Computational Courses Offered at The Graduate Center – Spring 2016

In Spring 2016, 26 computationally inflected courses will be offered across The Graduate Center’s doctoral, master’s, and certificate programs. Topics covered in these courses range from data visualization to natural language processing to statistical analysis and provide hands-on opportunities for GC students to develop their digital research skills.

The 26 courses are listed below in alphabetical order by program along with a brief description. Interested students should contact individual programs if they have questions regarding course content/prerequisites or if permission is required.

For more information about the GC Digital Initiatives Program, please visit http://cuny.is/gcdi and http://cuny.is/gcdigitalfellows.

BIOL 79303: Computational Molecular Bio
***Permission from instructor required.***
This course will introduce both bioinformatics theories and practices. Topics include database searching, sequence alignment, molecular phylogenetics, structure predication, and microarray analysis. The course is held in a UNIX-based instructional lab specifically configured for bioinformatics applications. Each session consists of a first-half instruction on bioinformatics theories and a second-half session of hands-on exercises.

CSc 86010: Massively Parallel Programming
***Some technical background required. Permission from CSc EO required.***
Rationale: Computationally complex problems such as graphical representations of movement cannot be processed in a reasonable amount of time on a single CPU. Currently, most graphical computations and many scientific calculations involving large datasets and complex systems are run in a massively parallel environment. Designing algorithms to efficiently execute in both time and memory usage in such environments requires an understanding of concurrency and the hardware requirements of massively parallel systems, for example, Graphical Processing Units (GPUs). This course is designed to give the students an introduction to the concepts and usage of GPUs and the CUDA extensions to the C/C++ languages.
Description: A survey of the approaches to massively parallel computer applications with emphasis on using graphical programming units (GPUs) and the CUDA extensions to the C/C++ programming languages. Comparisons between multicore CPUs and multi-processor GPUs will be given. Issues such as organization of large data sets, memory usage, and communication concerns will be addressed. Different levels of concurrency will also be discussed with most the focus on thread level-concurrency. Also multiple data streams on a single GPU and multiple GPUs will be covered with quick reviews of OpenMP and OpenMPI usage. Standard problems will be discussed.
CSc 74030: Computer Vision and Image Processing

***Some technical background required. Permission from CSc EO required.***

Rationale: Computer vision and image processing are important and fast evolving areas of computer science, and have been applied in many disciplines. This course will introduce students to the fascinating fields. Student will gain familiarity with both established and emergent methods, algorithms and architectures. This course will enable students to apply computer vision and image processing techniques to solve various real-world problems, and develop skills for research in the fields.

Description: This course introduces fundamental concepts and techniques for image processing and computer vision. Topics to be covered include image formation, image filtering, edge detection and segmentation, morphological processing, registration, object recognition, object detection and tracking, 3D vision, etc.

CSc 83040: Text Mining

***Some technical background required. Permission from CSc EO required.***

Rationale: With the explosion of textual data on the world-wide web, text mining has become an important area of research. Text sources such as blogs, literature, social media, web pages and news articles can be analyzed to learn patterns, opinions, trends, and ideas. Text mining is a sub-area of data mining that deals with unstructured text. Algorithms have been developed for learning from unstructured text, and these often have practical applications in areas such as healthcare, advertising and homeland security.

Description: Text mining can be defined as the process of finding or learning patterns from textual data to aid in decision making. This course will include the study of different representations of textual data and the algorithms used to glean new information from the data. It encompasses ideas from many other areas in computer science including artificial intelligence, machine learning, databases, information retrieval, and natural language processing. This class will primarily focus on the statistical methods for text mining, including machine learning techniques that are used to facilitate decision making.

CSc 72030: Database Systems

***Some technical background required. Permission from CSc EO required.***

Description: Database Management Systems (DBMS) are vital components of modern information systems. Database applications are pervasive and range in size from small in-memory databases to terabytes or even larger in various applications domains. The course focuses on the fundamentals of knowledgebase and relational database management systems, and the current developments in database theory and their practice. The course reviews topics such as conceptual data modeling, relational data model, relational query languages, relational database design and transaction processing and current technologies such as semantic web, parallel and noSQL databases. It exposes the student to the fundamental concepts and techniques in database use and development as well provides a foundation for research in databases. The course assumes prior exposure to databases, specifically to the relational data
model and it builds new technologies on this foundation. In the first half of the course the relational data model, relational query languages, relational database design and conceptual data modeling are reviewed. It then focuses on XML, RD, OWL, parallel, and noSQL databases. It also bridges databases and knowledgebases which is the current trend. The course requires a term project in which the student implements a database application or explores a database issue. We will use PostgreSQL as the database platform for doing the assignments.

CSc 84010: Advanced Natural Language Processing (Crosslisted with LING 83600)
***Some technical background required. Permission from CSc EO required.***
Description: A multimodal user interface for devices requires the integration of several recognition technologies together with sophisticated user interface and distinct tools for input and output of data. Multimodal interaction provides the mobile user with new complex multiple modalities of interacting with a system such as: speech, gestures and movements, touch, type and more. The course discusses the new world of multimodality User Interface, the technologies and design which are innovation and create a state of the art user interface. We will discuss the commercial challenges and try to offer new approaches to these issues. The objective of the course is to expose the students to state of the art multimodal user interface technologies and to have them face design challenges so they will become familiar with the area of Mobility and Multimodality both from the technological aspect and from the usability aspect. Student will be required to suggest and design the architecture and dialog flow for a multimodal application. The design plane will be done using the tools and best practices acquired in the class.

