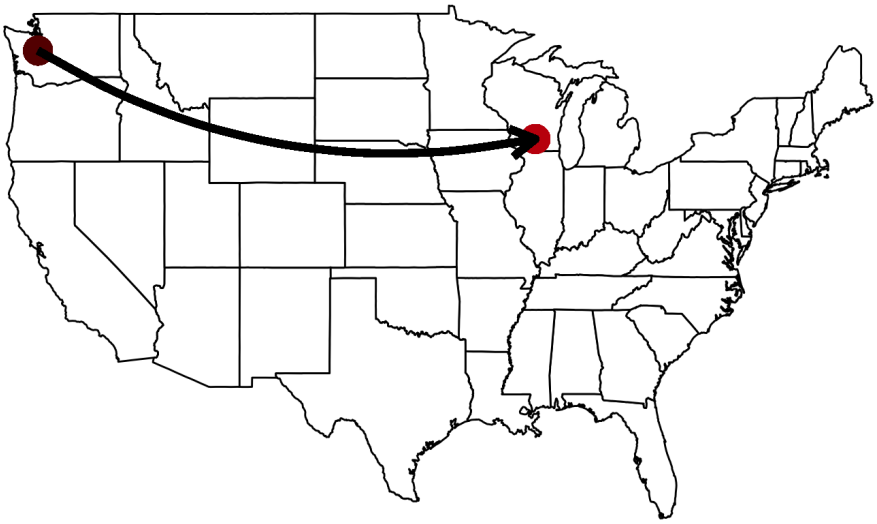
Improving data science education

Kim Dill-McFarland, PhD
April 16, 2019

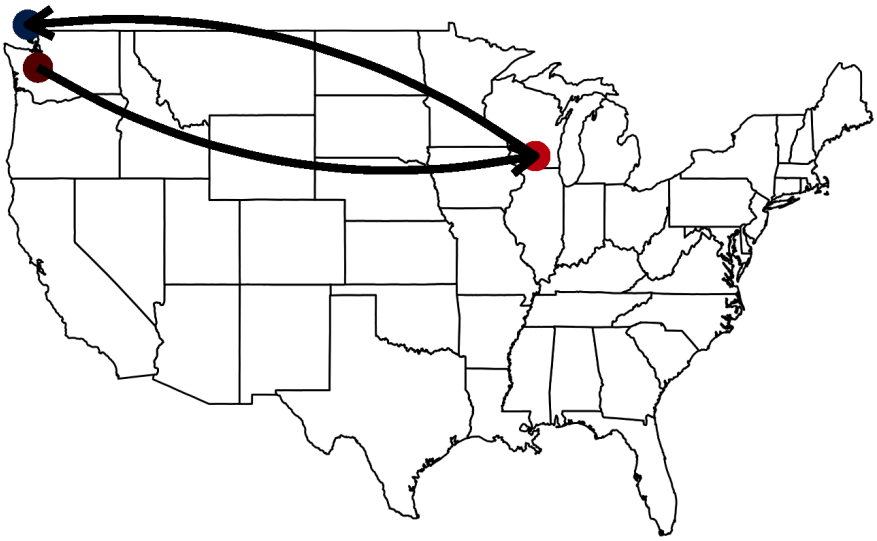
Background

Background

Background



Background



Complex microbial communities

Charis Tsevis

> 90% of researchers in the biological sciences work with or plan to work with big data

(Williams & Teal 2017)

> 60% of researchers in the biological sciences report a need for more training in data science

Meta-analysis 2013 - 2016
(Attwood *et al* 2017)

We need to teach data science in undergraduate life science curriculum.

