

Incorporating data science into undergraduate microbiology

Kim Dill-McFarland

April 2, 2019

Data science

- extracting knowledge and meaning from (big) data
- statistics, mathematics, computer science
- Where do the data come from?



(James Montgomery Flagg)

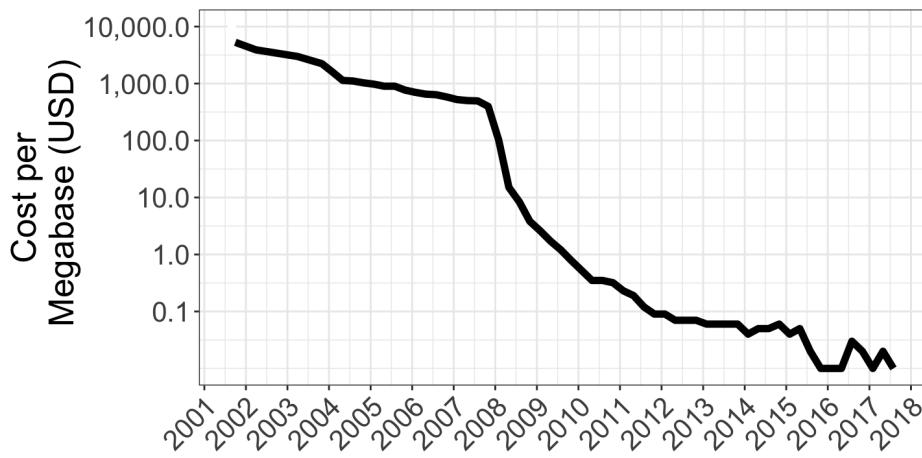
Slides at bit.ly/M&I_EDUCE 3/44

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

(Williams & Teal 2017)

Slides at bit.ly/M&I_EDUCE 4/44

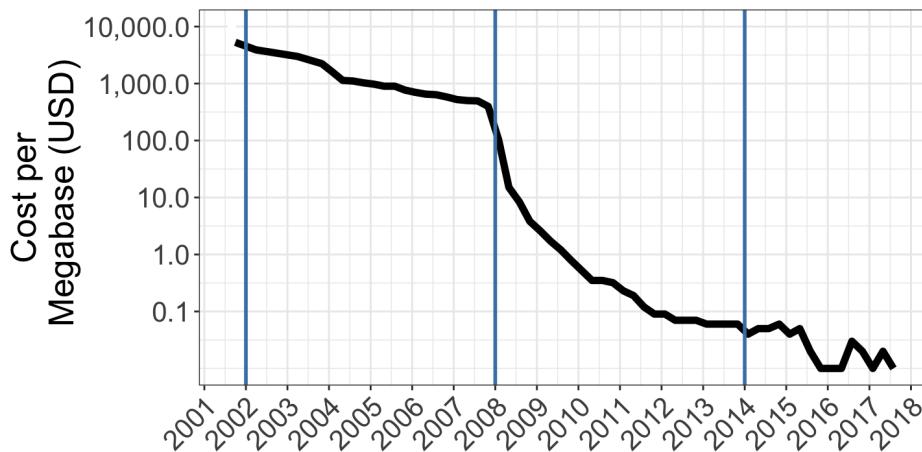
Next-generation sequencing



(NIH National Human Genome Research Institute)

