QUANTITATIVE POLITICAL & POLICY ANALYSIS
POLSCI 784
Term 2, Winter 2020

Instructor: Dr. Michelle L. Dion
Class: Thursday, 8:30 -11:20
Classroom: LRW 5001
Office: KTH 533
Office phone: 905-525-9140 ext. 24029
Skype: michellelynndion

Teaching Assistant: Rochelle Wijesingha
Office hours: TBD
Office: TBD
Email: rochelle.wijesingha@gmail.com

Office Hours:
- Tuesdays, 10:30am – 12:30pm, no appointment needed OR
- Other times by appointment. See me before or after class to book an appointment.

Contents

Course Description

This is an <u>introductory</u> graduate course in empirical research and statistical methods. 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 <u>all</u> 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 syllabus and a list of related web-based resources for the course can be found at the course websites:

- Course website for notes, discussion board, resources, etc.: https://github.com/polsci784.
  - 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.
- Curated collection of data, R, and related websites: https://bkmrk.michelledion.com/

Course Objectives

The course will provide an introduction to basic 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, you will complete a series of homework assignments and an independent research project employing linear regression. Student attendance and participation in class is required and constitute a significant portion of final grades.

Required Texts and Software

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 https://www.indiebound.org/indie-bookstore-finder, brick & mortar chain https://www.chapters.indigo.ca/en-ca/, or online https://www.amazon.ca/).

Statistical SAGE Little Green Books (and other SAGE textbooks)

Lewis-Beck, Michael. 1995. *Data Analysis*. Thousand Oaks, CA: SAGE. (hard copy on reserve, soft copy via: http://srmo.sagepub.com.libaccess.lib.mcmaster.ca/view/data-analysis/SAGE.xml or 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. (soft copy via: http://srmo.sagepub.com.libaccess.lib.mcmaster.ca/view/understanding-regression-analysis/SAGE.xml or https://dx.doi.org/10.4135/9781412986410).

Lewis-Beck, Colin, and Michael Lewis-Beck. 2016. *Applied Regression: An Introduction*. Thousand Oaks: SAGE. (soft copy via: http://methods.sagepub.com.libaccess.lib.mcmaster.ca/book/applied-regression-an-introduction-second-edition or https://doi.org/10.4135/9781483396774).

Healy, Kieran. 2018. *Data Visualization: A Practical Introduction*. 1st ed. Princeton, NJ: Princeton University Press. Online version as of 2018-04-25. https://socviz.co/index.html

Fox, John. 1991. *Regression Diagnostics*. Thousand Oaks, CA: SAGE. http://srmo.sagepub.com.libaccess.lib.mcmaster.ca/view/regression-diagnostics/SAGE.xml or https://dx.doi.org/10.4135/9781412985604.

Hardy, Melissa A. 1993. *Regression with Dummy Variables*. Thousand Oaks: SAGE. http://srmo.sagepub.com.libaccess.lib.mcmaster.ca/view/regression-with-dummy-variables/SAGE.xml or https://dx.doi.org/10.4135/9781412985628.

Kahane, Leo H. 2008. *Regression Basics*. 2nd ed. Thousand Oaks: SAGE. http://methods.sagepub.com.libaccess.lib.mcmaster.ca/book/regression-basics or https://dx.doi.org/10.4135/9781483385662.

Menard, Scott. 2002. *Applied Logistic Regression Analysis*. 2nd ed. Thousand Oaks: SAGE. http://methods.sagepub.com.libaccess.lib.mcmaster.ca/book/applied-logistic-regression-analysis or https://dx.doi.org/10.4135/9781412983433

Reference or alternative 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)

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. 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 (https://www.tidyverse.org/) 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. Swirl (https://swirlstats.com/) is a package that teaches R commands through an interactive interface. There are swirl modules to learn the tidy way to analyze data in R (see below).

RStudio also integrates with GitHub.com 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 homework 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.

