An integrated, modular approach to data science education in the life sciences

Kimberly A Dill-McFarland[1,2*], Steven J Hallam[1*],

1 Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada **2** Division of Allergy and Infectious Diseases, Department of Medicine, University of Washington, Seattle, WA, USA

* kadm@uw.edu * shallam@mail.ubc.ca

Abstract

We live in an increasingly data-driven world. This is strongly evident in the life sciences where high-throughput data generation platforms are transforming biology into information science. This has shifted major challenges in biological research from data generation to interpretation and knowledge translation. However, training in bioinformatics, or more generally data science for life scientists, lags behind current demands at multiple career levels. In particular, the development of accessible, undergraduate data science curricula has the potential to improve learning outcomes and better prepare students in the life sciences to address current social, environmental, and technological careers. Here, we describe the Experiential Data science for Undergraduate Cross-Disciplinary Education (EDUCE) initiative, which aims to progressively build data science competencies across two years of undergraduate coursework. In this program, students complete data science modules (2 - 17 hrs) integrated into required courses augmented with opportunities to participate in additional training through calibrated co-curricular activities. The EDUCE program draws on a cross-disciplinary community of practice to overcome several reported barriers to data science for life scientists, including instructor capacity, student prior knowledge, and relevance to discipline-specific problems. Preliminary survey results

indicate that even a single module improves student self-reported interest and/or experience in bioinformatics and computer science. Thus, EDUCE provides a flexible framework for integration of effective data science curriculum into a diverse range of undergraduate programs.

Introduction

The world is increasingly inundated with data. This is especially true in the life sciences, where recent advances in high-throughput sequencing and mass spectrometry have resulted in an explosion of multi-omic information (*e.g.* DNA, RNA, protein, etc.). For example, over 31 tera-bases of genomic data are created on average per second, and rates are expected to continue to increase with continued improvement in platform technologies [1].

Over the past decade, application of these multi-omic platforms to the study of microbial communities has changed our perception of the world, and the impacts of these tools can be seen across environmental [2], agricultural [3], and medical fields of study [4], as just some examples. Thus, major challenges in the life sciences are shifting away from data generation toward data processing and interpretation. This has resulted in a need for training in bioinformatics or more generally, data science.

However, life science training in these areas lags behind recognized needs. Despite calls to action as early as 20 years ago [5], there remains a sustained, unmet need for bioinformatics training in the life sciences [6]. A meta-analysis of surveys from organizations such as the Global Organisation for Bioinformatics Learning, Education and Training (GOBLET), the European life-sciences Infrastructure for biological Information (ELIXIR), the U.S. National Science Foundation (NSF), and the Australia Bioinformatics Resource (EMBL-ABR) found the desire for training to be widespread both in terms of geography and career level [6]. While a number of training formats were applicable, the most common ask for undergraduate students was integrated data science training within current degree programs [6]. This ensures that all students receive the necessary training and provides comprehensive, foundational knowledge for further studies or subsequent careers.

In order to meet this clear and present need for data science education in the life

sciences, the University of British Columbia has launched a number of initiatives in recent years. These include the Master of Data Science program (established 2016) as well as several specialized undergraduate majors like biotechnology (BIOT) and combined computer science-life science programs (established early 2000s). While these programs provide a subset of students with in-depth training in data science and bioinformatics, they are inherently self-selecting. This is concerning as there is continued under-representation of women and minorities in undergraduate degree programs in computer science, mathematics, and statistics [7]. Therefore, since data science training is necessary for the majority of life science graduates and specialized degree programs do not attract all students, this training needs to be incorporated into the fabric of undergraduate coursework in a progressive manner such that students are able to build their confidence and skillsets across the curriculum.

With this integrative objective in mind, the Experiential Data science for Undergraduate Cross-Disciplinary Education (EDUCE) initiative was launched in Fall 2017. The following provides an in-depth description of the EDUCE framework as well as preliminary findings from EDUCE in the Department of Microbiology and Immunology at the U. of British Columbia.

Experiential Data science for Undergraduate Cross-Disciplinary Education (EDUCE)

EDUCE seeks to develop a uniform, cross-disciplinary, and collaborative framework to equip all undergraduate students in the life sciences with basic competency in data science. The EDUCE curriculum provides students with the most commonly needed data science skills as defined by recent surveys [6]. Our foundational learning objectives are for students to learn to:

- Recognize and define uses of data science
- Explore and manipulate data
- Visualize data in tables and figures
- Apply and interpret statistical tests

For each program, sub-objectives within these foundational objectives are then framed and defined using discipline-specific data, questions, and software. These learning objectives are achieved through 1) modular curriculum in courses, 2) calibrated co-curricular activities, and 3) a cross-disciplinary community of practice. These diverse learning opportunities allow students to develop progressive data science competencies over several years of integrated coursework.

Modular curriculum

EDUCE curriculum is modular, thus providing a flexible framework for integration of data science curriculum across a diverse range of programs. Modules are self-contained, customizable bundles of learning materials that are scaffolded into existing courses. Each module can be delivered as a separate chunk within a course (like a 2-week intensive), a reoccurring section (like 'data science Fridays'), or a fully integrated curriculum wherein the students are not specifically aware of what is or is not part of a module.

