Welcome to Quantitative Biology

QB Bootcamp, Day 1
Tuesday, 5 September 2023
10:00am - 10:30am

2023 QB Bootcamp Syllabus

Instructors:

- Justin Kinney, jkinney@cshl.edu
- Ivan Iossifov, iossifov@cshl.edu

Teaching assistants:

- Chandana Rajesh, rajesh@cshl.edu
- Michael Passalacqua, passala@cshl.edu

GitHub Repository: https://github.com/jbkinney/23e_qbbootcamp

Day 1: Tuesday, 5 September 2023, 10am - 6pm, Hershey East Room, Hershey Bldg.

10:00am - 10:30am: **Overview of Quantitative Biology (lecture, Justin)**
10:30am - 12:00pm: **The Unix command line (tutorial, Ivan)**
12:00pm - 1:00pm: *Lunch (provided)*
1:00pm - 1:30pm: **Introduction to Python and Jupyter Notebooks (tutorial, Justin)**
1:30pm - 3:30pm: **Python: data types (tutorial, Justin)**
3:30pm - 4:00pm: *Break*
4:00pm - 6:00pm: **Python: flow control (tutorial, Ivan)**

Day 2: Wednesday, 6 September 2023, 1pm - 6pm, Samet Conference Room, Koch Bldg.

1:00pm - 1:30pm: **Overview of high-performance computing (lecture, Justin)**
1:30am - 3:00pm: **Read mapping using Elzar (tutorial, Justin)**
3:00pm - 3:30pm: Break
3:30pm - 3:00pm: **Introduction to Pandas (lecture, Justin)**
4:00pm - 5:30pm: **Pandas I, TF analysis (tutorial, Justin)**

Day 3: Thursday, 7 September 2023, 10am - 6pm, Plimpton Conference Room, Beckman Bldg.

10:00am - 11:30am: **Pandas II, Replication origin analysis (tutorial, Ivan)**
11:30pm - 12:00pm: **Introduction to Data Visualization (lecture, Justin)**
12:00pm - 1:00pm: *Lunch (provided)*
1:00pm - 2:30pm: **Matplotlib (tutorial, Ivan)**
2:30pm - 3:00pm: *Break*
3:00pm - 4:30pm: **Advanced visualization (tutorial, Justin)**

jbkinney (Justin B. Kinney)

https://github.com/jbkinney

Toggl Papers Biosafety Now! ChatGPT Facebook Twitter Supergroup Sche... SmartSheet My NCBI Evernote Github Overleaf

jbkinney

Type / to search

Overview Repositories 57 Projects Packages Stars

Justin B. Kinney

jbkinney

Edit profile

32 followers · **1** following

Cold Spring Harbor Laboratory
United States
jkinney@cshl.edu
http://kinneylab.labsites.cshl.edu/

Achievements

Pinned

Customize your pins

23e_urp Public
Code for the 2023 CSHL URP Python Programming Course
Jupyter Notebook 1

mavenn Public
MAVE-NN: genotype-phenotype maps from multiplex assays of variant effect
Jupyter Notebook 21 4

logomaker Public
Software for the visualization of sequence-function relationships
Jupyter Notebook 165 34

23e_qbbootcamp Public
2023 Quantitative Biology Bootcamp at Cold Spring Harbor Laboratory
Jupyter Notebook

