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)

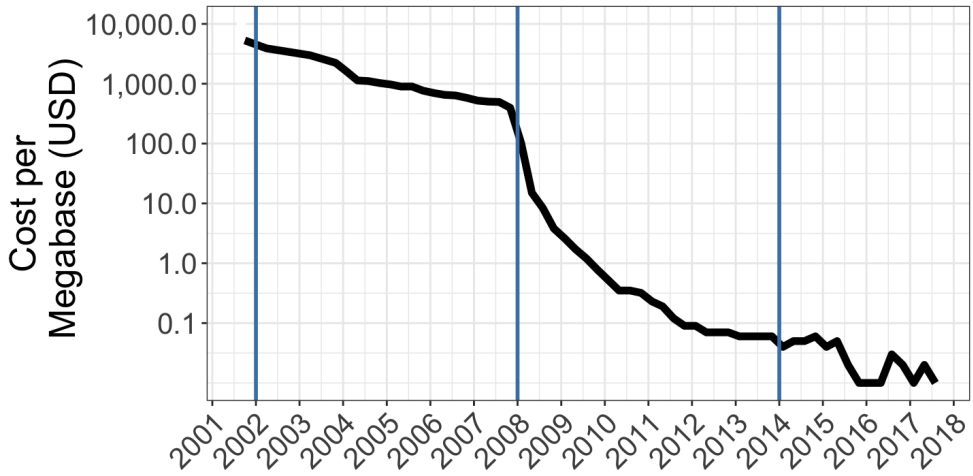
> 90% of researchers in the biological sciences work with or plan to work with big data

(Williams & Teal 2017)

Next-generation sequencing

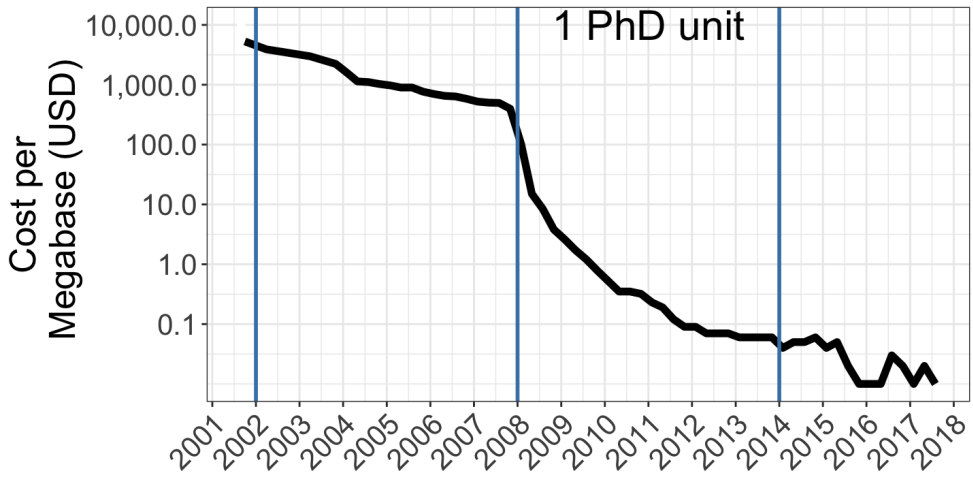
(NIH National Human Genome Research Institute)

Next-generation sequencing



(NIH National Human Genome Research Institute)

Next-generation sequencing

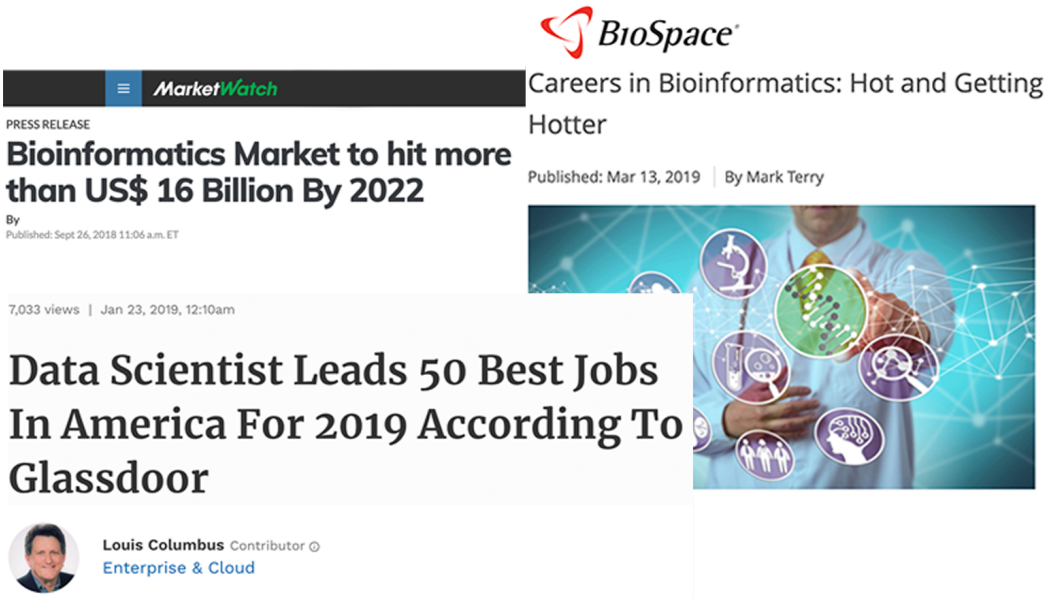


(NIH National Human Genome Research Institute)

