MA in Computational Social Science

Faculty Director

- James Evans, Sociology

Executive Committee

- Luc Anselin, Sociology
- Marc G. Berman, Psychology
- Kathleen Cagney, Sociology
- Guanglei Hong, Comparative Human Development
- Ali Hortaçsu, Economics
- Leslie M. Kay, Psychology
- Howard Nusbaum, Psychology
- John Padgett, Political Science
- Stephen W. Raudenbush, Sociology

Affiliated Faculty

- Stéphane Bonhomme, Economics
- Magne Mogstad, Economics
- Anna Mueller, Comparative Human Development
- James T. Sparrow, History
- Alessandra Voena, Economics
- Daniel Yurovsky, Psychology

Associate Director and Senior Lecturer

- Rick Evans

Assistant Director and Lecturer

- Benjamin Soltoff

Managing Director

- Chad Cyrenne

Director of Career Services

- Shelly Robinson

Career Services Coordinator

- Gözde Erdeniz

Employer Relations Manager

- Jeffrey Shoemaker

Student Affairs Administrator

- E.G. Enbar

Alumni, Staff, and Student Programming Administrator

- Stefani Metos

Business Administrator

- Tekeisha Yelton-Hunter

General Information

The Master of Arts in Computational Social Science is a two-year program of graduate study. It has a structured curriculum, with a total of 18 required and elective courses tailored to the disciplinary track a student follows. Students submit an article-length MA thesis in their second year, after completing a three-quarter research commitment working directly with a member of our Executive or Affiliated Faculty.

The program aims to produce leading social scientists in each of our core social science fields – economics, sociology, political science, psychology, history, and anthropology – producing competitive PhD applicants, well-
trained in computational approaches, who have mastered the research and analytical skills necessary to make important contributions.

Students receive close mentorship from the program’s Faculty Director, academic staff, and members of our Executive and Affiliated Faculty.

They receive full professional support from our Director of Career Services, with biweekly workshops, career planning, and employer recruitment.

Finally, all MA students may participate in an optional summer practicum between their first and second year, with internships drawn from academic and professional organizations. International students have three years of STEM work eligibility after they graduate.

**PROGRAM REQUIREMENTS AND COURSE WORK**

Students submit an article-length MA thesis in their second year, after completing a three-quarter research commitment working directly with a member of our Computation Faculty.

The courses are selected with the advice of our academic staff, and follow different disciplinary tracks, tailored to the research commitments of each student.

In their first year, all students take a three course core: Perspectives on Computational Analysis, Perspectives on Computational Modeling, and Perspectives on Computational Research.

Most take a three course sequence on Computer Science with Applications (with more advanced courses for students with prior exposure, and an optional sequence for psychology concentrators).

The remaining three courses vary, and depend on the student’s prior training and disciplinary path. Priority will go to any needed courses in statistics, linear algebra, or advanced math in particular disciplines (e.g. real analysis in economics). If those requirements are met, the student will take up to three social science electives in their area of research.

In their second year, all students complete a three course “research commitment,” producing an MA thesis modeled on a professional journal article. They take three advanced courses in computational methods, tailored to their disciplinary interest. And they complete three social science electives, in their area of research.

If students desire, they can petition to replace any portion of the three quarter research commitment with social science electives or other courses in computational methods.

Outside of their coursework, all MA students are expected to attend our weekly Computation Workshop, where advanced scholars and invited guests present drafts of their research for critique and discussion.

**ADMISSION**

MACSS applicants must meet the formal requirements of the Graduate Social Sciences Division.

All applicants must submit GRE scores, except for those applying for the joint BA/MA degree.

All financial aid is merit-based, and MACSS offers partial and full tuition scholarships at the time of admission.

Joint BA/MA applicants pay graduate tuition rates, and are eligible to receive the same aid they had in the College.

Applicants from non-English speaking countries must provide evidence of English proficiency by submitting scores from either the Test of English as a Foreign Language (TOEFL) or the International English Language Testing System (IELTS).

Some non-native English speakers are exempt, if they have studied in an English language University. Please contact our Dean of Students Office with any questions: ssd-admissions@uchicago.edu

**HOW TO APPLY**

The Application for Admission and Financial Aid, with instructions and deadlines, is available online at: https://apply-ssd.uchicago.edu/apply/.