Slides at bit.ly/M&I_EDUCE 5/44

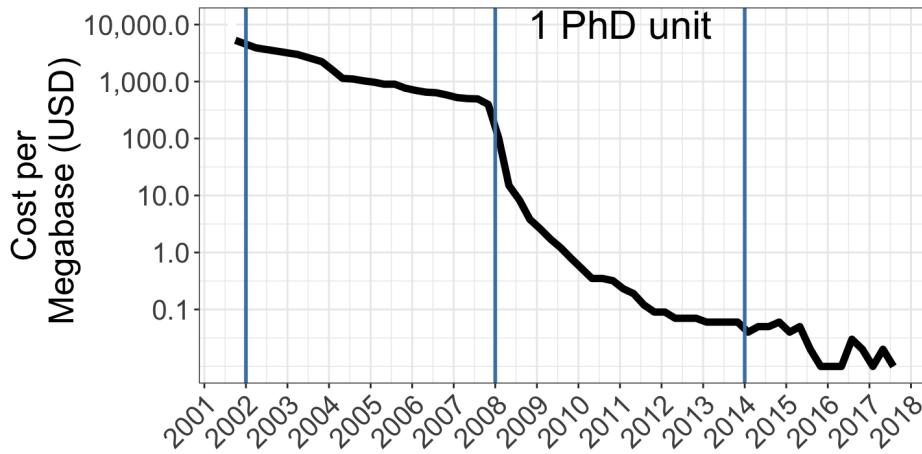
Next-generation sequencing



(NIH National Human Genome Research Institute)

Slides at bit.ly/M&I_EDUCE 6/44

Next-generation sequencing



(NIH National Human Genome Research Institute)
Slides at bit.ly/M&I_EDUCE

7/44

✓ 60% of researchers in the biological
✓ 80% of researchers in the biological
sciences report a need for more training
in data science

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

Slides at bit.ly/M&I_EDUCE 8/44

Not just academia

Bioinformatics Market to hit more than US\$ 16 Billion By 2022

By

Published: Sept 26, 2018 11:06 a.m. ET

Careers in Bioinformatics: Hot and Getting Hotter

Published: Mar 13, 2019 | By Mark Terry

Slides at bit.ly/M&I_EDUCE

9/44

7,033 views | Jan 23, 2019, 12:10am

Data Scientist Leads 50 Best Jobs In America For 2019 According To Glassdoor



Louis Columbus Contributor

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

Barriers to data science integration

1. Faculty training
2. Student interest
3. Student preparation in mathematics, statistics, and computer science
4. Already overly full curricula

5. Limited access to resources (hardware, software)

(Williams *et al* 2017)

Slides at bit.ly/M&I_EDUCE 11/44

Experiential
Data science for
Undergraduate



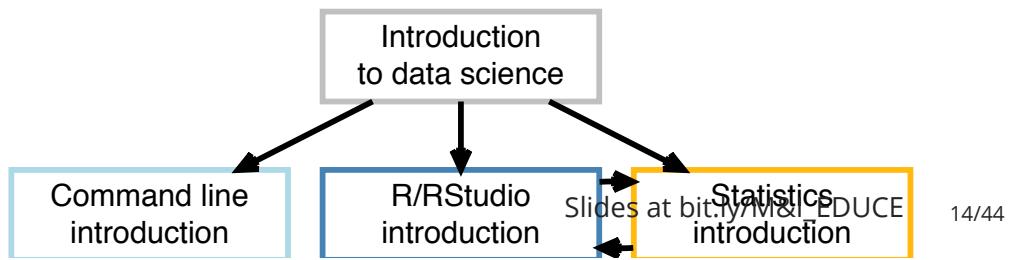
Cross-disciplinary
Education

Slides at bit.ly/M&I_EDUCE 12/44

Our goal

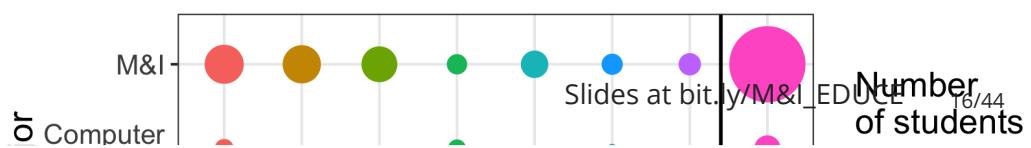
Modular integration of
data science curriculum into
existing courses

