

# Course outline: POLSCI 784: Quantitative Political & Policy Analysis

Term 2, Winter 2023

Thursdays, 2:30pm - 5:20pm EST

Classroom: LR Wilson 3001

Course website: [On GitHub.com](#)

Class calendar: [Outlook calendar](#)

## Instructor information

Dr. Michelle L. Dion

## Office Hours

**Normally: Mondays, 1:30pm – 3:30pm in KTH 533 or Zoom.** Please use [this form](#) to sign up for a 20 minute meeting, up to one week in advance.

However, I *may occasionally* need to shift these hours due to other commitments. The [Class Outlook Calendar](#) will always be up-to-date.

If you cannot make office hours due to your own prior commitments, please see me before class starts or during a break to make an appointment (rather than email me).

## Table of Contents

Course outline: POLSCI 784: Quantitative Political & Policy Analysis.....	1
Term 2, Winter 2023 .....	1
Thursdays, 2:30pm - 5:20pm EST .....	1
Classroom: LR Wilson 3001.....	1
Course website: On GitHub.com.....	1
Class calendar: Outlook calendar .....	1
Instructor information .....	1
Office Hours .....	1

Table of Contents .....	1
Course Description .....	3
Course Objectives .....	4
Assigned Texts .....	4
Computing and Statistical Software .....	4
Primary R textbooks & websites.....	5
Reference or alternative resources.....	5
Statistics textbooks (some available from the library, order from basic to advanced) .....	5
Other statistics reference materials online .....	6
Additional R reference and learning resources .....	6
Git & GitHub background & resources .....	6
Other class-related website.....	7
Getting help.....	7
Class notes .....	7
Learning Evaluation .....	7
Learning teams and engagement.....	8
Practice assignments (4 assignments, 24% total of final grade).....	8
Research project (6 assignments, 61% total of final grade).....	9
Weekly Course Schedule .....	9
January 12: Introduction.....	9
January 19: Setup & univariate descriptive statistics .....	10
January 26 Hypothesis testing and statistical significance.....	11
February 2 Exploratory data visualization .....	11
February 9 Association & simple, bivariate regression.....	12
February 16 Regression assumptions and statistical inference (intro to multivariate) .....	13
February 22 Reading week .....	13
March 2 Multiple, multivariate regression.....	13
March 9 Categorical independent variables, non-linear relationships, & interactions .....	14
March 16 Outliers and predicted outcomes.....	14
March 23 Collinearity, F-tests, adjusted $(R^2)$ , & model specification.....	14
March 30 Model fitting & assumptions .....	15
April 6 Logistic regression & other advanced models.....	15

April 13 Optional Virtual Lab .....	15
April 20 Final report due .....	15
Course Policies .....	15
Submission of Assignments .....	15
Grades .....	16
Late Assignments .....	16
Collaboration .....	16
A note on “reverse engineering” VS “copying” or “plagiarism” .....	16
Absences, Missed Work, Illness .....	17
Courses with an On-Line Element .....	17
University Policies .....	17
Copyright and Recording .....	17
Academic Accommodation for Religious, Indigenous or Spiritual Observances (RISO) .....	18
Academic Integrity Statement .....	18
Conduct Expectations .....	18
Academic Accommodation of Students with Disabilities .....	18
Faculty of Social Sciences E-mail Communication Policy .....	19
Course Modification .....	19
Extreme Circumstances .....	19

## Course Description

This is an *introductory* graduate course in empirical research and statistical methods in the social sciences. For MA students, the intention is to provide you with basic statistical skills and familiarity for use on the job market. For PhD students, the goal is to provide a basic foundation for more advanced coursework or applications in your research.

For some of you, the material presented in this course will be the beginning of a radically new way to approach research. To be successful in the course, you will NOT need to be a mathematician, statistician, or computer programmer, but you will need a desire to learn, to solve problems, and be open to new ways of thinking.

*Reading all* assignments and instructions and asking questions if you need additional explanation or clarification is also important. You will also need some basic algebra skills. A copy of this outline and related web-based resources for the course can be found at the course websites:

- [Course website](#) for notes, discussion board, resources, downloading and uploading assignments, etc.
  - You will receive an email invitation to join the course webpage.

- Once you accept the invitation, you will create a github.com account, if you don't already have one.

## Course Objectives

The course will provide an introduction to basic frequentist statistical methods in the social sciences through linear (and to a lesser extent logistic) regression. The emphasis will be on successful application of statistical methods and understanding the uses of such methods for public policy and social science. To gain experience in applying statistical analysis, students complete a series of practice assignments and an independent research project employing linear or logistic regression. Students will also gain experience using tools of data science, such as RStudio & GitHub. They will also learn how to plan, manage, and execute an applied quantitative research project.

## Assigned Texts

All textbooks and resources for this course are available without cost online. Some books and resources are available through McMaster University Library subscriptions, and you'll [either need to be on campus or] login to the library website for access from off campus. Some are open source books or websites. Many are also available for purchase in paper formats, including as used textbooks (e.g., from an [independent bookseller](#), [brick & mortar chain](#), or [online](#)).

**Link on item *title* goes to McMaster Library login for off-campus access.**

- Lewis-Beck, Michael. 1995. *Data Analysis*. Thousand Oaks, CA: SAGE. <https://dx.doi.org/10.4135/9781412983846>
- Schroeder, Larry D., David L. Sjoquist, and Paula E. Stephan. 1986. *Understanding Regression Analysis: An Introductory Guide*. Thousand Oaks: SAGE. <https://dx.doi.org/10.4135/9781412986410>
- Lewis-Beck, Colin, and Michael Lewis-Beck. 2016. *Applied Regression: An Introduction*. Thousand Oaks: SAGE. <https://doi.org/10.4135/9781483396774>
- Fox, John. 1991. *Regression Diagnostics*. Thousand Oaks, CA: SAGE. <https://dx.doi.org/10.4135/9781412985604>
- Hardy, Melissa A. 1993. *Regression with Dummy Variables*. Thousand Oaks: SAGE. <https://dx.doi.org/10.4135/9781412985628>
- Kahane, Leo H. 2008. *Regression Basics*. 2nd ed. Thousand Oaks: SAGE. <https://dx.doi.org/10.4135/9781483385662>
- Menard, Scott. 2002. *Applied Logistic Regression Analysis*. 2<sup>nd</sup> ed. Thousand Oaks: SAGE. <https://dx.doi.org/10.4135/9781412983433>

## Computing and Statistical Software

In this course, we will be using the R Statistical Computing language to analyze data. R is increasingly used in the private and public sectors for data analysis due to its flexibility and power (see [Long & Turner 2020](#)). Much of the power of R comes from user-contributed or developed packages that add new functions or types of analysis to “base” R.

The [tidyverse](#) is a collection of R packages that implement coordinated and consistent data analysis approaches and tools. The “tidy” approach to data analysis in R is quickly displacing “old school” approaches that use base R commands.

To interact with R, we will be using Desktop RStudio, which is a free Integrated Development Environment that provides an interface for R that is similar to interfaces for other statistical software, such as SPSS or STATA. In RStudio, you can run command scripts, interactively analyze your data, and view your data or results, including tables and plots. RStudio also has a range of other features that make it easy to prepare your results to share with collaborators or add to papers.

RStudio also integrates with GitHub.com :octocat: to track and archive your data and files. Git is an open source program that tracks versions of code and documents, and combined with GitHub.com, you can transfer your files to your account directly from RStudio.

We will be using GitHub as our course website, where we will have a private space to share class notes, access and submit practice assignments, and ask questions on a class discussion board. *You will receive an invitation to join the course on GitHub.com. If you don't already have an account on GitHub.com, you will be prompted to create one.*

## Primary R textbooks & websites

- Golemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. *Also available in hard copy.*
  - Solutions manual for *R for Data Science*: Arnold, Jeffrey B. n.d. *R for Data Science: Exercise Solutions*.
- Rodrigues, Bruno. 2020. *Modern R with the tidyverse*.
- Phillips, Nathaniel D. n.d. *YaRrr! The Pirate's Guide to R*. (solutions included in Chapter 18)

## Reference or alternative resources

### Statistics textbooks (some available from the library, order from basic to advanced)

- Pollock, Phillip. 2015. *The Essentials of Political Analysis*. 5th ed. edition. Los Angeles: CQ Press.
- Neil J. Salkind. *Statistics for People who (Think They) Hate Statistics*. 4th edition. Sage. ISBN 1412979595 (covers most of material in class)
- Alan Agresti and Barbara Finlay. *Statistical Methods for the Social Sciences*, 4th edition. Upper Saddle River, NJ: Prentice Hall. ISBN 0130272957 (covers class material)
- Earl Babbie. *The Practice of Social Research*, 13th ed. Wadsworth. ISBN: 1133049796. (covers class material and qualitative methods and data collection methods, like survey design)
- Janet Buttolph Johnson and H.T. Reynolds. *Political Science Research Methods*. CQPress. ISBN: 1608716899. (covers class material, qualitative methods, and general research strategies)
- Damodar N. Gujarati and Dawn Porter. *Essentials of Econometrics* 4th Ed. ISBN: 0073375845 (solid basic introduction to regression and some advanced topics, does not use matrix algebra)
- Damodar N. Gujarati and Dawn Porter. *Basic Econometrics*. 5th ed. ISBN: 0073375772 (solid basic introduction to regression and some advanced topics, used in many political science grad programs, uses matrix algebra)

- William H. Greene. *Econometric Analysis*. 7th ed. Prentice Hall. ISBN: 0131395386. (advanced econometric text that covers most advanced methods in more detail than Gujarati, used in many political science grad programs)

## Other statistics reference materials online

Before consulting Wikipedia or Google for basic statistical concepts or terms, search one of these methods encyclopedias available through the McMaster Library website:

- Lewis-Beck, Michael, Alan Bryman, and Tim Liao. 2004. *The SAGE Encyclopedia of Social Science Research Methods*. 3 vols. Thousand Oaks, California. <https://doi.org/10.4135/9781412950589>
- SAGE Little Green Books.

NB: SAGE also has a Little Blue Book series for qualitative methods.

## Additional R reference and learning resources

- RStudio [cheat sheets for various R and tidyverse packages](#)
- RStudio Education. [Finding Your Way To R](#).
- Michelle Dion's [collection of R resources](#)
- UCLA's IDRE. [R Statistical Computing](#).
- McConville, Chester Ismay and Albert Y. Kim Foreword by Kelly S. 2019. *Statistical Inference via Data Science*. New York: CRC Press. *Also available in hard copy*.
- Bruno Rodrigues. 2020. *Modern R with the tidyverse*.
- Mailund, Thomas. 2019. *R Data Science Quick Reference: A Pocket Guide to APIs, Libraries, and Packages*. Berkeley, CA: press. DOI. Separate chapters for different tidyverse packages, similar to *R for Data Science*.
- LinkedIn Learning (Lynda.com) via McMaster subscription. McMaster University maintains a subscription to various online courses hosted at LinkedIn Learning (formerly Lynda.com). These courses are not a substitute for class lectures or working through practice assignment exercises, but I include links to short (all < 10 min. and most < 5 min.) videos that cover some of each week's content as an additional resource. These are optional and supplemental. The full list of available courses is available at: <https://uts.mcmaster.ca/services/teaching-and-learning/linkedin-learning/>. To use the links listed in the [Weekly Schedule](#) below:

1. Login at: <https://uts.mcmaster.ca/services/teaching-and-learning/linkedin-learning/>
2. While your browser is still open, in another tab, click on the links to the videos on this page you want to view.

## Git & GitHub background & resources

Some of these may discuss some advanced features that we do not use.

- [JennyBC's instructions on installing git](#)
- [RStudio and Git - an Overview \(Part 1\)](#)
- [A simple overview](#)
- [RStudio and Git - an Example \(Part 2\)](#)

- [RStudio & Github Integration](#)
- [How to use Git and GitHub with R](#)
- [Happy Git and GitHub for the useR](#) - online book for students explaining git/github
- [Background on git, markdown, etc. by Jeff Racine \(Econ, McMaster\)](#)
- [Another nice overview](#)
- [Hadley Wickham's \(tidyverse creator\) page on git, GitHub & RStudio](#), lots of detail.

## Other class-related website

On [this website](#), I save links to other resources on the web, with a focus on: - [websites with data](#), tagged “data” + other keywords that describe the dataset; - [websites related to statistical methods and tools](#), tagged “stats” + other keywords that describe the method; and - websites related to statistical computing and related tools, tagged “stats” + name of the tool (e.g., [R](#), [R Studio](#), [tidyverse](#), etc.). The data links are particularly useful for finding data for projects.

## Getting help

If you've been working on the *same line of code* for more than 20 minutes or are seeing an error message that you haven't seen before, your classmates and instructor are the best source of help for specific assignments in this course. To post questions about assignments, we have a [discussion board](#).

- Use this space to ask questions about assignments, class content, or to organize yourselves into study groups.
- I prefer to answer most questions on the discussion board because then everyone has the same information about an assignment or class content. I find that if one student has a question, it is likely that at least one other (and often more!) student has the same or a similar question. You can always tag me [@michelledion](#) to make sure that I see your question.
- Answering each others' questions here also helps everyone practice asking for help online and providing it to others. Most people will find that using an online forum like this is useful not just for this class but for other technical issues (even outside of class or work) that you may encounter in the future.
- I will often leave a query unanswered for a day to give others a chance to answer, but then will either confirm advice provided by others or suggest additional strategies.
- I will sometimes suggest that you remind me to answer a question in class or to come to office hours. In those cases, the question may be better answered when I can “show” you the solution or add nuance to the answer.
- [This StackOverflow page](#) has some useful guidance about how to post a good question, and I would recommend trying to follow this guidance when you post to the class discussion board.

## Class notes

Class materials (including all the code and data used to generate the notes) are stored in this repository in the [notes folder](#).

I will demonstrate how to use the materials in this folder to take notes during class.

## Learning Evaluation

Assignment due dates are included in the [Weekly Schedule](#) below and on the [course Outlook Calendar](#).

## Learning teams and engagement

Students will be organized into learning teams of ~3 students each. Each learning team will exchange class notes. (More about the logistics of this later.)

### In-class activities and problems and team note exchange

Sometimes during class, we will have short examples or problems to work through, and learning teams will work together on these activities and the relevant updates to each student's class notes. In addition, members of each learning team will have "read only" access to their team members' class notes. This will help students see how others "solve" the same coding problem. They will also allow students to check their understanding by reviewing their class notes alongside those of other team members.

**However, each student's notes must be *in their own words* and under no circumstances should students copy/paste from a team members' notes into their own notes.**

### Class engagement (15% of final grade)

Each student will be *individually* evaluated based on their attendance, participation, collaboration and overall engagement in class. Students' engagement mark will be based on the following expectations:

- Students will be expected to *attend class regularly*.
- Students are expected to *read this Course Outline and the instructions* for all Practice Assignments and the Project in their entirety and ask questions, if anything is unclear.
- Students will be expected to *"push" their class notes at the end of class each week* to confirm their attendance and participation in any in-class activities.
- Students can update their class notes after class, *only if they also "pushed" their notes at the end of class*.
- Students who miss class will be expected to consult their learning team members' notes to review any information they may have missed and to complete any coding exercises done during class. *However, all notes must be in each student's "own words" and copy/paste across class notes is not allowed.*
- Students are encouraged to use the [GitHub team discussion board](#) to post questions. Answering others' questions will also be recognized as class engagement.
- Class notes may be reviewed by the instructor for the purposes of evaluating student engagement throughout the semester.

## Practice assignments (4 assignments, 24% total of final grade)

Practice assignments are designed to build skills and develop confidence and competence with collaborative assignments. We will sometimes use class time to complete some parts of practice assignments (e.g., some piece of code), and everyone will need to answer remaining questions (e.g., those that require interpretation) after class and assemble and submit the complete assignment for credit.



Collaboration to solve coding bugs and confirm understanding of course content is encouraged. However, the final submitted assignment should be each student's individual work. Students should explain with whom and how they collaborated on practice assignments.

To be effective as a learning support, practice assignments must be completed according to the course schedule, and late practice assignments will be accepted only *under exceptional circumstances and prior approval from the instructor*.

There will be 4 practice assignments, each worth the following portion of the final marks.

1. Descriptive statistics and basic visualization - 6%
2. Bivariate regression - 6%
3. Multivariate regression - 6%
4. Regression assumptions - 6% (Assignment does not require coding/data analysis)

## Research project (6 assignments, 61% total of final grade)

More than half of your final mark in this course will be based upon your completion of an original research project using quantitative data (e.g., a microdata file of survey responses or an aggregate dataset that you compile based on published sources) and linear (or logistic) regression.

The project will proceed in phases to give you guidance and feedback throughout the research process. You will submit a revised final report that assembles all the project assignments into one revised, final document.

Though students will complete the project in stages, there is only one project repository per student.

Detailed instructions for each project assignment are in the README.md file in the [project repository](#).

Each student will also have a private project repository shared with the instructor. Submission of research project assignments through the project repository will be discussed and demonstrated in class.

*I may edit project assignment templates or add other resources during term, so please "pull" updates to the repository before you begin working on a particular project assignment.*

Your final research project will proceed in phases:

1. Statement of research question with clear identification of dependent variable and preliminary bibliography (5%)
2. Description of research hypotheses (8%)
3. Diagram of research design and dataset (8%)
4. Descriptive statistics (10%)
5. Regression analysis (15%)
6. Final report (15%)

## Weekly Course Schedule

### January 12: Introduction

Due: Create GitHub.com [educational] account

## Required readings:

- Achen, Christopher H. 2002. “Advice for Students Taking a First Political Science Graduate Course in Statistical Methods,” *The Political Methodologist* 10 (2): 10–12.
- Lewis-Beck, Michael S. 1995. *Data Analysis: An Introduction*. Thousand Oaks: SAGE, pgs. 1-8.
- Chin, Monica. 2021. “Students Who Grew up with Search Engines Might Change STEM Education Forever.” *The Verge*. September 22, 2021. <https://www.theverge.com/22684730/students-file-folder-directory-structure-education-gen-z>.
- Golemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O’Reilly. Chapters 1-2.
- [Setting up your computing environment for class](#)

## Review mainly for structure & description of the data & analysis:

- Davis, Darren W., and Brian D. Silver. 2004. “Civil Liberties vs. Security: Public Opinion in the Context of the Terrorist Attacks on America.” *American Journal of Political Science* 48 (1): 28–46.
- Dion, Michelle L., and Catherine Russler. 2008. “Eradication Efforts, the State, Displacement and Poverty: Explaining Coca Cultivation in Colombia during Plan Colombia.” *Journal of Latin American Studies* 40 (3): 399–421.

