Sociology
UNIVERSITY OF TORONTO

SOC6302H Stats for Sociologists
(Fall 2024)

Instructor: Prof. Ethan Fosse (pronouns: he/him)
Interactive Workshop: Mondays (1pm - 3pm), Room 17020 in 700 University Avenue
Lab: Mondays (5pm - 6pm), Room 17020 in 700 University Avenue
Course Website: https://q.utoronto.ca/courses/353913
Office Hours: By appointment
Zoom Link: https://utoronto.zoom.us/my/ethanfosse
Instructor Email: ethan.fosse@utoronto.ca

Teaching Assistant (TA): Firdaous Sbai
TA Email: firdaous.sbai@mail.utoronto.ca

1   Course Description

This course introduces students to the rapidly developing field of data science[1] in sociology and related social sciences. It focuses on practical applications of data analysis, statistical computing, and visualization techniques across diverse data types. Students will learn to work with categorical, numerical, textual, relational, and spatial data, and will gain hands-on experience with real-world datasets. The course covers basic concepts such as descriptive statistics, correlation analysis, and linear regression, as well as more advanced topics such as sampling theory and causal inference. Emphasis is placed on developing critical thinking skills for evaluating research claims and conducting independent analyses. Through a combination of lectures, interactive workshops, and hands-on labs, students will gain proficiency in using programming for data manipulation, analysis, and visualization. By the end of the course, students will have acquired a solid foundation in data science methods, equipping them with valuable tools for their future research and professional endeavors in the social sciences.

2   Course Goals

The science fiction writer Arthur C. Clarke famously stated: "Any sufficiently advanced technology is indistinguishable from magic." Developed over the past 100 or so years, modern data science is arguably best thought of as a kind of advanced technology, empowering anyone who wishes to use it the ability to explore, interpret, and explain the world around us in new, remarkable ways. It thus follows that the magic of data science, unlike forms of magic commonly depicted in fiction, is **real** and well within your capabilities to understand and use for your needs.

This course is designed to demystify data science and equip you with the skills to harness the magic of data science in sociological research. To achieve this, the course is structured around three main goals:

---

[1]In this course, we will treat "statistics" as interchangeable with "data science," with "statistical inference" referring to the goal of making a claim about a population given a sample of that population. For a similar view, see the reading by the statistician Jeff Wu in Week 1.

1. *To inspire an appreciation for the power and potential of data science in sociology.* We will explore how data science can unlock insights and answer pressing questions about social reality. Through exposure to diverse applications across sociological subfields, you'll discover the transformative impact of these tools on our understanding of society.

2. *To develop your foundational skills in data science methods and tools.* You will learn to organize, visualize, and describe data; use regression analysis; understand sampling theory; and think more clearly about how to pose and answer causal questions. Through hands-on experience analyzing real-world data with statistical software, you will build a strong practical skill set that will help you answer important sociological questions.

3. *To cultivate your ability to communicate data-driven insights effectively.* As important as drawing insights from data is the ability to convey your findings to others. You will learn to present results through engaging oral presentations, compelling visualizations, and clear, concise writing. By examining how various social scientists communicate their work, you will develop the skills to translate complex data into accessible and impactful sociological narratives.

3   Course Content

The course is structured around two main parts, which can be likened to the roots and lower trunk of a tree (see Figure 1).[2] The first part focuses on the foundation of data science: understanding and working with the main types of data used by sociologists. This includes categorical, numerical, textual, relational (or network-based), and spatial data. In this part of the course, you will learn how to organize, describe, and visualize these different data types, forming the roots and soil that nourish the tree of data science.

Building upon this foundation, the second part of the course examines the core tool of the social sciences: regression analysis. Regression analysis serves as the trunk of the tree, providing stability and support for the various branches of data science. As you learn about regression analysis, you will also explore key concepts such as relative risk and odds ratios, correlation analysis, and the basics of sampling theory. Near the end of the course you will also encounter a gentle introduction to causal inference, represented by one of the branches on the tree. This introduction will provide you with a glimpse into the wider range of data science methods and their applications in the social sciences.

It is important to note that the tree analogy in Figure 1 is simply a visual representation and should not be taken too literally. The branches included in Figure 1 are not exhaustive, and other sociologists may present different branches based on their expertise and focus. For a more comprehensive overview of methods beyond regression that hint at some of these other branches, you might find it useful to explore the CRAN Task Views at `https://cran.r-project.org/web/views/`.

---

[2]This visual concept was inspired by discussions with Prof. Christopher Winship.