Because EDUCE is modular, it overcomes several of the largest barriers to data science education in the life sciences [8]. Specifically, the modules can be taught by the main course instructor(s), visiting lectures, or the students themselves. This lifts the need for all faculty to be well-versed in data science methods as a single, knowledgeable instructor can teach modules across several courses or even across an entire program. Also, the modules assume no prior knowledge as they incorporate all necessary introductory and background material, including materials unique to that module and review activities from previous modules. Finally, module integration circumvents traditional course creation, *i.e.* new course codes, thus allowing quicker, more broadly accessible curriculum deployment within departments or across faculties.

In addition, modular instruction allows for more timely and direct linkages to discipline-specific content than traditional course structures. For example, students cover the global impacts, thermodynamics, etc. of the nitrogen cycle in their regular course content and then immediately enter an EDUCE module in which they quantify and plot nitrogen species in the ocean using R [9]. This aids student learning in both the discipline and data science as students are able to reference prior knowledge [10] as

well as think about the meaning and application of a new skill as they learn that skill [11]. Thus, modules leverage student interest in their chosen area of study and maximize the relevancy of the data science content.

An EDUCE module consists of all materials related to a set of learning objectives and includes 1) introduction, 2) practice, 3) application, and 4) communication (Fig 1). As students progress from introductory to advanced modules across several courses, their focus transitions from introduction and practice to application and communication. Thus, while each module contains all four aspects, later modules build on previous knowledge to challenge students to apply their data science skills to more complex data and new scientific questions. This builds progressive competency in data science over several years of instruction without the need for additional course requirements.

Fig 1. EDUCE module overview. Each EDUCE module consists of introduction, practice, application, and communication. Early, introductory modules are focused on introduction and practice while later, more advanced modules challenge students to apply their data science skills to complex data and to communicate novel research findings.

Modules vary in size from a single activity to semester-long projects, and they can include a range of materials. The following pieces are collated together based on the module's learning objectives, available class time, and instructor needs. Examples of specific microbiology modules are shown in detail below and additional examples can be found on the EDUCE website (`educe-ubc.github.io`).

- Lecture slides
- Notes / handouts
- In-class tutorials
- At-home tutorials
- Screen capture videos
- Individual assignments
- Team projects

Example modules in Microbiology

EDUCE has or is in the process of being integrated into seven third and fourth year microbiology (MICB) courses selected based on internal MICB curriculum review and instructor interest (Table 1). These courses include lectures, wet labs, and dry labs (*i.e.* tutorials).

Table 1. EDUCE courses in microbiology (MICB)

MICB code	Name	L	W	D	Hrs/wk	EDUCE hrs
301	Microbial ecophysiology	X			3	5
322	Molecular microbiology laboratory	X	X		6	3
323	Molecular immunology and virology laboratory	X	X		7	3
405	Bioinformatics	X		X	4	17
425	Microbial ecological genomics	X			3	17
421, 447	Experimental microbiology/molecular biology (CURE)	X	X		7	2+

L: Lecture, W: Wet lab, D: Dry lab, CURE: Course-based undergraduate research experience

EDUCE content varies in each course and builds from the third (300+) to fourth-year (400+) (Fig 2). Ideally, students enter the program in MICB 301 where they master basic R/RStudio scripting to manipulate data, create figures, and perform *t*-tests. Then, students acquire additional practice in R/RStudio in 300-level labs (MICB 322, 323) where they employ specialized R packages to manipulate and plot data generated in the course. For example, MICB 322 uses the Sushi package [12] to visualize transposon insertions in the *Caulobacter* mutant library created by students in the course. Next, students progress to capstone 400-level courses. These include traditional 400-level courses (MICB 405, 425) and course-based undergraduate research experiences (CUREs, MICB 421, 447). In the traditional courses, EDUCE modules are built into the coursework and include capstone projects using published but under-explored metagenomic and metatranscriptomic data-sets [13]. In contrast, CUREs involve student-led research projects that generate novel data. Thus, EDUCE serves as a consultation resource to help students with R/RStudio packages, Unix tools, or statistical methods as required for their individual projects. Finally, all EDUCE capstone courses offer the opportunity for students to communicate their scientific findings through publication in the undergraduate Journal of Microbiology and Immunology (JEMI, `jemi.microbiology.ubc.ca`) and/or presentation at the MICB Undergraduate Research Symposium (URS, `jemi.microbiology.ubc.ca/UndergraduateResearchSymposium`).

Fig 2. EDUCE curriculum in microbiology. Students enter the program in MICB 301 at the U. of British Columbia or equivalent coursework at the British Columbia Institute of Technology (BCIT). Some students take additional 300-level courses, MICB 322 and 323. Then, all students progress to one or more capstone 400-level courses that culminate formal publication or presentation within the department. MICB: microbiology

While the above is an ideal module progression, there exists considerable variation and flexibility in the EDUCE program. This is necessary, because MICB prerequisite structure and major requirements do not yield a linear or conserved course progression for students. Microbiology (MICB) and Microbiology Honours (MBIM) degrees require the majority of EDUCE courses while the combined majors require only some EDUCE courses with little overlap across the three options. Specifically, the MICB/MBIM and computer science (+CPSC) or Earth, Ocean, and Atmospheric Sciences (+EOAS) majors require the first EDUCE course (MICB 301) but different 400-level capstones (MICB 405 and 425, respectively). MICB/MBIM + EOAS also requires an additional 300-level EDUCE 'practice' course. The Biotechnology (+BIOT) degree, on the other hand, is a joint program wherein students spend their second and third years at the British Columbia Institute of Technology (BCIT). Thus, +BIOT does not require any 300-level EDUCE courses. In addition, many students choose to take the optional or suggested 400-level capstone courses in addition to those required by their major (Table 2).

