

# COMPUTATIONAL SOCIAL SCIENCE

Program Website: <https://macss.uchicago.edu/programs-of-study/college-minor> (<https://macss.uchicago.edu/programs-of-study/college-minor/>)

## MINOR PROGRAM IN COMPUTATIONAL SOCIAL SCIENCE

Computational Social Science is an emergent field at the intersection of the social and computational sciences—leveraging computational advances, such as those in AI and machine learning, as well as the widespread availability of “big data,” to drive the next generation of social research.

The Minor in Computational Social Science enables students in the social sciences to build fundamental computational skills with the purpose of enriching existing approaches to social scientific inquiry. Additionally, for students outside of the social sciences (e.g., in the physical and biological sciences), it provides an opportunity to consolidate their computational, mathematical, and statistical skills, as well as apply these skills to questions about social and cultural life.

Over the course of the program, students will learn how computational approaches complement and expand upon traditional research designs in the social sciences, as well as gain the hands-on computer programming and data analysis skills necessary to implement these computationally enhanced research designs. For instance, students will learn how “big data” and AI technologies enhance traditional observational studies and explore simulation-based research designs, as well as investigate how to run large, computationally enhanced surveys and digital experiments. They will additionally have the opportunity to explore advanced approaches for applying their computational and analytical skills to social science problems in elective courses in Computational Social Science. For example, students may complete machine learning and high-performance computing coursework, with an emphasis on applicability to social scientific research.

## PROGRAM REQUIREMENTS

The minor consists of six courses: one course introducing contemporary approaches for performing Computational Social Science research, one data analysis course, two computational methods courses, and two electives in Computational Social Science.

Courses in the minor must be taken for quality grades and may not be double counted with the student’s major(s), other minors, or general education requirements.

Students who wish to declare a minor in Computational Social Science must reach out to the program contact listed below to indicate their intention to complete the minor and have their Consent to Complete a Minor Program ([https://humanities-web.s3.us-east-2.amazonaws.com/college-prod/s3fs-public/documents/Consent\\_Minor\\_Program.pdf](https://humanities-web.s3.us-east-2.amazonaws.com/college-prod/s3fs-public/documents/Consent_Minor_Program.pdf)) form signed. Students will need to submit the signed form to their College adviser before the end of the Spring Quarter of their third year.

Students interested in receiving email updates with information about courses, events, and opportunities relevant to the Computational Social Science Minor Program can sign up to join our listhost ([https://docs.google.com/forms/d/e/1FAIpQLSfFgR4ni9GknDP-q3zEx9b9Cm\\_mEcMonfm5O7C1r3PVOfx8Zg/viewform/](https://docs.google.com/forms/d/e/1FAIpQLSfFgR4ni9GknDP-q3zEx9b9Cm_mEcMonfm5O7C1r3PVOfx8Zg/viewform/)). Students are also encouraged to reach out to the program contact listed below to meet and discuss course plans, as well as obtain advice relevant to their interests.

## SUMMARY OF REQUIREMENTS

MACS 10000	Perspectives on Computational Analysis	100
STAT 22000	Statistical Methods and Applications (or approved equivalent coursework in data analysis)	100
MACS 10111 & MACS 10112	Principles of Computing 1: Computational Thinking for Social Scientists and Principles of Computing 2: Data Management for Social Scientists (or approved equivalent coursework in computational methods)	200
Two Computational Social Science Electives (any two courses with an MACS course number at the 20000-level or below)		200
<b>Total Units</b>		<b>600</b>

\* Students who take STAT 22000 or equivalent coursework for their major, another minor, or general education requirement will complete a five-course (500-unit) minor.

## APPROVED EQUIVALENT COURSEWORK IN DATA ANALYSIS

If students have not already completed a statistics or data analysis course for their major, another minor, or general education requirement, they must complete such a course for the Computational Social Science minor. Specifically, students should complete STAT 22000 Statistical Methods and Applications or equivalent coursework (at the same or higher level of mathematical requirements).

Examples of approved equivalent coursework include the following. Students should complete the course specific to their major, other minor, or general education requirements.

- Higher-level foundational STAT courses:
  - STAT 23400 Statistical Models and Methods
  - STAT 24400 Statistical Theory and Methods I
- DATA 11900 Introduction to Data Science II
- Other statistics offerings within the Social Sciences Division:
  - CHDV 20101 Applied Statistics in Human Development Research
  - ECON 11010 Introduction to Statistical Methods in Economics
  - ECON 21010 Statistical Methods in Economics
  - PSYC 20250 Introduction to Statistical Concepts and Methods
  - SOCI 20004 Statistical Methods of Research
  - SOSC 26009 Introductory Statistical Methods
- Other approved substitutions by petition

#### APPROVED EQUIVALENT COURSEWORK IN COMPUTATIONAL METHODS

Students must complete two introductory computer programming courses: MACS 10111 Principles of Computing 1: Computational Thinking for Social Scientists and MACS 10112 Principles of Computing 2: Data Management for Social Scientists, or equivalent coursework.