Figure 1: The Tree of Data Science
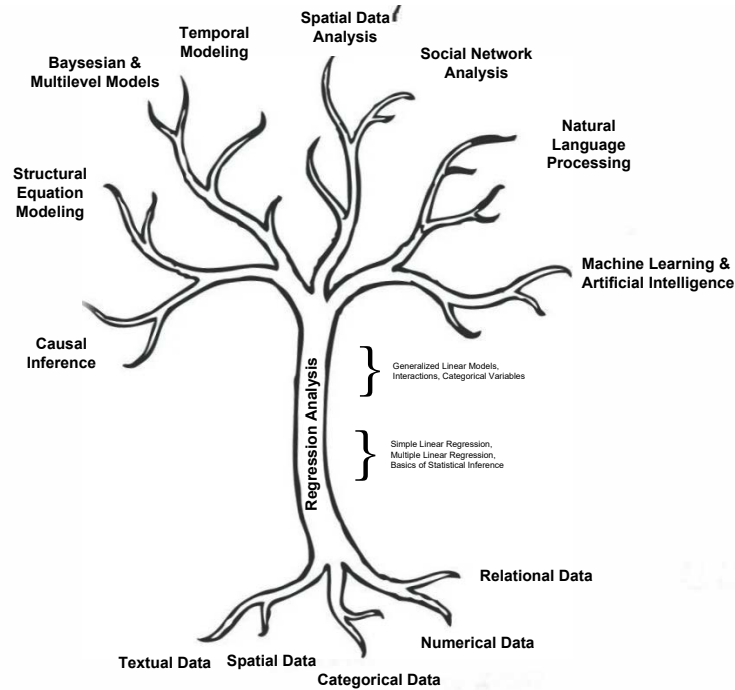
Notes: This diagram serves as a conceptual framework for understanding the structure and progression of topics in this course and more advanced courses. While it illustrates key relationships between foundational concepts and more advanced applications, it is not an exhaustive representation of the various methods available.

4   Course Philosophy & Principles

This course is built on six key principles:

1. *Data science is not just mathematics.* While mathematics is an essential tool for data scientists, it is not the only (or even the most important) skill or tool. Data scientists also rely on computer programming, graphic design, substantive expertise, and creativity, among other things. Mathematics can clarify concepts, but often the biggest issues and problems are conceptual in nature. For example, understanding the limitations of a dataset or the ethical implications of a study requires more than just mathematical knowledge.

2. *Substantive questions drive methodological choices.* As sociologists, many of us are more interested in the methods as powerful tools to answer interesting questions rather than in the methods themselves. To reflect this, each section of the course focuses on a single real-world dataset with an overarching question, and the tools for answering this question follow. (However, if you find the methods themselves fascinating independent of the subject matter, perhaps as interesting "puzzles" or "games," that is perfectly fine as well.)

3. *Multiple approaches facilitate learning.* For each tool or concept, I present various approaches, including analogies, mathematical formulas, verbal descriptions, visualizations, and practical examples based on coding and examining real data. I encourage you to explore multiple ways of actively learning the key concepts and tools covered in this course. However, from my experience most people learn best by doing, which is why coding and exploring data is so important.

4. *Playing with data is exciting and insightful.* Exploring data is a form of play that can be both fun and intellectually stimulating. Rather than viewing data analysis as a mundane task, I strongly encourage you to embrace the idea of "playing with data" instead of merely "working with data." When we approach data with a playful attitude, we open ourselves up to unexpected discoveries. We become more willing to explore unconventional paths, ask "what if" questions, and test unusual hypotheses. This exploratory mindset can lead to uncovering hidden patterns, revealing surprising relationships, and generating novel insights that might otherwise remain hidden.

5. *Computational techniques enhance understanding.* When appropriate, we use computational techniques to understand the concepts used in this course. For example, when introducing the concept of statistical inference, we use code to repeatedly draw samples from a population. This approach mirrors the mathematics of statistical inference but often provides a clearer understanding for most students. By seeing the process in action, students can better grasp the underlying principles.

6. *No prior knowledge is assumed.* This course assumes that you know virtually nothing about data science, statistics, or mathematics. We start from first principles and build upward, focusing on the most important concepts. Whether you are a complete beginner or have some experience, the course is designed to provide a solid foundation in data science for those with a sociological mindset.

5   Course Design

This course consists of three components: an online lecture delivered asynchoronously, a two-hour in-person interactive workshop, and a one-hour lab. These components are held most weeks during the semester. You are required to watch the online lecture before each class, and attendance is required for both the interactive workshop and the lab.

1. **Online Lecture**: Before most classes, there will be a lecture delivered asynchronously via Zoom and uploaded to Quercus. You may watch the online lecture live, but you are not required to do so. However, you are required to watch each lecture before class.

2. **Interactive Workshop**: The in-person interactive workshop consists of three parts. In the first part, I will ask and solicit questions about the week's material, focusing on that which is covered in the slides and required readings. In the second part, we will cover the R code and dataset. Make sure to bring a laptop computer, as you are expected to interact and program along. If you do not have access to a laptop, please contact me and we will work on a way to provide you with a laptop. **Note that you are expected to ask questions during the interactive workshop. If participation is low then I will require each student to ask at least three questions during each workshop.** In the third and final part of the interactive workshop we will have a short group activity that helps us apply the ideas and tools we have covered that week.

3. **Lab**: Besides the online lecture and in-person class, we also have a lab (or tutorial). This is a safe, supportive environment in which you are expected to work on the mission for that week. The lab begins with a short (approximately 5-10 minutes) overview of the week's mission. Although not required, you are encouraged to work

collaboratively. The TA will actively provide help on the mission, giving key tips and advice during the lab. If for whatever reason you have already completed the week's mission, then you are expected to help other students during the lab.

6   Course Resources

Quercus Website: All datasets, scripts, lecture slides, readings, and other resources are available on the course website. The course website will also have additional information on the course mechanics and specific due dates for the assignments. Note that all assignments will be submitted entirely through course website, which will be updated as the course begins. The course website may be accessed via the University of Toronto's Quercus platform.