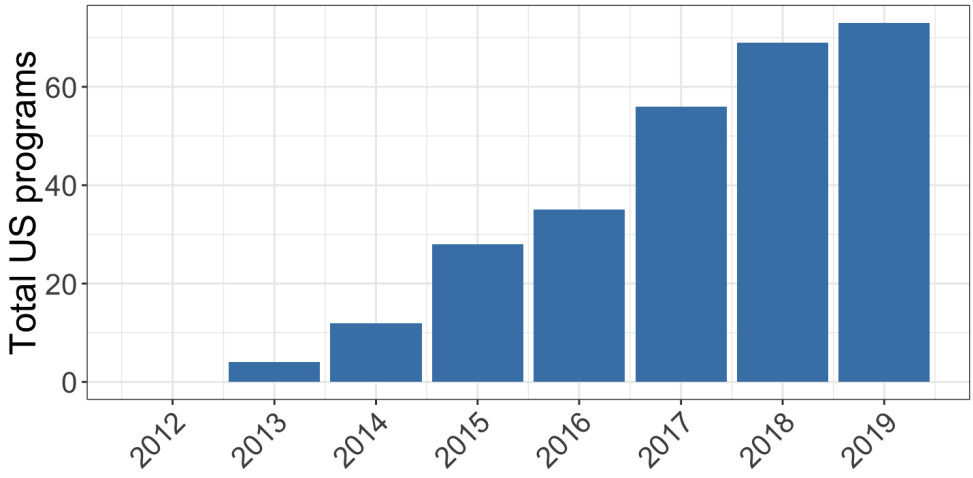
> 60% of researchers in the biological sciences report a need for more training in data science

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

Not just academia



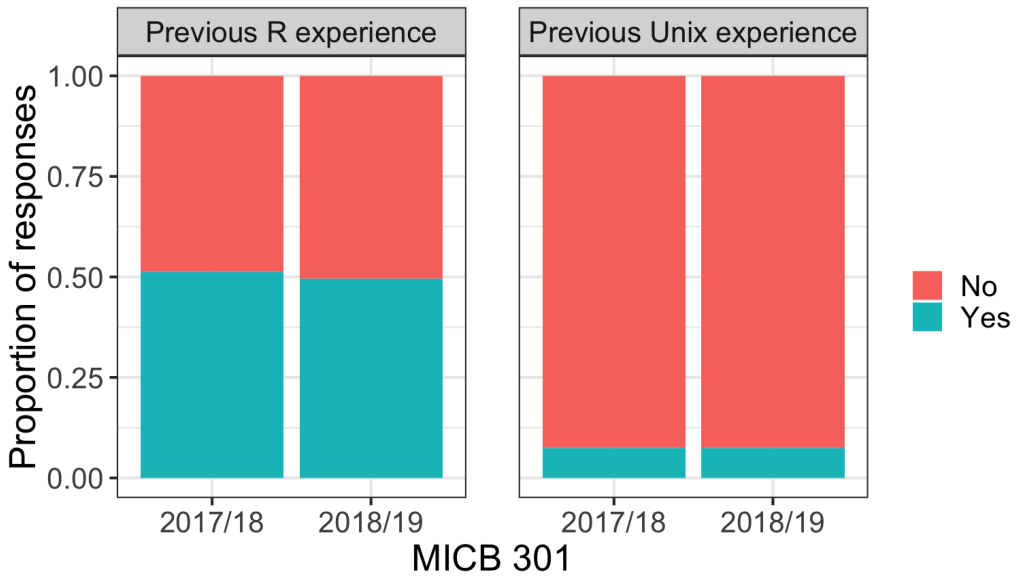
Masters of data science



(Michael Rappa, NC State University)