Table 2. EDUCE course requirements for microbiology majors

	Undergraduate major			
MICB code	**MICB/MBIM**	**+CPSC**	**+EOAS**	**+BIOT**
301	R	R	R	S
322	R	S	R	O
323	R	S	O	O
405	O	R	S	R
425	O	O	R	S
421/447	R	O	O	S

R: Required, S: Suggested, O: Optional. MICB: Microbiology, MBIM: Microbiology Honours, CPCS: Computer Science, EOAS: Earth, Ocean, and Atmospheric Sciences, BIOT: Biotechnology.

These course modules form the basis of the EDUCE program and address the foundational learning objectives. While all MICB students obtain some level of data literacy through EDUCE modules within their required courses, some students obtain

more practice or more advanced skills if they choose to take additional EDUCE courses. Additionally, all students can gain more experience and competencies through the remaining aspects of the program.

Co-curricular activities

EDUCE co-curricular activities reinforce course modules by providing students with diverse data science learning opportunities outside the classroom. These can include workshops, hackathons, independent research, etc. that incorporate one or more of the foundational learning objectives and meet the following criteria.

Co-curricular activities must first and foremost be accessible to undergraduate students in terms of cost, content, and schedule as these are common barriers. Specifically, undergraduate students are often unable to pay the high costs of data "boot camps" [14] and unwilling to pursue outside training opportunities as they feel intimidated by data science tools [15].

Secondarily, EDUCE co-curricular activities directly reiterate and build on learning objectives from related courses. As with the integrated course modules, this aims to improve student learning by referencing prior knowledge [10] and imparting discipline-specific meaning to the skills being learned [11]. This is particularly important for short-term training opportunities as fully stand-alone programs, such as "boot camps", do not promote long-term skill development even at the graduate level [16].

Finally, co-curriculars utilize the same or similar data sets and tools as courses. Course modules, thus, serve as pre-requisites for co-curriculars such that students are more prepared and less intimidated by data science training opportunities outside the classroom. This requires co-curriculars to be discipline-specific but does not necessarily exclude other participants. For example, an R [9] workshop using oceanic oxygen data may be designed for life or Earth science students but is no less accessible to other disciplines than the commonly used cars data [17].

EDUCE co-curricular activities fulfill a number of roles within the program. Workshops generally serve as additional practice with data science tools for students who want more experience with concurrent modules or to review concepts from past modules. They also allow students to explore modules from courses that do not fit in

their schedules. Other opportunities like hackathons and research challenge students to move beyond the course modules and to develop their data science skills to answer new, unexplored questions alongside other undergraduates as well as graduate students and even faculty. Thus, as with the course modules, EDUCE co-curriculars are flexible and integrated in order to provide students with varied types and levels of instruction from introductory to advanced (Fig 1).

Example co-curricular workshops in Microbiology

The EDUCE program at UBC has partnered with the Ecosystem Services, Commercialization Platforms and Entrepreneurship (ECOSCOPE, `ecoscope.ubc.ca`) training program and the Applied Statistics and Data Science Group (ASDa, `asda.stat.ubc.ca`) to deliver data science workshops accessible to all participants from undergraduate students to industry professionals.

The current workshop portfolio (`github.com/EDUCE-UBC/workshops_R`) consists of 32 hours of content in R [9] using the same oceanic geochemical data set [18] as most EDUCE course modules. These workshops target the EDUCE foundational learning objectives and include:

- Introduction to R
- The R tidyverse
- Reproducible research in R and Git
- Intermediate R programming
- Statistical models in R
- Phylogenetics and microbiomes in R

Introduction to R is a short refresher of the MICB 301 module or an introduction to the workshops for non-EDUCE students. The R tidyverse closely follows its respective module in MICB 405 and 425 (Fig 1) and thus, can be used as a refresher or additional practice. The remaining workshops have some overlap with course modules (like ANOVA in Statistical models in R and MICB 405) but mainly build on the fourth year modules to challenge students to continue their learning in data science.

Importantly, ECOSCOPE sponsorship allows undergraduates to take any workshop for 10 CAD as compared to regular registration fees of 100 CAD or more. This, along with scheduling to align with course modules and to accommodate the most common undergraduate schedules in MICB, facilitates undergraduate participation. Moreover, these opportunities are actively advertised in undergraduate courses in MICB and directly linked to concurrent and previous course modules.

