Overview

Students are encouraged to plan out their quantitative course sequences so as to build the prerequisite skills needed to enter more advanced courses and ultimately to pursue their own research agendas. Incoming GSE graduate students vary greatly in the prior statistical training they may have received. Moreover, they face a myriad of course choices both in the GSE and across the University when it comes to quantitative methods. The purpose of this document is to inform students of various course offerings and to assist them with their choices. The guide describes various introductory statistics sequences taken by GSE students; the additional and more advanced quantitative courses offered by GSE faculty; various workshops concerning quantitative methods; and a listing of potential mini-courses students may wish to take so as to develop their programming skills (in R, Python and Stata).

Table of contents:

While this guide offers a great deal of information, it does not entail information on all quantitative courses at Stanford. For example, if you have little programming background but plan to take courses in computer science or use data science for your own research, then it is highly recommended you take CS 106A: Programming Methodology.

Please also check Explore Courses (https://explorecourses.stanford.edu) as course offerings change from year to year and even quarterly. In addition, be sure to discuss with your academic advisors and more advanced students which courses and skills are most related to your interests.

Last, be sure to consider methods requirements for your degree when selecting courses. Should you wish to make substitutions, you readily can, but it requires getting approval from the required courses' instructors. This guide offers alternative training pathways many students have taken before, but course contents can change, so you will still need to get such a waiver (see your degree handbook).

This guide to quantitative course offerings was developed by Daniel McFarland with input from GSE graduate students, faculty and academic services (circa 2021). For any additional questions on courses, please contact Tommy Liu (liutommy@stanford.edu).[1]

[1] Thank you to the GSE's faculty, dean of students, graduate students, and members of academic services who provided input. Special thanks to Emma Kerr for her assistance compiling this information.

Example Introductory Statistics Sequences

To fulfill graduation requirements, GSE PhD students are asked to take two statistics courses, EDUC 400A and EDUC 400B; one qualitative course, EDUC 450A; and an additional advanced methods course. The sequences below reflect this main sequence as well as alternative sequences students substituted in the past, based on their prior training and interest, to meet requirements. To make such substitutions, email the instructors of EDUC 400A and 400B, identifying the proposed alternate course(s) and a brief rationale for the change (https://ed.stanford.edu/academics/doctoral-handbook/courses/waivers).

Stata Tracks

Main sequence: Conceptual understanding / predictive modeling emphasis

EDUC 400A: Introduction to Statistical Methods in Education (3-4 units, AUT) Schwartz, D.	EDUC 400B: Statistical Analysis in Education: Regression (5 units, WIN) Bettinger, E.	EDUC 326: Advanced Regression Analysis (3-5 units, SPR) Smith, S.

Alternative 1: Conceptual understanding / predictive modeling emphasis (intermediate)

SOC 381: Sociological Methodology I: Introduction (5 units, AUT) Dahir, N.; Torche, F.	SOC 382: Sociological Methodology II: Principles of Regression Analysis (4-5 units, WIN) Torche, F.	SOC 383: Sociological Methodology III: Models for Discrete Outcomes (5 units, SPR) Freese, J.

Alternative 2: Causal modeling emphasis (advanced, prerequisite basic statistics)

EDUC 430A: Experimental Research Design and Analysis (3-5 units, AUT) Bettinger, E.	EDUC 430B: Quasi-Experimental Research Design & Analysis (3-5 units, WIN) Dee, T.	EDUC 430C: Using Data to Describe the World: Descriptive Social Science Research Techniques (3-5 units, SPR) reardon, s.

R Tracks

Alternative 3: Data management to predictive modeling

ED 423A Introduction to Data Science I: Data Processing (3-4 units, AUT) S. Smith	PYSCH 252 Statistical Methods for Behavioral and Social Sciences (5 units, WIN) T. Gerstenberg	EDUC 326 Advanced Regression (3-5 units, SPR) S. Smith

Alternative 4: Experimentation to predictive modeling

PSYCH 251: Experimental Methods (3 units, AUT) Frank, M.	PSYCH 252: Statistical Methods for Behavioral and Social Sciences (5 units, WIN) Gerstenberg, T.	PSYCH 253: Advanced Statistical Modeling (3 units, SPR) Poldrack, R.; Yamins, D.

Alternative 5: Predictive modeling and experimentation (advanced, prerequisite application)

POLISCI 450A: Political Methodology I: Regression (5 units, AUT) Xu, Y.	POLISCI 450B: Political Methodology II: Causal Inference (3 units, WIN) Hainmueller, J.	POLISCI 450C: Political Methodology III: Model-Based Inference (3 units, SPR) Grimmer, J.