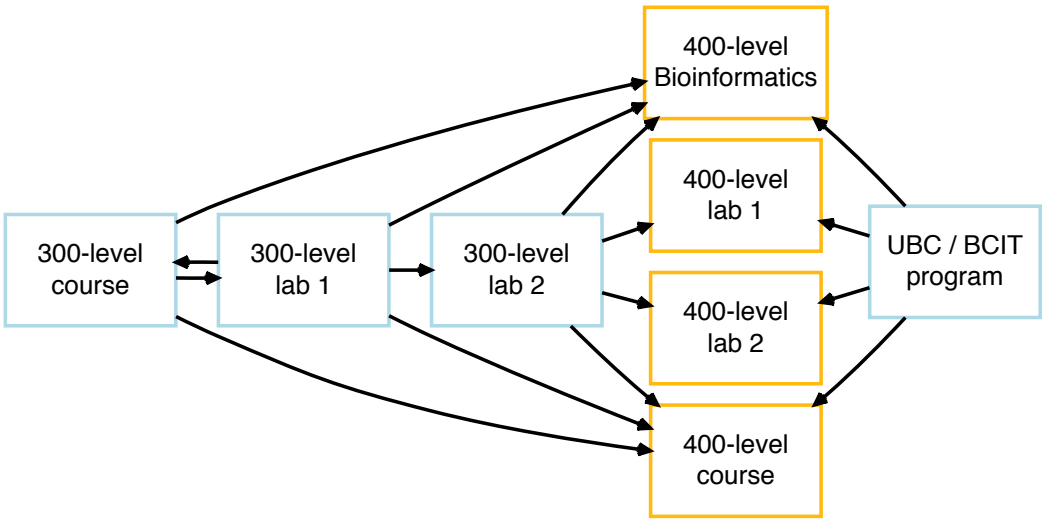
Experiential
Data science for
Undergraduate
Cross-disciplinary
Education