For additional information about our program, please contact E.G. Enbar, our Student Affairs Administrator, at 773-702-8312 or egenbar@uchicago.edu.

Please also visit our website: https://macss.uchicago.edu (https://macss.uchicago.edu/)
COURSES

MACS 30000. Perspectives on Computational Analysis. 100 Units.
Massive digital traces of human behavior and ubiquitous computation have both extended and altered classical social science inquiry. This course surveys successful social science applications of computational approaches to the representation of complex data, information visualization, and model construction and estimation. We will reexamine the scientific method in the social sciences in context of both theory development and testing, exploring how computation and digital data enables new answers to classic investigations, the posing of novel questions, and new ethical challenges and opportunities. Students will review fundamental research designs such as observational studies and experiments, statistical summaries, visualization of data, and how computational opportunities can enhance them. The focus of the course is on exploring the wide range of contemporary approaches to computational social science, with practical programming assignments to train with these approaches.
Instructor(s): Benjamin Soltoff Terms Offered: Autumn
Note(s): MACSS students have priority. Others admitted with instructor consent.

MACS 30100. Perspectives on Computational Modeling. 100 Units.
Students are often well trained in the details of specific models relevant to their respective fields. This course presents a generic definition of a model in the social sciences as well as a taxonomy of the wide range of different types of models used. We then cover principles of model building, including static versus dynamic models, linear versus nonlinear, simple versus complicated, and identification versus overfitting. Major types of models implemented in this course include linear and nonlinear regression, machine learning (e.g., parametric, Bayesian and nonparametric), agent-based and structural models. We will also explore the wide range of computational strategies used to estimate models from data and make statistical and causal inference. Students will study both good examples and bad examples of modeling and estimation and will have the opportunity to build their own model in their field of interest.
Instructor(s): Rick Evans Benjamin Soltoff Terms Offered: Winter
Prerequisite(s): MACSS students have priority. Others admitted with instructor consent.

MACS 30123. Large-Scale Computing for the Social Sciences. 100 Units.
Computational social scientists increasingly need to grapple with data that is either too big for a single machine and/or code that is too resource intensive to process on a single machine. In this course, students will learn how to effectively scale their computational methods beyond their local machines. The focus of the course will be social scientific applications, ranging from training machine learning models on large economic time series to processing and analyzing social media data in real-time. Students will be introduced to several large-scale computing frameworks such as MPI, MapReduce, Spark, and OpenCL, with a special emphasis on employing these frameworks using cloud resources and the Python programming language.
Instructor(s): Jonathan Clindaniel Terms Offered: Spring

MACS 30150. Perspectives on Computational Modeling for Economics. 100 Units.
Students are often well trained in the details of specific models relevant to their respective fields. This course presents a generic definition of a model in the social sciences as well as a taxonomy of the wide range of different types of models used. We then cover principles of model building, including static versus dynamic models, linear versus nonlinear, simple versus complicated, and identification versus overfitting. Major types of models implemented in this course include linear and nonlinear regression, machine learning (e.g., parametric, Bayesian and nonparametric), agent-based and structural models. We will also explore the wide range of computational strategies used to estimate models from data and make statistical and causal inference. Students will study both good examples and bad examples of modeling and estimation and will have the opportunity to build their own model in their field of interest.
Instructor(s): R. Evans Terms Offered: Winter
Note(s): MACSS students have priority.