GSE QUANTITATIVE COURSE GUIDE

List of Courses in Introductory Statistics Sequences

Course#	Course Name	Instructors	Description	Methods taught	Language	Units	QTR	Pre-requisites	Assignments
EDUC 400A	Introduction to Statistical Methods in Education	Schwartz, D.	(Formerly EDUC 160.) Basic techniques in descriptive and inferential statistics for educational research will be covered with an emphasis on rigorous preparation for intermediate and advanced courses. Topics include central tendency, variance, probability, distributions, confidence interval, t-test, F-test, correlation, regression, and analysis of variance. Non-parametric statistics and graphical principles for data representation will also be addressed. Students will also be introduced to STATA in preparation for subsequent higher level courses.	descriptive statistics; inferential statistics; entral tendency; variance; probability; distributions; confidence intervals; t-test; F-test; correlation; regression; analysis of variance	Excel	3-4	AUT	None	Six individual problem sets; 2 small group data gathering & analysis projects; final exam
EDUC 400B	Statistical Analysis in Education: Regression	Bettinger, E.	Primarily for doctoral students; part of doctoral research core; prerequisite for advanced statistical methods courses in School of Education. Basic regression, a widely used data-analytic procedure, including multiple and curvilinear regression, regression diagnostics, analysis of residuals and model selection, logistic regression. Proficiency with statistical computer packages.	basic regression; multiple regression; curvilinear regression; regression diagnostics; analysis of residuals; model selection; logistic regression	STATA	5	WIN	EDUC 400A or equiv	5 short assignments (CR/NC), 3 problem sets (graded), 1 report, and final exam (all ind)
EDUC 326	Advanced Regression Analysis	Smith, S.	Social science researchers often deal with complex data and research questions that traditional statistics models like linear regression cannot adequately address. This course offers the opportunity to understand and apply two widely used types of advanced regression analysis that allow the examination of 1) multilevel data structures (multilevel models) and 2) multivariate research questions (structural equation models).	advanced regression analysis; multilevel data structures; multilevel models; multivariate models	R (STATA if needed)	3-5	SPG	Basic stats, programming, regression	Required weekly labs (ind/ non-graded); 2 graded projects (ind or group)
EDUC 423A	Introduction to Data Science I: Data Processing	Smith, S.	Quantitative data require considerable work before they are ready to be analyzed: they are often messy, incomplete and potentially biased. This course is designed to help you thoughtfully collect, manage, clean and represent data so it can offer substantive information researchers can act upon. In our weekly sessions you will take a critical and reflective approach to these tasks and learn the technical skills needed to get your data into shape. Education and social science datasets will be our focus.	collect data; manage data; clean data; represent data	R	3-5	AUT	None	Required weekly labs (ind/ non-graded); 3 ind graded problem sets; 1 group assignment
EDUC 430A	Experimental Research Design and Analysis	Bettinger, E.	The course will cover the following topics: a) the logic of causal inference and the Fisher/Neyman/Rubin counterfactual causal model (Fisher, 1935; Heckman, 1979; Holland, 1986; Neyman, 1990; Rubin, 1978); b) randomized experiments; c) complex randomized experiments in education (cluster randomized trials, multi-site trials, staggered implementation via randomization, etc.); d) policy experiments with randomization; e) meta-analysis; and f) power in randomized experiments; g) the ethics and politics of randomized experiments.	causal inference; the Fisher/Neyman/Rubin counterfactual causal model; randomized experiments; complex randomized experiments; cluster randomized trials; multi-site trials; staggered implementation via randomization; policy experiments with randomization; meta-analysis; power in randomized experiments; ethics and politics of randomized experiments	STATA	3-5	AUT	Basic stats	Weekly or biweekly required (graded) individual problem sets; Required (graded) final. Group work permitted but individual write-ups required
EDUC 430B	Quasi-Experimental Research Design & Analysis	Dee, T.	This course surveys quantitative methods to make causal inferences in the absence of randomized experiment including the use of natural and quasi-experiments, instrumental variables, regression discontinuity, fixed effects estimators, and difference-in-differences. We emphasize the proper interpretation of these research designs and critical engagement with their key assumptions for applied researchers. Prerequisites: Prior training in multivariate regression (e.g., ECON 102B or the permission of the instructor).	causal inferences; natural experiments; quasi-experiments; instrumental variables; regression discontinuity; fixed effects estimators; difference-in-differences	STATA	3-5	WIN	EDUC 430A	Weekly or biweekly required (graded) problem sets (group and individual components); optional (graded) final presentation

EDUC 430C	Using Data to Describe the World: Descriptive Social Science Research Techniques	reardon, s.	This course focuses on the skills needed to conduct theoretically-informed and policy-relevant descriptive social science. Students read recent examples of rigorous descriptive quantitative research that exemplifies the use of data to describe important phenomena related to educational and social inequality. The course will help develop skills necessary to conceptualize, operationalize, and communicate descriptive research, including techniques related to measurement and measurement error, data harmonization, data reduction, and visualization. Students develop a descriptive project during the course. Prerequisite: satisfactory completion of a course in multivariate regression.	descriptive research; conceptualization; operationalization; communication of descriptive research; measurement and measurement error; data harmonization; data reduction; data visualization	STATA	3-5	SPG	EDUC 430B	3 memos that build on each other to answer a descriptive research question and present visualization (graded)
PSYCH 251	Experimental Methods	Frank, M.	Graduate laboratory class in experimental methods for psychology, with a focus on open science methods and best practices in behavioral research. Topics include experimental design, data collection, data management, data analysis, and the ethical conduct of research. The final project of the course is a replication experiment in which students collect new data following the procedures of a published paper. The course is designed for incoming graduate students in psychology, but is open to qualified students from other programs who have some working knowledge of the R statistical programming language. Requirement: Psych 10/ Stats 60 or equivalent	experimental design; data collection; data management; data analysis; research ethics	R	3	AUT	Basic stats	Four required (graded) individual problem sets; Required (graded) final project with required (graded) midterm and final presentations
PSYCH 252	Statistical Methods for Behavioral and Social Sciences	Gerstenberg, T.	This course offers an introduction to advanced topics in statistics with the focus of understanding data in the behavioral and social sciences. It is a practical course in which learning statistical concepts and building models in R go hand in hand. The course is organized into three parts: In the first part, we will learn how to visualize, wrangle, and simulate data in R. In the second part, we will cover topics in frequentist statistics (such as multiple regression, logistic regression, and mixed effects models) using the general linear model as an organizing framework. We will learn how to compare models using simulation methods such as bootstrapping and cross-validation. In the third part, we will focus on Bayesian data analysis as an alternative framework for answering statistical questions. Please view course website: https:// psych252.github.io/. Open to graduate students only. Requirement: Psych 10/ Stats 60 or equivalent	visualize data; manage data; simulate data; frequentist statistics; multiple regression; logistic regression; mediation; mixed models; bootstrapping; cross-validation; Bayesian data analysis	R	5	WIN	PSYCH 251 or equiv	Weekly or biweekly required (graded) individual problem sets; Required (graded) individual midterm exam; Required (graded) final project
PSYCH 253	Advanced Statistical Modeling	Poldrack, R.; Yamins, D.	Introduction to high-dimensional data analysis and machine learning methods for use in the behavioral and neurosciences, including: supervised methods such as SVMs, linear and nonlinear regression and classifiers, and regularization techniques; statistical methods such as bootstrapping, signal detection, factor analysis, and reliability theory; metrics for model/data comparison such as representational similarity analysis; and unsupervised methods such as clustering. Students will learn how to both use existing statistical data analysis packages (such as scikit-learn) as well to build, optimize, and estimate their own custom models using an optimization framework (such as Tensorflow or Pytorch). Requirement: Psych 251. Familiarity with python programming and multivariable calculus and linear algebra (Math 51) highly recommended.	SVMs; linear regression; nonlinear regression; regression classifiers; regularization techniques; bootstrapping; signal detection; factor analysis; reliability theory; model comparison; data comparison; representational similarity analysis; unsupervised methods; clustering	Python	3	SPG	PSYCH 252 or equiv, intermediate fluency in python	2 ind lab reports (Jupyter notebooks), and final ind project of relevant methods / data using Jupyter notebook and class presentation

