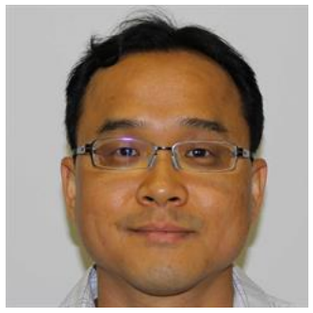
Welcome to Quantitative Biology

QB Bootcamp, Day 1
Wednesday, 28 August 2019
10:00am - 10:30am

QB Bootcamp team

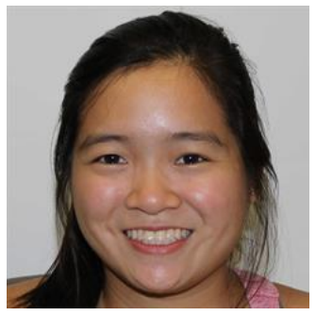
Justin B. Kinney
(Assistant Professor)



Wei-Chia Chen
Postdoc, Kinney Lab

Ben Harris
WSBS, Gillis Lab

Hussein Hijazi
Postdoc, Siepel Lab



Shaina Lu
WSBS, Gillis /
Zador Labs

2019 QB Bootcamp Schedule

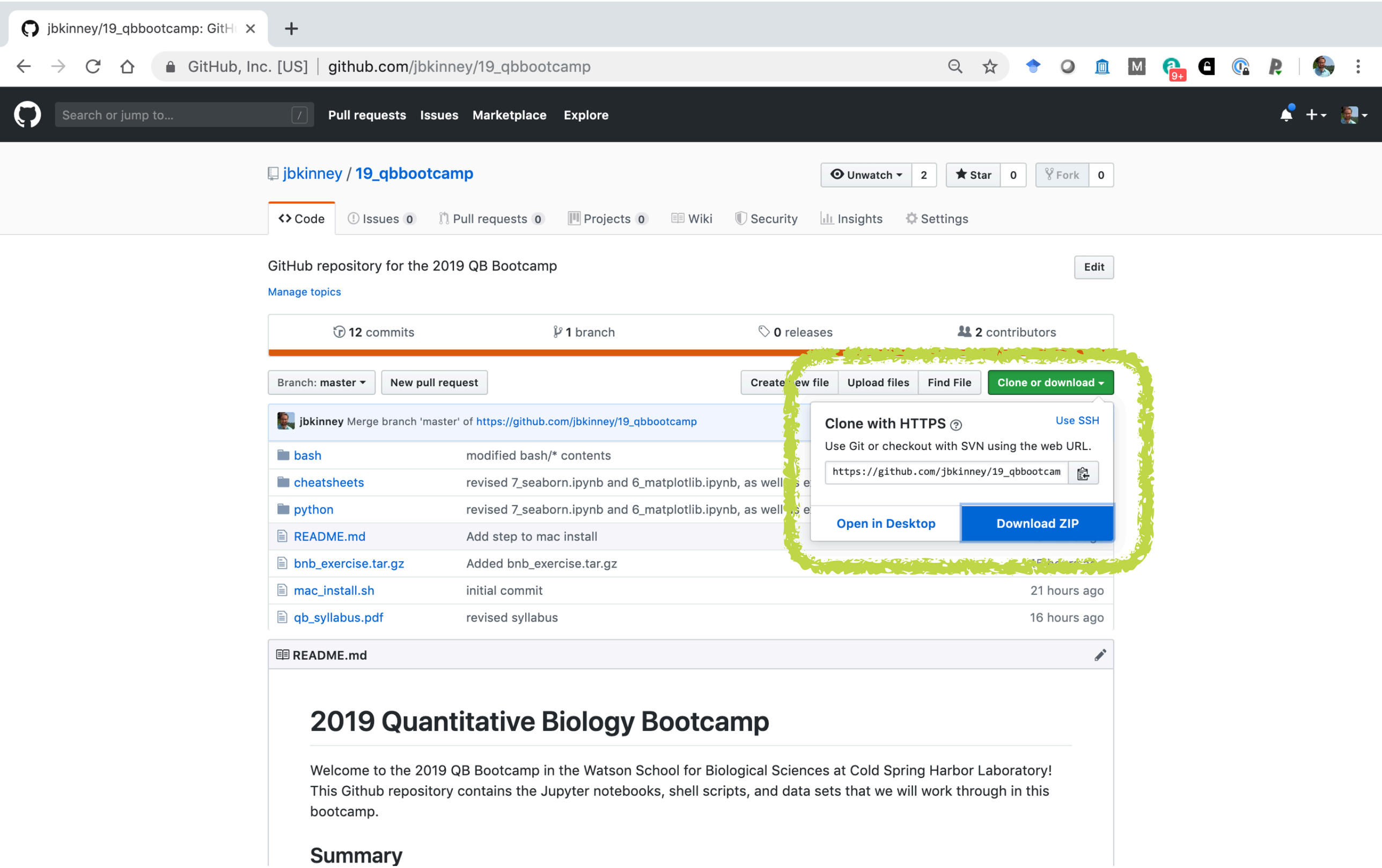
- **Wednesday, August 28, 10a-5p**
  - 10:00a - 10:45a: Welcome to quantitative biology (**Justin**)
  - 10:45a - 12:00p: The UNIX command line (**Justin**)
  - *12:00p - 1:00p: Lunch*
  - 1:00p - 1:30p: Introduction to Python and Jupyter Notebooks (**Justin**)
  - 1:30p - 3:00p: Python: data types (**Ben**)
  - *3:00p - 3:30p: Break*
  - 3:30p - 5:00p: Python: flow control (**Ben**)