MACS 30200. Perspectives on Computational Research. 100 Units.
This course focuses on applying computational methods to conducting social scientific research through a student-developed research project. Students will identify a research question of their own interest that involves a direct reference to social scientific theory, use of data, and a significant computational component. The students will collect data, develop, apply, and interpret statistical learning models, and generate a fully reproducible research paper. We will identify how computational methods can be used throughout the research process, from data collection and hiding, to exploration, visualization and modeling, to the final communication of results. The course will include modules on theoretical and practical considerations, including topics such as epistemological questions about research design, writing and critiquing papers, and additional computational tools for analysis.
Instructor(s): Richard Evans Benjamin Soltoff Terms Offered: Spring
Prerequisite(s): MACSS students have priority. Others admitted with instructor consent.
MACS 30250. Perspectives on Computational Research for Economics. 100 Units.
This course focuses on scaling up computational approaches to social science analysis and modeling with big data in context of opportunities afforded by high performance and cloud computing. We will begin by exploring various data structures encountered in social science research, how to deal with large or complex data storage and streaming data, and how to factor considerations of computational complexity into their analyses. We will also study social science applications of parallel computing, both on stand-alone machines and in supercomputing environments, to carry out complex computations. Students will learn to carry out parallel I/O and parallel computation on their own machines and on a cluster. We will also address API construction and access, and explore cloud configurations for social science research designs. We will also help students construct web-based outward facing data, analysis and visualization portals. Students will efficiently gather, structure, perform and present analysis on large-scale social science data. This course will be specifically tailored to students concentrating in Economics.
Instructor(s): R. Evans Terms Offered: Spring
Prerequisite(s): MACSS students have priority.

MACS 30301. Introduction to Bayesian Statistics. 100 Units.
The goal of this course is to give students an overview of the theory and methods for data analyses using the Bayesian paradigm. Topics include: (1) foundations of Bayesian inference, (2) development of Bayesian models and prior choices (3) analytical and simulation techniques for posterior estimation (4) model choice and diagnostics (5) sensitivity analysis, (6) an introduction to Monte Carlo Markov Chain (MCMC) simulations (7) intro to commonly used Bayesian estimation packages (R/JAGS/Bugs) (8) application of Bayesian analysis in real world and Political Science problems.
Instructor(s): Diogo Ferrari Terms Offered: Autumn
Equivalent Course(s): MAPS 30301

MACS 30501. Computational Anthropology. 100 Units.
This course exposes students to the methods and data of Computational Anthropology—the systematic, computational study of the human species, past and present. Such methods have been essential in recent years for simulating human behavior in different cultures and economic systems, uncovering ancient demographic changes that still have an influence into the present day, preserving cultural heritage, and much more. Anthropological data allows social science researchers to evaluate long term trends in the human condition, across a variety of cultures, with a unique combination of material, textual, and structured database data. Students will have the opportunity to evaluate state of the art approaches in computational anthropology and learn how to apply these methods to their own social scientific research agendas using open anthropological datasets and the Python programming language.
Instructor(s): Jonathan Clindaniel Terms Offered: Winter

MACS 33000. Computational Math Camp. 000 Units.
MACS 33001. Mathematics and Statistics for Computational Social Science. 100 Units.
This course aims to provide students with a core understanding of mathematics and statistics for computational social science. Students who complete this course should be prepared to take more advanced computational methods courses. Completion of the Computational Math Camp in September is recommended, but not required.
Instructor(s): Staff Terms Offered: Autumn

MACS 33002. Introduction to Machine Learning. 100 Units.
This course will introduce students to the foundations of machine learning. We will cover everything needed for getting up and running with any computational research project from a machine learning perspective, including the key concepts: classification, regression, and dimensionality reduction. Students will learn about the fundamental mathematical concepts underlying machine learning algorithms, but this course will equally focus on the practical use of machine learning using open source statistical computing (e.g., Python, R, Julia). We will cover a variety of topics, including: the mechanics of a model, (re)sampling methods, model training, testing and tuning, supervised vs. unsupervised learning, regularization, tree-based methods, and several other algorithms contributing to a solid foundation of inferential machine learning for addressing real social problems.
Instructor(s): Philip Waggoner Terms Offered: Winter
Equivalent Course(s): MAPS 33002, PLSC 43505

MACS 35000. MA Research Commitment. 100 Units.
Student Initiated research and writing for the MA research component.
Instructor(s): James Evans

