Name: Tiffany Timbers
Email: tiffany.timbers@stat.ubc.ca
Faculty, College, or Administrative Unit: Faculty of Science
Title: Instructor
Role: Faculty
Department, School, or Unit: Statistics
Co-Applicants and Team Members: Gabriela Cohen Freue, Associate Professor, Statistics, Faculty of Science, gcohen@stat.ubc.ca

Michael Gelbart, Instructor, Computer Science, Faculty of Science, mgelbart@cs.ubc.ca

Rachel Pottinger, Associate Professor, Computer Science, Faculty of Science,rap@cs.ubc.ca

Phil Austin, Associate Professor, Earth, Ocean & Atmospheric Sciences, Faculty of Science, paustin@eoas.ubc.ca

Patrick Walls, Instructor, Mathematics, Faculty of Science, pwalls@math.ubc.ca

Gülnur Birol, Director, Science Centre for Learning and Teaching, Faculty of Science,birol@science.ubc.ca

Warren Code, Associate Director, Science Centre for Learning and Teaching, Faculty of Science, warcode@science.ubc.ca

The Faculty of Science Associate Dean Academic (Sara Harris) has been involved in this project.

Department Head Approved: Yes

Project Summary: University graduates in many disciplines need skills to organize, interact with and extract meaning from data via reproducible and auditable methods. Through this project, we will create and sustain widely accessible learning opportunities in data science and contribute to a data-literate citizenry. Specifically, we will: (1) review and broaden DSCI 100: Introduction to Data Science, based on campus-wide consultation; (2) create a data science minor available to any undergraduate, which involves developing seven new courses; (3) create professional development opportunities for faculty and TAs to share student-focused pedagogies for teaching data science; and (4) develop and implement evaluation plans to assess the effectiveness of (1)–(3). This project would provide thousands of undergraduates from any discipline with foundational data skills and the opportunity to complement their degree with a relevant minor.

This project is synergistic with other TLEF proposals which address discipline-specific efforts in data science curriculum (https://bit.ly/2MIAC68).

Students Reached by Project

Course Code	Section	Academic Year	Term
DSCI 100: Introduction to Data Science	All (currently 2/yr, expected to grow # of sections)	2020/21, 2021/22, 2022/23	Sep/Jan
STAT 2XX: Statistical Inference for Data Science	New course, one section	2020/21	Jan
STAT 3XX: Statistical Modeling for Data Science	New course, one section	2021/22	Sep
CPSC 3XX: Databases for Data Science	New course, one section	2021/22	Sep
DSCI 3XXa: Visualization for Data Science	New course, one section	2021/22	Jan
DSCI 3XXb: Reproducible Data Science Workflows	New course, one section	2021/22	Jan
CPSC 4XX: Cloud Computing and Data Science Applications for Big Data	New course, one section	2022/23	Sep
DSCI 4XX: Ethics for Data Science	New course, one section	2023/24 (1st offering after end of TLEF. Existing course CPSC 430 will substitute in the minor until DSCI 4XX is offered)	Sep

Overall Student Outreach:

Number of Students Impacted: 3000

Project Details

Project Objectives: Rationale

Data science can be defined as extracting value from data through reproducible and auditable processes. A strategic priority for the Faculty of Science is to provide high-quality, coordinated, widely accessible learning opportunities in data science to a broad range of UBC-V undergraduates. As few opportunities to acquire these skills currently exist, we will build on UBC-V's expertise with the successful Master of Data Science (MDS) program, and the recently launched DSCI 100 course, to provide this curriculum.

Overall Objectives

1. Review & broaden (as needed) DSCI 100 to serve undergraduates in a wide range of disciplines, whether or not they will continue with additional data science learning.
2. Develop a minor in data science that could be paired with any major, incorporating DSCI 100 and seven new courses with short pre-requisite chains.
3. Create professional development for faculty and TAs to provide training in the foundational data science toolkit and share student-focused pedagogies for teaching data science.
4. Develop and implement evaluation plans to assess the effectiveness of the foundational data science course (DSCI 100), the minor and the professional development activities.

