

Data Science Foundations

1 Preamble

1.1 Description

Quantitative approaches have a common concern: How can others be confident that our statistical approaches have been brought to bear on appropriate datasets? This course focuses on the ‘data’ of the data sciences. It develops in students an appreciation for the many ways in which dealing with a dataset can get out-of-hand, and establishes approaches to ensure data science is conducted in ways that engenders trusted findings. It focuses not only on statistical modelling, but also on everything that comes before modelling, and by focusing on those steps places modelling and analysis on a more firm foundation. In assessment, students will conduct end-to-end data science projects using real-world data, enabling them to fully understand potential pitfalls.

1.2 Overview

The purpose of this course is to develop students who appreciate, and can iterate on, the foundations of data science.

The focus of the learning will be on:

1. actively reading and consider relevant literature;
2. actively using the statistical programming language R in real-world conditions;
3. gathering, cleaning, and preparing datasets; and
4. choosing and implementing statistical models and evaluating their estimates.

Essentially this course provides students with everything that they need to know to be able to do the most exciting thing in the world: use data to tell convincing stories.

1.3 FAQ

- Can I audit this course? Sure, but it is pointless, because the only way to learn this stuff is to do the work.
- What is a tutorial? You write a paper. Then you send it to your tutor. The next day you have a meeting, ‘tutorial’, where you discuss it with them.
- Why is there so much assessment? The only way to learn is to actually do the work, and students only do the work when they are assessed. It is unfortunate, but there is no way around it.

1.4 Learning objectives

The purpose of the course is to develop the core skills of data science that are applicable across academia and industry. By the end of the course, you should be able to:

1. Engage critically with ideas and readings in data science (demonstrated in all papers but also tutorials and quizzes).
2. Conduct research in data science in a reproducible and ethical way (demonstrated in all papers).
3. Clearly communicate what was done, what was found, and why in writing (demonstrated in all papers).
4. Understand what constitutes ethical high-quality data science practice, especially reproducibility and respect for those that underpin our data (demonstrated in all papers and selected quizzes).
5. Respectfully identify strengths and weaknesses in the data science research conducted by others (demonstrated in quizzes, and the peer review).
6. Develop the ability to appropriately choose and apply statistical models to real-world situations (demonstrated in the final paper)
7. Conduct all aspects of the typical data science workflow (demonstrated in all papers).
8. Reflect effectively on your own learning and professional development (demonstrated in some tutorials and quizzes).

1.5 Relationship to PhD Student Learning Outcomes

- Read broadly across data science to understand the extent of knowledge
- Conduct original research
- Work in an independent way
- Communicate work and findings in written form
- Be especially aware of the limitations of the data, and methods, they are using.

1.6 Pre-requisites

- None.

1.7 Textbook

[Telling Stories with Data](#)

2 Content

Before class starts you should go through Chapters 1 and 2 of *Telling Stories with Data*.

2.1 Week 1

- [R Essentials](#)

2.2 Week 2

- [Reproducible workflows](#)

2.3 Week 3

- [Writing research](#)

2.4 Week 4

- [Static communication](#)

2.5 Week 5

- [Farm data](#)

2.6 Week 6

- [Gather data](#)

2.7 Week 7

- [Hunt data](#)

2.8 Week 8

- [Clean and prepare](#)

2.9 Week 9

- [Store and share](#)

2.10 Week 10

- [Exploratory data analysis](#)

2.11 Week 11

- [Linear models](#)

2.12 Week 12

- [Generalized linear models](#)

3 Assessment

3.1 Summary

Item	Weight (%)	Due date
Quiz	10	Weekly, end of each week
Tutorial	10	Weekly, end of each week
Paper 1	25	Friday, noon, Week 3
Paper 2	25	Friday, noon, Week 6

Item	Weight (%)	Due date
Paper 3	25	Friday, noon, Week 8
Paper 4	25	Friday, noon, Week 10
Final Paper (initial submission)	2	Wednesday, noon, Week 12
Conduct peer review	3	Friday, noon, Week 12
Final Paper	25	Two weeks after that

You must submit Paper 1. And you must submit the Final Paper.

