Welcome to Quantitative Biology

QB Bootcamp, Day 1
Tuesday, 3 September 2024
10:00am - 10:30am

2024 QB Bootcamp Syllabus

2024 Quantitative Biology Bootcamp, CSHL School of Biological Sciences

Instructors:

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

Teaching assistants:

- Ari Benjamin, benjami@cshl.edu
- Kaiser Loell, loell@cshl.edu

GitHub Repository: https://github.com/jbkinney/24e_qbbootcamp/

Day 1: Tuesday, 3 September 2024, 10am - 6pm, Plimpton Conference Room, Beckman 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 Break*
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, 4 September 2024, 1pm - 6pm, Plimpton Conference Room, Beckman Bldg.
1:00pm - 1:30pm: **NextGen sequencing and high-performance computing (lecture, Justin)**
1:30am - 3:00pm: **Read mapping using Elzar (tutorial, Justin)**
3:00pm - 3:15pm: Break
3:15pm - 3:30pm: **Introduction to dataframes (lecture, Justin)**
3:30pm - 5:00pm: **Pandas I, TF analysis (tutorial, Ivan)**

Day 3: Thursday, 5 September 2024, 10am - 6pm, Plimpton Conference Room, Beckman Bldg.
10:00am - 11:30am: **Pandas II, Replication origin analysis (tutorial, Justin)**
11:30am - 12:00pm: **Introduction to Data Visualization (lecture, Justin)**
12:00pm - 1:00pm: *Lunch Break*
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

New Chrome available

Toggl | Overleaf | SmartSheet | Papers | Github | ChatGPT | CSHL: Lawson | Supergroup Sched... | Mimecast | eRA Commons | Concur

jbkinney

Type / to search

Overview | Repositories 59 | Projects | Packages | Stars 1

Justin B. Kinney

jbkinney

Edit profile

36 followers · **1** following

Cold Spring Harbor Laboratory

United States

jkinney@cshl.edu

http://kinneylab.labsites.cshl.edu/

Pinned Order updated

Customize your pins

24e_qbbootcamp Public

Repository for the 2024 QB Bootcamp

Jupyter Notebook

24e_urp Public

Code for the 2024 CSHL URP Python Programming Course

Jupyter Notebook 1 3

mavenn Public

MAVE-NN: genotype-phenotype maps from multiplex assays of variant effect

Jupyter Notebook 24 5

logomaker Public

Software for the visualization of sequence-function relationships

Jupyter Notebook 178 35

61 contributions in the last year

Contribution settings

Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul

Mon

Wed

Fri

Learn how we count contributions

Less More

2024

2023

2022

2021

2020

2019

Contribution activity

jbkinney/24e_qbbootcamp: R

https://github.com/jbkinney/24e_qbbootcamp

New Chrome available

Toggl | Overleaf | SmartSheet | Papers | Github | ChatGPT | CSHL: Lawson | Supergroup Sched... | Mimecast | eRA Commons | Concur

jbkinney / 24e_qbbootcamp

Type / to search

Code | Issues | Pull requests | Actions | Projects | Wiki | Security | Insights | Settings

24e_qbbootcamp

Public

Unpin | Unwatch 1 | Fork 0 | Star 0

main | 1 Branch | 0 Tags

Go to file t | + | Code

jbkinney Update README.md

bash	adde fi
lectures	adde fi
python	adde fi
.gitignore	adde fi
23e_qbbootcamp_syllabus.pdf	adde fi
LICENSE	adde fi
README.md	Upda README.md 15 minutes ago

Local | Codespaces

Clone

HTTPS | SSH | GitHub CLI

https://github.com/jbkinney/24e_qbbootcamp.

Clone using the web URL.