MACS 35001. Structured MA Research Commitment. 100 Units.
Student initiated research and writing for the MA research component.
Instructor(s): James Evans Terms Offered: Autumn
MACS 40000. Economic Policy Analysis with Overlapping Generation Models. 100 Units.
The course will study economic policy questions ideally addressed by the overlapping generations (OG) dynamic general equilibrium framework. OG models represent a rich class of macroeconomic general equilibrium model that is extremely useful for answering questions in which inequality, demographics, and individual heterogeneity are important. OG models are used extensively by the Joint Committee on Taxation, Congressional Budget Office, and Department of the Treasury. This course will train students how to set up and solve OG models. The standard nonlinear global solution method for these models—time path iteration—is a fixed point method that is similar to but significantly different from value function iteration. This course will take students through progressively richer versions of the model, which will include endogenous labor supply, nontrivial demographics, bequests, stochastic income, multiple industries, non-balanced government budget constraint, and household tax structure.
Instructor(s): Jonathan Clindaniel Terms Offered: Autumn

MACS 40100. Big Data and Society. 100 Units.
The massive explosion of information produced by computers and sophisticated computational methods capable of harnessing this data to generate inferences has led to an increasingly data-driven society. Businesses, governments, and individuals seek to leverage this data to develop and market products, formulate policy, and improve the human condition. Computational approaches to decision making have become increasingly prevalent in domains such as criminal justice, education, employment, finance, and politics. While decision making based on data mining and algorithms has the capacity to improve society, critics argue that these approaches strengthen socioeconomic class divisions, constitute an invasion of privacy, or violate the civil rights of minority groups. This course will survey some of the major uses of big data in society and assess the potential ethical, moral, and legal implications of these models.
Instructor(s): R. Evans Terms Offered: Winter
Prerequisite(s): MACSS students have priority. Others admitted with instructor consent.

MACS 40200. Structural Estimation. 100 Units.
Structural estimation refers to the estimation of model parameters by taking a theoretical model directly to the data. (This is in contrast to reduced form estimation, which often entails estimating a linear model that is either explicitly or implicitly a simplified, linear version of a related theoretical model.) This class will survey a range of structural models, then teach students estimation approaches including the generalized method of moments approach and maximum likelihood estimation. We will then examine the strengths and weaknesses of both approaches in a series of examples from the fields of economics, political science, and sociology. We will also learn the simulated method of moments approach. We will explore applications across the social sciences.
Instructor(s): Richard Evans Terms Offered: Winter

MACS 40300. Open Research Methods. 100 Units.
The purpose of this course is to give students experience in the broad set of skills and tools for managing, collaborating on, and contributing to open source research projects. Transparency and replicability of research have received renewed emphasis in recent years due to the increased prevalence and sophistication of empirical and computational methods as well as the increased availability of large high frequency data sources. This course focuses on the open source programming languages of Python and R, but the principles could be applied to projects using any language. The course will present the common open source software development workflow as an efficient structure for collaborative academic research. We will learn Git and GitHub basic tools and methods. We will practice multiple levels of documentation ranging from in-code docstrings to full PDF and HTML documentation tools. Students will implement continuous integration testing and regression testing in their own open source repositories. And students will learn how to set an environment with specific library and package versions. We will also discuss methods for anonymizing proprietary data or creating synthetic datasets that can be used by the general public.
Instructor(s): R. Evans Terms Offered: Autumn

MACS 40400. Computation and the Identification of Cultural Patterns. 100 Units.
Culture is increasingly becoming digital, making it more and more necessary for those in both academia and industry to use computational strategies to effectively identify, understand, and (in the case of industry) capitalize on emerging cultural patterns. In this course, students will explore interdisciplinary approaches for defining and mobilizing the concept of “culture” in their computational analyses, drawing on relevant literature from the fields of Anthropology, Psychology and Sociology. Additionally, they will receive hands-on experience applying computational approaches to identify and analyze a wide range of cultural patterns using the Python programming language. For instance, students will learn to identify emerging social movements using social media data, predict the next fashion trends, and even decipher ancient symbols using archaeological databases.
Instructor(s): Jonathan Clindaniel Terms Offered: Autumn
Prerequisite(s): No previous coding experience required. A Python boot camp will be held at the beginning of the quarter to teach the coding skills necessary to succeed in the course. Open to Advanced Undergraduates with Instructor Permission.
Equivalent Course(s): MAPS 40401, CHDV 40404, PSYC 40460
MACS 40500. Computational Methods for American Politics. 100 Units.
In this class, students will be introduced to several computational techniques aimed at exploring, understanding, and diagnosing substantive American political phenomena. Rather than focus on derivations and proofs of models, the main focus of the course will be applying and diagnosing model fit, along with computation and application in R. The goal of the class is twofold: first, to offer students a methodological toolbox to tackle complex questions of interest in the social sciences. The second goal, then, is to prepare students for applied quantitative research, offering modern data science techniques and computational training in the service of understanding and predicting American political behavior in a range of contexts. The course will be a combination seminar/applied, where we will read and discuss the latest developments as well as classical works related to a week’s topic, but also apply the concepts in R.
Instructor(s): Philip Waggoner Terms Offered: Autumn
Equivalent Course(s): MAPS 40500, PLSC 40525, PLSC 20525

