Laura Chelidonopoulos

laura.chelidonopoulos@pitt.edu　　　　　　　　MWF 1:30–3:45PM (Location TBA)

Office Hours: TBA

https://xkcd.com
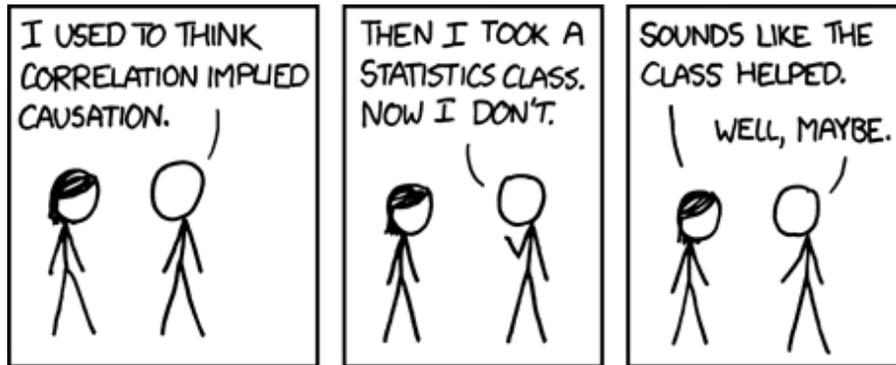
Catalog Description: Data analysis is quickly changing the way we understand and engage in politics, how we implement policy, and how organizations across the world make decisions. In this course, we will learn the fundamental principles of statistical inference and develop the necessary programming skills to answer a wide range of political and policy oriented questions with data analysis. Who is most likely to win the upcoming presidential election? Do countries become less democratic when leaders are assassinated? Is there racial discrimination in the labor market? These are just a few of the questions we will work on in the course.

Students are not expected to have any prior programming knowledge or experience. The course will be centered around bite–size assignments that will help build coding and statistical skills from scratch. Students will leave the course equipped for work in any setting that requires a social scientific approach to data analysis, from policy non–profits to government, from Silicon Valley to Wall Street and beyond.

Required Textbook: Elena Llaudet and Kosuke Imai. *Data Analysis for Social Science: A Friendly and Practical Introduction*. United States, Princeton University Press, 2022.

Required Statistical Softwares: R (www.r–project.org) and RStudio (www.rstudio.com). Both are free!

Tentative Course Schedule: The tentative course schedule is detailed in the table below. (Disclaimer: The schedule, policies, procedures, and assignments in this course are subject to change in the event of extenuating circumstances, by mutual agreement, and/or to ensure better student learning.)

Tentative Course Schedule (Subject to Change as Course Progresses):

Day	Topic	Readings	Key Concepts and *R* Code
		Week 1	
M 6/29	Course Introduction		
	R Installation Guide		
W 7/1	Introduction to R and RStudio	1–1.6	*R*: +, -, *, /, <-, ", (), sqrt(), #
	Observations and Variables + In-class Exercise	1.7	dataframes, observations, variables, unit of observation, i, character vs. numeric variables, binary vs. non-binary variables, n; *R*: setwd(), read.csv(), View(), head(), dim()
F 7/3	Computing and Interpreting Means + In-class Exercise	1.8–1.10	mean or average, $\sum$, unit of measurement; *R*: $, mean()
	Problem Set # 1		
		Week 2	
M 7/6	Estimating Causal Effects with Randomized Experiments	2–2.4	causal relationships, treatment (X) vs. outcome variables (Y), potential outcomes, factual vs. counterfactual outcomes, fundamental problem of causal inference, individual vs. average causal effects, randomized experiments, random treatment assignment, treatment and control groups, pre-treatment characteristics, the difference-in-means estimator
	Estimating Causal Effects (continued)	2.5–2.7	*R*: ==, ifelse(), []
W 7/8	In-class Exercise		
	Survey Research and Exploring One Variable + In-class Exercise	3–3.4 (skip: 3.2.2, 3.4.1, 3.4.2, 3.4.3, 3.4.5)	sample, representative sample, random sampling, table of frequencies, table of proportions, histogram, descriptive statistics (mean, median, standard deviation, and variance); *R*: table(), prop.table(), hist(), median(), sd(), var(), ^
F 7/10	Problem Set # 2		
		Week 3	
M 7/13	Exploring the Relationship Between Two Variables	3.5–3.7	scatter plot, correlation; *R*: plot(), cor()