Open with GitHub Desktop

Download ZIP

README | MIT license

2024 Quantitative Biology Bootcamp

About

Repository for the 2024 QB Bootcamp

Readme

MIT license

Activity

0 stars

1 watching

0 forks

Releases

No releases published

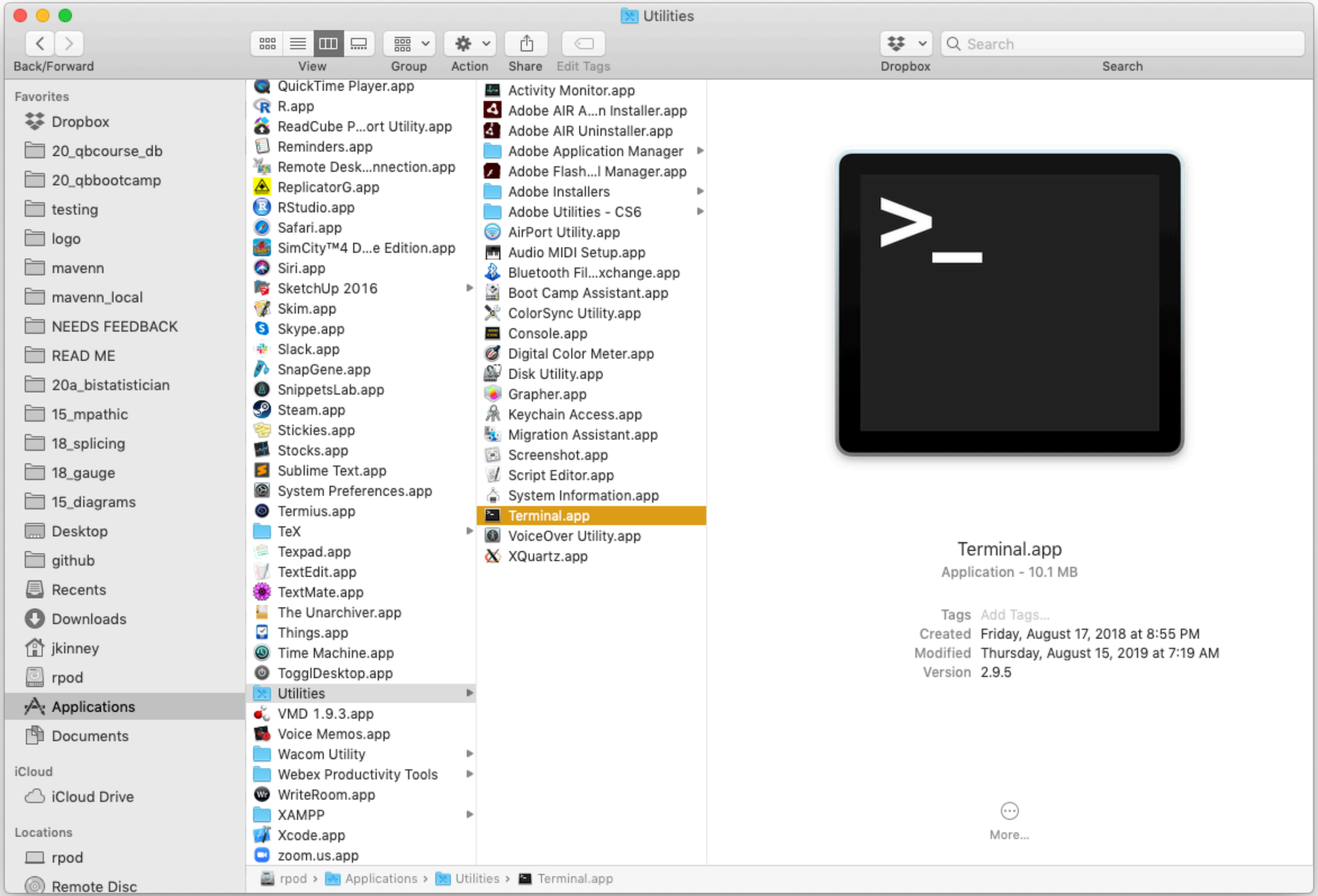
Create a new release

Packages

No packages published

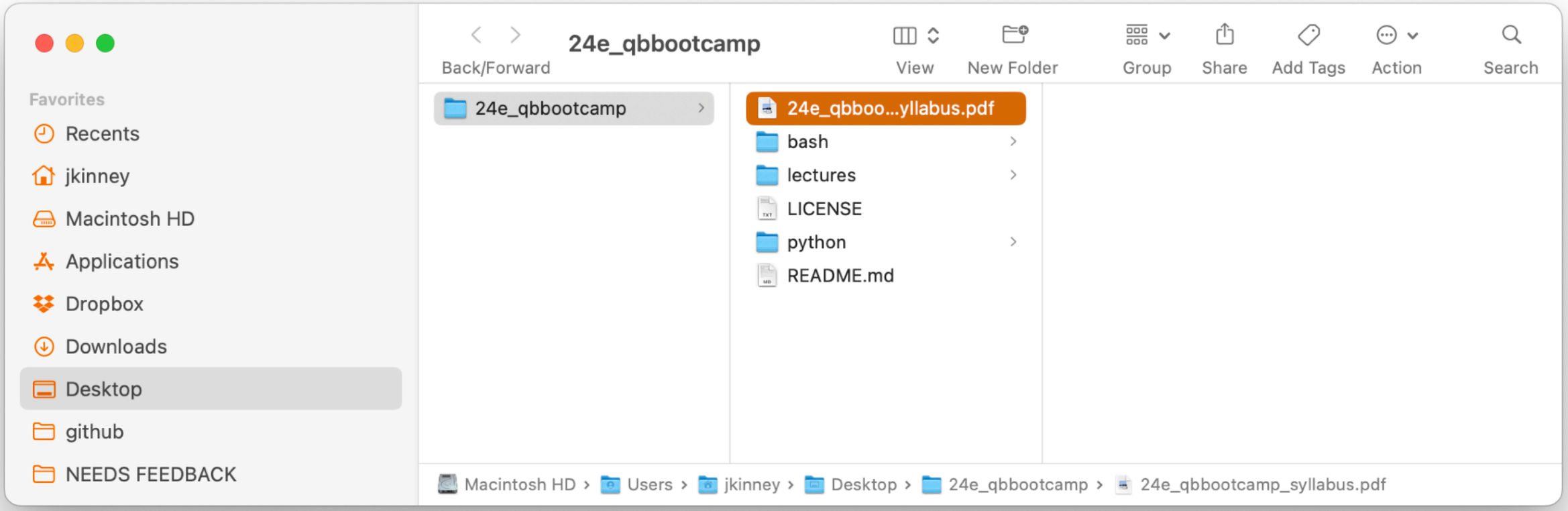
Publish your first package

Languages


Utilities
Back/Forward
View
Group
Action
Share
Edit Tags
Dropbox
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 — jkinney@bamdev1:/grid/kinney/home — -bash — 99×15

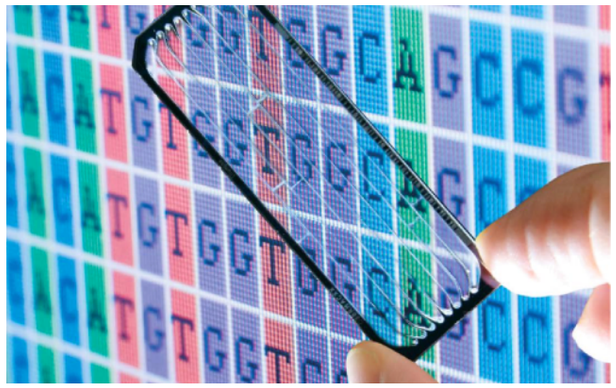
```
(base) jkinney@MJM-2023:~/Desktop$ git clone https://github.com/jbkinney/24e_qbbootcamp.git
Cloning into '24e_qbbootcamp'...
remote: Enumerating objects: 49, done.
remote: Counting objects: 100% (12/12), done.
remote: Compressing objects: 100% (11/11), done.
remote: Total 49 (delta 5), reused 2 (delta 0), pack-reused 37 (from 1)
Receiving objects: 100% (49/49), 75.97 MiB | 14.24 MiB/s, done.
Resolving deltas: 100% (17/17), done.
(base) jkinney@MJM-2023:~/Desktop$
```


Back/Forward
24e_qbbootcamp
View
New Folder
Group
Share
Add Tags
Action
Search
Favorites
Recents
jkinney
Macintosh HD
Applications
Dropbox
Downloads
Desktop
github
NEEDS FEEDBACK
24e_qbbootcamp
24e_qbboo...yllabus.pdf
bash
lectures
LICENSE
python
README.md
Macintosh HD › Users › jkinney › Desktop › 24e_qbbootcamp › 24e_qbbootcamp_syllabus.pdf

What is Quantitative Biology?

