



# CAPILANO UNIVERSITY

COURSE OUTLINE		
TERM: FALL 2022	COURSE NO: COMP 330	
INSTRUCTOR:	COURSE TITLE: DATA WRANGLING: SCRIPTING AND AUTOMATED DATA PROCESSING	
OFFICE: LOCAL: E-MAIL: @capilanou.ca	SECTION NO(S):	CREDITS: 4.0
OFFICE HOURS:		
COURSE WEBSITE:		

Capilano University acknowledges with respect the Lil'wat, Musqueam, Squamish, Sechelt, and Tsleil-Waututh people on whose territories our campuses are located.

## COURSE FORMAT

1.5 hours of class time, 3 hours lab time, plus an additional 1.5 hours delivered through on-line or other activities for a 15-week semester, which includes two weeks for final exams.

## COURSE PREREQUISITES

45 credits of 100-level or higher coursework and one of COMP 115 (A-) or COMP 215

## CALENDAR DESCRIPTION

Students are introduced to the principles, concepts, approaches, technologies, and practices of automated data acquisition, cleaning, transformation, merging, matching, storage, and analysis. Students learn to apply scripting languages to pattern matching, data pipelines, and other data wrangling techniques. The course focusses on practical uses for these technologies and techniques in a range of domains, including science, business, and computing; and includes an introduction to operating systems, command line shells, filesystems, and file formats.

## COURSE NOTES

COMP 301 is an approved Science and Technology course for Cap Core requirements.

COMP 330 is an approved Quantitative/Analytical course for baccalaureate degrees.

COMP 301 is an approved Science course.

## REQUIRED TEXTS AND/OR RESOURCES

Robbins and Beebe, Classic Shell Scripting, O'Reilly 2005.

Sweigart, Al. Automate the Boring Stuff with Python. No Starch Press, 2015. <https://automatetheboringstuff.com/> (Ch. 7 + 11-17)

Severance, Charles. Python for Everybody. CreateSpace, 2016. <https://www.py4e.com/book.php> (Ch. 11 – 16)

Shotts, William. The Linux Command Line (5<sup>th</sup> ed.) No Starch Press, 2019. <http://www.linuxcommand.org/tlcl.php/tlcl.php>

## COURSE STUDENT LEARNING OUTCOMES

**On successful completion of this course, students will be able to do the following:**

- identify the steps required to extract, clean, structure, and transform a raw data set to a useable form;
- acquire data from and load data to a variety of sources, including sensors, web API's, databases, filesystem, spreadsheets and other documents;
- work comfortably with a number of common structured data exchange formats, including CSV, XML, and JSON;
- apply regular expressions to solve complex pattern matching problems;
- write short Unix shell scripts to automate common O/S tasks;
- write small data processing scripts in a variety of languages (e.g., sed, awk, perl, and/or python) for extracting, matching, replacing, merging data from structured and unstructured text files;
- choose an appropriate scripting language for a particular data wrangling task;
- write the 'glue' code to construct a data processing pipeline from existing software tools;
- apply data wrangling techniques within their own field of study.

**Students who complete this Science and Technology course will be able to do the following:**

- Apply numerical and computational strategies to solve problems;
- Assess the cultural, economic, and political effects of technology;
- Evaluate scientific information (e.g., distinguish primary and secondary sources, assess credibility and validity of information);
- Demonstrate how a problem, concept, or process can be modelled numerically, graphically, or algorithmically;
- Participate in scientific inquiry and communicate the elements of the process, including making careful and systematic observations, developing and testing a hypothesis, analyzing evidence, and interpreting results.

## COURSE CONTENT

# of Weeks (Approx.)	Topics
1	Introduction to Unix Command Line Interface (CLI): Interactive shell, filesystem navigation, permissions, editors
2	Data sources: capture, extract, mine, and crawl Sensors, filesystems, web APIs, databases
1	Structured Data formats: CSV, XML, JSON
1	Unstructured and binary data formats
2	Regular expressions: search, match, replace

# of Weeks (Approx.)	Topics
2	Stream processing languages / tools: perl, sed, awk, grep
3	Data pipelines and glue code: Acquire, clean, structure, transform, analyse, store data streams
1	Review and Testing
(2)	Final Exam Period (weeks 14 and 15)

### EVALUATION PROFILE

Final grades for the course will be computed based on the following schedule:

Assignments + Quizzes	20%
Projects*	40%
Term Test	15%
Final Exam (comprehensive)	20%
Performance Evaluation	5%
<b>TOTAL</b>	<b>100%</b>

\*No single project will be worth more than 35% of the total marks.

- Projects are completed in small groups and are peer-reviewed. Students receive a grade for both their project work and for their evaluation of their peer's work.
- The default **performance evaluation** component in the evaluation profile is pro-rated to the grade earned on the remainder of the profile. In exceptional circumstances, a student's improved performance in the later part of the term may justify an elevated grade. The instructor has sole discretion, in such cases, to alter the performance evaluation to elevate the student's grade to better reflect their performance at the end of term.

### GRADING PROFILE

A+ 90 - 100%	B+ 77 - 79%	C+ 67 - 69%	D 50 - 59%
A 85 - 89%	B 73 - 76%	C 63 - 66%	F 0 - 49%
A- 80 - 84%	B- 70 - 72%	C- 60 - 62%	

### Incomplete Grades

Grades of Incomplete "I" are assigned only in exceptional circumstances when a student requests extra time to complete their coursework. Such agreements are made only at the request of the student, who is responsible to determine from the instructor the outstanding requirements of the course.

**Late Assignments**

Assignments are due at the beginning of the class on the due date listed. Late assignments are penalized 10% if submitted late on the due date, then 20% for each succeeding day until solutions are posted, and will not be accepted thereafter.

