Exploring Library Support for Data Science

Andrew Creamer, Bruce Boucek, Birkin Diana, Elli Mylonas and Patrick Rashleigh

Overview

In 2017 Brown University launched its Data Science Initiative (DSI). For the DSI kick-off event, different units on campus were invited to host roundtables with faculty & listen to their ideas for ways to partner with and support their research, teaching, & learning. This poster presents faculty feedback from the University Library (BUL) Data Services roundtable as well as two initiatives that reflect BUL responses to this feedback: 1) a pilot project with Brown's Instructional Technology Group to support undergraduates who are non-computer science concentrators in a data science course & 2) *Software Carpentry* (SwC) training for upskilling BUL staff.

Students using the BUL's Data Visualization Wall in BUL's Patrick Ma Digital Scholarship Lab (DSL); BUL Data Visualization Coordinator is designing a tool for faculty to integrate the wall into their teaching

Acknowledgements

Harriette Hemmasi, Joseph Hogan, Katherine Kinnaird, Jeffrey Brock, Nora Dimmock, Jean Rainwater, Catherine Zabriskie, Peter Tirrel, Phirum Peang, Robert Pelcovits, Kemp Plumb, Thomas Serre, Indra Neil Sarkar, Drew Linsley, TJ Kalaitzidis, John Kromer, Stacey Lawrence, NESCLIC & Software Carpentry, Joshua Dull and Kristin Lee, Kerri Hicks, Crystal Brusch, Joseph Rhoads, Ben Cail, and Ann Caldwell, Birkin Diana, Hector Correa, Steven McCauley, Sarah Evelyn, Carina Cournoyer, Mark Howison, Gurcharan Khanna, Kathi Fisler and Allison Levy

Current BUL Data Services

Research Data Management

- Data set curation & metadata
- Data set retention & archiving in *Brown Digital Repository* (BDR)
- Digital Object Identifiers (DOI) & copyleft licenses consults
- Data publication & citation
- Grant Data Management & Sharing Plans (DMPs)
- Funder Public Access data retention/sharing compliance
- Publisher Data Availability compliance
- Best practices education

Data Visualization

- Data visualization wall
- Software & hardware
- Methods/tools education
- Project consultation

Geographic Information Systems (GIS)

- Licensing software
- Methods/tools education
- Project consultation

Data Wrangling & Analysis

- Quantitative statistical software basics
- Qualitative analysis software (*Nvivo*) basics
- Data wrangling & cleaning (*OpenRefine* & regular expressions)

Other BUL Data-Related Services

- Data reference/searching
- Acquisitions, subscription, & licensing of data sets & databases/resources
- Online course & research guides
- Digital Humanities consultation
- Digital scholarship & pedagogy consultations
- 3D design, modeling and printing

Faculty Feedback

Theme (1): **Faculty** in non-computer science (CS) departments**, across all disciplines**, **want students to incorporate data literacies, tools & methods into their courses that are relevant for students' pre and post-graduate and career success**

Theme (2): **Faculty would like to be aware of data science-related methods & tools** being developed & applied by researchers & educators in their disciplines

Challenge (1)**: Upskilling & support of faculty** to apply data science methods & tools in their own research & teaching

Challenge (2)**: Upskilling & support of students** in non-CS concentrations to apply data science-related concepts, methods & tools to course assignments & projects, i.e., data science literacy

Challenge (3): **Upskilling & professional development of librarians** & hiring of more data science-savvy staff to address the challenges named above

Challenge (4): Workshop offerings distributed across campus units; **lack of awareness**, marketing & promotion, **& coordination among campus units** offering data science-related workshops for faculty, staff, and students

Example Library Responses

Fall Semester 2017 "Module 0"

"Module 0" is in-person support & online content aimed at helping students with minimal computational/programming backgrounds get prepared for a course incorporating data science methods/tools before the course begins.

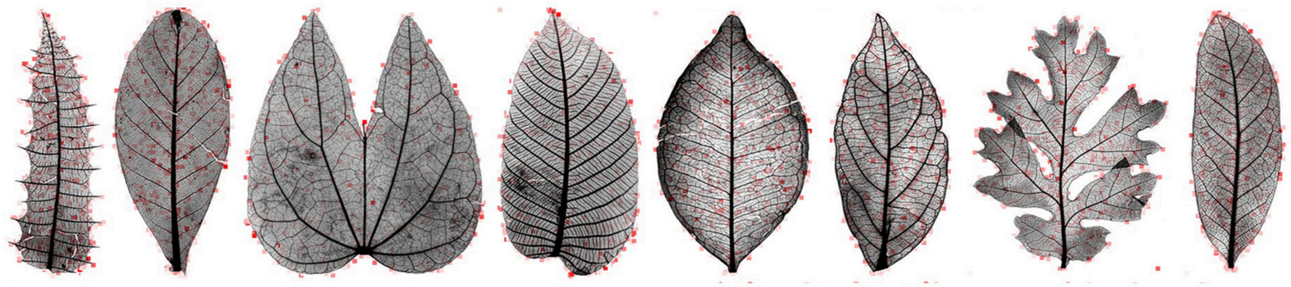
Pilot: A faculty teaching a cognitive science & computer vision course wants BUL and Instructional Technology (ITG) to help support & retain non-CS students in courses applying data science methods such as machine learning.

Module 0 Objectives: 1) download the specified version of *Python* using their preferred operating system, 2) download the *Python* packages required for course 3) set up their *Jupyter Notebooks* & 4) lead a *Python* basics tutorial

Challenges: Sustainability; Scalability; Customization; Scheduling & coordinating staff volunteers & students; Lack of staff with necessary skills to participate; **Lack of ownership: which campus unit should own/sustain this?**

Spring Semester 2018
New England Software Carpentry Libraries Consortium

As a part of NESCLIC, BUL will support two staff members' upskilling by funding their SwC Instructor training. SwC trained-staff members can then hold a Carpentry event for other BUL staff. The SwC curriculum aligns with identified skill gaps using command line for basic computer skills, such as downloading software & managing files & directories. Additionally programming languages *R* & *Python* and their libraries/ packages are increasingly being utilized in courses across disciplines.



Thomas Serre teaches machine learning by using *Python TensorFlow* and *SciKit-Learn* to auto-classify leaves