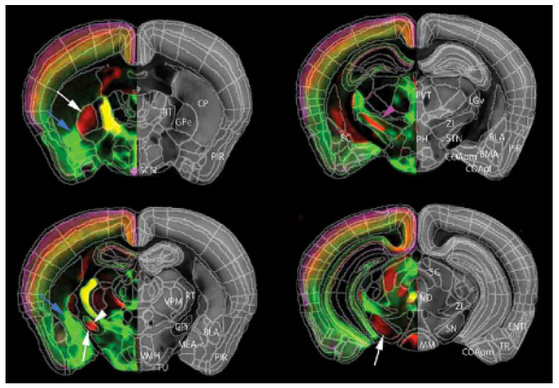
Quantitative biology is a vast field

Genomics



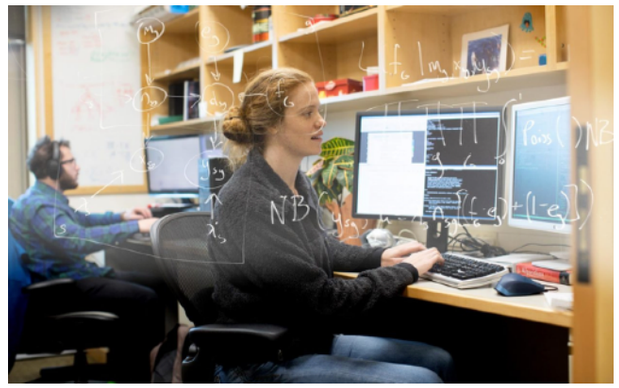
Functional genomics
Evolutionary genomics
Genome dynamics
Technology development

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

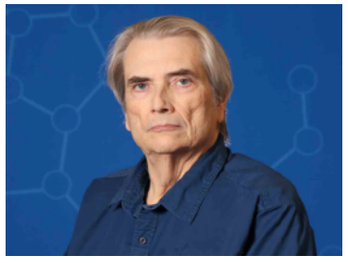
David McCandlish

Adam Siepel

Michael Wigler

QB Associated Faculty

Genomics



Richard McCombie

Tom Gingeras

Doreen Ware

Neuroscience

Ben Cowley

David Klindt

Alexei Koulakov

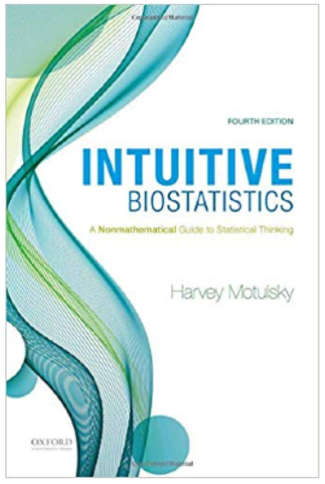
Partha Mitra

What QB skills should all biology researchers have?

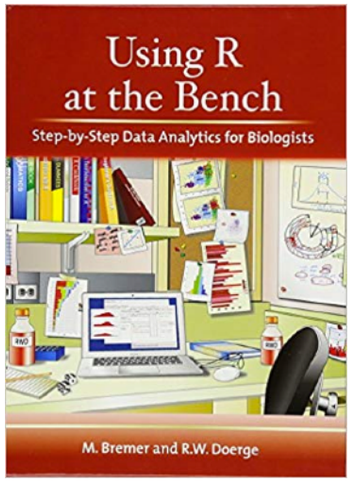
Learn to interpret standard statistics

Key statistical concepts:

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



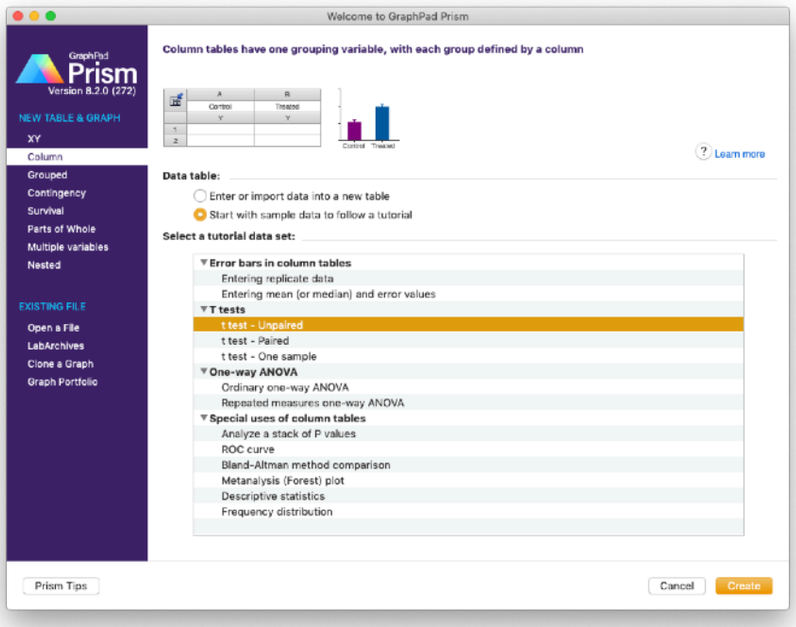
Intuitive Biostatistics, 4th ed
Motulsky (2018)
[CEO, GraphPad Software]