Content overview



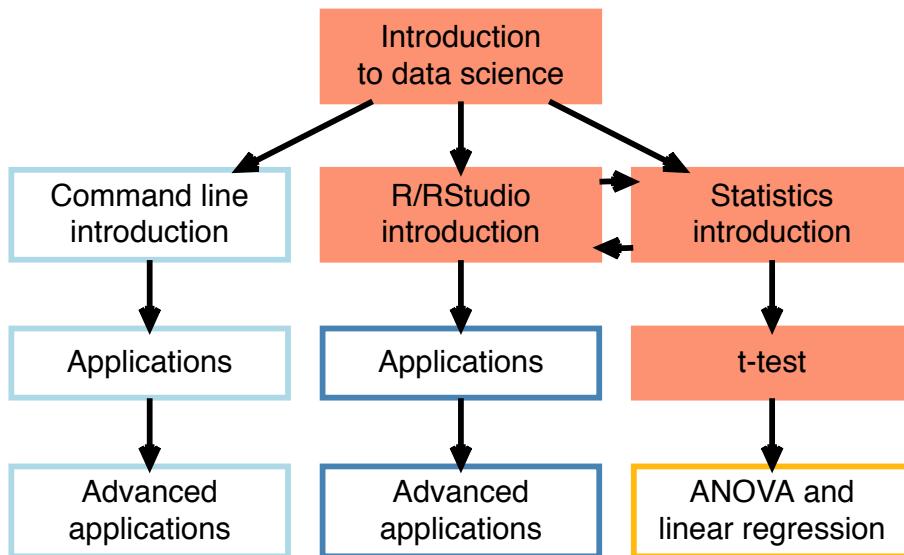
Course overview

Students impacted per year

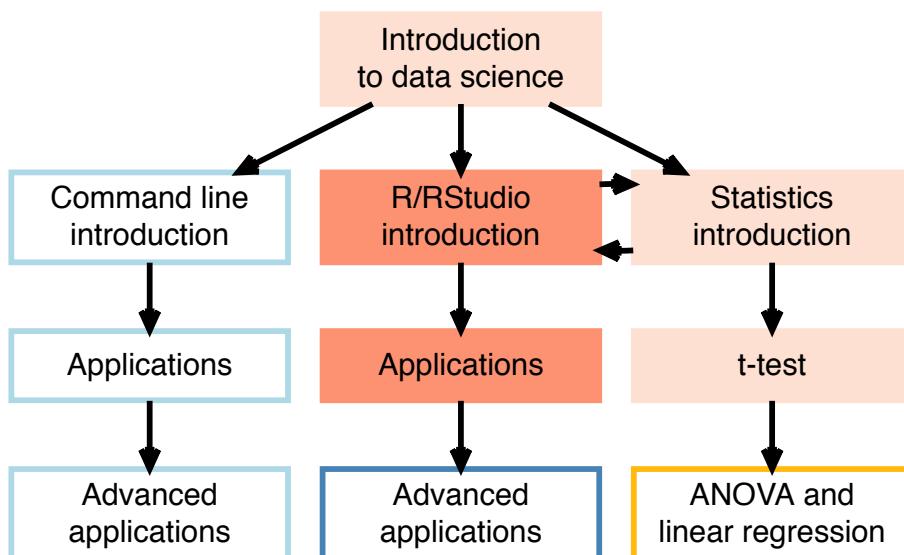


Example student

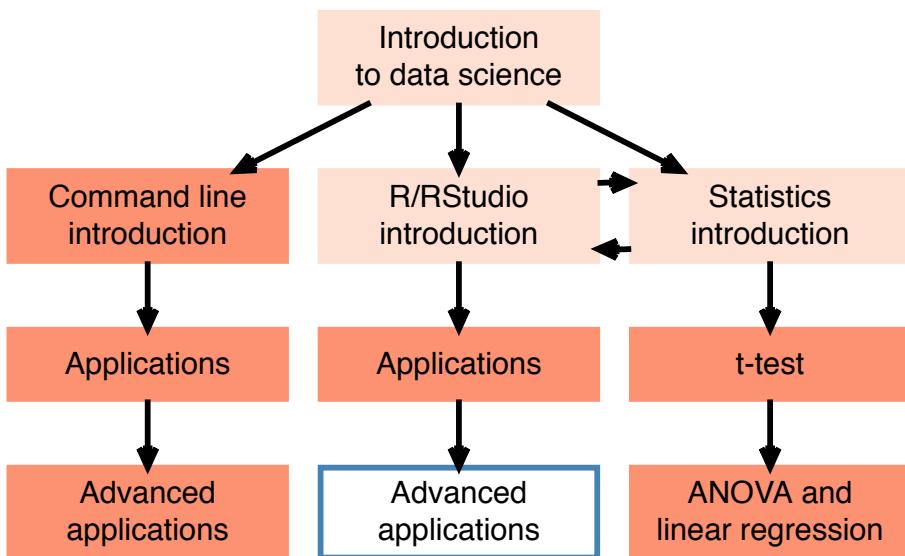
MICB 301



MICB 301 - 322

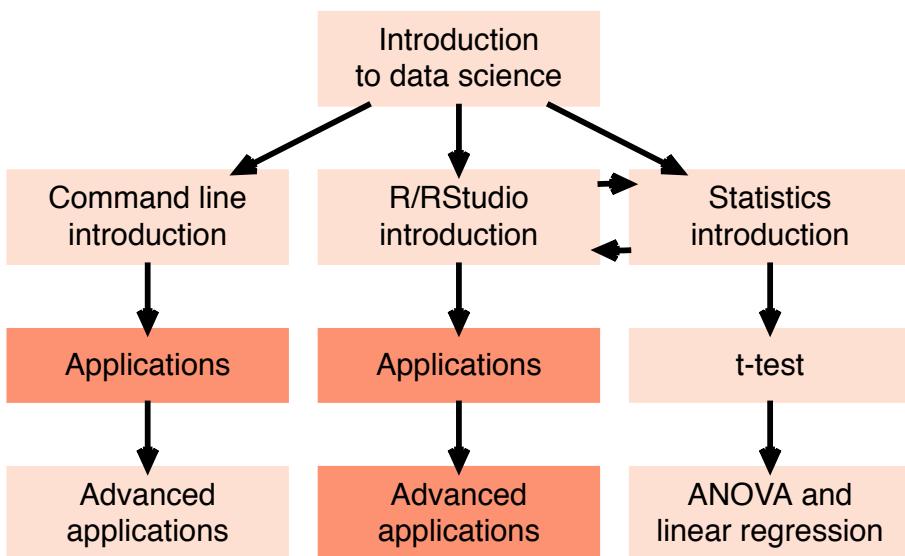


MICB 301 - 322 - 405



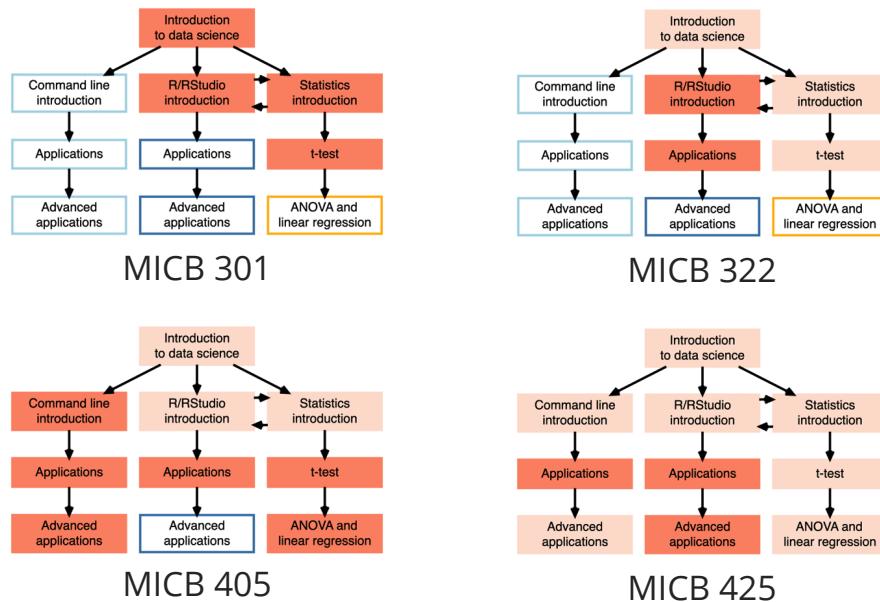
Slides at bit.ly/M&I_EDUCE 20/44

MICB 301 - 322 - 405 - 425



Slides at bit.ly/M&I_EDUCE 21/44

Example student



Slides at bit.ly/M&I_EDUCE 22/44

Solutions to integration

1. Faculty training

- *Dedicated Postdoctoral Teaching and Learning Fellow*
- *Cross-disciplinary TAs from multiple departments*

2. Student interest

- *Direct connections to other course curricula*
- *Hands-on, experiential learning*