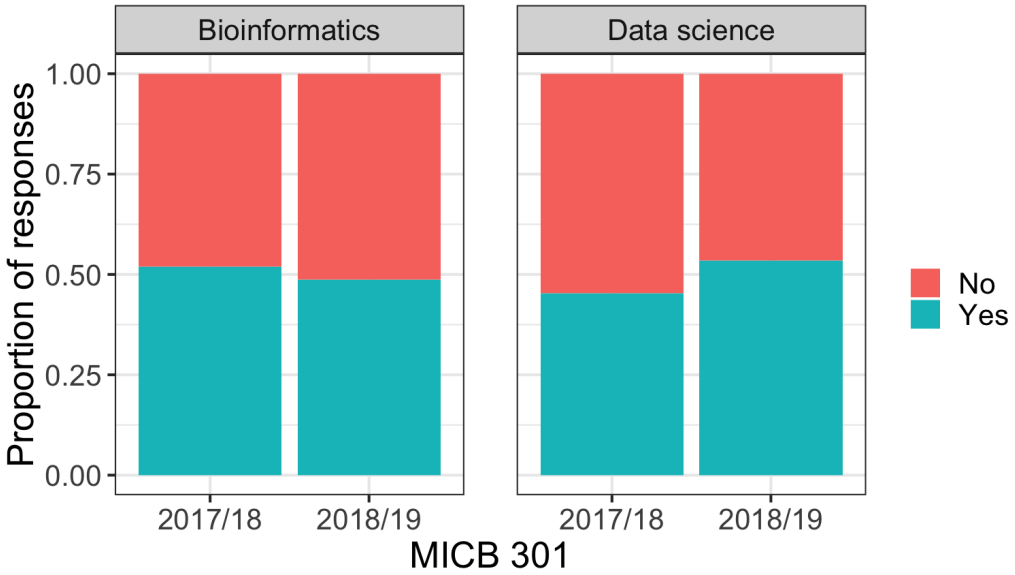
We need to teach data science in undergraduate life science curriculum.

An unmet need in M&I



An unmet need in M&I

Have you heard the term _____?



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)

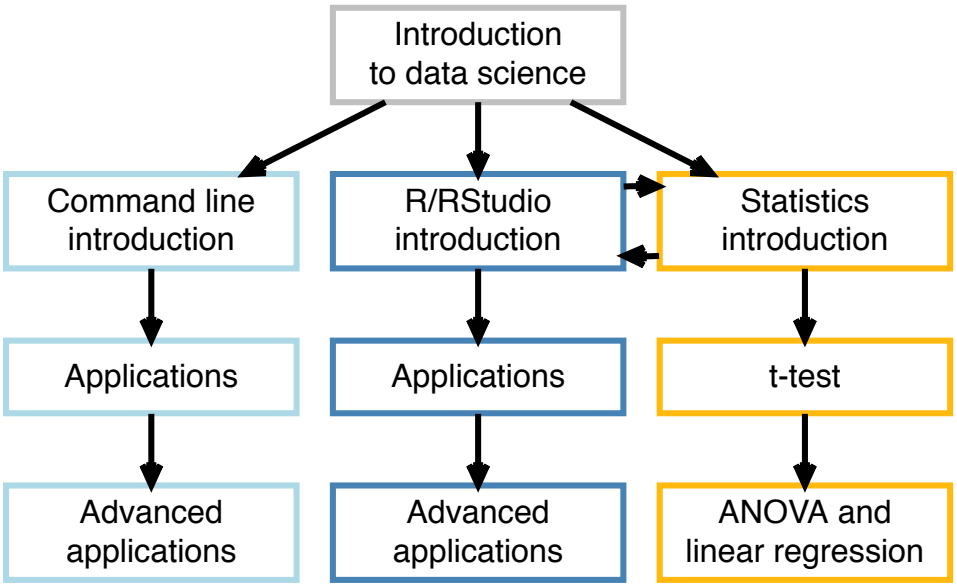
(Williams *et al* 2017)