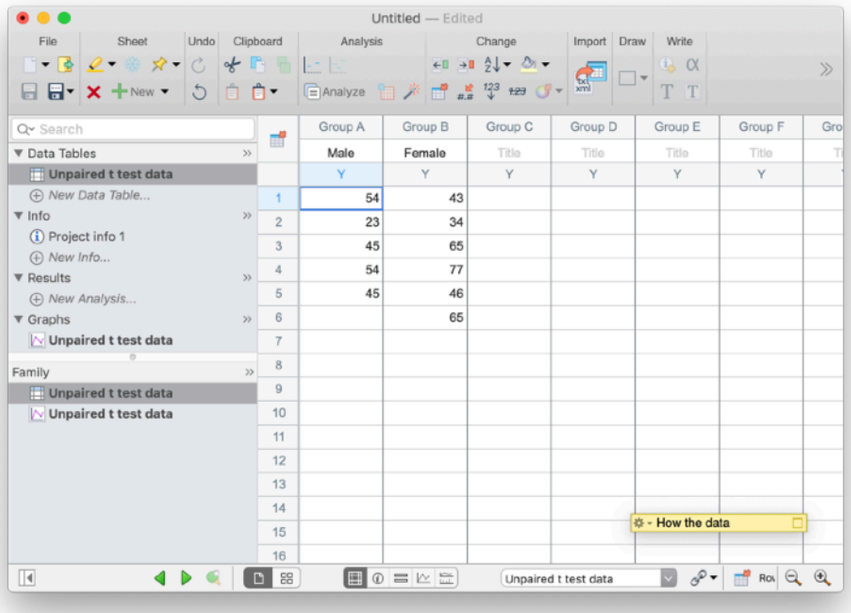


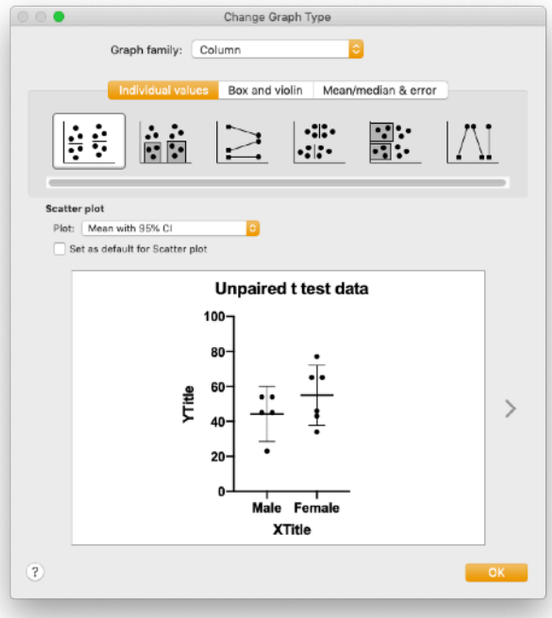
Using R at the Bench
Bremmer & Doerge (2015)

Learn to compute standard statistics

Prism 8







Alternatively:

python™

R

MATLAB

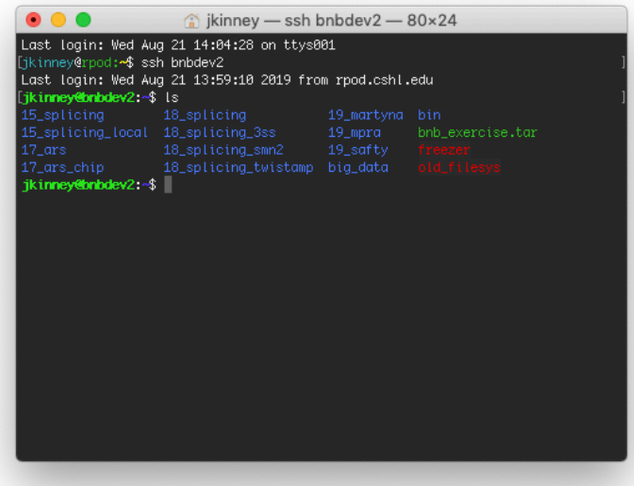
Excel

Learn to navigate UNIX systems

Sequencer

Microscope

High Performance
Computer Cluster

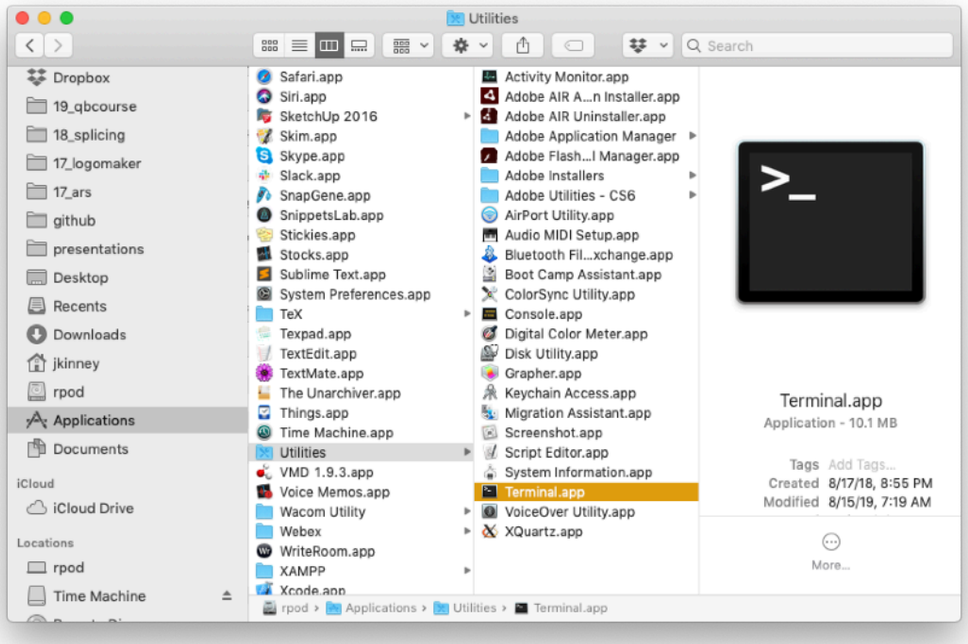


UNIX command line

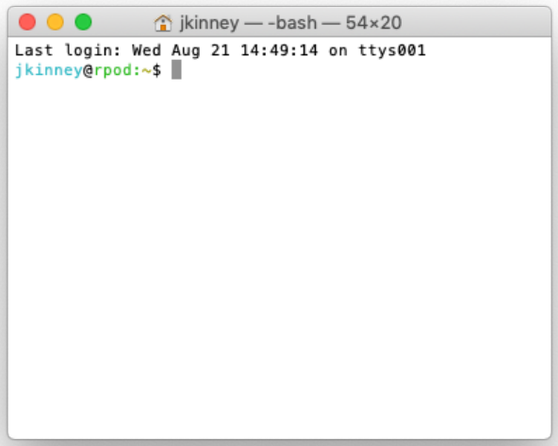
Databases

Mac OS X is based on UNIX