Slides at bit.ly/M&I_EDUCE 23/44

Solutions to integration

3. Student preparation

- *No prior knowledge assumed*

4. Already overly full curricula

- *No new courses required*

5. Limited access to resources

- *Stripped down datasets and use of cloud resources*
- *Open-source tools and curricula*

Slides at bit.ly/M&I_EDUCE 24/44

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

MICB 301 as a case study

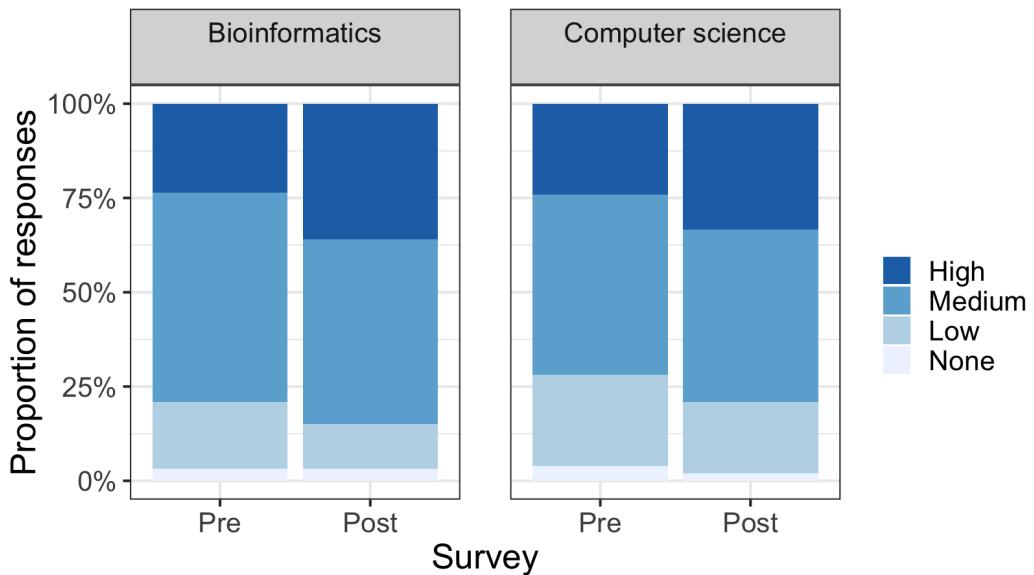
EDUCE in MICB 301

- 5 x 50 min class sessions across 5 weeks
- Weekly assignments and a final report
- Introduction to
 - data science
 - R/RStudio
 - statistics
- Simple plots and running a *t*-test in R

Slides at bit.ly/M&I_EDUCE 27/44

Increased interest in data science

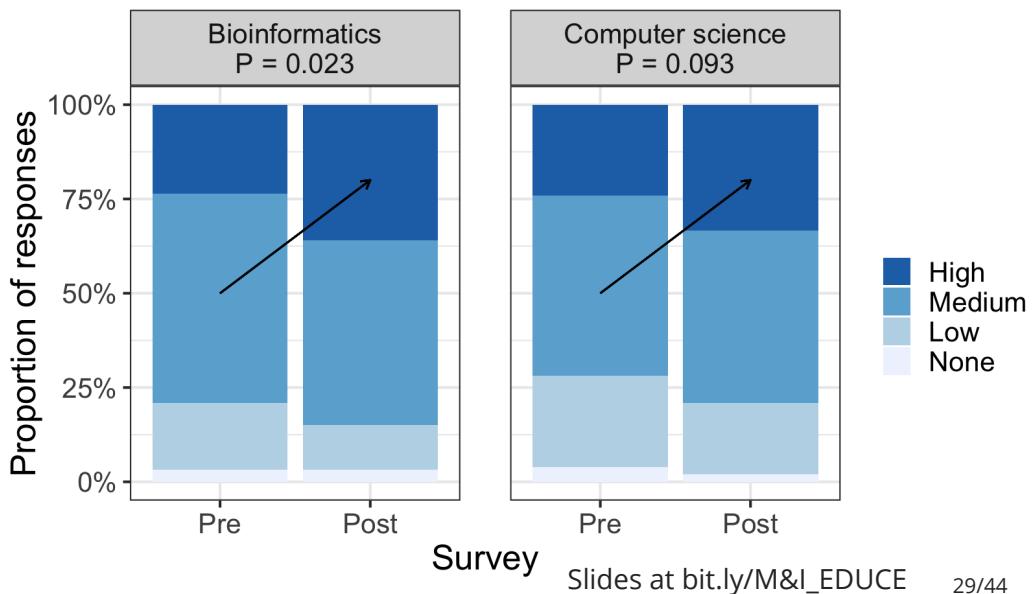
How would you rate your interest in...