Examples of approved equivalent coursework include:

- CMSC 12100 Computer Science with Applications I and CMSC 12200 Computer Science with Applications II
- CMSC 14100 Introduction to Computer Science I and CMSC 14200 Introduction to Computer Science II
- If students have already completed introductory computer programming courses (e.g., CMSC 14100-14200) as a part of their major, another minor, or general education requirements, they may substitute additional CMSC or MACS courses to meet the two-course Computational Methods minor requirement. For instance, a student who completed CMSC 14100-14200 to meet requirements in their major might take MACS 10113 Principles of Computing 3: Big Data and High Performance Computing for Social Scientists and CMSC 14300 Systems Programming I to meet the requirements (or any other combination of CMSC and MACS courses) of the minor.
- Other approved substitutions by petition

#### COMPUTATIONAL SOCIAL SCIENCE ELECTIVE OPTIONS

Students must complete at least two Computational Social Science electives as a part of the minor—any two Computational Social Science courses with an MACS course number at the 20000-level and below (see those listed below). Course availability varies by year.

A list of all MACS courses offered each quarter is also available on the program website (<https://macs.uchicago.edu/programs-of-study/courses/>).

#### COMPUTATIONAL SOCIAL SCIENCE COURSES

##### **MACS 10000. 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.

Instructor(s): Jean Clipperton Terms Offered: Autumn

Equivalent Course(s): MACS 30000

##### **MACS 10100. Perspectives on Computational Modeling. 100 Units.**

This is a core-course for the MACSS program and it requires Python programming experience (for non-MACSS students, please email the instructor for consultation). This course will teach fundamental skills of applying statistical machine learning models in computational social science tasks. It will focus on understanding the strengths and weaknesses of modern machine learning algorithms as well as their applications in real-world tasks. Topics will include the key techniques in standard machine learning pipelines: data processing (e.g., data representation, feature selection), classification models (e.g., decision trees, logistic regression, naive bayes), regression models (e.g., linear regression), model evaluation (e.g., cross-validation, confusion matrix, precision,

recall, and  $f_1$  for classification models; RMSE and Pearson correlation for regression models), and error analysis (e.g., data imbalance, bias-variance tradeoff, interpret model performance).

Instructor(s): Wang, Zhao Terms Offered: Winter

Prerequisite(s): Consent required for all Non MACSS students

Equivalent Course(s): MACS 30100

#### **MACS 10111. Principles of Computing 1: Computational Thinking for Social Scientists. 100 Units.**

This course is the first in a three-quarter sequence that teaches fundamentals of computational thinking to students in the social sciences. Lectures in the class will cover topics such as functions, data structures, as well as classes and objects. Assignments will give students the opportunity to practice these basic computing concepts using the Python programming language and get familiar with computational logic in real-world tasks.

Instructor(s): Clipperton, Jean Terms Offered: Autumn

Prerequisite(s): This course is intended for those who placed into it via MACS 30120 "Computing Fundamentals Boot Camp" or for those who are not otherwise prepared to independently and fluently write code using any programming language (e.g., R, Python, C, C++, Java). Note that this is the introductory version of MACS 30121. We will provide several test questions to help students better identify the suitability of this course at the beginning of the quarter.

Equivalent Course(s): MACS 30111

#### **MACS 10112. Principles of Computing 2: Data Management for Social Scientists. 100 Units.**

This course is the second in a three-quarter sequence that teaches computational thinking and programming skills to students in the social sciences. Specifically, this course equips students with a fundamental toolkit for working with social science data. Students will learn the basics of web-scraping, relational databases, record linkage, data cleaning, modeling, visualization, and data structures. The programming language of the course is Python.

Instructor(s): Nardin, Sabrina Terms Offered: Winter

Prerequisite(s): MACS 30111 or instructor consent. Note that this is the introductory version of MACS 30122.

Instructor consent required for all non-MACS students.

Equivalent Course(s): MACS 30112

#### **MACS 10113. Principles of Computing 3: Big Data and High Performance Computing for Social Scientists. 100 Units.**

Computational social scientists increasingly need to grapple with data that is too big and code that is too resource intensive to run on a local machine. Using Python, students in this course will learn how to effectively scale their computational methods beyond their local machines –optimizing and parallelizing their code across clusters of CPUs and GPUs, both on-premises and in the cloud. The focus of the course will be on social scientific applications, such as: accelerating social simulations by several orders of magnitude, processing large amounts of social media data in real-time, and training machine learning models on economic datasets that are too large for an average laptop to handle.

