Department of Computer Science and Department of Statistics, Faculty of Science, Vancouver Campus

Bringing Accessible Data Science Training to Undergraduates

Warren Code, Gabriela Cohen-Freue, Maria Florencia D'Andrea, Cinda Heeren, Varada Kolhatkar, Rodolfo Lourenzutti, Joel Östblom, Tiffany Timbers

Introducing: a Minor in Data Science at UBCV

To make data science accessible to students across campus, the Departments of Computer Science and Statistics have teamed up to develop a Minor in Data Science, which accepted its first cohort in 2021. The TLEF-funded project has included the development of seven new courses.

Program learning outcomes

- Identify & collect data necessary to answer a given research question through sampling and/or through extracting data from pre-existing sources (relational databases, web pages, web APIs, etc.)
- Manipulate messy, ill-formed data to extract meaningful insights.
- Map & apply an appropriate data analysis approach to a given research question and the data at hand.
- Select data science methods to work with diverse data types across diverse subject-area domains.
- Build statistical models that are appropriate given distribution(s) of the data, & appropriately quantify uncertainty of resulting estimates & predictions.
- Apply fundamental programming principles in the data analysis process to make analysis code readable, modular, accurate, & scalable.
- Communicate results of data science experiments to diverse audiences through data visualizations, written work, & oral presentations.
- Employ best practices for collaboration for projects that involve both code & people.
- Perform & communicate results from analyses that are fair, equitable, & honest.
- Employ workflows that facilitate reproducible & transparent data analyses.
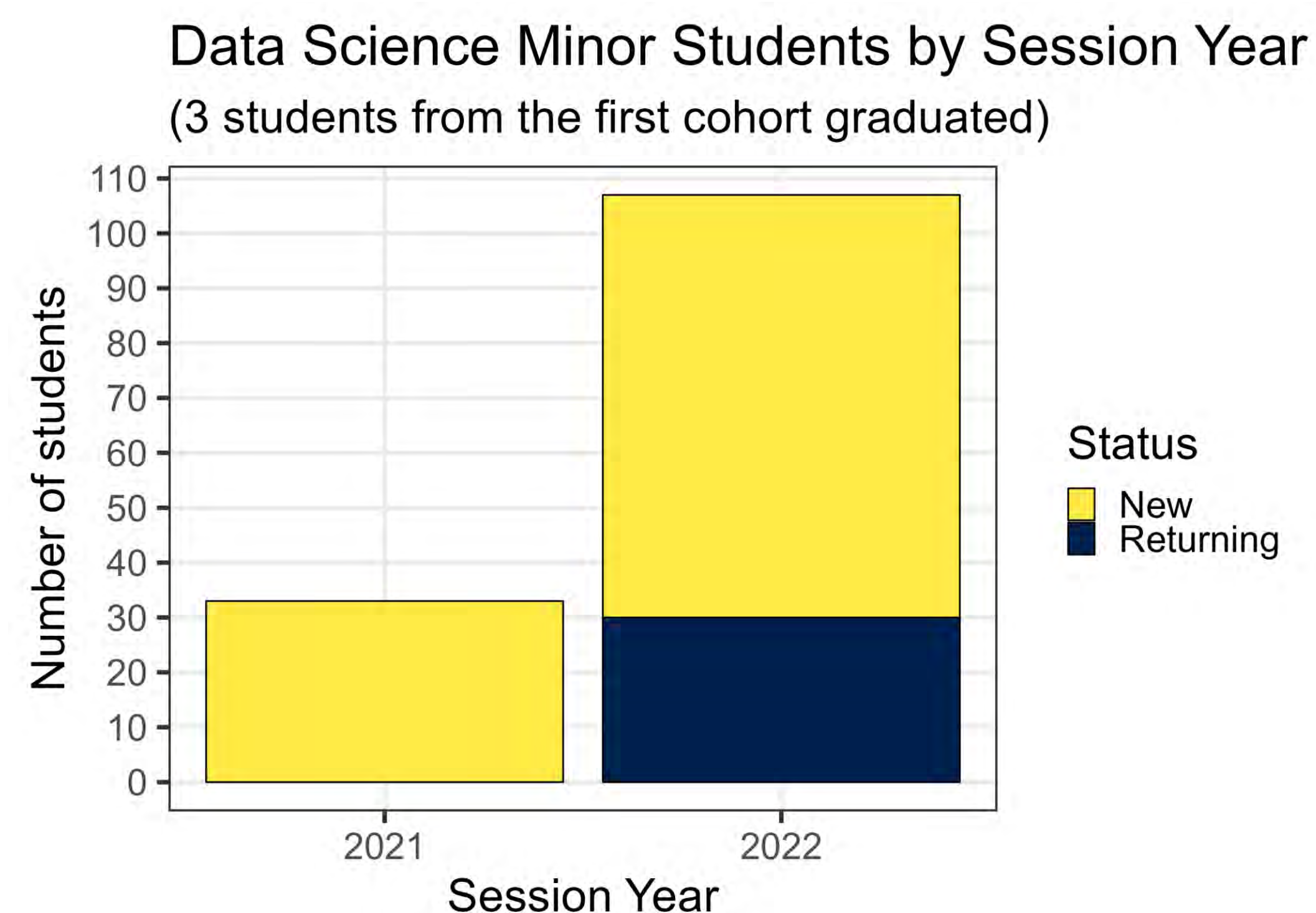
Accessible to many students

