Syllabus: EEB313 Quantitative Methods in R for Biology [24L, 12P]

This course covers statistics and data analysis for ecology and reproducible quantitative methods in R. Statistical analysis, modelling, simulation, and data analysis are essential skills for applying ecology concepts to data. This course is designed to meet a growing demand for reproducible, openly accessible, analytically thorough, and well documented science. Students will learn to develop ecological population models, analyze data, and document their research using the R programming language. No prerequisite programming experience is required.

Prerequisites: BIO220H1 and one of EEB225H1, STA288H1, or STA220H1

## Time

Tue and Thu 2:10 - 4:00 pm. Office hours are Tue 4:00 - 5:00 pm.

## Class locations

|  |  |
| --- | --- |
| Day | Room |
| Tue | [Ramsay Wright](http://map.utoronto.ca/utsg/building/072) (RW 109) |
| Thu | [Ramsay Wright](http://map.utoronto.ca/utsg/building/072) (RW 109) |

Office hours are in RW 107, right next to RW 109.

The lecture hall has access to individual computers for the students. To use the computer workstations, students can login with their UTORid and password. Programs and packages that you install, and files that you save, will be deleted from these computers daily. Please bring a USB key to save files onto or email them to yourself. Students can use any of the lecture halls when there are no classes scheduled. Lecture halls are usually open 9 am - 5 pm, see the [online schedules](http://lab.chass.utoronto.ca/carr.php) for available times.

## Contact info

Quercus is the preferred communication channel. If you need to use email instead, please address all general course-related issues to [ahmed.hasan@mail.utoronto.ca](mailto:ahmed.hasan@mail.utoronto.ca), and project specific communication to the respective TA of your group. Prefix the subject matter with “EEB313”. If you do not receive a reply within 48 hours (excluding week-ends), please send a reminder.

### Course Instructors

* James Santangelo, [james.santangelo@mail.utoronto.ca](mailto:james.santangelo@mail.utoronto.ca)
* Lindsay Coome, [lindsay.coome@mail.utoronto.ca](mailto:lindsay.coome@mail.utoronto.ca)
* Madeleine Bonsma-Fisher, [m.bonsma@mail.utoronto.ca](mailto:m.bonsma@mail.utoronto.ca)
* Ahmed Hasan, [ahmed.hasan@mail.utoronto.ca](mailto:ahmed.hasan@mail.utoronto.ca)

### Supervising professor

Prof. Asher Cutter, [asher.cutter@utoronto.ca](mailto:asher.cutter@utoronto.ca), 416-978-4602, ESC2052

## Course Website and Quercus

All course information is accessible [on its own website](https://uoftcoders.github.io/rcourse/) and on [Quercus](https://q.utoronto.ca), including the syllabus, assessments, and lecture slides. If you have any problem accessing the material, let us know via email right away so we can fix the problem.

## Recommended resources

* [R for Data science](http://r4ds.had.co.nz/), H Wickham, G Grolemund, 2017
  + Excellent open access resource for R.
* [RStudio cheat sheets](https://www.rstudio.com/resources/cheatsheets/), RStudio, 2017
  + As good as it sounds, great quick reference.
* [R for ecological data science](https://datacarpentry.org/R-ecology-lesson/index.html)
  + An inspiration for our lectures.

## Course learning outcomes

1. Develop proficiency in the programming language R.
2. Use R to apply statistics to analyze and interpret data.
3. Choose appropriate analysis techniques for a variety of data types and formats.
4. Learn and use techniques and best practices for reproducible, high-quality science.
5. Learn how to work as part of a research team to produce a scientific product.
6. Learn what is required to generate a scientific item ready for publishing.

## Improving your writing skills

Effective communication is crucial in science. The [University of Toronto provides services](http://writing.utoronto.ca/) to help you improve your writing, from general advices on effective writing to writing centers and writing courses. The Faculty of Arts & Science also offers an English Language Learning (ELL) program, which provides free individualized instruction in English skills. Take advantage of these!

## Academic integrity

You should be aware of the University of Toronto Code of Behaviour on Academic Matters. Also see [How Not to Plagiarize](http://advice.writing.utoronto.ca/using-sources/how-not-to-plagiarize/). Note that it is NOT appropriate to use large sections from internet sources, and inserting a few words here and there does not make it an original piece of writing. Be careful in using internet sources – there is no review of most online material and there are many errors out there. Use only academic or government internet sources when absolutely necessary. Make sure you read material from many sources (published, peer-reviewed, trusted internet sources) and that you write an original text using this information. Always cite your sources. In case of doubt about plagiarism, talk to your instructor. Please make sure that what you submit for the final project does not overlap with what you submit for other classes, such as the 4th year research project. We will not enforce this, but the department will.

