



DATA*6200 Data Manipulation and Visualization Fall 2023

0 COURSE PREAMBLE

Mental Health. One out of every five students in Canada experiences some sort of mental health issue at some point in their academic career. If you find yourself facing a mental health crisis, or just need to talk to someone, please consider taking advantage of one of the following resources available to University of Guelph students:

Counseling Services: (x53244) is located at Health Services (J.T. Powell Building) and offers individual and group counseling sessions by appointment or walk-in.

Student Support Network is located in the Wellness & Education Promotion Centre (inside the J.T. Powell Building) and offers confidential, peer-based, drop-in support.

Good2Talk: (1-866-925-5454) is a free, 24/7 student hotline that provides professional counseling and referrals for mental health, addictions and well-being.

Here 24/7: (1-844-437-3247) specializes in assessment, referral and appointment booking and is available 24/7 for crisis support.

You are not alone and you will not be judged for asking for help.

For information on current safety protocols, follow these links:

<https://news.uoguelph.ca/return-to-campuses/how-u-of-g-is-preparing-for-your-safe-return/>
<https://news.uoguelph.ca/return-to-campuses/spaces/\#ClassroomSpace>

Please note, these guidelines may be updated as required in response to evolving University, Public Health or government directives. Medical notes will not normally be required for singular instances of academic consideration, although students may be required to provide supporting documentation for multiple missed assessments or when involving a large part of a course (e.g. final exam or major assignment).

1 AIMS & OBJECTIVES

1.1 Calendar Description

This course provides a hands-on introduction to exploratory data analysis using Python. Manipulation and visualization techniques for evaluating statistical and machine learning models, data acquisition, data integrity, and good analysis practices are discussed. Students will gather insights on interpreting model accuracy metrics, selecting data subsets, addressing inherent biases in data, and the ethical considerations when summarizing data.

1.2 Course Description

An analysis is only as good as the data on which it is based and the tables and graphics used to summarize the data and analysis results. Data manipulation and visualization are key components to the data life cycle, and often where most of the analyst's time is spent. This course will cover the basics of accessing, cleaning, re-shaping, re-formatting, transforming, selecting, filtering and tidying data in preparation for effective analysis. Generating informative and powerful data tables/graphics will be introduced for checking and maintaining data integrity as well as for insightful exploratory data analysis. Critical assessment of data tables and visualizations as well as good analysis practices will be developed, including review of ethical considerations arising from manipulation and visualization of data. The course ends with the acquisition and manipulation of spatially structured data. Mastering these skills can facilitate and drastically simplify downstream analyses of data.

Course topics include:

- Review of data basics: data types, objects, data structures, and reading/writing data
- Fundamentals of data manipulation: select, filter, transform, arrange, mutate and summarize
- Tools and techniques for data manipulation: group-wise calculations, joining tables, piping, chain operators, working with distributed and backend databases, dealing with dates and character strings
- Fundamentals of data storytelling: visual thinking and visualization design including focus, data, aesthetics, scale, statistics, and facets
- Tools and techniques for visualization based on grammar of graphics for multi-dimensional data
- Introduction to spatially structured data: geoprocessing and mapping for raster data, including aggregation, reclassification, cropping, masking and extraction
- Ethical issues including intentional and inadvertent deception, P-hacking, cherry-picking data, and inadvertent and intentional violation of privacy.

Prerequisite(s): Enrollment in the data science program or permission of instructor.

Credit Weight: 0.5

Academic Department: Mathematics & Statistics

1.3 Learning Outcomes

Upon successful completion of this course, students will have demonstrated the ability to:

1. Perform data import and pre-processing of raw data: inspection, cleaning, verification, and reporting.
2. Process disparate multi-dimensional data using selection, filtering, rearranging, grouping, transformation.
3. Understand the fundamentals of effective tables and graphs for diverse data types.
4. Demonstrate understanding of good and ethical data practices.
5. Understand implications of data manipulation and visualization.
6. Create insightful data graphics using contemporary software tools and packages and appropriate design choices.
7. Understand what a dashboard is and its role in knowledge translation.
8. Demonstrate basics of importing, geoprocessing and mapping of spatial data.
9. Generate basic data graphics for spatial data over time.

2 TEACHING AND LEARNING ACTIVITIES

2.1 Course Topics and Schedule

Week	Topic
September 7 - September 21	Common statistical errors, interpreting tabular medical data and evaluating statistical models using visualizations, and the ethics of data manipulation.
September 26 - September 28	Exploratory data analysis of Co-operators Case Study.
October 3	Student presentations on Case Study 1.
October 5 - October 19	Manipulation of astronomy and medical image data, data bias, selection of data subsets, and interpreting and evaluating machine learning models with visualizations.
October 24 - October 26	Tutorial on creating a dashboard in Google Locker Studio.
October 31 - November 9	Visualization of spatial-temporal data and correlations in Google Earth Engine, data cleaning and processing methods for agricultural yield data, QGIS, interpreting and evaluating agriculture yield prediction model outcomes, and addressing inherent biases in data and ethical considerations.
November 14	Feedback on project progress.
November 16	Data summaries and Grammar of Graphic, ethical concerns, privacy, optics, and implications.
November 21 - November 30	Student presentations of final project progress reports.

3 LEARNING RESOURCES

3.1 Course Website

Course material, news, announcements, and grades will be regularly posted to the DATA*6200 CourseLink site. You are responsible for checking the site regularly.