SOC 381	Sociological Methodology I: Introduction	Dahir, N.; Torche, F.	Enrollment limited to first-year Sociology doctoral students. Other students by instructor permission only. This course provides a conceptual and applied introduction to quantitative social sciences methodology, including measurement, sampling and descriptive statistics, statistical inference, ANOVA, factor analysis, and ordinary least squares regression. Students will be introduced to both the methodological logic and techniques of statistical data analysis. The course will present the purpose, goals, and mathematical assumptions behind techniques of statistical analysis and will discuss applications to analyzing data and interpreting results. In addition to the lecture time, SOC381 includes a weekly lab section to learn statistical software and conduct applied research.n*Students enrolling in Soc381 are strongly encouraged to take a 1-week Math/Statistics refresher course from September 16 to September 20. Please contact the instructor at torche@stanford.edu for details	measurement; sampling; descriptive statistics; statistical inference; ANOVA; factor analysis; ordinary least squares regression; statistical data analysis	STATA	5	AUT	Basic math	Weekly or biweekly required (graded) problem sets. Midterm and final exam or project. All work individual (i.e. not group).
SOC 382	Sociological Methodology II: Principles of Regression Analysis	Torche, F.	Preference to Sociology doctoral students. Other students by instructor permission only. Required for Ph.D. in Sociology. Enrollment limited to first-year Sociology doctoral students. Rigorous treatment of linear regression models, model assumptions, and various remedies for when these assumptions are violated. Introduction to panel data analysis. Enrollment limited to 15. Prerequisites: 381.	linear regression models; model assumptions; panel data analysis	STATA	4-5	WIN	SOC 381 or equiv	Weekly or biweekly required (graded) problem sets
SOC 383	Sociological Methodology III: Models for Discrete Outcomes	Freese, J.	Required for Ph.D. in Sociology; other students by instructor permission only. enrollment limited to first-year Sociology doctoral students. The rationale for and interpretation of static and dynamic models for the analysis of discrete variables. Prerequisites: 381 and 382, or equivalents.	static models; dynamic models; analysis of discrete variables; event history models	STATA	5	SPG	SOC 382 or equiv	Weekly or bi-weekly problem sets (required, graded). Group conferencing on assignments encouraged but individual work submitted.
PolSci 450A	Political Methodology I: Regression	Xu, Y.	Introduction to statistical research in political science, with a focus on linear regression. Teaches students how to apply multiple regression models as used in much of political science research. Also covers elements of probability and sampling theory.	linear regression; multiple regression; probability; sampling theory	R	3-5	AUT	Math test for entry (linear algebra and calculus)	Weekly prob sets (R code from scratch; formal proofs), midterm, and final exam
PolSci 450B	Political Methodology II: Causal Inference	Hainmueller, J.	Survey of statistical methods for causal inference in political science research. Covers a variety of causal inference designs, including experiments, matching, regression, panel methods, difference-in-differences, synthetic control methods, instrumental variables, regression discontinuity designs, quantile regression, and bounds. Prerequisite: POLISCI 450A.	causal inference; experiments; matching; regression; panel methods; difference-in-differences; synthetic control methods; instrumental variables; regression discontinuity; quantile regression; bounds	R	3-5	WIN	PolSci450A	Weekly prob sets (R code from scratch; formal proofs), midterm, and final exam
PolSci 450C	Political Methodology III: Model-Based Inference	Grimmer, J.	Provides a survey of statistical tools for model-based inference in political science. Topics include generalized linear models for various data types and their extensions, such as discrete choice models, survival outcome models, mixed effects and multilevel models. Prerequisites: POLISCI 450A and POLISCI 450B.	model-based inference; generalized linear models; discrete choice models; survival outcome models; mixed effects models; multilevel models	R	3-5	SPG	PolSci450B	Weekly prob sets (R code from scratch; formal proofs), midterm, and final exam

Additional and More Advanced Quantitative Courses in GSE

In addition to introductory statistics sequences, students may want to take more advanced and specialized coursework in quantitative methods. Below is a list of additional quantitative courses offered by GSE faculty. Please note the different methods they cover, the programming languages they use, and their prerequisites. Some have no prerequisites, others require only basic statistics (e.g., one statistics course like EDUC 423A, EDUC 400A, PSYCH 251, etc), and yet others assume students have some familiarity with basic programming (often in R or Python) or multiple regression (e.g., completion of EDUC 423B, EDUC 400B, PSYCH 252, SOC 381, or POLSCI 450A). It is recommended you plan out your quantitative course taking so you are sure to have the prerequisite skills for these more advanced courses. Be sure talk with your program advisors and more advanced students to learn which will best suit you. If you are unsure whether you have the requisite skills to take a particular course, ask the instructor. They want to help you and are often flexible.