Community of practice

Data science, itself, is a cross-disciplinary field traditionally comprised of statistics, mathematics, and computer science. Thus, EDUCE operates through a cross-disciplinary community of practice including these and other fields. While EDUCE at UBC is led by a postdoctoral fellow in MICB, the team spans 10 departments across three faculties (Fig 3). For example, statistics consults on module development, mathematics provides resource support, and teaching assistants (TAs) have been recruited from the Faculties of Science, Applied Science, and Medicine. The community also includes scientists at many career levels including undergraduate and graduate students, postdoctoral fellows, instructors, research faculty, and staff. Overall, the EDUCE team brings together expertise from across the university to provide relevant curriculum grounded in both current data science best practices and scholarship of education leadership.

Fig 3. EDUCE team members at UBC. The EDUCE program is led by MICB in the Faculty of Science. However, TAs, consultants, collaborators, and other support come from across UBC, including 10 departments from 5 Faculties. Team members also represent multiple career stages from undergraduates to faculty to staff. Other includes the Faculties of Applied Science, Land and Food Systems, Medicine, and UBC Central.

The EDUCE community of practice reflects widespread changes in teaching in MICB. A growing number of MICB courses (24% in 2018/19) are taught by a team of faculty often including both tenure-track researchers and instructors. This team teaching brings additional expertise, styles, and perspectives into the classroom [19, 20] as well as fosters collaboration and community at the university [20].

Additionally, EDUCE TAs contribute to team teaching. Unlike traditional appointments, EDUCE TAs are not assigned to a specific course nor do they always

come from the course department. Instead, EDUCE TAs are recruited from any department and work across the entire program. With their varied expertise, TAs act as contributing members to all aspects of EDUCE from curriculum development to office hours to instruction. This provides the EDUCE students with diverse resources, viewpoints, and role models as they progress through modules and co-curriculars. Moreover, this provides the TAs will hands-on teaching experience as well as evaluation to build their teaching portfolios.

Preliminary outcomes of EDUCE

EDUCE launched at UBC in Term 1 of the 2017/18 academic year. Per year, approximately 300 (redundant) students are impacted through 40 total classroom hours in five MICB courses. Once modules have been deployed in all seven planned courses, this will increase to roughly 475 students and 45 hours. Also, over the first two years, co-curricular workshops were taken by 77 EDUCE students, which resulted in increasing undergraduate participation from 0% (2016/17, prior to partnership with EDUCE) to 8.8% (2017/18) to 23% (2018/19) of total registrants.

Evaluation of EDUCE is on-going and includes student surveys (BREB H17-02416) at the start (pre) and end (post) of courses (Text S1) as well as at the end of workshops. Courses with <5 EDUCE hours (Table 1) were excluded to avoid redundant surveying of students. For example, >95% of students in MICB 322 concurrently take MICB 301. To-date, 439 pre-course responses and 400 post-course responses have been collected with 91 students indicating willingness to complete post-graduation surveys or focus groups. We have also collected 44 responses from co-curricular workshops.

Preliminary analysis of the first EDUCE course, MICB 301, from 2017-18 revealed that student self-reported interest in 'bioinformatics' increased with a significant transition from medium to high interest (Monte Carlo FDR P = 1.82E-2, Fig 4). Similarly, student self-reported experience in 'bioinformatics' significantly increased from none to low (FDR P = 1.84E-7) or medium (FDR P = 2.74E-5) as well as from low to medium (FDR P = 2.69E-2). This is particularly important for the MICB program at UBC as the fourth year course MICB 405, Bioinformatics is available to these students.

In contrast, self-reported interest in 'computer science' remained constant but

Fig 4. Student interest and experience in data science. Student responses from matched pre- and post-surveys in MICB 301 indicating self-reported interest (N = 143 - 146) and experience (N = 136 -143) in data science areas including bioinformatics and computer science. Numerical responses from 2018/19 were converted to categorical groups from 2017/18 as none (0), low (1-3), medium (4-7), and high (8-10). Arrows and p-values indicate significant response changes by Monte Carlo symmetry tests for paired contingency tables.

showed significant gains in experience from none to low (FDR P = 2.44E-3), and 'statistics' had no significant changes in either interest or experience. Thus, it appears that even minimal exposure to data science in disciplinary coursework (5 hours, (Table 1)) may impact student interest and confidence in some data science areas.

??????RESULTS ON JEMI, URS

Full analysis details are available at `https://github.com/kdillmcfarland/EDUCE_desc`.

On-going challenges

The EDUCE program aims to improve undergraduate competency in data science through modular course curriculum and calibrated co-curricular activities offered by a cross-disciplinary teaching team. While course modules are designed to be extensible, some modification is required prior to implementation in a new course in order to accommodate individual teaching style, course schedule, and student prior knowledge. This process is made easier through collaboration across the EDUCE community of practice. However, as with any new curriculum, instructor preparation time remains a barrier to implementation of EDUCE modules in some courses.

Another on-going challenge is student's prior knowledge. Due to the non-linear nature of MICB degree requirements at UBC, some students come into an intermediate EDUCE course without prior EDUCE experience while others repeat EDUCE content in multiple courses. To overcome this, we are developing supplementary exercises for students entering the program in fourth-year as well as integrating new data sets in fourth-year modules.

