

University of Manitoba
Department of Statistics

SCI 2000—Introduction to Data Science

Winter Term 2021

Course Details

Course Number & Title: SCI 2000—Introduction to Data Science
Section & CRN: Section A01, CRN: 60485
Course Schedule: Tuesday/Thursday, 11:30 am to 12:45 pm,
on Webex.
Prerequisites: Permission from the instructor.

Instructor Contact Information

Instructor: Max Turgeon
Office Location: 373 Machray Hall
Email: Max.Turgeon@umanitoba.ca
Office Hours: By appointment only.

Student Learning Objectives

This course aims to provide students with an introduction to data science. Specifically, this course will introduce you to tools and hands-on experience needed to analyse data. By the end of the course, students will:

- Become proficient in R, to the level that they can analyse data using the tools from this class.
- Be able to describe and analyze data through visualization and simple statistical procedures.
- Be introduced to statistical thinking and be able to think critically about variation and biases.

The course material will be illustrated using the statistical programming language R, but students are allowed to submit their work in Python, if preferred.

Textbook and Other Materials

Textbook: There is **no** textbook for this course. Notes will be provided to students through UM Learn, along with additional resources.
Course Material: Lecture videos, lecture slides and additional resources will all be made available on UM Learn.

Outline of Covered Topics

The course is expected to cover the following topics, as time permits:

1. Data visualization
2. Data wrangling
3. Relational data
4. Web scraping and APIs
5. Introduction to regular expressions
6. (If time permits) Automation and Version control

Throughout the course, the applied topics above will be complemented with an introduction to statistical thinking: how to think about variability, what biases can occur in the data, and how to perform simple statistical procedures (e.g. comparing means, proportions, linear regression).

Course Assessment

- Assignments:** There will be four (4) assignments during the term, worth a total of 40% of the final grade.
Students are encouraged to form study groups to discuss assignment questions but not the answers. Each student must submit his or her own individual written work. Copying, in whole or in part, the work of another will not be tolerated and will result in disciplinary action (see Academic Integrity section). Assignments should be handed in electronically via UM Learn on the due date. **Late submissions will be accepted** up to five (5) days after the submission deadline, but 5% will be deducted from your assignment mark for each 24h period following the submission deadline.
- Term project:** There will be a term project, worth 50% of the final mark. This project will be done in teams of 2-3 students. Further details will be provided later.
- Final Exam:** There will be **no** final exam in this course.
- Summaries:** Three times during the semester, the instructor and the students will spend the lecture collaborating on a data analysis. Students will then have until the next day to submit a short summary about what they found out and what they learned during the data analysis. This assessment will represent 5% of the final mark.
- Class participation:** Students will be assessed according to their level of participation in the lectures and/or on the UM Learn Discussion groups. This assessment will represent 5% of the final mark.
- Grading Timeline:** Work will be graded and returned within two weeks of submission.

Course Evaluation and Grading Scheme

Final Mark: The final mark for the course will be obtained according to the following rule:

Assignments (4)	40%
Term project	50%
Summaries (3)	5%
Class Participation	5%

Letter Grade: I normally follow the following cutoffs when assigning letter grades:

Letter Grade	Mark out of 100
A+	[90 – 100]
A	[80 – 90)
B+	[75 – 80)
B	[70 – 75)
C+	[65 – 70)
C	[60 – 65)
D	[50 – 60)
F	below 50

However, I might elect to use lower thresholds for some letter grades if I think they are more appropriate (i.e. use a smaller lower bound for the ranges above). I will not use higher thresholds.

Important Dates

The following dates are important to how the course will progress throughout the term.

Date	Information
Jan 19	First lecture
Feb 15-19	Winter Term break–no class
Mar 31	Last day to VW the course
Apr 15	Last lecture
Apr 19-May 1	Final Examination Period

Changes are subject to Section 2.8 of the ROASS Procedure.

Tentative Schedule

The following schedule is tentative and may be adjusted as needed during the semester.

