University of Pennsylvania
Division of Biostatistics
Subject Guide

BSTA 670: Programming and Computation for Biomedical Data Science

Credit points: **1.0**
Semester: **Spring 2019**
Time: T/Th 9:00-10:20am
Location: 418 Blockley

Course Instructor:
Kristin A. Linn
Assistant Professor of Biostatistics
Email: klinn@pennmedicine.upenn.edu
Office: 222 Blockley Hall
Phone: 215-746-3519

Pre-requisites: BSTA 621, BSTA 651; or permission of instructor.

Subject Aims: The course will cover programming and computational fundamentals in Python and R. It will concentrate on computational tools that are useful for statistical research and computationally intensive analyses. The goal is for students to develop a knowledge base and skill set that includes a wide range of modern computational tools needed for statistical research and data science. Topics may include, but are not limited to:

1. Reproducible research and programming
2. Algorithms
3. Computer storage and arithmetic
4. Numerical Integration
5. Optimization
6. Simulation

The focus will be on practical applications and implementation.

Course Materials: All course materials will be available on Canvas. Canvas is assessable from the Penn library: https://canvas.upenn.edu

Software: A combination of R and Python will be used.

Textbook: None required.

Breaks: There will be no classes the week of March 4-8 (Spring Break).
Reading Days: Reading days will be held from May 2-5.

Assessment: Homework: 50% (5 @ 10% each)
Midterm Exam: 10% (tentatively scheduled for February 26)
Advanced Topic Report and Presentation: 15% (due April 25)
Final project: 25%

Final Project: Project proposals must be approved prior to starting work. Projects will be presented as a project report. Although the data used may be part of a previous paper or research project, the work presented must not be part of a prior research project. The report should be written using RMarkdown, and all files needed to reproduce the report must be submitted. Detailed information about the final project will be available in February on the course canvas website.

Proposal due: February 28 by 5:00pm
Report due: May 10 by 5:00pm

Classes and Topics: Specific topics may include: shell basics, Git, Python basics, Jupyter notebooks, functions, namespaces, classes, data structures, algorithms, R basics, R packages, testing, debugging, benchmarking, profiling, parallelization, simulations, Rcpp, computer arithmetic, condition, stability, numerical and Monte Carlo integration, solving linear systems, optimization, web scraping, advanced special topics.

Useful resources: *Python documentation*: https://docs.python.org/3/

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to algorithms*. MIT press.

Wickham, H (2015). *Advanced R*. CRC Press.

Matloff, N (2011). *The Art of R Programming*. No Starch Press.

Monahan, J (2011). *Numerical Methods of Statistics* (second edition). Cambridge University Press.

Givens, G.H., & Hoeting, J.A. (2013) *Computational Statistics*. Second edition. Wiley.

Cheney, W, & Kincaid D. (2008) *Numerical Mathematics and Computing*. Sixth edition. Thomson.