Finally, while preliminary results indicate that EDUCE may have positive impacts on student interest and learning in data science, further analyses of collected survey data are needed to determine if EDUCE curriculum is truly causative of these outcomes.

?????Do we need a discussion section that related EDUCE findings with other programs and papers?

Acknowledgments

The authors wish to thank faculty and instructors for access to courses and support during module integration, including Sean Crowe, Marcia Graves, Martin Hirst, William Mohn, David Oliver, and Jennifer Sibley.

We gratefully acknowledge the financial support for this project provided by UBC Vancouver students via the Teaching and Learning Enhancement Fund. This work was also supported the UBC Department of Microbiology and Immunology as well as the Natural Sciences and Engineering Research Council of Canada (NSERC), Collaborative Research and Training Experience Program (CREATE, ????Grant number?????).

References

1. Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, et al. Big Data: Astronomical or Genomical? PLOS Biology. 2015;13(7):e1002195.

2. Hawley AK, Nobu MK, Wright JJ, Durno WE, Morgan-Lang C, Sage B, et al. Diverse Marinimicrobia bacteria may mediate coupled biogeochemical cycles along eco-thermodynamic gradients. Nature communications. 2017;8(1):1507. doi:10.1038/s41467-017-01376-9.

3. Neumann AP, Suen G. The Phylogenomic Diversity of Herbivore-Associated Fibrobacter spp. Is Correlated to Lignocellulose-Degrading Potential. mSphere. 2018;3(6):e00593–18. doi:10.1128/mSphere.00593-18.

4. Sze MA, Schloss PD. Leveraging Existing 16S rRNA Gene Surveys To Identify Reproducible Biomarkers in Individuals with Colorectal Tumors. mBio. 2018;9(3):e00630–18. doi:10.1128/mBio.00630-18.

5. MacLean M, Miles C. Swift action needed to close the skills gap in bioinformatics. Nature. 1999;401(6748):10. doi:10.1038/43269.

6. Attwood TK, Blackford S, Brazas MD, Davies A, Schneider MV. A global perspective on evolving bioinformatics and data science training needs. Briefings in Bioinformatics. 2017; p. bbx100–bbx100.

7. National Science Foundation: National Center for Science and Engineering Statistics. Women, Minorities, and Persons with Disabilities in Science and Engineering: 2019. Alexandria, VA: Special Report NSF 19-304; 2019. Available from: `https://www.nsf.gov/statistics/wmpd`.

8. Williams JJ, Teal TK. A vision for collaborative training infrastructure for bioinformatics. Annals of the New York Academy of Sciences. 2017;1387(1):54–60. doi:10.1111/nyas.13207.

9. R Core Team. R: A language and environment for statistical computing; 2019. Available from: `http://www.r-project.org/`.

10. Bransford JD, Brown AL, Cocking RR. How People Learn: Brain, Mind, Experience, and School. Washington D.C.: National Academy Press; 2000.

11. Morris CD, Bransford JD, Franks JJ. Levels of processing versus transfer appropriate processing. Journal of Verbal Learning and Verbal Behavior. 1977;16(5):519–533. doi:10.1016/S0022-5371(77)80016-9.

12. Phanstiel DH. Sushi: Tools for visualizing genomics data; 2018.

13. Hawley AK, Torres-Beltrán M, Zaikova E, Walsh DA, Mueller A, Scofield M, et al. A compendium of multi-omic sequence information from the Saanich Inlet water column. Scientific Data. 2017;4:170160.

14. National Academies of Sciences and Medicine E. Data Science for Undergraduates: Opportunities and Options. Washington, DC: The National Academies Press; 2018. Available from: `https://www.nap.edu/catalog/25104/data-science-for-undergraduates-opportunities-and-options`.

15. Farrell KJ, Carey CC. Power, pitfalls, and potential for integrating computational literacy into undergraduate ecology courses. Ecology and evolution. 2018;8(16):7744–7751. doi:10.1002/ece3.4363.