## Additional R resources:

- Phillips, Nathaniel D. n.d. *YaRrr! The Pirate’s Guide to R*. Chapters 1-2.

## LinkedIn:

- [Navigating the RStudio Environment: Learning R](#)
- [Why Is the Tidyverse Unique?: R Programming in Data Science: Setup and Start](#)
- [How to Install the Tidyverse: R Programming in Data Science: Setup and Start](#)
- [Packages for R: Learning R](#)
- [The Tidyverse: Learning R](#)

## GitHub & RStudio videos:

- [Github Education Classroom](#).
- [RStudio and Git - an Overview \(Part 1\)](#)
- [RStudio and Git - an Example \(Part 2\)](#)
- [RStudio & Github Integration](#)
- [How to use Git and GitHub with R](#)

# January 19: Setup & univariate descriptive statistics

Due: Install R, RStudio, git & TeX

## Required readings:

- Lewis-Beck, Michael S. 1995. *Data Analysis: An Introduction*. Thousand Oaks: SAGE, pgs. 1-18.

- Grolemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. Chapters 4, 6, 8-9, 26-30.

### Additional R resources:

- Phillips, Nathaniel D. n.d. *YaRrr! The Pirate's Guide to R*. Chapters 3-4, 9.
- Irizarry, Rafael A. n.d. *Introduction to Data Science*. Chapter 39 Git and GitHub.
- Bryan, Jennifer, and Jim Hester. n.d. *What They Forgot to Teach You About R*. Chapters 1-2.
- Rodrigues, Bruno. 2020. *Modern R with the tidyverse*. Chapters 1-4.

### LinkedIn:

- [Importing Data from a Spreadsheet: Learning R](#)
- [What Are CSV Files?: Data Wrangling in R](#)
- [Importing CSV Files into R: Data Wrangling in R](#)
- [Importing Excel Files into R: Data Wrangling in R](#)
- [Loading Data Sets with Read\\_csv: Data Visualization in R with Ggplot2](#)
- [Recoding Variables: Learning R](#)
- [Computing New Variables: Learning R](#)
- [Computing Descriptives: Learning R](#)
- [Computing Frequencies: Learning R](#)

## January 26 Hypothesis testing and statistical significance

Due: Project research question, dataset location, & bibliography

### Required readings:

- Lewis-Beck, Michael S. 1995. *Data Analysis: An Introduction*. Thousand Oaks: SAGE, pgs. 31-41.
- Grolemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. Chapters 3, 5, 7.

### LinkedIn:

- [Using the Tidyverse: Data Wrangling in R](#)
- [Variables, Observations, and Values: Data Wrangling in R](#)
- [What Is Tidy Data?: Data Wrangling in R](#)
- [Subsetting Tibbles: Data Wrangling in R](#)
- [Filtering Tibbles: Data Wrangling in R](#)
- [R Data Types: Basic Types: R for Data Science: Lunchbreak Lessons](#)
- [R Data Types: Data Frame: R for Data Science: Lunchbreak Lessons](#)
- [R Data Types: Factor: R for Data Science: Lunchbreak Lessons](#)
- [Piping Commands with %>%: Learning R](#)
- [Selecting Cases and Subgroups: Learning R](#)

## February 2 Exploratory data visualization

Due: (You should have draft started of Practice Assignment 1 to ask questions in class about code/content.)

Recommended content readings:

- Healy, Kieran, and James Moody. 2014. "Data Visualization in Sociology." *Annual Review of Sociology* 40 (1): 105–28.

Additional R resources:

- Healy, Kieren. 2018. *Data Visualization*. Princeton: Princeton UP. Chapters 1, 3, 4.
- Golemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. Chapters 10-12, 14-15.
- Phillips, Nathaniel D. n.d. *YaRrr! The Pirate's Guide to R*. Chapter 11.
- Rodrigues, Bruno. 2020. *Modern R with the tidyverse*. Chapter 5.

LinkedIn:

- [Introducing Ggplot2: Data Visualization in R with Ggplot2](#)
- [Barplot: R for Data Science: Lunchbreak Lessons](#)
- [Bars and Columns: Data Visualization in R with Ggplot2](#)
- [Dotchart: R for Data Science: Lunchbreak Lessons](#)
- [Histogram: R for Data Science: Lunchbreak Lessons](#)
- [Histograms: Data Visualization in R with Ggplot2](#)
- [Scatterplots: Data Visualization in R with Ggplot2](#)

## February 9 Association & simple, bivariate regression

Due: Practice assignment 1

Required readings:

- Schroeder, Larry, David Sjoquist, and Paula Stephan. 1986. *Understanding Regression Analysis*. Thousand Oaks: SAGE, pgs. 12-29.
- Lewis-Beck, Michael S. 1995. *Data Analysis: An Introduction*. Thousand Oaks: SAGE, pgs. 19-30, 42-53.
- Lewis-Beck, Colin, and Michael Lewis-Beck. 2016. *Applied Regression: An Introduction*. Thousand Oaks: SAGE, pgs. 1-22.
- Kahane, Leo. 2008. *Regression Basics*. 2nd ed. Thousand Oaks: SAGE, pgs. 1-16.
- Golemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. Chapters 17-18, 22-23.

Additional R resources:

- Phillips, Nathaniel D. n.d. *YaRrr! The Pirate's Guide to R*. Chapters 13-15.4.
- McConville, Chester Ismay and Albert Y. Kim Foreword by Kelly S. 2019. *Statistical Inference via Data Science*. New York: CRC Press. Chapter 5.
- Rodrigues, Bruno. 2020. *Modern R with the tidyverse*. Chapters 3-4.

LinkedIn:

- [Computing a Linear Regression: Learning R](#)
- [Computing Contingency Tables: Learning R](#)
- [Computing Correlations: Learning R](#)

## February 16 Regression assumptions and statistical inference (intro to multivariate)

Due: Literature review

Required readings:

- Lewis-Beck, Michael S. 1995. *Data Analysis: An Introduction*. Thousand Oaks: SAGE, pgs. 54-74.
- Lewis-Beck, Colin, and Michael Lewis-Beck. 2016. *Applied Regression: An Introduction*. Thousand Oaks: SAGE, pgs. 23-54.
- Schroeder, Larry, David Sjoquist, and Paula Stephan. 1986. *Understanding Regression Analysis*. Thousand Oaks: SAGE, pgs. 30-53.
- Kahane, Leo. 2008. *Regression Basics*. 2nd ed. Thousand Oaks: SAGE, pgs. 17-58.

Additional R resources:

- McConville, Chester Ismay and Albert Y. Kim Foreword by Kelly S. 2019. *Statistical Inference via Data Science*. New York: CRC Press. Chapter 6.

## February 22 Reading week

## March 2 Multiple, multivariate regression

Due: Practice assignment 2

Required readings:

- Lewis-Beck, Colin, and Michael Lewis-Beck. 2016. *Applied Regression: An Introduction*. Thousand Oaks: SAGE, pgs. 55-74.
- Kahane, Leo. 2008. *Regression Basics*. 2nd ed. Thousand Oaks: SAGE, pgs. 59-78.
- Golemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. Chapters 24-25.

Additional R resources:

- McConville, Chester Ismay and Albert Y. Kim Foreword by Kelly S. 2019. *Statistical Inference via Data Science*. New York: CRC Press. Chapter 6.

## March 9 Categorical independent variables, non-linear relationships, & interactions

Due: Project research design diagram

Required readings:

- Lewis-Beck, Colin, and Michael Lewis-Beck. 2016. *Applied Regression: An Introduction*. Thousand Oaks: SAGE, pgs. 75-96.
- Schroeder, Larry, David Sjoquist, and Paula Stephan. 1986. *Understanding Regression Analysis*. Thousand Oaks: SAGE, pgs. 54-65.
- Kahane, Leo. 2008. *Regression Basics*. 2nd ed. Thousand Oaks: SAGE, pgs. 79-102.
- Hardy, Melissa A. 1993. *Regression with Dummy Variables*. Thousand Oaks: SAGE, pgs. 18-29.
- Fox, John. 1991. *Regression Diagnostics*. Thousand Oaks: SAGE, pgs. 53-65.
- Golemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. Chapters 15, 23.4.

## March 16 Outliers and predicted outcomes

Due: Project descriptive statistics

Required readings:

- Lewis-Beck, Michael S. 1995. *Data Analysis: An Introduction*. Thousand Oaks: SAGE, pgs. 54-72.
- Lewis-Beck, Colin, and Michael Lewis-Beck. 2016. *Applied Regression: An Introduction*. Thousand Oaks: SAGE, pgs. 55-74.
- Schroeder, Larry, David Sjoquist, and Paula Stephan. 1986. *Understanding Regression Analysis*. Thousand Oaks: SAGE, pgs. 54-65. (repeat)
- Fox, John. 1991. *Regression Diagnostics*. Thousand Oaks: SAGE, pgs. 21-40.
- Golemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. Chapters 23.3

LinkedIn:

- [Lines and Smoothers: Data Visualization in R with Ggplot2](#)
- [Plot to File: R for Data Science: Lunchbreak Lessons](#)

## March 23 Collinearity, F-tests, adjusted $(R^2)$ , & model specification

Due: Practice assignment 3

Required readings:

- Schroeder, Larry, David Sjoquist, and Paula Stephan. 1986. *Understanding Regression Analysis*. Thousand Oaks: SAGE, pgs. 66-80.
- Fox, John. 1991. *Regression Diagnostics*. Thousand Oaks: SAGE, pgs. 1-20.

