

Integrating data science across undergraduate STEM curriculum

Kim Dill-McFarland, Steven Hallam, Dave Oliver

May 25, 2018

CTLT Spring Institute

Who we are



Kim Dill-McFarland, PhD
Postdoctoral teaching fellow
EDUCE
Microbiology & Immunology



Steven Hallam, PhD
Associate Professor
ECOSCOPE Director
Microbiology & Immunology

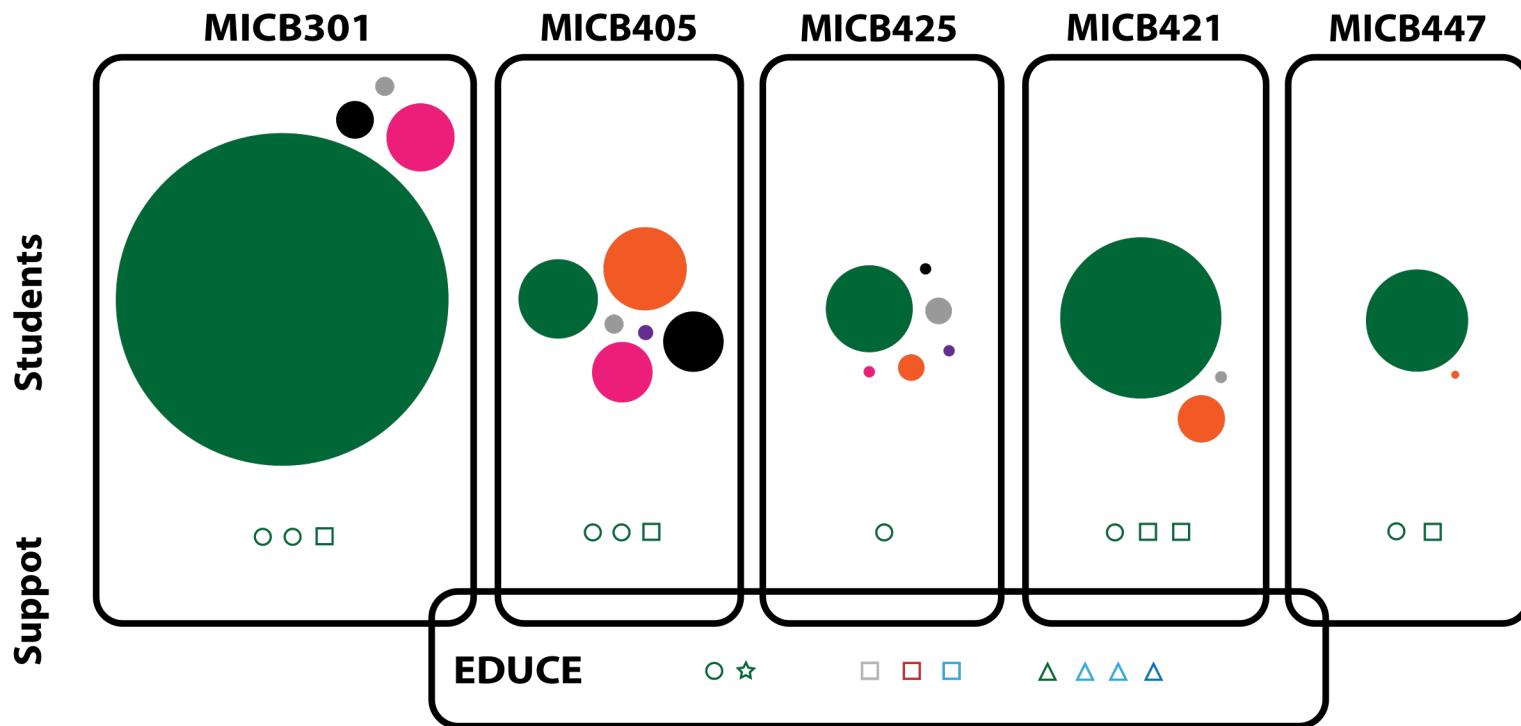


Dave Oliver, PhD
Instructor
Microbiology & Immunology



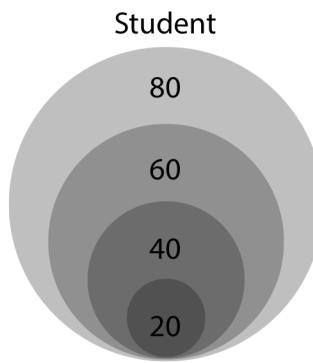
Experiential
Data science for
Undergraduate
Cross-disciplinary
Education

EDUCE



Faculty of Science
Microbiology & Immunology
Statistics
Mathematics
Computer Science
Biotechnology
Other

Faculty of Medicine
Medical Genetics
Other faculties
MS/PhD



- Faculty / Instructor
- ☆ TLF
- TA
- △ Consultant/other

Objectives

- Appreciate why students across many faculties should learn data science
- Describe a modular method for incorporating data science into existing courses
- Identify areas in your course(s) where data science should and can be effectively incorporated

What is data science?

Take 5 min to write down a definition or key words

Data science

- **Interdisciplinary field** of scientific methods, processes, algorithms and systems to extract knowledge or insights from **data in various forms**, either structured or unstructured, similar to data mining
- Includes big data, statistics, mathematics, and computer science

Applications in STEM

- Next-generation DNA, RNA, protein, metabolite analyses
- Global systems monitoring
- Personalized medicine

Applications in your daily life

- E-commerce and advertising
- City planning and traffic patterns
- Netflix, Hulu, etc.
- FitBit, Garmin, etc.

Importance of teaching data science

- Data is everywhere
- Skills in one data science application are highly transferable to other areas
- Critical thinking, complex questions, and no “right” answer
- Communication and collaboration across disciplines

Keys to data science curriculum success

- Hands-on, experiential learning
- Connections to domain knowledge
- Open and accessible
- Collaborative experiences
- Foundations for independent learning

As a student:

Introduction to
command line module

Module outline

- Day 1: Lecture + discussion – Why data science?
- **Day 2: Live demo + group work to practice using command line**
- Day 3: Utilize command line to run an analysis related to the course content
 - BLAST sequence aligner
- Day 4: Discuss biological meaning of the results

