September, 2019

Data Science Curriculum Group 2019 Report and Recommendations

Major contributors: Phil Austin, Gülnur Birol, Gaby Cohen-Freue, Kim Dill-MacFarland, Steven Hallam, Sara Harris, Rachel Pottinger, Tiffany Timbers, Patrick Walls
Additional contributors *to the discussions leading to this report are listed in Appendix C*
Contact*: Sara Harris, Associate Dean Academic, Faculty of Science*

Executive Summary

University graduates now and into the future will need skills to organize, interact with and extract meaning from data. The recommendations in this document offer a plan that would provide opportunities for all UBC-V undergraduates[1] to develop data skills. These skills are applicable to a wide and expanding range of career paths and will open future opportunities for UBC students, as well as contribute to a data-literate citizenry. This effort will advance our understanding of best practices in data science education, and provide professional development and educational leadership opportunities for faculty and TAs. These recommendations emerged from discussions and surveys between April 2018 and September 2019, involving faculty (largely from Science, but also Applied Science, Arts, and others (via surveys)), staff, and students. The recommendations align with UBC's strategic goals for transformative learning through program renewal, interdisciplinary learning, evidence-based pedagogies, and practice with real world problems. ***Note that these recommendations build on the excellent work done by faculty and departments already. Implementation of these recommendations all require further and ongoing discussion with stakeholders.***

Recommendations:

<u>Curriculum & Pedagogy</u>

1. **Create a data science minor available to any undergraduate**
    - *Develop new courses specifically for this minor, to best enable access by a wide range of students. These new courses could form the core of a future Data Science Major at UBC-V.*
    - *Integrate existing data-driven courses into this framework enabling students from diverse fields of inquiry to participate in the minor*
2. **Create/broaden/expand a foundational course in data science/literacy accessible to any/a broad range of undergraduate students**
    - *Existing course DSCI 100: Introduction to Data Science, or a modified version, is recommended.*
    - *A need for a Data Literacy course may emerge through additional consultation.*

[1] A subgroup is discussing approaches for research-based graduate students, which will be separately summarized.

3. **Infuse data skills throughout the curriculum**
    - *Incorporate skills into new or existing disciplinary courses, course modules and co-curricular activities.*
    - *With this infusion, students will have already learned, and be able to use, foundational data skills in their upper year specialization courses, particularly capstones.*
4. **Use evidence-based, active learning pedagogies, aligned with goals and assessments**
    - *Use and innovate on best practices to become a leader in data science learning.*

<u>**Resources & Tools**</u>

5. **Use open-source tools and resources**
    - *Minimize costs and maximize benefit to students after graduation by using common and standard open-source resources,.*
6. **Align/coordinate tools used across disciplines for data-intensive courses**
    - *Students will be able to use the same tools/languages in multiple courses with emphasis placed on reproducible workflows and best practices*

<u>**Infrastructure, Support & Professional Development**</u>

7. **Expand computational infrastructure and staff support to accommodate increased educational needs for cloud computing**
    - *Increasing numbers of students will need access to, and support for, cloud computing, e.g. JupyterHubs.*
8. **Create and sustain professional development opportunities for faculty, TAs, and staff to learn/share best practices in pedagogy for data science/literacy**
    - *Sharing pedagogical content knowledge across disciplines will benefit undergraduates, graduate training, and research efforts.*
    - *Recruit more interdisciplinary data science instructors*
9. **Create a "Centre for Excellence in Data Science Education" (CEDSE). The CEDSE would:**
    - *bring together people (faculty, staff, students) grounded in different disciplines with different home departments, and centred with expertise from Statistics and Computer Science.*
    - *aim to become an internationally recognized leader in creating and delivering data science-enabled curricula to diverse constituencies.*
    - *form a hub to implement and sustain recommendations 1-8*
    - *Details, and alignment with other units (e.g. departments, Skylight) to be discussed*

<u>**Evaluation**</u>

10. **Evaluate the impact on undergraduates learning data skills.**

Proposed Timeline (academic years):

2019W:

- Review and broaden DSCI 100 to serve foundational data science needs across a wide range of disciplines and Faculties.
- Begin design for one new course for the Data Science Minor.
- Expand (and sustain in subsequent years) computational infrastructure and staff support to accommodate increased educational needs for cloud computing.
- Begin (and sustain in subsequent years) professional development for faculty members who want to incorporate data skills into their existing disciplinary courses.
- Begin evaluation efforts, to be continued for the first three years at a minimum.
- Begin discussions regarding formation of CEDSE.

2020W:

- Implement broadened DSCI 100 with expanded number of sections.
- Continue to develop new courses for Data Science Minor
- Work with domain-specific courses to harmonize data science instruction within the framework of this report's recommendations (e.g. BIOL, EOSC, MICB)
- Establish on-going co-curricular activities including workshops in data visualization, statistical programming, experimental design and modeling, that reinforce classroom learning objectives using domain-specific data sets.

2021W:

- Offer the Data Science Minor, including enough upper-level new courses to provide 18 credits.
- Continue to work with domain-specific courses to expand infusion of data science skills to more disciplines
- CEDSE (or similar) formed by this time.

Beyond 2021W:

- Plan for Data Science Major
- Plan for a cross-disciplinary capstone experience.

Resources:

To fulfill these recommendations will take commitments of time and money. Faculty members and departments are the drivers of curricular change and at the forefront of shifts in our students' educational needs. Faculty members are already applying for UBC grants to moe curriculum changes along (e.g. TLEF sources). Hiring of new data-science-oriented faculty members is a priority in the Faculty of Science. Providing these learning experiences for UBC students creates one category of new demand for high performance computing (HPC) for education, among other applications across campus. We anticipate that a new investment by UBC in HPC for educational purposes will be needed.

Aims within ~5 years, with ramp-up for each item:

- *Foundational course serving 2000 students per year in steady state (or larger)*
- *Data science minor serving 150 students per year (or larger)*
- *20 (or more) discipline-specific courses incorporating data science skills, broadly aligned*
- *Cross-disciplinary capstone experience for 100 student groups*
- *Training for 20-30 faculty members and 40-50 TAs per year (or larger)*
- *~4700 course enrolments impacted per year in steady state (or larger)*

Background

Skills in working effectively with data have emerged as critical to successful participation in a wide range of disciplines including at least science, humanities, engineering, medicine, and business (e.g. https://www.hiringlab.org/2019/01/17/data-scientist-job-outlook/). Examples in which strong data skills are key include: designing real-time responses to epidemics, tracking census trends and human migrations, global climate modeling, personalized healthcare, identifying historical trends, and linguistic analyses. Successful research in many fields involves data skills, and, increasingly, working with large data sets. Although the tools and approaches used across disciplines may vary widely, there are some core concepts, skills, and ethics around working with data that are common across fields.

Many universities have developed data science educational opportunities in the past few years, including for undergraduate, graduate, and continuing education populations (Appendix A). At UBC, the 10-month Master of Data Science program has grown quickly since its inception in 2016, and graduates of that program are successful in obtaining employment post-graduation. UBC-O now offers a Data Science major and minor for undergraduates. UBC-V undergraduates could benefit from similar opportunities.

Despite efforts to integrate data science into undergraduate curriculum in some courses, a unifying teaching and learning framework does not yet exist at UBC-V. If undergraduate students had opportunities to develop core data science skills early in their university years and were then able to use and expand them in later years within their discipline, they would graduate better able to think through problems involving data and do the data wrangling and analysis needed to make decisions with data and explain it to others. They would have practice with state-of-the-art data science tools in the context of their in-discipline knowledge which would serve them well after finishing their UBC degrees.

This document proposes a plan to create a unifying data science framework for undergraduate teaching and learning, and has the following scope & goals:

- Define a coordinated curriculum structure (or structures) to provide appropriate opportunities for students to acquire needed data science skills
- Estimate the resources needed to develop this structure and sustain it long term
- Outline a possible implementation plan, with the assumption of ongoing discussions and collaborations across the university

Toward these goals, the report addresses the following questions:

- What data science skills should UBC-V undergraduates have when they graduate?
- Where do UBC-V undergraduate students currently gain data science skills, if they do?
- What curricular model(s) could most effectively provide appropriate opportunities for all UBC-V undergraduates to obtain data science skills useful to them?
- What would it take to create these opportunities?

Definitions

While **data science** (and data literacy) have various definitions, we define "data science" as ***"the act of extracting value from data using reproducible and auditable processes"*.**

In more detail, data science includes transparent and reproducible activities to find and manage relevant data (big or small) to answer complex questions, extract information to make meaningful interpretations and decisions from data, and communicate information from data to different audiences. Each of these aspects can be explored in varying depth.

What data skills should UBC-V undergraduates gain?

Through a series of workshops and discussions (details in Appendix B), Science faculty members, the Science Undergraduate Society, Skylight, as well as faculty members from Applied Science, Arts, and Medicine (Appendix C), defined nine data science themes that represented the ideal set of skills a data literate student would have upon graduation:

1. Communication of data and interpretations
2. Ethical data use
3. Using software (e.g. SPSS, Excel, open-source software…)
4. Mathematical foundations and reasoning (e.g. linear algebra, calculus, probability...)
5. Computational foundations and reasoning (e.g. algorithms, programming...)
6. Statistical foundations and reasoning
7. Scientific reasoning & process, in context
8. Data management
9. Data visualization

These nine themes align well with those defined by the National Academies of Sciences, Engineering, and Medicine (2017, 2018), and the results of a needs assessment survey regarding training for bioinformatics (Attwood et al., 2019). These can form a framework for building comprehensive data science curricula.

Regarding whether these themes resonate across campus, in March/April 2019, we conducted an open survey of UBC-V faculty members and received responses from Arts, Commerce, Education, Kinesiology, Pharmaceutical Sciences, and Science regarding 52 undergraduate courses (24 in Science, 28 in other Faculties). The survey is available at https://ubc.ca1.qualtrics.com/jfe/form/SV_dbBW0pIolmxSVNz. In that survey, each of these themes was rated by the majority of faculty respondents as "important" or "very important" for a data-literate graduate in their discipline, although non-Science faculty members rated "Mathematical foundations…" and "Computational foundations…" as less important than respondents from Science (Appendix D).

We also surveyed students graduating from the Faculty of Science in May 2019 regarding these themes and their importance for a data literate person in their discipline. Although the response rate was low (n=16), recent graduates also rated these themes "important" or "very important", with "Mathematical foundations…" and "Computational foundations…" receiving the most "Not important" responses (Appendix E).

Where do UBC-V undergraduate students currently gain data skills, if they do?