Slides at bit.ly/M&I_EDUCE 28/44

Increased interest in data science

How would you rate your interest in...



Slides at bit.ly/M&I_EDUCE 29/44

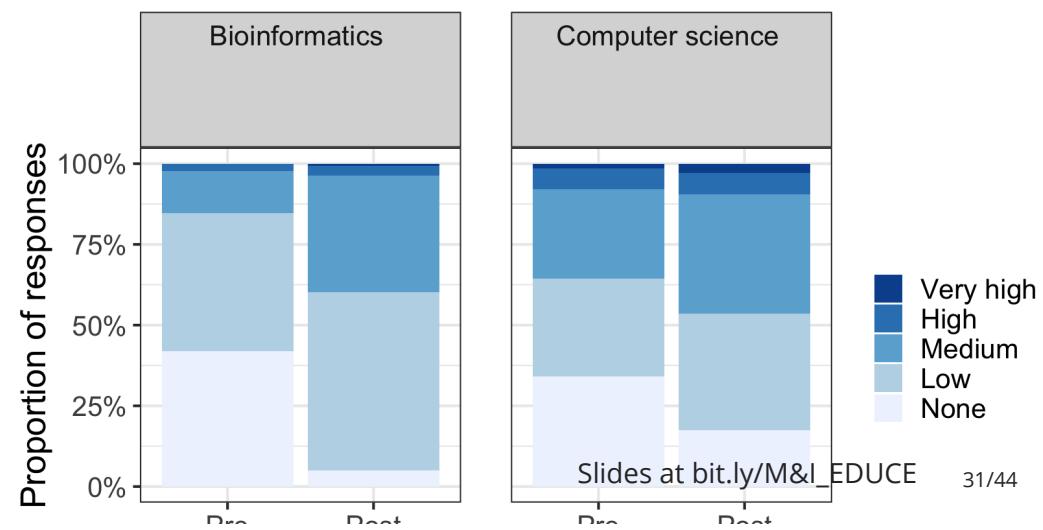
Except...

No significant changes in interest in statistics

Slides at bit.ly/M&I_EDUCE 30/44

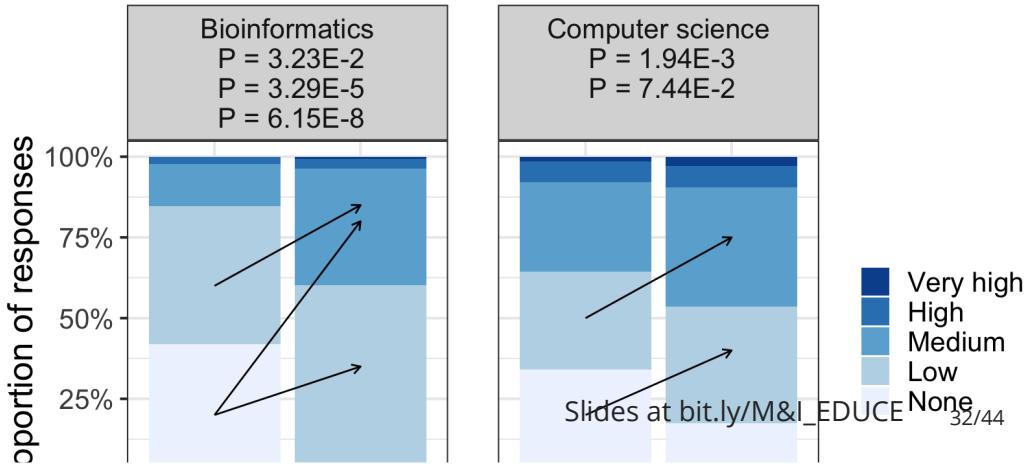
Increased experience in data science

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
bioinformatics

No significant changes in experience in
statistics

Slides at bit.ly/M&I_EDUCE 33/44

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 interest and experience in data science areas