If you anticipate handing in an assignment late, please consult with your instructor beforehand.

**Missed Exams/Quizzes/Labs/etc.**

A score of zero is normally given for missed work. Make-up exams or quizzes may be permitted, at the discretion of the instructor, and generally only in cases of medical emergency or severe personal crisis. In some cases, it may not be possible to accommodate a missed exam or quiz. Please consult with your instructor, ideally *before* the missed activity.

**Attendance**

Students are expected to attend and fully participate in all classes, labs, and associated activities. Students are responsible for all information given during lectures, labs, and tutorials, including exam dates and assignment deadlines, even if they were unable to attend for any reason.

**English Usage**

Students are expected to proofread all written work for any grammatical, spelling and stylistic errors. Instructors may deduct marks for incorrect grammar and spelling in written assignments.

**Electronic Devices**

Students may use electronic devices during class solely for class-related activities, such as note-taking, coding, and 'just in time' research to contribute to class discussions.

**On-line Communication**

Outside of the classroom, instructors will (if necessary) communicate with students using either their official Capilano University email or eLearn; please check both regularly. Official communication between Capilano University and students is delivered to students' Capilano University email addresses only.

**UNIVERSITY OPERATIONAL DETAILS:****Tools for Success**

Many services are available to support student success for Capilano University students. A central navigation point for all services can be found at: <https://www.capilanou.ca/student-life/>

**Capilano University Security: download the [CapU Mobile Safety App](#)**

**Policy Statement (S2009-06)**

Capilano University has policies on Academic Appeals (including appeal of final grade), Student Conduct, Academic Integrity, Academic Probation and other educational issues. These and other policies are available on the University website.

**Academic Integrity (S2017-05)**

Any instance of academic dishonesty or breach of the standards of academic integrity is serious and students will be held accountable for their actions, whether acting alone or in a group. See policy and procedures S2017-05 Academic Integrity for more information: <https://www.capilanou.ca/about-capu/governance/policies/>

Violations of academic integrity, including dishonesty in assignments, examinations, or other academic performances, are prohibited and will be handled in accordance with the Student Academic Integrity Procedures.

**Academic dishonesty** is any act that breaches one or more of the principles of academic integrity. Acts of academic dishonesty may include but are not limited to the following types:

**Cheating:** Using or providing unauthorized aids, assistance or materials while preparing or completing assessments, or when completing practical work (in clinical, practicum, or lab settings), including but not limited to the following:

- Copying or attempting to copy the work of another during an assessment;
- Communicating work to another student during an examination;
- Using unauthorized aids, notes, or electronic devices or means during an examination;
- Unauthorized possession of an assessment or answer key; and/or,
- Submitting of a substantially similar assessment by two or more students, except in the case where such submission is specifically authorized by the instructor.

**Fraud:** Creation or use of falsified documents.

**Misuse or misrepresentation of sources:** Presenting source material in such a way as to distort its original purpose or implication(s); misattributing words, ideas, etc. to someone other than the original source; misrepresenting or manipulating research findings or data; and/or suppressing aspects of findings or data in order to present conclusions in a light other than the research, taken as a whole, would support.

**Plagiarism:** Presenting or submitting, as one's own work, the research, words, ideas, artistic imagery, arguments, calculations, illustrations, or diagrams of another person or persons without explicit or accurate citation or credit.

**Self-Plagiarism:** Submitting one's own work for credit in more than one course without the permission of the instructors, or re-submitting work, in whole or in part, for which credit has already been granted without permission of the instructors.

**Prohibited Conduct:** The following are examples of other conduct specifically prohibited:

- Taking unauthorized possession of the work of another student (for example, intercepting and removing such work from a photocopier or printer, or collecting the graded work of another student from a stack of papers);
- Falsifying one's own and/or other students' attendance in a course;
- Impersonating or allowing the impersonation of an individual;

- Modifying a graded assessment then submitting it for re-grading; or,
- Assisting or attempting to assist another person to commit any breach of academic integrity.

### **Sexual Violence and Misconduct**

All Members of the University Community have the right to work, teach and study in an environment that is free from all forms of sexual violence and misconduct. Policy B401 defines sexual assault as follows:

Sexual assault is any form of sexual contact that occurs without ongoing and freely given consent, including the threat of sexual contact without consent. Sexual assault can be committed by a stranger, someone known to the survivor or an intimate partner.

Safety and security at the University are a priority and any form of sexual violence and misconduct will not be tolerated or condoned. The University expects all Students and Members of the University Community to abide by all laws and University policies, including B.401 Sexual Violence and Misconduct Policy and B.401.1 Sexual Violence and Misconduct Procedure (found on Policy page <https://www.capilanou.ca/about-capu/governance/policies/>)

**Emergencies:** Students are expected to familiarise themselves with the emergency policies where appropriate and the emergency procedures posted on the wall of the classroom.

### **DEPARTMENT OR PROGRAM OPERATIONAL DETAILS:**

#### **Computer Access**

Students may bring and use their own computing devices, running any modern OS (i.e., Windows, OSX, or Linux). Every effort is made to ensure that required course software can be freely downloaded and installed on student computer. However, it is the responsibility of each student to ensure their computer meets the minimum requirements of required course software, and to perform the installation and configuration of such software themselves. Computer labs at the University will have course-required software installed and configured; students may use lab computers to complete all their course work. Drop-in access to the University computers is available during the hours posted outside each lab, subject to availability. Please respect an instructor's directions if asked to leave the lab due to a class booking.

University policies on student conduct and use of University computer systems, available on the University website, will be strictly enforced.