Social Sciences Quantitative Laboratory
Workshops Offered
Default time: 2 hours

contact: Ella Foster-Molina

mfoster1@swarthmore.edu

ITEM	DESCRIPTION	NOTES
	DATA ANALYSIS WORKSHOPS (NO PROGRAMMING)	
Introduction to Interpreting Regressions	Many books and papers present regression results to support their arguments. There are some simple ways to scan these tables that let you quickly understand the main points and common pitfalls. We cover how to read regression tables and graphs efficiently, what the numbers for coefficients and statistical significance mean, and common ways regressions are misused.	Open attendance
		No prerequisite
Where to Find Data	Many research ideas are generated by creating questions answerable by existing data sources. We cover some commonly used data sites, how to navigate these sources, and how to search for specific questions in surveys. We also use Google Sheets to easily create some basic graphics to summarize data on public opinions of global warming.	Open attendance
		No prerequisite
Data Visualization: The Good, the Bad, and the Ugly	Data visualization is a powerful tool to communicate a point, but can also be misused. We discuss the history of graphics in data interpretation, how graphics can highlight important theories, how to create your own graphics using Google Sheets, and how graphics are used to misinform.	Open attendance
	Reference: Tufte, Edward, and P. Graves-Morris. "The visual display of quantitative information.; 1983." (2014).	No prerequisite
Sampling and the 2016 Election	Random samples are the basis for many statistical methods, including polling people to predict election winners. We use the 2016 election as an example of good and bad sampling methods, examining when the polls correctly predicted election results and why they sometimes got it wrong.	Open attendance
		No prerequisite
Digital Tools in Research	Text is everywhere, yet the systematic study of large quantities of text as a way to understand the world is still in its infancy. I demonstrate how I use some of these skills to reveal an aspect of why rich white men are better represented in policy. Tweets by members of Congress, the text of their budgetary documents detailing staffing resources, and the bills they create reveal that members of Congress from poor districts are distracted from policy because they spend more effort helping their constituents navigate the social safety net. Rich white men are represented, in part, because they require less help with the federal bureacracy, freeing their politicians to make policy.	Open attendance

ITEM	DESCRIPTION	NOTES
		No prerequisite
Cherry Picking Data: The Widespread Problem of P-Hacking	P-hacking is a common way to cherry pick your results by selecting regressions and t-tests that are artificially significant. The kicker is that many people don't know they are doing it. We look at the connection, or lack of connection, between political parties and national economic health. We discuss how p-hacking happens, why it's a problem in social science research, and how to avoid it.	Open attendance
		No prerequisite
Game Theory in Social Science: When Philosophy meets Math	Game theory is mathematical philosophy for social science. We start with the classic prisoners' dilemma. Instead of prison terms we use a tastier incentive to reach the best outcome considering your opponent's best strategy of play: brownies. We discuss the ways that game theory has improved social science theory, and where it still needs improvement.	Open attendance
		No prerequisite
Theory and Data Analysis: Ever the Twain Shall Meet	From data selection, data cleaning, to the final statistical results, all data analysis relies on solid theoretical grounding. We discuss the role of theory in statistical analyses, but historically and in the modern era. We discuss papers that have developed strong theoretical justifications for their techniques. as well as papers that have data mined their way to results that fail to match their theoretical basis.	Open attendance
		No prerequisite
Experimental Methods	Experiments are often the gold standard method for uncovering causal connections; i.e., does donating to campaigns make them more likely to win? Yet they are tricky to pull off, and even trickier to apply to the relevant population. One common pitfall is experimenting on college students, then assuming that college students behave like the rest of the world. We discuss the relevance of experiments, and how they are implemented.	Open attendance
		No prerequisite

