PH 290 SEM 003: Biostatistical Computing

Course Syllabus (Fall 2022)

Course Information

Meeting Dates and Times: Monday and Wednesday 8:00 am -- 9:29 am, Aug 24 2022 -- Dec 09 2022

Location: Berkeley Way West 1205

Instructor: Xiongtao Dai, PhD

E-mail: xdai@berkeley.edu

Office hour: TBD

Classroom Canvas/bCourses: https://bcourses.berkeley.edu/courses/1517122

Course Unit Value: 3

GSI: TBD

E-mail: TBD

Course Description

This course is for those who want to extend biostatistical computing techniques for working with involved projects and big datasets.

Topics include: Rules and principles for R programming, software reliability, efficiency, and scalability, writing R packages, visualization, shell scripting and linux tools, version control, interfacing R with a low-level language, database, and parallel programming. Statistical topics may include simulation-based procedures.

As far as programming prerequisites go,

- you should have some previous experience with R and data analysis, or
- you are an advanced beginner in R who is familiar with another programming language (say, python).

Basic knowledge in statistics (e.g., linear and logistic regression models), linear algebra, and calculus will be required, though we focus on the computation aspects.

Course Learning Objectives

After successfully completing this course, you will be able to:

1. Learn a roadmap for advanced R coding concepts
2. Learn how to make your workflow reliable, fast, and scalable
3. Learn how best to incorporate existing code/functions into packages
4. Be able to apply appropriate tools and methods to work with large volume of data and computation

Methods of Instruction

Weekly lectures (3 hours/week, in person)

Tentative Course Schedule

Week	Topic	Due
1	R fundamentals	
2	Functional programming	
3	Unit test and debugging in R	
4	Scoping rules, environment, lazy evaluation	Homework 1
5	Object-oriented programming	
6	Writing an R package	Homework 2
7	Visualization	
8	Bash tools and version control	Homework 3
9	Benchmarking and profiling	
10	C++ and R's interface to it	Homework 4
11	Introduction to algorithms	
12	Database and SQL	Homework 5
13	Computing cluster and parallel programming	
14	MapReduce, Hadoop, and Spark	Homework 6
15	RRR week	
16	Project presentation	Project report

Course Requirements

We will try to return all evaluated and graded material to students within 14 days.

Homework

There will be 6 homework assignments to familiarize yourselves with the concepts and tools covered in class.

Participation

You are expected to actively participate in class, office hours, and on Ed Discussion (detailed shortly). Participation is counted towards final evaluation.

Final Project

The final project is for you to dive deep into a topic covered by this course. This will be an individual project, but you can optionally form a team of two students if you find common interest. The project topic should be closely relevant to the course materials. For example, you can choose to create a full-fledged R package from your research, analyze a big dataset, or optimize an algorithm, etc.

The assessment consists of two parts, (i) a short project presentation outlining the motivation and main results of your investigation and (ii) a project report detailing what you have explored and found. The final presentation will be given from 7-10 pm on December 12 in the finals week.

Course Grading

Grading is based on the following:

- Homework 65%
- Class discussion/participation 5%
- Project presentation 15%
- Project report 15%

Course Materials

Bcourses website

All course materials and announcements will be posted on bcourses.berkeley.edu.

Optional Textbooks

(All are freely available online, some through the Berkeley Library)

Adler, J. *R in a Nutshell: A Desktop Quick Reference*; 2nd ed.; O'Reilly Media, Inc., 2012.

Wickham, H. *Advanced R*; 2nd ed.; CRC press, 2019.

Wickham, H. *R Packages: Organize, Test, Document, and Share Your Code*; O'Reilly Media, Inc., 2015.

Eddelbuettel, D. *Seamless R and C++ Integration with Rcpp*; Springer, 2013.

Cormen, T. H.; Leiserson, C. E.; Rivest, R. L.; Stein, C. *Introduction to Algorithms*, 4th ed.; MIT Press: Cambridge, MA, USA, 2022.

Course Communication

Course announcements will be sent out through bCourses' Announcement and also in Ed Discussion. Please check regularly for updates. Emails to the GSI and the instructor are reserved for personal matters.

Getting Help

Online Help Sites

Most of quick questions (e.g., where is my R installed?) find solution online in seconds. A good way to start is to frame the question and put it into a search engine (Google, DuckDuckgo, Bing, etc). The following websites are some likely places where you may end up finding your answers:

- Stack Overflow
- R/other documentations
- Tutorial websites
- Blogs
- R mailing list

Course Assignments

You should feel free to discuss the course materials (homework, project, etc) with anyone. That said, you must write the homework and project by yourself.

Ed Discussion

We will be using Ed Discussion for class discussion, which is integrated into bCourses. The system is highly catered to getting you help fast and efficiently from classmates, the GSI, and myself. Rather than emailing questions to the GSI and the instructor, you will post your questions on Ed Discussion. Emails to the GSI and the instructor are reserved for personal matters.

When posting on Ed Discussion, please follow adequate netiquettes:

- Be polite and respectful to others.
- Search before you post. Your question may have already been asked and answered.
- When you post a question, please explain the context and give an example of what you have issue with. Posting screenshots and asking "What is going wrong?" is unacceptable.
- Posting short snippet of code is fine, but please refrain from posting a complete solution to a question.

Policies

Due Dates

Please communicate with instructors through email if you will not be able to meet course deadlines *ahead* of the deadlines. We will try to accommodate requests to extend the deadlines because of health, family, and work issues, etc. However, only requests that are made 48 hours before the deadline will be considered, unless an emergency occurs.

Late Assignments

Homework up to three days (72 hours) late will be accepted but will receive a penalty; homework more than three days late will not be accepted.

Attendance

Lecture attendance is expected. If you are not able to attend, you will be responsible for learning the materials yourself by, e.g., walking through the slides and code or learning from a classmate's notes.

Technology

Bring your laptop to the lecture for taking notes and following the instructor's demonstration. Please refrain from using electronics for non-class purposes.

SPH Course Policies

Descriptions of and relevant campus links to SPH school wide course policies on Disability Support Services, Accommodation of Religions Creed, Course Evaluations, Academic Integrity can be found at: https://berkeley.box.com/s/knh3rbk9ikgvmca4ymy93msgj9bkebq5

Dedication to Diversity, Equity, and Inclusion

Berkeley Public Health strives to create a learning environment that is inclusive of students of all backgrounds and identities. We strive in this course to highlight how our study, work, and interactions with others can be done with empathy, a striving for understanding, and used for improvement of us as people, of our institutions and the members of our communities. We acknowledge that we will make mistakes as we are all learning together. Your perspective is important to us and the knowledge you bring to help make this an enriching learning environment for all participants. We welcome input at any time and invite constructive feedback on any suggested modifications that may help improve the course now or in the future.

Students are the experts of their own experiences. Your world lens is welcomed and encouraged; and as students, you are invited to lift up information and data that is relevant to the course material. We cannot speak on behalf of all groups, or fully understand the issues, concerns, and

history of students from diverse backgrounds and identities. However, we are willing to listen and learn, admit mistakes, and engage in the ongoing work of intellectual humility. Racism, stubborn inequities, and the continuing degradation of our environment result in injustices, which are perpetuated by silence. We commit to try to turn uncomfortable conversations into teachable moments, and invite all students to do the same even though we may not all be confident or fully-skilled in doing so. Language or comments that alienate, demean, and denigrate other students in the classroom will not be ignored, but confronted.