Course-based opportunities in many disciplines exist for undergraduate students to acquire data skills, and we offer specializations in statistics and computer science. As yet, there are no undergraduate specializations or credential opportunities specifically in data science. Information about existing courses that incorporate data skills comes from (1) the Data Science Curriculum Group (Appendix C), which generated a list of courses that teach data skills in some way, and (2) the faculty survey mentioned above, with responses from Arts, Commerce, Education, Kinesiology, Pharmaceutical Sciences, and Science. The current situation is that faculty and students are responding to the need for students to acquire data science or literacy skills, in particular disciplines. Existing opportunities at present include:

1. DSCI 100: Introduction to Data Science was offered for the first time in 2018WT2. This course is a potential foundational course for a coordinated curriculum effort.
2. The computer science and statistics departments offer many courses that are appropriate for their majors/honours students but are inaccessible to other students because of long pre-requisite chains.
3. Many departments beyond Statistics and Computer Science teach courses that incorporate data skills (e.g., the task of classification is taught in commerce, geology, earth and ocean sciences, and other departments). At least 52 undergraduate courses contain components that faculty teaching them consider to be "data literacy".
4. There are currently upper-level courses at UBC incorporating fairly low-level data skills because those skills are needed and not acquired earlier. As data skills become increasingly important, we can expect more courses will do this, unless there is coordinated and collaborative action to consolidate some of that effort.
5. There are several NSERC CREATE and TLEF-related initiatives focused on data science teaching and learning in the Life Sciences (e.g. The Experiential Data science for Undergraduate Cross-disciplinary Education (*EDUCE*) initiative)

Based on the faculty survey responses, all nine themes are addressed in disciplinary courses across UBC-V. The themes most commonly addressed are "Communication of data…" and "Scientific reasoning…". The theme that is least addressed is, not surprisingly, "Ethics around data use" (Appendix D).

Surveys from the EDUCE initiative indicated limited familiarity with data science among MICB undergraduate students prior to EDUCE modules. After course modules, students showed increased interest and experience in data science topics like "bioinformatics" and "computer science" (paper in preparation). Thus, EDUCE demonstrated achievement of one of the project's fundamental goals: to expose undergraduate students to introductory data science within the context of current MICB courses and curriculum (see recent presentation for further details: https://github.com/EDUCE-UBC/presentations).

In our current state, some undergraduate students are likely getting good practice with data skills, in context, depending on which courses they take, or other experiences they have (e.g. Co-op, work, or research opportunities). Some are learning basic data management skills in upper level courses. Faculty open-ended responses to the survey indicate time could be better spent in upper level courses on discipline-specific problems if students arrived in those courses with stronger data science skills. Other students are likely graduating with little experience learning data skills. High enrolments of Year 3 and Year 4 students in the first three offerings of DSCI 100 imply latent demand for this experience. This enrolment pattern likely is transitional, and as more students access a foundational course in their early years, enrolment in the foundational course should shift toward Year 1.

What curricular models could be effective?

Our goal is to provide effective data science educational opportunities for large numbers of UBC undergraduates, not only for students with data-intensive specializations. For curricular models, we looked at examples external to UBC (Appendix A), UBC-driven teaching and learning initiatives, and used survey data within UBC.

Many universities already have major and/or minor undergraduate programs in Data Science. Some also have broad data science courses introduced before specialization, which can serve both as a foundational gateway course and also as the only data science experience for some students.

Given the diversity of UBC, one possible path is that each specialization develops and teaches data science skills, starting from the very basics. The advantages of this approach are that students will learn the specific skills needed in the context in which they specialize and any uniqueness in that discipline will be part of that education from the beginning. The disadvantages of this approach are that it leads to situations in which low-level data skills are taught in upper-level courses because students don't have the skills they need yet. This limits opportunities for more data-intensive problem solving or course-based learning applications. In addition, students may not grasp the generality of the intellectual ideas in data science. This approach also leads to replication of effort and discontinuities in best practices.

A second approach is to coordinate among disciplines and provide broad-based lower-level learning opportunities in data science that will serve the needs of many disciplines. The advantages of this approach are consistency among fundamental data skills taught, opportunities to hone pedagogical content knowledge specifically around the learning of data science skills, and the opportunity for faculty teaching upper level courses to go farther in their disciplinary courses because their students already have some foundational capacity. The challenge with this approach is defining the learning goals for foundational experiences so that many disciplines are indeed well-served. **We think this second approach is the one to pursue, for consistency, for development of data science pedagogical expertise, and to reach the most students.**

The curricular structure proposed in this document is similar to, but clearly differentiated from, that implemented at UC Berkeley:

- A foundational course with zero prerequisites and open to all students.
- "Data-enabled" courses that connect the foundational course with discipline-specific courses.

We are not proposing what are called "connector courses" at Berkeley which are designed to be taken concurrently with the foundational course because (1) UBC's curriculum approval process makes this impractical, and (2) such courses are difficult to maintain (based on Berkeley's experience). Instead, datasets from many different disciplines can be incorporated into the foundational course with input from faculty disciplinary experts.

Curricular Structure Recommendations

Based on a survey of existing approaches and information from the UBC community, we recommend the following (Figure 1):

1. A foundational data science course that may satisfy the needs of those wishing to learn only the basics of data science, and also serve as a prerequisite for higher-level courses using data skills.
2. A minor that is accessible to students from many majors that would allow those wishing for greater depth to have the classes that they need (Details in Appendix F)
3. Opportunities for students who wish to fall somewhere in the middle to take courses that are more accessible to them than existing computer science and statistics offerings. These needs would be met with courses designed for the proposed data science minor.
4. Infuse data science into discipline-specific courses and specializations, extent to be determined by faculty and departments.
5. Interdisciplinary collaborative capstone experiences in which teams of students from different disciplines collaborate on a data-intensive project. This would benefit student careers, prospects for graduate school, and UBC research labs.
6. Use common tools throughout the flow of courses (e.g. Python, R, Jupyter Notebooks, Github) and share materials.
7. Develop and support co-curricular activities

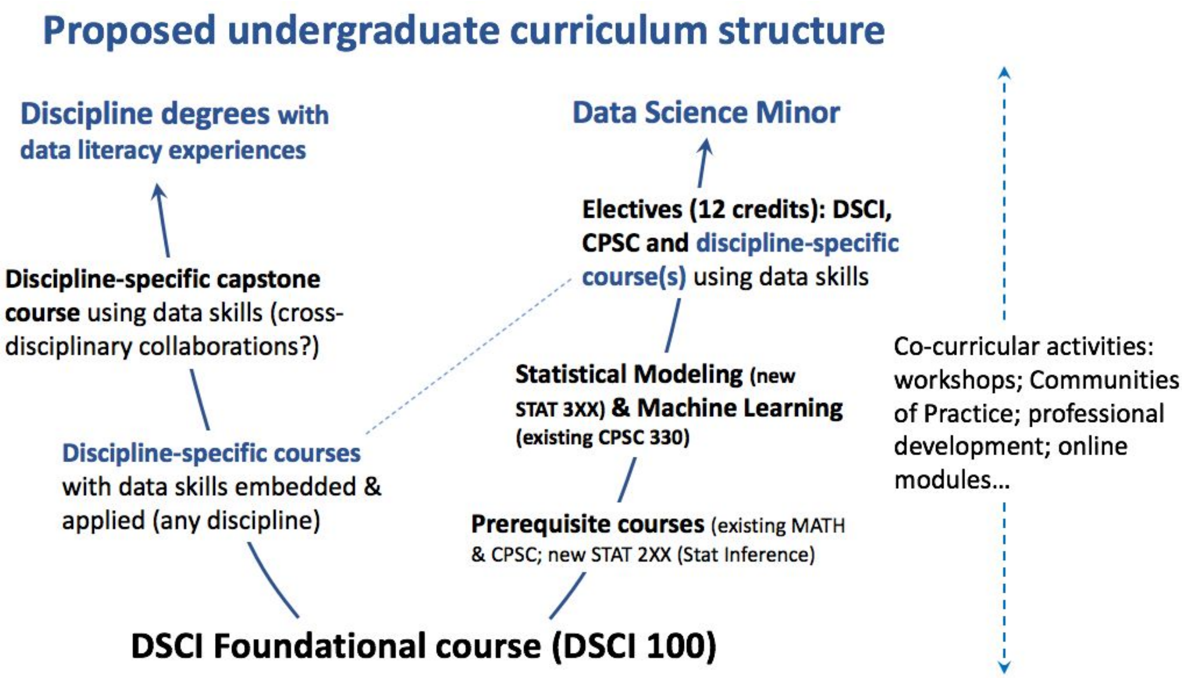


Figure 1: Proposed general model for undergraduate data science curriculum

The proposed Data Science Minor allows any student from across the university to add this credential to their degree. The new courses proposed for the minor will form the core of a future Data Science Major. The Data Science Curriculum Group thought that additional credentials (e.g., a certificate) would not currently add value beyond the options above. For students doing fewer data-science-related courses than the minor, prospective employers would be able to see courses taken, on a transcript, and/or student work (e.g. a portfolio) without need for an additional credential.

Recommendations for the foundational course

Survey responses from faculty indicate that having a foundational data science course would provide students opportunities to achieve more in many (not all) existing higher-level courses that rely on, or currently teach, data skills. It would allow faculty to “save a lot of time”, “go into greater depth”, and “teach more specialized topics”.

A question remains regarding what should be in the foundational course in order to serve many students from many disciplines well. As part of the faculty survey in March/April 2019, we asked faculty to categorize the current DSCI 100 learning goals as whether they should be in a foundational course, a higher-level course, a stand-alone resource, or didn’t need to be addressed at all (results in Appendix D). Those most highly rated to be appropriate for a foundational course are:

- Download and scrape data off the World Wide Web.
- Wrangle data from their original format into a fit-for-purpose format.
- Create, and interpret, meaningful tables from wrangled data.
- Create, and interpret, impactful figures from wrangled data.
- Accomplish all of the above using workflows and communication strategies that are sensible, clear, reproducible, and shareable.

Some more specific DSCI 100 learning goals were rated by many as appropriate for a higher-level course. Based on these results, we recommend:

- Further discussions about the goals of a foundational course, to inform modifications, broadening, and whether there should be 2 foundational courses. These discussions began in August 2019, with >80 people across campus invited to join a discussion (>30 participated).
- Modification of DSCI 100 to align well with needs across campus.
- Offering a modified DSCI 100 as the foundational course (or one of the two).

We also propose that instructors and teaching assistants from multiple disciplines collaborate in teaching different sections of the foundational course, to be actively involved in the community teaching the foundational course and to bring authentic problems from their disciplines. Students would get credit for the foundational course as a pre-requisite even if there were multiple “flavours” of the course.

Recommendations for the Data Science Minor

A scan of more than 30 existing “Data Science” majors and minors in North America show that some are largely compilations of courses from existing units (e.g. computer science, statistics, engineering, math…) and others are structured with new, specifically designed “data science” courses. These imply two possible models: (1) an approach where courses that exist for other reasons are used to form a larger initiative (e.g., a minor) or (2) a coordinated approach where representatives from multiple disciplines come together to discuss what is needed in a data science curriculum. Speakers at a conference of statistics education faculty communicated that those who choose the first approach are generally less satisfied with their eventual offering and had far less flexibility to modify the program or courses within it when needed. Based on this feedback, we recommend a purpose-built Data Science Minor (details in Appendix F), albeit one that builds on experience and materials designed for other contexts.

The proposed data science minor aligns with and covers the nine themes identified for a data-literate graduate. It requires some pre-requisite courses, such as the foundational course, computer science and math, followed by two required courses (Statistical Modeling and Machine Learning), then a suite of offerings among which students may choose in order to complete their 18 credits of upper level courses. Importantly, an in-discipline data-intensive course within a student’s major specialization may count toward the minor.