Finder



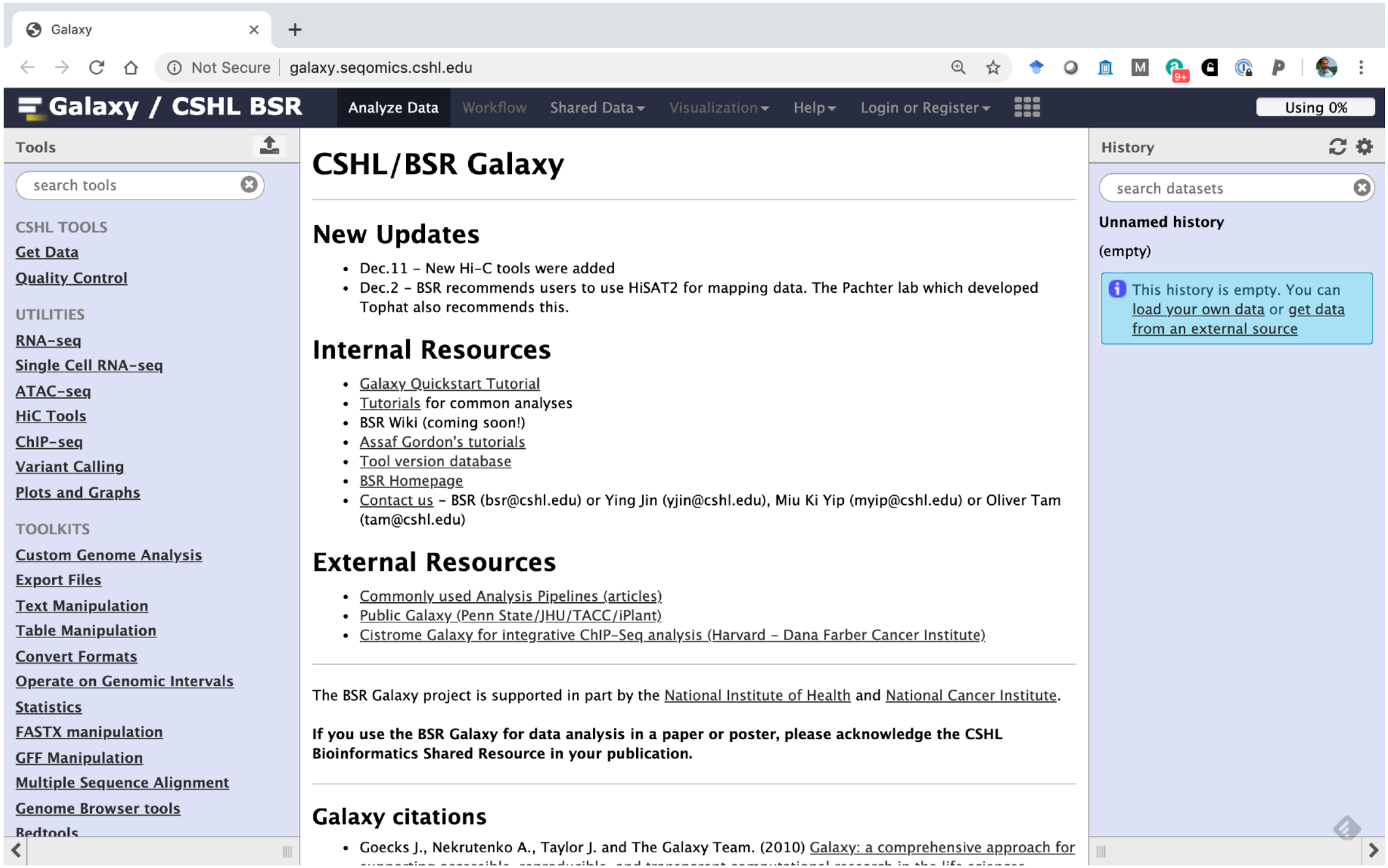
Applications/Utilities/Terminal.app



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 - **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

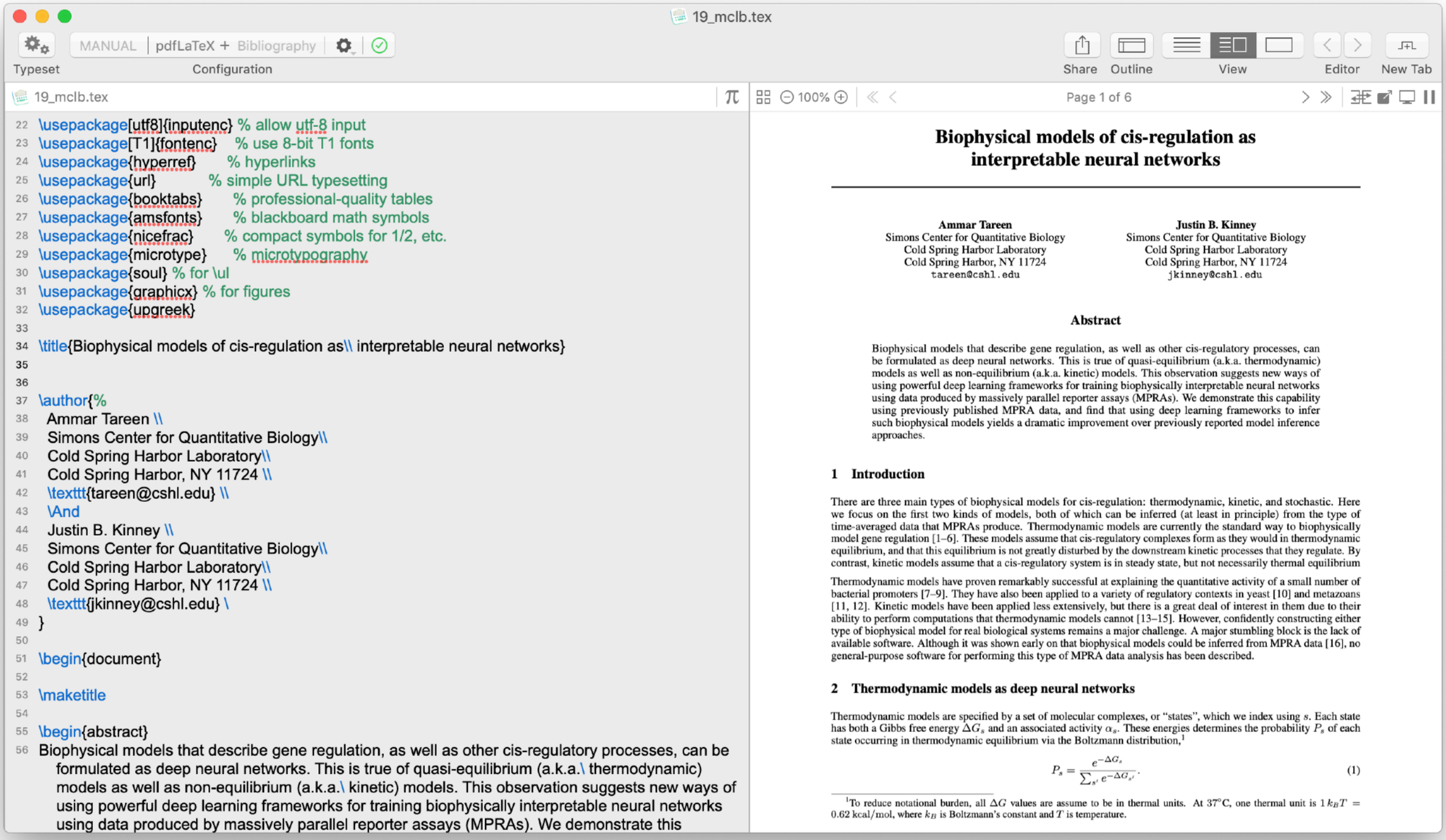
BEST ONLINE COURSES FOR PYTHON AT A GLANCE

