

QUEEN'S UNIVERSITY -- SCHOOL OF POLICY STUDIES

**MPA 8XX: Data Science and Analytics for Public Policy**

Winter 2019

Someday: 2:X0 to 5:X0 pm in **Room 334** (Robert Sutherland Hall)

**Instructor:** Steven Lehrer

**Office:** Sutherland Hall 324

**Telephone:** 613-533-6692 (office)

**Hours:** Monday, Tuesday 2:45 – 3:50pm  
and by appointment

**E-mail:** lehrers@queensu.ca

**Course Description:** This course aims to provide an understanding of the fundamental concepts and frameworks in the interdisciplinary field of data science and data analytics. The primary goal of this course is for students to learn data analysis concepts and techniques that facilitate making decisions from a rich (and potentially large) data set as well as practical computational skills. The course is also designed to act as a primer for continued study. It is not specifically an introduction to computer science or machine learning or data mining course, nor a class on high-dimensional econometrics and statistics; rather, like a good data scientist, the class borrows from multiple disciplines. Techniques covered include an advanced overview of linear and logistic regression, model choice and false discovery rates, information criteria and cross validation, regularized regression and the lasso, bagging and the bootstrap, causal estimation and treatment effect heterogeneity, binary regression, classification, latent variable models, principal component analysis, topic models, decision trees and random forests, text analysis and natural language processing. Many of these methods have the potential to dramatically improve the public welfare by guiding policy decisions and interventions, and their incorporation into intelligent information systems will improve public services in domains ranging from medicine and public health to law enforcement and security.

Throughout the course, heavy emphasis is placed on analysis of actual datasets and the course will provide an opportunity to utilize an open source data analysis tool, R, for data manipulation, analysis, and visualization. Finally, in this course we will discuss diverse issues around data including tradeoffs in ICTC policies.

**Course Prerequisites:** Students are expected to be comfortable with basic statistics and microeconomics at an introductory level. Students are expected to be able to undertake data analysis with R.

That said, any course that seeks to provide knowledge in data science, will contain mathematics and statistics and require the ability for algorithmic thinking. Individuals who are unfamiliar with math notation, symbols and basic algebra rules should not take this course. To be clear, even though this course is not designed to transform you into a data scientist, your success will be tied to your foundation and ability to develop from your foundation in programming and both mathematical and statistical concepts/theory.

If you have any questions about these requirements, please see me.

**Website:** Additional information, lecture outlines, links to readings, the calendar/reading list, and other useful information about this class can be found on onQ.

**Objectives:** After completing this course, you will be able to:

- 
- Improve critical thinking skills
- 
- Realize that data science has many practical applications within public policy.

**Required Readings:** A reading list will appear on the course calendar. A copy of all the readings listed on the calendar in this syllabus will be placed on reserve at the library. You may be assigned additional readings dealing with topics covered in class. **Specifically assigned** supplementary readings are integral parts of the course, and therefore, exam questions dealing with those readings are **highly likely**.

In general, readings are assigned for each lecture period, and the material in these readings will be discussed in class. Please complete all readings prior to the class in which they will be discussed. The lectures will also cover material not included in the readings. To a large degree, the readings and lectures are not substitutes – they are chosen and designed to complement each other.

A solid primer to much of the material we will cover is An Introduction to Statistical Learning, by James, Witten, Hastie, and Tibshirani. However, it takes a very different approach from us and only partially overlaps on material.

**Supplementary Readings:** You may be assigned additional readings in the form of articles dealing with topics covered in class. **Specifically assigned** supplementary readings are integral parts of the course, and therefore, exam questions dealing with those readings are **highly likely**.