- **Thursday, August 29, 10a-5p**
  - 10:00a - 10:30a: Overview of high-performance computing (**Justin**)
  - 10:30a - 12:00p: BlackNBlue, read mapping (**Justin**)
  - *12:00p - 1:00p: Lunch*
  - 1:00p - 1:30p: Introduction to dataframes (**Justin**)
  - 1:30p - 3:00p: Pandas I, binding site analysis (**Hussein**)
  - *3:00p - 3:30p: Break*
  - 3:30p - 5:00p: Pandas II, replication origin analysis (**Hussein**)

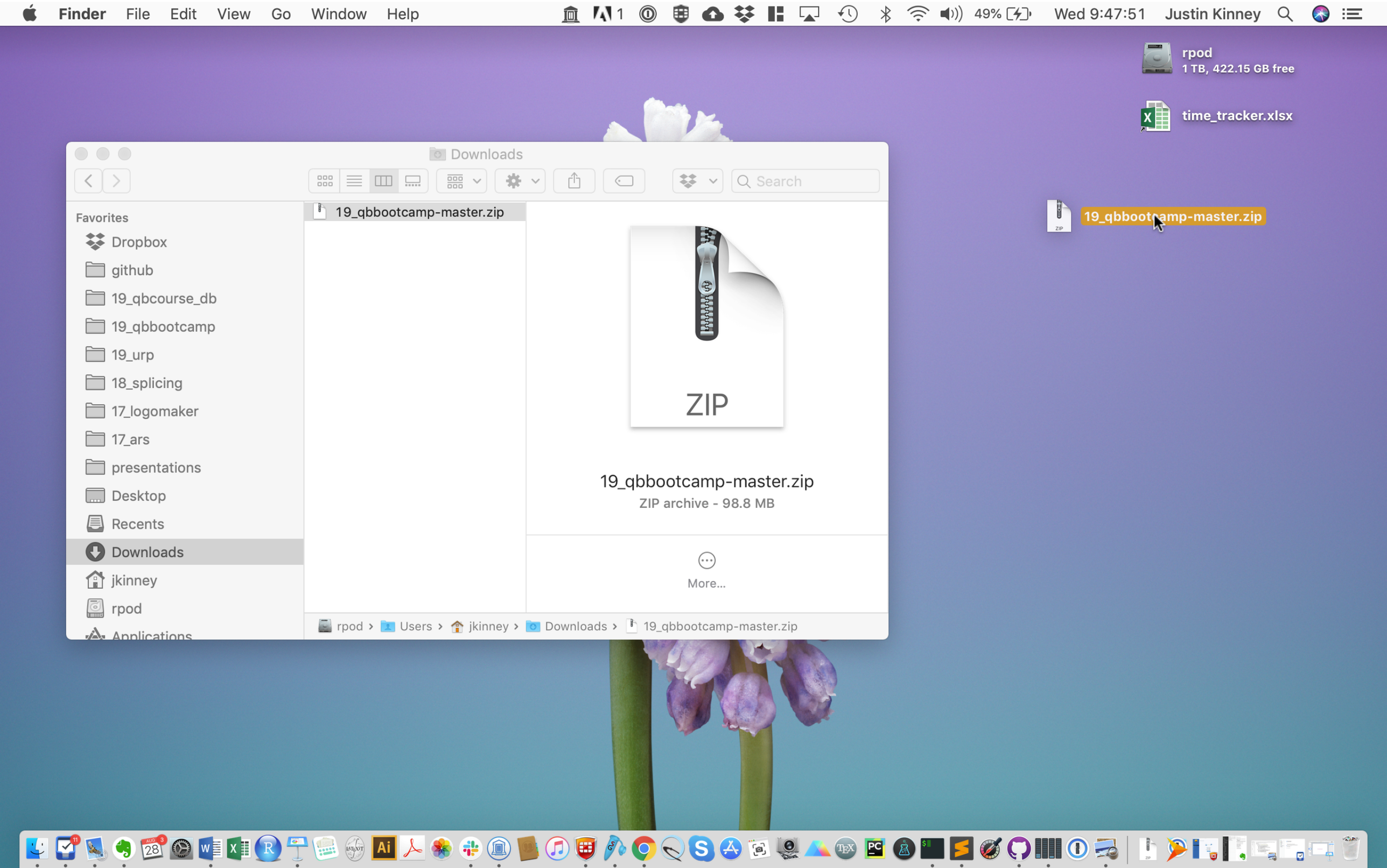
- **Friday, August 30, 2p-6p**
  - 2:00p - 2:30p: Overview of data visualization (**Justin**)
  - 2:30p - 4:00p: Matplotlib (**Shaina**)
  - 4:00p - 4:30p: *Break*
  - 4:30p - 6:00p: Seaborn (**Shaina**)

