Computational Skills for Biostatistics I

This course is about programming and other computing skills for biostatistics. It is intended as an optional "add-on" course to Biost 514 (https://canvas.uw.edu/courses/1063494) - in which students see basic R commands, and many examples using those commands. Biost 561 introduces more sophisticated use of R, and other programming techniques and tools.

See links (left) for submitting homework assignments and the course's Canvas "syllabus" - really a schedule of events. For everything else, see the pages on;

- About 561 (https://canvas.uw.edu/courses/1063499/pages/about-561) - course content, grading, etc
- Lecture and Homework Material (https://canvas.uw.edu/courses/1063499/pages/Lectures%20and%20Homework?titleize=0)
- Other Resources (https://canvas.uw.edu/courses/1063499/pages/other-resources)

Logistic details

- Instructor: Ken Rice (http://faculty.washington.edu/kenrice/)
- Questions? Discussion with other students is encouraged, but you should code and write up your work on your own. Please contact the instructor either in person or at office hours (10.30-Noon, Tuesday) or via email (kenrice@uw.edu). Bring a laptop if you want to show problems "live".
- Lectures: Th, 2:30 - 3:20pm. You should attend - and ask questions
- Classroom: Health Sciences Building, Room I132
- Homework: There will be a short exercise following each class, due the next week in class. To obtain credit for the class, students must make a good faith effort to answer every part of each exercise. Solutions will be read and corrected but not graded. Students who will miss the deadline for any exercise should contact the instructor, who will make other arrangements.
- There are no exams for this course
About 561

Course Objectives

This course provides an introduction to statistical computing. Emphasis is placed on good programming techniques useful in statistical analysis. At the end of the course, students are expected to be familiar with:

1. Object-oriented programming practice in R
2. Use of R on departmental clusters
3. Tools for good practice in data analysis (e.g. ggplot, markdown) and coding (e.g. Github)
4. How programming in R compares to other languages (e.g. C, Python)

Prerequisites: Any graduate student in Biostatistics/Statistics can register for the course, but some familiarity with statistical analysis and very introductory R will be useful. Contact the instructor for permission to take the course if you are not a graduate student in Biostat/Stat and if you are not certain whether this course is right for you.

Computing Software: R will be the primary language; other free languages and tools will be introduced as needed

Text: There is no course text, but several texts are mentioned in Class 1. Other reading will be made available on the Other Resources page.

Notes: Slides for each lecture will be posted on the class site before class, so students can bring a copy, in some format. Hard copies will not be provided by the instructor.

Assessment: Short exercises are assigned weekly, after each class, and are due by the following class. Students must submit a good-faith effort at all questions in all exercises in order to receive credit for the course. Students who miss a deadline should contact the instructor without delay: deadlines may be arranged.

Grading: CR/NC

Other Important Notes:

1. Please check the webpage regularly for new/revised material
2. The instructor reserves the right to modify the course plan and the syllabus as conditions require
3. Questions and discussions are welcome, and encouraged throughout the class; keep in mind that if there is something that is not clear to you, it most likely is unclear to others as well.

Tentative Course Schedule:

1. 10/6 Introductions. Debugging, the object systems, generic functions
2. 10/13 Writing functions, high level loops
3. 10/20 ggplot
4. 10/27 LaTeX - for later Markdown material
5. 11/3 Markdown
6. 11/10 Unix system, shell scripts, cluster computing
7. 11/17 C and C from R
8. 11/24 NO CLASS - Thanksgiving
9. 12/1 Github, making packages
10. 12/8 Intro to Python

There will be several guest lectures, including several by students, as they have the most recent experience using tools relevant for thesis/dissertation research.

History of this course: For several years prior to 2016, Biost 514/515 and Biost 511/12/13 were taught primarily in Stata. During this time, Biost students were primarily taught R commands (with some other programming skills and topics relevant to research) in Biost 561/562/563. Beginning in 2016, Biost 514/515 and 511/12/13 are now taught primarily in R, thus eliminating the need to teach R extensively in Biost 561/562. The more sophisticated programming methods students now see in 561/562 have been chosen because of their broad relevance to many statistical researchers.
The class is provides a more direct link between 1st year Biostatistics students and faculty (and other instructors) than is possible in large classes like 514/515.