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

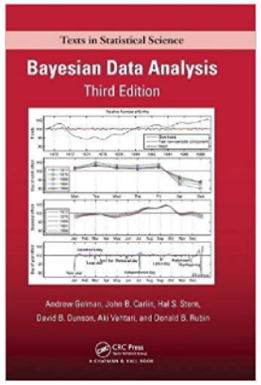


Develop core quantitative knowledge

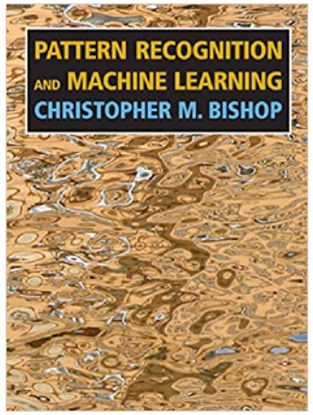
Fundamentals	**Intermediate material**	**Advanced material**
Calculus	Bayesian inference	Molecular biophysics
Linear Algebra	Introductory machine learning	Stochastic processes
Algorithms (basic)	Deep learning	Dynamical systems
Statistics (basic)	Sequence analysis	Information theory
	Population genetics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

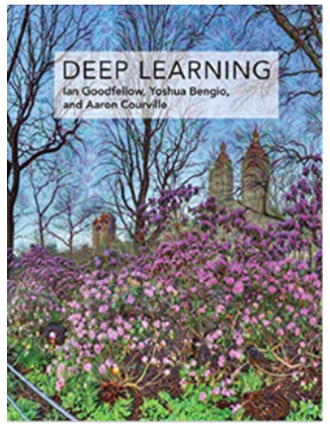
Mentored independent study in QB:
Faculty will help interested students pursue directed reading of graduate-level material.
Email me <jkinney@cshl.edu> if interested.



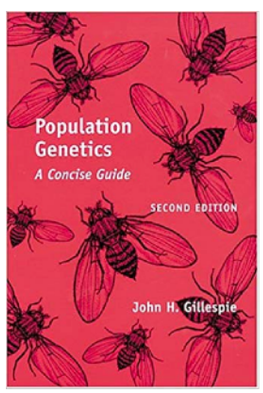
Bayesian Data Analysis, 3rd ed
Gelman et al., 2013



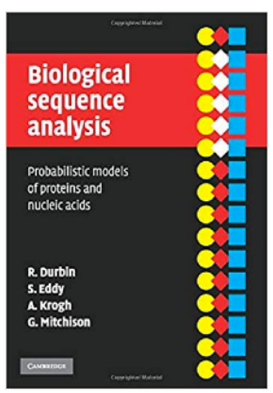
Pattern Recognition and Machine Learning
Bishop, 2006



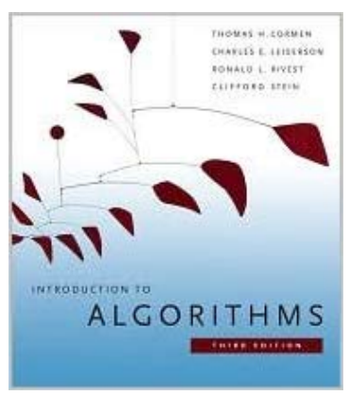
Deep Learning
Goodfellow et al., 2016



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.