Readings: You are required to thoroughly read and study the slides before each class. The slides are essentially the core textbook. To supplement the slides, each week we have two additional sets of readings: **required** and **supplemental**. The required readings reiterate the methods covered in the slides, often providing a more in-depth, technical discussion. By contrast, the supplemental readings are intended primarily to help you learn about how sociologists and other social scientists use data to answer questions about the world around us. For example, when we analyze data on passengers aboard the Titanic in Mission 1, one of the supplemental readings is an actual study whether or not there were gender and social class differences between those who survived the disaster and those who did not. Each week you are required to read **all** of the required readings and **at least one** of the supplemental readings. **All readings are available for free on the course website.**

Data Analysis Software: While reading and studying is helpful for understanding the tools used in this course, the philosophy of this course is that you will learn best by "doing." This means using software to analyze real data sets to answer real-world questions. Because of its popularity and applicability, this course will focus on using R with RStudio. R is the underlying programming language, while RStudio is a graphical user interface (GUI) that makes working with R much easier. Both are free, open-source, and used widely by statisticians. To install R with RStudio, go to the following link and click on the installer for your computer's operational system: `https://www.rstudio.com/products/rstudio/download/`

7   Evaluation Components

The course grade follows the standard criteria set by the university. For all students, your grade will be based on the following:

- Active **participation** in the interactive workshops and labs, which will count for 10% of your grade
- Completion of a **stats anxiety reflection memo** that will count for 5% of your grade
- Regularly-assigned **missions** that will count for 30% of your grade
- A **data memo** that will count for 5% of your grade
- A short **flash presentation** that will count for 10% of your grade
- A **final paper** that will count for 40% of your grade

Interactive Workshop and Lab Participation (10%**)**: Attendance is mandatory unless there are exigent circumstances. You must not only attend but actively participate in class workshops and labs throughout course.

Stats Anxiety Reflection Memo (5%**)**: You are required to write a short (approximately 1-3 page) reflection piece on your experiences, emotions, and/or anxieties on learning and/or using mathematics and statistics. The piece will be submitted on the course website. You will *not* be evaluated on the content of your memo, only on whether or not you completed it. Anyone who completes the memo will receive the full grade. Critical reflection is encouraged. Additional details on the stats anxiety reflection memo will be given during the first in-person interactive workshop.

Missions (30%**)**: Missions (or "problem sets") are assigned and submitted online through the course website. Each mission is organized around a key question and the analysis of a real dataset For each mission, you are encouraged to "role play" as a particular kind of sociologist or social scientist. For example, in Week 2 you are asked to role play as a historical sociologist using data from the Titanic disaster to examine who survived and who did not. Details on the missions will be published on the course website as the semester starts. Late missions are not accepted for any reason. Working with other students on the missions is allowed and encouraged but only as long as you hand in your own work and do not simply copy the work of someone else. For these missions, you will be graded on whether or not you have written the correct answer.

Data Memo (5%**)**: To help you prepare for the final paper, you will be required to submit a 1-3 page document that describes your dataset and preliminary research question. This memo should provide a brief description of your chosen dataset, including its source, sample size, and key variables. You should clearly state your preliminary research question and explain its sociological importance. The memo should also include a description of at least three key variables you plan to use. Additionally, discuss any potential limitations or challenges you foresee with the dataset or research question. Additional details on the data memo will be given near the beginning of the semester.

Flash Presentation (10%**)**: To help you prepare for your final paper, you will be required to give a short (approximately 5-7 minutes) presentation on your research project. In doing so, you will be required to submit the slides for your presentation to Quercus. Although there is no limit for the number of slides, it is generally recommended that you use between 3-10 slides. Details on the flash presentation will be given near the beginning of the semester.

Final Paper (40%**)**: The final paper will consist of a paper (no more than 20 pages) describing the analysis of a dataset using one or more of the tools or methods discussed in the course. You must use a dataset that is **not** used in the course. The final paper will be submitted online through the course website. Details on the final paper will be given near the beginning of the semester.

Instructions on Submitting Papers & Memos: All papers and memos must be submitted via the course website. **Only Adobe pdf documents will be allowed.** Papers and memos should be written using standard 12 point font, double spacing, and without a separate title page. All works in the papers should be referenced using American Sociological Association (ASA) style.

Late Penalty: Unless there is a documented reason beyond one's control (e.g., an illness or emergency), there will be a 5% deduction for each day the assignment is late.

8 Course Schedule

The **tentative** course schedule is shown in detail below (see also Table 1). It is subject to change throughout the semester depending on students' needs. More specific due dates and information for the assignments will be available on the course website.