CSc 86030: Simulation Methodology
***Some technical background required. Permission from CSc EO required.***
Rationale: Systems have become so complex that it is often the case that understanding them cannot be done analytically. Therefore, their behavior can be observed by modeling them and simulating them. This course will introduce the theories and applications of computer modeling and simulation, focusing on discrete event system modeling and simulation.
Description: Basic concepts of systems modeling, in-depth discussions of modeling elements, simulation protocols, and their relationships are covered. The modeling and simulation techniques will be illustrated by examples in DEVSJAVA, which is a Java implementation of the systematic and generic DEVS (Discrete Event System Specification) approach to modeling and simulation. Related application domains of this course include communication, manufacturing, social/biological systems, and business. Some advanced concepts and practices will be presented to attract students’ interests in a seminar format.

CSc 84010: Pattern Matching Algorithms
***Some technical background required. Permission from CSc EO required.***
Rationale: The advent of the worldwide web, next generation sequencing, and increased use of satellite imaging have all contributed to the current information explosion. One of the most basic
tasks common to many applications is the discovery of patterns in the available data. To render the searching of big data feasible, it is imperative that the underlying algorithms be efficient, both in terms of time and space. Pattern Matching is a branch of theoretical computer science whose ideas are used in practice daily in many different data-driven areas, including (but not limited to) word processors, web search engines, biological sequence alignments, intrusion detection systems, data compression, database retrieval, and music analysis. This course gives a student training in the process of developing and analyzing efficient algorithms through the study of pattern matching algorithms that are used for searching and indexing large textual data. Description: Pattern Matching is one of the fundamental problems in Computer Science. In its classical form, the problem consists of 1-dimensional string matching. Given a string (or text) $T$ and a shorter string (or pattern) $P$, find all occurrences of $P$ in $T$. Over the last four decades, research in Pattern Matching has developed the field into a rich area of algorithmics. This course covers several variants of the pattern matching problem. Emphasis is placed on the algorithmic techniques used to speed up naive solutions, and on the time complexity analysis of the algorithms.

**CSc 86005: Big Data**

***Some technical background required. Permission from CSc EO required.***

Rationale: In addition to constantly growing volumes of proprietary transaction, product, inventory, customer, competitor, and industry data collected from enterprise systems, organizations are also faced with overwhelming amounts data from the Web, social media, mobile sources, and sensor networks that do not fit into traditional databases in terms of volume, velocity and variety (the three Vs of Big Data). This Big Data flood poses challenges as well as opportunities, if managed and analyzed properly, to derive new actionable knowledge and intelligence in a timely manner. This course will explore existing and emerging methods to manage, integrate, analyze and visualize domain-specific Big Data, to identify and provide domain specific solutions.

Description: This course covers the research issues and practical methods of managing and analyzing Big Data to gain and discover insights, patterns, and knowledge nuggets that can support decision makers.

**CSc 85040: Algorithms for Big Data**

***Some technical background required. Permission from CSc EO required.***

Rationale: Traditional analysis of algorithms generally assumes full storage of data and considers running times polynomial in input size to be efficient. Operating on massive-scale data sets such as those of tech companies such as Google, Facebook, etc., or on indefinitely large data streams, such as those generated by sensor networks and security applications, leads to fundamentally different algorithmic models. MapReduce/Hadoop in particular has seen widespread adoption in industry.

Description: This course addresses algorithmic problems in a world of big data, i.e., problems in settings where the algorithm’s input [the data] is too large to fit within a single computer’s
Computational Courses Offered at The Graduate Center – Spring 2016

memory. Traditional analysis of algorithms generally assumes full storage of data and considers running times polynomial in input size to be efficient. Operating on massive-scale data sets such as those of tech companies such as Google, Facebook, etc., or on indefinitely large data streams, such as those generated by sensor networks and security applications, leads to fundamentally different algorithmic models. In previous decades, DBMS settings where the data sets on a machine’s disk but not in memory motivated the external memory or I/O model (e.g. external mergesort and B-trees). More recently, models such as MapReduce/Hadoop have appeared for computing on data distributed across many machines (e.g. PageRank computation or matrix multiplication).

CSc 74011: Artificial Intelligence
***Some technical background required. Permission from CSc EO required.***
Rationale: Artificial intelligence (AI) develops programmed agents (systems) that match or outperform people’s abilities to make decisions, to learn, and to plan. To do so, AI develops algorithms and methodologies that sense a system’s environment, decide what to do given that data, and effect its chosen actions in its environment.
Description: This is a graduate-level course on artificial intelligence. It emphasizes fast and clever search heuristics, thoughtful ways to represent knowledge, and incisive techniques that support rational decision making. Application areas will include game playing, natural language processing, and robotics. Students are expected to have a solid background in the analysis of algorithms, proofs in propositional and first-order logic, discrete mathematics, and elementary probability.

