POLI 618: Techniques of Empirical Research

Professor Aaron Erlich

Autumn 2017

E-mail: aaron.erlich@mcgill.ca

Office Hours: W 4-5 F 10-11 (drop-in)

Office: 3610 McTavish, Room 26-4

Phone: 514-398-4756

Class Hours: W,F 8:55 AM - 9:55 AM

Class Room: MUSIC A-412

TA: **Aengus Bridgman**

Office Hours: TBD

Office: TBD

E-mail: aengus.bridgman@mail.mcgill.ca

Conference Hours: TBD

Conference Room: Check your schedule

1   Course Description

This course is designed to introduce graduate students to the new and exciting world of data driven political analysis. The course employs examples from across political science sub-disciplines and is generally relevant to all social science research. The course is divided into lectures and labs.

1.1   Who this course is for?

- This is your first semester-long course on graduate quantitative methods course for data analysis.

- You are willing to spend time considerable outside of the classroom to learn the course materials, as data analysis is a skill learned by doing.

2   Course Objectives

1. Learn the basic tools of empirical research in political science.

2. Obtain skills in R, a highly powerful and FREE programming language used extensively by academics in political science across the world, as well as the open source and data science community.

3. Understand scientific replicability.

4. Gain real world skills that will help you obtain jobs in careers of the future.

3   Textbooks

Given people's various backgrounds, we will have three required and several optional textbooks in addition to the assigned articles. The Bailey book is a great book and very applied. Everyone should read it. For those of you seriously interested in pursuing quantitative analysis, you should then read the Fox book on top of the Bailey book. Unless otherwise cleared with me, everyone is required to do the math review problems assigned from the Moore book. The Grolemund and Wickham book is a great tool (that is online and free) and can often be used in lieu of videos or to help with coding.

Bailey, Michael A. (2015). *Real Stats: Using Econometrics for Political Science and Public Policy*. 1st. New York: Oxford University Press.

Moore, Will H. and David A. Siegel (2013). *A Mathematics Course for Political and Social Research*. Princeton University Press.

Grolemund, Garrett and Hadley Wickham (2016). *R for Data Science*. http://r4ds.had.co.nz/.

4   Recommended

Fox, John (2015). *Applied Regression Analysis and Generalized Linear Models*. 3rd ed.. Los Angeles: SAGE Publications, Inc. ISBN: 978-1-4522-0566-3.

5   R and R-Studio

R is a FREE open source programming language uses by data scientists and statisticians across the world. We will be using a FREE graphical user interface (GUI) called **RStudio** that makes learning R easier. While the learning curve in R is steeper than expensive programs, there is much more you can do with it!

There are many free online tutorial for downloading and installing R and RStudio. R will also be available on the conference computers.

6 Evaluation Policy

A description of the means of evaluation to be used in the course:

There are 1000 points available in the class. Therefore, for each 10% of the grade, 100 points are allotted.

- **10%**. *Attendance and participation.*

- **40%**. *Problem sets*. A quarter of this grade will be from peer-reviewing each others' assignments on *GitHub*.

- **10%** *Final Presentation*. In class presentation of research.

- **40%** *Final Paper*. A paper — likely a replication paper or part of your dissertation or MA Thesis.

6.1 Re-Grading

Students who wish to contest a grade for an assignment or exam must do so in writing (by email, sent to me) providing the reasoning behind their challenge to the grade received within two weeks of the day on which the assignments are returned. I will re-evaluate the paper, but also reserve the right to **raise or lower the grade**. Please also see (http://www.mcgill.ca/politicalscience/files/politicalscience/assessment_and_re-read_policy_final.pdf).

7 Assignment Submission

Problem sets must be submitted via *GitHub*. You must use a variant `.Rmd` otherwise known as `rmarkdown` files that is GitHub compatible to complete your homework. Do not submit your homework using Microsft Word or any other document editor. It will not be graded. Part of your homework grade will be peer-reviewing your colleagues' homework assignments via the course *GitHub*. There will be approximately 5 problem sets. I reserve the right to lower the number of problem sets if needed. In principle, these assignments will be due every 2 weeks starting week 2 for the proceeding 12 weeks of the term. Collaboration is part of learning how to code. I encourage you to collaborate! But you do not learn how to do statistical programming if you DO NOT write your own code. Please feel free to collaborate with colleagues, but please DO NOT copy each others' code verbatim. You must also write your own interpretations of the results.