Course#	Course Name	Instructors	Description	Methods taught	Language	Units	QTR	Prerequisites	Assignments
EDUC 200A	Introduction to Data Analysis and Interpretation	Smith, S.; Solano-Flores, W.	Primarily for master's students in the School of Education. Focus is on reading literature and interpreting descriptive and inferential statistics, especially those commonly found in education. Topics: basic research design, instrument reliability and validity, descriptive statistics, correlation, t-tests, one-way analysis of variance, and simple and multiple regression. All offerings of this course (whether meeting on Mon & Weds or Tues & Thurs) will be taught identically.	basic research design; instrument reliability and validity; descriptive statistics; correlation; t-tests; one-way analysis of variance; simple and multiple regression	No software application	4	AUT	None	Weekly ind quizzes (graded), 3 ind assignments (graded)
EDUC 222	Resource Allocation in Education	Loyalka, P.	This course covers economic principles and tools for informing resource allocation decisions in education. Students will review concepts related to educational goods and values; the costs and benefits of different levels and types of schooling; public versus private schooling; as well as adequacy and equity in education financing. Students will also learn about the use of educational production functions, teacher value-added estimation, cost effectiveness analysis, experimental program evaluation, systematic reviews, and causal chain analysis. Prerequisites: introductory statistics and regression analysis.	educational production functions; teacher value-added estimation; cost effectiveness analysis; experimental program evaluation; systematic reviews; causal chain analysis	R, STATA	3-5	SPG	basic stats	2 short papers, 2 problem sets, 1 policy brief, and 1 small group presentation
EDUC 234	Curiosity in Artificial Intelligence	Haber, N.	How do we design artificial systems that learn as we do early in life -- as "scientists in the crib" who explore and experiment with our surroundings? How do we make AI "curious" so that it explores without explicit external feedback? Topics draw from cognitive science (intuitive physics and psychology, developmental differences), computational theory (active learning, optimal experiment design), and AI practice (self-supervised learning, deep reinforcement learning). Students present readings and complete both an introductory computational project (e.g. train a neural network on a self-supervised task) and a deeper-dive project in either cognitive science (e.g. design a novel human subject experiment) or AI (e.g. implement and test a curiosity variant in an RL environment). Prerequisites: python familiarity and practical data science (e.g. sklearn or R).	cognitive science; intuitive physics and psychology; developmental differences; computational theory; active learning; optimal experiment design; artificial intelligence; self-supervised learning; deep reinforcement learning	Python	3	WIN	basic stats and programming	TBD
EDUC 252	Introduction to Test Theory	Domingue, B.	Concepts of reliability and validity; derivation and use of test scales and norms; mathematical models and procedures for test validation, scoring, and interpretation.	reliability and validity; derivation; test scales and norms; mathematical models and procedures for test validation; scoring; interpretation	No software application	3	WIN	basic stats	Weekly written responses to assigned readings, final paper (graded)