Table 1: Summary Course Schedule

Date	Online Lecture & Interactive Workshop	Lab	Application	Week's Assignment
Sept. 9	Week 1: Introduction to Data			Stats Anxiety Reflection Memo
Sept. 16	Week 2: Categorical Data	Lab 1	Historical Sociology	Mission 1: Who Survived the Titanic Disaster?
Sept. 23	Week 3: Numerical Data	Lab 2	Social Psychology	Mission 2: What Predicts Social Status Among Kids?
Sept. 30	Week 4: Textual, Relational, and Spatial Data	Lab 3	Computational Social Science	Mission 3: What are the Partisan Differences in Defining Contested Concepts?
Oct. 7	Week 5: Relative Risk and Odds Ratios	Lab 4	Social Epidemiology	Mission 4: What are the Health Risks of Living in Poverty?
Oct. 14	**Thanksgiving Break**			
Oct. 21	Week 6: Correlation Analysis	Lab 5	Political Sociology	Mission 5: Why are Some States in the U.S. "Blue" vs. "Red"?
Oct. 28	Week 7: Linear Regression	Lab 6	Criminology	Mission 6: Why are Some Societies More "Violent" than Other Societies? Data Memo
Nov. 4	Week 8: Basics of Sampling and the Sampling Distribution	Lab 7	Demography	Mission 7: How Can We Use a Sample to Say Something About the Characteristics of a Population?
Nov. 11	Week 9: Inferences for Correlation and Linear Regression	Lab 8	Sociology of Health and Illness	Mission 8: Why are Some Societies "Happier" than Others?
Nov. 18	Week 10: Introduction to Multiple Linear Regression	Lab 9	Cultural Sociology	Mission 9: Does Socioeconomic Development Predict Changes in a Society's Culture?
Nov. 25	Week 11: Extending Regression with Categorical Variables and Interactions	Lab 10	Environmental Sociology	Mission 10: How Do Countries Vary in their "Ecological Footprint"?
Dec. 2	Week 12: Regression and Causal Inference	Lab 11		None
Dec. 9	Week 13: Flash Presentations			Final Project (Due Near End of Term)

Week 1: Introduction to Data

Before Class:

Make sure that you have R with RStudio installed. To install R with RStudio, go to the following link and click on the installer for your computer's operational system: `https://www.rstudio.com/products/rstudio/download/`

If you are using Windows 11, you can find more information on how to install R and Studio <u>here</u> on YouTube. If you are using a Mac, you can find more information on how to install R and Studio <u>here</u> on YouTube.

Required Readings:

Pp. 9-25 of Diez, David, Mine Çetinkaya-Rundel, and Christopher D. Barr. 2019. *OpenIntro Statistics.* Available online for free via Leanpub.

Assignments:

- Stats Anxiety Reflection Memo

Supplemental Readings (Pick One):

Fienberg, Stephen. 2014. "What is Statistics?" *Annual Review of Statistics and Its Application*, 1(1): 1-9.

Loukides, Mike. 2012. "What is Data Science?" Sebastopol, CA: O'Reilly Media.

Wu, Jeff. 1997. "Statistics = Data Science?" Presentation. Ann Arbor, Michigan: University of Michigan.

Week 2: Categorical Data

Before Class:

Watch online lecture titled "Module 2: Categorical Data" via Quercus. Download and explore the `Titanic.RData` dataset as well as the `Mission1.R` script. Study the required readings as well as the lecture slides, and read at least one of the supplemental readings.

Required Readings:

Pp. 61-66 of Diez, David, Mine Çetinkaya-Rundel, and Christopher D. Barr. 2019. *OpenIntro Statistics.* Available online for free via Leanpub.

Assignment(s):

- Mission 1: Who Survived the Titanic Disaster?
    - Your role: Imagine you are a historical sociologist examining the Titanic disaster

Supplemental Readings (Pick One):

Anderson, Margo. 2007. "Quantitative History." Pp. 248-64 in *The Sage Handbook of Social Science Methodology*, edited by William Outhwaite and Stephen P. Turner. London, UK: Sage Publications.

Hall, Wayne. 1986. "Social Class and Survival on the S.S. Titanic." *Social Science & Medicine* 22(6):687–90.

Mullins, Daniel Austin, Daniel Hoyer, Christina Collins, Thomas Currie, Kevin Feeney, Pieter François, Patrick E. Savage, Harvey Whitehouse, and Peter Turchin. 2018. "A Systematic Assessment of 'Axial Age' Proposals Using Global Comparative Historical Evidence." *American Sociological Review* 83(3):596-626.

Spinney, Laura. 2012. "History as Science." *Nature* 488(2): 24-26.

Week 3: Numerical Data

Before Class:

Watch online lecture titled "Week 3: Numerical Data" via Quercus. Download and explore the `PopularKids.RData` dataset as well as the `Mission2.R` script. Study the required readings as well as the lecture slides, and read at least one of the supplemental readings.

Assignment(s):

- Mission 2: What Predicts Social Status Among Kids?
  - Your role: Imagine you are a social psychologist studying social status among children

Required Readings:

Pp. 42-51 of Diez, David, Mine Çetinkaya-Rundel, and Christopher D. Barr. 2019. *OpenIntro Statistics.* Available online for free via Leanpub.

Supplemental Readings (Pick One):

Chase, Melissa A. and Gail M. Dummer. 1992. "The Role of Sports as a Social Status Determinant for Children." *Research Quarterly for Exercise and Sport* 63(4):418–24.

Chase, Melissa A., and Moe Machida. 2011. "The Role of Sport as a Social Status Determinant for Children: Thirty Years Later." *Research Quarterly for Exercise and Sport* 82(4):731-39.

Cillessen, Antonius H. N., and Amanda J. Rose. 2005. "Understanding Popularity in the Peer System." *Current Directions in Psychological Science* 14(2):102-5.

Goethals, George R. 2007. "A Century of Social Psychology: Individuals, Ideas, and Investigations." Pp. 3-23 in *The Sage Handbook of Social Psychology*, edited by Michael Hogg and Joel M. Cooper.

Week 4: Textual, Relational, and Spatial Data

Before Class:

Watch online lecture titled "Week 4: Textual, Spatial, and Relational Data" via Quercus. Download and explore the `ContestedPolitics.RData` dataset as well as the `Mission3.R` script. Study the required readings as well as the lecture slides, and read at least one of the supplemental readings.

Assignment(s):

- Mission 3: What are the Partisan Differences in Defining Contested Concepts?
  - Your role: Imagine you are a computational social scientist examining political discourse on "socialism" in the United States

Required Readings:

Pp. 189-236 of Imai, Kosuke. 2018. *Quantitative Social Science.* Princeton, NJ: Princeton University Press.

Supplemental Readings (Pick One):

Edelmann, Achim, Tom Wolff, Danielle Montagne, and Christopher A. Bail. 2020. "Computational Social Science and Sociology." *Annual Review of Sociology* 46(1):61-81.

Greenfield, Patricia M. 2013. "The Changing Psychology of Culture From 1800 Through 2000." *Psychological Science* 24(9):1722-31.

Pp. 13-32 and 48-59 of Grimmer, Justin, Margaret E. Roberts, and Brandom M. Stewart. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences.* Princeton, NJ: Princeton University Press.

Lazer, David, and Katya Ognyanova. 2024. "The Future of Computational Social Science" in *Oxford Handbook of Engaged Methodological Pluralism in Political Science (Volume 1)*, edited by Box-Steffensmeier, Janet M., Dino Christenson, and Valeria Sinclair-Chapman. Oxford, UK: Oxford University Press.

Macanovic, Ana. 2022. "Text Mining for Social Science: The State and the Future of Computational Text Analysis in Sociology." *Social Science Research* 108:102784.

Pp. 7-30 and 31-56 of Scott, John. 2012. *What is Social Network Analysis?* Bloomsbury Academic: New York, NY.

Week 5: Relative Risk and Odds Ratios

Before Class:

Watch online lecture titled "Week 5: Relative Risk and Odds Ratios" via Quercus. Download and explore the `ContestedPolitics.RData` datasets as well as the `Mission4.R` script. Study the required readings as well as the lecture slides, and read at least one of the supplemental readings.

Assignment(s):

- Mission 4: What are the Health Risks of Living in Poverty?
  - Your role: Imagine you are a social epidemiologist examining the relationship between poverty and health

Required Readings:

Pp. 79-108 of Diez, David, Mine Çetinkaya-Rundel, and Christopher D. Barr. 2019. *OpenIntro Statistics.* Available online for free via Leanpub.

Pp. 242-247 of Imai, Kosuke. 2018. *Quantitative Social Science.* Princeton, NJ: Princeton University Press.

Supplemental Readings (Pick One):

Andrade, Chittaranjan. 2015. "Understanding Relative Risk, Odds Ratio, and Related Terms: As Simple as It Can Get." *The Journal of Clinical Psychiatry* 76(7):21865.

Cummings, Jason L., and Pamela Braboy Jackson. 2008. "Race, Gender, and SES Disparities in Self-Assessed Health, 1974-2004." *Research on Aging* 30(2):137-67.

Krieger, Nancy. 2001. "A Glossary for Social Epidemiology." *Journal of Epidemiology and Community Health* 55: 693-700.

Marmot, Michael. 2005. "Social Determinants of Health Inequalities." *The Lancet* 365:1099-1104.

Oakes, J. Michael, and Kate N. Andrade. 2014. "Methodologic Innovations and Advances in Social Epidemiology." *Current Epidemiology Reports* 1(1):38-44.

Week 6: Correlation Analysis

Before Class:

Watch online lecture titled "Week 6: Correlation Analysis" via Quercus. Download and explore the `States.RData` datasets as well as the `Mission5.R` script. Study the required readings as well as the lecture slides, and read at least one of the supplemental readings.

Assignment(s):

- Mission 5: Why are Some States in the U.S. are "Blue" vs. "Red"?
  - Your role: Imagine you are a   political sociologist investigating geographic patterns of political behavior

Required Readings:

Pp. 79-108 of Diez, David, Mine Çetinkaya-Rundel, and Christopher D. Barr. 2019. *OpenIntro Statistics*. Available online for free via Leanpub.

Pp. 242-247 of Imai, Kosuke. 2018. *Quantitative Social Science*. Princeton, NJ: Princeton University Press.

Supplemental Readings (Pick One):

Feller, Avi, Andrew Gelman, and Boris Shor. 2013. "Red State/Blue State Divisions in the 2012 Presidential Election." *The Forum* 10(4).

Gilens, Martin. 2015. "Descriptive Representation, Money, and Political Inequality in the United States." *Swiss Political Science Review* 21(2):222-28.

Jost, John T. 2006. "The End of the End of Ideology." *American Psychologist* 61(7): 651-70.

Tufte, Edward R. 1969. "Improving Data Analysis in Political Science." *World Politics* 21(4):641-54.

Week 7: Linear Regression

Before Class:

Watch online lecture titled "Week 7: Linear Regression" via Quercus. Download and explore the `Violence.RData` datasets as well as the `Mission6.R` script. Study the required readings as well as the lecture slides, and read at least one of the supplemental readings.

Assignment(s):

- Mission 6: Why are Some Societies More "Violent" than Other Societies?
  - Your role: Imagine you are a   criminologist studying cross-national variation in homicide rates
- Data Memo

Required Readings:

Pp. 305-310, 317-323, and 328-329 of Diez, David, Mine Çetinkaya-Rundel, and Christopher D. Barr. 2019. *OpenIntro Statistics.* Available online for free via Leanpub.

Pp. 59-70 of Gareth, James et al. 2023. *An Introduction to Statistical Learning with Applications in R (Second Edition).* Available online for free at `https://www.statlearning.com/`.

Supplemental Readings (Pick One):

Krahn, Harvey, Timothy F. Hartnagel, and John W. Gartrell. 1986. "Income Inequality and Homicide Rates: Cross-National Data and Criminological Theories." *Criminology* 24(2): 269-94.

Neuman, W. Lawrence, and Ronald J. Berger. 1988. "Competing Perspectives on Cross-National Crime: An Evaluation of Theory and Evidence." *The Sociological Quarterly* 29(2):281-313.

Sampson, Robert J., and John H. Laub. 2005. "A Life-Course View of the Development of Crime." *The ANNALS of the American Academy of Political and Social Science* 602(1):12-45.

Pp. 530-551 of Weisburd, David, Chester Britt, David B. Wilson, and Alese Wooditch. 2020. *Basic Statistics in Criminology and Criminal Justice.* Cham, Switzerland: Springer International Publishing.

Week 8: Basics of Sampling and the Sampling Distribution

Before Class:

Watch online lecture titled "Week 8: Basics of Sampling and the Sampling Distribution" via Quercus. Download and explore the `IowaCensus.RData` datasets as well as the `Mission7.R` script. Study the required readings as well as the lecture slides, and read at least one of the supplemental readings.

Assignment(s):

- Mission 7: How Can We Use a Sample to Say Something About the Characteristics of a Population?
  - Your role: Imagine you are a demographer studying educational and age differences in the rural United States

Required Readings:

Pp. 25-28; 133-141; 170-180; 181-186 of Diez, David, Mine Çetinkaya-Rundel, and Christopher D. Barr. 2019. *OpenIntro Statistics.* Available online for free via Leanpub.

Supplemental Readings (Pick One):

McFalls, Joseph A., Jr. 2003. "Population: A Lively Introduction (4th Edition)." *Population Bulletin* 58(4): 3-40.

Swanson, David A, Thomas K. Burch, and Lucky M. Tedrow. 1996. "What is Applied Demography?" *Population Research and Policy Review* 15:403-418.

Pp. 1-29 and 108-145 of Weeks, John R. 2008. *Population: An Introduction to Concepts and Issues, Tenth Edition.* Belmont, CA: Thomson Higher Education

Week 9: Inferences for Correlation and Regression

Before Class:

Watch online lecture titled "Week 9: Inferences for Correlation and Regression" via Quercus. Download and explore the `World.RData` datasets as well as the `Mission8.R` script. Study the required readings as well as the lecture slides, and read at least one of the supplemental readings.

Assignment(s):

- Mission 8: Why are Some Societies "Happier" than Others?
    - Your role: Imagine you are a health sociologist investigating the predictors of subjective well-being

Required Readings:

Pp. 331-334 of Diez, David, Mine Çetinkaya-Rundel, and Christopher D. Barr. 2019. *OpenIntro Statistics.* Available online for free via Leanpub.

Pp. 295-307 of Carey, Thomas M. and Jeffrey J. Harden. *Monte Carlo Simulation and Resampling Methods for Social Science.* Thousand Oaks, CA: Sage Publications.

Supplemental Readings (Pick One):

Chen, Ying, Dorota Weziak-Bialowolska, Matthew T. Lee, Piotr Bialowolski, Eileen McNeely, and Tyler J. VanderWeele. 2022. "Longitudinal Associations between Domains of Flourishing." *Scientific Reports* 12(1):2740.

Land, Kenneth C., and Alex C. Michalos. 2018. "Fifty Years After the Social Indicators Movement: Has the Promise Been Fulfilled? An Assessment an Agenda for the Future." *Social Indicators Research* 135(3):835-68.

VanderWeele, Tyler J. 2017. "On the Promotion of Human Flourishing." *Proceedings of the National Academy of Sciences* 114(31):8148-56.

Willen, Sarah S., Abigail Fisher Williamson, Colleen C. Walsh, Mikayla Hyman, and William Tootle. 2022. "Rethinking Flourishing: Critical Insights and Qualitative Perspectives from the U.S. Midwest." *SSM - Mental Health* 2:100057.

Week 10: Introduction to Multiple Linear Regression

Before Class:

Watch online lecture titled "Week 10: Introduction to Multiple Linear Regression" via Quercus. Download and explore the `PostMat.RData` datasets as well as the `Mission9.R` script. Study the required readings as well as the lecture slides, and read at least one of the supplemental readings.

Assignment(s):

- Mission 9: Does Socioeconomic Development Predict Changes in a Society's Culture?
  - Your role: Imagine you are a cultural sociologist studying "post-materalist" cultural values

Required Readings:

Pp. 343-352; 358-370 of Diez, David, Mine Çetinkaya-Rundel, and Christopher D. Barr. 2019. *OpenIntro Statistics.* Available online for free via Leanpub.