7.1 Interim Data Set and Check-in

All students must submit a one page write up of the data set they are going to use and the research question they are going to ask by Monday, Oct 23. This should be a one page write up in `rmarkdown` explaining the data set/s which you are going to use and the question you will ask. You should also highlight your outcome variable. Submission will be via *MyCourses*.

7.2 Presentation

Each individual will present their work in class to the rest of the course. All presentations must be completed in .Rpres, revealjs_presentation, or Beamer (LaTeX).

7.3 Paper

All students will submit a final paper that is of article length. This will be done via *MyCourses*. This can be *either* a replication paper with an extension of the original paper *or* a new paper. It is highly recommend you do a replication. Please come talk me as early in the term as possible if you want to write an original paper. An original paper must contain 1) a clear theory proposing a relationship between explanatory variable(s) on an outcome variable; 2) use of linear regression (or some other model cleared with professor); 3) a clear discussion of both findings and limitations of the paper. Students may use a chapter of their master or Ph.D. thesis as a research paper. The paper is due by the end of the exam period Wednesday, Dec 21.

8 Class Discussion List and E-mail Policy

I have set up a class discussion list on *MyCourses*. I encourage you to use this mailing list to ask questions you may have. *NEVER* post your code or specific homework questions on the course list server. Please post general questions! If you post homework code on the website, it will be taken down and your grade may be lowered.

9 Make-Up Work Policy

If you are absent for documented emergency medical or family reasons, an alternative homework submission date will be arranged. The alternative arrangement is only open to those who can provide a valid medical/family reason for missing the quizzes. If you cannot provide a valid reason for your absence, you will receive 0 points for the missed quiz or homework submission.

Students who need to miss a class due to a religious holiday should notify me at least fourteen days prior to the holiday. If you must miss a class, an examination, a work assignment, or a project in order to observe a religious holy day, you will be given an opportunity to complete the missed work within a reasonable time after the absence.

10   Technology Policy

10.1   Recording Policy

- No audio or video recording of any kind is allowed in class without the explicit permission of the instructor.
- Mobile Computing devices are not to be used for voice communication without the explicit permission of the instructor.

11   Academic Integrity

11.1   Course Policy on Computer Code

As discussed in the problems set section, verbatim copying other people's computer code constitutes plagiarism. Moreover, data programming is learned through trial and error. **Please do not under any circumstances copy another students code.** You may of course collaborate with colleagues, but please write your own code! If you are found to have plagiarized, you may be referred to the appropriate Dean. The instructors reserve the right to use software to compare the code that has been written by different students.

11.2   McGill Policy

"McGill University values academic integrity. Therefore, all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures" (see www.mcgill.ca/students/srr/honest/ for more information).

12   Other Policies

12.1   Language of Submission:

In accord with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French any written work that is to be graded.

12.2 Disabilities Policy

As the instructor of this course I endeavor to provide an inclusive learning environment. However, if you experience barriers to learning in this course, do not hesitate to discuss them with me and the Office for Students with Disabilities, 514-398-6009.

12.3 End of Course Evaluations

End-of-course evaluations are one of the ways that McGill works towards maintaining and improving the quality of courses and the student's learning experience. You will be notified by e-mail when the evaluations are available. Please note that a minimum number of responses must be received for results to be available to students.

13   Class Schedule

Week 01, 09/04 - 09/08: Introduction, Prediction and Causation, Intro to R

- VIDEOS:
  - Data camp: Intro R lectures Chapters 1 (Intro to Basics), 2 (Vectors), 4 (Factors), and 5 (Data Frames)
  - Data camp: R-markdown lectures
- READING:
  - Moore Ch. 1
  - Bailey Ch. 1
  - Grolemund & Wikham Ch. 27
- TASKS
  - Complete tryGit
  - Install git on your computer and push a test file to GitHub
  - Fill out Pre-course Survey (Google Form)

Week 02, 09/11 - 09/15: Introduction to R-Markdown, Managing Data, Replication, Git and GitHub