52 contributions in the last year

Contribution settings

Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug

Mon
Wed
Fri

Learn how we count contributions

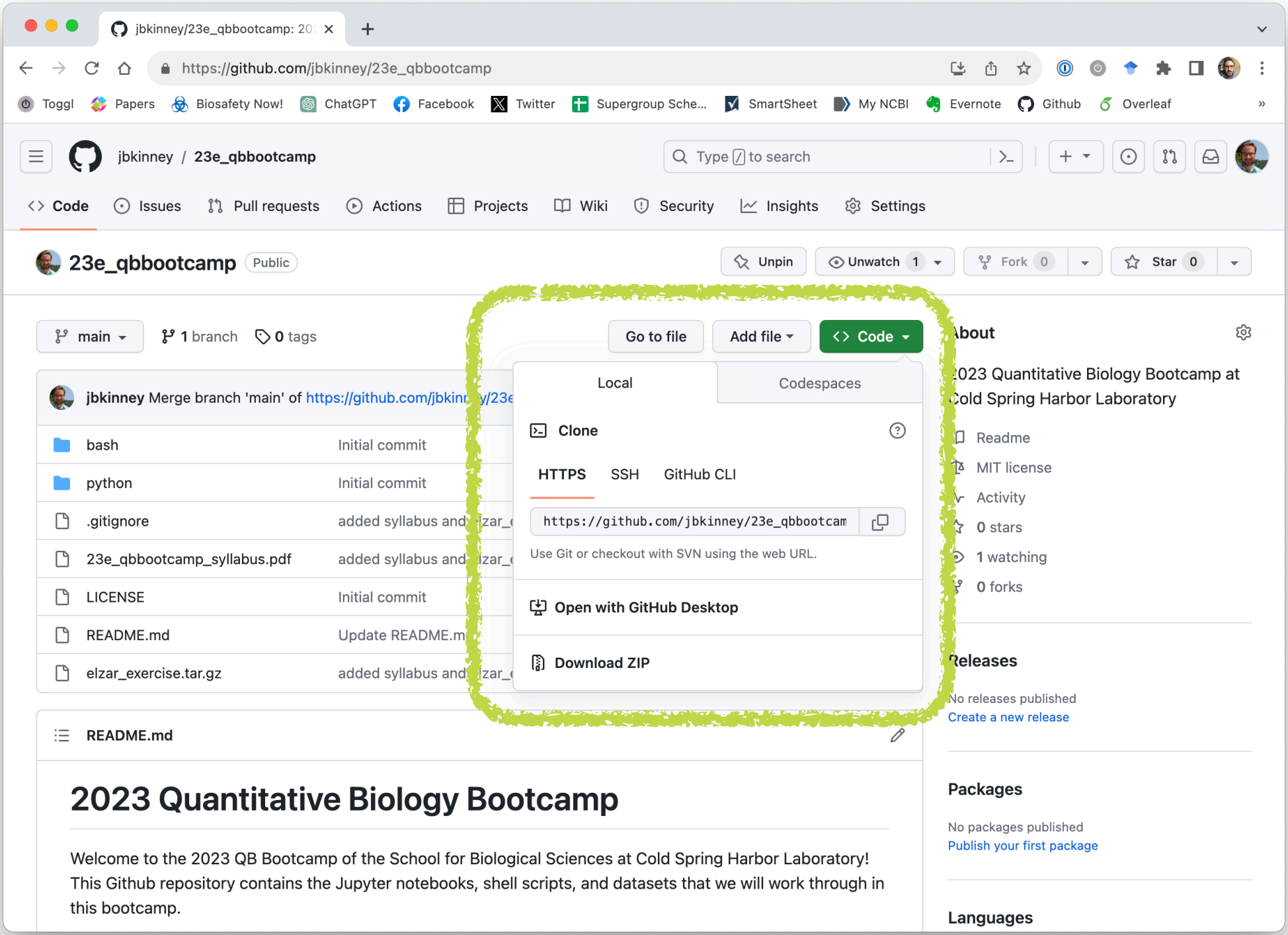
Less More

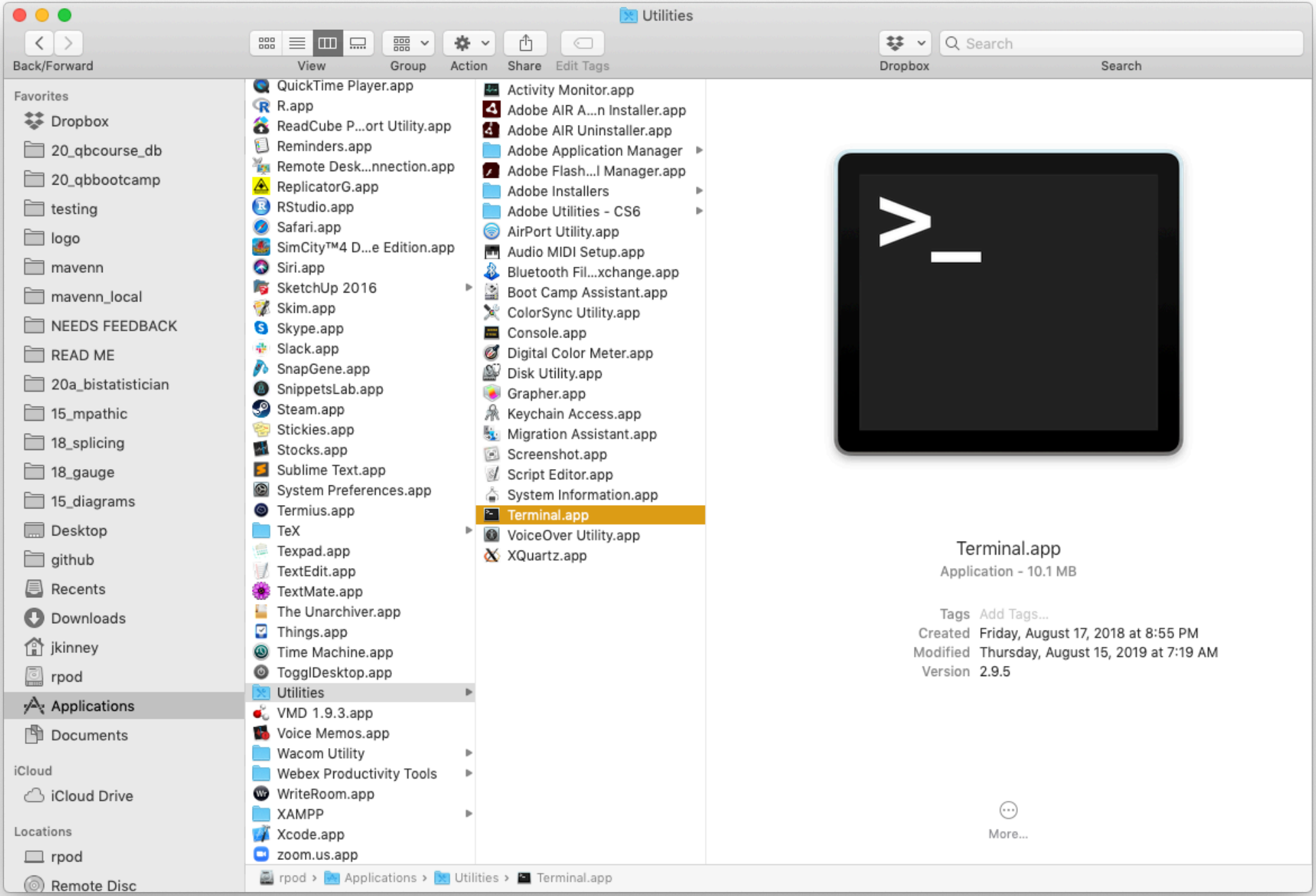
Contribution activity

2023

2022

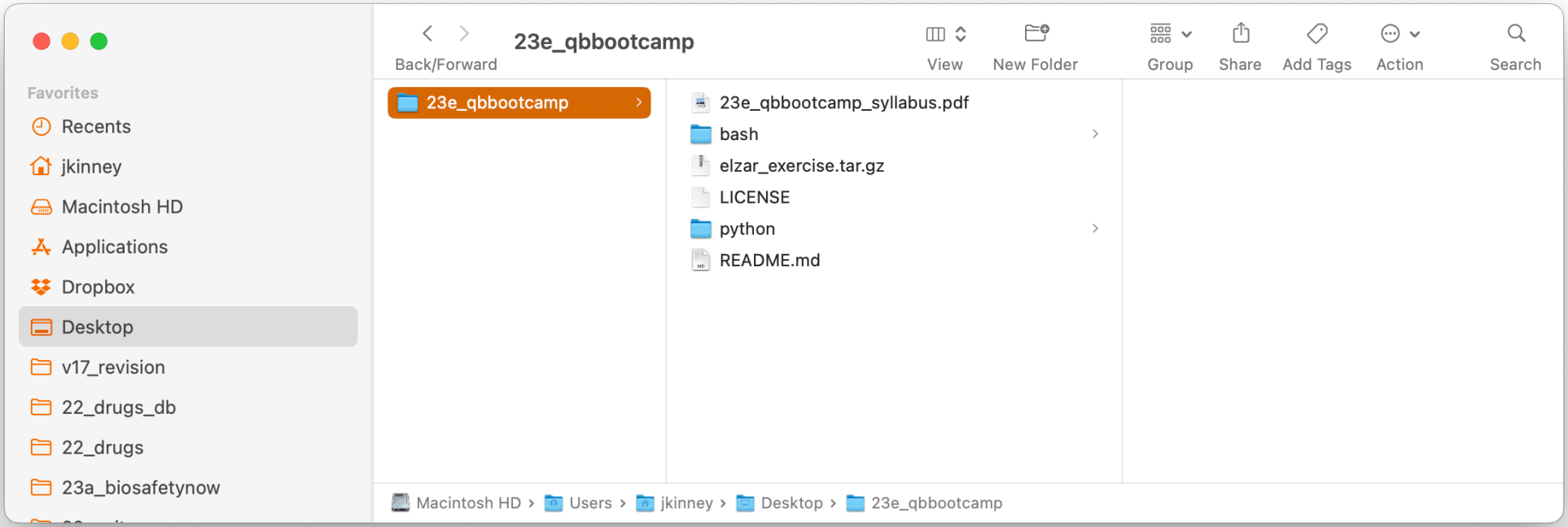
September 2023


jbkinney/23e_qbbootcamp: 20
https://github.com/jbkinney/23e_qbbootcamp
Toggl
Papers
Biosafety Now!
ChatGPT
Facebook
Twitter
Supergroup Sche...
SmartSheet
My NCBI
Evernote
Github
Overleaf
jbkinney / 23e_qbbootcamp
Type / to search
Code
Issues
Pull requests
Actions
Projects
Wiki
Security
Insights
Settings
23e_qbbootcamp
Public
Unpin
Unwatch 1
Fork 0
Star 0
main
1 branch
0 tags
Go to file
Add file
Code
jbkinney Merge branch 'main' of https://github.com/jbkinney/23e
bash
Initial commit
python
Initial commit
.gitignore
added syllabus and elzar_e
23e_qbbootcamp_syllabus.pdf
added syllabus and elzar_e
LICENSE
Initial commit
README.md
Update README.m
elzar_exercise.tar.gz
added syllabus and elzar_e
Local
Codespaces
Clone
HTTPS
SSH
GitHub CLI
https://github.com/jbkinney/23e_qbbootcam
Use Git or checkout with SVN using the web URL.
Open with GitHub Desktop
Download ZIP
About
2023 Quantitative Biology Bootcamp at Cold Spring Harbor Laboratory
Readme
MIT license
Activity
0 stars
1 watching
0 forks
Releases
No releases published
Create a new release
Packages
No packages published
Publish your first package
Languages
README.md
2023 Quantitative Biology Bootcamp
Welcome to the 2023 QB Bootcamp of the School for Biological Sciences at Cold Spring Harbor Laboratory! This Github repository contains the Jupyter notebooks, shell scripts, and datasets that we will work through in this bootcamp.