Recommendations for infusing data science into existing specializations

Decision-makers for each specialization are the ones who would decide whether to introduce data science into disciplinary courses and specializations. Faculty in Microbiology and Immunology have piloted incorporating data science modules in courses, through EDUCE (Experiential Data science for Undergraduate Cross-Disciplinary Education) with co-curricular data science workshop support through the NSERC CREATE training program ECOSCOPE. Earth, Ocean and Atmospheric Sciences is planning to incorporate data science, and data science-specific tools and pedagogies into >10 courses. Biology plans to incorporate data science skills into all lab courses such that students progressing through biology specializations practice data skills multiple times.

We recommend that collaborative structures, e.g. a Centre for Excellence in Data Science Education (CEDSE), could provide opportunities for sharing innovation and best practices that build on and expand these departmental efforts into an integrated community of practice.

Recommendations for interdisciplinary capstone experiences

This idea needs development by faculty members who are keen to collaborate with those in other disciplines to provide teams of students with authentic research experiences involving data science. The Masters of Data Science program has relevant experience with capstones, which will be beneficial to developing interdisciplinary capstones for undergraduates. Future discussions are needed to flesh out this idea.

Recommendations for co-curricular activities

Existing co-curricular activities such as statistical programming and data visualization workshops, hackathons and bootcamps are ongoing and new ones can and should be defined. Useful activities would include: communities of practice; speaker series; events that connect students to employers. Co-curricular activities are an excellent way to test co-curricular modules designed for integration into more intensive data science courses and help socialize the learning experience across disciplines. Some Co-op positions will provide students with highly valuable experience in data science, and students who have a head start with data science before going on Co-op may have a better chance of landing those jobs.

Pedagogical Recommendations

We recommend adoption of evidence-based active learning strategies and coordination of pedagogical practices and learning tools across data-science-intensive courses. Currently, DSCI 100 uses a flipped classroom approach with decreased class time. Worksheets, problem sets, case studies, and group projects give students hands-on experience with the skills they are learning and opportunities to iterate. Students receive timely, sometimes immediate, feedback, with considerable auto-grading. Learning activities are set up such that scaffolding is prevalent toward the start of term, and gradually removed to support students becoming autonomous. Active learning is encouraged throughout the learning process, especially in the context of group projects that focus on domain-specific knowledge. Students have many opportunities for practice and development throughout with the integration of co-curricular activities tuned to reinforce learning objectives within the data science core curriculum.

What would it take to create these opportunities?

Both resources and principles/policy will be needed to offer these opportunities to learn data skills. The goal to build specific data science pathways (e.g. the minor) plus infuse data science through specializations in domain-specific courses, using common tools, will require a good deal of cooperation and coordination among different units. Statistics and Computer Science have taken leadership roles in data science education at UBC-V in general, and house much expertise in this area. Our overarching recommendation is to create a collaborative community hub for this work, a Centre for Excellence in Data Science Education (CEDSE). A recommendation is for the initial core of this group to include five Educational Leadership faculty members, who would each have specific departmental homes. They would design, build, and offer courses in the data science minor in collaboration with other faculty with links to data science. They would also contribute to teaching data-intensive courses in their home departments, support graduate education, and offer professional development for faculty and TAs in data science. CEDSE is envisioned as involving people beyond core EL faculty members, and its structure would need considerable future discussion. Skylight, the Science Centre for Teaching and Learning, has extensive expertise in effective pedagogy and professional development, from which CEDSE could benefit.

Specific categories of resources and principles/policy include:

- Course development and revisions
- Instructional resources (faculty, TAs, course coordination)
- Computational infrastructure and support
- Physical infrastructure: data science learning spaces
- Professional development for faculty and TAs
- Evaluation resources
- Curricular principles to be consulted as new curriculum proposals come forward.

Course Development and Revisions

This proposal involves revising DSCI 100 to become the/a foundational course, design and development of new courses for the data science minor, and redesign of existing courses to infuse data skills into domain-specific courses. This course development would require

- faculty time (and course buyouts)
- GTA support
- professional development for faculty to learn to teach with open source tools, e.g. Jupyter, Python/R, and Git with nbgrader and Canvas integration (see professional development below).

Additional resources would be needed for:

- Legal support for licensing, IP and data security, aiming for open access for course materials and student work.
- Resources/time for faculty to contribute to open source data science community

- Travel support for faculty & staff to network with data science communities, to build and maintain open source tools
- Instructor training to design and implement new course designs (see professional development below).

Instructional Resources

The numbers in this section are ballpark recommendations, recognizing that specific departments would need to develop courses and ramp up capacity over time. In these recommendations, new course offerings include additional sections of the foundational course, plus each of the new courses for the data science minor, for an estimated total of 17 new sections, after ramping-up.

- Offering a foundational course to serve about 2000 students per year, would require 12 lecture sections of 180 students, and the equivalent of 58 tutorial sections of 35 students. This is 10 more sections than currently offered per year for DSCI 100. Instructors and TAs would be needed for these. Ideally, both faculty and graduate students from multiple disciplines could be involved in this course. A plan for ramp-up will be needed.
- Offering 7 new courses for the data science minor will require one section each (150 students for 2xx courses and 100 students for 3xx and 4xx courses)

Some of the sections of the foundational course may be taught by faculty from any domain, in collaboration with the foundational course team. The net number of new course sections is approximately 17. Accounting for sabbatical years, this is approximately the equivalent of five EL faculty members with course loads of 4 (plus other duties), which aligns with the proposed hires for the CEDSE.

Computational Infrastructure and Support

Cloud-based computational infrastructure that can handle a few thousand concurrent users is critical for this proposed curriculum to succeed. Decisions about how to set up, access, and support this infrastructure need to be made and discussions are ongoing. Below, we describe current infrastructure resources at UBC, and a possible model for the future.

Current resources: Jupyter and Syzygy

The Jupyter Notebook is an open-source web-application for interactive computing. JupyterHub runs in the cloud and hosts Jupyter notebooks for a community of users. Syzygy is a collection of JupyterHubs deployed by the Pacific Institute for the Mathematical Sciences (PIMS) with authentication by university credentials. PIMS launched Syzygy in 2016 to serve researchers and students at UBC and has now grown to over 20 JupyterHubs serving universities in Canada and the US with 15,000 users (with 5000 users at UBC alone). For example, anyone at UBC with a CWL can login to ubc.syzygy.ca, open a Jupyter notebook and start computing in Python or R. PIMS (in partnership with Cybera and CanCode) also launched Callysto — educational modules for K-12 in Alberta and a JupyterHub to host Jupyter notebooks.

Current and Projected Jupyter Users at UBC

There is a huge demand for accessible, cloud-based computational tools in undergraduate courses at UBC. Jupyter notebooks with Python/R are already being used in many courses in MATH, DSCI, CPSC, ECON, EOAS, and CHBE, as well as student workshops. Syzygy is the primary tool for several of these courses. There are at least 3 different use cases for Jupyter at UBC:

1. *JupyterHub for all*. The JupyterHub ubc.syzygy.ca is available to all with a CWL. It is equipped with standard Python and R packages for machine learning, data science, data visualization, and more. It is suitable for novice users doing moderate computations. The service is being used in MATH 210 (Introduction to Mathematical Computing), CPSC 103 (Introduction to Systematic Program Design), EOSC 350 (Environmental, Geotechnical, and Exploration Geophysics I), and more.

2. *Custom JupyterHubs for Teaching*. DSCI 100 is a new foundational course designed by Tiffany Timbers in collaboration with faculty in the Department of Statistics. The course has a very structured architecture including a JupyterHub server, nbgrader server and GitHub enterprise server which are all integrated with Canvas. The specific structure removes all complexity for users where content and assessments are pushed and pulled seamlessly behind the scenes.

3. *Custom JupyterHubs for Research*. Collaborative research groups require more computational power and specific kernels and packages for their work.

With the increasing push to provide data science training for students at UBC, it is not unreasonable to expect that there will be 30,000 Jupyter users at UBC (1/2 of the undergraduate population) in the coming years. At present, ubc.syzygy.ca serves 5,000 users and can handle approximately 500 simultaneous users. The Pacific Institute for Mathematical Sciences (PIMS) is relaunching Syzygy in Fall 2019 to scale-up to serve 10,000 users at UBC with 1000+ simultaneous users, as a short-term solution to accommodate near-term growth. The requirements below are based on estimates by PIMS to meet these projections. Sources and systems for long-term funding and support is an on-going discussion.

Requirements (short-term, 2019W)

- Update hardware/cloud computing to serve 1000+ concurrent users

Requirements (long-term)

- Cloud-computing resources, per student
- Salaries for IT support engineers
- LT staff support
- LT Hub Rovers
- GitHub Enterprise license

Physical Infrastructure: Data Science Learning Spaces

Design Features of Data Science Learning Spaces

- Robust WiFi
- Table surfaces
- AC power at all seats
- Multi-input, multi-output audio-visual system including wireless screen sharing
- SCALE-UP design (see Student-Centred Active Learning Environment for Undergraduate Programs)

UBC Learning Spaces provides a searchable database of learning spaces at UBC. Examples of data science learning spaces which currently (or will soon) exist at UBC:

Room	Occupancy
UCLL 109	30
LSK 121	64
ESB 2012	80
DMP 310	160
DMP 301	80
DMP 110	120
HENN 200	180
DL 005	60
BIOL 1000	240
ORCH 4074	72

Note - ORCH 4074, LSK 121 & HENN 200 are the most desirable rooms on this list due to their pod-like table set-ups that enable and promote collaborative learning. Collaboration is a key component in Data Science and how to work collaboratively is explicitly taught in Data Science courses. Also, not all rooms on this list have wireless screen sharing capabilities

Recommendations and Requirements

- Designate select spaces for data science courses that meet the design features listed above.
  - HENN 200 and BIOL 1000 for 2000 students in foundational course
  - DMP 110, DMP 301 and DMP 310 for 100-150 students in Data Science Minor

    - Similar spaces for 1000 students in discipline-specific data-enabled courses
- Data science spaces for 30 sections of data enabled courses including foundational course, Data Science Minor and domain specific courses
- Work with Classroom Services to gain priority access to rooms identified as Data Science Learning Spaces
- Work with Learning Spaces team to add features to existing spaces to transform into data science learning spaces
- Work with UBC IT to upgrade WiFi bandwidth for data science learning spaces

Professional Development for Faculty and TAs.

Expanding data science in the undergraduate education space at UBC will provide many professional development opportunities for faculty and TAs, such as:

- Redesign existing courses
- Develop new courses
- Participate in collaborative communities to advance data science learning
- Mentoring
- Redesign undergraduate specializations
- Contribute learning materials to the broader community
- Educational leadership

To support instruction in data science intensive courses, we recommend:

- Instructor and TA training programs:
    - Delivered through CTLT Institutes and organized similar to CTLT Course Design Intensive
    - Modelled on UC Berkeley's Data Science Pedagogy and Practice Workshop
    - Learning goals:
        - Python/R programming
        - Design learning activities in Jupyter notebooks
        - Deploy course content in GitHub repositories
        - Design assessments with nbgrader
        - Design for Canvas integration
        - Select open source license for course materials
- Continuous support:
    - Delivered through LT Hub
    - Canvas, Jupyter, Python, R, nbgrader and GitHub integration

Requirements (continuing)

- Educational programmer/LT Hub staff member for materials design support and scheduling/logistics.
- Instructors to design workshop activities and materials
- Instructors each cycle to deliver workshop

Evaluation Plans

Introduction to Data Science (DSCI 100)

DSCI 100 is the first undergraduate course at UBC with the potential to serve as a broad introduction to data science for students in the Faculty of Science. To assess if DSCI 100 should serve as this foundational course in its entirety or with modification, pre-, mid-, and post-surveys were administered to students in DSCI 100 in 2018/19 (N = 59). These surveys of the inaugural run of the course assess student knowledge, interest, and confidence in data science as well as specific course logistics like workload and difficulty. Responses show that students' confidence in their abilities to use R strongly increased, and show what topics students found most valuable. Course observations was also completed by Alice Campbell (UBC Science Education Specialist, CPSC). Select results from surveys and observation are included in Appendix G of this report. Surveys will continue in this course during 2019W terms.