Beyond that, you have scope to pick an assessment schedule that works for you. We will take your best three of the twelve tutorials for that 10 per cent, and your best five of twelve quizzes for that 10 per cent. And we will take your two best papers from Papers 1-4 for that 50 per cent (25 per cent for each). The remainder is made up of 2 per cent for submitting a draft of the Final Paper, 3 per cent for conducting peer review of other people's drafts of the Final Paper, and 25 per cent for the Final Paper.

Additional details:

- Quiz questions are drawn from those in the Quiz section that follows each chapter of *Telling Stories with Data*. Some of them are multiple choice, and you should expect to know the mark within a few days of submission.
- Tutorial questions are drawn from those in the Tutorial section that follows each chapter of *Telling Stories with Data*. The general expectation (although this differs from week to week) is about two pages of written content, which the tutor will read, discuss with you, and then provide a mark. You should expect to know the mark within a few days of the tutorial.
- In general papers require a considerable amount of work, and are due after the material has been covered in quizzes and tutorials (i.e. you would draw on knowledge tested in the quizzes, and potentially material could be re-used from the tutorial material). In general, they require original work to some extent. Papers are taken from the Papers appendix of *Telling Stories with Data* and students have access to the grading rubrics before submission.

3.2 Quiz

- Due date: Friday, noon, weekly (with grace period through to Sunday, midnight, to submit without penalty).

- Weight: 10 per cent. Only best five out of twelve count.
- Task: Please complete a weekly quiz.

3.3 Tutorial

- Due date: Friday, noon, weekly (with grace period through to Sunday, midnight, to submit without penalty).
- Weight: 10 per cent. Only best three out of twelve count.
- Task: Please complete a tutorial question.
- Rubric:
 - 0 - Any typos, major grammatical errors, other table stakes issues for this level. Too short.
 - 0.25 - Grammatical errors, if relevant: tables/graphs not properly labeled, no references, other aspects that affect credibility.
 - 0.6 - Makes some interesting and relevant points, related to course material (including required materials), but lacking in terms of structure and story/argument.
 - 0.80 - Interesting paper that is well-structured, coherent, and credible.
 - 1 - As with 0.80, but exceptional in some way.

3.4 Paper #1

- You must submit this paper.
- Task: [Paper One](#)
- Due date: Friday, noon, Week 3 (with grace period through to Sunday, midnight, to submit without penalty).
- Weight: 25 per cent (for Papers #1-#4 the best two of four count).

3.5 Paper #2

- Due date: Friday, noon, Week 6 (with grace period through to Sunday, midnight, to submit without penalty).
- Task: [Paper Two](#)
- Weight: 25 per cent (for Papers #1-#4 the best two of four counts).

3.6 Paper #3

- Due date: Friday, noon, Week 8 (with grace period through to Sunday, midnight, to submit without penalty).
- Task: [Paper Three](#)
- Weight: 25 per cent (for Papers #1-#4 the best two of four counts).

3.7 Paper #4

- Due date: Friday, noon, Week 10 (with grace period through to Sunday, midnight, to submit without penalty).
- Task: [Paper Four](#)
- Weight: 25 per cent (for Papers #1-#4 the best two of four counts).

3.8 Final Paper

- Task: [Final Paper](#)
- You must submit this paper.
- Due dates:
 - Initial submission: Wednesday, noon, Week 12 (no grace period and no late submissions accepted).
 - Conduct peer review: Friday, noon, Week 12 (no grace period and no late submissions accepted).
 - Final Paper: Two weeks after that (with grace period through to Sunday, midnight, to submit without penalty).
- Weight: 30 per cent
 - Initial submission: 2 per cent
 - Conduct peer review: 3 per cent
 - Final Paper: 25 per cent

4 Other

4.1 Children in the classroom

Babies (bottle-feeding, nursing, etc) are welcome in class as often as necessary. You are welcome to take breaks to feed your infant or express milk as needed, either in the classroom or elsewhere including: <https://familycare.utoronto.ca/childcare/breastfeeding-at-u-of-t/>. A list of baby change stations is also available: <https://familycare.utoronto.ca/childcare/baby-change-stations-at-u-of-t/>. Please communicate with me so that I can make sure that we have regular breaks to accommodate this. For older children, I understand that unexpected disruptions in childcare can happen. You are welcome to bring your child to class in order to cover unforeseeable gaps in childcare.