R textbooks & websites
- Grolemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. https://r4ds.had.co.nz/. Also available in hard copy.
  - Solutions manual for *R for Data Science*: Arnold, Jeffrey B. n.d. *R for Data Science: Exercise Solutions*. http://jrnold.github.io/r4ds-exercise-solutions.
- IDRE. n.d. "R." Statistical Computing. https://stats.idre.ucla.edu/r/.
- MD's collection of R resources: https://bkmrk.michelledion.com/r/
- RStudio cheat sheets for various R and tidyverse packages: https://rstudio.com/resources/cheatsheets/
- Phillips, Nathaniel D. n.d. *YaRrr! The Pirate's Guide to R*. https://bookdown.org/ndphillips/YaRrr/. (solutions included in Chapter 18)
- Healy, Kieren. 2018. "Appendix". In *Data Visualization*. Princeton: Princeton UP. https://socviz.co/index.html. Also available in hard copy.
- McConville, Chester Ismay and Albert Y. Kim Foreword by Kelly S. 2019. *Statistical Inference via Data Science*. New York: CRC Press. https://moderndive.com/. Also available in hard copy.

Statistics reference books 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 or http://sk.sagepub.com.libaccess.lib.mcmaster.ca/reference/socialscience .
- SAGE Little Green Books: https://methods-sagepub-com.libaccess.lib.mcmaster.ca/Search/Results?booktypes=qass&IncludeSegments=false&IncludeParts=false&sort=pubyear&keywords[0].Field=FullText&keywords[0].Text=&noback=true.
- NB: SAGE also has a Little Blue Book series for qualitative methods.

swirl interactive learning exercises
Swirl (https://swirlstats.com/) is an interactive R package that walks users through fundamental concepts and basic commands of the R Statistical Computing language. In addition to the original core swirl modules that cover basic data types and related concepts in R, several new collections of modules are interactive learning tools for probability and statistical inference, data cleaning using the tidyverse, linear regression, and data visualization using ggplot. Swirl modules vary in length and are not a substitute for class or homework exercises. Instead, they are useful reference tools for those who want to develop more confidence and intuition using R.

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 homework 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://www.mcmaster.ca/uts/linkedinlearning/

Course Evaluation

Class participation, attendance, and pop quizzes, 15%. To get the full benefit of our class meetings and to be able to participate actively, you must have done the content reading prior to class and you must attend class regularly. (Indeed, the norm in graduate school is that you attend every class.) Asking thoughtful or insightful questions is just as important as answering questions posed by others in the class. Pop quizzes during class on major concepts may be used as needed. Participation through the class discussion thread will also count positively toward your participation grade. This includes asking questions and posting answers to others' questions. Absences, tardiness, cell phone disruptions, and distracting computer use during class will adversely affect your participation grade. Turn off digital notifications and restrict social media/communication use to official class breaks.

Homework assignments, 25%. Homework assignments are designed to build skills and develop confidence and competence with collaborative assignments that are marked based on completeness and effort. Students are encouraged to discuss homework assignments and work together to solve coding bugs and confirm understanding of course content. However, each student must turn in their own homework file that includes all output and answers in their own words. Since some assignments require more effort and time than others, the relative weights of each homework assignment vary.

The TA will <u>correct</u> assignments, identifying corrections in understanding or code. I will then assign <u>marks</u> based on completeness / effort, rather than accuracy. This approach is meant to provide feedback to students, help me quickly identify material that requires collective review, and minimize everyone's worry about marks in order to facilitate student engagement and learning. However, this approach to marking requires that students complete assignments according to the course schedule, and <u>no late homework assignments will be accepted</u>.

Final research project, 60% total. 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 logistics) regression. The project will proceed in phases to give you guidance and feedback throughout the research process. The final product of your research project will be presented in a poster session the final week of classes. Other graduate students and faculty will be invited to attend. Students are strongly encouraged to discuss their research projects with me early and often to make sure the projects meet the assignment's requirements and are feasible. Please review the Project Instructions for detailed guidance for each of the assignments below.

Your final research project will proceed in phases:
1. Statement of research question with clear identification of dependent variable (5%)
2. Description of research hypotheses and bibliography (10%)
3. Diagram of research design (5%)
4. Description of data and sources bibliography (10%)
5. Description of analysis and results (10%)
6. Final poster with results (20%)

Course Policies

Submission of Assignments
All assignments will be submitted online at our private organization on GitHub.com: https://github.com/polsci784.