The future

- A wealth of survey data to mine
- Repetition across 3 years for statistical analyses
- More courses? Other departments?
- Faculty of Science Data Science Committee

Acknowledgements

**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)

Course instructors

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

Connor Morgan-Lang (BINFO)
Nolan Shelley (Botany)
David Yin (CPSC, STAT)

Collaborators

Gaby Cohen-Freue (STAT)
Patrick Walls (MATH)
Biljana Stojkova (ASDA)
Slides at bit.ly/M&I_EDUCE

36/44

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)

Slides at bit.ly/M&I_EDUCE 37/44

Opportunities at UBC

Postdoctoral Teaching and Learning Fellow

- EDUCE
<http://ecoscope.ubc.ca/program-structure/educe/educe-postdoctoral-application>

- Master of Data Science
<https://www.stat.ubc.ca/postdoctoral-teaching-and-learning-fellow-ubc-master-data-science-program-0>

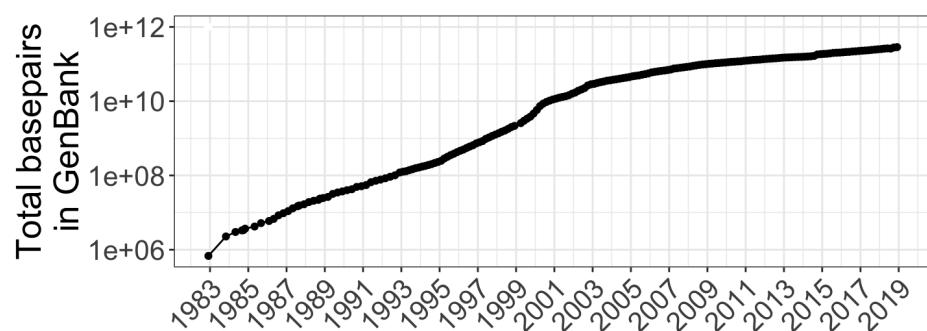
References

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](https://doi.org/10.1093/bib/bbx100)

Williams JJ *et al* 2017. *Barriers to integration of bioinformatics into undergraduate life sciences education*. BioRxiv. doi: [10.1101/204420](https://doi.org/10.1101/204420)

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](https://doi.org/10.1111/nyas.13207)

GenBank sequences



Undergraduate programs

BSc in Bioinformatics

- U. of Montreal
- U. Saskatchewan
- U. Calgary
- Carleton U.

Joint BSc degrees

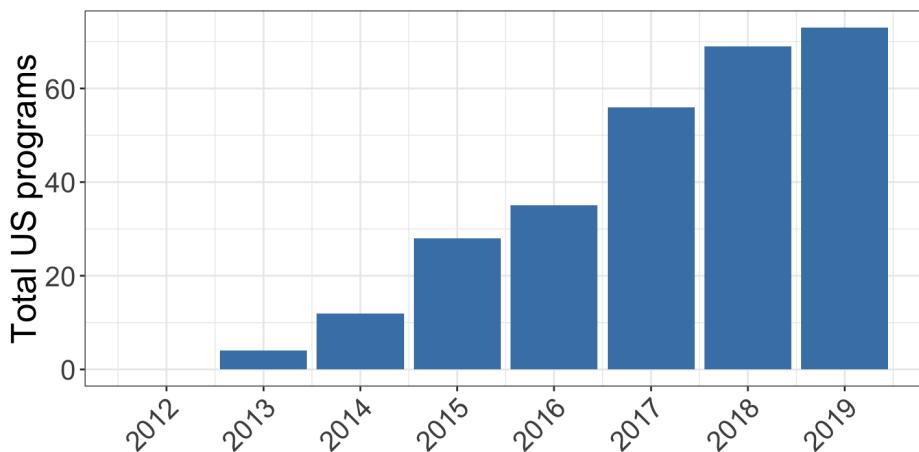
- Simon Fraser U.
- U. of British Columbia

Specializations / minors

- Dalhousie U.
- McGill U.
- U. of Toronto
- U. of Victoria
- U. of Waterloo
- U. of Western Ontario