Slides at bit.ly/FredHutch_Apr16

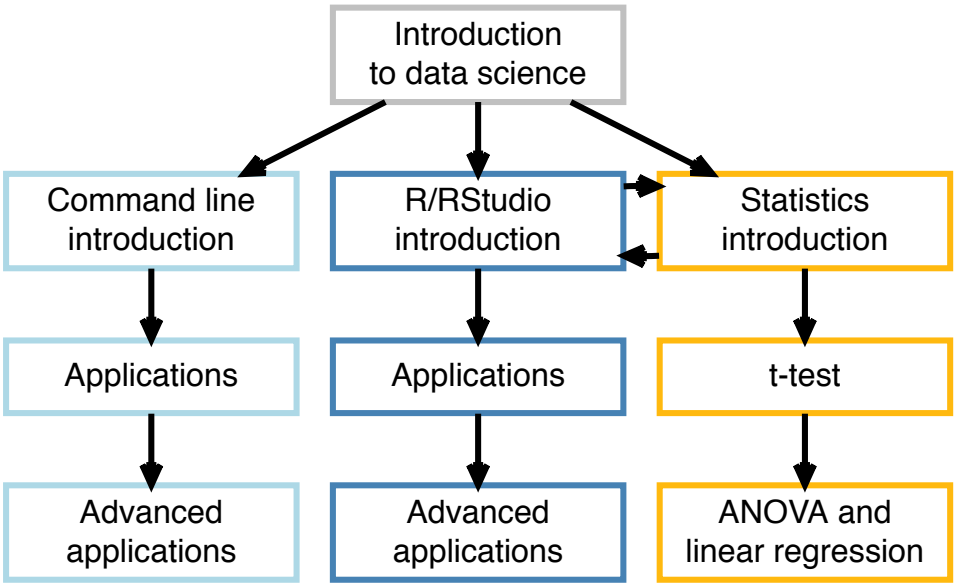
Our goal

Modular integration of
data science curriculum into
existing courses

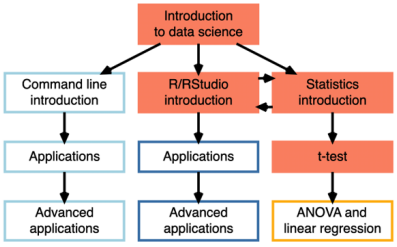
Course overview



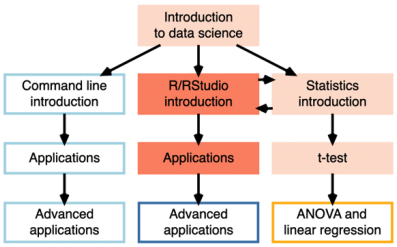
Content overview



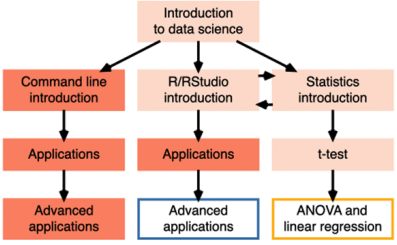
Example student



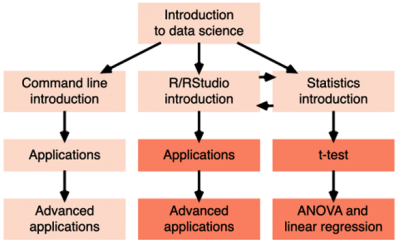
300-level course



300-level lab

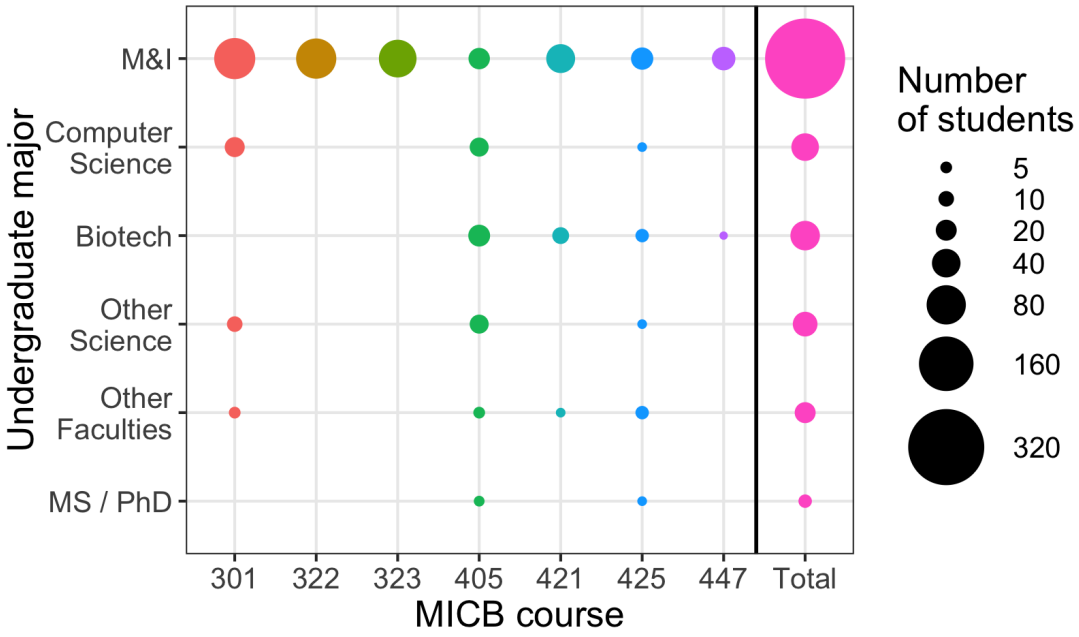


400-level course



400-level lab

Students impacted per year



Does EDUCE effectively teach data science skills to M&I students?

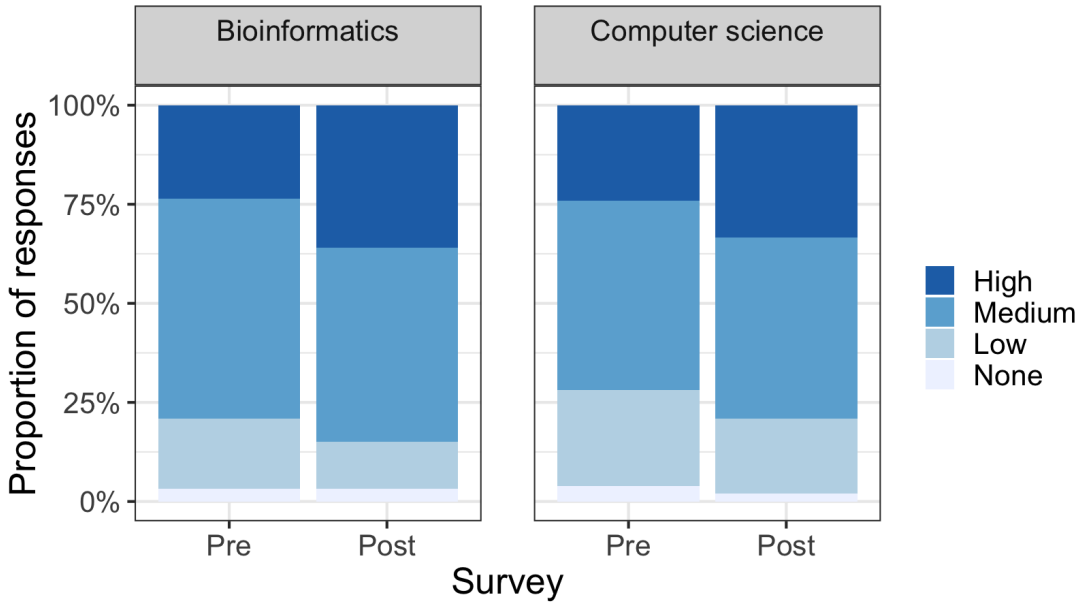
EDUCE in MICB 301

- 300-level course, no lab
- ~120 students / yr

- 5 x 50 min class sessions across 5 weeks
- Introduction to
  - data science
  - command line
  - R/RStudio
  - statistics
- BLAST, simple plots, and running a *t*-test in R

