Session topics	Objectives	Readings	Activities/ Assignments
Week of: Jan. 7-9			
Module 1: Research computing for biological data analysis * Command line computing environments * General command line data manipulation principles: files, redirects, pipes, and more * Built-in command line data manipulation tools * Accessing grid and cloud research computing environments **Module 2: Introductory biological data analysis in Python** * Basics of Python as a general data manipulation environment * Biological data types in Python * Text, numerical, and structured data types * Data input and output * Data visualization	**Upon Successful completion of this week, you should be able to:** 1. Understand the capabilities of and navigate local command line environments 2. Access basic grid and cloud environments for biological computing 3. Carry out basic biological data manipulation in Python	**Required** Review syllabus in advance of class **Recommended** A Quick Guide for Developing Effective Bioinformatics Programming Skills (https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000589) A Quick Guide to Organizing Computational Biology Projects (https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000424) An Introduction to Programming for Bioscientists: A Python-Based Primer (https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004867)	**Activity #1:** system-specific command line setup (Windows/Mac/Linux), local command line playground, Linux Journey interactive command line tutorial, Cannon-Odyssey/AWS/GCP **Homework #1:** DNA sequence and transcriptomic data analysis using built-in command line tools and Python (on the cloud if needed) **Activity #2:** local Python/Jupyter setup, Data Camp Python basics, FASTA/FASTQ genomic data handling, RNA-seq transcriptional data handling, plotting and visualization

HARVARD T.H. CHAN | SCHOOL OF PUBLIC HEALTH

Week of: Jan. 14-16			
Module 3: Quantitative data analysis in Python and R * Python libraries for scientific data analysis: NumPy, SciPy, Pandas, Matplotlib, Seaborn * Basics of R and RStudio * Statistical methods and visualization in R * Capabilities, pros, and cons of different biological data manipulations in different environments **Module 4: Quantitative data structures and finding additional resources** * Data structures and how to use them: vectors, lists, matrices, sets, dictionaries, etc. * Statistical methods and terminology: probabilities, probability distributions, and hypothesis tests * Generative model fitting, discriminative predictors * Online resources for quantitative biology and how to use them	**Upon Successful completion of this week, you should be able to:** 1. Understand quantitative data handling and visualization capabilities of Python and R 2. Recognize and use terminology for general quantitative data structures, models, and methods 3. Know how to find more help and documentation for computational biology tools	**Recommended** A Quick Guide to Teaching R Programming to Computational Biology Students (https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000482) Ten quick tips for getting the most scientific value out of numerical data (https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006141) Ten simple rules for biologists learning to program (https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005871) Ten Simple Rules for Effective Computational Research (https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003506)	**Activity #3:** NumPy/SciPy/Pandas/Seaborn in Jupyter, RStudio setup, Data Camp R basics **Homework #2:** 'omics data statistics and visualization using Jupyter and knitr **Activity #4:** RStudio knitr for basic model fitting, prediction, hypothesis testing, and permutation tests; documentation scavenger hunts for Python, Python libraries, R/CRAN, Bioconductor, Stack Overflow, and general Googling