Pp. 71-82 of Gareth, James et al. 2023. *An Introduction to Statistical Learning with Applications in R (Second Edition).* Available online for free at `https://www.statlearning.com/`.

Supplemental Readings (Pick One):

Dunlap, Riley E., and Richard York. 2008. "The Globalization of Environmental Concern and The Limits of The Postmaterialist Values Explanation: Evidence from Four Multinational Surveys." *The Sociological Quarterly* 49(3):529-63.

Pp. 8-35 of Inglehart, Ronald. 2018. *Cultural Evolution: People's Motivations are Changing, and Reshaping the World.* New York, NY: Cambridge University Press.

Inglehart, Ronald F., and Pippa Norris. 2012. "The Four Horsemen of the Apocalypse: Understanding Human Security." *Scandinavian Political Studies* 35(1):71-96.

Week 11: Extending Regression with Categorical Variables and Interactions

Before Class:

Watch online lecture titled "Week 11: Extending Regression with Categorical Variables and Interactions" via Quercus. Download and explore the `Footprint.RData` datasets as well as the `ExtendingRegression.R` script. Study the required readings as well as the lecture slides, and read at least one of the supplemental readings.

Assignment(s):

- Mission 10: How Do Countries Vary in their "Ecological Footprint"?
  - Your role: Imagine you are an environmental sociologist investigating cross-national variation in resource use

Required Readings:

Pp. 343-352; 358-370 of Diez, David, Mine Çetinkaya-Rundel, and Christopher D. Barr. 2019. *OpenIntro Statistics.* Available online for free via Leanpub.

Pp. 83-102 of Gareth, James et al. 2023. *An Introduction to Statistical Learning with Applications in R (Second Edition).* Available online for free at `https://www.statlearning.com/`.

Supplemental Readings (Pick One):

Pp. 26-46 of Angus, Ian. 2016. *Facing the Anthropocene: Fossil Capitalism and the Crisis of the Earth System.* New York, NY: Monthly Review Press.

Pp. 1-30 of Collins, Andrea, and Andrew Flynn. 2015. *The Ecological Footprint: New Developments in Policy and Practice.* Cheltenham, UK: Edward Elgar Publishing.

Lynch, Michael J., Michael A. Long, Paul B. Stretesky, and Kimberly L. Barrett. 2019. "Measuring the Ecological Impact of the Wealthy: Excessive Consumption, Ecological Disorganization, Green Crime, and Justice." *Social Currents* 6(4):377-95.

Steffen, Will, Wendy Broadgate, Lisa Deutsch, Owen Gaffney, and Cornelia Ludwig. 2015. "The Trajectory of the Anthropocene: The Great Acceleration." *The Anthropocene Review* 2(1):81-98.

Week 12: Regression and Causal Inference

Before Class:

Watch online lecture titled "Week 12: Regression and Causal Inference" via Quercus. Download and explore the `Duncan.RData` datasets as well as the `Causal.R` script. Study the required readings as well as the lecture slides, and read at least one of the supplemental readings.

Assignment(s):

- None.

Required Readings:

Pp. 77-95 and 105-117 of Morgan, Stephen L. and Christopher Winship. 2015. *Counterfactuals and Causal Inference: Methods and Principles for Social Research.* New York, NY: Cambridge University Press.

Supplemental Readings (Pick One):

Elwert, Felix. 2013. "Chapter 13: Graphical Causal Models" in *Handbook of Causal Analysis for Social Research*, edited by Stephen L. Morgan. New York, NY: Springer.

Pp. 447-454 of Hayes, Andrew F. 2018. *Introduction to Mediation, Moderation, and Conditional Process Analysis, Second Edition: A Regression-Based Approach.* New York, NY: The Guildford Press.

VanderWeele, Tyler J. 2016. "Mediation Analysis: A Practitioner's Guide." *Annual Review of Public Health* 37(1):17-32.

Week 13: Flash Presentations

Assignment(s):

- Short (approximately 5-7 minute) presentation on your research
- Final paper due near the end of the term

9   Use of Artificial Intelligence (AI)

Generative Artificial Intelligence (AI), and specifically <u>foundation models</u> that can create writing, computer code, and/or images using minimal human prompting, are proliferating and becoming ubiquitous. This includes not only GPT-4, Gemini, and other foundation models, but many writing assistants that are built on this or similar AI technologies. There are now hundreds of these systems that are readily available. AI assistants are becoming more proficient at:

1. Creating an outline for a paper, or bullet points and graphics for slides

2. Writing longer coherent prose in multiple languages

3. Providing explanations or ideas for a literature review with mostly accurate citations

4. Summarizing longer articles, text, or a corpus of texts

5. Suggesting a response to a question, such as on a short answer or multiple-choice test, or for a discussion board posting

6. Translating text more accurately

7. Creating computer code in multiple languages

8. Assisting users with formulas inside applications such as Excel

These are only a few examples. Many AI assistant applications give the user a choice of templates (e.g., email, essay, memo, plan) and a choice of tone to tailor the generated text to the user's need.

