

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)		
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.	No prerequisite
How to Read a Quantitative Paper	This 1 hour session focuses on reading and interpreting quantitative papers. This includes skimming techniques as well as examining regression tables, connecting them to scatterplots and regressions lines, and interpreting other common statistical summaries.	No prerequisite; 1 hour
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.	No prerequisite
Manipulating Data in Google Sheets	Google sheets provides an intuitive tool to create multiple numeric descriptions of data, as well as scatterplots, bar charts, maps, and more. This workshop uses a sample dataset on world life expectancy, population, and GDP by country over 50 years to demonstrate these functionalities.	No prerequisite
Finding Data and Using It in Google Sheets	Using Google Sheets to create descriptions of your own data requires knowledge of data formatting issues, sorting data, and merging data by hand. We cover some of the common issues in finding your own data, how to efficiently select data to answer a general topic, and how to manipulate your own data in Google Sheets.	Prerequisite: Manipulating Data in Google Sheets
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.	No prerequisite

ITEM	DESCRIPTION	NOTES
Experimental Methods	Experiments are often the gold standard method for uncovering causal connections; i.e., are campaigns that are randomly assigned to receive higher donations 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 lab and field experiments, and how they are implemented.	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.	No prerequisite
	Reference: Tufte, Edward, and P. Graves-Morris. "The visual display of quantitative information.; 1983." (2014).	
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.	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 bureaucracy, freeing their politicians to make policy.	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.	No prerequisite
Theory and Data Analysis: Ever the Twain Shall Meet	From data selection and data cleaning, to the final statistical results, all data analysis relies on solid theoretical grounding. We discuss the role of theory in statistical analyses, both 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.	No prerequisite

ITEM	DESCRIPTION	NOTES
PROGRAMMING WORKSHOPS		
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.	No prerequisite
Introduction to Visualizing Data in R with Data on Partisanship in Presidential Elections	Two packages in R provide powerful tools for visualizing data: ggplot and plotly. This workshop uses a dataset on partisanship, county demographics, and deaths of despair to demonstrate the power of these packages. You will learn to create maps, barplots, scatterplots, and 3 dimensional scatterplots, as well as the potential to publish this data online. In the process we discover the correlations between unemployment, deaths of despair, voting for Trump, and more.	Prerequisite: Experience programming in R
Stata 1: Introduction	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.	No prerequisite
Stata 2: Describing Data, Distributions	Using a dataset on county demographics, presidential voting, and deaths of despair, this workshop covers several additional basics of data analysis. This includes scatterplots, boxplots, standard errors/standard deviations/the normal distribution, and the fundamentals of developing useful theories.	Prerequisite: Stata 1
Stata 3: Regression Analysis	Using a dataset on county demographics, presidential voting, and deaths of despair, this workshop covers how to use Stata to create regressions, test regression assumptions, manipulate data to make it satisfy regression assumptions, and understand model specification and interaction terms.	Prerequisite: Stata 2
Stata 3a: Replicating Empirical Analyses in Stata	We uncover the complex relationship between unemployment and voter turnout. This session implements 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.	Prerequisite: Stata 2
	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.	

ITEM	DESCRIPTION	NOTES
Stata 4: 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.	Prerequisite: Stata 3a or instructor permission
	Paper: Burden, Barry C., and Amber Wichowsky. "Economic discontent as a mobilizer: unemployment and voter turnout." <i>The Journal of Politics</i> 76.4 (2014): 887-898.	
Stata 5: 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" is introduced.	Prerequisite: Stata 4
	Paper: Burden, Barry C., and Amber Wichowsky. "Economic discontent as a mobilizer: unemployment and voter turnout." <i>The Journal of Politics</i> 76.4 (2014): 887-898.	
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.	
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.	
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, and 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.	Prerequisite: experience in Stata or R
Manipulating Data	Data rarely comes in a clean format; erroneous data need to be fixed, the format needs to be changed, variables need to be combined or split, and data need to be merged together. This workshop covers how to reshape and reorganize your data.	Prerequisite: experience in Stata or R
Mapping Data in R	R is a data analysis language that can be used to create impressive visuals. Participants will learn how to import data from the census and create cartographs (maps displaying quantitative data) of US areas that interest them.	Prerequisite: none