EDUC 252L	Introduction to Test Theory – Lab	Domingue, B.	This course will cover the material from 252A in an applied setting. Emphasis will be in developing a capacity for applying and interpreting psychometrics techniques to real-world and simulated data.	applying and interpreting psychometrics; real world data; simulated data	R	2	WIN	None	Problem sets every other week (graded)
EDUC 260A	Applications of Causal Inference Methods	Rogosa, D.	See http://rogosateaching.com/stat209/ Application of potential outcomes formulation for causal inference to research settings including: mediation, compliance adjustments, time-1 time-2 designs, encouragement designs, heterogeneous treatment effects, aggregated data, instrumental variables, analysis of covariance regression adjustments, and implementations of matching methods. Prerequisite: STATS 209A/MSE 327 or other introduction to causal inference methods. (Formerly HRP 239)	mediation; compliance adjustments; time-1 time-2 designs; encouragement designs; heterogeneous treatment effects; aggregated data; instrumental variables; analysis of covariance regression adjustments; implementations of matching methods	R	2	WIN	prior course in causal inference	Required weekly ind assignments (non-graded), 2 ind problem sets (graded), and ind final exam (graded)
EDUC 316	Social Network Methods	McFarland, D.	Survey course of network methods, theories and research applications. Arguably all social data is relational. This course trains students to study any relational phenomena -- from links across webpages, to social relations and interactions, to co-occurrence of terms in texts, and even memberships held over time. Students learn how to richly describe and portray these social networks and to statistically model their form, their formation, and peer influence on individual outcomes (statistical network models).	network data management; network survey design, collection and sampling; visualization (static and dynamic); dyadic and triadic analysis; clustering & community detection; centrality measures; blockmodeling; bipartite networks; correspondence analysis; latent Dirichlet allocation; exponential graph models; separable temporal exponential graph models; stochastic actor-oriented models; relational event models; diffusion modeling & simulation	R	3-5	SPG	familiarity with regression and basic R programming (EDUC 423B or equiv)	Required weekly labs in R (ind / non-graded); individual prospectus and final project (graded); individual presentation
EDUC 326	Advanced Regression Analysis	Smith, S.	Social science researchers often deal with complex data and research questions that traditional statistics models like linear regression cannot adequately address. This course offers the opportunity to understand and apply two widely used types of advanced regression analysis that allow the examination of 1) multilevel data structures (multilevel models) and 2) multivariate research questions (structural equation models).	advanced regression analysis; multilevel data structures; multilevel models; multivariate models	R (STATA if needed)	3-5	SPG	familiarity with regression and basic R programming (EDUC 423B or equiv)	Required weekly labs in R (ind/ non-graded); 2 graded projects (ind or group)
EDUC 401D	Multilevel Modeling Using R	Rogosa, D.	See http://rogosateaching.com/stat196/ . Multilevel data analysis examples using R. Topics include: two-level nested data, growth curve modeling, generalized linear models for counts and categorical data, nonlinear models, three-level analyses.	two-level nested data; growth curve modeling; generalized linear models for counts and categorical data; nonlinear models; three-level analyses	R	1	SPG	basic stats	Engage class sessions, present 10 min on relevant data analysis
EDUC 423A	Introduction to Data Science I: Data Processing	Smith, S.	Quantitative data require considerable work before they are ready to be analyzed: they are often messy, incomplete and potentially biased. This course is designed to help you thoughtfully collect, manage, clean and represent data so it can offer substantive information researchers can act upon. In our weekly sessions you will take a critical and reflective approach to these tasks and learn the technical skills needed to get your data into shape. Education and social science datasets will be our focus.	collect data; manage data; clean data; represent data	R	3-4	AUT	None	Required weekly labs in R (ind/ non-graded); 3 ind graded problem sets; 1 group assignment
EDUC 423B	Introduction to Data Science II: Machine learning	Smith, S.	This course centers on the question of how you can use various data science techniques to understand social phenomena. Applied to education and social science topics, the course will introduce you to supervised and unsupervised machine learning algorithms, new data, and provide you the skills to thoughtfully evaluate and assess machine learning performance and implications.	supervised machine learning; unsupervised machine learning; evaluate and assess machine learning	R	3-4	WIN	basic stats and programming (EDUC 423A or equiv)	Required weekly labs in R (ind / non-graded); 3 ind graded problem sets; 1 group assignment

EDUC 452	Simulation in Education Research	Domingue, B.	Simulation is a valuable tool for understanding the structure of data. We will use simulation to study three classic educational research datasets: data from an experimental educational intervention, administrative data used to understand the role of schools and teachers, and item response data collected to understand students abilities. We will discuss the underlying rationale for the data collection and then use simulation to understand the statistical models used to analyze the data and the real-world implications of the data.	simulation designs; data management and processing	R	3	SPG	basic stats and programming	TBD
EDUC 463	Computer Vision for Education and Social Science Research	Haber, N.	Computer vision -- the study of how to design artificial systems that can perform high-level tasks related to image or video data (e.g. recognizing and locating objects in images and behaviors in videos) -- has seen recent dramatic success. In this course, we seek to give education and social science researchers the know-how needed to apply cutting edge computer vision algorithms in their work as well as an opportunity to workshop applications. Prerequisite: python familiarity and some experience with data.	computer vision; machine learning	Python	3	SPG	basic stats and programming	Individual project with milestones, iPython notebook
EDUC 464	Measuring Learning in the Brain	Yeatman, J.	Everything we learn - be it a historical fact, the meaning of a new word, or a skill like reading, math, programming or playing the piano - depends on brain plasticity. The human brain's incredible capacity for learning is served by a variety of learning mechanisms that all result in changes in brain structure and function over different time scales. The goal of this course is to (a) provide an overview of different learning systems in the brain, (b) introduce methodologies and experiments that have led to new discoveries linking human brain plasticity and learning, (3) design an experiment, collect neuroimaging data, and measure the neurobiological underpinnings of learning in your own brain with MRI. The first section of the course will involve a series of lectures and discussions on the foundations of plasticity and learning with particular attention to experimental methods used in human neuroimaging studies. The second part of the course will involve workshops on designing and implementing experiments in MATLAB/Psychtoolbox or Python/PsychoPy. During this part of the course students will design, present and implement their own experiments as group projects. Finally, students will learn how to collect and analyze MRI data by being participants in their own fMRI experiments or analyzing publicly available datasets. Requirements: This class is designed for students who are interested in gaining hands-on experience with measuring the neurobiological underpinnings of learning. Student projects will involve designing experiments, collecting and analyzing data. So some experience with MATLAB/Python or an equivalent programming language is required. Some background in neuroscience (at least 1 course) is also required as we will assume basic knowledge.	experiment design; neuroimaging data collection; measurement; designing and implementing experiments; collecting and analyzing MRI data	R, Python, MATLAB	3	SPG	basic programming and prior neuroscience course taking	Individual and group projects (proposal, writeup, and presentation)

Quantitative Methods Workshops

The GSE offers workshops focused on developing further skills in quantitative methods. These workshops are not required but they can be useful for developing methodological skills and learning more advanced quantitative methods. Students are welcome to explore these with instructor permission.