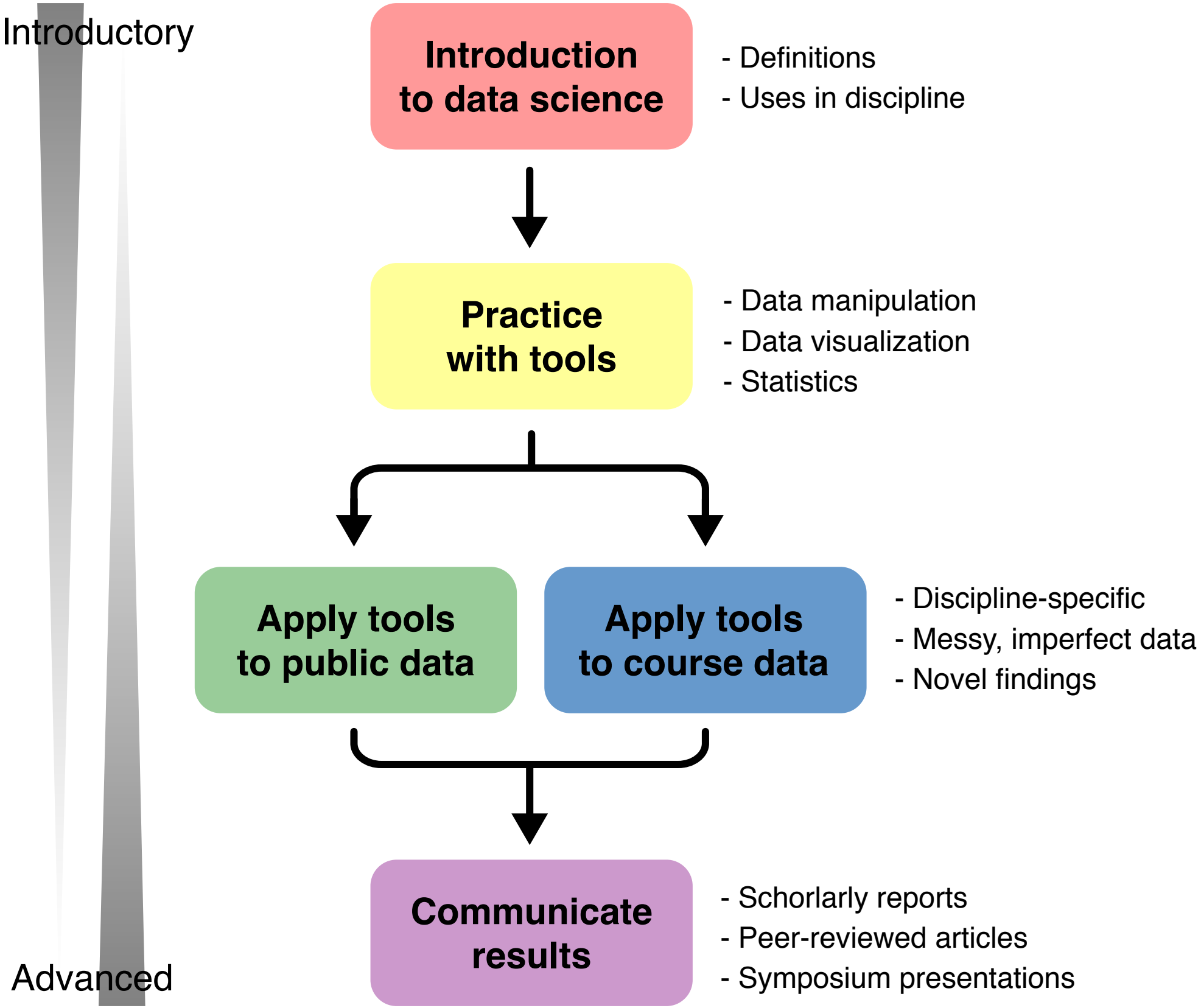
16. Feldon DF, Jeong S, Peugh J, Roksa J, Maahs-Fladung C, Shenoy A, et al. Null effects of boot camps and short-format training for PhD students in life sciences. Proceedings of the National Academy of Sciences of the United States of America. 2017;114(37):9854–9858. doi:10.1073/pnas.1705783114.

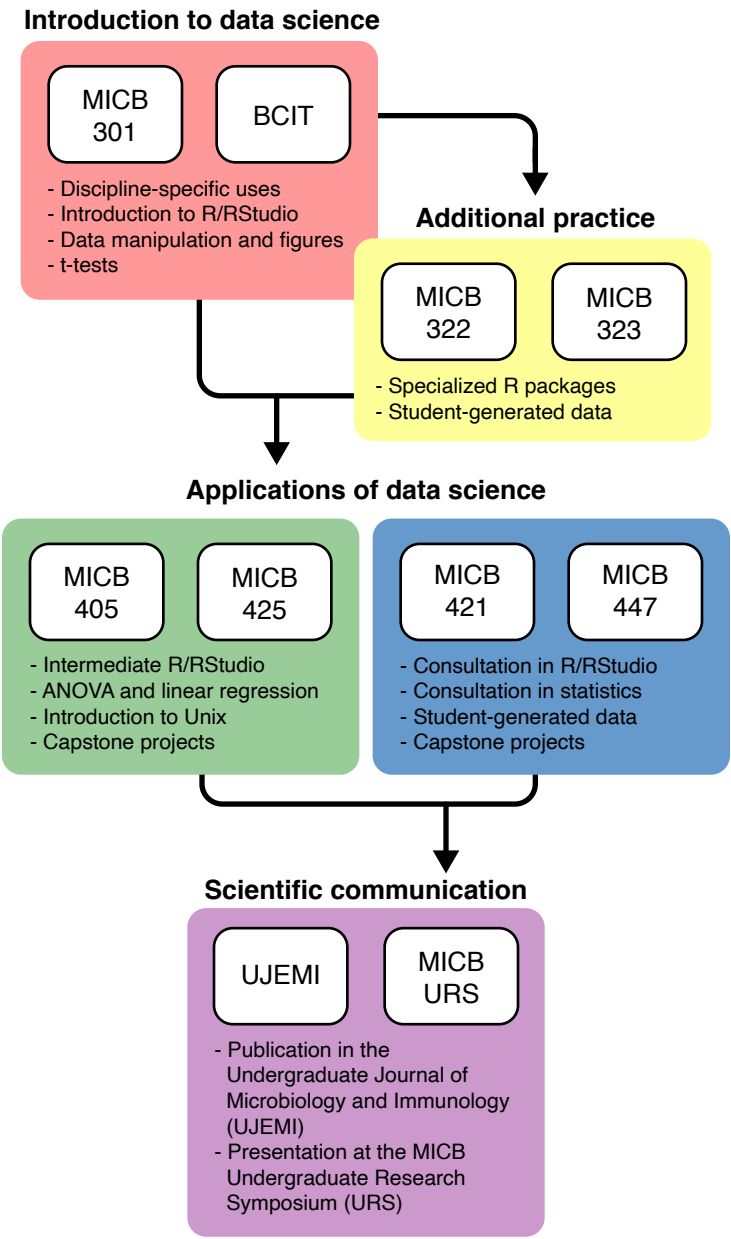
17. Fox J, Weisberg S. An {R} Companion to Applied Regression. 3rd ed. Thousand Oaks, Ca: Sage; 2019.

