WILFRID LAURIER UNIVERSITY

LAURIER *Inspiring Lives.*

Course Syllabus

GG606 Scientific Data Wrangling
Geography and Environmental Studies, Faculty of Science, Waterloo Campus
Winter 2022

I acknowledge that in Kitchener, Waterloo, Cambridge and Brantford we are on the traditional territory of the Neutral, Anishnawbe, and Haudenosaunee peoples.

The course schedule, evaluation scheme, methodology, assessments, lab and tutorial attendance, testing and final exam policies have been planned based on current public health guidelines. Should these guidelines change, any adjustments will be communicated to students.

Instructor Information

Colin Robertson | Arts 3C10B†
Email: crobertson@wlu.ca
Weekly Office Hours† (after class) or online By Appointment

Course Information

This course covers the data science skills comprising data visualization, data wrangling (cleaning, combining, modelling, etc.), and methodological and statistical design, which are an important part of the scientific method.

Peters Building P331, Thursdays 2:30 – 5:20 pm

Course Overview and Approach

This is a skills-based course that aims to get students up and running with data processing that can support their own independent quantitative research. Lectures will primarily be collaborative and example-based, examining code, data, and tools in an interactive, discussion-based format. Students will be expected to review reading material before each class, and be prepared to work through exercises and discuss readings during class time. The philosophy underpinning this approach to learning is that learning-by-doing is much more effective than lecturing when it comes to working with data. The purpose of this course is applied and practical skill development, as such we focus on this by working with datasets directly during classes.

Course Goals and Learning Outcomes

This course will develop skills that allow students to work confidently with data processing and modelling with a focus on reproducible and modern data workflows. Students will gain hands-on experience with open-source statistical software and real-world environmental datasets.

By the end of this course students should be able to:

- describe the characteristics of datasets to plan for data wrangling and visualisation

- develop workflows for dealing with disparate data types
- apply knowledge to tidy, transform, visualize, model datasets similar to thesis/project data
- develop a template for a reproducible workflow including metadata

Course Materials

Required Text

- Wickham H, Grolemund G. 2017. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media. Chicago, available online at http://r4ds.had.co.nz/
- Timbers, T-A., Campbell, T., and Lee, M. 2021. Introduction to Data Science https://ubc-dsci.github.io/introduction-to-datascience/

Additional Readings TBA

Course Software

- This course will use R, RStudio, and the tidyverse packages. Other software will also be discussed.
- Students need to have a computer with R and RStudio Desktop installed.
- Windows users:
  - Download and install R from CRAN https://cran.r-project.org/bin/windows/base/release.htm
  - Download and install RStudio Desktop from RStudio https://www.rstudio.com/products/rstudio/download/
  - Open RStudio to check that there are no error messages
- Mac OS X users:
  - Go to CRAN https://cran.r-project.org/
  - Click "Download R for (Mac) OS X"
  - Download and install the appropriate pkg file for your version of OS X
  - Download and install RStudio Desktop from RStudio https://www.rstudio.com/products/rstudio/download/
  - Open RStudio to check that there are no error messages
- Linux users:
  - R is available through most Linux package managers. You can download the binary files for your distribution from CRAN https://cran.r-project.org/. Or you can use your package manager (e.g. for Debian/Ubuntu run sudo apt-get install r-base and for Fedora run sudo yum install R).
  - Download and install RStudio Desktop from RStudio for your distrobution https://www.rstudio.com/products/rstudio/download/
  - Open RStudio to check that there are no error messages
- The tidyverse packages https://www.tidyverse.org/packages/ can be installed from inside RStudio by running install.packages("tidyverse") in the R Console or from the Packages tab in the lower right quarter of RStudio.

Student Evaluation

Assessment	Weighting	Due Date
Assignments (2x15%)	30%	January 27; March 03
Analytics Studio Demo	40%	March 31; April 07
Participation	15%	N/A
Course notebook	15%	April 07
Total	**100%**	

Learning Activities/Assignments

Assignments: Two assignments will require students to demonstrate data wrangling skills learned in the course. This will include a code and written component.

Analytics Studio Demo: Students will work on a major project which will require a technical overview of an analysis of their choosing. This will be presented in the last weeks of the course as a technical demonstration in class.

Participation: Students will be expected to attend and participate in designated course times. Participate includes contributing to discussions and working collaboratively with other students when needed.

Course Notebook: Students will keep a notebook to keep track of their learning throughout the course. These can be as digital or hardcopy, and will be graded at the end of the course for dept and quality of notes, and coverage of topics covered in the course. This notebook will be returned to students after grading.

Weekly Schedule(s) (subject to change)

Week # - Date	Topic	Tools	Delivery Mode
1-Jan 6	Data/Science workflows	R	Remote
2-Jan 13	I/O	R (base, tidyverse)	Remote
3-Jan 20	Data objects	R (base, tidyverse)	Remote
4-Jan 27	Databases	R (tidyverse, dbplyr), SQLite	Remote
5-Feb 03	Spatial Data	R (rgdal, sp, sf, raster, terra)	Remote
6-Feb 10	Spatial Data	R (spatstat, gstat)	Remote
-Feb 17	Reading Week		No Class
7-Feb 24	Work Period / Catch up		Remote Office Hours / Help Session
8-Mar 03	Temporal Data	R (ts, zoo, lubridate)	TBD
9-Mar 10	Reproducible Research	Git, Github	TBD
10-Mar 17	Data Sharing and Collaboration	R Markdown, Docker	TBD
11-Mar 24	Project Work Period		TBD Office Hours / Help Session
12-Mar 31	R in Production	Docker, R-Shiny	TBD
13-Apr 07	Project Presentations		TBD

University and Course Policies

1. **Academic Calendars:** Students are encouraged to review the Academic Calendar for information regarding all important dates, deadlines, and services available on campus.