Comprehensive, Faculty of Science

The overarching goal for project evaluation is to capture student learning and experience in DSCI 100 and newly developed data science courses to inform course effectiveness during pilot offerings, and to inform further revisions. So far, surveys of faculty across campus, and graduating BSc students have informed curricular recommendations.

Student learning will be assessed by course specific assessment tools (e.g., assignments, exams) that will be created as part of the course development work. Student experience in terms of their expectations, interactions with course materials and specific activities, and their satisfaction with the course(s) will be captured through carefully designed evaluation tools such as surveys or focus group questions. We will use validated evaluation tools when/if available, otherwise we will develop homegrown tools. We anticipate tailoring the evaluation tools for each course to evaluate their unique contributions to student experience.

We will also evaluate the effectiveness of the faculty and TA professional development workshops to promote implementation of best practices (e.g., hands-on analysis of real-world data and using state of the art techniques/tools) and to uncover further needs around professional development and support. The process that we go through developing and implementing this curriculum in an interdisciplinary setting presents itself as an opportunity to learn. To this end, we will attempt to capture the process of putting together a program of this scale to document lessons learned and to share with the broader community.

Resources for evaluation:

- Educational programmer for evaluation
- GRAs to help with data analysis

Curricular Principles

There is presently keen interest in data science across UBC. To help support building a cohesive and collaborative data science curricular structure, we propose the following (draft) curricular principles:

Encourage:

1. Revision of existing domain-based courses to use/build on data skills from the Foundational course.
2. Creation of domain-based courses to use/build on data skills from the Foundational course.
3. Revisions/creation of capstone courses that use/build on data skills learned earlier in a student's program.
4. Interdisciplinary capstone courses that bring together students from different disciplines, including inter-departmental collaborations.
5. Use of open source, accessible tools
6. Revisions of specialization requirements to include courses in which students learn data skills. This means deciding what to remove from specializations to make room.
7. Cooperation & consultation across departments about curriculum/courses.
8. Courses which integrate instruction in important but non-technical aspects of data science, such as ethics, law, sustainability, scientific method, public communication, etc. More information and discussion is needed in these areas.
9. More flexibility and elective space in specializations, more opportunities for students from different specializations to interact in formal coursework.

Discourage:

1. Use of proprietary tools
2. Domain specific upper-level courses in which a significant fraction of the instruction duplicates basic programming, statistics and/or mathematics available in lower-level courses, recognizing that until a data curriculum structure is in place, then we need to support students in what they need to learn, when they need it.

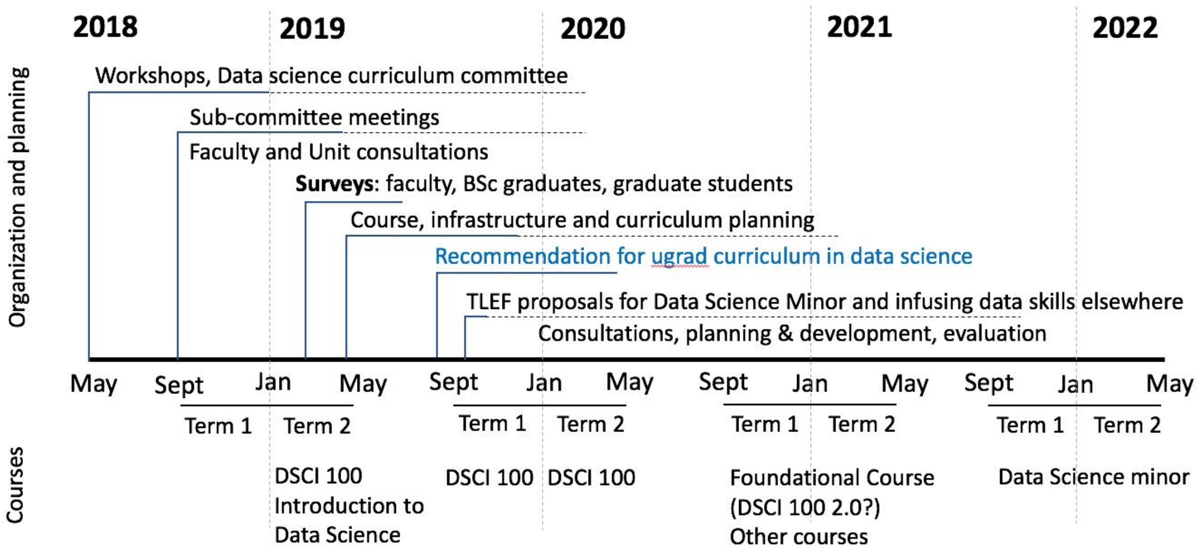
Next steps

- Two new and one returning large TLEF proposals will be submitted in October 2019:
  - (1) New: Development of the data science minor (PI: Tiffany Timbers)
  - (2) New: Infusion of data science in EOAS courses (PI: Tara Ivanochko)
  - (3) Returning: EDUCE (PI: Steven Hallam)
- Further work planning the data science minor:
  - Map learning objectives for the minor to proposed courses and to 9 data science themes.
  - Take new courses through the curriculum approval process during 2019W

- Continuing consultation regarding DSCI 100, evaluating its applicability to students from a broad range of disciplines, building on work done in summer 2019.
- Continued evaluation of student experience and learning in DSCI 100.
- Identify a long-term strategy for computational infrastructure to support data science education broadly across campus.
- Identify/secure funding for new faculty, instructional costs, computational infrastructure, staff support.

Timeline

The timeline below shows events since spring of 2018, and anticipated future activities through the launch of the data science minor in 2021W Term 1.



Appendices:

Appendix A: Data Science curriculum examples in North America

Examples of data science offerings for undergraduates in Canada and the US.

Undergraduate Data Science in Canada

- UBCO
  - BSc Major in Data Science
    - Curriculum: DATA 301, DATA 311, DATA 405, DATA 407 + 25 courses from MATH, STAT, CPSC, … + electives
  - DATA 101: broad, no-prerequisites, open to all Science students, not required by Data Science program
  - Note: Major program introduces DATA 301 in 3rd year

- University of Calgary
  - Minor in Data Science
    - Curriculum: DATA 201, DATA 211, DATA 305, DATA 311, DATA 501 + 5 courses from MATH, STAT, CPSC, …
  - DATA 201: broad, no-prerequisites, open to all Science students
- University of Toronto
  - Data Science Specialist (BSc, Hon.)
    - Curriculum: JSC270, JSC370, JSC470 + 23 courses from MATH, STAT, CPSC, … + electives
  - JSC270: 2nd year data science course with several prerequisites, restricted to Data Science Specialist program
  - Note: Specialist program introduces JSC270 in 2nd year

Undergraduate Data Science in US

- UC Berkeley
  - Major in Data Science
    - Data8: Foundations of Data Science with connector courses
    - Data100: Principles and Techniques of Data Science
    - Standard courses in MATH, STAT, CPSC, …
    - Electives
  - Data-enabled courses are domain-specific courses which align with Data8 and Data100 (but do not require them as prerequisites).
- Other: UW, Stanford, Penn, UCSD

Appendix B: Workshops & discussions, 2018-2019

In April 2018, ~30 faculty members and one student from all departments in Science plus a few faculty members from Applied Science, Arts, and Medicine (Appendix F) came together in two workshops to brainstorm and summarize thoughts about the ideal set of data skills a BSc student would have on graduation. Through that process, and later discussion, we defined nine themes:

1. Communication of data and interpretations
2. Ethics around data use
3. Using software (e.g. SPSS, Excel…)
4. Mathematical foundations and reasoning (e.g. linear algebra, calculus, probability...)
5. Computational foundations and reasoning (e.g. algorithms, programming...)
6. Statistical foundations and reasoning
7. Scientific reasoning & process, in context
8. Data management
9. Data visualization

Following these workshops, we established a smaller ad hoc committee (~15 people) with at least one representative from each Science department (2 each from Computer Science and Statistics), plus representation from Skylight and interested faculty members from Applied Science and Arts. This group met monthly until June 2019. Our goals were to:

- Define a curriculum structure (or structures) to provide appropriate opportunities for students [in Science, or more broadly] to acquire the data skills they need. Given this structure:
    - Estimate resources needed to develop this structure initially
    - Estimate resources needed to support this structure in the long term
    - Plan for making opportunities available for people to participate in development and teaching
    - Outline an implementation plan

To guide the work, we defined a logic model and formed five subgroups:
- Planning, curriculum principles, communications
- Course content
- Infrastructure needs (computational, physical, people support)
- Evaluation
- Graduate education (as graduate students have different needs, this subgroup is not included in the remainder of this report

These subgroups met separately and reported back to the ad hoc committee.

Appendix C: List of contributors

This list includes people who attended one of the two workshops, or at least one of the meetings after the workshops. People in **bold** are those most consistently involved in discussions from 2018-2019 on the Data Science Curriculum Group.
(Apologies if someone who was at one of these meetings has been inadvertently left off this list - let me know and I'll add you)

In addition to those people listed below, faculty, graduate and undergraduate students from across UBC contributed ideas and experiences via surveys.