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

MARK	GRADE
90-100	A+

MARK	GRADE
85-90	A
80-84	A-
77-79	B+
73-76	B
70-72	B-
69-0	F

Late Project Assignments

- No late <u>homework</u> assignments are accepted.
- If you cannot submit a <u>project assignment</u> before its deadline, it will be accepted up to 48 hours after its original due date and time. When submitted after the deadline but within the first 48 hours, the <u>project assignment</u> will be worth 90% of the maximum assignment value.
- Assignments will be marked in the order that they are received. Assignments submitted after the original deadline may not be marked before the next assignment is due.
- If you anticipate needing more than 48 hours past a project assignment deadline to complete it, you must communicate with me during office hours or an appointment to discuss your situation and identify an alternative submission schedule for your project going forward.

Collaboration

By submitting written homework assignments and your final project, you are pledging that you have not received unauthorized aid on the assignments and project. While you may discuss homework assignments with other students, you must generate your own output and write up your own answers. If computer analysis is required for an assignment, you must analyze your own data separately from your peers. While you are encouraged to discuss your projects with peers and the instructor, you must be the only author of your written assignments. This means that though you may discuss an assignment with peers, the write-up should be done alone and separate from them. Meet to discuss the assignment, then go your separate ways to write up your answers. All references to or paraphrasing of course readings or outside readings must be properly documented to avoid plagiarism. If you have any doubts, please ask me before turning in the assignment.

Absences, Missed Work, Illness

Regular attendance is crucial to your success in this course and is expected of all graduate students. Attendance is incorporated into your participation grade, which is a substantial portion of your final grade (15%). 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.

GitHub.com & GitHub Classroom

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 TA and instructor. Students should be aware that, when they access the electronic components of this course, private information such as first and last names and email addresses for McMaster e-mail accounts may be apparent to other students. The available information is dependent on the technology used. Continuation in this course will be deemed consent to this disclosure. If you have any questions or concerns about such disclosure please discuss this with the course instructor.

<u>University Policies</u>

Academic Integrity Statement

You are expected to exhibit honesty and use ethical behavior in all aspects of the learning process. Academic credentials you earn are rooted in principles of honesty and academic integrity.

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 behavior 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.

It is your responsibility to understand what constitutes academic dishonesty. For information on the various types of academic dishonesty please refer to the Academic Integrity Policy, located at www.mcmaster.ca/academicintegrity.

The following illustrates only three forms of academic dishonesty:
1. Plagiarism, e.g. the submission of work that is not one's own or for which credit has been obtained.
2. Improper collaboration in group work.
3. Copying or using unauthorized aids in tests and examinations.

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. 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.

Weekly Course Schedule

January 9	Introduction
Due:	Install R, RStudio, swirl, git & create GitHub.com educational account