Course#	Course Name	Instructors	Description	Methods taught	Language	Units	QTR	Prerequisites
EDUC 259	Education Data Science Seminar	Smith, S.; McFarland, D.	Professional development and training seminar for masters students in EDS.	Data science techniques	R, Python	1	AUT; WIN; SPG	Instructor permission
EDUC 317	Computational Sociology	Hoffman, M.; McFarland, D.	For doctoral students. Professional development; engagement in full research cycle culminating in presentation and submission of publishable work; use of computational techniques and methods. Yearlong workshop where doctoral students are encouraged to collaborate with peers and faculty who share an interest in employing computational techniques in the pursuit of researching social network dynamics, text analysis, histories, and theories of action that help explain social phenomena. Students present their own research and provide helpful feedback on others' work. Presentations may concern dissertation proposals, grants, article submissions, book proposals, datasets, methodologies and other texts. Repeatable for credit.	Any – draws on participants' expertise in many different methods	R, Python	1-2	AUT; WIN; SPG	Instructor permission (familiarity with social networks and/or NLP)
EDUC 339	Advanced Topics in Quantitative Policy Analysis	Dee, T.; Domingue, B.; pearman, f.; reardon, s.	For doctoral students. How to develop a researchable question and research design, identify data sources, construct conceptual frameworks, and interpret empirical results. Presentation by student participants and scholars in the field. May be repeated for credit.	research design; data sources; conceptual frameworks; interpret empirical results	No software application	1-2	SPG	Instructor permission

Minicourses in R, Python and Stata (*Summer fun!*)

Many students want to know if there are mini-courses they can take during summer or over break so as to learn a programming language. This is not required nor sufficient for mastering code or methodologies, but it can be helpful. Here are some possible options. Be sure to search for more as they open up all the time!

Course	Description / URL (when available)	Topics Covered	Language	Time Constraints
CME 195: Introduction to R	This short course runs for four weeks and is offered in fall and spring. It is recommended for students who want to use R in statistics, science or engineering courses, and for students who want to learn the basics of data science with R. The goal of the short course is to familiarize students with some of the most important R tools for data analysis. Lectures will focus on learning by example and assignments will be application-driven. No prior programming experience is assumed.	Many!	R	Mini-course, 4 weeks
Data Analysis Examples Using R - EDUC 401C	We will do basic and intermediate level data analysis examples, likenthose that students will have seen in their courses, in R. Examples include: descriptive statistics and plots, analysis of variance, ncorrelation and regression, categorical variables, multilevel data. See http://rogosateaching.com/ed401/	Descriptive Statistics, Group Comparisons (including anova, factorial designs); Correlation and Regression; Categorical Data, Generalized Linear Models; Overflow and Extensions	R	Original was a 5 week, live course at Stanford
STATS 32: Introduction to R for Undergraduates	This short course runs for weeks one through five of the quarter. It is recommended for undergraduate students who want to use R in the humanities or social sciences and for students who want to learn the basics of R programming. The goal of the short course is to familiarize students with R's tools for data analysis. Lectures will be interactive with a focus on learning by example, and assignments will be application-driven. No prior programming experience is needed. Topics covered include basic data structures, File I/O, data transformation and visualization, simple statistical tests, etc, and some useful packages in R. Prerequisite: undergraduate student. Priority given to non-engineering students. Laptops necessary for use in class.	Many!	R	Mini-course, 5 weeks
Basics	The first in our Professional Certificate Program in Data Science, this course will introduce you to the basics of R programming. You can better retain R when you learn it to solve a specific problem, so you'll use a real-world dataset about crime in the United States. You will learn the R skills needed to answer essential questions about differences in crime across the different states. We'll cover R's functions and data types, then tackle how to operate on vectors and when to use advanced functions like sorting. You'll learn how to apply general programming features like "if-else," and "for loop" commands, and how to wrangle, analyze and visualize data. Rather than covering every R skill you might need, you'll build a strong foundation to prepare you for the more in-depth courses later in the series, where we cover concepts like probability, inference, regression, and machine learning. We help you develop a skill set that includes R programming, data wrangling with dplyr, data visualization with ggplot2, file organization with UNIX/Linux, version control with git and GitHub, and reproducible document preparation with RStudio. The demand for skilled data science practitioners is rapidly growing, and this series prepares you to tackle real-world data analysis challenges. https://online-learning.harvard.edu/course/data-science-r-basics?delta=2	Basic R syntax; Foundational R programming concepts such as data types, vectors arithmetic, and indexing; How to perform operations in R including sorting, data wrangling using dplyr, and making plots	R	Self-Paced, 8 weeks, 1-2 hours per week
R Programming for Beginners | Complete Tutorial | R & RStudio	R programming tutorial with everything you need to know to start coding in RStats and RStudio. All the basics and fundamentals for non-coders and beginners in R programming! This is the perfect first step in your journey to master Data Science. https://www.youtube.com/watch?v=BvKETZ6kr9Q	Base R overview; RStudio overview; Variable assignment; Numerics; Logicals; Characters; Factors; Vectors; Lists; Data frames; Matrices; Indexing; Functions; Packages	R	Single video, 50 minutes
R Programming Tutorial - Learn the Basics of Statistical Computing	Learn the R programming language in this tutorial course. This is a hands-on overview of the statistical programming language R, one of the most important tools in data science. https://www.youtube.com/watch?v=BvKETZ6kr9Q	Installing R; RStudio; Packages; plot(); Bar Charts; Histograms; Scatterplots; Overlaying Plots; summary(); describe(); Selecting Cases; Data Formats; Factors; Entering Data; Importing Data; Hierarchical Clustering; Principal Components; Regression; Next Steps	R	Single video, 2 hours
Learn R	In this course, you'll be exposed to fundamental programming concepts in R. After the basics, you'll learn how to organize, modify and clean data frames, a useful data structure in R. Then you'll learn how to create data visualizations to showcase insights in data! Finish up with	Data frames; Data cleaning; ggplot2; Aggregates; Joining tables; Summary statistics; Hypothesis testing	R	Self-Paced, 20 hours, multiple lessons/chapters