This program is open to any undergraduate student at UBCV. Admissions are structured so that no one program's students will dominate a cohort: if there are more applicants than available spots, then each program is allowed up to a certain proportion of the cohort's slots, and admission is by lottery within each program's group of Minor applicants meeting the minimum lower level requirements.

Flexible program requirements

After an initial set of 100- and 200-level programming, statistics, and mathematics courses, plus the foundational DSCI 100 course, the upper level requirements for the Minor have three required courses (including one in ethics) and for the other three courses students have a wide selection of topics and can include a course from their major subject that integrates data science in a substantial way. This program structure is described on the Minor in Data Science main website:
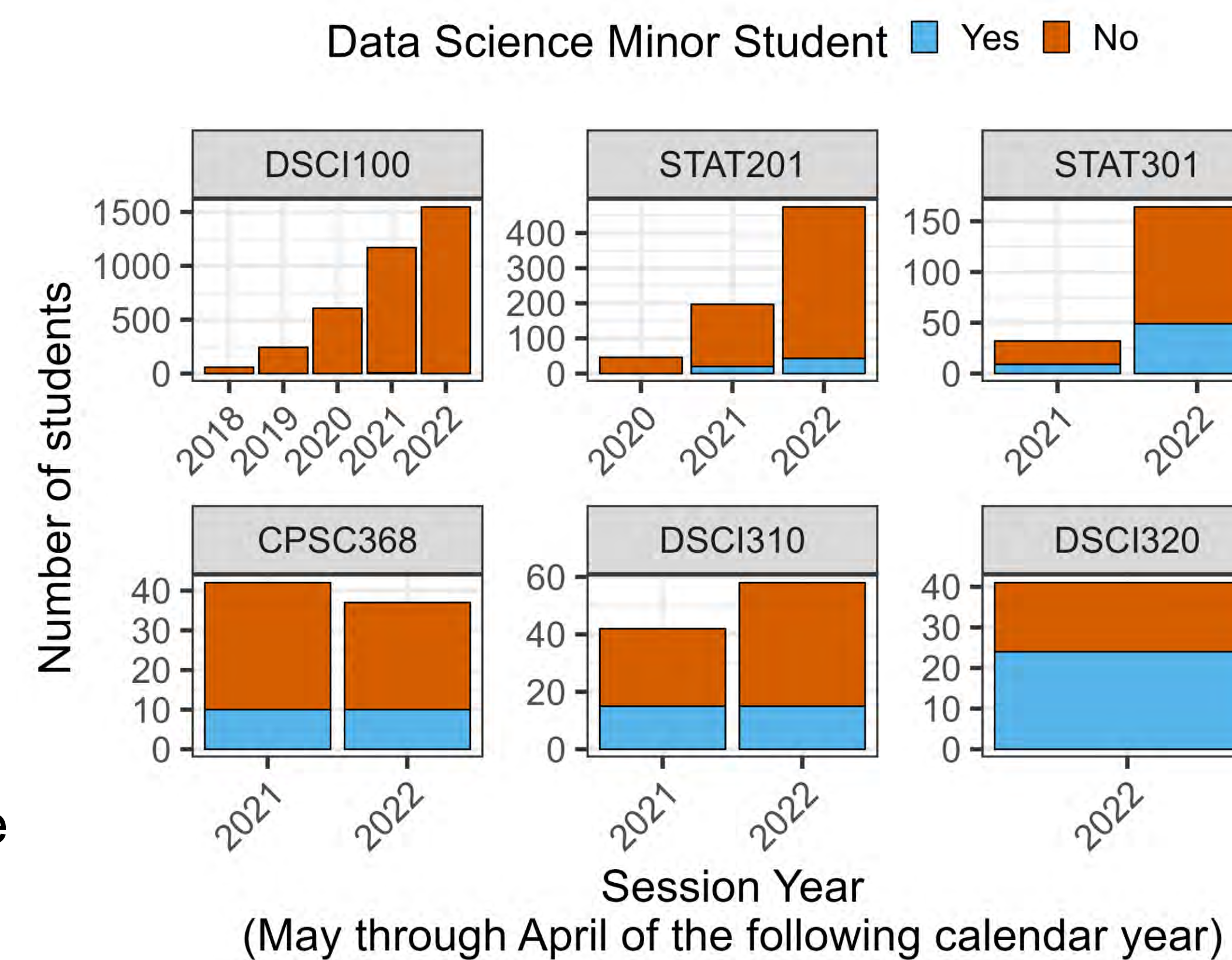https://datascience.ubc.ca/minor

