

CHL7001H: Statistical Programming and Computation for Health Data

Dalla Lana School of Public Health

Fall 2023

1 Details

Contact Information

Instructor

NAME: Aya Mitani
EMAIL: aya.mitani@utoronto.ca
OFFICE: Health Sciences Building 628
OFFICE HOURS: TBD

Teaching Assistant

NAME: TBD
EMAIL: TBD
OFFICE: TBD
OFFICE HOURS: TBD

Course Meeting Information

DATE AND TIME: Monday 10:00am – 1:00pm
LOCATION: [SS 1087](#)
WEBSITE: <https://q.utoronto.ca/courses/317202>

Acknowledgement of Territory

We wish to acknowledge this land on which the University of Toronto operates. For thousands of years, it has been the traditional land of the Huron-Wendat, the Seneca, and the Mississaugas of the Credit. Today, this meeting place is still home to many Indigenous people from across Turtle Island and we are grateful to have the opportunity to work on this land.

Pre-requisites

Open to MSc and PhD Biostatistics students. Others with permission from the instructor. Students are expected to have some experience with a statistical programming language, preferably R, and a basic understanding of linear regression and logistic regression.

Course Description

This is a pilot course that covers essential R programming and computational tools in health sciences research. Topics include data manipulation, data visualization, loops and functions, optimization, and package development with an emphasis on reproducibility and replicability.

Learning Outcomes

Students who complete this course will be able to:

- Develop a reproducible workflow while integrating version control for conducting research in health data sciences.
- Use R to prepare an analytical data set and perform descriptive analysis.
- Produce a visual (graphical or tabular) display of the data that effectively communicates the trend or pattern based on the research question of interest.
- Write efficient reproducible code throughout the data science project.
- Write an efficient simulation program.
- Develop an R package and deploy it on GitHub.
- Create reproducible professional grade documents (reports, articles, blog posts, presentation slides) using R.
- Use computational tools (bootstrap, multiple imputation, etc.) to aid in statistical analyses.

Computing

We will use R for all statistical programming and computation. R is a free open-source statistical software available for Macs, Windows, and Linux machines. R can be downloaded free from <https://www.r-project.org/>. Other software you need for the course are:

- Quarto, which is an open-source scientific and technical publishing system. Download for free from <https://quarto.org/>.
- RStudio, which is an integrated development environment for R. It can be downloaded for free from <https://rstudio.com/products/rstudio/>. See the course website for more information, including links to tutorials.
- Git, which is a free and open source distributed version control system. It can be downloaded from <https://git-scm.com/>.
- Download the Rtools package from <https://cran.r-project.org/bin/windows/Rtools/>

Finally, create a GitHub account (<https://github.com/>) If this is your first time creating a GitHub account, here are a few tips on choosing your username, <https://happygitwithr.com/github-acct.html>

Notes and Texts

Electronic copies of lecture notes and assignments, as well as code and data sets, will be posted on the course website. Although there are no required texts, the following are very good references for the material that will be covered in class:

1. Peng, R. D. *R Programming for Data Science* <https://bookdown.org/rdpeng/rprogdatascience>.
2. Wickham, H. & Golemund, G. *R for Data Science* <https://r4ds.had.co.nz/>.
3. Alexander, R. *Telling Stories with Data* <https://tellingstorieswithdata.com/>.
4. Baumer, B. S., Kaplan, D. T. & Horton, N. J. *Modern Data Science with R* <https://mdsr-book.github.io/mdsr2e/>.

Absences

If you have to miss class

If you have to miss a class, then try to complete the in-class activities in your own time. If you need to talk to me, please schedule a meeting with me. I would be happy to fill in the details of the missed lecture(s), within reason (i.e., I will not be your personal tutor every week). Due to the structure of this course that employs active learning, attendance is part of the evaluation. However, you are allowed to skip two classes without penalty.

If I have to miss class

If I need to stay home due to my own or a family member's illness, then I will move the class to Zoom. I have two young children, and unfortunately they get sick often. I will do my best to run this course smoothly but please understand if we experience any disruptions.

