Course: Data Science Basics in Neuroscience

Course Advisor: Dr. Amy Kuceyeski | Instructor: Xihe Xie

Spring, 2021

E-mail: xix2007@med.cornell.edu
Web: **wcm-datascibasics.github.io**
Class Room: Virtual (Zoom)
Class Hours: Mondays 9:30-11:00am

Course Description

The aim of this course is to improve the working efficiency of graduate students on data analysis projects. There are numerous powerful tools available for scientists to automate manual tasks, and this essential topic is often left for the students to discover while tackling a newly collected dataset in the lab. By learning and taking advantage of these tools, students will become more efficient in handling data heavy projects, avoid problems such as data loss or inability to reproduce certain tasks.

Additionally, the NIH has begun requiring data management and sharing plans in all grants, and funded research must preserve and share their data in a FAIR (Findable, Accessible, Interoperable, and Reusable) manner. How do we integrate good analysis pipelines with datasets? What are the skills required and tools available to help you achieve these principles? At the end of this course, students will be able to answer all these questions, dissect any type of data, and take any new analysis workflow head-on.

Course Objectives

1. Become knowledgeable about the best practices and tools available for managing data science projects.
2. Be able to proficiently use the shell command line for research data wrangling needs.
3. Be able to manipulate computing environments based on the needs of research projects.
4. Be able to design, create, replicate, and publish fully reproducible research projects.
5. Become proficient with version control of data analysis projects.
6. Integrates FAIR data principles with version controlled analysis.

Prerequisites/Corequisites

While not required, it is preferred if students have prior experience with a shell command-line terminal before. Before the first day of class, student's computer needs to be set up as follows:

1. Administrative access to the computer for permissions to modify and install software.
2. For Windows laptop users, the 64 bit version of Windows 10 if required.
3. Follow this year's Brainhack OS-Specific Setup Guidelines for detailed setup instructions.

Course Structure

The class will consist of weekly one and a half hour lectures with demonstrations. The demonstrations will highlight the fundamental concepts, tools, and skills necessary to complete a short weekly assignment to put the lectured materials into practice. The final 15-30 minutes of class will be used to answer questions and address concerns. All course materials and assignments will be distributed via GitHub, and your assignments will be submitted and reviewed on GitHub repositories.

Assessments

Grading will be based upon completion of weekly assignments graded on a scale of 0-100%. The assignments will not require complex algorithms or materials not covered in class, so there is no need for prior programming experience to complete them. This class is designed for both experienced programmers looking to learn better practices and newcomers who have never touched a command-line before.

Course Policies

Academic Integrity and Honesty

Each student in this course is expected to abide by the Cornell University Code of Academic Integrity. Any work submitted by a student in this course for academic credit will be the student's own work.

Should copying occur, both the student who copied work from another student and the student who gave material to be copied will both automatically receive a zero for the assignment. Penalty for violation of this Code can also be extended to include failure of the course and University disciplinary action.

Students are encouraged to collaborate via Github for assignments and discussions of course materials. This can be through code reviews, issues, and comments recorded in Github repos. We will discuss details of collaborative projects during the version control and git class.

Accommodations for Disabilities

In compliance with the Cornell University policy and equal access laws, we are available to discuss appropriate academic accommodations that may be required for student with disabilities. Requests for academic accommodations are to be made during the first three weeks of the semester, except for unusual circumstances, so arrangements can be made. Students are encouraged to register with Student Disability Services to verify their eligibility for appropriate accommodations.

Inclusivity Statement

WCM is a diverse community of researchers devoted to science and medicine. Therefore, we are committed to creating an equitable environment where human diversity is welcomed and respected. While no list can hope to be comprehensive, we explicitly honor diversity in: age, culture, ethnicity, gender identity or expression, language, national origin, political beliefs, profession, race, religion, sexual orientation, and socioeconomic status.

Schedule and weekly learning goals

The schedule is tentative and subject to change. The learning goals below should be viewed as the key concepts which should be grasped after each week.

Week 01, 01/04 - 01/08: **Dr. David N. Kennedy (UMass)**
Course Overview & Introduction to Data Reproducibility in Sciences

Week 02, 01/11 - 01/15:
No Class - Virtual SfN Conference

Week 03, 01/18 - 01/22: **Dr. Hagen Tilgner**
What is the Shell?

Week 04, 01/25 - 01/29: **Dr. Hagen Tilgner**
The Shell Continued: Tools, and Environments

Week 05, 02/01 - 02/05: **Dr. Hagen Tilgner**
Version Control

Week 06, 02/08 - 02/12: **Dr. Hagen Tilgner**
Conda, Python, & Binder - How to start a reproducible analysis?

Week 07, 02/15 - 02/19: **Dr. Morgan Huse (MSK)**
Data Wrangling: Extract data from confocal images with Python

Week 08, 02/22 - 02/26: **Dr. Amy Kuceyeski**
Data Wrangling - Time series

Week 09, 03/01 - 03/05: **Dr. Ben Huang**
Data Wrangling - Calcium traces

Week 10, 03/08 - 03/12: **Dr. Hagen Tilgner & Dr. Georgia Frost (MSK)**
Data Wrangling - RNA-Seq

Week 11, 03/15 - 03/19: **Dr. Amy Kuceyeski**
Moving forward with Docker, Singularity, DataLad, & Beyond