Data science curriculum across M&I

2018 Faculty retreat

Kim Dill-McFarland, Steven Hallam
version September 22, 2018

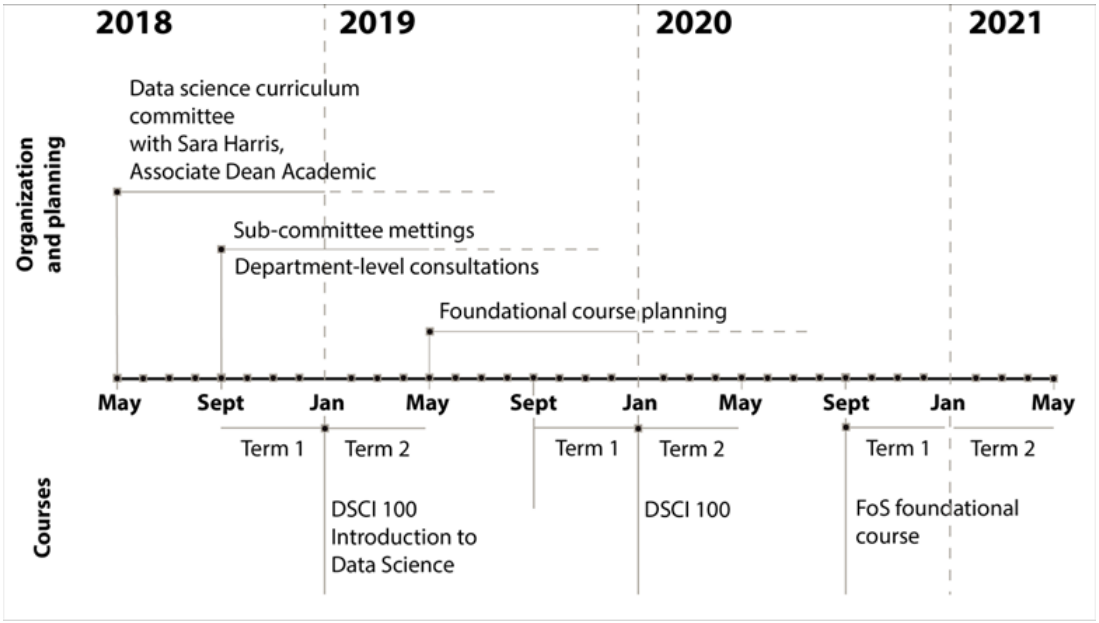
Outline

- Faculty of Science data science initiative
- Current data science in MICB courses
- Future directions in M&I

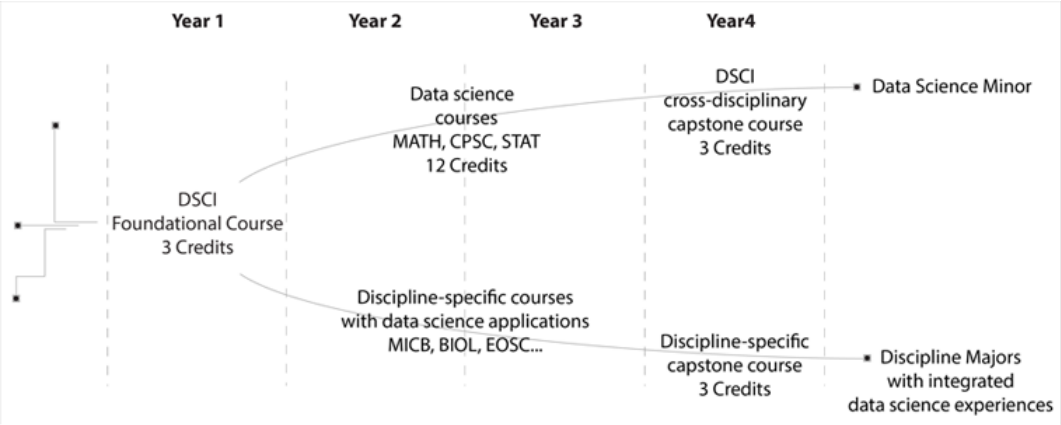
Faculty of Science data science initiative

1. Evaluate existing models of undergraduate data science education
2. Develop a unified and strategic vision for data science education within the Faculty of Science (FoS)
3. Determine curriculum structure needed to provide undergraduates with core data science competencies
4. Establish an inter-disciplinary community of practice to implement this curriculum in coordination with FoS

