

Teaching Team

Office Hours

Class Times and Location

Course Description

Learning objectives

Course Websites

Course Assignment Calendar

Computing

Optional Course Textbooks

Evaluation

Course Policies




Your Responsibilities

JSC270H1S: Data Science I

Syllabus

Winter 2019

Teaching Team

-  Prof. N. Taback (<http://utstat.toronto.edu/nathan/>), SS6027C.
-  Prof. F. Chevalier (<http://fannychevalier.net>), BA5226.
-  Email instructors: jsc270-winter2019@cs.toronto.edu (<mailto:jsc270-winter2019@cs.toronto.edu>)
- Teaching assistants: Nicole Sultanum and Michael Moon.

Office Hours

- Prof. Taback, Tuesday 13:00 - 14:00 (SS6027C).
- Prof. Chevalier, Tuesday, 14:00 - 15:00 (BA5226)
- TA office hours, Friday, 14:00 - 16:00 (BA5287).

Class Times and Location

Lectures

- Mondays, 12:00-14:00 in WI523 (<http://map.utoronto.ca/utsg/building/032A>).
- There is no lecture on February 18 (Reading Week)

Tutorials

- Wednesdays, 12:00-14:00 in BA3175.
- There are no tutorials on the following Mondays:
 - February 20 (Reading Week)
 - March 13 (Mid-Term Test)

Course Description

Data Science is a relatively new interdisciplinary field that encompasses statistical methods, the computational aspects of carrying out a data analysis, including acquisition, management, and analysis of data, and the communication of analysis processes and extracted knowledge. Statistical reasoning, computing with data, and visualization play important roles in this emerging discipline. The purpose of this course is to give you a broad introduction to many of the ways data scientists learn from data, including statistical reasoning, computation and communication. We will use both Python and R programming languages. Tutorial labs will give students hands-on experience in executing and communicating data science problems and solutions.

Through this course, you will gain experience working on data science projects that involve using data from industry, science, or the humanities to help answer salient questions; interact with data scientists, researchers, or other professionals from academia or industry; learn how to translate data science skills across domains and think critically about data and models of data; develop strong oral and written communication skills and the ability to work in multidisciplinary team.

Learning objectives

By the end of this course, you should be able to:

- Apply and evaluate statistical methods and study designs to develop solutions to questions based on real data.

- Understand current software infrastructure for handling data.
- Write computer programs to wrangle and analyse data.
- Understand the fundamental data structures needed for storing and processing data efficiently.
- Understand ethical issues related to data analysis and software development.
- Identify and answer questions that involve applying statistical methods or machine learning algorithms to complex data, and communicating the results.
- Work in a team to solve data science problems.
- Present the results and limitations of a data analysis at appropriate technical levels for the intended audience.

□ Course Websites

- Class slides, notes, and other important information can be found on the course website. (<https://jsc270.github.io/>)
- Questions and Answers can be posted on the course discussion board (Piazza) (<https://piazza.com/class/jpdcniqs4r36rj#>)

Course Assignment Calendar

- Assignment #1: Monday, February 4
- Assignment #2: Monday, February 25
- Assignment #3: Monday, April 3
- Term test: Monday, March 11

□ Computing

The course will use Python 3 (<https://www.python.org/>), R (<https://www.r-project.org/>) and R markdown (<https://rmarkdown.rstudio.com/>) for computing.

Jupyter and RStudio are both available on <https://www.teach.cs.toronto.edu> (<https://www.teach.cs.toronto.edu>).

Click here for Jupyter (<https://jupyter.teach.cs.toronto.edu/hub/login>)

Click here for RStudio (<https://rstudio.teach.cs.toronto.edu/auth-sign-in>)

Optional Course Textbooks

References that use Python

Hajba, Gábor László. "Website Scraping with Python." Electronic copy is available from the UofT library (<http://go.utlib.ca/cat/11936372>)

Igual, Laura, and Santi Seguí. "Introduction to Data Science." Introduction to Data Science. Springer, Cham, 2017. Electronic copy is available from the UofT library (<http://go.utlib.ca/cat/11433860>)

McKinney, Wes. "Python for data analysis: data wrangling with Pandas, NumPy, and IPython." (2017). <http://wesmckinney.com/pages/book.html> (<http://wesmckinney.com/pages/book.html>)

Skiena, Steven S. The Data Science Design Manual. Springer, 2017. Electronic copy is available from the UofT library (<http://go.utlib.ca/cat/11634033>)

References that use R

Efron, Bradley, and Trevor Hastie. Computer age statistical inference. Vol. 5. Cambridge University Press, 2016. Available here. (<https://web.stanford.edu/~hastie/CASI/>)

James, Gareth, et al. An introduction to statistical learning. Vol. 112. New York: springer, 2013. Available here. (<http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf>)