ITEM	DESCRIPTION	NOTES
	PROGRAMMING WORKSHOPS	
Introduction to Stata I	This introductory session is based on a survey conducted on workshop participants. We cover the basics of opening files, running commands, creating summary statistics and graphics, dealing with outliers, manipulating variables, and labelling variables. We also work through how to deal with errors (which are inevitable, but manageable, in any statistical package). This will be hands on, practical experience: each participant will be using the program in real time.	By reservation only
		No prerequisite
Introduction to R	This introductory session is based on a survey conducted on workshop participants. We cover the basics of opening files, running commands, creating summary statistics and graphics, dealing with outliers, and manipulating variables.. We also work through how to deal with errors (which are inevitable, but manageable, in any statistical package). This will be hands on, practical experience: each participant will be using the program in real time.	By reservation only
		No prerequisite
Introduction to Stata II	In this session we cover several additional basics of data analysis: scatterplots, boxplots, standard errors/standard deviations/the normal distribution, and the fundamentals of developing useful theories.	By reservation only
	*Note: Any programming workshop can be delivered in R or Stata	Prerequisite: Introduction to Stata I
Replicating Empirical Analyses in Stata	We uncover the complex relationship between unemployment and voter turnout. This session mplements and interprets regressions in Stata, as well as delves deeper into the interpretation of p-values, standard errors, and standard deviations. This replication highlights the importance of omitted variable bias and outliers: the results are fundamentally wrong when we exclude educational attainment, or include invalid data points.	By reservation only
	Paper: Burden, Barry C., and Amber Wichowsky. "Economic discontent as a mobilizer: unemployment and voter turnout." The Journal of Politics 76.4 (2014): 887-898.	Prerequisite: Introduction to Stata II or instructor permission
Stata: Visualizing data	Based on the classic work on visualization by Edward Tufte, this session extends the analysis of unemployment and voter turnout by looking at graphical methods for interpreting the relationship between education, unemployment, and voter turnout. This also serves as additional exploration of omitted variable bias in regressions. In addition, we explore the famous data in Anscombe's quartet to better understand how necessary data visualization is to any data analysis project.	By reservation only
	Paper: Burden, Barry C., and Amber Wichowsky. "Economic discontent as a mobilizer: unemployment and voter turnout." The Journal of Politics 76.4 (2014): 887-898.	Prerequisite: Replicating Empirical Analyses or instructor permission

ITEM	DESCRIPTION	NOTES
Stata: Exporting analyses into Tex/etc and for loops	We have learned how to run regressions and create graphics. The next step, covered in this workshop, is to present those results to others. This workshop introduces Latex, a versatile text formatting package, as one way to present regression results and images. We also use Word and Excel to format regression results and graphics. In the process we examine the importance of variable selection. Finally, a versatile programming method called for loops are introduced.	By reservation only
	Paper: Burden, Barry C., and Amber Wichowsky. "Economic discontent as a mobilizer: unemployment and voter turnout." The Journal of Politics 76.4 (2014): 887-898.	Prerequisite: Visualizing Data or instructor permission
Preparing for a Data Analysis Interview and Take Home Assignment, Part I: Data analysis	It is common to receive a take home assignment as part of a data science/data analysis interview. We cover many common requirements of such an assignment using a dataset from IMDB. This includes technical skills such as importing data, recoding variables, and creating graphics and regressions. More importantly, we discuss what to focus on (data validity, limitations, and theoretical development) and how to present both your results and code.	By reservation only
		Prerequisite: Exporting Analyses or instructor permission
Preparing for a Data Analysis Interview and Take Home Assignment, Part II: Data manipulation	This workshop covers a second type of take home assignment: data manipulation. Again using the IMDB dataset, we discuss how to create a new version of an existing dataset and why people would find this useful. We cover for loops and conditionals in more depth.	By reservation only
		Prerequisite:Preparing for a Data Analysis Inteview Part I or instructor permission
Best Practices in Data Analysis and Coding	This workshop covers the tools required to produce easily interpretable, reproducible code. This includes smart data cleaning practices, the importance of consistent output, how to structure a do file and comment smartly, variable naming conventions. Most importantly, we cover the importance of theoretical development for any data analysis project, and how to concisely describe the purpose of your code as it relates to your theory.	By reservation only
		Prerequisite: Instructor Permission