Programming for data analyses

Jason Stajich

https://biodataprog.github.io/2018_programming-intro/

What to expect

- This course will cover basics of using UNIX and Python for data analyses

- The topics will focus on skills for processing, filtering data with an emphasis on biological, genomic or transcriptomic sequence data.

- No prior programming experience required.

- You must have access to a computer that allows you to either run the terminal (e.g OSX or Linux) or run programs like SSH to log into UCR Biocluster.

- Some amount biology background will be assumed here. When in doubt ask questions. Others will be likely to have same questions!

Course Logistics

1. Meeting in University Laboratory Building (ULB) Bioinformatics suite

2. Wednesdays scheduled for 1 hrs, Fridays for 2 hrs

3. Expect to have a mix of in-class workshops and lectures.

4. Homework problem sets will be assigned to give you more practice and improve skills. If you don't practice you won't retain or master these skills.

5. Laptops are available during class setting, but you will need to have access to a computer outside of class to complete assignments

6. Accounts to the IIGB Biocluster will be provided during the course.

Resources

1. *Bioinformatics Data Skills: Reproducible and Robust Research with Open Source Tools*. Vince Buffalo. 2015 O'Reilly & Associates. Available from O'Reilly and Associates, Amazon

2. *Unix and Perl to the Rescue: A Primer*. Keith Bradnam and Ian Korf. Free PDF

3. *Unix and Perl to the rescue!* Bradnam and Korf. Amazon

4. Rosalind - An online platform to learn bioinformatics and programming in Python.

5. Software Carpentry - https://software-carpentry.org/ and Data Carpentry - http://www.datacarpentry.org/.

6. Berk Ekmekci, Charles E. McAnany, Cameron Mura. An Introduction to Programming for Bioscientists: A Python-Based Primer. PLoS Comp Bio. DOI: 10.1371/journal.pcbi.1004867

Grading

1. Homeworks (5 in total) will be worth 50% of the grade

2. Team project - presentation, written report and code repository worth 50% of grade.

3. You are expected to attend class. If you need be absent do let me know. Material will be presented from the slides, but also in-class explanations.

4. Your work is expected to be your own. Google, StackExchange are useful tools to find solutions, and resuse is part of coding. However, you need to work on understanding the steps in your code so what you turn in should be a reflection of your own efforts.

Homework

1. Code should be runnable as turned in. You will deposit your code in your github repository. You can make one private personal repository to deposit and should organize a folder for each homework assignment (e.g. hw1, hw2, hw3).

2. Homework is due BEFORE class on the Friday it is due.

Projects

1. Topics to be selected from a set of choices or the team's choosing (with approval from instructor). Selection of topics will occur by end of October.

2. Project teams will be 1-2 individuals working together.

3. A presentation will be made by each team - last day(s) of class but may have to extend class time to accomodate this.

4. A final report with the details will be turned in by the group.

5. The report needs to detail what each person's contribution is to the project.