Utilities
Back/Forward
View
Group
Action
Share
Edit Tags
Dropbox
Search
Search
Favorites
Dropbox
20_qbcourse_db
20_qbbootcamp
testing
logo
mavenn
mavenn_local
NEEDS FEEDBACK
READ ME
20a_bistatistician
15_mpathic
18_splicing
18_gauge
15_diagrams
Desktop
github
Recents
Downloads
jkinney
rpod
Applications
Documents
iCloud
iCloud Drive
Locations
rpod
Remote Disc
QuickTime Player.app
R.app
ReadCube P...ort Utility.app
Reminders.app
Remote Desk...nnection.app
ReplicatorG.app
RStudio.app
Safari.app
SimCity™4 D...e Edition.app
Siri.app
SketchUp 2016
Skim.app
Skype.app
Slack.app
SnapGene.app
SnippetsLab.app
Steam.app
Stickies.app
Stocks.app
Sublime Text.app
System Preferences.app
Termius.app
TeX
Texpad.app
TextEdit.app
TextMate.app
The Unarchiver.app
Things.app
Time Machine.app
TogglDesktop.app
Utilities
VMD 1.9.3.app
Voice Memos.app
Wacom Utility
Webex Productivity Tools
WriteRoom.app
XAMPP
Xcode.app
zoom.us.app
Activity Monitor.app
Adobe AIR A...n Installer.app
Adobe AIR Uninstaller.app
Adobe Application Manager
Adobe Flash...l Manager.app
Adobe Installers
Adobe Utilities - CS6
AirPort Utility.app
Audio MIDI Setup.app
Bluetooth Fil...xchange.app
Boot Camp Assistant.app
ColorSync Utility.app
Console.app
Digital Color Meter.app
Disk Utility.app
Grapher.app
Keychain Access.app
Migration Assistant.app
Screenshot.app
Script Editor.app
System Information.app
Terminal.app
VoiceOver Utility.app
XQuartz.app
Terminal.app
Application - 10.1 MB
Tags Add Tags...
Created Friday, August 17, 2018 at 8:55 PM
Modified Thursday, August 15, 2019 at 7:19 AM
Version 2.9.5
More...
rpod › Applications › Utilities › Terminal.app

Desktop — -bash — 104×17

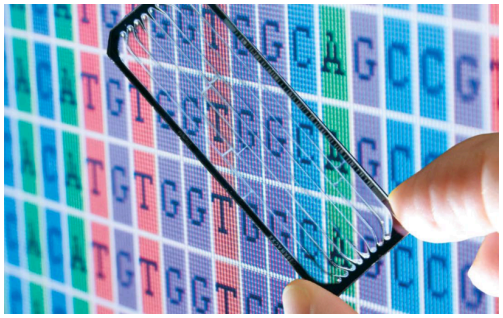
```
(base) jkinney@dobbs:~/Desktop$ git clone https://github.com/jbkinney/23e_qbbootcamp.git
Cloning into '23e_qbbootcamp'...
remote: Enumerating objects: 34, done.
remote: Counting objects: 100% (10/10), done.
remote: Compressing objects: 100% (9/9), done.
remote: Total 34 (delta 1), reused 8 (delta 1), pack-reused 24
Receiving objects: 100% (34/34), 79.19 MiB | 6.34 MiB/s, done.
Resolving deltas: 100% (7/7), done.
(base) jkinney@dobbs:~/Desktop$
```


Back/Forward
23e_qbbootcamp
View
New Folder
Group
Share
Add Tags
Action
Search
Favorites
Recents
jkinney
Macintosh HD
Applications
Dropbox
Desktop
v17_revision
22_drugs_db
22_drugs
23a_biosafetynow
23e_qbbootcamp
23e_qbbootcamp_syllabus.pdf
bash
elzar_exercise.tar.gz
LICENSE
python
README.md
Macintosh HD › Users › jkinney › Desktop › 23e_qbbootcamp

What is Quantitative Biology?

