The Data Science Toolbox
Syllabus

The Data Science Toolbox

Time and Place

Online, October 11 to December 12th, 2022

Description

Welcome to The Data Science Toolbox! This course is part of the Key Capabilities for Data Science program and covers topics related to workflows, platforms and tools used in data analysis.

In this course, we will dive into the world of data science tools and utilities. While these are not strictly required for data analysis, they are necessary for efficient, reproducible, and collaborative data science practices, and are all important building blocks for a successful and sustained data science career.

Prerequisites

Programming in Python for Data Science

Content

We are using an interactive platform that contains a series of slide decks accompanied by multiple-choice questions and interactive coding exercises to apply what you learned during the module: https://toolbox-learn.mds.ubc.ca/. We recommend using Chrome, Firefox or Safari to complete the content as other browsers are incompatible with the content platform.

Laptop requirements

Students will require a laptop meeting the following specifications (details on how to set up your laptop will follow in assignment 1):

- Runs one of the following operating systems: Ubuntu 22.04 (20.04 will also likely work), macOS Big Sur or above (>= 11.4.x), Windows 10 or 11 Home, Professional, Enterprise or Education (version 2004, 20H2, 21H1, or above).
  - When installing Ubuntu, checking the box "Install third party. . . " will (among other things) install proprietary drivers, which can be helpful for wifi and graphics cards.
- Can connect to networks via a wireless connection for on campus work
- Has access to an internet connection that is fast and stable enough for video calling and conducting online quizzes

- Has at least 50 GB disk space available
- Has at least 8 GB of RAM
- Uses a 64-bit CPU
- Is at most 6 years old at the start of the course (4 years old or newer is recommended)
- Uses English as the default language
- Student user has full administrative access to the computer

Learning Outcomes

By the end of the course, students will be able to:

1. Explain why it is important to use the right tool for the job.
2. Communicate with the computer via a text-based shell.
3. Work with projects under version control using Git and GitHub.
4. Use additional version control features such as git history, reverting repositories to a previous state, and resolving with merge conflicts.
5. Collaborate with Git and GitHub by using branches, forks and pull requests.
6. Organize coding projects with proper file names, folder structures, and virtual environments.
7. Work more effectively with JupyterLab.
8. Build books and create websites using Jupyter Book.

Course Facilitator

Position	Name	Email	Office Hours	Office Location
Course Facilitator	Analise Hofmann	analise.hofmann@ubc.ca	Mon/Wed 6:00-7:00 pm PST	Zoom
Course Facilitator	Matt Connell	matthew.connell@ubc.ca	Tue/Thr 6:00-7:00 pm PST	Zoom

Course Structure

- There are 8 modules in total for this course. Each module will be accompanied by an autograded multiple choice assignment to be submitted on Canvas.
- Students will have one attempt with unlimited time to submit this assignment and are expected to walk through the scenarios in the assignment questions on their own machines before answering.
- The students will be responsible to complete the material of 1 module every week
- In weeks 4 and 8, students will need to take an online 45-minute open-book quiz testing, material from Modules 1-4 and 5-8 respectively. The quizzes can be taken at any point during the week.
- At the end of the 8th module, students will complete a more comprehensive assignment where they are required to use the skills they learned throughout the course to produce an online site with technical material using the Jupyter Book platform and GitHub Pages.

Course Completion and Grading

You are assessed on each course according to UBC assessment guidelines to ensure the successful acquisition of required skills and concepts.

The Data Science Toolbox is part of the Key Capabilities in Data Science program. You must obtain 70% or higher in Data Visualization and complete the final assignment (assignment 8) to obtain a course completion

grade. To successfully complete the Key Capabilities in Data Science program, you must complete three of the four courses and achieve 70% or higher for each of the three courses.

Assignments

Each week students will submit 1 assignment for a grade. **Assignment due dates are posted on Canvas.** To submit your assignment click submit at the end of each assignment quiz. You will only be able to submit once so make sure to take notes of your answers as you go along so you can enter them when you return to complete your assignment.