2 Evaluation

- **Attendance and participation** (10% of final grade): This course uses a hands-on active learning style, which makes attendance and participation crucial. Class time will be dedicated to individual and group work. You are allowed to skip two classes without penalty. Your in-class contributions will be assessed weekly on a scale of 0 to 2, given in Appendix A.
- **Pre-class assignments and quizzes** (40% of final grade): Each week (except for Week 1), there will be a pre-class assignment, either in the form of reading or video watching. There will be a total of ten short online quizzes, each due before class time. Each quiz will be based on the pre-class assignment. The pre-class assignments and quizzes will be posted a week before they are due. You will have one attempt to complete the quiz in one hour at any time during the week. You can choose to skip two quizzes without penalty. If you choose to complete all ten quizzes, then the two lowest marked quizzes will be excluded from calculating the final grade.
- **Final project** (30% of final grade): You will develop an R package for the final project. In late October, you will give me a short proposal on the R package you plan to develop. I will provide some feedback. The proposal is not part of the final grade, but is required. The evaluation criteria for the project are given in Appendix B.
- **Final presentation** (20% of final grade): You will give a short tutorial on how to use your R package that you developed for your final project. The tutorial can be in any format; you can give a live demonstration, create slides, go through a vignette document, etc. The evaluation criteria for the presentation are given in Appendix C.

3 Course Schedule

Week	Date	Topic	Assignment due
1	Sep 11	Introduction - R, RStudio, Quarto, GitHub and version control	
2	Sep 18	Understanding data – observations and variables	Quiz 1
3	Sep 25	Preparing data for analysis – functions and loops	Quiz 2
4	Oct 2	Exploratory analysis and data visualization	Quiz 3
5	Oct 9	Fall break – no class	
6	Oct 16	Professional quality tables and figures	Quiz 4
7	Oct 23	Developing an R package	Quiz 5
9	Oct 30	Numerical methods and optimization	Quiz 6
10	Nov 6	Bootstrap standard errors and confidence intervals	Quiz 7
11	Nov 13	Multiple imputation for handling missing data	Quiz 8
8	Nov 20	Conducting a simulation study	Quiz 9
12	Nov 27	Creating a website	Quiz 10
13	Dec 4	Final presentations I	
14	Dec 11	Final presentations II	Final project due on Dec 15

*The schedule is subject to change.

4 Policies

Below are some of my and the university's teaching policies. Please read them at least once and refer to them when you need to. Feel free to ask me any questions.

Academic integrity

Academic integrity is essential to the pursuit of learning and scholarship in a university, and to ensure that a degree from the University of Toronto is a strong signal of each student's individual academic achievement. As a result, the University treats cases of cheating and plagiarism very seriously. The University of Toronto's Code of Behaviour on Academic Matters outlines the behaviours that constitute academic dishonesty and the processes for addressing academic offences: (<http://www.governingcouncil.utoronto.ca/Assets/Governing+Council+Digital+Assets/Policies/PDF/ppjun011995.pdf>)

University of Toronto's policy regarding plagiarism: <http://www.writing.utoronto.ca/advice/using-sources/how-not-to-plagiarize>

Potential offences include, but are not limited to:

In papers and assignments:

- Using someone else's ideas or words without appropriate acknowledgement.
- Submitting your own work in more than one course without the permission of the instructor.
- Making up sources or facts.
- Obtaining or providing unauthorized assistance on any assignment.

On tests and exams:

- Using or possessing unauthorized aids.
- Looking at someone else's answers during an exam or test.
- Misrepresenting your identity.

Normally, students will be required to submit their course essays to [Turnitin.com](http://www.turnitin.com) for a review of textual similarity and detection of possible plagiarism. In doing so, students will allow their essays to be included as source documents in the [Turnitin.com](http://www.turnitin.com) reference database, where they will be used solely for the purpose of detecting plagiarism. The terms that apply to the University's use of the [Turnitin.com](http://www.turnitin.com) service are described on the [Turnitin.com](http://www.turnitin.com) web site

Accessibility

Students with diverse learning styles and needs are welcome in this course. In particular, if you have a disability or health considerations that may require accommodations, please feel free to approach me and/or the Accessibility Services Office as soon as possible. The Accessibility Services staff are available by appointment to assess specific needs, provide referrals, and arrange appropriate accommodations. The sooner you let them and me know your needs, the quicker we can assist you in achieving your learning goals in this course. For more information, or to register with Accessibility Services, please visit: <http://studentlife.utoronto.ca/as>.