Expected Impacts

DSCI 100 was launched in 2018WT2 with 59 students as a foundational data science course with student-active design. In 2019W, 250 students are registered and the two offered sections are full. The course could grow to serve ~2000 students, based both on experience from Berkeley and anticipation that some large specializations/majors in Science may require DSCI 100. We expect the course to be able to meet the foundational data science needs of multiple disciplines.

Although several other undergraduate courses at UBC-V currently offer opportunities to learn some data science content, they also typically cover theoretical approaches and thus involve long pre-requisite chains (e.g., CPSC 304 & 340, STAT 406). Building courses tailored to students whose goal is to practice data science can lead to the creation of courses that are more accessible and targeted to data science content and pedagogies.

The seven new courses we propose to build in this project will be created with this student goal in-mind. They, along with DSCI 100, will form the core of a new data science minor (which may potentially inspire a data science major). Additionally, students will also be able to count one data-intensive disciplinary course in the minor. These courses will provide broad access to data science education with fewer barriers, serving both students pursuing the proposed data science minor as well as those who only complete a portion of the courses in the minor.

All courses will incorporate state-of-the-art tools used in research and industry (R & Python, Jupyter, GitHub, etc.), authentic projects with real data sets, and research-informed pedagogies. To enable this, the project will provide professional development (PD) for faculty and TAs in hands-on workshops to train them in the foundational data science toolkit (e.g., similar to UBC's Jupyter Day) and share evidence-based, student-focused pedagogies for teaching data science (e.g., similar to UBC's Master of Data Science Instructor training workshop).

These efforts will increase UBC's pedagogical capacity to help students develop data skills. This work aligns with UBC's Strategic plan (Strategies 11 & 13), TLEF criteria for students and faculty reached and consulted, and complements existing and emerging efforts to infuse data skills throughout specializations (e.g. see TLEF proposals from Ivanochko and Hallam).

Project Focus Areas:

True	False
False	False
True	False
True	False
True	True
True	False

Work Plan, Timeline & Milestones: Since 2018, over 35 UBC faculty members (from Science, Applied Science, and Arts), two staff members, and a Science Undergraduate Society (SUS) representative have contributed to discussions envisioning data science educational opportunities for UBC-V undergraduates. This group gathered input from surveys (described below) and developed a set of recommendations for creating accessible data science learning opportunities. 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.

The project proposed here addresses specific recommendations to broaden a foundational course; create a minor; use evidence-informed, active pedagogy; use open-source tools and align tool-use across disciplines; create and sustain professional development opportunities for faculty, TAs, and staff to learn/share best practices in data science pedagogy; and evaluate the impact on student learning.
A faculty sub-committee from Statistics and Computer Science (all co-PIs on this project) created a proposal for an undergraduate data science minor accessible to any UBC undergraduate. Like all DSCI courses and the MDS program, the minor will be the joint responsibility of these two departments. With this preparation, TLEF funding is timely to bring these plans to fruition.

Summary of completed work:

1. Compiled a list of institutions offering similar programs in North America.

2. Conducted workshops & discussions involving faculty, staff, and a student from Science, Applied Science and Arts.

3. Piloted DSCI 100, a foundational course for a data science minor, starting in 2018W2. This course incorporated evidence-based teaching practices and technologies. The piloting and preliminary evaluation offered invaluable information on how to structure and organize future courses in this area.

4. Surveyed faculty (64 participants from 5 Faculties, in March 2019), graduating BSc students (in May 2019), graduate students (in June 2019) across campus, and students in DSCI 100. The responses reveal which DSCI 100 learning goals are considered most key, and indicate that students in upper-level courses could achieve more in discipline-specific data-rich settings if they had prior foundational data science experience.

5. Held an open event (August 2019) for faculty interested in learning more about DSCI 100. Attended by faculty members from at least 4 Faculties.