4.2 Accommodations with regard to assessment

Please do **not** reveal your personal or medical information to me. I understand that illness or personal emergencies can happen from time to time. The following accommodations to assessment requirements exist to provide for those situations.

Straight-forward (will automatically apply to all students - there's no need to ask for these):

- Quiz: Only best five quizzes count.
- Tutorial: Only best three tutorials count.
- Papers #1-#4: Worst two are dropped.

So for those (with the exception of Paper #1), if you have a situation, then just don't submit.

Slightly more involved:

- Paper #1: You must submit something for Paper #1, even if it gets zero. If you have a medical reason that makes it impossible for you to submit Paper #1, then you are welcome to continue with the class, but then one of the remaining term papers (Papers #2 - #4), must be done individually to ensure fairness with the rest of the class.
- Peer review: No accommodation or late submission is possible for this because it would hold up the rest of the class. If you cannot submit then email me before the deadline and the weight will be shifted to the final paper.
- Final paper: The final paper is a critical piece of assessment. It's also up against deadlines for submission of grades. Extensions for valid reasons may be granted for a maximum of three days, however this isn't possible for all students (i.e. there are restrictions around graduating students). Hence, the exact extension need to be at my discretion. To be considered, an extension request must be sent to rohan.alexander@utoronto.ca by the business day before the due date so there is time to get advice from the Faculty about your particular circumstance.

4.3 Re-grading

Requests to have your work re-graded will not be accepted within 24 hours of the release of grades. This is to give you a chance to reflect. Similarly, requests to have your work re-graded more than seven days after the release of the grades will not be accepted. This is to ensure the course runs smoothly.

Inside that 1-7 day period if you would like to request a re-grade, please email rohan.alexander@utoronto.ca. Please specify where the marking error was made in relation to the marking guide. The entire assessment will be re-marked and it is possible that your grade could reduce.

4.4 Plagiarism and integrity

Please do not plagiarize. In particular, be careful to acknowledge the source of code - if it's extensive then through proper citation and if it's just a couple of lines from Stack Overflow then in a comment immediately next to the code.

You are responsible for knowing the content of the University of Toronto's Code of Behaviour on Academic Matters.

Academic offenses includes (but is not limited to) plagiarism, cheating, copying R code, communication/extra resources during closed book assessments, purchasing labor for assessments (of any kind). Academic offenses will be taken seriously and dealt with accordingly. If you have any questions about what is or is not permitted in this course, please contact me.

Please consult the University's site on Academic Integrity <http://academicintegrity.utoronto.ca/>.

Please also see the definition of plagiarism in section B.I.1.(d) of the University's Code of Behaviour on Academic Matters <http://www.governingcouncil.utoronto.ca/Assets/Governing+Council+Digital+A>

Please read the Code. Please review Cite it Right and if you require further clarification, consult the site How Not to Plagiarize <http://advice.writing.utoronto.ca/wp-content/uploads/sites/2/how-not-to-plagiarize.pdf>.

4.5 Late policy

You are expected to manage your time effectively. If no extension has been granted and no accommodation applies, then the late submission of an assessment item carries a penalty of 10 percentage points per day to a maximum of one week after which it will no longer be accepted, e.g. a problem set submitted a day late that would have otherwise received 8/10 will receive 7/10, if that same problem set was submitted two days late then it would receive 6/10.

4.6 Writing

Papers and reports should be well-written, well-organized, and easy to follow. They should flow easily from one point to the next. They should have proper sentence structure, spelling, vocabulary, and grammar. Each point should be articulated clearly and completely without being overly verbose. Papers should demonstrate your understanding of the topics you are studying in the course and your confidence in using the terms, techniques and issues you have learned. As always, references must be properly included and cited. If you have concerns about your ability to do any of this then please make use of the writing support provided to the faculty, colleges and the SGS Graduate Centre for Academic Communication.

4.7 Minimum submission requirement

If you are going to not be able to submit at least two term papers, and/or be unable to submit the final paper then it would be unfair on the other students to allow you to pass the course. Please ensure you and your registrar get in touch with me as early as possible if this may be the case for you.