Course Organization

DATA 201: Thinking With Data
Winter 2021

Jonathan Hudson, Ph.D
Instructor
Department of Computer Science
University of Calgary

Thursday, January 7, 2021

UNIVERSITY OF CALGARY

Welcome!

Jonathan Hudson, Ph.D (he/him)

Lectures: Mondays and Wednesdays 13:00-13:50

(D2L->Communication->Zoom, recorded)

Office: ICT 712 (I will not be in it!)

Office hours: 14:00-14:50 PM Mondays and Wednesdays

(D2L->Communication->Zoom, not recorded)

or by email-scheduled appointments.

jwhudson@ucalgary.ca

https://pages.cpsc.ucalgary.ca/~jwhudson/DATA201W21/

Tutorials

Start next week Monday.

Through D2L, will be recorded.

Point is active interaction with TA for material, exercises, and assignment help.

Use link to your tutorial only.

Your enrollment tutorial TA will mark your assignment material and they are only responsible for the students enrolled in their tutorial.

Why Data Science?

- In computer science we like to define **information** and **data** and a definition for computer science is **information theory** using algorithmic tools.
- Data is numbers, and information is data with context.
- Data science is about how do we get data, clean data, transform data, explore data, create new data, and communicate data to others. All of these are informed by the context of the data.

- Almost every field of study stores information as data for different reasons.
- Skills to handle data are highly transferrable.
- Opportunities for multidisciplinary study, work, and research

Course Goal

From the calendar:

- “An introduction to tools and techniques for managing, visualizing, and making sense of data. Includes an introduction to data cleaning, basic statistics, exploratory visualization, sensemaking, and data presentation.

Goals:

- Collecting data and understanding it.
- Cleaning and transforming data to make it more useful.
- Data analysis to create more data from existing data.
- Visualization of data so it can be communicated.
- Ethical considerations through-out data science.

Course Goal

Critical Thinking with Data

Data Analysis Pipeline

Collection => Cleaning => Analysis => Presentation

Lectures

We will learn fundamentals of data science (there is no computer programming)

We will cover:

- Visualization (non-digital/digital)
- Obtaining data
- Data cleaning
- Charts
- Presentations
- Quantitative/Qualitative Analysis

Out of lecture?

There is no attendance at labs but they are highly recommended

- Start Monday/Tuesday next week
- TAs will use labs to cover assignment material in hands-on environment
- Particularly technology tools to be used in assignments/project
- Your groups will be formed within your lab for the group project.

Grading

Component	Weighting %
Assignments (4)	15% each
Group Project	40%

- Each piece of work (reports, assignments, quizzes, midterm exam(s) or final examination) submitted by the student will be assigned a grade. The student's grade for each component listed above will be combined with the indicated weights to produce an overall percentage for the course, which will be used to determine the course letter grade.
- The conversion between a percentage grade and letter grade is as follows.

	A+	A	A-	B+	B	B-	C+	C	C-	D+	D
Minimum % Required	95 %	90 %	85 %	80%	75%	70 %	65 %	60%	55%	50 %	45 %

Assignments

- Four individual assignment (15%) consists of different data science skills
- Each assignment is due at 11:59 pm on the Friday due date.
- Current deadlines

Assignments	Due at 23:59
Assignment 1	Jan 29
Assignment 2	Feb 12
Assignment 3	Mar 5
Assignment 4	Mar 19

Group Projects

- Groups Formed: Tuesday, March 2nd, 2021: 23:59 (11:59PM)
- Proposal: Friday, March 12th, 2021: 23:59 (11:59PM)
- Presentations: March 24th-April 14th
- Report: Thursday, April 15th, 2021: 23:59 (11:59PM)
- Topic: Data Analysis Pipeline

Course Policies

- When you email include your first name, and last name.
- Please use “DATA201W21” as the prefix in the subject line

- Assignments are accepted within two days late with a -10 mark per day penalty. Assignments are out of 50 so this is -20% penalty per day. And 0 after 48 hours. Penalty steps are by day, 1 hour late is 1 day late.
- Submit early and double check after submitting. You can submit multiple times on D2L with no issue, so excuses will not be accepted. You can also download your submission from D2L to check it submitted correctly.

Zoom Norms

- Respect others:
  - Keep your zoom muted unless asking a question. (please indicate in chat you have a question and I'll make time to let you ask)
  - Video is not necessary. However, for office hours and even smaller tutorials it is recommended.
  - You can ask questions via chat at any time. Ability to answer will be time and class pace dependent.
  - Arrive on time.
  - Refrain from using the chat for topics not related to the current material.
  - Use directed chat if you chat with someone you know. (Be aware that the directed chat is not private and can be seen by owner of class recording!)
  - Avoid any activity that might disturb your classmates.

Academic Dishonesty

- *“A single offence of cheating, plagiarism, or other academic misconduct, on term work, tests, or final examinations, etc., may lead to disciplinary probation or a student's suspension or expulsion from the faculty by the Dean, if it is determined that the offence warrants such action. ”*

- Please refer to the University Calendar for more details.

- Paid or free online services that complete work for students is plagiarism. Your work must be your own on assignments, and on group projects must be only from the members of your group.

Getting Help

- Do your part: Attend the lectures and tutorials
- Act early!
- First try it yourself→
  - Study the material carefully
  - Break the problem down
  - Try to narrow down the question
  - Search on google for your answer
- Still unclear?
- Ask your TA
- Come to my office ☺

Crisis line!

- If you think:
  - You suck at data science!
  - You suck at data science tool!
  - You are not sure about this course!
  - You are OK with only a passing mark!!!
  - You tried but you didn’t understand!

- Come to my office → I’ll prove to you that you are wrong!

- Come early before things piled up!

Technology?

Surveys

- Qualtrics
- Microsoft Forms
- Google Forms

Data Cleaning

- Microsoft Excel
- OpenRefine

Data Visualization

- Microsoft Excel
- Tableau

Onward to …
Overview.

Jonathan Hudson
jwhudson@ucalgary.ca
https://pages.cpsc.ucalgary.ca/~jwhudson/

UNIVERSITY OF CALGARY