Instructor(s): TBD Terms Offered: Spring

Prerequisite(s): MACS 30111 and MACS 30112, or equivalent

Equivalent Course(s): MACS 30113

#### **MACS 10150. Perspectives on Computational Modeling for Economics. 100 Units.**

In this course students will learn several computational methodologies and tools to solve, simulate, and analyze models that are the backbone of current macroeconomic analysis. While learning the relevant computational methods is the main objective, the theoretical economic aspects of the model will be stressed and the students will be required to apply their economic knowledge and skills to interpret and analyze the results. We will examine non-stochastic and stochastic general equilibrium models, both under local and global approximations. The main part of the course will deal with representative agent models, but a significant part will be devoted to introducing students to the solution of heterogeneous agent models as well.

Instructor(s): Salas, Sergio Terms Offered: Winter

Prerequisite(s): PQ for undergraduates: Econ 20200/20210 and MATH 19620/Stat 24300/Math 20250 and Stat 23400/Stat 24400/Stat 24410.

Note(s): MACSS students have priority.

Equivalent Course(s): ECMA 31140, MACS 30150

#### **MACS 10200. Perspectives on Computational Research. 100 Units.**

This course focuses on applying computational methods to conducting social scientific research and the development of a strong research proposal. Students will identify a research question of their own interest that involves a direct reference to social scientific theory, the use of data, and a significant computational component. The students will review existing literature, identify an appropriate data source and conduct exploratory analysis, or develop a method (e.g. survey, experiment) through which they plan on collecting data, and generate a complete and well-motivated research proposal. The course will include modules on theoretical and practical considerations, including topics such as epistemological questions about research design, conducting literature reviews, data visualization and interpretation, reproducible research, writing and rhetoric, as well as presenting work to an audience.

Instructor(s): Pedro Arroyo, Shilin Jia, Ali Sanaei Terms Offered: Spring

Prerequisite(s): MACSS students have priority. Others admitted with instructor consent.

Equivalent Course(s): MACS 30200

**MACS 20101. Social Network Analysis. 100 Units.**

This course introduces students to concepts and techniques of Social Network Analysis ("SNA"). Social Network Analysis is a theoretical approach and a set of methods to study the structure of relationships among entities (e.g., people, organizations, ideas, words, etc.). Students will learn concepts and tools to identify network nodes, groups, and structures in different types of networks. Specifically, the class will focus on a number of social network concepts, such as social capital, homophily, contagion, etc., and on how to operationalize them using network measures, such as centrality, structural holes, and others.

Instructor(s): Sabrina Nardin Terms Offered: Autumn

Equivalent Course(s): MACS 40101, SOCI 40248

**MACS 20400. 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): Clindaniel, Jon Terms Offered: Autumn Winter

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): MACS 40400, PSYC 40460, CHDV 40404, MAPS 40401

**MACS 20500. Computing for the Social Sciences. 100 Units.**

This is an applied course for social scientists with little-to-no programming experience who wish to harness growing digital and computational resources. The focus of the course is on learning the basics of programming and on generating reproducible research. Topics include coding concepts (e.g., data structures, control structures, functions, etc.), data visualization, data wrangling and cleaning, version control software, exploratory data analysis, etc. Students will leave the course with basic computational skills implemented through many methods and approaches to social science; while students will not become expert programmers, they will gain the knowledge of how to adapt and expand these skills as they are presented with new questions, methods, and data. The course will be taught in R.

Instructor(s): Jean Clipperton Terms Offered: Autumn Spring Winter

Note(s): MACS students have priority.

Equivalent Course(s): ENST 20550, CHDV 30511, SOCI 40176, SOCI 20278, MACS 30500, SOSC 26032, PSYC 30510, MAPS 30500, PLSC 30235

**MACS 20519. Spatial Cluster Analysis. 100 Units.**

This course provides an overview of methods to identify interesting patterns in geographic data, so-called spatial clusters. Cluster concepts come in many different forms and can generally be differentiated between the search for interesting locations and the grouping of similar locations. The first category consists of the identification of extreme concentrations of locations (events), such as hot spots of crime events, and the location of geographical concentrations of observations with similar values for one or more variables, such as areas with elevated disease incidence. The second group consists of the combination of spatial observations into larger (aggregate) areas such that internal similarity is maximized (regionalization). The methods covered come from the fields of spatial statistics as well as machine learning (unsupervised learning) and operations research. Topics include point pattern analysis, spatial scan statistics, local spatial autocorrelation, dimension reduction, as well as spatially explicit hierarchical, agglomerative and density-based clustering. Applications range from criminology and public health to politics and marketing. An important aspect of the course is the analysis of actual data sets by means of open source software, such as GeoDa, R or Python.

Instructor(s): L. Anselin and P. Amaral Terms Offered: Winter

Prerequisite(s): STAT 22000 or equivalent; SOCI 20253/30253 (or equivalent) Introduction to Spatial Data Science required.