Experiential
Data science for
Undergraduate
Cross-disciplinary
Education

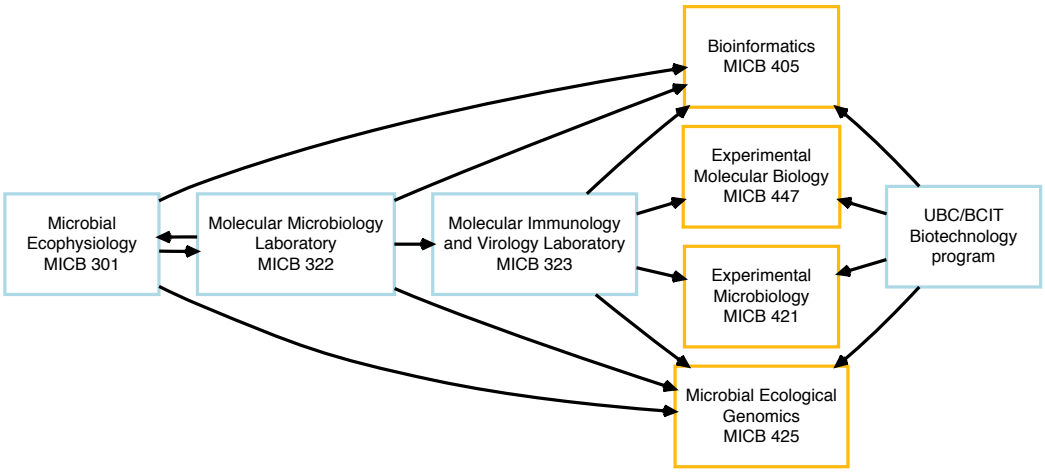
Our goal

Modular integration of
data science curriculum into
existing courses

Content overview

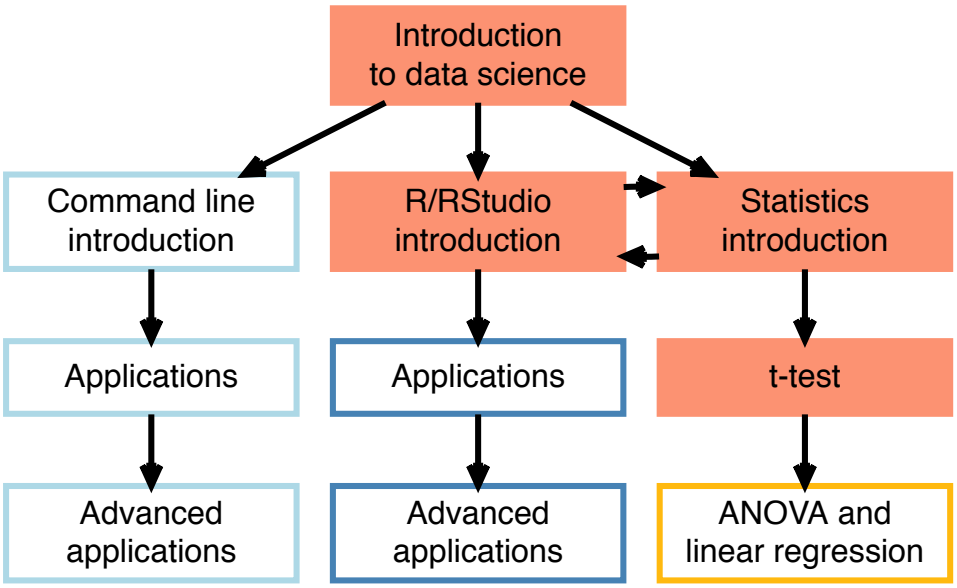


Course overview

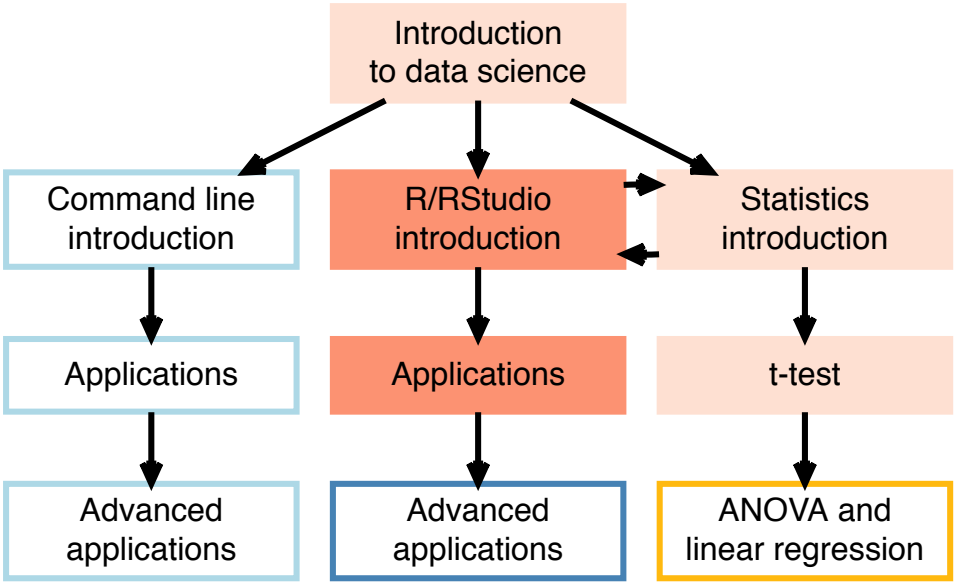


Example student

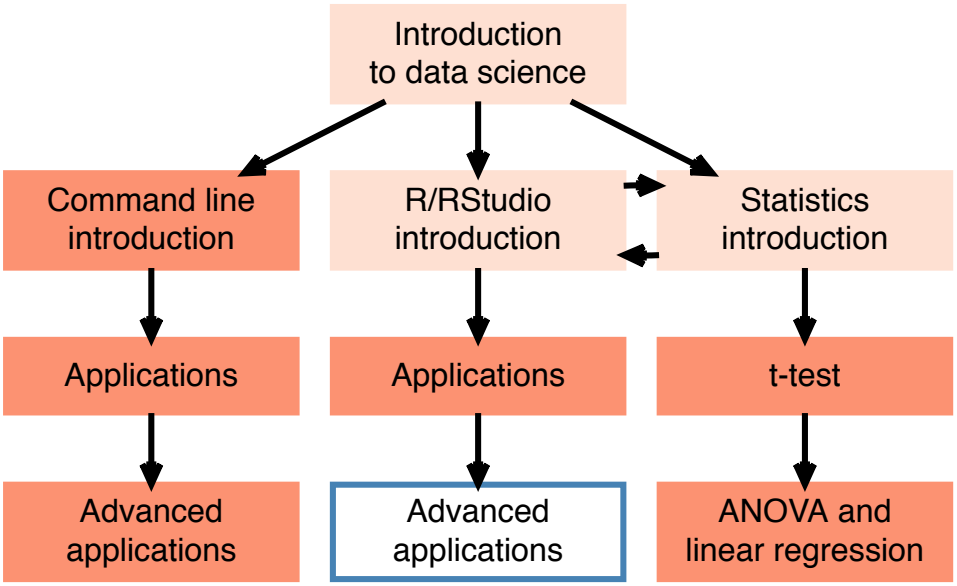
MICB 301



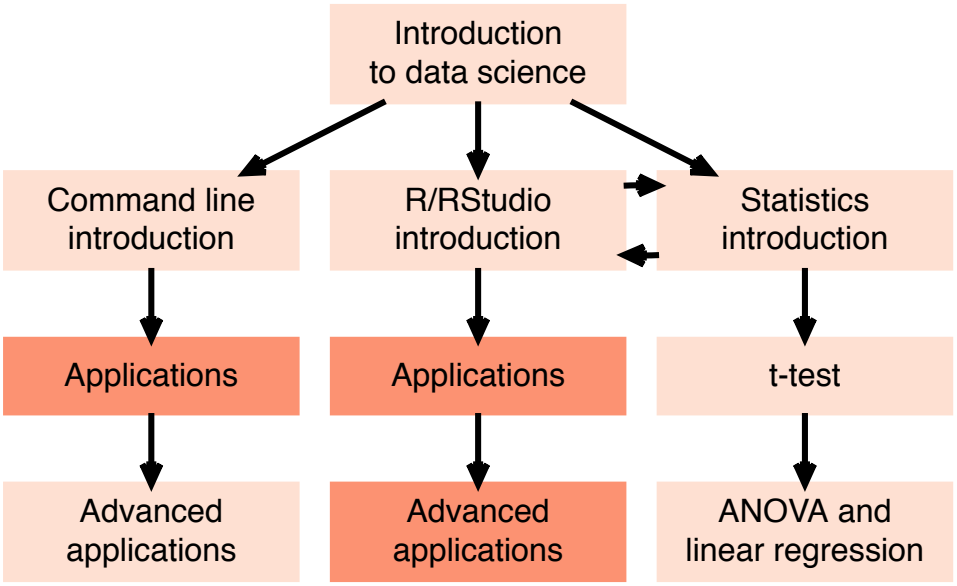
MICB 301 - 322



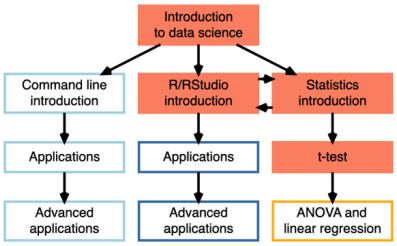