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

Download this repository



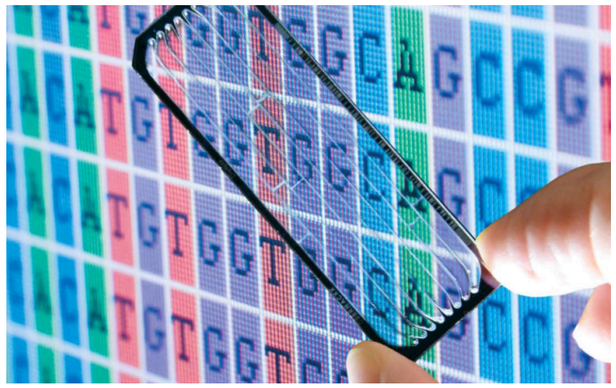
Move this repository from Downloads to your Desktop and decompress by double-clicking



What is Quantitative Biology?

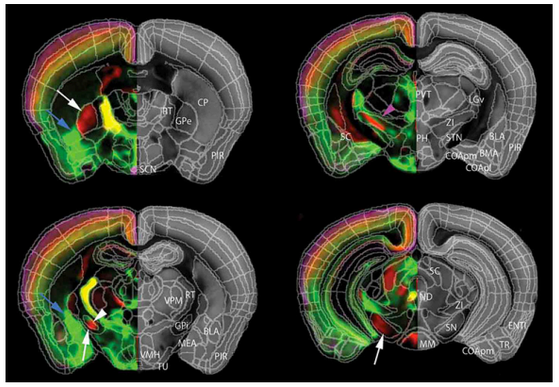
Quantitative biology is a vast field

Genomics



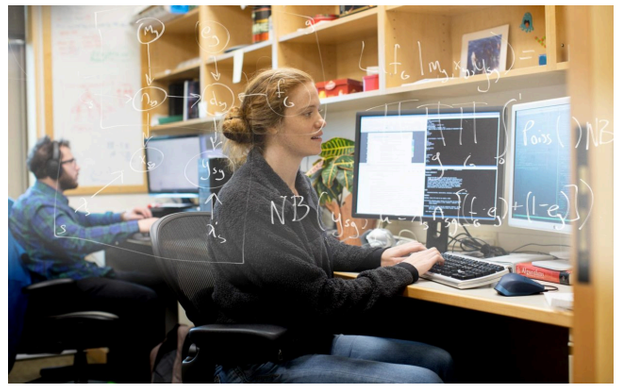
Functional genomics
Evolutionary genomics
Genome dynamics
Technology development

Neuroscience



Data analysis
Modeling neural systems
Behavioral modeling

Other



Biophysics
Machine learning
Software development

Who does Quantitative Biology at CSHL?