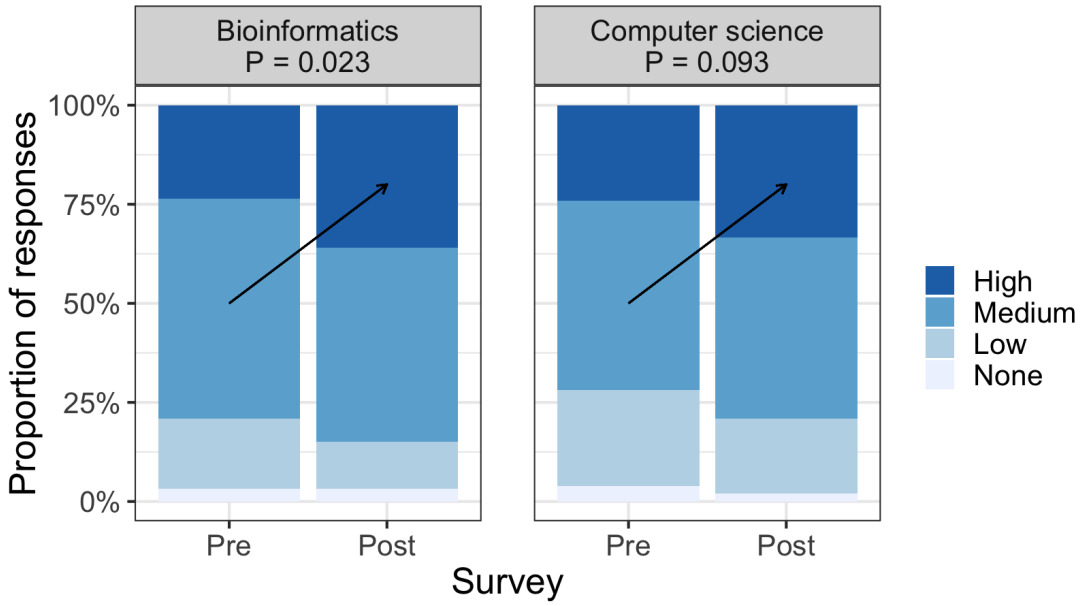
Increased interest in data science

How would you rate your interest in…



Increased interest in data science

How would you rate your interest in…

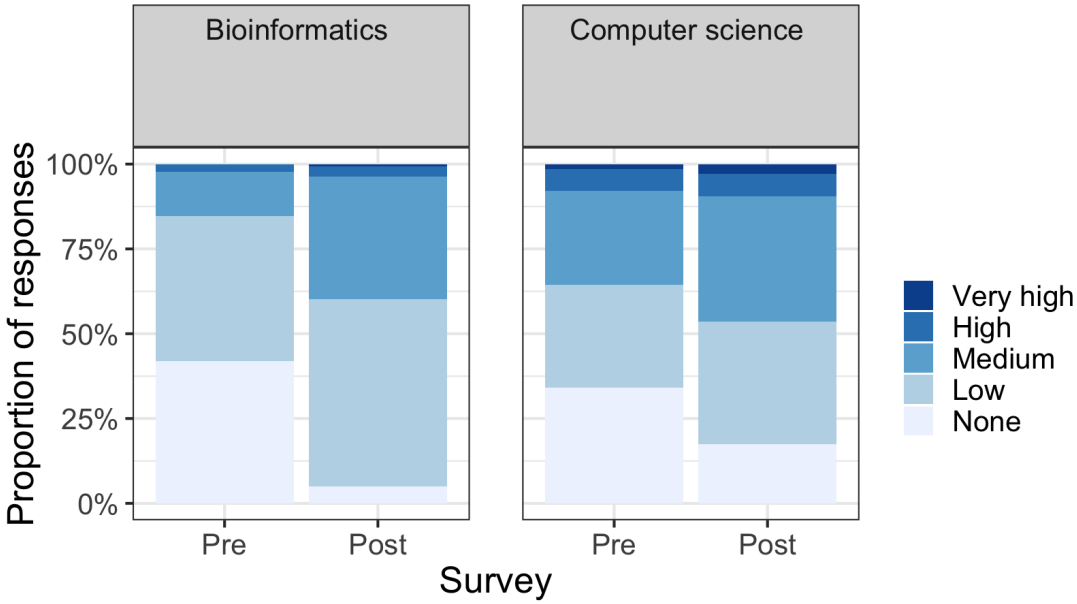


Except…

No significant changes in interest in statistics

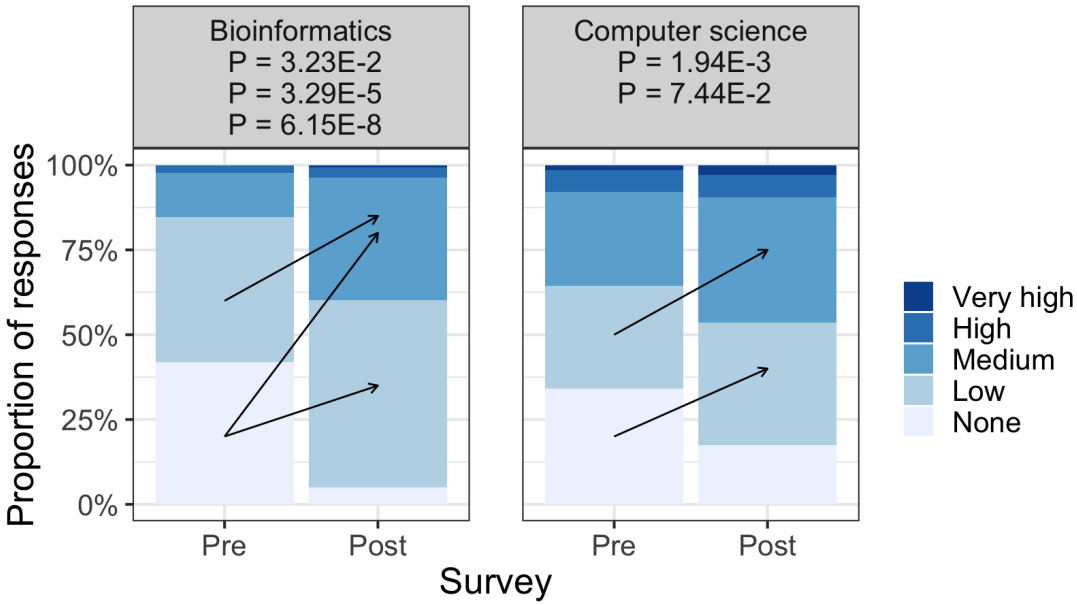
Increased experience in data science

What level of experience do you have in …



Increased experience in data science

What level of experience do you have in …



Except…

No significant changes in experience in statistics

EDUCE in MICB 301

- 300-level course, no lab
- ~120 students / yr

- 5 x 50 min class sessions across 5 weeks
- Introduction to
  - data science
  - ~~command line~~
  - R/RStudio
  - **statistics**
- ~~BLAST,~~ simple plots and running a *t*-test in R

Conclusions

- Data science literacy is needed in the life sciences
- EDUCE provides a flexible, modular approach for integrating data science into undergraduate curriculum
- Even minimal exposure (5 hours) can increase student self-reported knowledge, interest, and experience in data science areas

The future

- A wealth of survey data