6. Developed a proposal for a data science minor (http://bit.ly/dsci-minor), including the following draft learning goals:

By completing this minor, students will be able to:

- Identify and collect data necessary to answer a given research question through sampling and/or through extracting data from pre-existing sources (databases, html web pages, web APIs, etc)
- Manipulate messy, ill-formed data to extract meaningful insights.
- Map and apply an appropriate data analysis approach to a given research question.
- Tailor data science methods to deal with diverse data types across diverse subject-area domains.
- Apply fundamental statistical thinking in the data analysis process.
- Apply fundamental programming principles in the data analysis process.
- Communicate results of data science experiments to diverse audiences through data visualizations, written work and oral presentations.
- Employ best practices for collaboration for projects that involve both code and people.

This preliminary work serves as a solid foundation for the proposed work, the bulk of which involves developing courses for a data science minor. The planned new courses are designed to allow students to meet the learning goals listed above. The first two years mainly concentrate on new course development and piloting, leading to the availability of the minor mid-Year 2. The third year focuses on completing the last course development and full implementation of most new courses. Evaluation will be an integral part of the work and will inform decisions throughout the project's life span.

The novelty of this data science minor presents itself as an excellent opportunity to exchange ideas with colleagues within and beyond UBC and to contribute to the growing body of knowledge in data science education. Therefore, when possible, we will disseminate our work to gather input and share our experiences and findings broadly.

Drawing on the best practices in data science, this project will proceed as outlined below. In our three-year plan, we have staggered sequences of course development such that each course has one 4-month term for curriculum proposal development, followed by two terms of course materials development before its first offering. The maximum number of courses starting the sequence in any term is 2. Details of the courses listed in this work plan are below in the Project Outputs section (course abbreviations are used here).

Preparation year (2019W):

Develop and submit curriculum proposal for new course STAT 2XX.

Year 1 (2020/21):

1. Iterate DSCI 100, as needed, based on consultations and student feedback.
2. Build one new course, STAT 2XX (development summer & T1); offer it (T2); conduct early evaluation.
3. Evaluate redesigned DSCI 100, and STAT 2XX.
4. Develop and submit curriculum proposals for four additional new courses, staggering with two in T1 (STAT 3XX, CPSC 3XX) and two in T2 (DSCI 3XXa, DSCI 3XXb). Begin building STAT 3XX, and CPSC 3XX in T2.
5. Develop and submit curriculum proposal for the data science minor (T2). This will include broader consultation with undergraduate programs beyond those in the initial workshops.
6. Develop, offer and evaluate PD workshops for faculty and TAs in teaching data science with open source tools (e.g. Jupyter notebooks, nbgrader, effective data science pedagogies etc.)
7. Establish LT/IT needs associated with expansion in courses and enrolment in anticipation of Years 2 and 3.

Year 2 (2021/22):

1. Finish building four new courses; offer two of them in T1 (STAT 3XX, CPSC 3XX) and two in T2 (DSCI 3XXa, DSCI 3XXb).
2. Begin offering the data science minor, using the four new upper level courses, and two existing courses (CPSC 330, CPSC 430), and any existing data-intensive in-discipline courses.
3. Develop and submit curriculum proposal and begin course development for new course CPSC 4XX (an additional elective in the minor).
4. Iterate the earlier new course (STAT 2XX) as needed, based on early evaluation
5. Conduct early evaluations for the four new offerings (STAT 3XX, CPSC 3XX, DSCI 3XXa and DSCI 3XXb).
6. Iterate on (based on evaluation) the PD workshops for faculty and TAs in teaching with open source tools.

Year 3 (2022/23):

1. Finish building CPSC 4XX; offer it (T1); conduct early evaluation.
2. Develop and submit curriculum proposal and begin course development for new course DSCI 4XX (course development will finish after the end of the TLEF).
3. Iterate the four courses first offered in Year 2 as needed, based on early evaluation.
4. Evaluate the overall minor.

Project Outputs: The primary deliverables from this project are a course revision (DSCI 100), 7 new courses, pedagogies and open educational resources (OER) for teaching data science, and professional development for teaching data science.