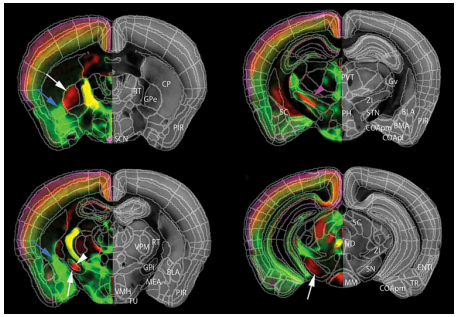
Quantitative biology is a vast field

Genomics



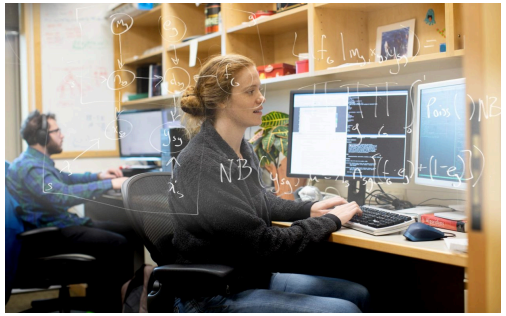
Functional genomics
Evolutionary genomics
Genome dynamics
Technology development

Neuroscience



Data analysis
Modeling neural systems
Behavioral modeling

Other



Biophysics
Machine learning
Software development

Who does Quantitative Biology at CSHL?

Core QB program

Ivan Iossofov

Peter Koo

Hannah Meyer

Justin Kinney

Dan Levy

Saket Navlakah

Alexander Krasnitz

David McCandlish

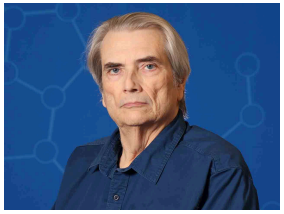
Adam Siepel

Michael Wigler

QB Associated Faculty

Genomics

Alexander Dobin



Richard McCombie

Tom Gingeras

Doreen Ware

Neuroscience

Ben Cowley

Alexei Koulakov

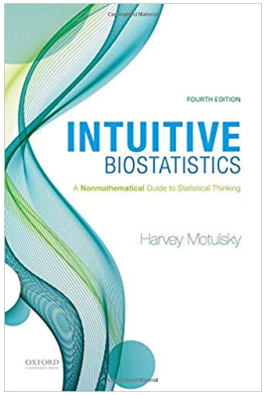
Partha Mitra

What QB skills should all biology researchers have?

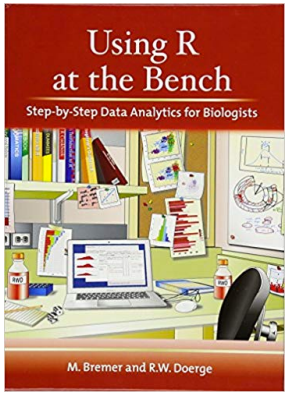
Learn to interpret standard statistics

Key statistical concepts:

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



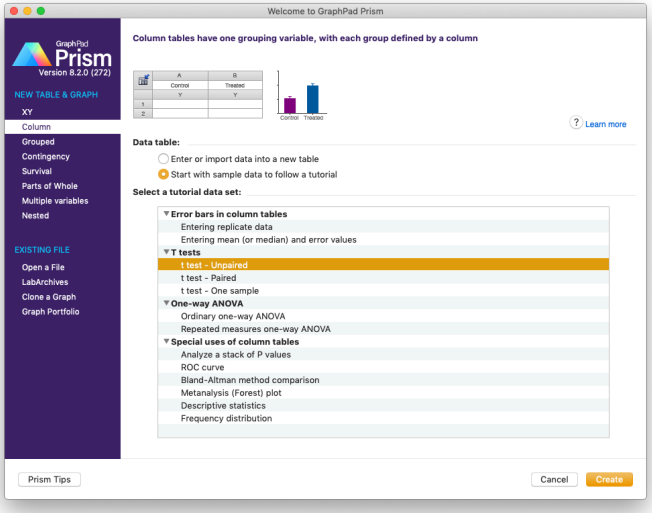
Intuitive Biostatistics, 4th ed
Motulsky (2018)
[CEO, GraphPad Software]