**Software:** All computing is conducted in R, a platform for statistical analysis. R, which is available for free via [www.r-project.org](http://www.r-project.org). You can download and install the software following directions at [cran.us.r-project.org](http://cran.us.r-project.org) (do this ASAP). R is a widely used and hugely flexible analysis platform. It has a command line interface (you type commands to get what you want). Some students find the learning curve for such ‘programming’ to be very steep. I provide limited software instruction, in-class demonstration, and code to accompany lectures and assignments. I assume that you have not used R in a previous class. However, this is not a class on R. Like any language, R is only learned by doing. You should install R as soon as possible and familiarize yourself with basic operations. Ideally, you would start this course able to replicate any analysis from previous classes (e.g., MPA 805) in R. A great way to start learning is to buy a book and start working through tutorials. A good guide is Adler’s R in a Nutshell. They have many tutorials to help you get up to speed. You can browse other options by searching ‘R statistics’ on Amazon. To make it possible to focus on data science concepts, I strongly encourage students to learn the basics of the language and software BEFORE starting the second week of the class. See the next section for detail and resources.

R studio is a free platform for both writing and running R, available at [www.rstudio.org](http://www.rstudio.org). Some students find it friendlier than basic R (especially in windows OS)

Additional R resources

- Tutorials at [data.princeton.edu/R](http://data.princeton.edu/R) are fantastic (and there are many others out there).
- youtube intros to R
- Me and your classmates: work together, and chat on the discussion board.

**Grading:** This course is designed to be very rigorous and demanding. You are expected to work hard, actively participate in class, ask questions when you have any doubts, and perform to the very best of your ability. Although the material is challenging, the purpose of this course is to teach you something about program evaluation, not to destroy your GPA.

Grading will follow QSPS guidelines. The course grade will be computed using the following weights

Class Participation/Attendance	6%
Assignments	50%
Final Exam	<u>44%</u>
Total	100%

**Exam:** The final exam will be cumulative and account for 44 percent of the course grade. Missing the exam without advance notice will result in a zero grade. As well, evidence of the calamity must be provided before as discussed below.

**Exam Policy:** During the exam students are allowed to use calculators, rulers, pens, pencils and erasers. No other materials will be permitted without prior permission from the instructor.

**On class participation:** Class discussion is important for both individual and collective learning. The quality of a student's participation is at least as important as the quantity, and the following points characterize effective participation: . Do comments draw on the text and materials from this and other courses? Do they show evidence of analysis? Does the student distinguish between positive and normative analysis? Does the student distinguish between opinion and well-supported analysis? . Are the points made substantive? Are they linked to the comments of others? Do they advance or deepen the discussion? Do they deepen the analysis? Do comments clarify and highlight the important aspects of earlier comments and lead to a clearer statement of the concepts being considered? Is there an attempt to synthesize the discussion?

**On assignments:** Assignments will be created and graded to ensure that each student can better assess their progress in the course.

**If you need help:** If you find that you are having difficulty with any of the material in this course:

(1) DO NOT let it build up. The material is very cumulative in nature and you are likely to find yourself only falling further behind.

(2) DO come and see me, either after class or by making an appointment. Be forewarned: I expect that you have read the appropriate sections of the textbook and reviewed your notes BEFORE you come to my office.

**Attendance and Lateness:** All students are expected to attend class regularly. Although attendance will not be taken each and every class, be warned that you are responsible for all material covered in class including that, which is not in the text. You are expected to make every effort to be on time to class. If for some reason you must be late, try not to disturb others while entering the classroom.

If you are unable to write the exam (i.e. if you are sick, personal issues, etc.), you must let me know BEFORE the exam time. If you cannot write the exam for some reason, do not sit down to write the

exam. In the interest of fairness, you will be graded if you come into the exam room and see the exam. If you are attending a Queen's activity, you must provide me with notice at least two weeks prior to the exam date so that an equivalent make-up exam can be created.

**Academic Integrity:** Academic Integrity is constituted by the five core fundamental values of honesty, trust, fairness, respect and responsibility (see [www.academicintegrity.org](http://www.academicintegrity.org)). These values are central to the building, nurturing and sustaining of an academic community in which all members of the community will thrive. Adherence to the values expressed through academic integrity forms a foundation for the "freedom of inquiry and exchange of ideas" essential to the intellectual life of the University (see the Senate Report on Principles and Priorities <http://www.queensu.ca/secretariat/policies/senate/report-principlesand-priorities>).