Core QB program

Molly Gale Hammell

Dan Levy

Saket Navlakah

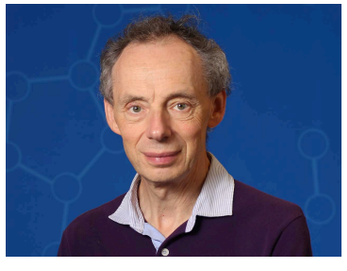
Ivan Iossofov

David McCandlish

Justin Kinney

Hannah Meyer

Peter Koo



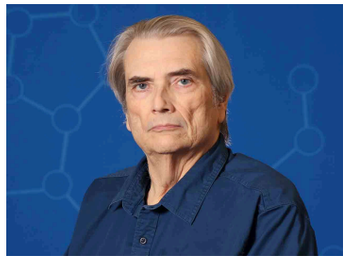
Alexander Krasnitz

Adam Siepel

QB Associated Faculty

Genomics

Alexander Dobin



Richard McCombie

Jesse Gillis

Doreen Ware

Neuroscience

Tatiana Engel

Alexei Koulakov

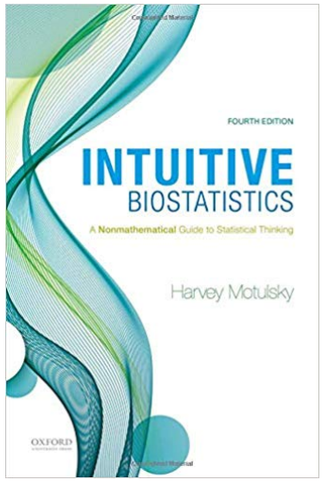
Partha Mitra

What QB skills should all biology researchers have?

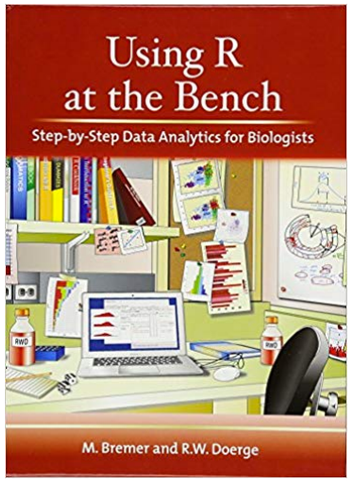
Learn to interpret standard statistics

Key statistical concepts:

- P-values
- Multiple hypothesis testing
- Confidence intervals
- Regression
- ANOVA
- Survival analysis



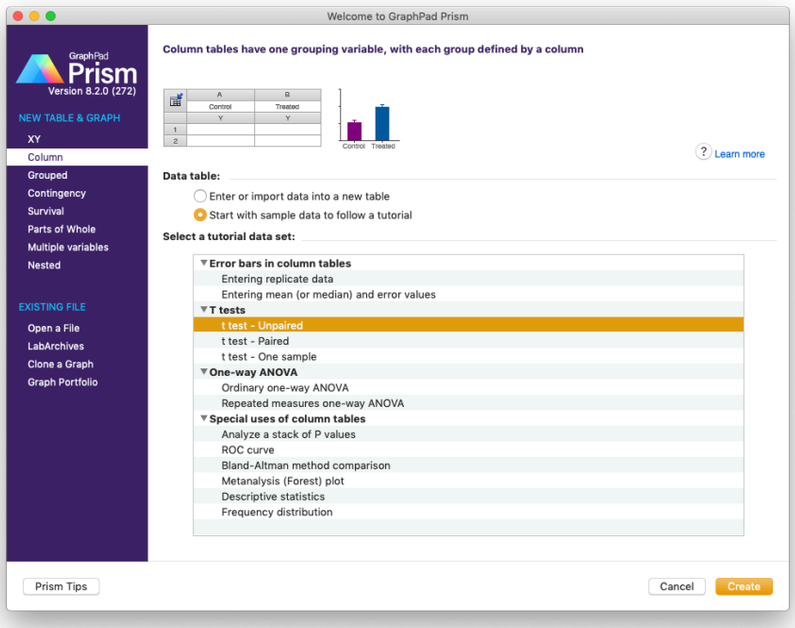
Intuitive Biostatistics, 4th ed
Motulsky (2018)
[CEO, GraphPad Software]