Name	**Department**	**Contributor to April 2018 Workshops**	**Contributor to ongoing discussions 2018-2019**
Amanda Giang	Institute for Resources, Environment & Sustainability	x	x
Anne Condon	Computer Science	x	
Colin Gay	Physics and Astronomy	x	x
Curtis Berlinguette	Chemistry	x	
Diane Srivastava	Zoology	x	

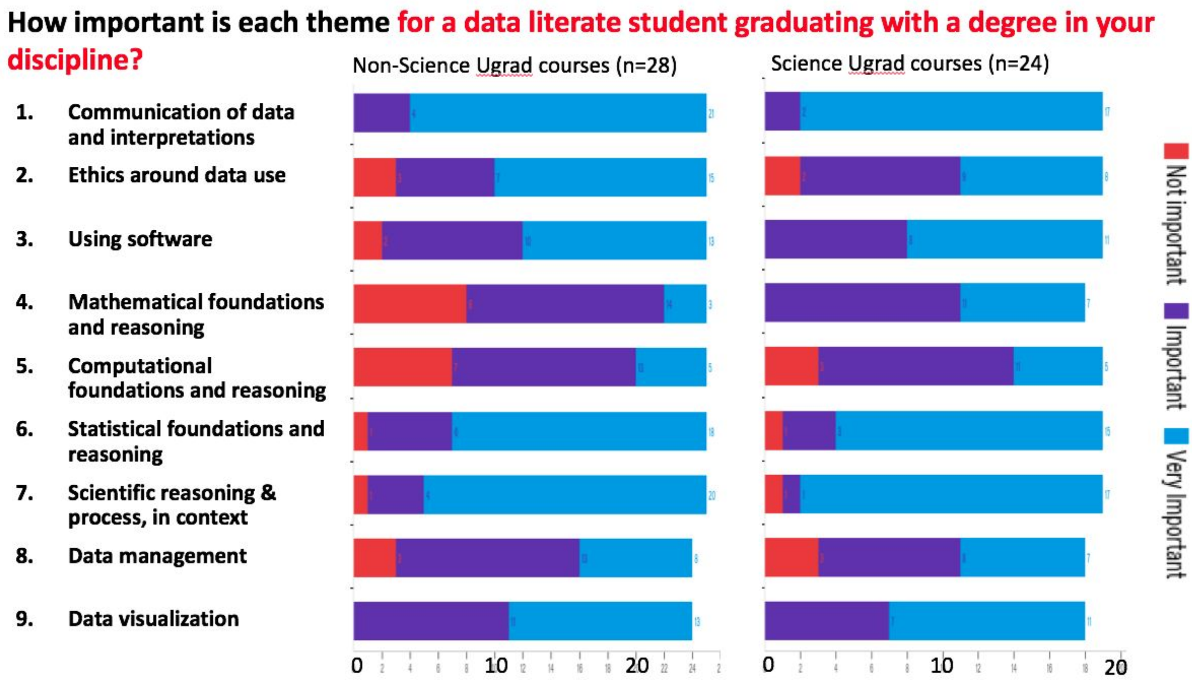
Eldad Haber	Earth, Ocean & Atmospheric Science	x	
Elvin Wyly	Geography	x	x
Gaby Cohen-Freue	Statistics	x	x
Guiseppe Carenini	Computer Science		x
Gulnur Birol	Skylight: Science Centre for Learning and Teaching	x	x
Harish Krishnan	Sauder School of Business		x
Harvey Richer	Physics and Astronomy	x	
Ian Mitchell	Computer Science	x	x
Julia Chai	Science Undergraduate Society	x	x
Karen Cheung	Biomedical Engineering	x	x
Keith Adams	Botany	x	x
Kim Dill-McFarland	Microbiology and Immunology	x	x
Kurt Haas	Cellular, Anatomical, and Physiological Sciences		x
Liang Song	Botany		x
Luke Bergmann	Geography		x
Marie Auger-Methe	Institute for the Oceans & Fisheries; Statistics	x	x
Mark Johnson	Institute for Resources, Environment & Sustainability; Earth, Ocean & Atmospheric Sciences	x	
Martin Hirst	Microbiology and Immunology	x	

Matt Pennell	Zoology	x	
Matthias Görges	Anesthesiology, Pharmacology & Therapeutics	x	
Mieszko Lis	Electrical and Computer Engineering		x
Mike Gelbart	Computer Science		x
Mike Whitlock	Zoology		x
Nancy Heckman	Statistics	x	x
Norm Hutchinson	Computer Science	x	
Paul Gustafson	Statistics		x
Patrick Walls	Mathematics	x	x
Phil Austin	Earth, Ocean & Atmospheric Science	x	x
Rachel Pottinger	Computer Science		x
Raymond Ng	Data Science Institute; Computer Science		x
Sara Harris	Faculty of Science Dean's Office; Earth, Ocean & Atmospheric Sciences	x	x
Sean Graham	Botany		x
Steven Barnes	Psychology	x	x
Steven Hallam	Microbiology and Immunology	x	x
Sunita Chowrira	Botany, Zoology		x
Tiffany Timbers	Statistics		x
Valentina Radic	Earth, Ocean & Atmospheric Science	x	

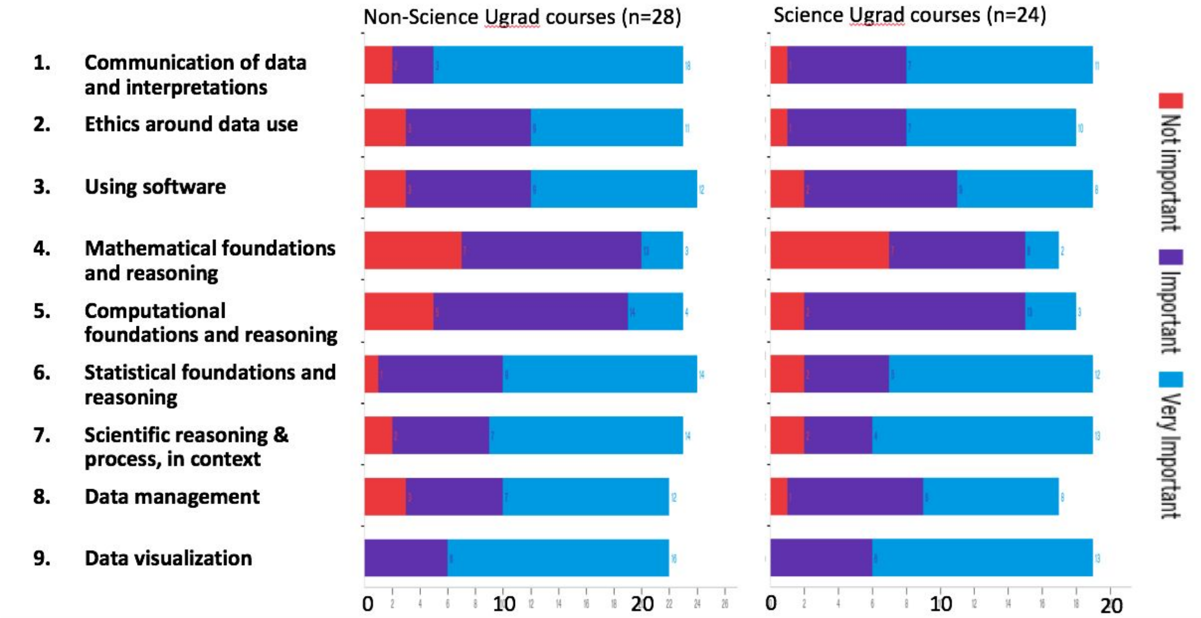
Warren Williams	Biochemistry; Faculty of Science Curriculum Chair	x	
Wayne Maddison	Botany, Zoology	x	x
Xiaojun Li	Political Science		x

Appendix D: Selected survey results of faculty

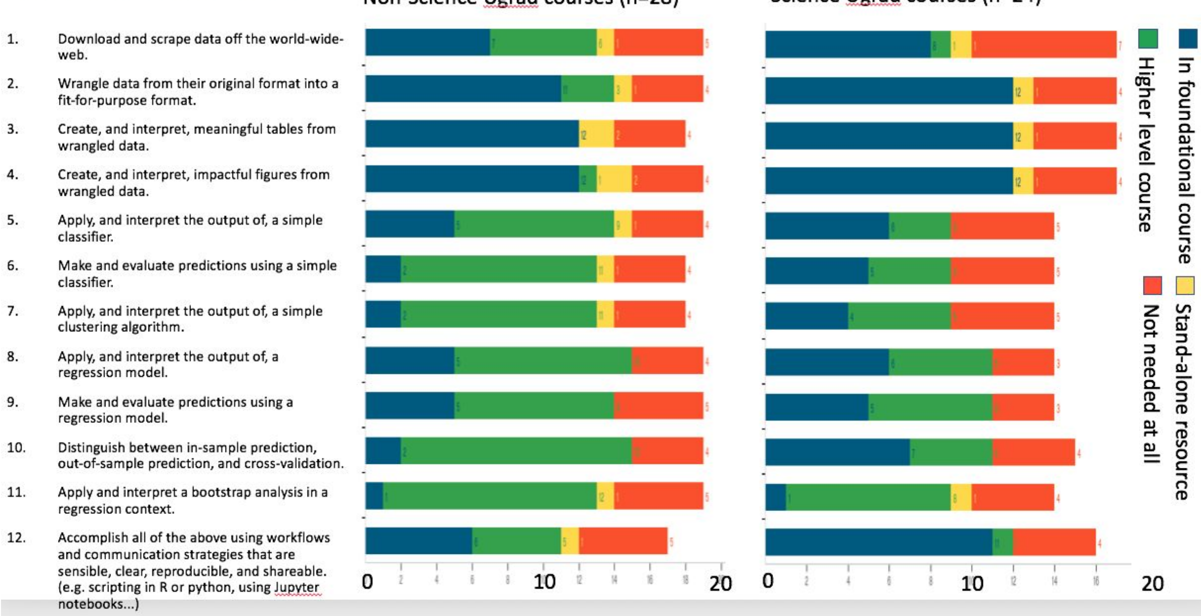
Selected results from a survey of faculty members across UBC-V. We received responses from 64 participants, representing 75 existing courses 52 of which are undergraduate courses. The full survey (questions asked) can be found here: https://ubc.ca1.qualtrics.com/jfe/form/SV_dbBW0pIolmxSVNz



Which of these themes should be in a 3-credit, 13-week foundational data literacy course?



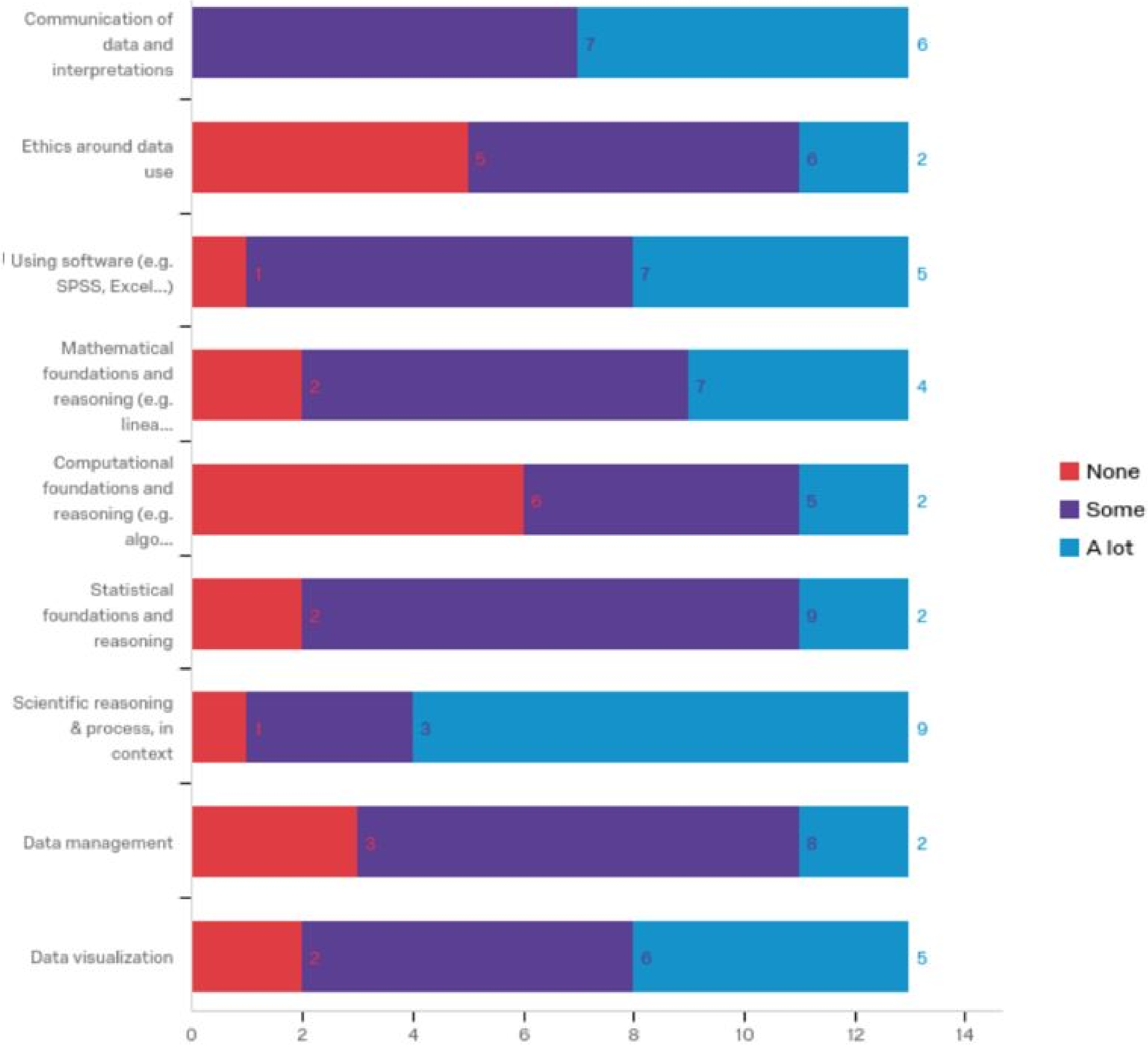
Which of these goals, if any, from the current DSCI 100 course should be in a 3-credit, 13-week foundational course?