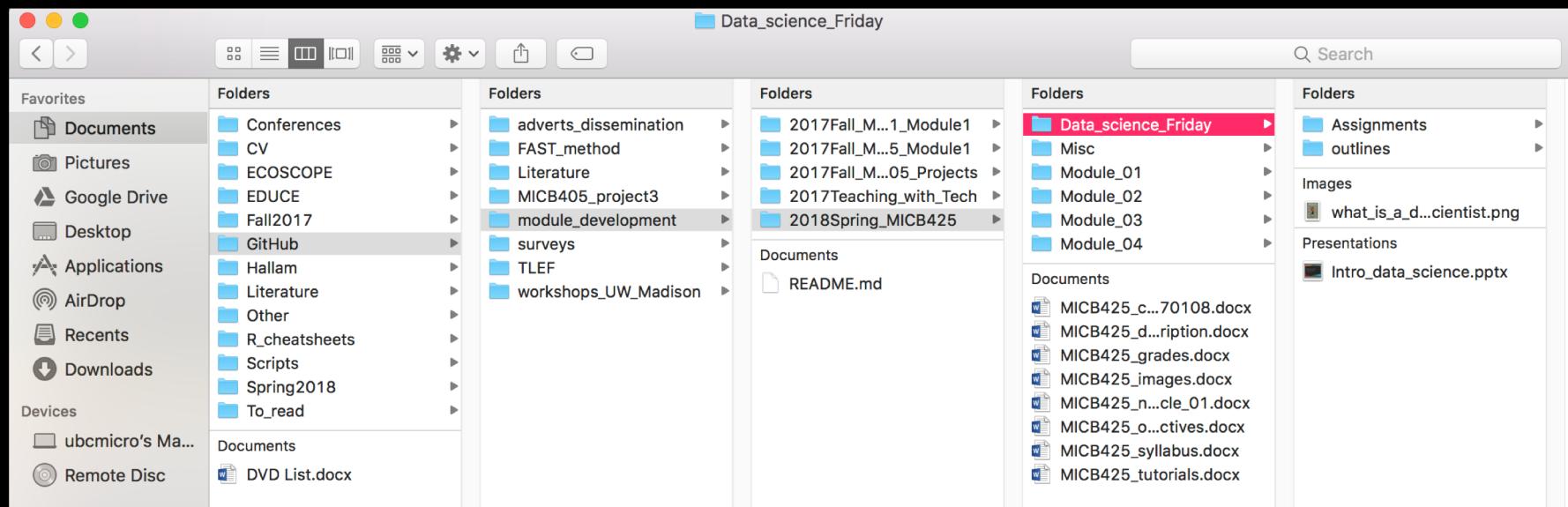
Objectives

- Identify uses of command line inside and outside the classroom
- Apply basic syntax and functions in command line
- Use command line to navigate your computer's file structure

Command line terminal

A screenshot of a macOS terminal window titled "kim — -bash — 47x5". The window shows the following text:

```
Last login: Mon Oct 23 13:48:36 on ttys000
dhcp-206-87-155-44:~ kim$
```



Command line can...

- Do everything you do in Finder/File explorer
 - Make, move, copy, delete, rename files and folders
- Run specific programs
 - Text editors, R, sequence analysis tools
- Connect to and transfer files from remote servers like GitHub
- And more...

Why command line?

- Yes, everything we are doing in this class in command line could be done using the GUI (point and click).
- But...

Why command line?

- There are processes that are easier, faster, or can only be accomplished using the command line
 - Super computing resources can only be accessed via command line
- Command line allows you to provide reproducible code for your future self as well as supervisors and co-workers



Past you is your closest, most consistent collaborator – and past you doesn't answer email.

~Karl Broman, UW-Madison

Why command line?

- Commands are the same across different OS types and versions
- Command line knowledge makes you more competitive not only for careers in biology but also in other STEM areas, finance, etc...



Hands-on activity:

Follow along on your own
computer

As an *instructor*:

Assess the module



What did you like? Dislike?

Take 10 min to discuss with your table.

Break

As an *instructor*:

Create a module

Foundational tools

Command line

- Syntax applicable in other programs
- Start a conversation on good computer habits (file names, organization, etc.)
- Use an outside computing source like Amazon Web Services (AWS)
- Feel like a data scientist

Foundational tools

Git/GitHub

- Utilizes command line
- Version control
- Collaboration

Foundational tools

R/RStudio (python/Jupyter)

- Anything done in Excel
- Statistics
- Data manipulation
- Data visualization

Your own classroom

- Where could data science be incorporated into your own course?
- Is there big data in your field? Could it be leveraged in your classroom?

Brainstorm & outline a potential module (20 min)
Then share with a partner and provide feedback (15 min)

Data science at UBC

- Experiential Data science for Undergraduate Cross-disciplinary Education (EDUCE)
<http://ecoscope.ubc.ca/program-structure/educe/>
<https://github.com/orgs/EDUCE-UBC/dashboard>
- R workshops by ECOSCOPE
<http://ecoscope.ubc.ca/workshop-series/>
- Masters in Data Science
<https://masterdatascience.science.ubc.ca/>
- Skylight, CTLT, TLEFs <https://tlef.ubc.ca/events/>

Other data science resources

- Swirl <http://swirlstats.com/>
- Codecademy
<https://www.codecademy.com/catalog/subject/all>
- edX <https://www.edx.org/>
- The Carpentries (data and software)
<https://carpentries.org/>
- Riffomonas <http://www.riffomonas.org/>
- Compute Canada Bioinformatics Helpdesk
<https://bioinformatics.computecanada.ca/>