

Fall 2018 CMSE 890:Section 301

Programming Foundation for Bioinformatics

Course Description

This course is intended to be an introduction to programming in R for people who want to do bioinformatics. The first half of the course will cover the basics of programming and R syntax; the second half will put programming in the context of data science and bioinformatics research. The primary goal is to have students able to take a dataset and use R programming to ask questions about the data.

Course Prerequisites

This course is the first in the bioinformatics modules series in the CMSE department. It does not have a prerequisite other than that students have basic computer skills (e.g. web-surfing, Excel, etc). It can be used as a prerequisite for CMSE890:302 Statistical Analysis and Data Visualization and as a co-requisite for CMSE890:303 Intro to Data Handling: Unix and Python for those students who lack prior programming experience. It also can be used as a prerequisite for the CMSE course Image Processing as taught by Dirk Colbry.

Course Instructional Objectives

The following skills are developed in this course:

Analytical thinking in terms of learning individual code commands, structures, and units and putting them together to create programs that achieve a desired goal.

Effective communication in terms of not only learning how to make graphs to display data, but to create those graphs from the perspective of the audience that is trying to understand it. Also in terms

of how to communicate results from computational work to colleagues who have no background in computation.

Integrated reasoning in terms of writing and editing programs so students can ask questions of their data and update as needed to focus in on a desired goal.

Required reading materials

This class has no required book or course pack, but The Book of R by Tilman Davies is a good resource for the introduction to programming sessions. This book will also be used in CMSE890:302 Statistical Analysis and Data Visualization and so is recommended for students who take both modules. The applied programming sections will be from R for Data Science by Hadley Wickham & Garrett Golemund (O'Reilly Media, Inc). Students are encouraged to seek out these and other reference books if they are new to programming.

Other required materials

Recitations are a critical for gaining hands-on experience with the course material. You are expected to bring your charged laptop to every class session. If you do not have a laptop or it does not run the required software, contact Dr. Black. If arrangements are made in advance, loaner laptops might be made available.

Course activities

Flipped classroom-style lectures: This course is offered in the flipped classroom-style with students watching online lecture videos for homework and then coming to class to ask questions, discuss the information, and solve problems in pairs/groups. The videos are between 10 and 40 minutes each and usually there will be three videos per class session. Students are encouraged to watch the videos as many

times as needed to understand the material and to work the examples on their laptop along with the videos. In-class assignments are usually finished in class, but some students need another hour or two to finish as homework.

Pre-class homework: A short homework will be due before class that will test whether students watched the video lectures and attempted to gain a basic understanding of the material. The homework will be available as a quiz in D2L (3 questions) and needs to be completed before the start of each class session. These homework assignments will be graded as pass/fail and the lowest grade will be dropped. Answers will be given in class.

Class participation: Active class participation is critical to learning the course material. As such, you are expected to attend class every week, bring your charged computer, and to actively participate in the recitation. Every Tuesday the recitation from Monday's class must be turned in by noon on D2L to receive credit. The lowest grade will be dropped.

Quizzes: There will be a quiz every Wednesday covering the material from the week. While these quizzes are not intentionally made cumulative, understanding material from previous weeks may be essential to completing the quizzes. Quizzes will be on paper and test conceptual understanding. Students will not be asked to write code by hand, but may be asked to debug or comment on provided code. The lowest grade will be dropped. Answers will be discussed in class the following session.

Activity and Grade Percentage:

Pre-class homework (bi-weekly) (6) : 9.6% total (1.6% each)

In-class recitation (weekly) (2) : 30% total (15% each)

Quizzes (weekly) (3) : 60% total (20% each)

Survey (1): 0.4%

Grading scale:

4.0 \geq 90%

3.5 \geq 85%

3.0 \geq 80%

2.5 \geq 75%

2.0 \geq 70%

1.5 \geq 65%

1.0 \geq 60%

0.0 < 60%

Note: grades will not be curved – your grade is based on your own progress. It is hoped that you will collaborate with your classmates, not compete.

Course meeting time and location

Class will meet in A158 Plant & Soil Sciences on Mondays and Wednesdays from 3:00 – 4:50 PM.

Class attendance

This class is only eight sessions or one month long. With the volume of material presented, most students will need to attend all eight sessions to gain the necessary understanding to use the tools and understanding to eventually apply them to their research. However, it is understood that graduate student researchers may have conferences, last-minute group meetings, and disease preventing them from attending a class session. The grading policy has been designed so students can miss one class session and still complete the course. It is the responsibility of the student to get notes and materials to catch up *on their own*. If a student

must miss more than one class session, they need to contact the instructor to determine if they can still complete the one-month module. If a student needs to be absent due to an emergency (e.g. car accident, funeral, etc), email the instructor for possible quiz make-up dates. If a student is tardy, be respectful of the students who were in class on time and talk to the instructor after the in-class assignment work time has started.