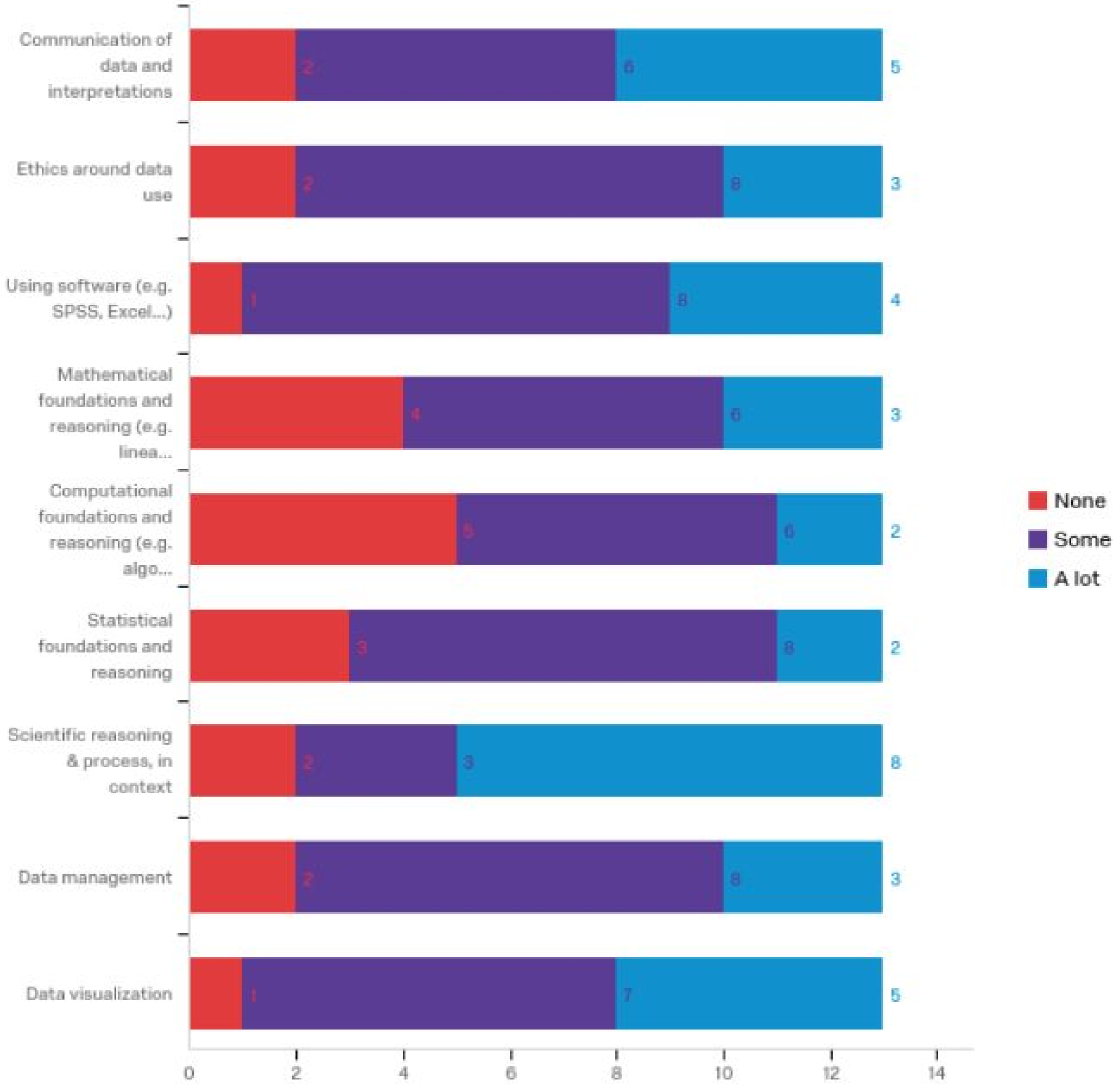
Appendix E: Selected survey responses from graduating BSc students, May 2019

Selected results from a survey of students graduating with a BSc in May 2019 from UBC-V. We received responses from 16 participants (13 completed most of the questions) from 8 of 9 Science departments, plus Integrated Sciences and Combined Major in Science. The full survey (questions asked) can be found here: https://ubc.ca1.qualtrics.com/jfe/form/SV_djxJ7dZZEaLmv7n

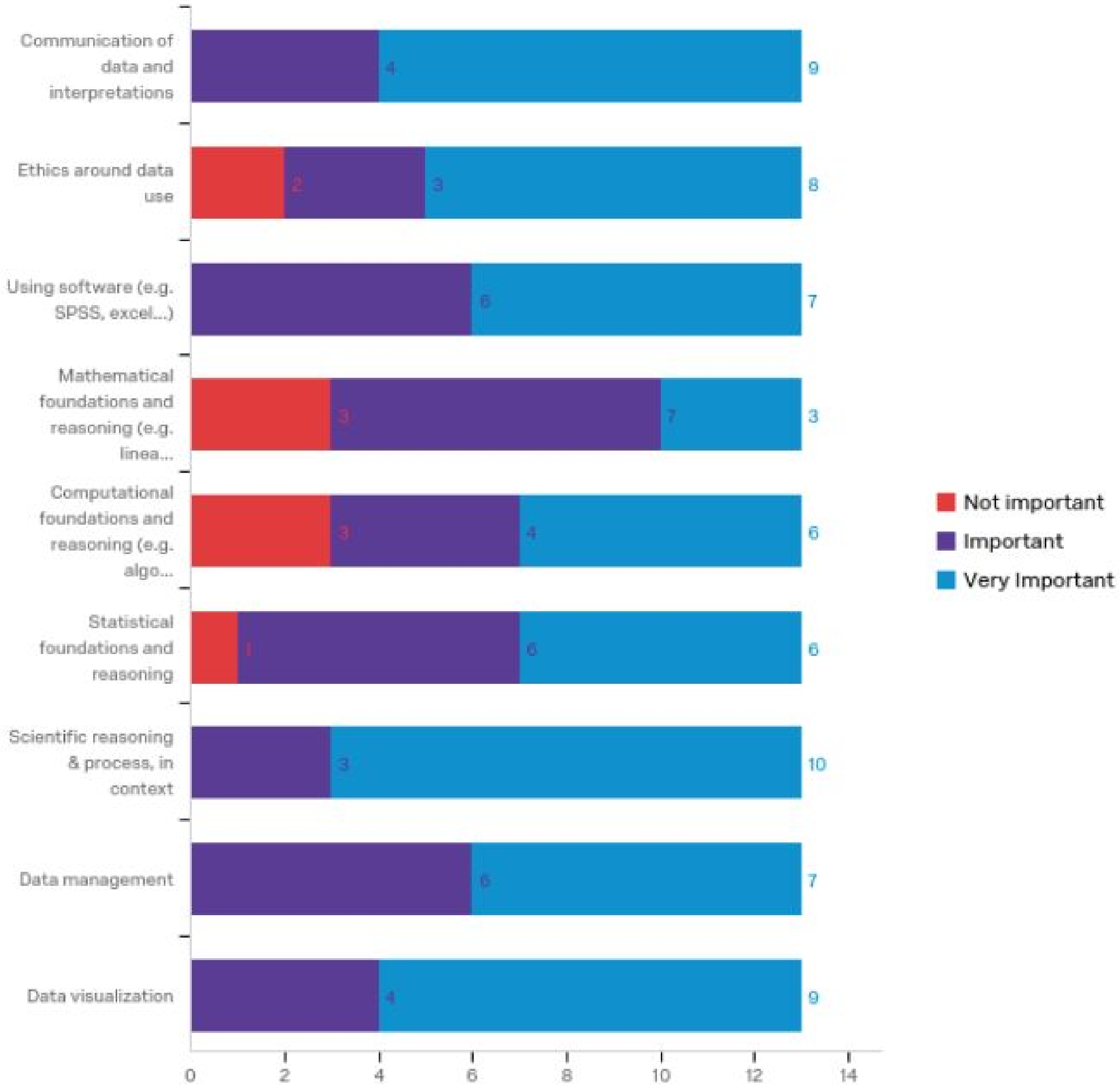
Below is a list of data literacy themes that the Data Science Committee determined are relevant to existing courses, programs, and departments with the Faculty of Science. Based on your experiences as part of your degree program, please rate your level of exposure in these themes.