	statistics and hypothesis testing to become a data analysis expert. https://www.codecademy.com/learn/learn-r			
R Programming	In this course you will learn how to program in R and how to use R for effective data analysis. You will learn how to install and configure software necessary for a statistical programming environment and describe generic programming language concepts as they are implemented in a high-level statistical language. The course covers practical issues in statistical computing which includes programming in R, reading data into R, accessing R packages, writing R functions, debugging, profiling R code, and organizing and commenting R code. Topics in statistical data analysis will provide working examples. https://www.coursera.org/learn/r-programming	Understand critical programming language concepts; Configure statistical programming software; Make use of R loop functions and debugging tools; Collect detailed information using R profiler	R	Self-Paced, Approx. 57 hours to complete
Visualization	As part of our Professional Certificate Program in Data Science, this course covers the basics of data visualization and exploratory data analysis. We will use three motivating examples and ggplot2, a data visualization package for the statistical programming language R. We will start with simple datasets and then graduate to case studies about world health, economics, and infectious disease trends in the United States. We'll also be looking at how mistakes, biases, systematic errors, and other unexpected problems often lead to data that should be handled with care. The fact that it can be difficult or impossible to notice a mistake within a dataset makes data visualization particularly important. The growing availability of informative datasets and software tools has led to increased reliance on data visualizations across many areas. Data visualization provides a powerful way to communicate data-driven findings, motivate analyses, and detect flaws. This course will give you the skills you need to leverage data to reveal valuable insights and advance your career. https://online-learning.harvard.edu/course/data-science-visualization?delta=2	Data visualization principles; How to communicate data-driven findings; How to use ggplot2 to create custom plots; The weaknesses of several widely-used plots and why you should avoid them	R	Self-Paced, 8 weeks, 1-2 hours per week
Productivity Tools	A typical data analysis project may involve several parts, each including several data files and different scripts with code. Keeping all this organized can be challenging. Part of our Professional Certificate Program in Data Science, this course explains how to use Unix/Linux as a tool for managing files and directories on your computer and how to keep the file system organized. You will be introduced to the version control systems git, a powerful tool for keeping track of changes in your scripts and reports. We also introduce you to GitHub and demonstrate how you can use this service to keep your work in a repository that facilitates collaborations. Finally, you will learn to write reports in R markdown which permits you to incorporate text and code into a document. We'll put it all together using the powerful integrated desktop environment RStudio. https://online-learning.harvard.edu/course/data-science-productivity-tools?delta=2	How to use Unix/Linux to manage your file system; How to perform version control with git; How to start a repository on GitHub; How to leverage the many useful features provided by Rstudio	R	Self-Paced, 8 weeks, 1-2 hours per week
Data Wrangling	In this course, part of our Professional Certificate Program in Data Science, we cover several standard steps of the data wrangling process like importing data into R, tidying data, string processing, HTML parsing, working with dates and times, and text mining. Rarely are all these wrangling steps necessary in a single analysis, but a data scientist will likely face them all at some point. Very rarely is data easily accessible in a data science project. It's more likely for the data to be in a file, a database, or extracted from documents such as web pages, tweets, or PDFs. In these cases, the first step is to import the data into R and tidy the data, using the tidyverse package. The steps that convert data from its raw form to the tidy form is called data wrangling. This process is a critical step for any data scientist. Knowing how to wrangle and clean data will enable you to make critical insights that would otherwise be hidden. https://online-learning.harvard.edu/course/data-science-wrangling?delta=2	Importing data into R from different file formats; Web scraping; How to tidy data using the tidyverse to better facilitate analysis; String processing with regular expressions (regex); Wrangling data using dplyr; How to work with dates and times as file formats, and text mining	R	Self-Paced, 8 weeks, 1-2 hours per week
Linear Regression	Linear regression is commonly used to quantify the relationship between two or more variables. It is also used to adjust for confounding. This course, part of our Professional Certificate Program in Data Science, covers how to implement linear regression and adjust for confounding in practice using R. In data science applications, it is very common to be interested in the relationship between two or more variables. The motivating case study we examine in this course relates to the data-driven approach used to construct baseball teams described in Moneyball. We will try to determine which measured outcomes best predict baseball runs by using linear regression. We will also examine confounding, where extraneous variables affect the relationship between two or more other variables, leading to spurious associations. Linear regression is a powerful technique for removing confounders, but it is not a magical process. It is essential to understand when it is appropriate to use, and this course will teach you when to apply this technique. https://online-learning.harvard.edu/course/data-science-linear-regression?delta=2	How linear regression was originally developed by Galton; What is confounding and how to detect it; How to examine the relationships between variables by implementing linear regression in R	R	Self-Paced, 8 weeks, 1-2 hours per week