Course breakdown

Deliverable	Percent Grade
7 Assignments	56 (8% each)
2 Quizzes	32 (16% each)
Final Assessment (Assignment 8)	12 %

Schedule

Module	Title	Description
1	Introduction to the Data Science Toolbox	Intro to the tools that we will be using in this course, as well as to computing in general
2	The shell	How to use the shell to navigate your filesystem and to execute commands.
3	Git and GitHub intro	The basics of version control with Git and GitHub.
4	Getting groovy with Git and GitHub	View your git history, travel back in time, deal with merge conflicts and other useful tools.
5	Branches, forks, and streams. . . Welcome to the Git nature walk!	Discover how to efficiently collaborate with Git and GitHub by using branches, forks ad pull requests.
6	File Names, Project Organization, Virtual Environments	An overview of how to effectively manage files, projects, and virtual environments.
7	JupyterLab	Get more in depth-knowledge about JupyterLab
8	Jupyter Book	Learn how to create beautiful, publication-ready books and websites using Jupyter Book.

Submission Schedule

Assignments are due on Sunday at 6 pm. Office hours will be held twice a week to support and answer questions regarding the concepts learned in the lecture. Quizzes will be opened for up to 7 days, please check Canvas for quiz submission dates.

Week	Modules to Complete	Submissions Due
Weeks 1 - 2	Module 1, Module 2,	Assignment 1, Assignment 2,
Weeks 3 - 4	Module 3, Module 4	Assignment 3, Assignment 4, Quiz 1
Weeks 5 - 6	Module 5, Module 6,	Assignment 5, Assignment 6
Week 7	Module 7	Assignment 7
Weeks 8 - 9	Module 8	Final Assessment (Assignment 8), Quiz 2

Policies

Late Assignments

Students should submit all assignments by the due date. Assignment 1 must be completed by it's due date to continue in the course. If for any reason a student is unable to submit Assignments 2 to 7 on time, they can submit the assignments before the course end date. This is the final allowed submission date and is non-negotiable.

Late Quizzes

We allow 2 weeks to complete the first quiz and 1 week to complete the second. if you've submitted a quiz late or did not submit a quiz at all, please contact the course facilitator and we will try to make accommodations.

Re-grading

If you have concerns about the way your work was graded, please contact the course facilitator.

Device/Browser

Students are responsible for using a device and browser compatible with all functionality of Canvas. Chrome or Firefox browsers are recommended; Safari has had issues with Canvas quizzes in the past.

Academic Integrity

The academic enterprise is founded on honesty, civility, and integrity. As members of this enterprise, all students are expected to know, understand, and follow the codes of conduct regarding academic integrity. At the most basic level, this means submitting only original work done by you and acknowledging all sources of information or ideas and attributing them to others as required. This also means you should not cheat, copy, or mislead others about what is your work. Violations of academic integrity (i.e., misconduct) lead to the breakdown of the academic enterprise, and therefore serious consequences arise and harsh sanctions are imposed. For example, incidences of plagiarism or cheating may result in a mark of zero on the assignment or exam and more serious consequences may apply if the matter is referred to the President's Advisory Committee on Student Discipline. Careful records are kept in order to monitor and prevent recurrences.

A more detailed description of academic integrity, including the University's policies and procedures, may be found in the Academic Calendar at http://calendar.ubc.ca/vancouver/index.cfm?tree=3,54,111,0.

Plagiarism

Students must correctly cite any code or text that has been authored by someone else or by the student themselves for other assignments. Cases of plagiarism may include, but are not limited to:

- the reproduction (copying and pasting) of code or text with none or minimal reformatting (e.g., changing the name of the variables)
- the translation of an algorithm or a script from a language to another
- the generation of code by automatic code-generation software

An "adequate acknowledgment" requires a detailed identification of the (parts of the) code or text reused and a full citation of the original source code that has been reused.

The above attribution policy applies only to assignments. **No code or text may be copied (with or without attribution) from any source during a quiz or exam. Answers must always be in your own words. At a minimum, copying will result in a grade of 0 for the related question.**

Repeated plagiarism of any form could result in larger penalties

Attribution

Parts of this syllabus (particularly the policies) have been copied and derived from the UBC MDS Policies as well as the Syllabus from UBC's DSCI 100 course.