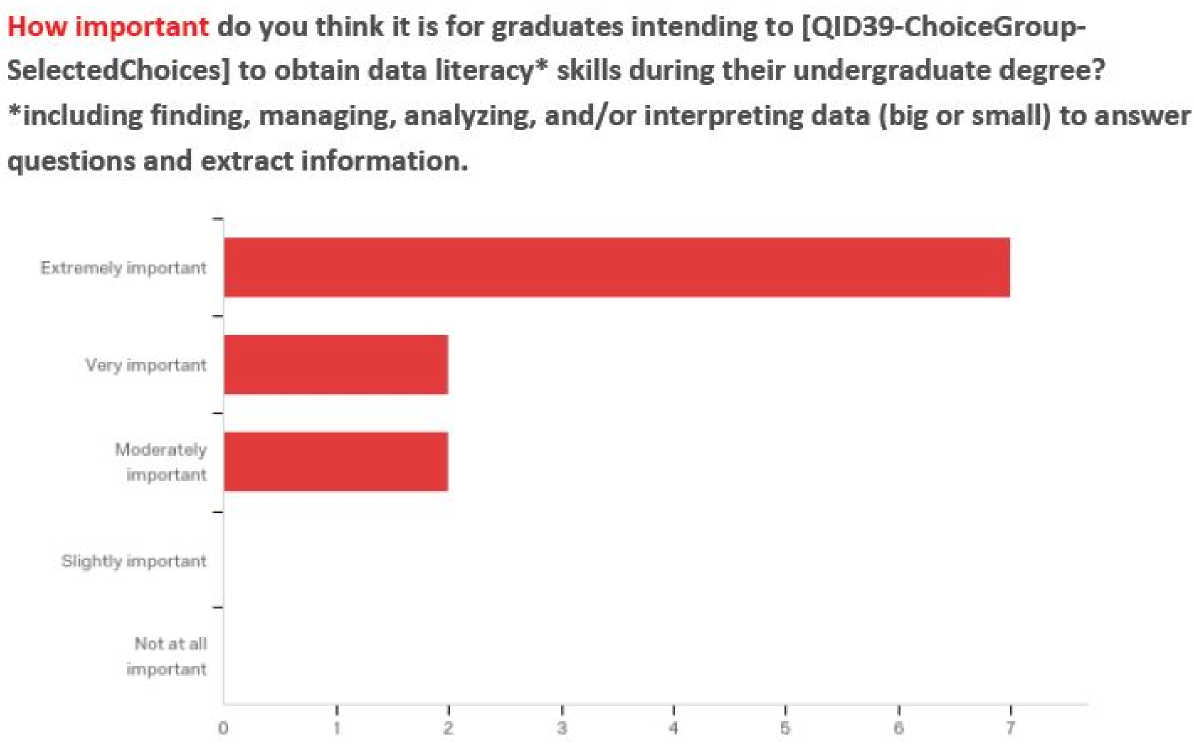


Below is a list of data literacy themes that the Data Science Committee determined are relevant to existing courses, programs, and departments with the Faculty of Science. Based on your experiences as part of your degree program, please rate your level of confidence in your ability in these themes.



Thinking about these same themes, how important is each theme for a data literate student graduating with a degree in your discipline?





Appendix F: Proposed Data Science Minor

Disclaimer: this is meant as a first pass at one *possible* data science minor. The data science minor will be more fully developed during the large TLEF proposal process. Additional hires in the departments of Statistics and Computer Science will be needed to ensure the expertise to make the minor high quality.

Course requirements

The prerequisites for the DSCI minor are 1st year data science, 2nd year programming and 2nd year statistics, 1st year calculus and 2nd year linear algebra. These are shown in the table below, under “Prerequisite” type. In addition to the prerequisites, the minor requires 6 courses (18 credits) of upper-level courses. These 6 courses are broken down into 2 required areas and 4 elective areas; for each area, there may be multiple UBC courses that fulfill the requirement.

The two required areas are statistical modeling and machine learning. These are shown in the table below, under “Required” type. For the remaining 4 courses, students must complete a course in 4 different areas, such as Visualization, Databases, Ethics, etc. Currently, there are 6 elective areas listed, but we may add more in the future.

Area	**Recommended courses**	**Example acceptable alternatives**	**Type**
Data science	DSCI 100		Prerequisite

Programming	CPSC 103, 203	CPSC 110 (or 107) and CPSC 210	Prerequisite
Statistical inference	STAT 2xx (new course)	STAT 302 and STAT 200	Prerequisite
Math	One of MATH 101, MATH 103, MATH 105, MATH 121, SCIE 001 and MATH 221		Prerequisite
Statistical modeling	STAT 3xx.a (new course)	STAT 306	Required
Machine learning	CPSC 330	CPSC 340 or STAT 406	Required
Visualization	DSCI 3xx.b (new course)	CPSC 4xx (new vis course for majors, requires CPSC 310) or STAT 545 & 547	Counts toward minor
Databases	CPSC 3xx.c (new course)	CPSC 304	Counts toward minor
Reproducible Workflows	DSCI 3xx.d (new course)	STAT 545 & 547 or 450	Counts toward minor
Ethics	DSCI 4xx.a (new course)	CPSC 430 or CPSC 100	Counts toward minor
Cloud Computing	CPSC 4xx.b (new course)	CPSC 416	Counts toward minor
In-discipline data science courses	(we will compile a list of acceptable courses across the University)	n/a	Counts toward minor

Courses listed above that fall clearly into STAT or CPSC have been coded with those course codes. Others would be DSCI and responsibility would be shared among units (likely Stats and CS, but could include others). We give more detail about each of the proposed new courses below.

In general, we expect students to be taking courses from the “Recommended course” column. However, we recognize that we will not be able to develop all of these new courses at once. Therefore, for each area we have made a list of other courses (“Example acceptable alternatives" column) that students can take in the meantime. Over time we aim to shift the program more and more towards the recommended courses.

In some cases, these alternative courses will be inaccessible to most students. For example, the alternative Visualization course (CPSC 4xx) will require CPSC 310 as a prerequisite, and so is mainly just accessible to CS majors. For non-CS majors in the Data Science minor, if the recommended course has not yet been created and the acceptable alternatives are totally inaccessible, we expect students to not pursue that particular area.

As an example path, consider a student in Psychology who is interested in the Data Science minor. For the sake of the example, let's say new courses have been created in statistical inference, statistical modeling, reproducible workflows, and databases. Then, this example student might take the following courses:

- Prerequisite (data science): DSCI 100
- Prerequisite (programming): CPSC 103 and 203
- Prerequisite (statistical inference): DSCI 2xx (new course)
- Prerequisite (math): MATH 103 and MATH 221
- Statistical modeling: DSCI 3xx.a (new course)
- Machine learning: CPSC 330
- Databases: DSCI 3xx.c (new course)
- Ethics: CPSC 430
- Workflows: DSCI 3xx.d (new course)
- In-discipline: PSYC xxx (new or existing course to be approved for inclusion into the minor)

Examples of potential in-discipline courses with data science incorporated.

ATSC 301: Atmospheric Radiation and Remote Sensing
EOSC 410: Geoscientific Data Analysis and Empirical Modeling
EOSC 442: Climate Measurement and Analysis
PSYC 359: Advanced Research Methods in Behavioural Sciences
MICB 405: Bioinformatics
MICB 425: Microbial ecological genomics

Details of proposed new courses

STAT 2xx – Statistical inference for Data Science

Fundamental concepts in probability. Describing data generated from a probability distribution. Statistical view of data coming from a probability distribution. The statistical and probabilistic foundations of inference, the frequentist paradigm. Concepts first introduced through simulation, followed with an introduction and exploration of asymptotic theory.

STAT 3xx.a – Statistical modelling for Data Science

Linear models for a quantitative response variable, with multiple categorical and/or quantitative predictors. Matrix formulation of linear regression. Model assessment and prediction. Useful extensions to basic regression (e.g., generalized linear models). Model fitting and prediction in the presence of correlation due to temporal and/or spatial association.

DSCI 3xx.b – Visualization for Data Science

Exploratory data analysis. Design of effective static visualizations. Plotting tools in R and Python. How to make principled and effective choices with respect to marks, spatial arrangement and colour. Analysis, design, and implementation of interactive figures. How to provide multiple views, deal with complexity, and make difficult decisions about data reduction.

CPSC 3xx.c - Databases for Data Science

Databases for those who are likely to use a lot of databases rather than for those who are likely to be called upon to set up a lot of databases (which is the subject matter of CPSC 304). Topics include Relational database basics, querying relational databases in SQL, NoSQL databases, how to write applications to query databases, how to write applications to insert data into databases, including the Extract, Transform, and Load process.

DSCI 3xx.d – Reproducible Data Science Workflows

Data science projects can quickly grow out-of-hand and become irreproducible in the absence of deliberate effort at organization, tool choice, and process. Reproducible research, where the scientific findings are published with the data, and software used to generate the findings, is critically important for the scientific process as it allows others to verify and easily extend the scientific findings. This course will teach principles of reproducible and sound data science workflows and will develop skills implementing them in appropriate state-of-the-art systems and languages. As tools and practices in data science change and evolve rapidly, a focus of this course will also be on the criteria which make data science workflows reproducible such that students can use this knowledge to adopt new reproducible workflow tools and practices beyond this course.

DSCI 4xx.a – Ethics for Data Science

Legal and ethical issues concerning data, including aggregated data. Proactive compliance with rules, and in their absence, principles for the responsible management of sensitive data. Case studies.

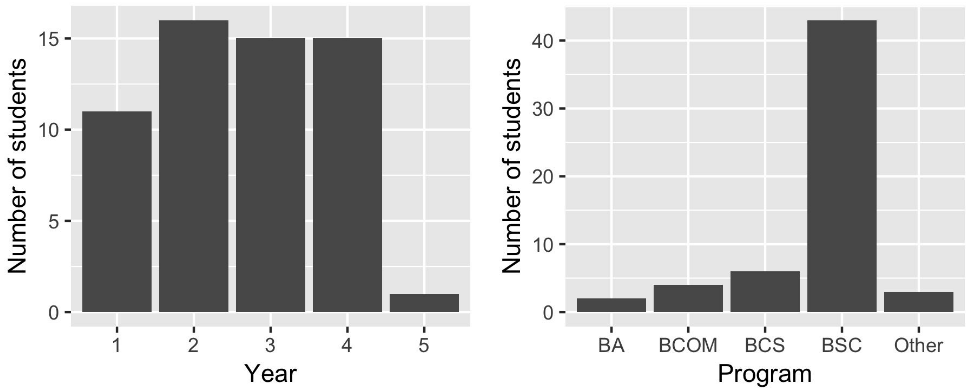
DSCI 4xx.b – Cloud computing and Data Science applications for Big Data

How to use the web as a platform for data collection, computation, and publishing. Accessing Data via scraping and APIs. Using the cloud for tasks that are beyond the capability of local computing resources.

Appendix G: DSCI 100 Evaluation

First offering in 2018-19 T2:

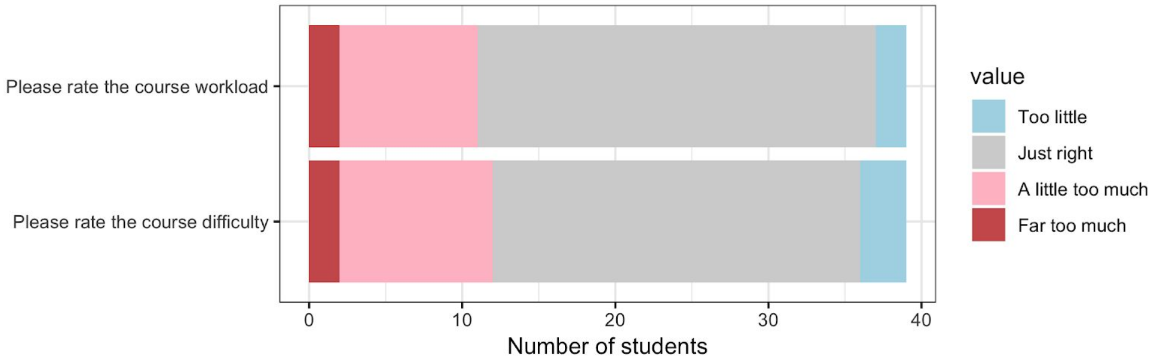
Student Demographics:



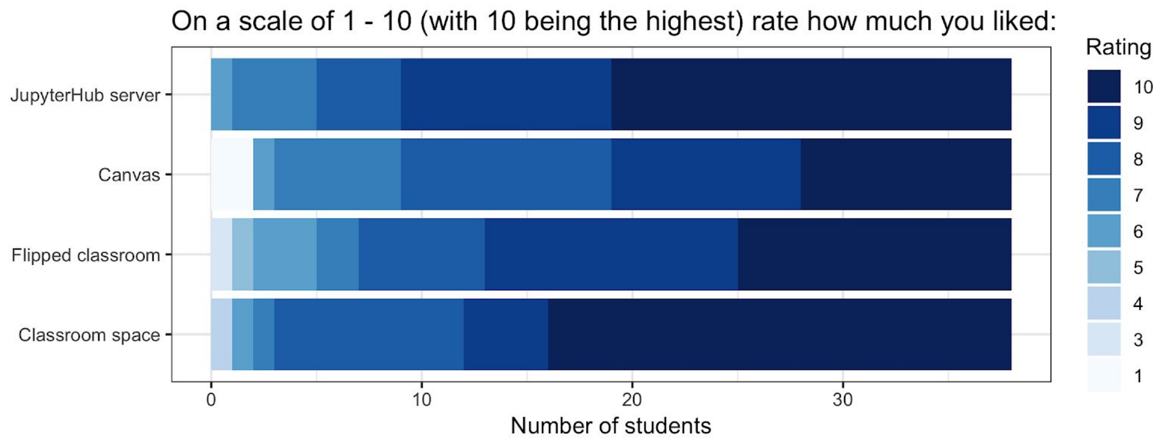
note - in this offering only limited seats were open to non-Science students.