MICB 301 - 322 - 405



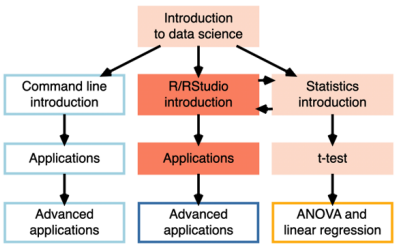
MICB 301 - 322 - 405 - 425



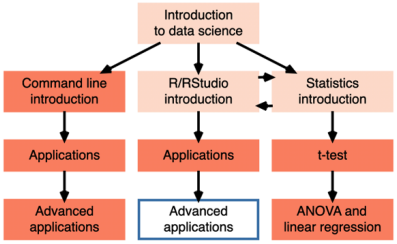
Example student



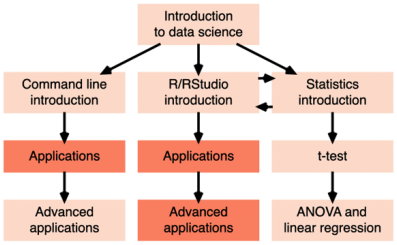
MICB 301



MICB 322



MICB 405



MICB 425

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*

Solutions to integration

3. Student preparation

- *No prior knowledge assumed*

4. Already overly full curricula

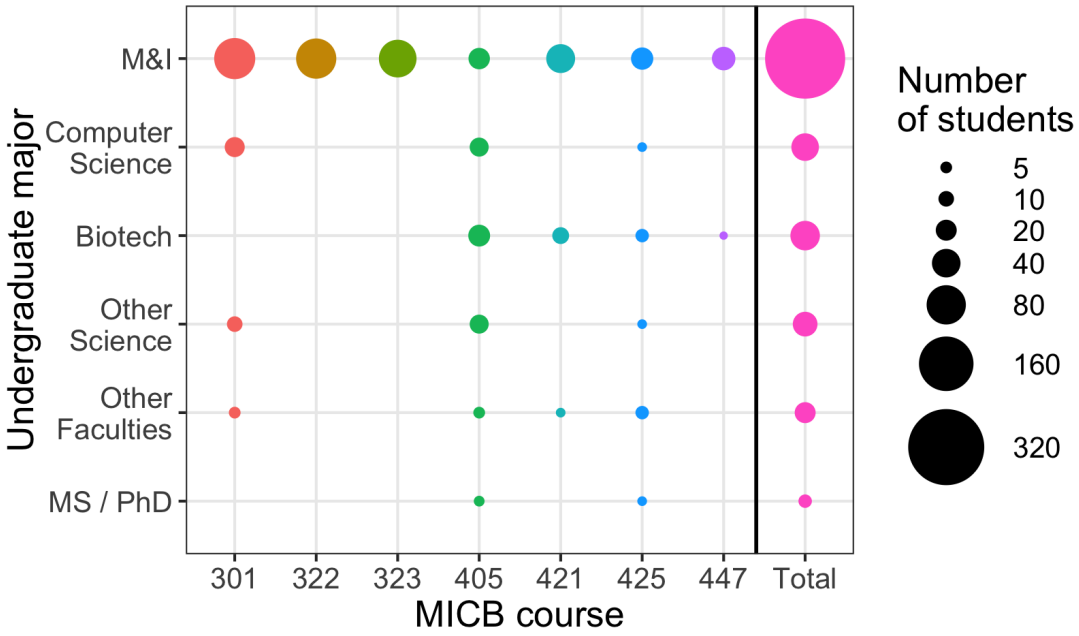