FoS initiative timeline



FoS concept map



Data science in M&I

Bioinformatics

- **2013-17**: MICB 405 over-subscribed; course size incrementally increased from 47 students
- **2018**: MICB 405 now at 80 students, still over-subscribed

Data science

- **2016**: MICB curriculum review reveals limited data science; need for content related to microbiome analyses
- **2017/18**: EDUCE (Experiential Data science for Undergraduate Cross-disciplinary Education) in 3 MICB courses
- **2018/19**: EDUCE in 7 MICB courses

EDUCE

- Modular integration of data science curriculum into existing 3rd and 4th year courses
- Collaboration of MICB and STAT with support from TLEF, MICB, ECOSCOPE

Teaching data science through

- Connections to domain knowledge and questions
- Hands-on practice
- Accessible and open-source
- Connections to co-curriculars

The EDUCE team

- Steven Hallam
- Kim Dill-McFarland

Course instructors

- Lindsay Eltis
- Jennifer Gardy
- Marcia Graves
- Martin Hirst
- Bill Mohn
- Dave Oliver
- Jen Sibley

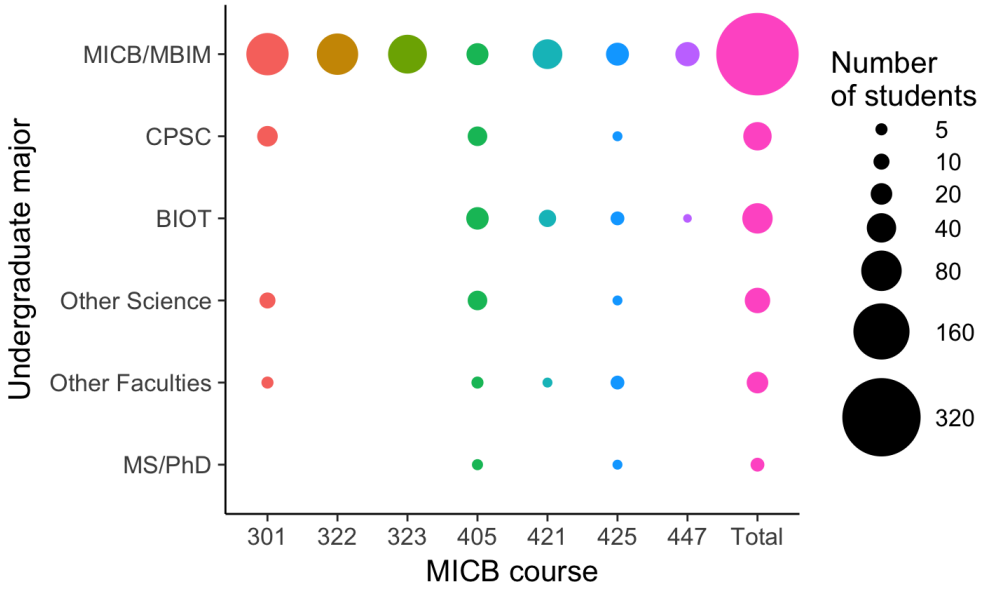
TAs

- Yue Liu (App MATH)
- Nolan Shelley (Botany)
- David Yin (CPSC, STAT)

Other partners

- Jennifer Bonderoff (ECOSCOPE)
- Gaby Cohen-Freue (STAT)
- Patrick Walls (MATH)
- Biljana Stojkova (STAT, ASDa)
- Carolyn Taylor (STAT, ASDa)

Current EDUCE impact



Data science modules

Introduction to data science
Align expectations and frame within the discipline

Command line
Unix functions and applications in BLAST, remote computing, version control, microbiome sequence data analysis

R/RStudio
Import, manipulation, visualization, and statistics of various data types

Statistics
Terminology and underlying hypotheses, assumptions, uses, and limitations of t-test, ANOVA, linear regression, etc.

See handout for more details.

Co-curriculars

- Reinforce and build on course modules

ECOSCOPE workshops

- R/RStudio
- Git/GitHub
- Software and Data Carpentry
- Compute Canada

Social learning opportunities

- Hackathons
- Open office hours

Future directions

- Identify areas for data science integration in immunology courses
- Resolve MICB 405 (Bioinformatics) over-subscription
- Pursue a sustainable model for EDUCE and/or other data science curriculum in M&I
- Integrate data science into the M&I strategic plan?