	Date	Topic	Comments
1	Tue	Jan 19	First lecture
2	Thu	Jan 21	Descriptive statistics
3	Tue	Jan 26	Data Visualization
4	Thu	Jan 28	Data Visualization
5	Tue	Feb 2	Data Analysis #1 Summary 1 due on Feb 3
6	Thu	Feb 4	Linear Regression
7	Tue	Feb 9	Linear Regression
8	Thu	Feb 11	Linear Regression Assignment 1 due on Feb 12
9	Tue	Feb 23	Relational data
10	Thu	Feb 25	Relational data
11	Tue	Mar 2	Intro to RDBMS
12	Thu	Mar 4	Intro to Regex Assignment 2 due on Mar 5
13	Tue	Mar 9	Sentiment Analysis
14	Thu	Mar 11	Data Analysis #2 Summary 2 due on Mar 12
15	Tue	Mar 16	Web scraping
16	Thu	Mar 18	Web scraping
17	Tue	Mar 23	Web scraping
18	Thu	Mar 25	Web scraping Assignment 3 due on Mar 26
19	Tue	Mar 30	Data Analysis #3 Summary 3 due on Mar 31
20	Thu	Apr 2	Project
21	Tue	Apr 6	Extra Topics #1
22	Thu	Apr 8	Project Assignment 4 due on Apr 9
23	Tue	Apr 13	Extra Topics #2
24	Thu	Apr 15	Project

Extra topics may include automation, version control, and/or logistic regression.

Class Communications

The University requires all students to activate an official U of M email account, which should be used for all communications between yourself and the university (including all your instructors). All these email communications should comply with the University's policy on electronic communication with students, which can be found at: http://umanitoba.ca/admin/governance/governing_documents/community/electronic_communication_with_students_policy.html

Questions of an administrative nature should be directed to me via email. Questions related to the course content should be directed to the Discussion Groups on UM Learn (which I will regularly visit). This is in order to provide an opportunity for learning and collaboration between the students.

In general, when a faculty or staff member receives an email from a student they will make an effort to reply within one business day of receipt of the email. It is understood that sometimes the reply may come sooner or in some instances later than this, with a normal response within three business days. Sometimes it will be useful for the faculty or staff member to reply on a slightly longer time scale so they can create a reply that covers multiple inquiries efficiently and effectively, in some instances by posting a response to UM Learn and pointing students to that response. In other cases the replies will be individualized, and a shorter turn-around time may be feasible.

Copyrights

We will use copyrighted content in this course. I have ensured that the content I use is appropriately acknowledged and is copied in accordance with copyright laws and University guidelines. Copyrighted works, including those created by me, are made available for private study and research and must not be distributed in any format without permission.

More details are available online at <http://umanitoba.ca/copyright/>.

Student Accessibility Services

If you are a student with a disability, please contact Student Accessibility Services (SAS) for academic accommodation supports and services such as note-taking, interpreting, assistive technology and exam accommodations. Students who have, or think they may have, a disability (e.g. mental illness, learning, medical, hearing, injury-related, visual) are invited to contact SAS to arrange a confidential consultation.

ROASS Schedule A

Schedule “A” of the *Responsibilities of Academic Staff with regards to Students* (ROASS) policies of the University of Manitoba lists resources and policies for students. It is important that you familiarize yourself with these resources and policies. This document is available from the Department of Statistics web page at: <http://umanitoba.ca/science/statistics/>.

Students enrolled in this course must ensure they satisfy the following minimum technological requirements:

1. A computing device where one can create and edit documents,
2. An internet connection capable of streaming videos and downloading software, and
3. Access to a web-cam and microphone.

University of Manitoba Acknowledgement of Traditional Territories

The University of Manitoba campuses are located on original lands of Anishinaabeg, Cree, Oji-Cree, Dakota, and Dene peoples, and on the homeland of the Métis Nation.

We respect the Treaties that were made on these territories, we acknowledge the harms and mistakes of the past, and we dedicate ourselves to move forward in partnership with Indigenous communities in a spirit of reconciliation and collaboration.