Below we outline existing and proposed courses for this minor (http://bit.ly/dsci-minor), that would be affected by this project. DSCI 100 forms the foundation for the minor, followed by three other required courses, plus electives. Students may choose one of their electives to be a data-intensive course in their chosen discipline. The minor will include 18 upper-level credits. We anticipate increasing demand for DSCI 100 over time. Courses listed below are new, unless otherwise noted.

Required:

1. DSCI 100: Introduction to Data Science (existing course, to be reviewed and broadened)
2. STAT 2XX: Statistical Inference for Data Science
3. STAT 3XX: Statistical Modeling for Data Science
4. CPSC 330: Applied Machine Learning (existing course)

Electives: Students choose 4 from this list or may substitute one discipline-specific data-intensive course:

1. CPSC 3XX: Databases for Data Science
2. DSCI 3XXa: Visualization for Data Science
3. DSCI 3XXb: Reproducible Data Science Workflows
4. CPSC 4XX: Cloud Computing and Data Science Applications for Big Data
5. DSCI 4XX: Ethics for Data Science (1st offering after end of TLEF. Existing course CPSC 430 will substitute in the minor until DSCI 4XX is offered)

Pedagogies

In these newly developed courses we will incorporate evidence-based, inclusive, active learning pedagogies to engage with real-world data using technological tools for visualization and analysis. This will allow students to use methods to think critically about the world, draw conclusions from data considering ethical implications, and visualize and communicate results effectively.

Open Educational Resources

Custom learning materials will be developed for each of the 7 new courses and deployed online under an open license following the standard model for OER development in data science: literate code documents (e.g., Jupyter Notebooks or R Markdown) rendered with a static website generator (such as Git book) and deployed on GitHub. For example, see the existing DSCI 100 textbook https://ubc-dsci.github.io/introduction-to-datascience/.

Professional Development

Workshops for instructors on teaching data science will be developed and run annually. Learning materials for the workshops will be deployed on GitHub for easy re-use by future workshop facilitators. Several of the co-PI's on this proposal have experience creating and delivering such workshops. For example, in August 2019 Patrick Walls, a co-PI on this TLEF proposal, ran JupyterDay. This event attracted over 60 students, fellows and faculty for instructor workshops and presentations on teaching data science (see https://github.com/patrickwalls/jupyterday2019). Additionally, the lead PI and a co-PI on this proposal, Tiffany Timbers and Michael Gelbart, ran a Master of Data Science Instructor training two-day workshop in August 2019, which focused on best pedagogical practices for teaching Data Science. This event was attended by 12 faculty members (https://ubc-mds.github.io/MDS_instructor_training/).

Project Impact: Data science skills are in great demand in research and industry across virtually all disciplines. This is because nearly all disciplines now involve bigger, or more complex data and as such, students need additional data skills to succeed and be competitive within these disciplines. Thus to have UBC-V graduates stand out in both academic research and industry, regardless of discipline, they require quality opportunities to acquire data science skills.

Tangible benefits to undergraduates (aligns with Objectives 1, 2, and 4):

1. Data science knowledge and skills to handle complex data and use it to tackle a wide-variety of questions relevant to many fields in a reproducible and auditable manner.
2. Exposure to, and practice with, open access educational resources created using authentic, open source data science tools (e.g., Git book textbooks, Jupyter notebook electronic worksheets).
3. Evidence-based, inclusive, active learning pedagogies suitable for data science (e.g., live and paired programming, flipped classrooms, peer code review, auto- and peer grading, group projects).
4. Data science minor undergraduates will be well qualified for future work with a data science or data analysis component. This could take place in a variety of fields and sectors spanning industry, research, government, etc. For those who do not continue directly in the field, they will bring a data science perspective to their career path of choice.
5. Evaluation framework and accompanying tools that can be used to assess the impact of new or existing strategies on student experience in relevant courses.

Tangible benefits extending beyond undergraduates (aligns with Objectives 3 and 4):

1. Data science skills injected into UBC research labs through the training of graduate students as TAs for undergraduate data science courses; faculty upskilling through co-teaching opportunities.
2. A community of practice around data science at UBC-V for faculty to engage in peer-to-peer sharing of best practices (e.g., 32 of the faculty survey respondents from 5 Faculties indicated interest in being part of discussions for this project).
3. A higher profile for the Faculty of Science and UBC-V in data science education and potential to become a model for a data science curriculum for other institutions.
4. Findings that can be shared with UBC and broader community through dissemination efforts to inform best practices and receive feedback.

