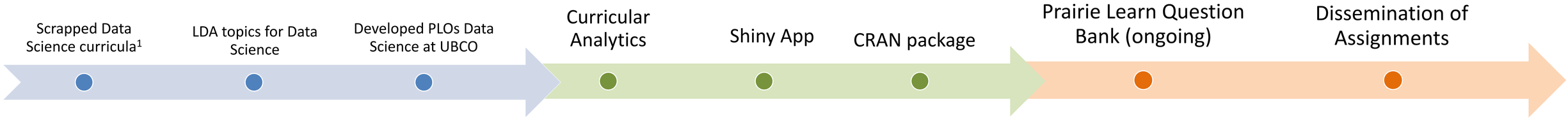
UBCO: Department of Computer Science, Mathematics, Physics, and Statistics (CMPS)

Review and redesign of UBCO's Data Science programs

Dr. Irene Vrbik, Dr. Firas Moosvi, Daniel Krasnov, Seth Akins, Ross Cooper, and Megan Lochhead



Scrapped Data Science Curricula[1]

Program requirements and course calendars for Data Science (DS) curricula from universities across North America were scraped from the web. The information collected includes: course descriptions, credit amounts, prerequisites, and various requirement indicators. For more details see: https://github.com/Hedgemon4/course-scraping

LDA topics for Data Science[2]

LDA, or Latent Dirichlet Allocation [1] was applied to the scrapped data to identify underlying topics in DS curricula. In this context, a topic represents a cluster of words that frequently occur together. The most common terms within these clusters were linked to primary topics: Computer Science, Statistics, and Mathematics (Topics 1, 2, and 3, respectively in Table 1). Details: https://github.com/Danyulll/Curricula-Topic-Modeling

Table 1

The top 3 relevant terms for identifying the 3-group LDA topics.

Topic 1	Topic 2	Topic 3
programming	statistics	mathematics
algorithms	models	applications
software	regression	linear

Developed PLOs

A committee comprised of curriculum specialists[3] and faculty members from Computer Science, Statistics, and Mathematics developed the Program Learning Outcomes (PLOs) for the undergraduate Data Science program at UBCO (approved at the department level in June 2024).

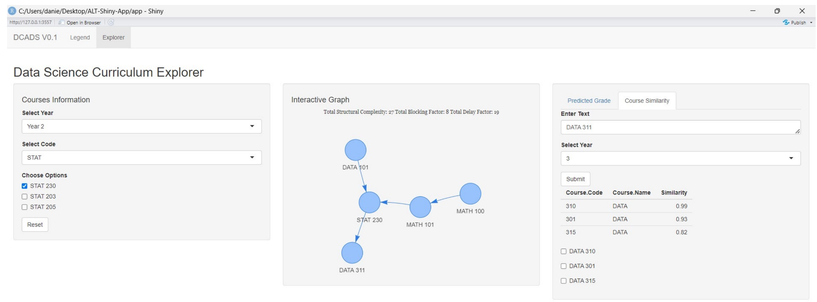
Curricular Analytics (CA)

The undergraduate UBCO DS curriculum was represented and analyzed as a directed acyclic graphs (DAGs) wherein nodes represent courses and directed edges represent pre- or co-requisite relationships. Curricular Analytics [2] are used to quantify the complexity of the program. Metrics include:

- Blocking factor
- Delay factor
- Complexity Score
- Centrality

Figure 1

The prerequisite course structure organized as a DAG. Image was created using the R Shiny App, DCADS, developed by Daniel Krasnov.



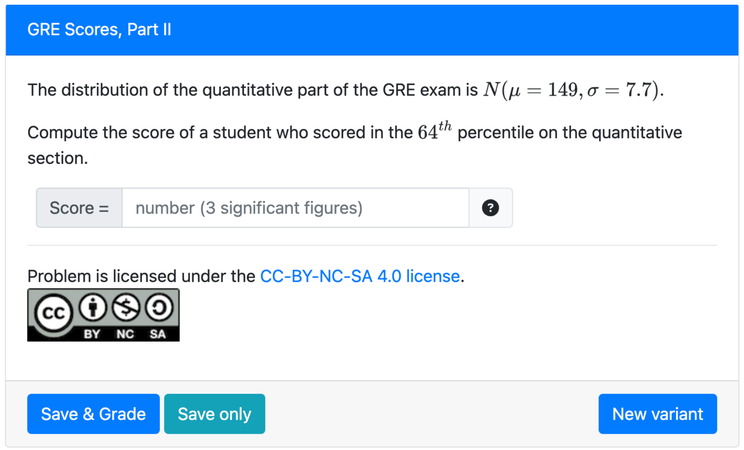
R Shiny App

An interactive R shiny application was developed to easily create visual maps of different pathways through the curriculum and evaluate its complexity through various metrics. Students may also benefit from an integrated recommendation systems[4].

CRAN Package

The implementation of the CA framework was release as an R package on the Comprehensive R Archive Network (CRAN). See: https://cran.r-project.org/web/packages/CurricularAnalytics/

Figure 2: *An example of a particular variant of a question developed for the open-problem bank for introductory statistics.*



Prairie Learn Question Bank[5]

The team[5] has been working over the last few years at growing the open-source question bank to for introductory Statistics courses and Data Science courses in Prairie Learn (PL).

Dissemination of Assignments

We expect to launch PL assignments in STAT 205 in 2024 W2.

Reference / Bibliography

1. Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan): 993–1022.
2. Heileman, G.L., Abdallah, C.T., Slim, A., Hickman, M.: Curricular analytics: A framework for quantifying the impact of curricular reforms and pedagogical innovations. arXiv preprint arXiv:1811.09676. (2018).

Acknowledgement

We gratefully acknowledge the financial support provided for this project by the Aspire-2040 Learning Transformations Fund at UBC Okanagan. Please see the footnotes for credits to individual contributors.

1. Work completed by Seth Akins (Computer Science major; Data Science Minor)
2. Work completed by Daniel Krasnov (Honours Data Science graduate)
3. Facilitated by Meghan Lochhead (Manager Curriculum & Academic Programs) and Jasmine Haley (Curriculum & Acadmic Programs Coordinator). Committee members: Irene Vrbik, Jeff Andrews, Gema Rodriguez Perez, Khalad Mohammad, Pourang Irani, and Eric Foxall.
4. Work in progress by Daniel Krasnov (Honours Data Science graduate) and Ross Cooper (Data Science graduate)
5. Contributors include: Dr. Firas Moosvi, Gavin Kendal-Freedman, Alejandro Builes, Larita Kipkeu, Christiina , Christina Yang and Paula Wong-Chung.