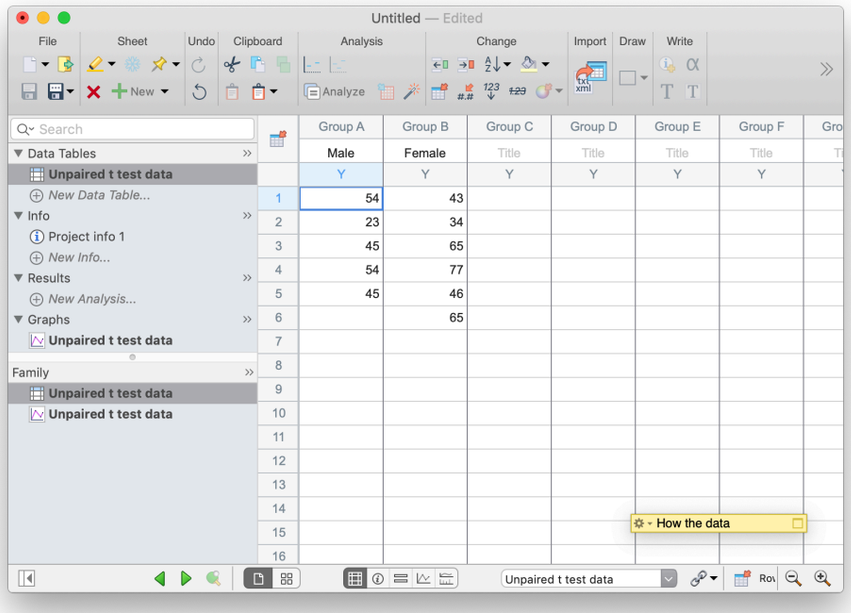


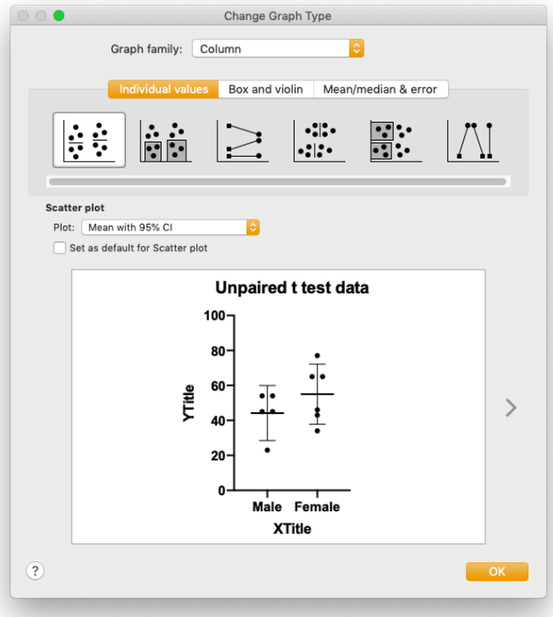
Using R at the Bench
Bremmer & Doerge (2015)

Learn to compute standard statistics

Prism 8







Alternatively:

python™

R

MATLAB

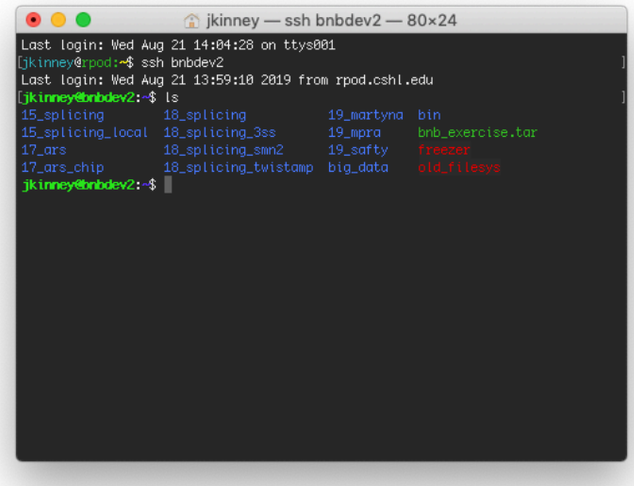
Excel

Learn to navigate UNIX systems

Sequencer

Microscope

High Performance
Computer Cluster

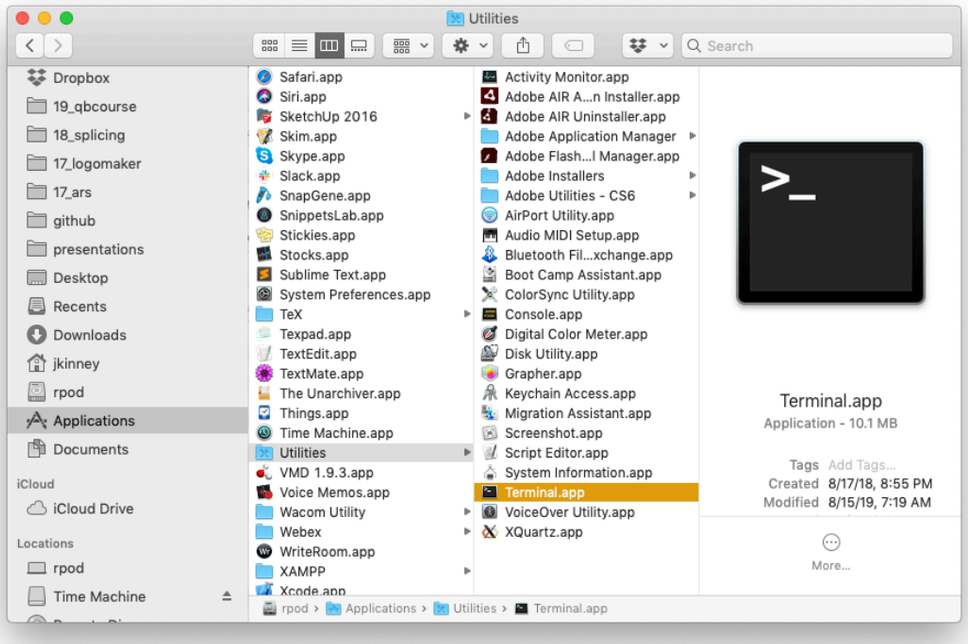


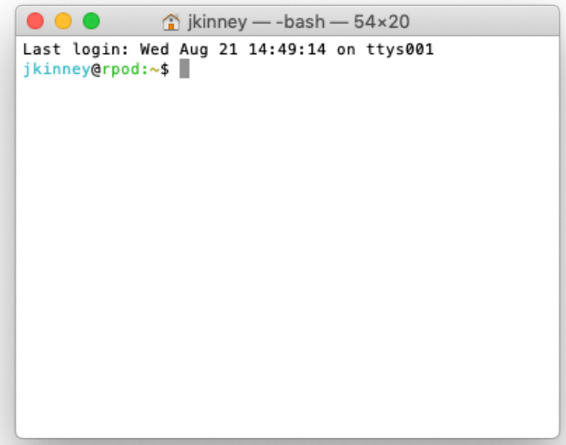
UNIX command line

Databases

Mac OS X is based on UNIX

Finder



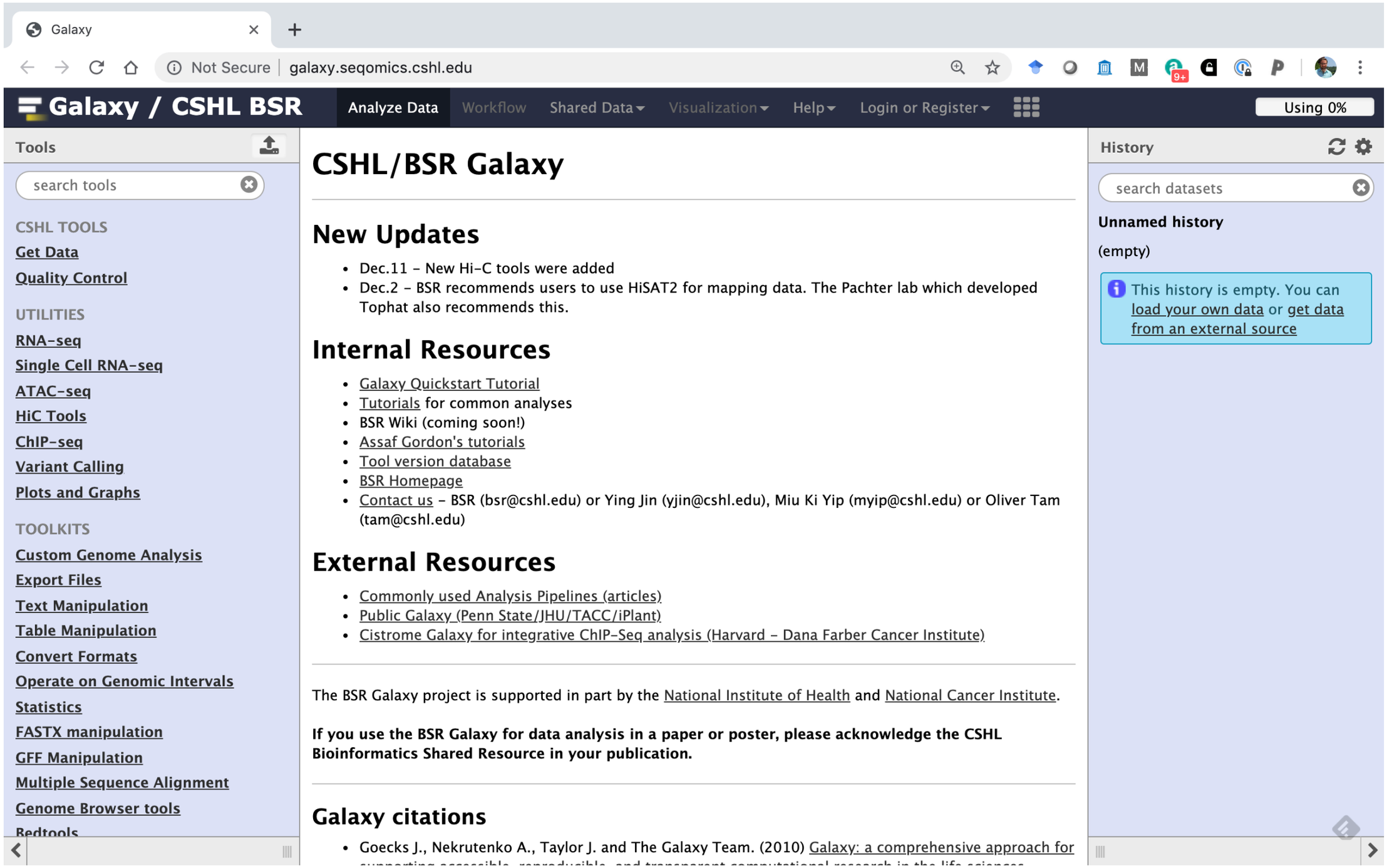


Applications/Utilities/Terminal.app

Become familiar with at least one programming language

language	strengths	weaknesses
	- elegant language - easy to learn - flexibility: use for large pipelines or local data analysis - highly valued skill in industry	- clunky dataframes - clunky statistics - clunky graphics
	- streamlined for statistics - highly developed for genomics - great graphics	- strange language - not great for building pipelines
	- used heavily in neuroscience and by old people	- proprietary - poorly supported - bad graphics - bad for strings