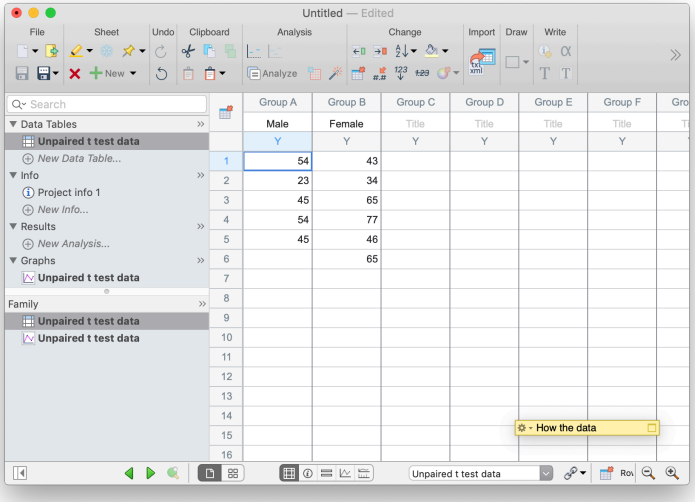


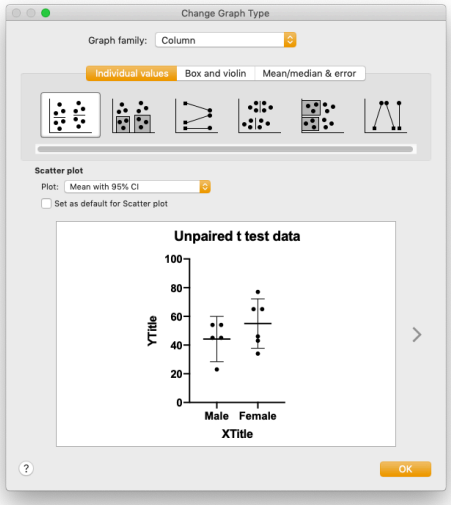
Using R at the Bench
Bremmer & Doerge (2015)

Learn to compute standard statistics

Prism 8







Alternatively:

python™

R

MATLAB

Excel

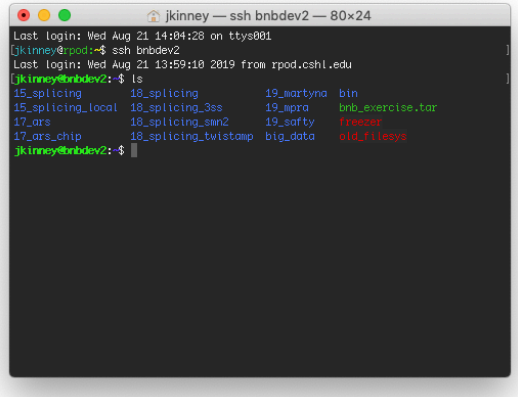
Learn to navigate UNIX systems

Sequencer

Microscope

Databases

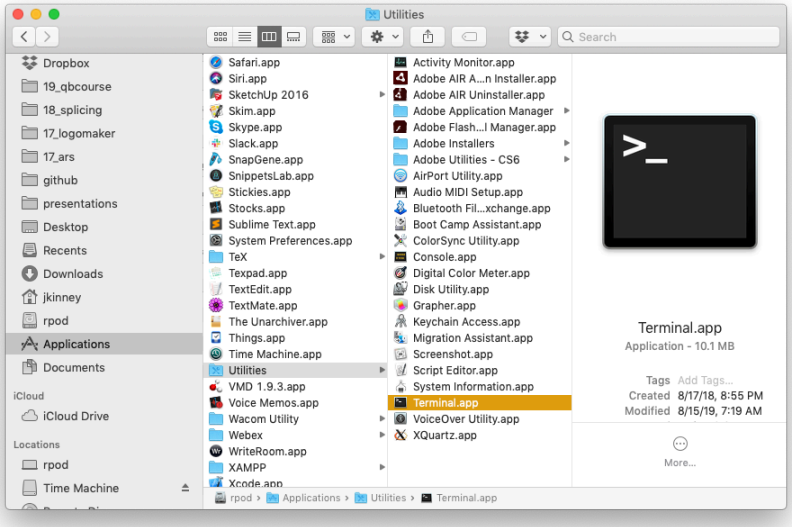
High Performance
Computer Cluster

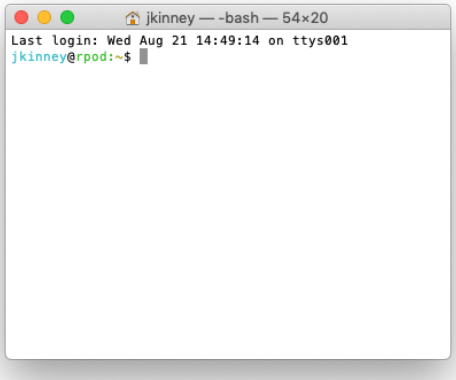


UNIX command line

Mac OS X is based on UNIX

Finder



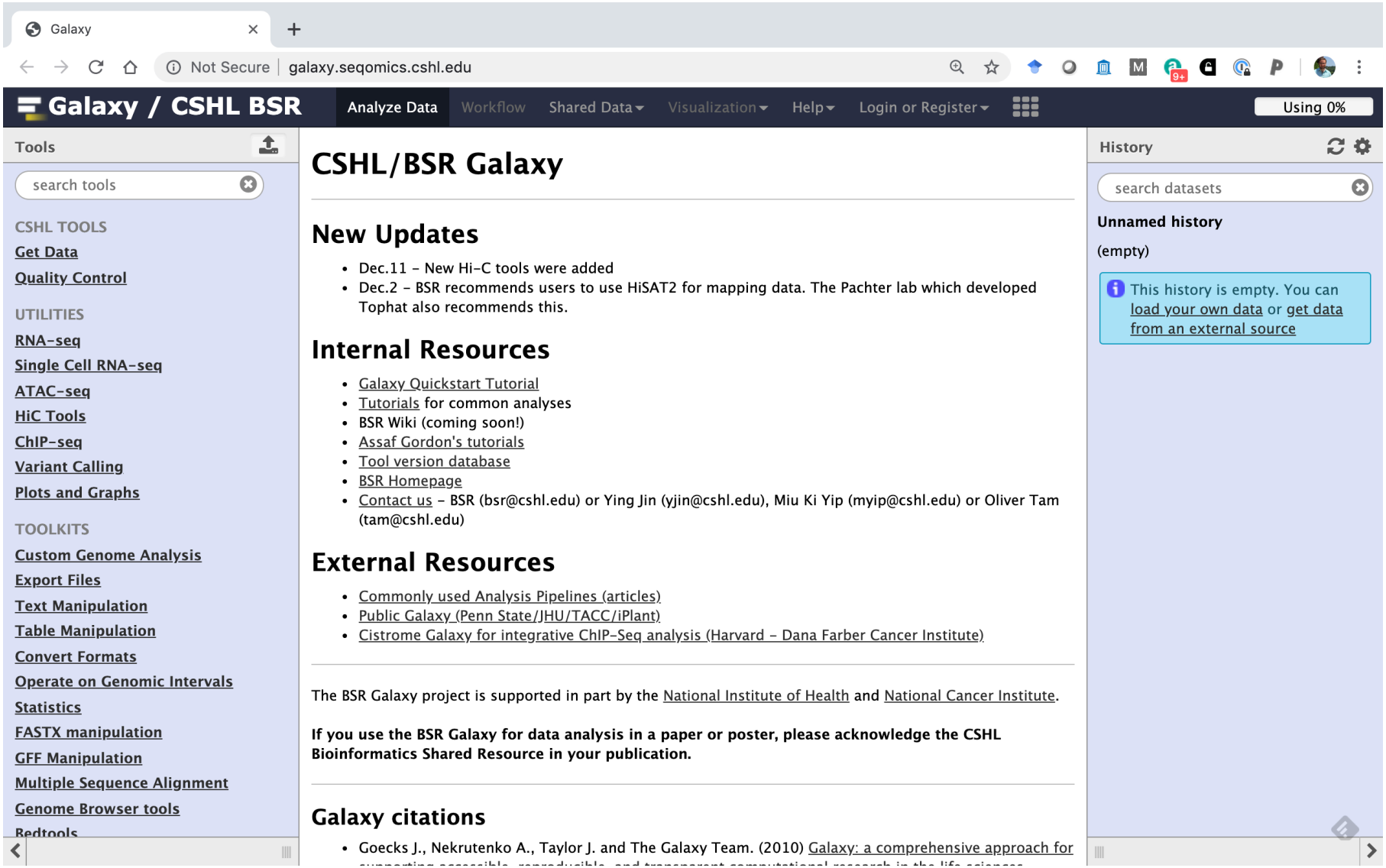


Applications/Utilities/Terminal.app

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 - **primary language for deep learning**	- clunky dataframes - clunky graphics - clunky statistics
	- streamlined for statistics - highly developed for genomics - great graphics	- strange language - not great for building pipelines
	- used heavily in neuroscience	- 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