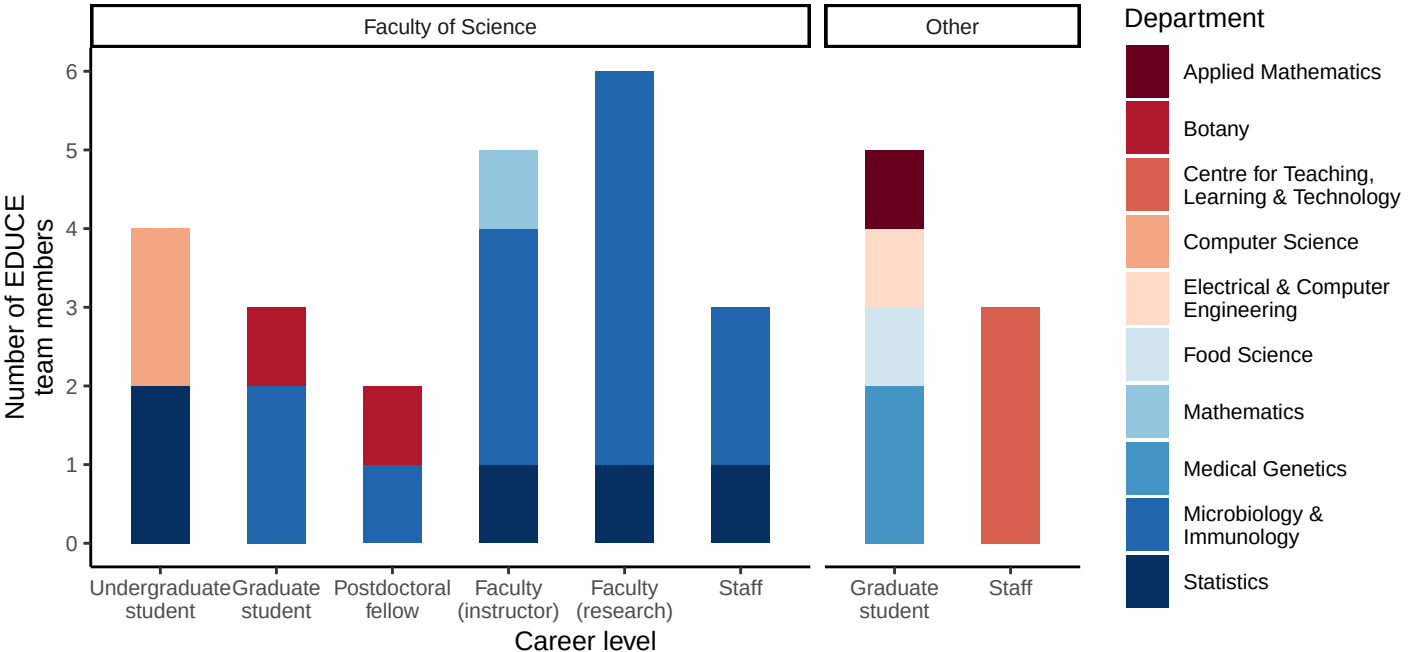
18. Torres-Beltrán M, Hawley AK, Capelle D, Zaikova E, Walsh DA, Mueller A, et al. A compendium of geochemical information from the Saanich Inlet water column. Scientific Data. 2017;4:170159.