Student Mental Health Resources

- U of T [Central Hub for Student Mental Health Resources](http://www.utoronto.ca/centralhub)
- [SGS Graduate Wellness Services](http://www.sgs.utoronto.ca/wellness)
- [Student Life Health and Wellness](http://www.studentlife.utoronto.ca/health)
- MySSP: 1-844-451-9700 (or use the app)
- Good2Talk: Call: 1-866-925-5454 or Text: GOODTOTALKON to 686868
- 8th Floor Lounge on Quercus

Key Resources and Supports for DSLPH Graduate Students

- U of T [Graduate Student Union](#)
- Biostatistics Union of Graduate Students ([BUGS](#))
- [DLSPH Covid Information](#)
- [DLSPH Student Resources \(Policies, Financial Aid, Health and Wellness, etc.\)](#)
- [DLSPH Student Handbook](#)

Respect for classmates

The University of Toronto is committed to equity, human rights, and respect for diversity. All members of the learning environment in this course should strive to create an atmosphere of mutual respect where all members of our community can express themselves, engage with each other, and respect one another's differences. U of T does not condone discrimination or harassment against any persons or communities.

Collaboration and external resources

You may use external resources such as textbooks, journal articles, or the internet to complete the quizzes, but you may not collaborate or consult with your classmates. You may collaborate and consult with your classmates to complete the final project and prepare for the final presentation. It may be a good idea to have your classmates test out the package before the due date.

Assignment submission

Please submit each assignment (quizzes and final project) on time. You are responsible for ensuring that I receive each assignment before it is due. If you need a deadline extension for the quizzes, let me know as soon as possible, but before the deadline. In general, late quizzes will not be accepted without prior approval from me. For the final project, 5% will be deducted for each day it is late, and it will not be accepted after four days from the due date.

Children in the classroom

Dependents, including children of any age (bottle-feeding, nursing, etc) are welcome in class as often as necessary. Please feel free to take breaks to feed or express milk for your child as needed, either in the classroom or elsewhere: <https://familycare.utoronto.ca/childcare/breastfeeding-at-u-of-t/>. Please let me know your needs so that I can make proper accommodations. There may be times when I need to bring *my* children to class too.

Talk to me!

I want for you to learn the material in a fun, collaborative, and stimulating environment. Please provide feedback – positive or negative – at any time during the semester. If you are having trouble learning, come talk to me before it is too late. If you enjoyed something we did in class, tell me so that we can do it again. This is the first time I'm teaching this course in-person. Your feedback will also benefit future students taking this course. The feedback you provide will *not* affect your grade in any way.

Appendix A: Weekly assessment for attendance and participation

- 0 = Absent
- 1 = Present but not respectfully participating in class discussion or individual/group activities
- 2 = Present and respectfully participating in class discussion and individual/group activities

Appendix B: Evaluation criteria for the final project

Item	Details	Score per item (1-10)
Accuracy	Package does what it intends to do, correct methods, user-friendly, clear outputs and results with correct labels	
Code	Free of bugs, organized, separate scripts when necessary, good documentation within code, efficient use of functions and loops	
Originality and creativity	Name of package, name of function, uniqueness of idea	
Documentation	Describes each component of package, each function has a help file, clear and concise, provides examples, free of typos and grammar errors	
Repository	User can download package without error, describes package, has simple instructions, organized	
Total		/50

Appendix C: Evaluation criteria for the final presentation

Item	Details	Score per item (1-10)
Delivery	Speed, eye contact, audibility, tone, composure, posture, facial expressions, completes within allotted time frame	
Content and structure	Objectives and goals clearly stated, organization, knowledge of material	
Materials	Easy to read, tables and figures are clearly described and labeled, free of typos and grammar errors	
Response to questions	Understands question that was asked, answers questions appropriately	
Total		/40