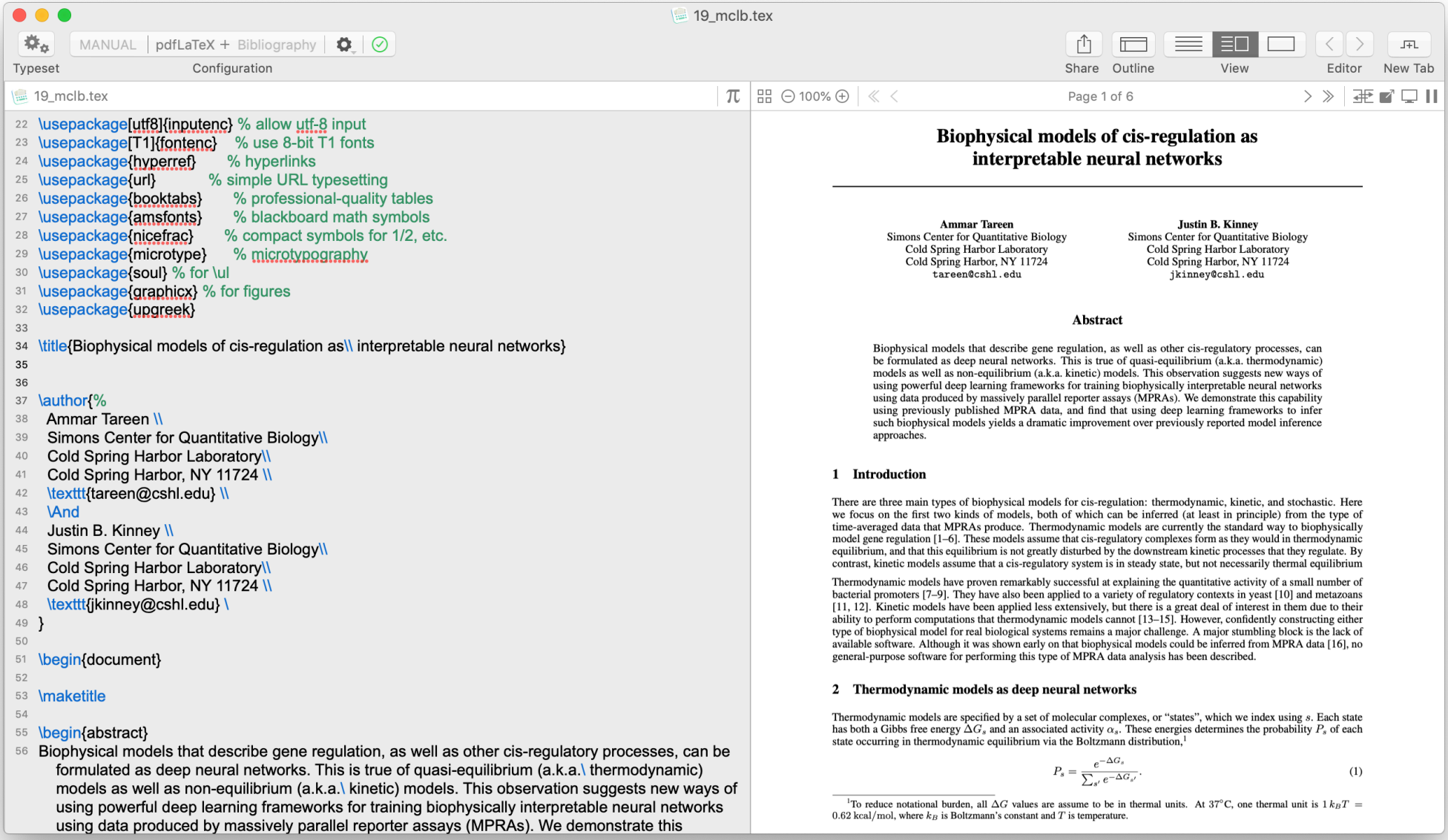
BEST ONLINE COURSES FOR PYTHON AT A GLANCE

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

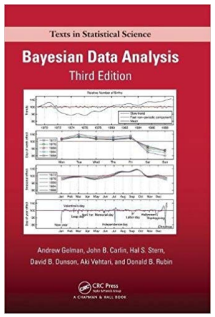


Develop core quantitative knowledge

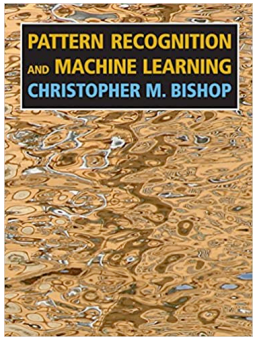
Fundamentals	**Intermediate material**	**Advanced material**
Calculus	Bayesian inference	Molecular biophysics
Linear Algebra	Introductory machine learning	Stochastic processes
Algorithms (basic)	Deep learning	Dynamical systems
Statistics (basic)	Sequence analysis	Information theory
	Population genetics	…
	Theoretical neuroscience	
	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

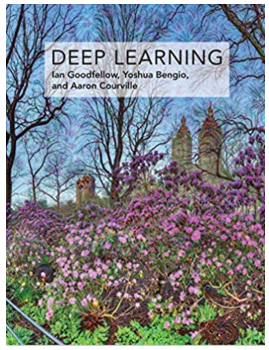
Mentored independent study in QB:
Faculty will help interested students pursue directed reading of graduate-level material.
Email me <jkinney@cshl.edu> if interested.



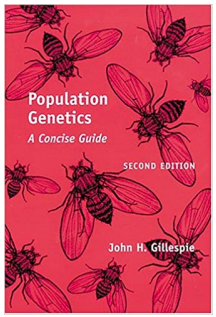
Bayesian Data Analysis, 3rd ed
Gelman et al., 2013



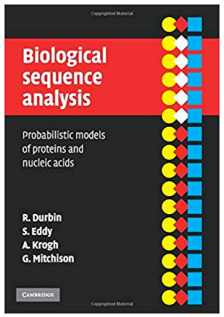
Pattern Recognition and Machine Learning
Bishop, 2006



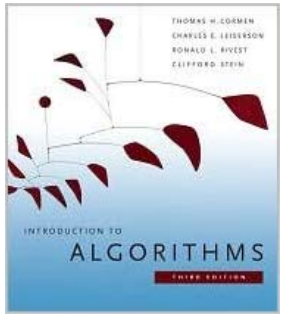
Deep Learning
Goodfellow et al., 2016



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



Biological Sequence Analysis
Durbin et al., 1998



Introduction to Algorithms
Cormen et al., 2009

Other tips

Attend the weekly QB seminars

Wednesdays at 12pm, Hawkins.

Attend QB Tea Time

Tuesday at 3pm, Samet.

Email Peter Koo to get on mailing list.