- *No new courses required*

5. Limited access to resources

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

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

Students impacted per year



MICB 301 as a case study

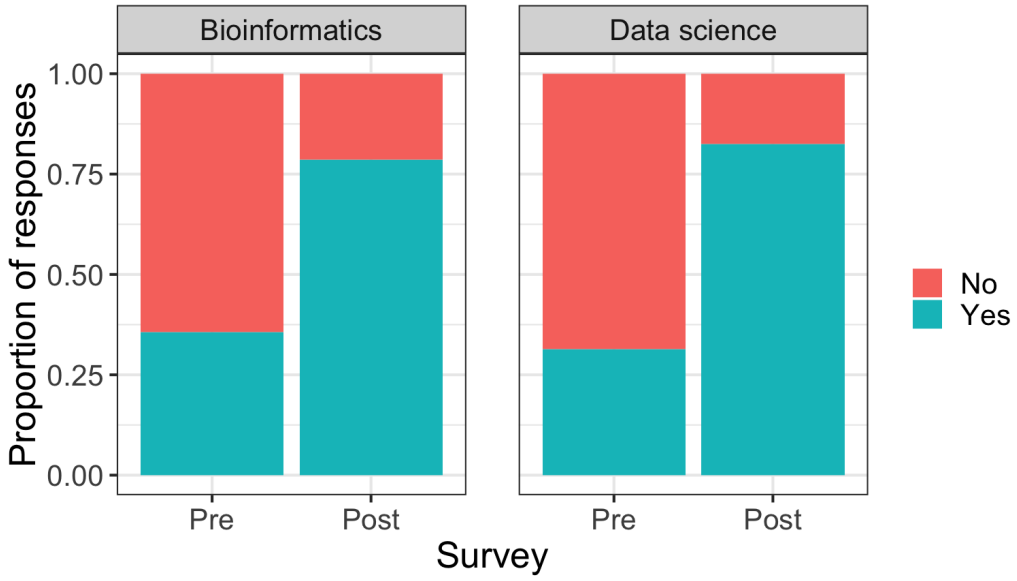
EDUCE in MICB 301

- ~120 students / yr
- 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

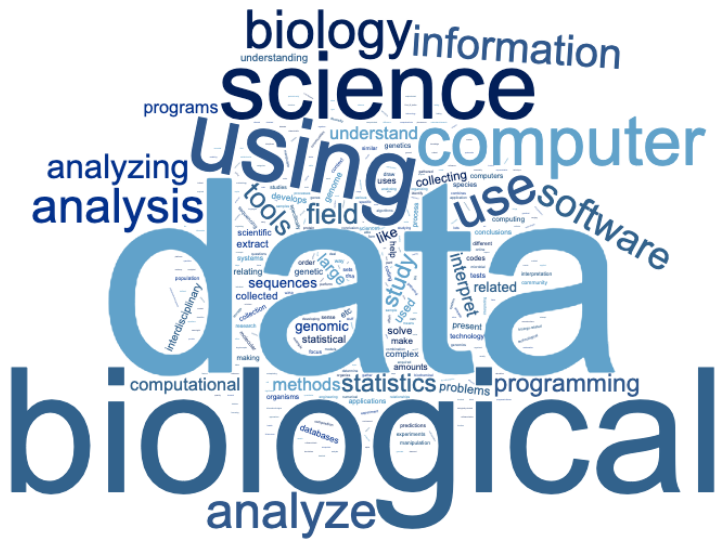
Increased awareness of data science

Have you heard the term _____?