Select mid-course survey results:
39 students completed our mid-course survey.

Students generally viewed workload and difficulty as just right:



Students generally gave high ratings to the technologies used, flipped classroom approach, and classroom space:



Teaching evaluations - Faculty of Science Questions

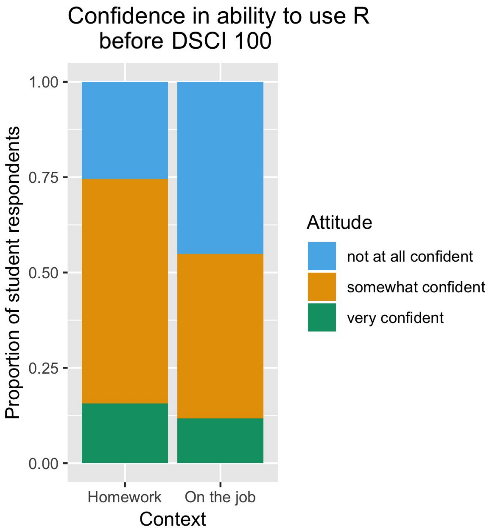
Faculty Questions

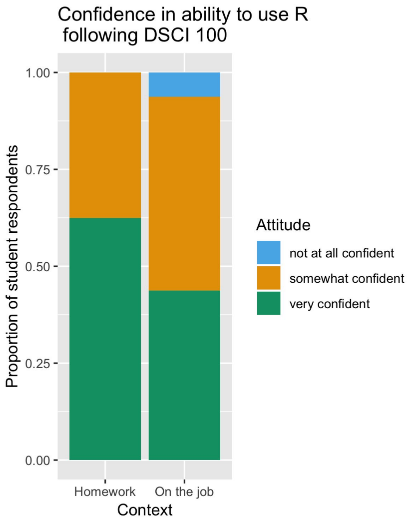
Course Questions

Question	N	n	SD	D	N	A	SA	N/A	IM	DI	Mean	STDEV
My academic background provided sufficient preparation for this course.	59	24	1	0	1	4	16	2	4.81	0.37	4.55	0.96
This course promoted conceptual understanding.	59	24	1	0	0	6	17	0	4.79	0.33	4.58	0.88
The learning activities helped me to succeed in this course.	59	24	0	0	1	3	20	0	4.90	0.18	4.79	0.51
The workload for this course was appropriate.	59	24	0	1	1	4	18	0	4.83	0.30	4.63	0.77
I received sufficient feedback on my progress during this course.	59	24	0	0	2	8	14	0	4.64	0.32	4.50	0.66

Question	%Favourable
My academic background provided sufficient preparation for this course.	90.91%
This course promoted conceptual understanding.	95.83%
The learning activities helped me to succeed in this course.	95.83%
The workload for this course was appropriate.	91.67%
I received sufficient feedback on my progress during this course.	91.67%

Comparison of students' confidence using R pre- and post-course:





Select additional exit-survey results:

Topics reported as most valuable

- Data Wrangling
- Data visualization
- Modelling/prediction/machine learning
- Practical/hands on work with data
- R programming language
- Working in a group
- Jupyter notebooks
- Git & GitHub
- Scraping data off the web

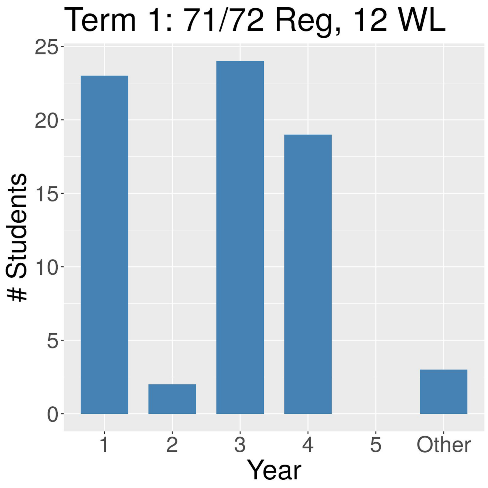
Topics reported as least valuable

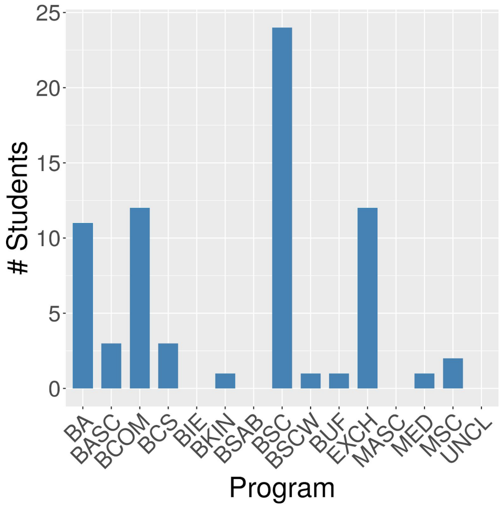
- R programming language
- Git & GitHub
- **Visualizing high dimensional data visualization**
- Scraping data off the web

bolded terms do not intersect these two lists

Second offering in 2019-20 T1:

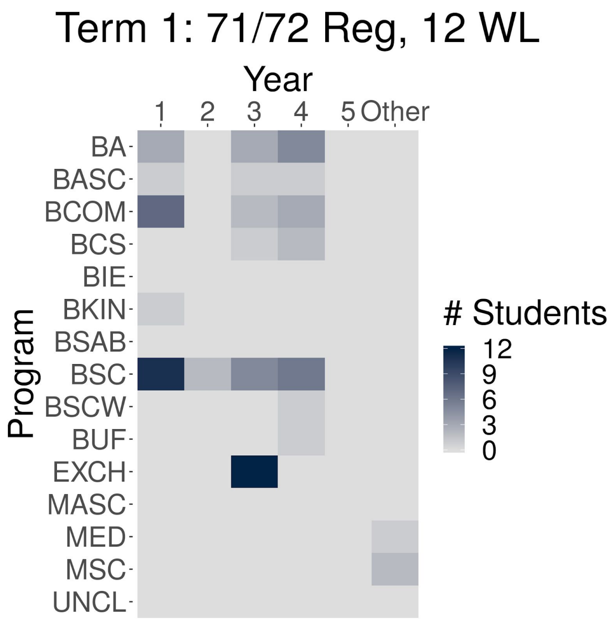
Student Demographics:



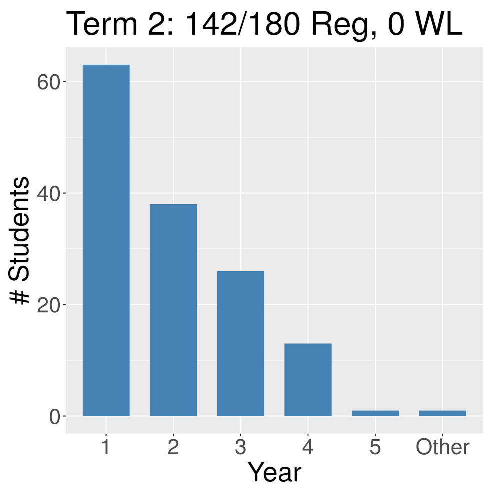


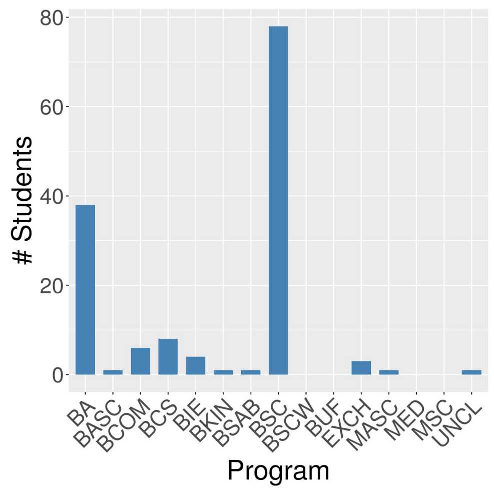
note - in this offering seats were open to all students.

Intersection of Year and Program:

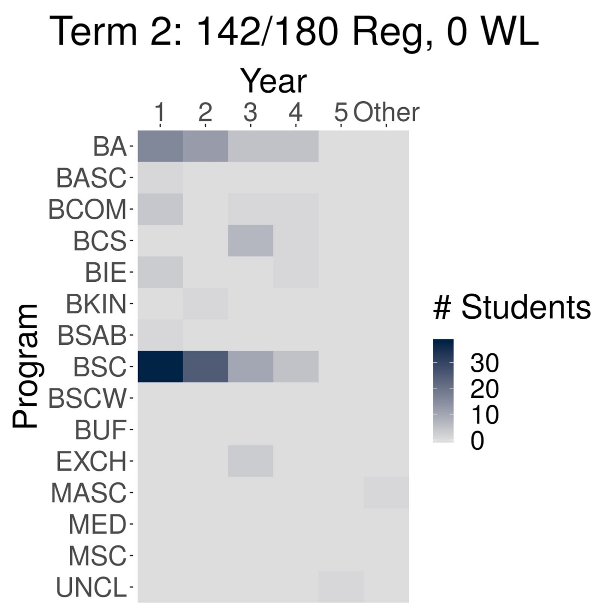


Third offering in 2019-20 T2:





note - in this offering seats were open to all students.



References:

Attwood TK, Blackford S, Brazas MD, Davies A, Schneider MV, 2019. A global perspective on evolving bioinformatics and data science training needs. *Brief Bioinform*. 20(2):398–404. doi:10.1093/bib/bbx100

National Academies of Sciences, Engineering, and Medicine, 2017. *Envisioning The Data Science Discipline: The Undergraduate Perspective: Interim Report*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24886.

National Academies of Sciences, Engineering, and Medicine, 2018. *Data Science for Undergraduates: Opportunities and Options*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25104