To maintain these benefits, the developed resources will be integrated with existing learning technology systems as well as being available as OERs, while the created professional development will enable these courses and practices to continue beyond the development team.

Evaluation Plan: The success of the overall project will be assessed by meeting the anticipated project outcomes in a timely manner and evaluation will be an integral part of the project to inform progress. In preparation for this proposal, we have developed a logic model to describe the overall project and the evaluation plan emerged from that exercise. We

also reached out to undergraduate and graduate students and faculty in Science, as well as other Faculties, to gather input and feedback. Institutions across North America offering comparable programs were reviewed to create benchmarks for evaluation. Furthermore, the Skylight team that will support the evaluation activities of this TLEF project has worked with the Master of Data Science (MDS) program for their program evaluation since its inception. Their experience from evaluating the MDS will help inform evaluation this data science TLEF project.

The overarching goal for project evaluation is to capture student learning and experience in DSCI 100 and the newly developed courses in the minor to inform course effectiveness during the pilot offerings, and to inform further revisions. We also plan to capture instructor and teaching assistant experience to ensure they are well-supported in teaching these newly developed courses.

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 in the minor to evaluate their unique contributions to student experience. The results of this project will be shared back with instructors and TAs to further refine the newly developed courses.

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 minor in an interdisciplinary setting presents itself as an opportunity to learn from. 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.

Student Involvement: Students are key stakeholders in this project and the team welcomes their input. Students (both May 2019 Science graduates and DSCI 100 students) were part of the initial evaluation efforts and will continue to be involved throughout the project providing their input and feedback by participating in surveys and focus groups. In DSCI 100 our surveys have (and will continue to) collect information on students' learning, experience (with course technology and structure) and confidence in applying the data science skills they have acquired. For Science graduates our survey focused on whether recent graduates acquired data skills, and if they did, how they acquired them, as well as their confidence in those skills.

During the preliminary work, we received invaluable input from the Science Undergraduate Society (SUS) representative who attended several meetings during the past year and received all the email communications. We will continue to have an undergraduate student representative (SUS member) in future discussions. The SUS representative will be our link to other students and will receive all correspondence, will be invited to key meetings and asked to provide input.

As part of the overall project evaluation, graduate students will be hired to help the project's TLF (see below) collect and analyze data to support the evaluation activities, presenting itself as a unique learning opportunity for students to gain skills in evaluation and to be part of a professional community.

Professional development workshops will be a great opportunity for these and other graduate students to take part and develop skills in teaching data science and being part of the broader teaching community. These experiences will benefit graduate students in their post-UBC careers.

Facilities Requirements: Ideal data science learning spaces facilitate the use of technology (by the Instructors and students) as well as collaboration. We identify critical design features of a data science learning space:

* Robust WiFi
* Table surfaces
* AC power at all seats
* Multi-input, multi-output audio-visual system

Examples of innovative learning spaces (which include wireless screen sharing): LSK 121 (capacity of 64), ORCH 4074 (capacity of 72), HENN 200 (capacity of 180), LIFE 2201 (capacity of 294).

Standard examples: ORCH 3018 (capacity of 48), DMP 110 (capacity of 120), ESB 1012 (capacity of 150), DMP 310 (capacity of 160).

We also anticipate additional needs in learning technology (LT) capacity associated with these courses as enrolment expands in the second and third years of the project. Use of existing LT (e.g., Canvas with existing methods for linking course components) is expected for 2020, while part of Year 1 activity will be to investigate further needs and integrations to be undertaken in Year 2 of the project.

Project Budget

Year of Application: 1

Title of Previous Funded Project (if any):

Year	Funding Requested
2020/21	115675
2021/22	115675
2022/23	17492

Total Budget: 248842

Other Funding: 100,000

Additional Existing TLEF Projects, If Any: