BIOST 2094 - Advanced R Computing

Spring 2017

Time and Place: 3-4:55pm every Friday, Crabtree Hall A425

Instructor: George C. Tseng, PhD
E-mail: ctseng@pitt.edu

Instructor: Zhiguang Huo (Caleb) (Lecturer 1)
E-mail: zhh18@pitt.edu

Instructor: Tianzhou Ma (Charles) (Lecturer 2)
E-mail: tim28@pitt.edu

Instructor: Chien-wei Lin (Masaki) (Guest Lecturer)
E-mail: chl169@pitt.edu

Office Hours:
Caleb: Wednesday; 3-4pm
Charles: Monday; 3-4pm

Course Description

BIOST 2094 is an advanced statistical computing course using R designed for graduate level biostatistics students with basic R programming background. The course will cover topics, including but not limited to, R in modeling and optimization, advanced R graphics, functional programming, object-oriented field guide, efficient computing in R, GUI for R-shiny, embedding C/C++, R package/documentation, Julia, Github, etc. The course will also include real life application for students to practice the programming techniques learnt in class.

Prerequisites: BIOST 2041, 2042, Programming experience in R or other low-level languages such as C, C++, Java, Fortran, etc. Experience in SAS/Stata does not qualify.

Intended Audience

- Intermediate R programmers who want to dive deeper into R and learn new strategies for solving diverse statistical problems.
- Those who want to explore higher level features of R (Markdown document, interactive website, bridge with other language, how to efficiently maintain a R package).

**Note: This is not a basic R class so if you are not familiar with basic features of R, you may find it hard to follow later in the class. In addition, the course may cover topics including linear model, generalized linear model, survival, optimization, etc., and may require students to have learnt some basic knowledge of these topics.

Auditing Policy

Auditing is allowed in this class, but the seat priority is given to those who registered since the room capacity is limited. Homework and final project will not be graded for auditing students.

Learning Objectives

1. Write reliable, transparent and efficient programs in R
2. Produce informative graphical descriptions of data using advanced R graphical tools
3. Understand what functional programming means and why it is useful
4. Analyze data using descriptive and inferential statistics and through model fitting
5. Know which operations in R are slow and how to improve computation
6. Know how to debug, organize the codes and create R package with clear documentation and how to maintain R package.
7. Learn R markdown and reproducible research.

Textbooks:

There is no required textbook. The following books are recommended references:

- Hadley Wickham, *Advanced R*, CRC Press, 2014 (http://adv-r.had.co.nz/)
- John Chambers, *Software for Data Analysis: Programming with R*, Springer, 2008.

Websites

- R Website http://www.r-project.org
- R Manuals http://cran.r-project.org/manuals.html
- Patrick Burns, The R Inferno http://www.burns-stat.com/
- R Journal http://journal.r-project.org/
- R Search engine (Internet) http://www.rseek.org/
- R Search engine (R documents) http://search.r-project.org/nmz.html
- R-bloggers http://www.r-bloggers.com
- Quick-R http://www.statmethods.net
- Google R Style Guide https://google.github.io/styleguide/Rguide.xml
- Advanced R book website http://adv-r.had.co.nz/

Software:

R, Rstudio (free online)

Student Evaluation and Grades

Course grades will be based on a weighted average of,

- Homework assignments 40%
- Project proposal (due Feb 17th, 2017) 10%
- Project presentation (last two classes) 20%
- Project report and package (due last class, Apr 7th, 2017) 30%

The cut-offs for computing letter grades will be: A, 100%-90%; B, 89%-80%; C, 79%-70%; D, 69%-60%; and F, <60%. Plus-minus grades will be assigned by dividing the respective intervals into thirds. Discussions of homework and project among students are allowed but each student has to write their own solution. Cheating and plagiarism is strictly not allowed and may be reported to the university. See the University of Pittsburgh's Policy on Academic Integrity at http://www.provost.pitt.edu/info/ai1.html

Grading Criteria: Homework will be graded according to the correctness of the answer. Grading of project presentation and report will depend on teamwork, contribution of each team member, content and quality of the project, coherence and organization of the presentation.

Homework

There will be 4 homework assignments in total. Homework needs to be done using R markdown and only the compiled html file is needed for submission. The detailed instructions will be provided in the first class. Students will submit an electronic copy via Courseweb (check the deadline, please!). **No hard copy is required**. *Please include clear comments to make code readable.*

Late Assignment policy:
Full credit will be given for assignments turned in on the due date (by 11:59pm).
80% credit for one day late.
Assignments turned in the next school day after the due date will have a maximum possible credit of 80%.
50% credit for two days late.
Assignment turned in two school days after the due date will have a maximum credit of 50%.
NO credit given after two days late.
If you are out sick, no deduction will be taken if you inform me before the homework is due that you are ill. Please stay home and do not get other people sick. Just turn in your homework to Courseweb as soon as you can.
If you are going to miss school on the day the homework is due (going out of town, religious holiday, etc.) please turn your homework in early.

Project

The goal of the project is to develop an R package or application that will be useful to other statisticians and R users. Students are encouraged to form groups (3-4 students) themselves. Otherwise, instructors will assign groups for them. Students can also work individually, which may add their workload.
Groups are encouraged to find project topic themselves. Otherwise, they can select from a list of suggested topics provided by the instructor.
For the project, we strongly recommend groups use what they have learnt from class, including but not limited to, defining classes and write a set of methods to accompany the functions, incorporate C++ to improve computing, R GUI, etc. The R documentation is required for the final R package. Each group will give a class presentation on their package in the last two classes.

By the end of the 3rd class (Jan 20th, 2017), students need to inform the instructors of their own group assignment or they need a group assignment by the instructors.
By the end of the 7th class (Feb 17th, 2017), students need to turn in a proposal for their project.
Proposals should at the minimum include the following information:

- Project title and list of group members
- Overview of project idea
- Description of the dataset planned to include in the project
- Supporting literature if available

Instructors will evaluate the appropriateness of the proposal and give comments/feedback to students.

By the last class (April 7th, 2017), students need to turn in a final report and the associated well-documented package for their project.

Lecture Schedule

Date	Due	Topic
January 6, 2017		Lecture 1: Basics in R (Lecturer 1)
January 13, 2017		Lecture 2: R in modeling and optimization (Lecturer 2)
January 20, 2017	Homework 1 Distributed	Lecture 3: R graphics (Guest Lecturer)
January 27, 2017		Lecture 4: Functional programming (Lecturer 1)
February 3, 2017	Homework 1 Due, Homework 2 Distributed	Lecture 5: R coding style, R for simulation (Lecturer 2)
February 10, 2017		Lecture 6: Debugging and exception handling (Lecturer 2)
February 17, 2017	Project proposal Due, Homework 2 Due, Homework 3 Distributed	Lecture 7: Object-Oriented Programming (Lecturer 2)

February 24, 2017		Lecture 8: Speed up R computation (Lecturer 1)
March 3, 2017	Homework 3 Due	Lecture 9: R GUI (Guest Lecturer)
March 10, 2017	Spring Break	No Class
March 17, 2017	Homework 4 Distributed	Lecture 10: R documentation and package, Reproducible research (Lecturer 1)
March 24, 2017		Lecture 11: Introduction to Julia (Lecturer 1)
March 31, 2017	Homework 4 Due	Lecture 12: Student presentation
April 7, 2017	Project report/package Due	Lecture 13: Student presentation

Required Equipment

Courseweb

Academic Integrity

Students in this course will be expected to comply with the University of Pittsburgh's Policy on Academic Integrity (http://www.provost.pitt.edu/info/ai1.html). Any student suspected of violating this obligation for any reason during the semester will be required to participate in the procedural process, initiated at the instructor level, as outlined in the University Guidelines on Academic Integrity. This may include, but is not limited to, the confiscation of the examination of any individual suspected of violating University Policy. Furthermore, no student may bring any unauthorized materials to an exam, including dictionaries and programmable calculators.

Disability Services

If you have a disability that requires special testing accommodations or other classroom modifications, you need to notify both the instructor and Disability Resources and Services (http://www.studenta_airs.pitt.edu/drswelcome) no later than the second week of the term. You may be asked to provide documentation of your disability to determine the appropriateness of accommodations. To notify Disability Resources and Services, call (412) 648-7890 (Voice or TTD) to schedule an appointment. The Disability Resources and Services office is located in 140 William Pitt Union on the Oakland campus.