Slides at bit.ly/M&I_EDUCE 41/44

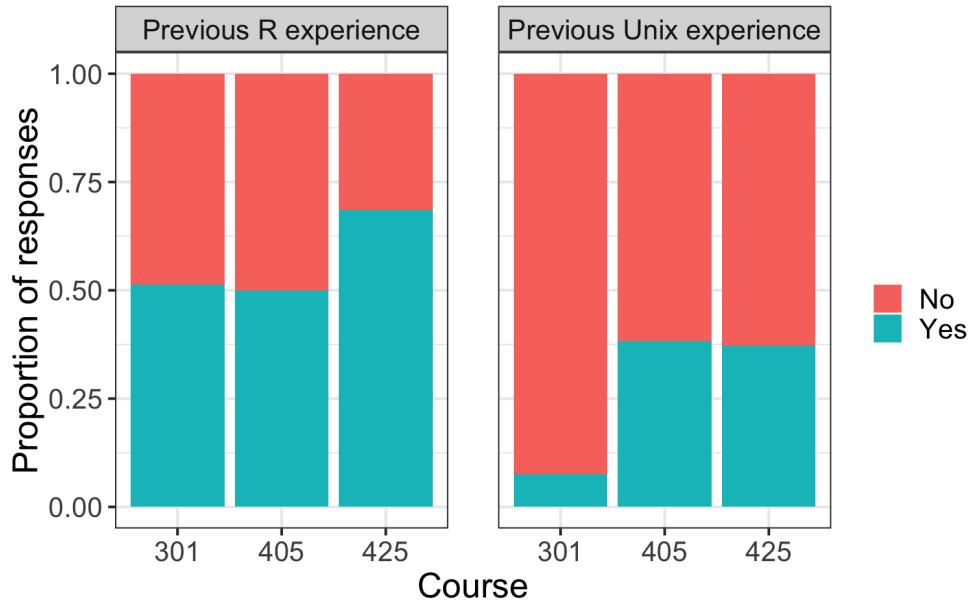
MDS programs



(Michael Rappa, NC State University)

Slides at bit.ly/M&I_EDUCE 42/44

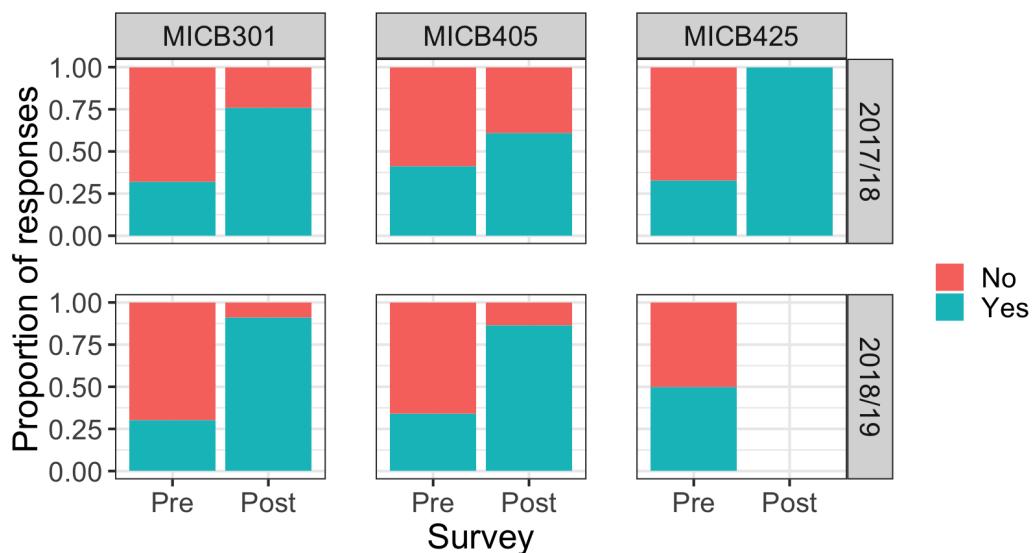
Some prior experience



Slides at bit.ly/M&I_EDUCE 43/44

Minimal prior knowledge

Have you heard the term 'data science'?



Slides at bit.ly/M&I_EDUCE 44/44