Enrolment is growing

Data Science Minor Students by Session Year
(3 students from the first cohort graduated)

Initial cohorts are going well

Evaluation includes surveys in the core Minor courses; preliminary results suggest that students value the course sequences that have been introduced and find DSCI 100 to be an important foundation for data science.

Enrolment counts in key courses
Includes data through Jan-Apr 2023 term

Data Science Minor Student — Yes / No

Number of students

Session Year
(May through April of the following calendar year)

Course development

Existing course revised as part of the project:
DSCI 100: Introduction to Data Science

New courses developed as part of the project and now in place (see enrolment graphs above):
STAT 201: Statistical Inference for Data Science
STAT 301: Statistical Modeling for Data Science
CPSC 368: Databases for Data Science
DSCI 310: Workflows for reproducible and trustworthy Data Science
DSCI 320: Communicating through Data Visualization

Development underway, with first offerings to come in 2023 Winter Session:
DSCI 4XX: Ethics for Data Science
CPSC 4XX (pilot 2023W): Cloud Computing

Governance and Structure

The Data Science Minor is jointly managed by the two departments via representation from both on a governance committee for course alignment, admissions process, and other program management. They share teaching responsibility for DSCI courses and each have some key courses of their own (CPSC and STAT) in the Minor.

Challenges

- Starting the project in mid-2020 made hiring difficult and available time from all team members was very constrained; the project was able to proceed with small delays.
- Offering home program credit towards the minor is tricky to manage as data science content is expanding all over campus; this flexibility for students requires monitoring and updating as courses continue to be created.
- The diverse set of major programs also makes scheduling conflicts with the Minor's core courses inevitable, but offering more sections of key courses has helped.

What's next?

- Complete development of the final two courses, with first offerings to occur in 2023W.
- Look into a program assistant for this growing program now that the development project is wrapping up and enrolment is growing.
- Complete the evaluation of the project.

Acknowledgement

THE UNIVERSITY OF BRITISH COLUMBIA