MACS 40700. Data Visualization. 100 Units.
Social scientists frequently wish to convey information to a broader audience in a cohesive and interpretable manner. Visualizations are an excellent method to summarize information and report analysis and conclusions in a compelling format. This course introduces the theory and applications of data visualization. Students will learn techniques and methods for developing rich, informative and interactive, web-facing visualizations based on principles from graphic design and perceptual psychology. Students will practice these techniques on many types of social science data, including multivariate, temporal, geospatial, text, hierarchical, and network data. These techniques will be developed using a variety of software implementations such as R, ggplot2, D3, and Tableau.
Instructor(s): Benjamin Soltoff Terms Offered: Spring

MACS 40800. Unsupervised Machine Learning. 100 Units.
Though armed with rich datasets, many researchers are confronted with a lack of understanding of the structure of their data. Unsupervised machine learning offers researchers a suite of computational tools for uncovering the underlying, non-random structure that is assumed to exist in feature space. This course will cover prominent unsupervised machine learning techniques such as clustering, item response theory (IRT) models, multidimensional scaling, factor analysis, and other dimension reduction techniques. Further, mechanics involved in unsupervised machine learning will also be covered, such as diagnosing clusterability of a feature space (visually and mathematically), measures of distance and distance matrices, different algorithms based on data size (k-medoids/k-means vs. PAM vs. CLARA), visualizing patterns, and methods of validation (e.g., internal vs. external validation).
Instructor(s): Philip Waggoner Terms Offered: Autumn
Equivalent Course(s): MAPS 40800, PLSC 40825, PLSC 20825

MACS 41200. Advanced Machine Learning. 100 Units.
This is an intermediate-to-advanced introduction to the mathematical and computational aspects of the core statistical and machine learning techniques. The goal is to equip students with a knowledge of the theoretical and practical aspects of four groups of machine learning methods which are widely used in applied research: (1) dimension reduction (PCA, MDS, and their extensions) (2) classification methods (SVM, Bayes classifiers, and other classification methods) (3) clustering procedures (K-means, FMM, non- and semi-parametric Bayesian methods) (4) categorical data analysis (with brief introduction to probabilistic graphical models). The course includes some applications in Political Science, such as FMM to estimate fraud in elections, PCA to construct indices to measure democracy, and text classification.
Instructor(s): Diogo Ferrari Terms Offered: Spring

MACS 41300. Computational Methods for Comparative Politics. 100 Units.
Comparative Politics is one of the most traditional areas in Political Science. In this course, students are exposed to some of the methodological challenges of studying politics from a comparative perspective. The course draws on canonical substantive and methodological debates in Comparative Politics and discusses some modern machine learning, latent variable analysis, and computational methods to overcome some of those difficulties. With instructor guidance, students will have the opportunity to develop their project and apply computational methods to study a topic of their choice in comparative politics.
Instructor(s): Diogo Ferrari Terms Offered: Spring
MACS 51000. Introduction to Causal Inference. 100 Units.
This course is designed for graduate students and advanced undergraduate students from the social sciences, education, public health science, public policy, social service administration, and statistics who are interested in quantitative research and are interested in studying causality. The goal of this course is to equip students with basic knowledge of and analytic skills in causal inference. Topics for the course will include the potential outcomes framework for causal inference; experimental and observational studies; identification assumptions for causal parameters; potential pitfalls of using ANCOVA to estimate a causal effect; propensity score based methods including matching, stratification, inverse-probability-of-treatment-weighting (IPTW), marginal mean weighting through stratification (MMWS), and doubly robust estimation; the instrumental variable (IV) method; regression discontinuity design (RDD) including sharp RDD and fuzzy RDD; difference in difference (DID) and generalized DID methods for cross-section and panel data, and fixed effects model. Intermediate Statistics or equivalent such as STAT 224/PBHS 324, PP 31301, BUS 41100, or SOC 30005 is a prerequisite. This course is a prerequisite for "Advanced Topics in Causal Inference" and "Mediation, moderation, and spillover effects."
Instructor(s): G. Hong, K. Yamaguchi Terms Offered: Winter
Prerequisite(s): Intermediate Statistics or equivalent such as STAT 224/PBHS 324, PP 31301, BUS 41100, or SOC 30005
Note(s): CHDV Distribution: M; M
Equivalent Course(s): PLSC 30102, CHDV 30102, SOCI 30315, PBHS 43201, STAT 31900