Equivalent Course(s): SOCI 20519, DATA 20519, ENST 20519, GISC 20519, MACS 30519, SOCI 30519, GISC 30519

**MACS 20550. Agent-Based Modeling. 100 Units.**

Social science problems often have so many details and moving parts that it can be difficult for researchers to gain traction without models. In this course, we explore agent-based modeling approaches to understand these social science problems including cooperation and the development of culture. Agent-based models enable us to build an understanding from the bottom up, starting with simple assumptions and analyzing how patterns emerge at a larger scale. Through the term, we'll cover the fundamentals of modeling, including basic principles of model design, data extraction, and canonical examples like Conway's Game of Life, Schelling's segregation model, and Boids/flocking. The course is balanced between social science readings and applications and hands-on coding. It culminates in a final project consisting of an agent-based model designed by students to apply to a social science phenomenon.

Instructor(s): Clipperton, Jean Terms Offered: Winter

Prerequisite(s): Background in Python at the level of MACS 30111 or equivalent. Consent required for all undergrads not meeting prerequisite and all non MACSS students.

Equivalent Course(s): MACS 40550

**MACS 20700. 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): Jean Clipperton Terms Offered: Spring

Equivalent Course(s): MACS 40700

**MACS 21000. 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 involved 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 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): STAT 31900, CHDV 30102, PLSC 30102, SOCI 30315, PBHS 43201, MACS 51000, CHDV 20102

**MACS 23002. Introduction to Machine Learning. 100 Units.**

This course requires Python programming experience. The course will train students to gain the fundamental skills of machine learning. It will cover knowledge and skills of running with computational research projects from a machine learning perspective, including the key techniques used in standard machine learning pipelines: data processing (e.g., data cleaning, feature selection, feature engineering), classification models (e.g., logistic regression, decision trees, naive bayes), regression models (e.g., linear regression, polynomial regression), parameter tuning (e.g., grid-search), model evaluation (e.g., cross-validation, confusion matrix, precision, recall, and f1 for classification models; RMSE and Pearson correlation for regression models), and error analysis (e.g., data imbalance, bias-variance tradeoff). Students will learn simple and efficient machine learning algorithms for predictive data analysis as well as gain hands-on experience by applying machine learning algorithms in social science tasks. The ultimate goal of this course is to prepare students with essential machine learning skills that are in demand both in research and industry.

Instructor(s): Wang, Zhao Terms Offered: Winter

Prerequisite(s): Python programming experience required.

Equivalent Course(s): PLSC 43505, MACS 33002, MAPS 33002

**MACS 26000. Computational Methods Using Online Social Media Data. 100 Units.**

This course will discuss a broad range of computational social science topics that leverage large-scale data from online communication platforms to gain insights into social issues. We will start from collecting and processing data from social media platforms (e.g., Twitter, IMDB, Airbnb, Yelp), and then introduce computational research topics that include but are not limited to: sentiment analysis, deceptive marketing, recommendation system, fake news detection, spam detection, bot detection, demographic inference, public health, political attitude analysis, personality and behavior analysis, and cyberbullying. We will use version control techniques (e.g., git, Github) to keep track of the class projects. The ultimate goal of this course is to provide a broad introduction of computational social science research areas and train students to be familiar with the pipelines of doing computational research.

Instructor(s): Wang, Zhao Terms Offered: Winter

Equivalent Course(s): MACS 36000, MAPS 36001

**MACS 27000. Thinking with Deep Learning for Complex Social & Cultural Data Analysis. 100 Units.**

A deluge of digital content is generated daily by web-based platforms and sensors that capture digital traces of human communication and connection, and complex states of society, culture, economy, and the world.

Emerging deep learning methods enable the integration of these complex data into unified social and cultural "spaces" that enable new answers to classic social and cultural questions, and also pose novel questions. From the perspective of deep learning, everything can be viewed as data—novels, field notes, photographs, lists of transactions, networks of interaction, theories, epistemic styles—and our treatment examines how to configure deep learning architectures and multi-modal data pipelines to improve the capacity of representations, the accuracy of complex predictions, and the relevance of insights to substantial social and cultural questions. This class is for anyone wishing to analyse textual, network, image or arbitrary structured and unstructured data, especially in concert with one another to solve complex social and cultural analysis problems (e.g., characterize a culture; predict next year's ideology).

Instructor(s): James Evans Terms Offered: Spring

Prerequisite(s): The course uses Python and the widely popular PyData ecosystem to demonstrate all motivating examples and includes working code, accompanying exercises, relevant datasets and additional analytics and visualization that facilitate social and cultural interpretation and communication. Familiarity with Python is required.

Equivalent Course(s): MACS 37000, SOCI 30332, DATA 30332

