 1-4</i>	

Watch UFRC Training Video

UF Research Computing Intro & getting started

1 *Read TLCL Ch 5-8*

2 *Read Notes on Regular Expressions and TLCL Ch 19*

Problem Set 1

2 *Quiz 1*

GitHub Account assignment

Read TLCL Ch 20

3 *Quiz 2*

Watch Learn the Linux Command Line

Register for github.com account

Read TLCL Ch 24-26

3 *Continue working on shell scripts and git/github*

3 *Problem Set 1*

Read TLCL Ch 27, 29 & 33

4 *Problem Set 2 Google and Documentation*

4 *Using UF Research Computing resources*

Running batch jobs

4 *Quiz 2*

Read TLCL Ch 23

More bash scripting

5 *Quiz 3*

Read Py4E Ch 1

5 *Read Py4E Ch 2 Python data types*

5 *Problem Set 2*

Read Py4E Ch 3

6 *Problem Set 3*

Read Py4E Ch 4

6 *Read Py4E Ch 5 Python: Iteration*

6 *Quiz 3*

Read Py4E Ch 6 & Ch 7

7 *Quiz 4*

Read Py4E Ch 8, and Chs 9-10

7 *Read Ch 11 RegEx in Python*

7 *Problem Set 3*

Read Py4E Ch 12 & Ch 13

8 *Problem Set 4 SciPy, NumPy, Pandas*

8 *Pandas with Messy Data* *Data visualization with Pandas*

8 *Quiz 4*

Matplotlib and data visualization

9 *Quiz 5 More data visualization*

9 *Scan Py4E Ch 16*

Py4E Ch 14: Object oriented Programming

Project 1

9 *Problem Set 4 Py4E Ch 14: Object oriented programming*

10 *Programming Foundations Databases* *Work on Project 1*

10 *Database intro*

Flight DB Example

10 *Project 1*

Overview of databases

Database design

11 *Read Py4E Ch. 15, through 15.5 and my notes* *Problem set 5* *Py4E Ch. 15, through*
15.5 and my notes

Databases, SQL and sqlite

11 *Veteran's Day, no class*

11 *Quiz 5 More on databases and Joins*

12 *Quiz 6, due Friday, Dec. 4* *SQLAlchemy*

12 *Project 2, Due Wed, Dec 9* *SQLAlchemy and Pandas*

13 *Work on project 2*

13 *Thanksgiving, no class*

13 *Thanksgiving, no class*

14 *Work on project 2*

14 *Argparse*

14 *Problem Set 5*

Quiz 6

15 *Graphics*

15 *Project 2*