Learn to analyze your own sequencing data



Don’t be shy about asking QB labs to help you learn.

What skills do you need to do research in Quantitative Biology?

Learn to program well

Tip: it is better to know one language well than many languages superficially.

How to learn to program

1. Ask for guidance
2. Work on projects that require it
3. Google your questions & read help threads
4. Read package documentation
5. Read select books
6. Take online courses (don't worry about cost)

PCWorld
FROM IDG

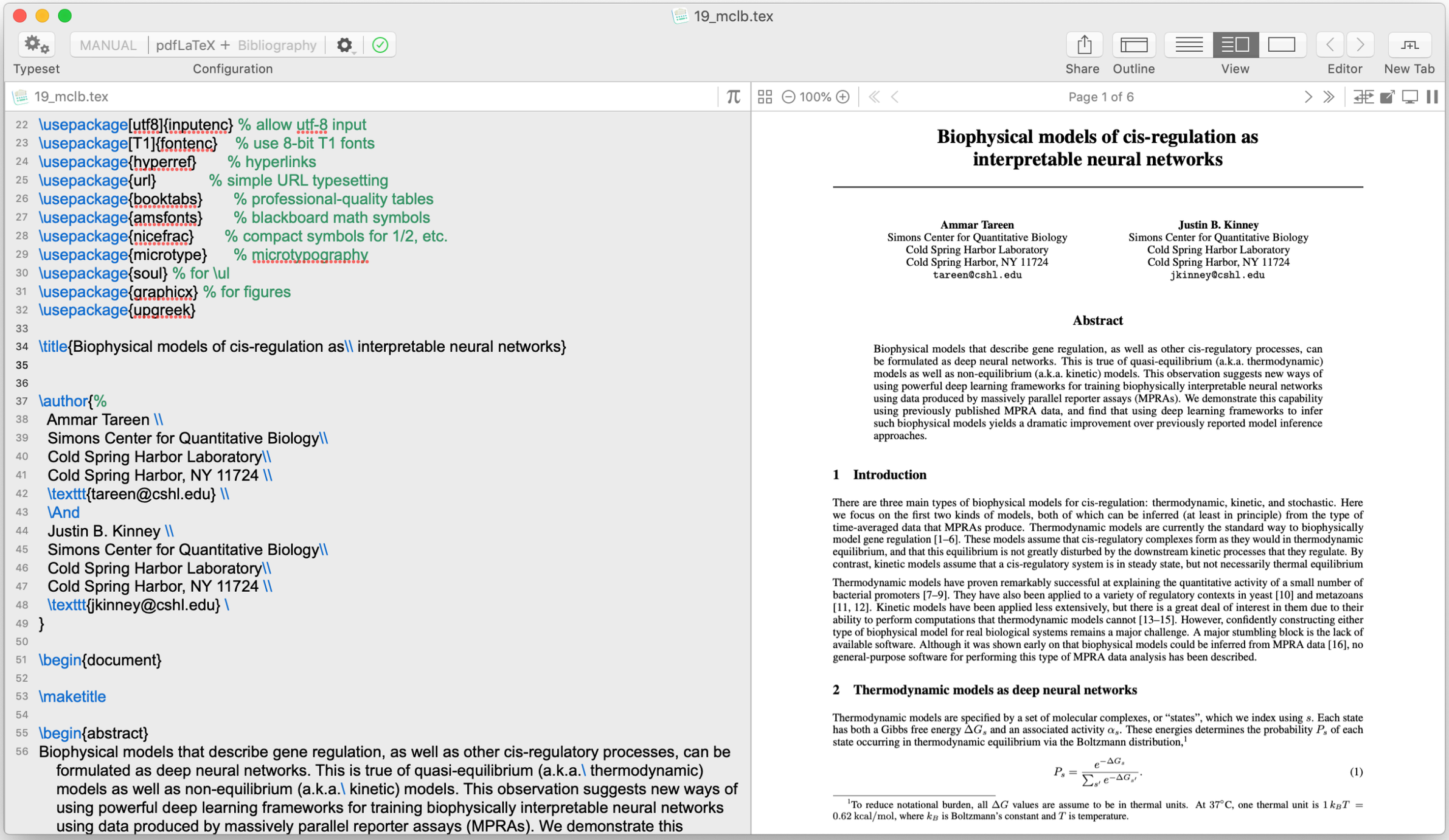
BEST ONLINE COURSES FOR PYTHON AT A GLANCE