- VIDEOS:
  - Data camp: Data minipulation lectures
- READING:
  - Moore Chapter 5, 6
  - Bailey Chapter 2
  - Dafoe, Allan (2014). "Science Deserves Better: The Imperative to Share Complete Replication Files". In: *PS: Political Science & Politics* 47.1, pp. 60–66.
  - Eubank, Nicholas (2016). "Embrace your Fallibility: Thoughts on Code Integrity". In: *The Political Methodologist* 23.2, pp. 10–15. (Visited on Nov. 21, 2016).
  - Moravcsik, Andrew (2010). "Active Citation: A Precondition for Replicable Qualitative Research". In: *PS: Political Science & Politics* 43.01, pp. 29–35. (Visited on Nov. 30, 2016).

Week 03, 09/18 - 09/22: Probability and Statistical Inference

- READING:
  - Moore Chapter 9
  - Galimard Chapter 3 (photocopy)

Week 04, 09/25 - 09/29: Intro to graphics, Linear Regression Basics Review

- VIDEOS
  - Data camp: ggplot lectures
- READING:
  - Moore 2 (if you need algebra review)
  - Bailey Chapters 3, 4, 5
  - Fox Chapter 2
  - Grolemund Chapter 3
  - Kastellec, Jonathan P. and Eduardo L. Leoni (2007). "Using Graphs Instead of Tables in Political Science". In: *Perspectives on Politics* 5.4, pp. 755–771. ISSN: 1541-0986, 1537-5927. (Visited on Oct. 25, 2016).
  - King, Gary, Michael Tomz and Jason Wittenberg (2000). "Making the Most of Statistical Analyses: Improving Interpretation and Presentation". In: *American Journal of Political Science* 44.2, pp. 347–361. ISSN: 0092-5853. (Visited on Apr. 28, 2011).

Week 05, 10/02 - 10/06: Data Transformation

- READING:
  - Moore Chapter 3
  - Bailey Chapters 6.1-6.3, 7
  - Fox Chapters 4, 7.1-7.2

Week 06, 10/09 - 10/13: Interaction Terms, Simulation

- READING:
  - Moore Chapter 7
  - Bailey 6.4
  - Fox 7.3
  - Berry, William D, Matt Golder and Daniel Milton (2012). "Improving Tests of Theories Positing Interaction". En. In: *The Journal of Politics* 74.3, pp. 653–671. ISSN: 0022-3816, 1468-2508.
  - Brambor, Thomas, William Roberts Clark and Matt Golder (2006). "Understanding Interaction Models: Improving Empirical Analyses". En. In: *Political Analysis*, pp. 63–82.
- **PAPER CHECK IN : Monday, Oct 23**

Week 07, 10/16 - 10/20: Assumptions & Properties of Linear Regression

- READING:

- Bailey Ch. 14
- Fox Ch.6

Week 09, 10/30 - 11/03 : LS and Causal Inference

- READING:
  - TBD
  - TBD
- NO CLASS NOVEMBER 23, WORK ON YOUR PAPERS

Week 08, 10/23 - 10/27: Matrix Presentation of LS

- READING:
  - Moore Chapter 12
  - Fox Chapter 9.1-9.2

Week 10, 11/06 - 11/10: Linear Regression Diagnostics and Fixes

- READING:
  - Moore Chapter 12
  - Fox Chapter 9.1-9.2

Week 11, 11/13 - 11/17: Logit/Probit and Linear Probability Model

- READING:
  - Bailey Ch. 12
  - Fox Ch. 14

Week 12, 11/20 - 11/24: Presentations 1

Week 13, 11/27 - 12/01: Presentations 2

14   Recommended Textbooks

There are many other important textbooks and at some point you may find yourself looking for a different explanation of something you didn't understand – or looking to go deeper. Here are some places to start.

Angrist, Joshua D. and Jörn-Steffen Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. ISBN: 978-1-4008-2982-8.

Gailmard, Sean (2014). *Statistical Modeling and Inference for Social Science*. New York, NY: Cambridge University Press.

Gelman, Andrew and Jennifer Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Greene, William H. (2003). *Econometric analysis*. Pearson Education.