Continued on next page

Day	Topic	Readings	Key Concepts and R Code
	Predicting Non–Binary Outcomes Using Linear Regression	4–4.4.1	prediction and correlation, predicted ($\widehat{Y}$) vs. actual outcome (Y), prediction errors ($\widehat{\epsilon}$), the least squares method, the linear regression model, $\widehat{Y} = \widehat{\alpha} + \widehat{\beta}X$, interpretation of coefficients, intercept ($\widehat{\alpha}$) and slope ($\widehat{\beta}$), $\triangle\widehat{Y} = \widehat{\beta}\triangle X$; R: lm($Y \sim X$), abline()
W 7/15	Predicting Binary Outcomes Using Linear Regression	4.6–4.9 (skip 4.8)	R^2, relationship between R^2 and correlation
	What is the Effect of the Death of the Leader on the Level of Democracy? + In-class Exercise		
F 7/17	Problem Set # 3		
	Week 4		
M 7/20	Estimating Causal Effects with Observational Data and the Problem of Confounders	5–5.3.1	observational studies vs. randomized experiments, confounders (Z), interpretation of $\widehat{\alpha}$ and $\widehat{\beta}$ when X is binary and identifies treatment assignment
	Controlling for Confounders Using Multiple Linear Regression + In-class Exercise	5.3.2–5.4.2	multiple vs. simple linear regression models, new interpretation of coefficients
W 7/22	Does Having a Black Candidate Running Increase Black Turnout? + In-class Exercise		
	Internal vs. External Validity	5.5–5.7	internal validity, external validity
F 7/24	Problem Set # 4		
	Week 5		
M 7/27	Probability	6–6.8 (skim, skip 6.7 and ignore all code)	probability, random variables, probability distributions, Bernoulli vs. normal distribution, the standard normal distribution, population parameters vs. sample statistics, the law of large numbers, the central limit theorem
	Hypothesis Testing with Coefficients	7–7.1 (skim), 7.3–7.6 (skip 7.3.1)	hypothesis testing, test statistic, standard error of $\widehat{\beta}$; R: summary()$coef
W 7/29	Is There Racial Discrimination in the Labor Market? + In-class Exercise		
	Does Social Pressure Affect Turnout? + In-class Exercise		
F 7/31	Problem Sets # 5		

Day	Topic	Readings	Key Concepts and R Code
		Week 6	
M 8/3	Final Project Instructions		
	Does Social Pressure Affect Turnout? + In-class Exercise		
W 8/5	Is There Racial Discrimination in the Labor Market? + In-class Exercise		
	Final Project Work Time		
F 8/7	Final Project Work Time		

Evaluation: The final grade will be based on the following:

1. *Class Participation* (*10%*). Active engagement and participation in class discussions and activities.

2. *In-class Exercises* (*10%*). In-class exercises must be submitted on Canvas after class. The due date is the same day as the class in which the exercise is completed.

3. *Problem Sets* (*50%*). Students will work on problem sets during Friday classes and are encouraged to work in groups, consult with me, and collaborate with classmates. The goal is to create a cooperative learning environment where everyone can learn together. Problem sets must be submitted through Canvas by the end of the Friday class (3:45PM). Late submissions will not be accepted unless special permission is granted in advance.

4. *Final Project* (*30%*). This is an individual assignment and is not to be done in groups. Students will be given two different datasets to choose from, analyze the data, and answer a set of questions based on the course content covered throughout the weeks. This project will test students' ability to conduct data analysis (a very important real-world skill and, from my view, the most important takeaway from this course).

Class Policies:

E-mail Policy

The University of Pittsburgh has established that e-mails are the primary form of communication. It is your responsibility to be up to date of any information posted on Canvas or sent via e-mail throughout the semester. I will check and respond to e-mails from Monday to Friday, from 9-5pm, and please allow up to 24h for me to respond. Emails sent on weekends, overnight before class, or on holidays will be returned only during working hours.

Conduct

Respect for diversity is essential to how we learn and engage in this class. I believe that meaningful learning happens when everyone's experiences, perspectives, and voices are welcomed and valued. Importantly, I recognize that we do not all start from the same place, so please feel free to reach out to me if there are any questions or concerns. I especially encourage an environment with active participation, where disagreements and debates will be approached respectfully with an emphasis on listening and understanding.

Course Goals and Learning Objectives:

Goals	Objectives	Assessments
1. Students will know how to recognize and interpret quantitative information	(a) Students will be able to read and understand quantitative data in various formats (b) Students will be able to communicate the meaning of quantitative data and the results of data analysis	Participation In-classes Problem sets Final Project
2. Students will understand the theoretical basis of quantitative reasoning	(a) Students will be able to explain the basic concepts of quantitative reasoning, such as variables, constants, and estimates (b) Students will be able to understand how inferences are drawn from quantitative analysis (c) Students will be able to recognize the limitations of quantitative methods	Participation In-classes Problem sets Final Project
3. Students will understand the practical application of quantitative data analysis	(a) Students will be able to determine and use appropriate quantitative methods to solve problems (b) Students will be able to accurately interpret the results of data analyses (c) Students will be able to assess results for reasonableness	Participation In-classes Problem sets Final Project