Specialized Discipline Course, Quantitative Biology, 2019

Instructors:	Justin Kinney (lead), jkinney@cshl.edu
	Alex Dobin
	Adam Siepel
	David McCandlish
	Hannah Meyer
	Molly Gale Hammell
	Tatiana Engel
Bootcamp assistants:	Ben Harris
	Shaina Lu
Course TAs:	Wei-Chia Chen, chenw@cshl.edu
	Hussein Hijazi, hijazi@cshl.edu

Quantitative reasoning is a powerful tool for uncovering and characterizing biological principles, ranging from the molecular scale all the way to the ecological. With the advent of high throughput technologies in genomics and neuroscience it has become increasingly necessary for biological researchers to be able to analyze and interpret large data sets and frame biological hypotheses quantitatively. To this end, this course will aim to equip the students with a working knowledge of standard statistics and Python programming, as well as provide exposure to more advanced topics in machine learning, genomics, population genetics, neuroscience, and biophysics.

Homework: Twelve problem sets will be assigned. Each will be due on the Friday after the lectures that cover the relevant material (see schedule). Optional problem sessions will be held by the TAs (typically on Tuesdays from 4pm - 6pm or 5pm - 7pm in Samet) each week that a problem set is due. Unless otherwise stated, problem sets should be completed as Jupyter notebooks and emailed to Hussein Hijazi by 11:59pm on the due date. An assignment submitted after the specified date will be accepted with a late penalty of 30% off of the final grade. Assignments will not be accepted after one week past the deadline date. In case of a documented crisis, such as illness or a family emergency, the student should submit an official document to arrange for alternate grading. Advance notification is required for late submission unless this is impossible.

Student Evaluation: Problem sets: 85%, Lecture participation: 15%

QB Bootcamp:

GitHub Repository: https://github.com/jbkinney/19_qbbootcamp

Day 1: Wednesday, August 28, Plimpton (Beckman), 10am - 5pm
- 10:00am - 10:30am: **Overview of Quantitative Biology (Justin)**
- 10:30am - 12:00pm: **The Unix command line (Justin)**
- 12:00pm - 1:00pm: *Lunch (provided)*
- 1:00pm - 1:30pm: **Introduction to Python and Jupyter Notebooks (Justin)**
- 1:30pm - 3:00pm: **Python: data types (Ben)**
- 3:00pm - 3:30pm: *Break*
- 3:30pm - 5:00pm: **Python: flow control (Ben)**

Day 2: Thursday, August 29, Plimpton (Beckman), 10am - 5pm
- 10:00am - 10:30am: **Overview of High-Performance Computing (Justin)**
- 10:30am - 12:00pm: **Read mapping using BlackNBlue (Justin)**
- 12:00pm - 1:00pm: *Lunch (provided)*
- 1:00pm - 1:30pm: **Introduction to Pandas (Justin)**
- 1:30pm - 3:00pm: **Pandas I, TF analysis (Hussein)**

3:00pm - 3:30pm: *Break*
3:30pm - 5:00pm: **Pandas II, Replication origin analysis (Hussein)**

Day 3: Friday, August 30, Plimpton (Beckman), 2pm - 6pm
2:00pm - 2:30pm: Introduction to Data Visualization (**Justin**)
2:30pm - 4:00pm: Matplotlib (**Shaina**)
4:00pm - 4:30pm: *Break*
4:30pm - 6:00pm: Seaborn (**Shaina**)

QB Course:

1. **Statistics: Introduction to probability and statistics (Justin)**
   Tuesday, September 3, Wendt (Wendt), 2pm-4pm

2. **Statistics: Probability distributions and their origins (Justin)**
   Wednesday, September 4, Samet (Koch), 2pm-4pm

   Homework 1 (Bootcamp) *due Friday, September 6*

3. **Statistics: Confidence intervals and null hypotheses (Justin)**
   Wednesday, September 11, Samet (Koch), 2pm-4pm

4. **Statistics: Common statistical tests (Justin)**
   Thursday, September 12, Samet (Koch), 2pm-4pm

   Homework 2 (Lectures 1 & 2) *due Friday, September 13*

5. **Statistics: Perils of multiple hypotheses (Justin)**
   Wednesday, September 18, Samet (Koch), 2pm-4pm

6. **Statistics: Designing powerful experiments (Justin)**
   Friday, September 20, Samet (Koch), 2pm-4pm

   Homework 3 (Lectures 3 & 4) *due Friday, September 20*

7. **Statistics: Curve fitting (Justin)**
   Wednesday, October 2, Samet (Koch), 10am-12pm

8. **Statistics: Survival analysis (Justin)**
   Thursday, October 3, Plimpton (Beckman), 10am-12pm

   Homework 4 (Lectures 5 & 6) *due Friday, October 4*

9. **Machine learning: Unsupervised methods (Alex)**
   Wednesday, October 9, Samet (Koch), 2pm-4pm

10. **Machine learning: Supervised methods (Alex)**
    Thursday, October 10, Samet (Koch), 2pm-4pm

    Homework 5 (Lectures 7 & 8) *due Friday, October 11*

11. **Algorithms: Introduction (Adam)**
    Wednesday, October 23, Samet (Koch, 2pm-4pm)

12. **Algorithms: Sequence alignment (Adam)**
    Thursday, October 24, Samet (Koch, 2pm-4pm)

    Homework 6 (Lectures 9 & 10) *due Friday, October 25*

13. **Algorithms: Hidden Markov models (Adam)**
    Thursday, October 31, Samet (Koch), 2pm-4pm

14. **Algorithms: Phylogenetics (Adam)**
    Friday, November 1, Samet (Koch), 2pm-4pm

    Homework 7 (Lectures 11 & 12) *due Friday, November 1*

15. **Popgen: Evolution (David)**
    Tuesday, November 5, Wendt (Wendt), 2pm-4pm

16. **Popgen: Genomics (David)**
    Thursday, November 7, Samet (Koch), 2pm-4pm

    Homework 8 (Lectures 13 & 14) *due Friday, November 8*

17. **Genomics: GWAS (Hannah)**
    Tuesday, November 12, Wendt (Wendt), 2pm-4pm

18. **Genomics: ChIP-Seq (Hannah)**
    Thursday, November 14, Samet (Koch), 2pm-4pm

    Homework 9 (Lectures 15 & 16) *due Friday, November 15*

19. **Genomics: RNA-Seq (Molly)**
    Thursday, November 21, Samet (Koch), 2pm-4pm

20. **Genomics: Single cell analysis (Molly)**
    Friday, November 22, Samet (Koch), 2pm-4pm

    Homework 10 (Lectures 17 & 18) *due Friday, November 22*

21. **Neuroscience I (Tatiana)**
    Monday, November 25, Samet (Koch), 2pm-4pm

22. **Neuroscience II (Tatiana)**
    Tuesday, November 26, Wendt (Wendt), 2pm-4pm

    Homework 11 (Lectures 19 & 20) *due Friday, November 29*

23. **Image analysis (Justin)**
    Monday, December 2, Samet (Koch), 2pm-4pm

24. Molecular biophysics (Justin)
Tuesday, December 3, Wendt (Wendt), 2pm-4pm

Homework 12 (Lectures 21 & 22) *due Friday, December 6*