Other important information (CMSE Standard Syllabus)

Course Website, Calendar, and discussion channel: This course uses a Desire 2 Learn page for course organization, which can be found at d2l.msu.edu. Accompanying course information, including the course calendar, can be found at this website. All assignments will be handed in via the **D2L Dropbox**. If you have a problem with D2L, first contact the D2L Help Desk (link on the page) as they are more likely to be able to resolve it. If you receive no answer or an unhelpful answer, contact the instructor. If you lose access to D2L such that you cannot access videos or turn in homework, contact the instructor for possible workarounds.

Classroom behavior: Respectful and responsible behavior is expected at all times, which includes not interrupting other students, turning your cell phone off, refraining from off-topic use of electronic devices, and not using offensive or demeaning language in our discussions. Flagrant or repeated violations of this expectation will result in ejection from the classroom. Using technology for purposes other than in-class assignments and note taking is prohibited until after the student has finished the in-class assignment. Avoid eating in class or bringing drinks in unsealed containers as there will not be resources for effectively cleaning up spills and crumbs.

Email: At times, we will send out important course information via email. This email is sent to your MSU email address (the one that ends in “@msu.edu”). You are responsible for all information sent out to your

University email account, and for checking this account on a regular (daily) basis.

Academic Honesty: Intellectual integrity is the foundation of the scientific enterprise. In all instances, you must do your own work and give proper credit to all sources that you use in your papers and oral presentations – any instance of submitting another person's work, ideas, or wording as your own counts as plagiarism. This includes failing to cite any direct quotations in your essays, research paper, class debate, or written presentation. The MSU College of Natural Science adheres to the policies of academic honesty as specified in the General Student Regulations 1.0, Protection of Scholarship and Grades, and in the all-University statement on Integrity of Scholarship and Grades, which are included in Spartan Life: Student Handbook and Resource Guide. Students who plagiarize will receive a 0.0 in the course. In addition, University policy requires that any cheating offense, regardless of the magnitude of the infraction or punishment decided upon by the professor, be reported immediately to the dean of the student's college.

It is important to note that plagiarism in the context of this course includes, but is not limited to, directly copying another student's solutions to in-class or homework problems; copying materials from online sources, textbooks, or other reference materials *without citing those references in your source code or documentation*, or having somebody else do your in-class work or homework on your behalf. Any work that is done in collaboration with other students should state this explicitly, and have their names as well as yours listed clearly.

Commercialization or Public Release of Course Materials: MSU policy prohibits students commercializing their notes of lectures and University-provided class materials without the written consent of the instructor. If you decide to hack the videos in some way, keep them strictly for your own use.

Emergency Procedures: If there is severe weather, a bomb threat, or other unforeseeable emergencies that might necessitate a cancellation of class, the announcement will be made on D2L.

Accommodations: If you have a university-documented learning difficulty or require other accommodations, please provide us with your VISA as soon as possible and speak with the instructor about how they can assist you in your learning. If you do not have a VISA but have been documented with a learning difficulty or other problems for which you may still require accommodation, please contact MSU's Resource Center for People with Disabilities (355-9642) to acquire current documentation.

Contact Information Course Instructor:

Alexis Black Pyrkosz

Bioinformatics Program Coordinator

Department of Computational Mathematics, Science and Engineering

Office: 2507 Engineering Building

Office Phone: (517) 432-0634

Email: ablackpz@msu.edu

Web: <https://cmse.msu.edu/directory/faculty/alexis-black-pyrkosz/>

Office hours: Monday/Wednesday 1:30 PM – 2:30 PM in Eng 2507K

Course Schedule

This schedule is tentative and subject to change. Changes will be announced on D2L.

Session 1: 2018/08/29: Syllabus, Overview, History of R

Session 2: 2018/09/05: R/R-Studio, Programming Fundamentals

Assessment: Pre-Class Homework 1, Quiz 1

Session 3: 2018/09/10: File Input/Output, Data Structures

Assessment: Pre-Class Homework 2, Recitation Activity 1

Session 4: 2018/09/12: Control Structures, Strings

Assessment: Pre-Class Homework 3, Quiz 2

Session 5: 2018/09/17: Basic R Graphing, GGPlot2

Assessment: Pre-Class Homework 4, Recitation Activity 2

Session 6: 2018/09/19: Tidyverse—Data Transformations

Assessment: Pre-Class Homework 5, Quiz 3

Session 7: 2018/09/24: Tidyverse—Tidying Data

Assessment: Pre-Class Homework 6, Recitation Activity 3

Session 8: 2018/09/26: StringR, Data Modeling, Markdown

Assessment: Pre-Class Homework 7, Quiz 4