MACS 54000. Introduction to Spatial Data Science. 100 Units.
Spatial data science consists of a collection of concepts and methods drawn from both statistics and computer science that deal with accessing, manipulating, visualizing, exploring and reasoning about geographical data. The course introduces the types of spatial data relevant in social science inquiry and reviews a range of methods to explore these data. Topics covered include formal spatial data structures, geovisualization and visual analytics, rate smoothing, spatial autocorrelation, cluster detection and spatial data mining. An important aspect of the course is to learn and apply open source software tools, including R and GeoDa.
Instructor(s): L. Anselin and M. Kolak Terms Offered: Autumn
Prerequisite(s): STAT 22000 (or equivalent), familiarity with GIS is helpful, but not necessary
Equivalent Course(s): ENST 20510, GEOG 30500, SOCI 20253, GEOG 20500, SOCI 30253

MACS 55000. Spatial Regression Analysis. 100 Units.
This course covers statistical and econometric methods specifically geared to the problems of spatial dependence and spatial heterogeneity in cross-sectional data. The main objective of the course is to gain insight into the scope of spatial regression methods, to be able to apply them in an empirical setting, and to properly interpret the results of spatial regression analysis. While the focus is on spatial aspects, the types of methods covered have general validity in statistical practice. The course covers the specification of spatial regression models in order to incorporate spatial dependence and spatial heterogeneity, as well as different estimation methods and specification tests to detect the presence of spatial autocorrelation and spatial heterogeneity. Special attention is paid to the application to spatial models of generic statistical paradigms, such as Maximum Likelihood, Generalized Methods of Moments and the Bayesian perspective. An important aspect of the course is the application of open source software tools such as R, GeoDa and PySal to solve empirical problems.
Instructor(s): Staff Terms Offered: Spring
Equivalent Course(s): ENST 20510, GEOG 30500, SOCI 20253, GEOG 20500, SOCI 30253

MACS 60000. Computational Content Analysis. 100 Units.
A vast expanse of information about what people do, know, think, and feel lies embedded in text, and more of the contemporary social world lives natively within electronic text than ever before. These textual traces range from collective activity on the web, social media, instant messaging and automatically transcribed YouTube videos to online transactions, medical records, digitized libraries and government intelligence. This supply of text has elicited demand for natural language processing and machine learning tools to filter, search, and translate text into valuable data. The course will survey and practically apply many of the most exciting computational approaches to text analysis, highlighting both supervised methods that extend old theories to new data and unsupervised techniques that discover hidden regularities worth theorizing. These will be examined and evaluated on their own merits, and relative to the validity and reliability concerns of classical content analysis, the interpretive concerns of qualitative content analysis, and the interactional concerns of conversation analysis. We will also consider how these approaches can be adapted to content beyond text, including audio, images, and video. We will simultaneously review recent research that uses these approaches to develop social insight by exploring (a) collective attention and reasoning through the content of communication; (b) social relationships through the process of communication; and (c) social state
Instructor(s): James Evans Terms Offered: Winter
Equivalent Course(s): SOCI 40133, CHDV 30510
MACS 95000. Computation MA Internship. 000 Units.
All MACS students participating in the The Computational Social Science Internship Program will be required to enroll in this non-credit summer quarter field research course. The course will appear on the transcript, and will be evaluated on a pass/fail basis, in consultation with the employer.
Instructor(s): James Evans