CME 193: Introduction to Scientific Python	This short course runs for the first four weeks of the quarter. It is recommended for students who are familiar with programming at least at the level of CS106A and want to translate their programming knowledge to Python with the goal of becoming proficient in the scientific computing and data science stack. Lectures will be interactive with a focus on real world applications of scientific computing. Technologies covered include Numpy, SciPy, Pandas, Scikit-learn, and others. Topics will be chosen from Linear Algebra, Optimization, Machine Learning, and Data Science. Prior knowledge of programming will be assumed, and some familiarity with Python is helpful, but not mandatory.	Many!	Python	Mini-course, 4 weeks
CS 193Q: Introduction to Python Programming	CS193Q teaches basic Python programming with a similar end-condition to CS106AP: strings, lists, numbers, dicts, loops, logic, functions, testings, decomposition and style, and modules. CS193Q assumes knowledge of some programming language, and proceeds by showing how each common programming idea is expressed in Python. CS193Q moves very quickly, meeting 3 times for 4 hours for a total of 12 hours which is a mixture of lecture and lab time.	Many!	Python	Mini-course, 3 meetings, 12 hours
Learn Python - Full Course for Beginners [Tutorial]	This course will give you a full introduction into all of the core concepts in python. Follow along with the videos and you'll be a python programmer in no time! https://www.youtube.com/watch?v=rfscVS0vtbw	Installing Python & PyCharm; Setup & Hello World; Drawing a Shape; Variables & Data Types; Working With Strings; Working With Numbers; Getting Input From Users; Building a Basic Calculator; Mad Libs Game; Lists; List Functions; Tuples; Functions; Return Statement; If Statements; If Statements & Comparisons; Building a better Calculator; Dictionaries; While Loop; Building a Guessing Game; For Loops; Exponent Function; 2D Lists & Nested Loops; Building a Translator; Comments; Try / Except; Reading Files; Writing to Files; Modules & Pip; Classes & Objects; Building a Multiple Choice Quiz; Object Functions; Inheritance; Python Interpreter	Python	Single video, 4.5 hours
Python Tutorial for Beginners - Learn Python in 1 Hour	This Python tutorial for beginners show how to get started with Python quickly. Learn to code in 1 hour! Watch this tutorial get started! https://www.youtube.com/watch?v=kqtD5dpn9C8	Introduction; What You Can Do With Python; Your First Python Program; Variables; Receiving Input; Type Conversion; Strings; Arithmetic Operators; Operator Precedence; Comparison Operators; Logical Operators; If Statements; Exercise; While Loops; Lists; List Methods; For Loops; The range() Function; Tuples	Python	Single video, 1 hour
Python As Fast as Possible - Learn Python in ~75 Minutes	This python tutorial aims to teach you python as fast as possible. This python speed course will cover all the fundamentals of python and give you a quick overview of all of the main python features. https://www.youtube.com/watch?v=VchuKL44s6E	Introduction; Setup & Installation; What Python is Used For; Data Scientist Master's Program; Data Types; Output & Printing; Variables; User Input; Arithmetic Operator; String Methods; Conditional Operators; Chained Conditionals; If/Else/Elif; List/Tuples; For Loops; While Loops; Slice Operator; Sets; Dicts; Comprehensions; Functions; *args & **kwargs; Scope & Globals; Exceptions; Handling Exceptions; Lambda; Map and Filter; F Strings	Python	Single video, 1.25 hours
Introduction to Computer Science and Programming Using Python	This course is the first of a two-course sequence: Introduction to Computer Science and Programming Using Python, and Introduction to Computational Thinking and Data Science. Together, they are designed to help people with no prior exposure to computer science or programming learn to think computationally and write programs to tackle useful problems. Some of the people taking the two courses will use them as a stepping stone to more advanced computer science courses, but for many it will be their first and last computer science courses. This run features lecture videos, lecture exercises, and problem sets using Python 3.5. Even if you previously took the course with Python 2.7, you will be able to easily transition to Python 3.5 in future courses, or enroll now to refresh your learning. Since these courses may be the only formal computer science courses many of the students take, we have chosen to focus on breadth rather than depth. The goal is to provide students with a brief introduction to many topics so they will have an idea of what is possible when they need to think about how to use computation to accomplish some goal later in their career. That said, they are not "computation appreciation" courses. They are challenging and rigorous courses in which the students spend a lot of time and effort learning to bend the computer to their will. https://www.edx.org/course/introduction-to-computer-science-and-programming-7	A Notion of computation; The Python programming language; Some simple algorithms; Testing and debugging; An informal introduction to algorithmic complexity; Data structures	Python	Self-Paced, 9 weeks, 14-16 hours per week
Learn STATA in 15 minutes	The goal of this video is to get you up and running using STATA for data analysis in just 15 minutes. https://www.youtube.com/watch?v=rdFw-fBfygQ	***Very basic	STATA	Single video, 15 minutes

Getting Started in STATA - PLAYLIST	Want to know how to conduct a basic data analysis using Stata? We show you just that. Watch as we show you how to import data into Stata from Excel, how to change the display format of a variable, how to label variables and the values of categorical variables, how to create new variables, how to describe and summarize variables, and how to create several kinds of graphs. Next we demonstrate how to fit a linear regression model, how to calculate and graph marginal predictions from the model, and how to send those commands to Stata's Do-file editor. The video ends with a discussion of how to learn more about Stata using the Help files, PDF documentation, and the -search- command. https://www.youtube.com/watch?v=FQ1MBQw_MTI&list=PLN5IskQdgXWnHC_5-ebmFZUNdpKcoLtDT	Copy and paste data from Excel to Stata; Import Excel data to Stata; Change the display format of a variable; Define and add value labels to a variable; Create a new variable; Label a variable; Describe a dataset; Summarize a dataset; Create a histogram; Create a pie chart; Create a boxplot; Fit a linear regression model; Calculate marginal predictions; Create interaction plots; Creating do-files for reproducible analyses; How to open a Help File; How to open the PDF documentation; How to search for topics	STATA	Playlist, 3 videos, 30 minutes total
STATA Quick Tips Playlist	NA. https://www.youtube.com/watch?v=-brVdCA7NYs&list=PLN5IskQdgXWmJxPDydvhuRpdLA2H5VEUo	Drag and drop; Date function; Finding duplicates; Partial dataset; Project manager; Margins	STATA	Playlist, 6 videos, 20 minutes