2. **Intellectual Property:** The educational materials developed for this course, including, but not limited to, lecture notes and slides, handout materials, examinations and assignments, and any materials posted to MyLearningSpace, are the intellectual property of the course instructors. These materials have been developed for student use only and they are not intended for wider dissemination and/or communication outside of a given course. Posting or providing unauthorized audio, video, or textual material of course content to third-party websites violates instructors' intellectual property rights, and the Canadian Copyright Act. Recording lectures in any way is prohibited in this course unless specific permission has been granted by instructors. Failure to follow these instructions may be in contravention of the university's Student Non-Academic Code of Conduct and/or Code of Academic Conduct, and will result in appropriate penalties. Participation in this course constitutes an agreement by all parties to abide by the relevant University Policies, and to respect the intellectual property of others during and after their association with Wilfrid Laurier University.

3. **Accessibility:** Students requiring accommodation are advised to contact Laurier's Accessible Learning Centre for information regarding its services and resources.

4. **Plagiarism:** Wilfrid Laurier University uses software that can check for plagiarism. If requested to do so by course instructors, students are required to submit their written work in electronic form and have it checked for plagiarism. (Approved by Senate May 14, 2002).

5. **Academic Integrity:** Laurier is committed to a culture of integrity within and beyond the classroom. This culture values trustworthiness (e.g., honesty, integrity, reliability), fairness, caring, respect, responsibility and citizenship. Together, we have a shared responsibility to uphold this culture in our academic and nonacademic behaviour. The University has a defined policy with respect to academic misconduct. As a Laurier student you are responsible for familiarizing yourself with this policy and the accompanying penalty guidelines, some of which may appear on your transcript if there is a finding of misconduct. The relevant policy can be found at Laurier's academic integrity website along with resources to educate and support you in upholding a culture of integrity. Ignorance is not a defense.

6. **Use of Zoom for Instructional Purposes**: Wilfrid Laurier University uses a range of technologies to facilitate in-person and remote instruction. Zoom is currently used for remote course delivery, including lectures, seminars, and group office hours, which may be recorded, stored and shared through MyLearningSpace for access by students in the course. For these course activities, students are permitted to turn off their cameras or use an alternative name to maintain their privacy after they have confirmed this with their course instructors. Student personal information is collected and used in the course in accordance with University policies and the Notice of Collection, Use or Disclosure of Personal Information. All exams and mid-terms in the course that are conducted online will be proctored using only technologies approved for assessment at Laurier as outlined on this page.

7. **Late Assignment Policy:** Late assignments will be penalized at 5% per day for a period up to 1 week.

8. **Foot Patrol, the Wellness Centre, and the Student Food Bank:** The University approved the inclusion of information about select wellness and safety services and supports on campus in the course information provided to students. (Approved by Senate November 28, 2011.) Specific language (by campus) is provided below.

Multi-campus Resource:

- Good2Talk is a postsecondary school helpline that provides free, professional and confidential counselling support for students in Ontario. Call 1-866-925-5454 or through 2-1-1. Available 24-7.
-

Kitchener/Waterloo Resources:

- Waterloo Student Food Bank: All students are eligible to use this service to ensure they're eating healthy when overwhelmed, stressed or financially strained. Anonymously request a package online 24-7. All dietary restrictions accommodated.

- Waterloo Foot Patrol: 519.886.FOOT (3668). A volunteer operated safe-walk program, available Fall and Winter daily from 6:30 pm to 3 am. Teams of two are assigned to escort students to and from campus by foot or by van.

- Waterloo Student Wellness Centre: 519-884-0710, x3146. The Centre supports the physical, emotional, and mental health needs of students. Located on the 2nd floor of the Student Services Building, booked and same-day appointments are available Mondays and Wednesdays from 8:30 am to 7:30 pm, and Tuesdays, Thursdays and Fridays from 8:30 am to 4:15 pm. Contact the Centre at x3146, wellness@wlu.ca or @LaurierWellness. After hours crisis support available 24/7. Call 1-844-437-3247 (HERE247).

Brantford Resources:

- Brantford Student Food Bank: All students are eligible to use this service to ensure they're eating healthy when overwhelmed, stressed or financially strained. Anonymously request a package online 24-7. All dietary restrictions accommodated.

- Brantford Foot Patrol: 519-751-PTRL (7875). A volunteer operated safe-walk program, available Fall and Winter, Monday through Thursday from 6:30 pm to 1 am; Friday through Sunday 6:30 pm to 11 pm. Teams of two are assigned to escort students to and from campus by foot or by van.

- Brantford Wellness Centre: 519-756-8228, x5803. Students have access to support for all their physical, emotional, and mental health needs at the Wellness Centre. Location: Student Centre, 2nd floor. Hours: 8:30 am to 4:15 pm Monday through Friday. After hours crisis support available 24/7. Call 1-884-437-3247 (HERE247).