Our picks for the best subscription / fee-based Python courses and tutorials

- Python For Everybody *[coursera.com]*
- Learning Python with PyCharm *[lynda.com]*
- DataCamp *[datacamp.com]*
- Introduction to Python: Absolute Beginner *[edx.com]*
- Introduction to Computer Science and Programming Using Python *[edx.com]*
- Python and Django Full Stack Web Developer Bootcamp *[udemy.com]*
- AI Programming with Python *[udacity.com]*
- Introduction to Computing in Python *[edx.com]*
- Python I: Essentials *[quickstart.com]*

https://www.pcworld.com/article/3287981/best-python-courses.html

Learn to use LaTeX

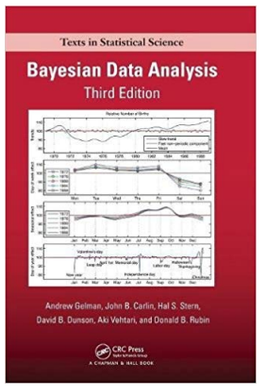


Develop core quantitative knowledge

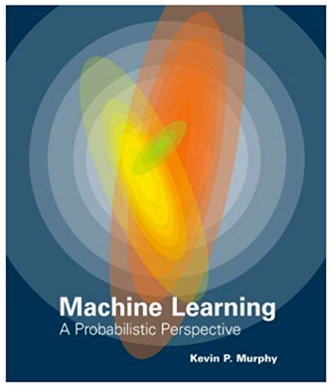
Fundamentals	**Intermediate material**	**Advanced material**
Calculus	Bayesian inference	Molecular biophysics
Linear Algebra	Machine learning	Stochastic processes
Algorithms (basic)	Sequence analysis	Dynamical systems
Statistics (basic)	Population genetics	Information theory
	Theoretical neuroscience	Deep learning
	Algorithms (intermediate)	…
Master all of these topics	**Master at least one of these topics**	**Learn selected topics as needed**

Learn to work through technical books systematically and independently

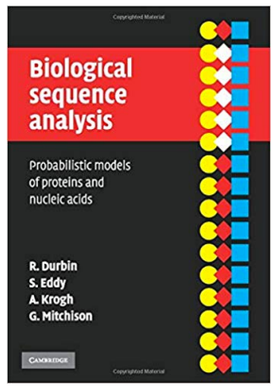
New QB initiative: faculty will help interested students pursue directed independent study of graduate-level material.
Email me <jkinney@cshl.edu> if interested.



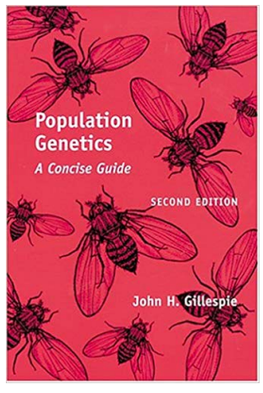
Bayesian Data Analysis, 3rd ed
Gelman et al., 2013



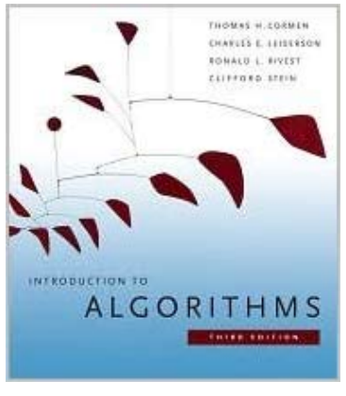
Machine Learning:
A Probabilistic Perspective
Murphy, 2012



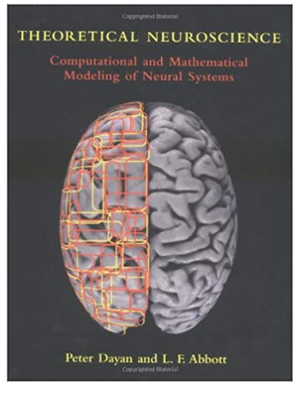
Biological Sequence Analysis
Durbin et al., 1998



Population Genetics:
A concise guide, 2nd ed
Gillespie, 2004



Introduction to Algorithms
Cormen et al., 2009



Theoretical Neuroscience
Dayan and Abbott, 2001

Other tips

Attend the weekly QB seminars
Wednesdays at 12pm, Hawkins.

Attend QB Tea Time
Wednesdays at 4pm, Samet.
Email me to get on mailing list.