## Lecture schedule

|  |  |  |  |
| --- | --- | --- | --- |
| Week | Date | Topic | Instructor |
| 1 | Sep 06 | Intro to course, programming, RStudio, R Markdown | Everyone |
| 2 | Sep 11 | Assignment, vectors, functions | Ahmed |
| 2 | Sep 13 | Data frames, intro to dplyr | Ahmed |
| 3 | Sep 18 | Data wrangling in dplyr, ggplot, tidy data | Ahmed |
| 3 | Sep 20 | More dplyr and ggplot | Ahmed |
| 4 | Sep 25 | Exploratory data analysis | Lindsay |
| 4 | Sep 27 | Linear models and statistical modelling | Lindsay |
| 5 | Oct 02 | Mixed effects models | James |
| 5 | Oct 04 | Simulating data | James |
| 6 | Oct 09 | Multivariate stats | Lindsay |
| 6 | Oct 11 | Model selection | James |
| 7 | Oct 16 | Numerically solving population models | Madeleine |
| 7 | Oct 18 | Time series analysis | Madeleine |
| 8 | Oct 23 | Datasets, hypothesis, begin projects | Everyone |
| 8 | Oct 25 | Reproducible science | Lina or Sara |
| 9 | Oct 30 | Reproducible science | Lina or Sara |
| 9 | Nov 01 | Project work | 1 TA/group |
| - | Nov 06 | Fall break | - |
| - | Nov 08 | Fall break | - |
| 10 | Nov 13 | Project work | 1 TA/group |
| 10 | Nov 15 | Project work | 1 TA/group |
| 11 | Nov 20 | Project work | 1 TA/group |
| 11 | Nov 22 | Project work | 1 TA/group |
| 12 | Nov 27 | Project work | 1 TA/group |
| 12 | Nov 29 | Project work | Everyone |
| 13 | Dec 04 | Project work | Everyone |

## Assessment schedule

|  |  |  |  |
| --- | --- | --- | --- |
| Assignment | Type | Due date | Marks |
| Getting set up | Individual | Sep 18 | 4 |
| Basic R and dplyr | Individual | Sep 25 | 8 |
| dplyr and tidy data | Individual | Oct 02 | 8 |
| Statistical modelling | Individual | Oct 09 | 8 |
| Bootstrapping & multivariate stats | Individual | Oct 16 | 8 |
| Model selection and time series analysis | Individual | Oct 23 | 8 |
| Project setup | Project, Individual | Oct 30 | 8 |
| Data visualization | Project, Individual | Nov 13 | 8 |
| Final project work | Project, Group | Dec 05 | 32 |
| Weekly project updates | Individual | - | 8 |

There are 100 marks in total. Your final course mark will be the sum of your assignment scores, which will be translated to a letter grade according to the [official grading scale](http://www.artsci.utoronto.ca/faculty-staff/teacher-info/academic-handbook-for-instructors/sections-9-11#official) of the Faculty of Arts and Science.

Assignments will be distributed and submitted in the R Markdown format via Quercus. Assignments will be handed out on Thursdays and are due 11:59 pm on the Tuesday eight weekdays later. *There will be a penalty of 5% per day (including week-ends) for late submissions*.

### Final project grading rubric

|  |  |  |  |
| --- | --- | --- | --- |
|  | Inadequate (0 marks) | Adequate (4 marks) | Excellent (8 marks) |
| Contribution to group work | Student contributed little to project; self-assessed contributions are low in quality and/or quantity; self-assessment is not consistent with actual contribution. | Student contributed adequately to project; made some significant contributions | Student substantially contributed to project to ensure success; self-assessed contributions are crucial to project; self-assessment is consistent with actual contribution. |
| Content | Missing crucial information; methods and results are inconsistent, not logical, or not adequately explained; conclusions are confusing or unsupported by results; unnecessary information included as clutter | Most essential information included; methods and results are adequately described; conclusions supported by results; most included material is relevant to report | All essential information included; methods and results are succinct, clear, logical, and scientifically valid; conclusions are creative and meaningful; project is concise throughout |
| Style and reproducibility | Code and writing are poorly organized, poorly formatted, missing units, difficult to read, poorly documented, difficult to reproduce analyses | Code and writing are well-organized, well-formatted, consistent use of units and significant figures | Code and writing are precise and clear throughout, free of errors, well-organized, well-documented, easily reproducible analyses, publication-ready |
| Presentation | Presentation is poorly organized; much too long or much too short; presentation is unclear; presentation is missing information; presentation is not scientific and professional; presentation uses too much jargon; not all team members participate; does not adequately address audience questions | Presentation is adequately organized; timing is appropriate; most information is presented logically; presentation is scientific and professional; most jargon is avoided; all team members participate but equally; audience questions are sometimes addressed well | Presentation is clearly and logically organized; presentation flows and is easy to follow; presentation includes appropriate information without jargon; presentation is well-rehearsed and high-quality; all team members participate equally; audience questions are clearly addressed |

As the final project is a team effort, all members within a group will receive the same mark. A final project that is considered to lie between two of the defined levels will be marked accordingly, e.g. between “Adequate” and “Excellent” would be 5, 6, or 7 marks.