In this course you can use generative AI but with some restrictions. While powerful, it is absolutely crucial to recognize that these tools are not infallible and can make mistakes. They should complement, not replace, your own independent thinking and writing. As of the date of this syllabus (September 2024), generative AI can be likened to a **friendly, helpful, but unreliable research assistant**, prone to obvious errors and occasionally providing misleading information. **Current AI tools in particular struggle with mathematical reasoning, factual accuracy, and generating truly unique ideas.**

Permitted Uses

You are allowed to use generative AI for:

- Outlining the data memo and final paper
- Preparing presentations
- Summarizing and reinterpreting course readings
- Generating analogies of formulas and ideas
- Learning how to code
- Brainstorming ideas
- Proofreading text

Prohibitions

The primary reason for allowing AI use is to enhance your learning of the course material, not to have the AI complete work for you. Use these tools to deepen your understanding and improve your skills, rather than as a shortcut to avoid engaging with the course content. Along these lines, the following uses of generative AI are **strictly prohibited**:

- Copying and pasting AI-generated text directly into your work without modification or critical engagement
- Using AI to generate answers for missions directly with no effort on your part to learn the material

If you use generative AI in your work, **you must include a statement at the end of your submission detailing how you used these tools.** If you have any doubts about how to use AI tools appropriately in this course, please contact me for clarification.

10   Academic Integrity

Copying, plagiarizing, falsifying medical certificates, or other forms of academic misconduct will not be tolerated. Any student caught engaging in such activities will be referred to the Dean's office for adjudication. Any student abetting or otherwise assisting in such misconduct will also be subject to academic penalties. Students are expected to cite sources in all written work and presentations. See this link for tips for how to use sources effectively.

According to Section B.I.1.(e) of the Code of Behaviour on Academic Matters it is an offense "to submit, without the knowledge and approval of the instructor to whom it is submitted, any academic work for which credit has previously been obtained or is being sought in another course or program of study in the University or elsewhere." By enrolling in this course, you agree to abide by the university's rules regarding academic conduct. You are expected to be familiar with the Code of Behaviour on Academic Matters and Code of Student Conduct, which spell out your rights and provide all relevant details on academic responsibilities at the University of Toronto.

11   Copyright Statement

Course materials prepared by the instructor are considered by the University to be an instructor's intellectual property covered by the *Copyright Act*, RSC 1985, c C-42. These materials are made available to you for your own study purposes, and cannot be shared outside of the class or "published" in any way. Lectures, whether in person or online, cannot be recorded without the instructor's permission. Posting course materials or any recordings you may make to other websites without the express permission of the instructor will constitute copyright infringement.

12   Accessibility Services

It is the University of Toronto's stated goal to create a community that is inclusive of all persons and treats all members of the community in an equitable manner. In creating such a community, the University aims to foster a climate of understanding and mutual respect for the dignity and worth of all persons. Please see the University of Toronto

Governing Council's Statement of Commitment Regarding Persons with Disabilities. In working toward this goal, the University has committed to supporting and facilitating the accommodation of individuals with disabilities so that all may share the same level of access to opportunities, participate in the full range of activities that the University offers, and achieve their full potential as members of the community.

Students seeking support must have an interview with a disability adviser to discuss their individual needs. In many instances it is easier to arrange certain accommodations with advance notice, so you are strongly encouraged to act as quickly as possible. To schedule a registration appointment with a disability adviser, please visit Accessibility Services or call 416-978-8060. The office is located at 455 Spadina Avenue, 4th Floor, Suite 400. Additional student resources for distressed or emergency situations can be found here. You may also contact the Health & Wellness Centre at 416-978-8030 or Student Crisis Response at 416-946-7111.

13   Course Extensions: Extenuating Circumstances

Students are expected to submit course work on time. Occasionally, students may not be able to make agreed upon deadlines due to extenuating circumstances. Students are required to make arrangements with their instructors about how to submit late course work. The graduate office highly recommends that course work extensions remain within the term dates in which the course was taught.

Note: submitting work beyond the term end date (not the last day of instruction but the actual end of term, e.g., the last day of a winter term class may be April 3, but the term ends April 30) requires a discussion with the instructor and the graduate office, as well as completion of an SGS request for an extension of course work form. These forms will be considered by the graduate office and are not automatically approved.

14   Equity and Diversity

The University of Toronto has a public commitment to equity and respect for diversity. All members of the learning environment in this course should strive to create an atmosphere of mutual respect. As the course instructor, I will neither condone nor tolerate behavior that undermines the dignity or self-esteem of any individual in this course and wish to be alerted to any attempt to create an intimidating or hostile environment. It is our collective responsibility to create a space that is inclusive and welcomes discussion. Discrimination, harassment, and hate speech of any kind will not be tolerated. Additional information on Equity and Diversity at the University of Toronto is available here.

15   Land Acknowledgment

The land on which we gather is the traditional territory of the Wendat, Anishinabek (ah-nish-nah-bek) Nation, the Haudenosaunee (ho-den-oh-sho-nee) Confederacy, the Mississaugas of Scugog (skoo-gog), Hiawatha (hi-ah-wah-tha), and Alderville First Nations and the Métis (may-tee) Nation. This territory was the subject of the Dish With One Spoon Wampum Belt Covenant, an agreement between the Iroquois Confederacy and the Ojibwe and allied nations to peaceably share and care for the resources around the Great Lakes. Today, the meeting place of Toronto is still the home to many Indigenous people from across Turtle Island and we are grateful to have the opportunity to work in the community, on this territory.

To learn more about the meaning of this land acknowledgement, please see: `https://native-land.ca/resources/territory-acknowledgement/`