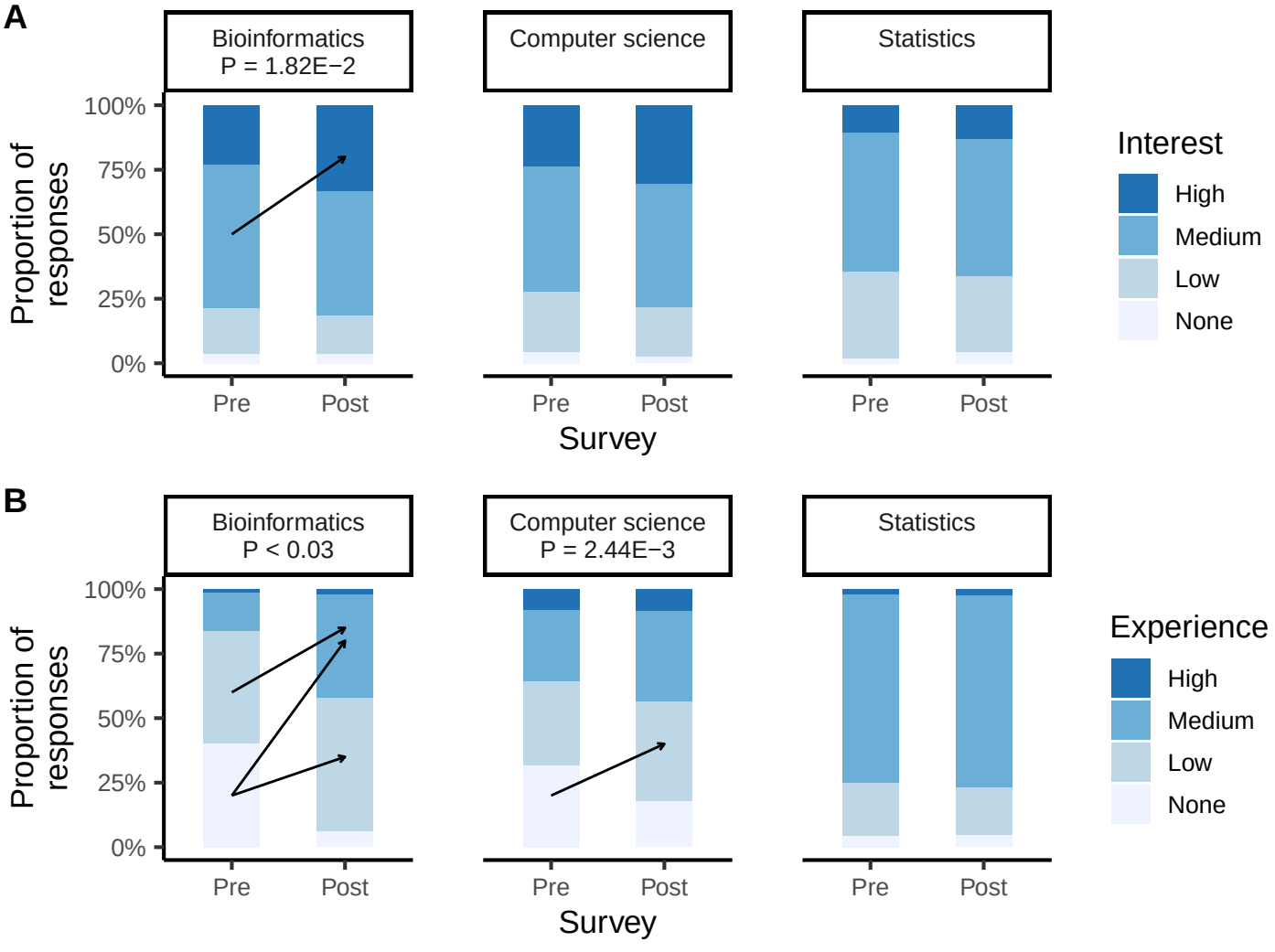
19. Jones F, Harris S. Benefits and Drawbacks of Using Multiple Instructors to Teach Single Courses. College Teaching. 2012;60:132–139.

20. Bess JL. Teaching Alone, Teaching Together : Transforming the Structure of Teams for Teaching. 1st ed. San Francisco, USA: Jossey-Bass; 2000.


Introductory
Advanced
Introduction to data science
- Definitions
- Uses in discipline
Practice with tools
- Data manipulation
- Data visualization
- Statistics
Apply tools to public data
Apply tools to course data
- Discipline-specific
- Messy, imperfect data
- Novel findings
Communicate results
- Schorlarly reports
- Peer-reviewed articles
- Symposium presentations


Introduction to data science
MICB 301
BCIT
- Discipline-specific uses
- Introduction to R/RStudio
- Data manipulation and figures
- t-tests
Additional practice
MICB 322
MICB 323
- Specialized R packages
- Student-generated data
Applications of data science
MICB 405
MICB 425
- Intermediate R/RStudio
- ANOVA and linear regression
- Introduction to Unix
- Capstone projects
MICB 421
MICB 447
- Consultation in R/RStudio
- Consultation in statistics
- Student-generated data
- Capstone projects
Scientific communication
UJEMI
MICB URS
- Publication in the Undergraduate Journal of Microbiology and Immunology (UJEMI)
- Presentation at the MICB Undergraduate Research Symposium (URS)


Faculty of Science
Other
Number of EDUCE team members
0
1
2
3
4
5
6
Undergraduate student
Graduate student
Postdoctoral fellow
Faculty (instructor)
Faculty (research)
Staff
Graduate student
Staff
Career level
Department
Applied Mathematics
Botany
Centre for Teaching, Learning & Technology
Computer Science
Electrical & Computer Engineering
Food Science
Mathematics
Medical Genetics
Microbiology & Immunology
Statistics


A
Bioinformatics
P = 1.82E−2
Computer science
Statistics
Proportion of responses
100%
75%
50%
25%
0%
Pre
Post
Survey
Interest
High
Medium
Low
None
B
Bioinformatics
P < 0.03
Computer science
P = 2.44E−3
Statistics
Experience
High
Medium
Low
None