- More courses? Other departments?
- Independent learning tools https://ubc-educe.shinyapps.io/course_knitter/
- Other open-source resources https://github.com/EDUCE-UBC

Community of practice

Steven Hallam
Jennifer Bonderoff

EDUCE TAs

Yue Liu (App MATH)
Julia Beni (U. Minnesota)
Kris Hong (CPSC, STAT)
Jonah Lin (MICB, CPSC)
Lisa McEwen (MedGen)
Ryan McLaughlin (BINFO)
Connor Morgan-Lang (BINFO)
Nolan Shelley (Botany)
David Yin (CPSC, STAT)

Course instructors

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

Collaborators

Gaby Cohen-Freue (STAT)
Patrick Walls (MATH)
Biljana Stojkova (ASDa)

Funding

UBC Teaching and Learning Enhancement Fund (TLEF)

NSERC CREATE Program (ECOSCOPE)

Department of Microbiology & Immunology

UBC Skylight and the Center for Teaching, Learning and Technology (CTLT)

References

Attwood TK *et al* 2017. *A global perspective on evolving bioinformatics and data science training needs.* Brief Bioinform. 20(2):398-404. doi: 10.1093/bib/bbx100

Williams JJ & Teal TK. 2017. *A vision for collaborative training infrastructure for bioinformatics.* Ann N Y Acad Sci. 1387(1):54-60_ doi: 10.1111/nyas.13207