BSTA 670: Programming and Computation for Biomedical Data Science

Credit points:	**1.0**
Semester:	**Spring 2020**
Time:	T/Th 9:00-10:20am
Location:	701 Blockley

Course Instructor: Kristin A. Linn
Assistant Professor of Biostatistics
Email: klinn@pennmedicine.upenn.edu
Office: 220 Blockley Hall
Phone: 215-746-3519
Office hours: by appointment

TA Justin Lakkis
Email: jlakkis@pennmedicine.upenn.edu
Office hours: time and location TBD

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. Simulation
5. Optimization
6. Numerical Integration

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 9-13 (Spring Break) and no class on March 24 (ENAR).
	Class will be cancelled on April 28 to give students more time to work on the final simulation project. The TA will hold office hours that day (time and location TBD) for project-related questions and help.
Rescheduled classes:	The April 21 and 23 lectures will be **moved to Friday, April 3 and 10, 2020 from 12:00-1:30pm**. Thus, there will be **no classes on April 21 and 23**. Note also the change in location: **class on April 3 will be held in 1311 Blockley**. April 10 will be in 701 Blockley.
	These two classes will be student presentation days, and **attendance is required both days** in order to get full credit for your presentation grade. If you have a pre-existing conflict with either date, please email or speak with me about it by January 30, 2020. As a thank you for being flexible with these schedule changes, I will provide lunch for the class on both April 3 and 10. **If you have any restrictions or allergies, please let me know ASAP.**
Reading Days:	Reading days will be held from April 30 - May 3.
Assessment:	Homework: 40% (4 @ 10% each) Midterm Exam: 15% (tentatively scheduled for February 26) Advanced Topic Report and Presentation: 15% (due April 25) Final project: 30%
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.
	<u>Proposal due:</u> February 27, 2020 by 5:00pm <u>Report due:</u> May 8, 2020 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, debugging, benchmarking, profiling,

parallelization, simulations, Rcpp, computer arithmetic, condition, stability, numerical and Monte Carlo integration, solving linear systems, optimization, advanced special topics.

Useful resources: *Git documentation and book by Chacon and Straub:* https://git-scm.com/book/en/v2

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.