3.2 Readings and Resources

Required Text: None

Recommended Texts:

1. Hadley Wickham and Garrett Grolemund. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*, 1st Ed., O'reilly Media Inc., 2017. Claus O.
2. Wilke, *Fundamentals of Data Visualization*, O'reilly Media Inc., 2019.
3. Other relevant readings may/will be provided throughout the semester.

4 ASSESSMENT

4.1 Dates and Distribution

Assignment	Due Date	Weighting
Case Study 1 Presentation	Oct 3	10%
Case Study 1 Report	Oct 3	20%
Project - Proposal	Oct 31	10%
Project - Report	Nov 20	25%
Project - Presentation	Nov 21 - Nov 30	30%
Project - Q and A Participation	Nov 21 - Nov 30	5%

4.2 Course Grading Policies

Passing grade: In order to pass the course, students must obtain a grade of 65% or higher on the total mark of all assessments.

Course Policy on Group Work: Students are encouraged to work together but must each write up individually the material they turn in. Work on the final project may be done in consultation with other students or with faculty but the final project must be completely written by the student.

Accommodation of Religious Obligations: If you are unable to meet an in-course requirement due to religious obligations, please email the course instructor within two weeks of the start of the semester to make alternate arrangements. See the graduate calendar for information on regulations and procedures for Academic Accommodation of Religious Obligations:

[https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/sec_d0e2228.shtml]

5 UNIVERSITY STATEMENTS

5.1 Email Communication

As per university regulations, all students are required to check their e-mail account regularly; e-mail is the official route of communication between the University and its students.

5.2 When You Cannot Meet a Course Requirement

When you find yourself unable to meet an in-course requirement because of illness or compassionate reasons, please advise the course instructor (or designated person, such as a teaching assistant) in writing, with your name, id#, and e-mail contact. See the Undergraduate Calendar for information on regulations and procedures for [Academic Consideration](#).

Academic Consideration, Appeals and Petitions

[<https://calendar.uoguelph.ca/undergraduate-calendar/undergraduate-degree-regulations-procedures/academic-consideration-appeals-petitions/>]

5.3 Drop Date

Courses that are one semester long must be dropped by the end of the last class day; two-semester courses must be dropped by the last day of class in the second semester. The regulations and procedures for changing graduate course registration are available in the Undergraduate and Graduate Calendars.

Undergraduate Calendar - Dropping Courses

[<https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-drop.shtml>]

Graduate Calendar - Registration Changes

[<https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/genreg-reg-regchg.shtml>]

5.4 Copies of Out-of-class Assignments

Keep paper and/or other reliable back-up copies of all out-of-class assignments; you may be asked to resubmit work at any time.

5.5 Accessibility

The University promotes the full participation of students who experience disabilities in their academic programs. To that end, the provision of academic accommodation is a shared responsibility between the University and the student.

When accommodations are needed, the student is required to first register with Student Accessibility Services (SAS). Documentation to substantiate the existence of a disability is required, however, interim accommodations may be possible while that process is underway.

Accommodations are available for both permanent and temporary disabilities. It should be

noted that common illnesses such as a cold or the flu do not constitute a disability.

Use of the SAS Exam Centre requires students to make a booking at least 14 days in advance, and no later than November 1 (fall), March 1 (winter) or July 1 (summer). Similarly, new or changed accommodations for online quizzes, tests and exams must be approved at least a week ahead of time.

More information: www.uoguelph.ca/sas

5.6 Academic Integrity

The University of Guelph is committed to upholding the highest standards of academic integrity, and it is the responsibility of all members of the University community—faculty, staff, and students—to be aware of what constitutes academic misconduct and to do as much as possible to prevent academic offenses from occurring. University of Guelph students have the responsibility of abiding by the University's policy on academic misconduct regardless of their location of study; faculty, staff, and students have the responsibility of supporting an environment that discourages misconduct. Students need to remain aware that instructors have access to and the right to use electronic and other means of detection.

Please note: Whether or not a student intended to commit academic misconduct is not relevant for a finding of guilt. Hurried or careless submission of assignments does not excuse students from responsibility for verifying the academic integrity of their work before submitting it.

Students who are in any doubt as to whether an action on their part could be construed as an academic offense should consult with a faculty member or faculty advisor.

Undergraduate Calendar - Academic Misconduct

[<https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-amisconduct.shtml>]

Graduate Calendar - Academic Misconduct

[https://www.uoguelph.ca/registrar/calendars/graduate/2018-2019/genreg/sec_d0e2632.shtml]

Inappropriate online behavior will not be tolerated. Examples of inappropriate online behavior include:

- Posting inflammatory messages about your instructor or fellow students
- Using obscene or offensive language online
- Copying or presenting someone else's work as your own
- Adapting information from the Internet without using proper citations or references
- Buying or selling term papers or assignments
- Posting or selling course materials to course notes websites
- Having someone else complete your quiz or completing a quiz for/with another student
- Making false claims about lost quiz answers or other assignment submissions
- Threatening or harassing a student or instructor online
- Discriminating against fellow students, instructors or TAs
- Using the course website to promote profit-driven products or services

- Attempting to compromise the security or functionality of the learning management system
- Sharing your username and password
- Recording lectures without the permission of the instructor

5.7 Recording of Materials

Presentations that are made in relation to course work—including lectures—cannot be recorded or copied without the permission of the presenter, whether the instructor, a student, or guest lecturer. Material recorded with permission is restricted to use for that course unless further permission is granted.

5.8 Other University Resources

The Academic Calendars [<https://www.uoguelph.ca/academics/calendars>] are the source of information about the University of Guelph's procedures, policies, and regulations that apply to undergraduate, graduate, and diploma programs.