Content 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. https://thepoliticalmethodologist.files.wordpress.com/2013/09/tpm_v10_n21.pdf Lewis-Beck, Michael S. 1995. *Data Analysis: An Introduction*. Thousand Oaks: SAGE, https://dx.doi.org/10.4135/9781412983846, pgs. 1-8. Zigerell, L. J. 2011. "Of Publishable Quality: Ideas for Political Science Seminar Papers." *PS: Political Science & Politics* 44 (3): 629–33. https://doi.org/10.1017/S1049096511000734. 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. https://doi.org/10.2307/1519895. • 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. https://doi.org/10.1017/S0022216X08004380.
R resources:	*Dion, Michelle L. "Install R & RStudio on your laptop or desktop." https://github.com/polsci784/outline-notes-handouts/blob/master/install_01_R-RStudio.md *Dion, Michelle L. "Install swirl & explore R." https://github.com/polsci784/outline-notes-handouts/blob/master/install_02_swirl.md *Dion, Michelle L. "Install TeX & git." https://github.com/polsci784/outline-notes-handouts/blob/master/install_03_TeX-git.md * These are part of our "private" class repository. You can only see these after you join our class team. You will receive an email invitation from GitHub.com. Grolemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. https://r4ds.had.co.nz/. Chapter 1-2 Healy, Kieren. 2018. *Data Visualization*. Princeton: Princeton UP. https://socviz.co/index.html. Chapter 2 Github Education Classroom: https://www.youtube.com/watch?v=ChA_zph7aao. Phillips, Nathaniel D. n.d. *YaRrr! The Pirate's Guide to R*. https://bookdown.org/ndphillips/YaRrr/. Chapters 1-2.
swirl:	Swirl 1-2
LinkedIn:	Navigating the RStudio Environment: Learning R. n.d. https://www.linkedin.com/learning/learning-r-2/navigating-the-rstudio-environment. Why Is the Tidyverse Unique?: R Programming in Data Science: Setup and Start. n.d. https://www.linkedin.com/learning/r-programming-in-data-science-setup-and-start/why-is-the-tidyverse-unique. Swirl: R for Data Science: Lunchbreak Lessons. n.d. https://www.linkedin.com/learning/r-for-data-science-lunchbreak-lessons/swirl. Packages for R: Learning R. n.d. https://www.linkedin.com/learning/learning-r-2/packages-for-r. The Tidyverse: Learning R. n.d. https://www.linkedin.com/learning/learning-r-2/the-tidyverse.
January 16	**Univariate descriptive statistics**
Due:	• Homework 1 (1x) • Statement of research question
Content readings:	Lewis-Beck, Michael S. 1995. *Data Analysis: An Introduction*. Thousand Oaks: SAGE, https://dx.doi.org/10.4135/9781412983846, pgs. 9-18.
R resources:	Grolemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. https://r4ds.had.co.nz/. Chapters 4, 6, 8-9, 26-30. Phillips, Nathaniel D. n.d. *YaRrr! The Pirate's Guide to R*. https://bookdown.org/ndphillips/YaRrr/. Chapters 3-4, 9. Irizarry, Rafael A. n.d. *Introduction to Data Science*. https://rafalab.github.io/dsbook/git.html#rstudio-git. Chapter 39 Git and GitHub.
swirl:	• Swirl: 3-7 • Statistical inference: 1-3

LinkedIn:	How to Install the Tidyverse: R Programming in Data Science: Setup and Start. n.d. https://www.linkedin.com/learning/r-programming-in-data-science-setup-and-start/how-to-install-the-tidyverse. Importing Data from a Spreadsheet: Learning R. n.d. https://www.linkedin.com/learning/learning-r-2/importing-data-from-a-spreadsheet. What Are CSV Files?: Data Wrangling in R. n.d. https://www.linkedin.com/learning/data-wrangling-in-r/what-are-csv-files. Importing CSV Files into R: Data Wrangling in R. n.d. https://www.linkedin.com/learning/data-wrangling-in-r/importing-csv-files-into-r. Importing Excel Files into R: Data Wrangling in R. n.d. https://www.linkedin.com/learning/data-wrangling-in-r/importing-excel-files-into-r. Loading Data Sets with Read_csv: Data Visualization in R with Ggplot2. n.d. https://www.linkedin.com/learning/data-visualization-in-r-with-ggplot2/loading-data-sets-with-read-csv. Recoding Variables: Learning R. n.d. https://www.linkedin.com/learning/learning-r-2/recoding-variables. Computing New Variables: Learning R. n.d. https://www.linkedin.com/learning/learning-r-2/computing-new-variables. Computing Descriptives: Learning R. n.d. https://www.linkedin.com/learning/learning-r-2/computing-descriptives. Computing Frequencies: Learning R. n.d. https://www.linkedin.com/learning/learning-r-2/computing-frequencies.
January 23	**Hypothesis testing and statistical significance**
Due:	[Make sure your RStudio & GitHub workflow is set up before class]
Content readings:	Lewis-Beck, Michael S. 1995. *Data Analysis: An Introduction*. Thousand Oaks: SAGE, https://dx.doi.org/10.4135/9781412983846, pgs. 31-41.
R resources:	Grolemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. https://r4ds.had.co.nz/. Chapters 3, 5, 7.
swirl:	• Getting and Cleaning Data: 1-3 • Statistical inference: 4-11
LinkedIn:	Using the Tidyverse: Data Wrangling in R. n.d. https://www.linkedin.com/learning/data-wrangling-in-r/using-the-tidyverse. Variables, Observations, and Values: Data Wrangling in R. n.d. https://www.linkedin.com/learning/data-wrangling-in-r/variables-observations-and-values. What Is Tidy Data?: Data Wrangling in R. n.d. https://www.linkedin.com/learning/data-wrangling-in-r/what-is-tidy-data. Subsetting Tibbles: Data Wrangling in R. n.d. https://www.linkedin.com/learning/data-wrangling-in-r/subsetting-tibbles. Filtering Tibbles: Data Wrangling in R. n.d. https://www.linkedin.com/learning/data-wrangling-in-r/filtering-tibbles. R Data Types: Basic Types: R for Data Science: Lunchbreak Lessons. n.d. https://www.linkedin.com/learning/r-for-data-science-lunchbreak-lessons/r-data-types-basic-types. R Data Types: Data Frame: R for Data Science: Lunchbreak Lessons. n.d. https://www.linkedin.com/learning/r-for-data-science-lunchbreak-lessons/r-data-types-data-frame. R Data Types: Factor: R for Data Science: Lunchbreak Lessons. n.d. https://www.linkedin.com/learning/r-for-data-science-lunchbreak-lessons/r-data-types-factor. Piping Commands with %>%: Learning R. n.d. https://www.linkedin.com/learning/learning-r-2/2929003. Selecting Cases and Subgroups: Learning R. n.d. https://www.linkedin.com/learning/learning-r-2/selecting-cases-and-subgroups.
January 30	**Exploratory data visualization**
Due:	[should have draft started of homework 2 to ask questions in class about code / content]

Content readings:	Healy, Kieran, and James Moody. 2014. "Data Visualization in Sociology." *Annual Review of Sociology* 40 (1): 105–28. https://doi.org/10.1146/annurev-soc-071312-145551.
R resources:	Healy, Kieren. 2018. *Data Visualization.* Princeton: Princeton UP. https://socviz.co/index.html. Chapters 1, 3, 4. Grolemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. https://r4ds.had.co.nz/. Chapter 10-12, 14-15. Phillips, Nathaniel D. n.d. *YaRrr! The Pirate's Guide to R*. https://bookdown.org/ndphillips/YaRrr/. Chapter 11.
swirl:	• Swirl: 8-9,12 • Exploratory Data Analysis: 1-4, 7-10
LinkedIn:	Introducing Ggplot2: Data Visualization in R with Ggplot2. n.d. https://www.linkedin.com/learning/data-visualization-in-r-with-ggplot2/introducing-ggplot2. Barplot: R for Data Science: Lunchbreak Lessons. n.d. https://www.linkedin.com/learning/r-for-data-science-lunchbreak-lessons/barplot. Bars and Columns: Data Visualization in R with Ggplot2. n.d. https://www.linkedin.com/learning/data-visualization-in-r-with-ggplot2/bars-and-columns. Dotchart: R for Data Science: Lunchbreak Lessons. n.d. https://www.linkedin.com/learning/r-for-data-science-lunchbreak-lessons/dotchart. Histogram: R for Data Science: Lunchbreak Lessons. n.d. https://www.linkedin.com/learning/r-for-data-science-lunchbreak-lessons/histogram. Histograms: Data Visualization in R with Ggplot2. n.d. https://www.linkedin.com/learning/data-visualization-in-r-with-ggplot2/histograms. Scatterplots: Data Visualization in R with Ggplot2. n.d. https://www.linkedin.com/learning/data-visualization-in-r-with-ggplot2/scatterplots.
February 6	**Association & simple, bivariate regression**
Due:	Homework 2 (2x)
Content readings:	Schroeder, Larry, David Sjoquist, and Paula Stephan. 1986. *Understanding Regression Analysis*. Thousand Oaks: SAGE, https://doi.org/10.4135/9781412986410, pgs. 12-29. Lewis-Beck, Michael S. 1995. *Data Analysis: An Introduction*. Thousand Oaks: SAGE, https://dx.doi.org/10.4135/9781412983846, pgs. 19-30, 42-53. Lewis-Beck, Colin, and Michael Lewis-Beck. 2016. *Applied Regression: An Introduction*. Thousand Oaks: SAGE. https://doi.org/10.4135/9781483396774. pgs. 1-22. Kahane, Leo. 2008. *Regression Basics*. 2nd ed. Thousand Oaks: SAGE. https://doi.org/10.4135/9781483385662. pgs. 1-16.
R resources:	Phillips, Nathaniel D. n.d. *YaRrr! The Pirate's Guide to R*. https://bookdown.org/ndphillips/YaRrr/. Chapters 13-15.4. Grolemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. https://r4ds.had.co.nz/. Chapter 17-18, 22-23. McConville, Chester Ismay and Albert Y. Kim Foreword by Kelly S. 2019. *Statistical Inference via Data Science*. New York: CRC Press. https://moderndive.com/. Chapter 5.
swirl:	Regression models: 1-4
LinkedIn:	Computing a Linear Regression: Learning R. n.d. https://www.linkedin.com/learning/learning-r-2/computing-a-linear-regression Computing Contingency Tables: Learning R. n.d. https://www.linkedin.com/learning/learning-r-2/computing-contingency-tables. Computing Correlations: Learning R. n.d. https://www.linkedin.com/learning/learning-r-2/computing-correlations.
February 13	**Regression assumptions and statistical inference (intro to multivariate)**
Due:	Literature review