Students are responsible for familiarizing themselves with the regulations concerning academic integrity and for ensuring that their assignments conform to the principles of academic integrity. Information on academic integrity is available in the Graduate Studies Calendar ([http://www.queensu.ca/calendars/sgsr/Academic\\_Integrity\\_Policy.html](http://www.queensu.ca/calendars/sgsr/Academic_Integrity_Policy.html)). Departures from academic integrity include plagiarism, use of unauthorized materials, facilitation, forgery and falsification, and are antithetical to the development of an academic community at Queen's. Given the seriousness of these matters, actions which contravene the regulation on academic integrity carry sanctions that can range from a warning or the loss of grades on an assignment to the failure of a course to a requirement to withdraw from the university.

**Disability Accommodations:** Queen's University is committed to achieving full accessibility for persons with disabilities. Part of this commitment includes arranging academic accommodations for students with disabilities to ensure they have an equitable opportunity to participate in all of their academic activities. If you are a student with a disability and think you may need accommodations, you are strongly encouraged to contact Student Wellness Services (SWS) and register as early as possible. For more information, including important deadlines, please visit the Student Wellness website at: <http://www.queensu.ca/studentwellness/accessibility-services/>

**Teaching Style:** There are many ways to learn. And different styles are more effective for some students than others. Therefore, we will utilize several different approaches: straight lectures, Powerpoint slides, problem sets, and exam preparation.

**Accommodation after the fact:** Once a student has written an exam or submitted an assignment, they may not subsequently be granted accommodation such as being offered a second opportunity to write the exam or assignment or have it count for less than originally specified in the course syllabus (reweighted). Students who cannot perform to the best of their abilities due a serious, extenuating circumstance must inform their instructor before attempting an exam or completing a course to arrange appropriate accommodation.

**Receiving a grade below 70:** In the event of receiving a final grade between 60 and 69.5 in this course, the School of Policy Studies will allow the student to write a qualifying examination to assess if the students' knowledge of this field meet the minimum standard for the respective degree. Passing the qualification exam will result in receiving a passing grade of B- for the course. The qualifying exam is not a make-up exam. The qualifying exam must be written by May 31, 2019.

Receiving a grade below 60 will result in a course failure. In this case, students will be required to either repeat the course or take an approved substitute. Note that The School of Graduate Studies regulations require students to have no more than one course failure.

**Travel during exams:** According to university regulations, students are expected to be available to write scheduled exams at any time during the official December and April examination periods as well as during any scheduled class times. Requests to write a make-up exam because of conflicting travel plans (e.g. flight bookings) or requests to miss an in class exam due to other plans will NOT be considered except under extraordinary circumstances. Students are advised to wait until the final exam schedules are posted before making any travel arrangements.

**Office Hours:** The office hours are tentatively set for XYZ. If you have a class conflict, please email me so that we can come up with another mutually convenient slot. Please internalize externalities and avoid unscheduled visits to my office.

**References:** In general, I am happy to provide references for employers or write letters of reference for students who plan to attend graduate school. The strength of my recommendation remains positively correlated with your performance in my course. For job references, please email me with a heads up that a potential employer might call or email. Please also let me know if there are any skills of yours that I should highlight in my reply to them. Naturally, make sure that these claims are credible as my reputation is on the line. If you would like a letter of reference for graduate school please place an addressed and stamped enveloped in my mailbox along with a short note explaining what the reference is for and when it is due. Also attach a statement of purpose (if relevant) as well as a current CV. Please allow 3 weeks for the completion of letters.

## Tentative Calendar and Reading List

(Subject to Change)

Date	Topic	Assigned Reading
Week 1	Course Introduction	
Week 2	Regression Review and Binary Responses	
Week 3	Model Selection	
Week 4	Clustering	
Week 5	Text Analysis	
Week 6	The bootstrap	
Week 7	Classification	
Week 8	Regression Trees	
Week 9	Forests	
Week 10	Networks	
Week 11	Data Regulation and Policy	
Week 12	Catch up and Review	