- Kahane, Leo. 2008. *Regression Basics*. 2nd ed. Thousand Oaks: SAGE, pgs. 119-142.
- Golemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. Chapters 24-25.

## March 30 Model fitting & assumptions

Due: nothing

Required readings:

- Fox, John. 1991. *Regression Diagnostics*. Thousand Oaks: SAGE, pgs. 40-53.
- Achen, Christopher H. 2005. "Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong." *Conflict Management and Peace Science* 22 (4): 327–39.
- Schrodtt, Philip A. 2014. "Seven Deadly Sins of Contemporary Quantitative Political Analysis." *Journal of Peace Research* 51 (2): 287–300.
- Golemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. Chapter 25.

## April 6 Logistic regression & other advanced models

Due: Project results

Required readings:

- Kahane, Leo. 2008. *Regression Basics*. 2nd ed. Thousand Oaks: SAGE, pgs. 143-46.
- Menard, Scott. 2002. *Applied Logistic Regression Analysis*. 2nd ed. Thousand Oaks: SAGE, pgs. 1-67.

Additional R resources:

- Phillips, Nathaniel D. n.d. *YaRrr! The Pirate's Guide to R*. Chapter 15.5.

LinkedIn:

- [Using Colors in R: Learning R](#) and next 4 modules are all about customizing your figures

## April 13 Optional Virtual Lab

Due: Practice assignment 4

April 20 Final report due

## Course Policies

## Submission of Assignments

All assignments will be submitted online at [our private organization on GitHub.com](#) in either the project or a practice assignment repository.

## Grades

Grades will be based on the **McMaster University grading scale**:

### MARKGRADE

90-100A+

85-90 A

80-84 A-

77-79 B+

73-76 B

70-72 B-

69-0 F

## Late Assignments

- No late *practice* assignments are accepted.
- Project assignments *may* be turned in after the deadline without an evaluation penalty. However, given the scaffolded nature of class assignments, late project assignments will not be marked before additional project assignments are due. In some cases, late project assignments may not be marked until the end of term, depending on how many students turn in late work.
- Assignments will be marked in the order that they are received.

## Collaboration

By submitting your weekly notes and your final project, you are pledging that you have not received *unauthorized aid* on the notes and project.

While you are encouraged to discuss your projects with peers and the instructor, you must be the only author of your written assignments and the related code.

All references to or paraphrasing of course readings or outside readings must be properly documented to avoid plagiarism. Educate yourself about *patchwriting* and other common types of plagiarism. If you have any doubts, please ask me before turning in the assignment.

## A note on “reverse engineering” VS “copying” or “plagiarism”

Encouraging students to “reverse engineer” code is a common way to teach coding, and working analysts or coders use this strategy to troubleshoot problems in their code or figure out how to apply a solution from an example to their own problem. Reverse engineering entails taking the code apart to see how each line works and then substituting your own variable names or making other adjustments to make the code do what you want it to do. In this course, reverse engineering class notes to do assignments is encouraged and is not considered “plagiarism.” Working with another student to reverse engineer code for practice assignments is acceptable because it is part of the learning process.



What is *not* acceptable? Borrowing a classmate's .Rmd file and substituting your name. This is copying their work and presenting it as if it is your own. This is discouraged both because it is unfair to the student whose work was copied but also because you do not learn how to write the code yourself.

It is also not acceptable to copy, paraphrase, or plagiarize another scholar's (whether student or published work) work on:

- answers to a practice assignment question (e.g., each team should use their own words, not those of another student or of a published source without attribution) or
- *ANY* written work, including descriptions of your data or analysis, as part of your project.

If you have questions about the line between reverse engineering code and unauthorized copying or plagiarism, please ask before you proceed.

## Absences, Missed Work, Illness

Regular attendance is crucial to your success in this course and is expected of all graduate students. In the past, students who have missed even one class have had trouble catching up with the material, and students who have missed more than one class usually have had significant trouble completing the final project to their satisfaction.

## Courses with an On-Line Element

Some courses may use on-line elements (e.g. e-mail, Avenue to Learn (A2L), LearnLink, web pages, capa, Moodle, ThinkingCap, etc.). Students should be aware that, when they access the electronic components of a course using these elements, private information such as first and last names, user names for the McMaster e-mail accounts, and program affiliation may become apparent to all other students in the same course. The available information is dependent on the technology used. Continuation in a course that uses on-line elements will be deemed consent to this disclosure. If you have any questions or concerns about such disclosure please discuss this with the course instructor.