Content readings:	Lewis-Beck, Michael S. 1995. *Data Analysis: An Introduction*. Thousand Oaks: SAGE, https://dx.doi.org/10.4135/9781412983846, pgs. 54-74. Lewis-Beck, Colin, and Michael Lewis-Beck. 2016. *Applied Regression: An Introduction*. Thousand Oaks: SAGE. https://doi.org/10.4135/9781483396774. pgs. 23-54. Schroeder, Larry, David Sjoquist, and Paula Stephan. 1986. *Understanding Regression Analysis*. Thousand Oaks: SAGE, https://doi.org/10.4135/9781412986410, pgs. 30-36, 37-53. Kahane, Leo. 2008. *Regression Basics*. 2nd ed. Thousand Oaks: SAGE. https://doi.org/10.4135/9781483385662. pgs. 17-58.
R resources:	Healy, Kieren. 2018. *Data Visualization*. Princeton: Princeton UP. https://socviz.co/index.html. Chapter 6.
swirl:	Regression models: 5
February 20	**Mid-term recess, NO CLASS**
February 27	**Multiple, multivariate regression**
Due:	Homework 3 (2x)
Content readings:	Lewis-Beck, Colin, and Michael Lewis-Beck. 2016. *Applied Regression: An Introduction*. Thousand Oaks: SAGE. https://doi.org/10.4135/9781483396774. Pages 55-74. Kahane, Leo. 2008. *Regression Basics*. 2nd ed. Thousand Oaks: SAGE. https://doi.org/10.4135/9781483385662. pgs. 59-78.
R resources:	Grolemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. https://r4ds.had.co.nz/. Chapters 24-25. McConville, Chester Ismay and Albert Y. Kim Foreword by Kelly S. 2019. *Statistical Inference via Data Science*. New York: CRC Press. https://moderndive.com/. Chapter 6.
swirl:	Regression models: 5 (repeat), 9
March 5	**Categorical independent variables, non-linear relationships, & interactions**
Due:	Research design
Content readings:	Lewis-Beck, Colin, and Michael Lewis-Beck. 2016. *Applied Regression: An Introduction*. Thousand Oaks: SAGE. https://doi.org/10.4135/9781483396774. pgs. 75-96. Schroeder, Larry, David Sjoquist, and Paula Stephan. 1986. *Understanding Regression Analysis*. Thousand Oaks: SAGE, https://doi.org/10.4135/9781412986410, pgs. 54-65. Kahane, Leo. 2008. *Regression Basics*. 2nd ed. Thousand Oaks: SAGE. https://doi.org/10.4135/9781483385662. pgs. 79-102. Hardy, Melissa A. 1993. *Regression with Dummy Variables*. Thousand Oaks: SAGE. https://doi.org/10.4135/9781412985628. pgs. 18-29. Fox, John. 1991. *Regression Diagnostics*. Thousand Oaks: SAGE. https://doi.org/10.4135/9781412985604. pgs. 53-65.
R resources:	Grolemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. https://r4ds.had.co.nz/. Chapter 15, 23.4.
March 12	**Outliers and predicted outcomes**
Due:	[Come to class with questions about HW 4]
Content readings:	Lewis-Beck, Michael S. 1995. *Data Analysis: An Introduction*. Thousand Oaks: SAGE, https://dx.doi.org/10.4135/9781412983846, pgs. 54-72. Lewis-Beck, Colin, and Michael Lewis-Beck. 2016. *Applied Regression: An Introduction*. Thousand Oaks: SAGE. https://doi.org/10.4135/9781483396774. pgs. 55-74. Schroeder, Larry, David Sjoquist, and Paula Stephan. 1986. *Understanding Regression Analysis*. Thousand Oaks: SAGE, https://doi.org/10.4135/9781412986410, pgs. 54-65. (repeat) Fox, John. 1991. *Regression Diagnostics*. Thousand Oaks: SAGE. https://doi.org/10.4135/9781412985604. pgs. 21-40.
R resources:	Grolemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. https://r4ds.had.co.nz/. Chapter 23.3 Healy, Kieren. 2018. *Data Visualization*. Princeton: Princeton UP. https://socviz.co/index.html. Chapter 5-6.

