# JSC 370: Data Science II

UNIVERSITY OF TORONTO

| | |
|---|---|
| Term: | Winter/Spring 2024 |
| Time: | M, W 1-3pm |
| Location: | ES B142 |
| Instructor: | Meredith Franklin |
| Contact | meredith.franklin@utoronto.ca |
| Office: | 700 University #9087 |
| Office Hours: | By Appointment |
| TA: | Jenny Du, junni.du@mail.utoronto.ca |

**Course Description**

This course serves as the second in a series of courses on data science. We will focus on the acquisition and analysis of real-life data. Students will learn the toolsets needed to 1) create workable and reproducible data by accessing, scraping, sampling and cleaning data; 2) conduct exploratory data analysis and data visualizations; 3) apply statistical and machine learning tools to learn from data. Coding languages R and Python will be used.

**Learning Objectives**

Through this course, students will become familiar with the techniques used in Data Science. Students will learn:

- Programming in R, and tools Markdown, Git
- Data visualization – summarizing data through interpretable summaries
- Data collection – data scraping, wrangling, cleaning, and sampling
- Exploratory data analysis – generating hypotheses and building intuition
- Statistical algorithms
- Building a github website

**Prerequisite(s):** JSC270H1, STA261H1, MAT237Y1/MAT257Y1, CSC263H1, STA302H1, CSC343H1

**Course Notes**

Lecture notes presented in class will be posted on the course website https://jsc370.github.io/jsc370-2023/

**Technological Proficiency and Hardware/Software Required**

Computation using R (downloaded from http://cran.r-project.org), and development tools including Git (https://github.com/) and Markdown will be used throughout the semester.

**Readings**

1) **R Programming for Data Science**, 2022. Roger Peng. https://bookdown.org/rdpeng/rprogdatascience/

**Supplementary References**

1) **R for Data Science (2e),** 2023 *Garrett Grolemund and Hadley Wickham.* https://r4ds.hadley.nz/
2) **Exploratory Data Analysis with R**, 2020 *Roger Peng* https://bookdown.org/rdpeng/exdata/
3) **Mastering Software Development in R,** 2020 *Roger Peng, Sean Kross, Brooke Anderson* https://bookdown.org/rdpeng/RProgDA/

**Description and Assessment of Assignments**

***Assignments***: There will be 5 assignments given throughout the semester. Students may discuss the problems with one another, however, individual solutions must be submitted and copying will not be tolerated. All assignments must be completed in R Markdown, and submitted through the Github classes portal of the course. Late assignments will be penalized by 20% for each day past the due date.

***Midterm:*** In preparation for the final project you will provide a mid-semester report that details the data you will be using for your final project. Exploratory data analysis, visualizations and summaries of the data will be presented.

***Final Project***: The final project will be to apply the concepts learned in the course to analyze a dataset that you have chosen.

***Labs:*** Lab attendance is mandatory and participation in the lab is required and counts as part of the overall lab grade. The lab assignment will be handed in at the end of the lab or at the end of the lab day if more time is needed.

**Grading Breakdown**

| **Assignment** | **% of Grade** |
|---|---|
| Labs | 10% |
| Homework (5) | 25% |
| Midterm Report | 30% |
| Final Project | 35% |

**Assignment Submission Policy**

Late homework assignments will not be accepted without penalty (20% per day late), except when verifiable extenuating circumstances can be demonstrated.

## Course Schedule: A Weekly Breakdown

| | Topics/Weekly Activities | Due Dates |
|---|---|---|
| **Week 1**<br>January 8 (lecture)<br>January 10 (lab) | Introduction to Data Science tools: R, markdown | |
| **Week 2**<br>January 15 (lecture)<br>January 17 (lab) | Version Control & Reproducible Research, Git | |
| **Week 3**<br>January 22 (lecture)<br>January 24 (lab) | Exploratory Data Analysis | **HW1 Due** |
| **Week 4**<br>January 29 (lecture)<br>January 31 (lab) | Data visualization | |
| **Week 5**<br>February 5 (lecture)<br>February 7 (lab) | Data cleaning and wrangling<br>Machine Leaning 1 | **HW2 Due** |
| **Week 6**<br>February 12 (lecture)<br>February 14 (lab) | Regular Expressions, Big Data, Data scraping, using APIs | |
| **Week 7**<br>February 19/21 | Reading Week | |
| **Week 8**<br>February 26(lecture)<br>February 28 (lab) | Text mining | **HW3 Due** |
| **Week 9**<br>March 5 (lecture)<br>March 7 (lab) | High performance computing, cloud computing | |
| **Week 10**<br>March 12 (lecture)<br>March 14 (lab) | SQL<br>Machine Learning 2 | **HW4 Due** |
| **Week 11**<br>March 19 (lecture)<br>March 21 (lab) | Interactive visualization and effective data communication I | |
| **Week 12**<br>March 26 (lecture)<br>March 28 (lab) | Interactive visualization and effective data communication II | **HW5 Due** |
| **Week 13**<br>April 2 (lecture)<br>April 4 (lab) | Final Project Workshop | |

## Statements on Academic Conduct and Support Systems

Academic Conduct

All suspected cases of academic dishonesty will be investigated following procedures outlined in the *Code of Behaviour on Academic Matters*. If you have questions or concerns about what constitutes appropriate academic behaviour or appropriate research and citation methods, please reach out to me. Note that you are expected to seek out additional information on academic integrity from me or from other institutional resources (for example, the University of Toronto website on Academic Integrity).

Accommodations

The University provides academic accommodations for students with disabilities in accordance with the terms of the Ontario Human Rights Code. This occurs through a collaborative process that acknowledges a collective obligation to develop an accessible learning environment that both meets the needs of students and preserves the essential academic requirements of the University's courses and programs.
Students with diverse learning styles and needs are welcome in this course. If you have a disability that may require accommodations, please feel free to approach me and/or the Accessibility Services* office.
Accessibility Services on the St. George campus

Religious Observances

The University provides reasonable accommodation of the needs of students who observe religious holy days other than those already accommodated by ordinary scheduling and statutory holidays. Students have a responsibility to alert members of the teaching staff in a timely fashion to upcoming religious observances and anticipated absences and instructors will make every reasonable effort to avoid scheduling tests, examinations or other compulsory activities at these times. Please reach out to me as early as possible to communicate any anticipated absences related to religious observances, and to discuss any possible related implications for course work.

Family Care Responsibilities

The University of Toronto strives to provide a family-friendly environment. You may wish to inform me if you are a student with family responsibilities. If you are a student parent or have family responsibilities, you also may wish to visit the Family Care Office website at familycare.utoronto.ca.

Intellectual Property Statement

Course material that has been created by your instructor (i.e. lecture slides, term test questions/solutions and any other course material and resources made available to you on Quercus) is the intellectual property of your instructors and is made available to you for your personal use in this course. Sharing, posting, selling or using this material outside of your personal use in this course is not permitted under any circumstances and is considered an infringement of intellectual property rights.

Land Acknowledgement

A land acknowledgement is a way of honouring the Indigenous people who have lived and worked here for thousands of years, and whose land was colonised. It is also an invitation to reflect on the history of this land and we encourage you to consider the history of the land wherever you are now. https://native-land.ca/