In this course we will be using GitHub.com to distribute course content and host a class discussion board, and GitHub Classroom to submit assignments. Site content will be restricted to course members, and student assignments will all be only visible to the instructor (and TA when there is a TA).

## University Policies

### Copyright and Recording

Students are advised that lectures, demonstrations, performances, and any other course material provided by an instructor include copyright protected works. The Copyright Act and copyright law protect every original literary, dramatic, musical and artistic work, including lectures by University instructors.

The recording of lectures, tutorials, or other methods of instruction may occur during a course. Recording may be done by either the instructor for the purpose of authorized distribution, or by a student for the purpose of personal study *when they have an accommodation that allows them to do so*. Students authorized to record class sessions should not post or redistribute the recordings.

Students should be aware that their voice and/or image may be recorded by others during the class. Please speak with the instructor if this is a concern for you.

## Academic Accommodation for Religious, Indigenous or Spiritual Observances (RISO)

Students requiring academic accommodation based on religious, indigenous or spiritual observances should follow the procedures set out in the [RISO policy](#). Students should submit their request to their Faculty Office normally *within 10 working days* of the beginning of term in which they anticipate a need for accommodation or to the Registrar's Office prior to their examinations. Students should also contact their instructors as soon as possible to make alternative arrangements for classes, assignments, and tests.

## Academic Integrity Statement

You are expected to exhibit honesty and use ethical behaviour in all aspects of the learning process. Academic credentials you earn are rooted in principles of honesty and academic integrity. **It is your responsibility to understand what constitutes academic dishonesty.**

Academic dishonesty is to knowingly act or fail to act in a way that results or could result in unearned academic credit or advantage. This behaviour can result in serious consequences, e.g. the grade of zero on an assignment, loss of credit with a notation on the transcript (notation reads: "Grade of F assigned for academic dishonesty"), and/or suspension or expulsion from the university. For information on the various types of academic dishonesty please refer to the [Academic Integrity Policy](#), located at <https://secretariat.mcmaster.ca/university-policies-procedures-guidelines/>

The following illustrates only three forms of academic dishonesty:

- plagiarism, e.g. the submission of work that is not one's own or for which other credit has been obtained.
- improper collaboration in group work.
- copying or using unauthorized aids in tests and examinations.

## Conduct Expectations

As a McMaster student, you have the right to experience, and the responsibility to demonstrate, respectful and dignified interactions within all of our living, learning and working communities. These expectations are described in the [Code of Student Rights & Responsibilities \(the "Code"\)](#). All students share the responsibility of maintaining a positive environment for the academic and personal growth of all McMaster community members, whether in person or online.

It is essential that students be mindful of their interactions online, as the Code remains in effect in virtual learning environments. The Code applies to any interactions that adversely affect, disrupt, or interfere with reasonable participation in University activities. Student disruptions or behaviours that interfere with university functions on online platforms (e.g. use of Avenue 2 Learn, WebEx or Zoom for delivery), will be taken very seriously and will be investigated. Outcomes may include restriction or removal of the involved students' access to these platforms

## Academic Accommodation of Students with Disabilities

Students who require academic accommodation must contact Student Accessibility Services (SAS) to make arrangements with a Program Coordinator. Academic accommodations must be arranged for each term of study. Student Accessibility Services can be contacted by phone 905-525-9140 ext. 28652 or e-mail [sas@mcmaster.ca](mailto:sas@mcmaster.ca). For further information, consult McMaster University's Policy for [Academic Accommodation of Students with Disabilities](#).

## Faculty of Social Sciences E-mail Communication Policy

Effective September 1, 2010, it is the policy of the Faculty of Social Sciences that all e-mail communication sent from students to instructors (including TAs), and from students to staff, must originate from the student's own McMaster University e-mail account. This policy protects confidentiality and confirms the identity of the student. It is the student's responsibility to ensure that communication is sent to the university from a McMaster account. If an instructor becomes aware that a communication has come from an alternate address, the instructor may not reply at their discretion.

## Course Modification

The instructor and university reserve the right to modify elements of the course during the term. The university may change the dates and deadlines for any or all courses in extreme circumstances. If either type of modification becomes necessary, reasonable notice and communication with the students will be given with explanation and the opportunity to comment on changes. It is the responsibility of the student to check their McMaster email and course websites weekly during the term and to note any changes.

## Extreme Circumstances

The University reserves the right to change the dates and deadlines for any or all courses in extreme circumstances (e.g., severe weather, labour disruptions, etc.). Changes will be communicated through regular McMaster communication channels, such as McMaster Daily News, A2L and/or McMaster email.