LinkedIn:	Lines and Smoothers: Data Visualization in R with Ggplot2. n.d. https://www.linkedin.com/learning/data-visualization-in-r-with-ggplot2/lines-and-smoothers. Plot to File: R for Data Science: Lunchbreak Lessons. n.d. https://www.linkedin.com/learning/r-for-data-science-lunchbreak-lessons/plot-to-file.
March 19	**Collinearity, F-tests, adjusted R^2, & model specification**
Due:	Homework 4 (x2)
Content readings:	Schroeder, Larry, David Sjoquist, and Paula Stephan. 1986. *Understanding Regression Analysis*. Thousand Oaks: SAGE, https://doi.org/10.4135/9781412986410, pgs. 66-80. Fox, John. 1991. *Regression Diagnostics*. Thousand Oaks: SAGE. https://doi.org/10.4135/9781412985604. pgs. 1-20. Kahane, Leo. 2008. *Regression Basics*. 2nd ed. Thousand Oaks: SAGE. https://doi.org/10.4135/9781483385662. pgs. 119-142.
R resources:	Grolemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. https://r4ds.had.co.nz/. Chapters 24-25.
swirl:	Regression models: 10-11
March 26	**Model fitting & assumptions**
Due:	Results
Content readings:	Fox, John. 1991. *Regression Diagnostics*. Thousand Oaks: SAGE. https://doi.org/10.4135/9781412985604. 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. https://doi.org/10.1080/07388940500339167. Schrodt, Philip A. 2014. "Seven Deadly Sins of Contemporary Quantitative Political Analysis." *Journal of Peace Research* 51 (2): 287–300. https://doi.org/10.1177/0022343313499597.
R resources:	Grolemund, Garrett, and Hadley Wickham. n.d. *R for Data Science*. O'Reilley. https://r4ds.had.co.nz/. Chapter 25.
April 2	**Logistic regression & other advanced models**
Due:	Homework 5 (2x)
Content readings:	Kahane, Leo. 2008. *Regression Basics*. 2nd ed. Thousand Oaks: SAGE. https://doi.org/10.4135/9781483385662. pgs. 143-46. Menard, Scott. 2002. *Applied Logistic Regression Analysis*. 2nd ed. Thousand Oaks: SAGE. https://doi.org/10.4135/9781412983433. pgs. 1-67.
R resources:	Phillips, Nathaniel D. n.d. *YaRrr! The Pirate's Guide to R*. https://bookdown.org/ndphillips/YaRrr/. Chapter 15.5. Healy, Kieren. 2018. D*ata Visualization*. Princeton: Princeton UP. https://socviz.co/index.html. Chapter 8.
swirl:	Regression models: 11-12
LinkedIn:	Using Colors in R: Learning R. n.d. https://www.linkedin.com/learning/learning-r-2/using-colors-in-r. [And next 4 modules are all about customizing your figures]
April 9	**Poster session**
Due:	poster