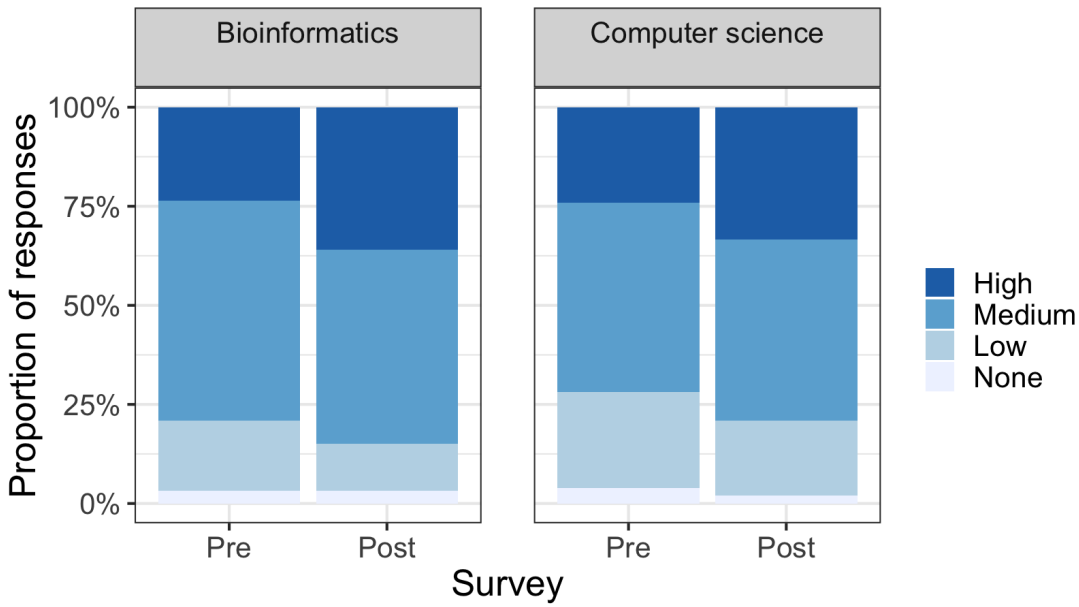
Definitions of data science

Definitions of bioinformatics



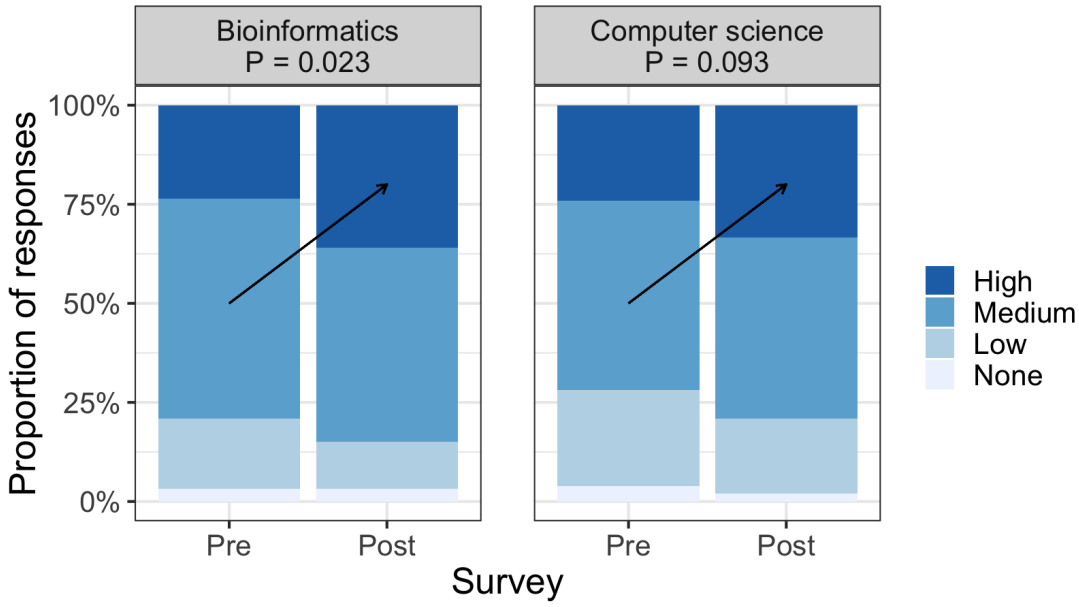
Increased interest in data science

How would you rate your interest in…



Increased interest in data science

How would you rate your interest in…

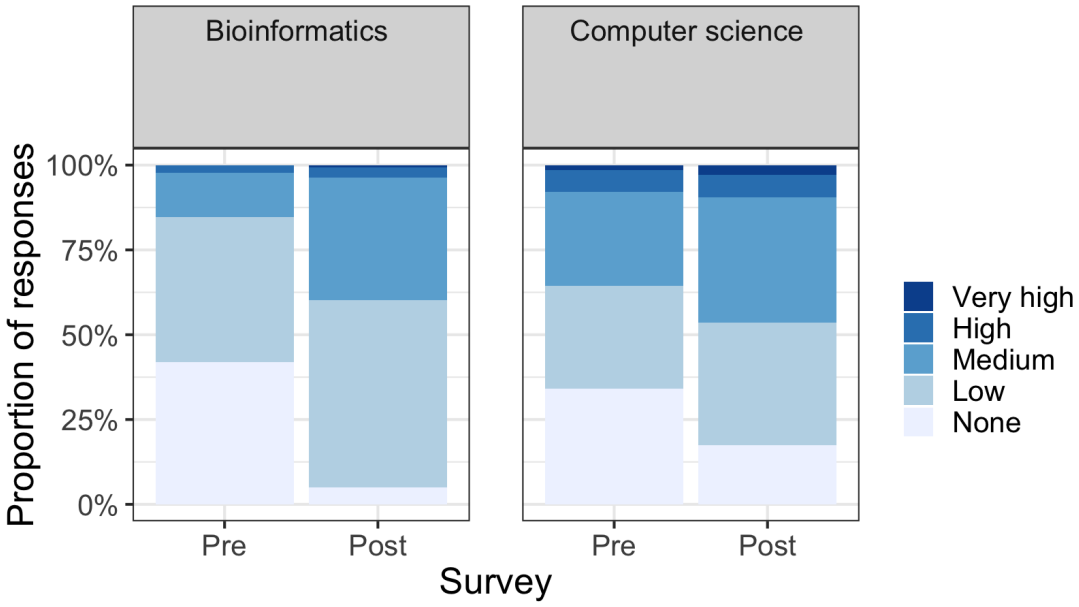


Except…

No significant changes in interest in statistics

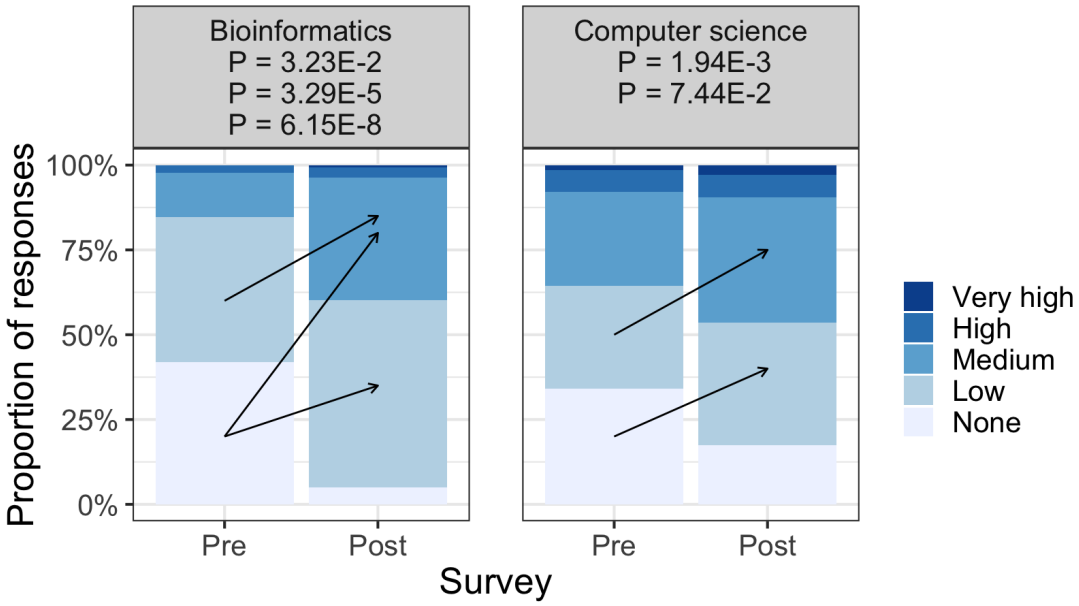
Increased experience in data science

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

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?
- Faculty of Science Data Science Curriculum Committee
- Independent learning tools https://ubc-educe.shinyapps.io/course_knitter/

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)
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)

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

Williams JJ *et al* 2017. *Barriers to integration of bioinformatics into undergraduate life sciences education.* BioRxiv. doi: 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