Wickham, Hadley, and Garrett Grolemund. R for data science: import, tidy, transform, visualize, and model data. " O'Reilly Media, Inc.", 2016. Available here. (<https://r4ds.had.co.nz>)

Evaluation

Evaluation	Percentage of Final Grade
Assignment 1	10%
Assignment 2	15%
Assignment 3	35%
Term test	20%
Tutorial Assignments	10%
End of course survey/Reflection	5%
Class participation	5%

Course Policies

Missed Tests or Assignments

- If a test or tutorial work is missed for a valid medical reason, you must submit the University of Toronto Verification of Student Illness or Injury form (<http://www.illnessverification.utoronto.ca>) within one week. Submit the form to your TA if it is related to missed tutorial work. Submit the form to both of your instructors if it is related to the test.
- The form will only be accepted as valid if the form is filled out according to the instructions on the form.
- The form must indicate that the degree of incapacitation on academic functioning is moderate, serious, or severe in order to be considered a valid medical reason for missing the term test. If the form indicates that the degree of incapacitation on academic functioning is negligible or mild then this will **NOT** be considered a valid medical reason.
- If a tutorial is missed for a valid reason then the weight for that in-lab portion of the tutorial will be shifted to the tutorial assignment completion.
- If the term test is missed then an oral exam may be used as the makeup test.
- If a group presentation is missed then an individual presentation may be used as makeup presentation.
- The late penalties for submitting a completed tutorial assignment are: 10% if the assignment is submitted within 24 hours of the due date; 30% if the assignment is submitted within 48 hours of the due date; and 100% if the assignment is submitted after 48 hours.
- Other reasons for missing an assignment, tutorial work, or the term test will require prior approval by your instructors. If prior approval is not received for non-medical reasons then you will receive a grade of zero for the missed assignment, tutorial work, or test.

Marking Concerns

Any requests to have your work remarked must contain a written justification for consideration. Remarking requests should be made within one week of receiving your marked work. Re-evaluation appeals are at the discretion of the instructors. Note that adjustments in marks will be rare and could equally result in a lowering or raising of the mark. If a re-evaluation is completed by the instructors, the student must accept the resulting mark as the new mark, whether it goes up or down or remains the same. When appealing a re-evaluation decision, the student accepts this condition.

How to Communicate with Your Instructors

Always use your utoronto.ca e-mail account to ensure that your message doesn't automatically go to a Junk folder and include your full name and student number. Violation to this policy may incur severe response delays. Allow up to 72 hours for a reply.

Questions about course material or organization, such as,

- What do I change the colour of my plotting symbol?
- What is the difference between supervised and unsupervised learning?
- When is the term test?

should be posted on the discussion forums on Piazza (<https://piazza.com/class/jpdcniqs4r36rj#>) or asked in person. Questions can be posted anonymously (so that the author is anonymous to other students but not to the instructors), if desired.

If your communication is private, such as, "I missed the test because I was ill", then e-mail both instructors (jsc270-winter2019@cs.toronto.edu (<mailto:jsc270-winter2019@cs.toronto.edu>)). If you missed a tutorial then e-mail your TAs.

Academic Integrity

You are responsible for knowing the content of the University of Toronto's Code of Behaviour on Academic Matters (<http://www.governingcouncil.utoronto.ca/policies/behaveac.htm>).

As a general rule, we encourage you to discuss course material with each other and ask others for advice. However, it is not permitted to share complete solutions or to directly share code for anything that is to be handed in. When an assignment is required to be completed as a team, you may share solutions and code with other members of your team, but not with another team in the class. For example, "For question 2.1 what R function did you use?" is a fair question; "Please show me your R code for question 2.1" is not.

If you have any questions about what is or is not permitted in this course, please do not hesitate to contact your instructors (<mailto:jsc270-winter2019@cs.toronto.edu>).

Accessibility Needs

The University of Toronto is committed to accessibility. If you require accommodations for a disability, or have any accessibility concerns about the course, the class room, or course materials, please contact Accessibility Services as soon as possible: accessibility.services@utoronto.ca (<mailto:accessibility.services@utoronto.ca>) or <http://accessibility.utoronto.ca> (<http://accessibility.utoronto.ca>)

Your Responsibilities

The course is designed to actively engage you in the course material. We hope you'll find the methods of data science interesting, challenging, and fun. In order for classroom sessions and tutorials to be effective, prepare by learning about the week's concepts through completing the recommended problems and readings.



(<http://creativecommons.org/licenses/by-nc-sa/4.0/>)

This website is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (<http://creativecommons.org/licenses/by-nc-sa/4.0/>).

R Markdown source for this website  (<https://github.com/JSC270/JSC270.github.io>)