CSc 83060: Data Visualization
***Open to all GC students.***
Rationale: Today quantitative and symbolic data are easily collected in computer format, from databases, websites, smart devices, and anything that has interconnect capabilities. When such large amounts of data are put in spreadsheets or tabular reports, it becomes difficult to see the patterns, structure, trends, or relationships inherent in the data. Effective data visualization exposes these inherent relationships, consolidating and illustrating them in graphics.
Description: A visualization organizes data in a way that the structure and relationships in the data that may not be so easily understood becomes easily understood and interpreted with the visualization. Visualizations of a data set give the reader a narrative that tells the story of the data. The purpose of data visualization is to convey information contained in data to clearly and efficiently communicate an accurate picture of what the data says through understandable and context appropriate visualizations. To do a visualization can be just exploratory or entails using Machine Learning techniques that determine the structure of the data. The visualizations are then matched to the data structure. The course will explore how principles of information graphics and design and how principles of visual perception, can be used with machine learning techniques to make effective data visualizations. Each student will make a presentation of some principles of data visualizations or do a visualization project. The course is open to PhD
Computational Courses Offered at The Graduate Center – Spring 2016

students in all programs. Non-computer science students will be paired with computer science students for the visualization project.

ECON 82200: Econometrics II
***Permission from instructor required. Prerequisite: ECON 82100 Econometrics I***
This is the second part of the standard first year graduate econometrics class. Topics include: instrumental variables estimation, generalized method of moments, static and dynamic panel data, specification tests, limited dependent variables models, selection models, count data models and (time permitting) duration models. Examples from the literature are discussed and students gain practice in estimating and interpreting empirical results.

ECON 82300: Applied Microeconometrics
***Permission from instructor required. Prerequisite: ECON 82100 and ECON 82200***
This course will cover the analytical methods often used in applied microeconomic research, in areas such as health and labor economics, and is particularly important for students planning to write an empirical dissertation in one of these areas. We will cover discrete choice models, censored regression models, selection models, and panel data models. In addition to covering the statistical properties of these models, the course will emphasize the applications of these models in a variety of studies. Students will be asked to carry out analysis of data using the methods covered by the course, and will also learn to critically analyze empirical studies that implement these methods.

ECON 82900: Spatial Econometrics
***Permission from instructor required. Prerequisite: ECON 82100 and ECON 82200***
This course provides a theoretical and empirical overview of econometric techniques that may be used when studying spatial data. Spatial data consist of observations that interact in some sense: geographically, economically, socially, politically, etc. This interaction may express itself in behavioral patterns or in exposure to common shocks. Lectures cover economic theories that motivate interactive behavioral patterns, examine spatial-econometric techniques of estimation and hypothesis testing, discuss statistical software, and review empirical studies that apply those techniques. Evaluation is based on exams that focus on theory and interpretation, and assignments that focus primarily on empirical implementation.

EPSY 70600: Statistics and Computer Programming II (Crosslisted with PSYC 70600)
***Prerequisite: An undergraduate statistics course***
This is the second course in a one-year sequence teaching the basic aspects of statistical theory and methods, as well as computing, for the most common data analysis techniques in the social sciences. EPSY 70500 and 70600 form an integrated sequence covering descriptive statistics, point and interval estimation, hypothesis testing, t-tests, analysis of variance,
correlation, regression (including elementary matrix algebra), repeated measures designs, cross-classified data, and the use of computer packages for these analyses. This course will use SPSS statistical analysis software for Windows.

**EPSY 83500: Categorical Data Analysis**  
***Permission from instructor required. Prerequisite: Course on applied regression, EPSY 70600 or equivalent***  
The goal of this course is to introduce statistical models for categorical data, which are common throughout the behavioral and health sciences. These include binary data (sick vs. not), ordinal data (coarse Likert scales), nominal data (answers “yes”, answers “no”, answers “don’t know”), and count data (how many events in a given amount of time). The distributions of these data are markedly non-Gaussian and the linear model usually fails to model them well. This course will: Introduce relevant statistical theory for categorical distributions and the basics of maximum likelihood estimation; Cover models such as logistic, Poisson, ordinal or nominal regression, etc., and show how they fit in the framework of the generalized linear model (GLM); Consider some extensions to these models for longitudinal and clustered data; Deal with the interpretation of these models, which differ substantially from that of the linear model.

**ITCP 70020: Interactive Technology and the University: Theory, Design, and Practice**  
***Prerequisite: ITCP 70010 (Core I) or permission from ITP EO.***  
This second core course introduces students to IT in the classroom and in academic research, focusing on cognition and design. Interest areas include research in digital media; visualization and design; modes of learning within and outside the classroom; and conceptualization and production of educational media products. The course provides a hands-on introduction to key educational uses of digital media applications, including on-line writing tools, electronic archives, and experimentation in virtual spaces. Core II employs an interdisciplinary approach to the application of digital media to classroom teaching and scholarly research and presentations. Students will learn skills and concepts and then will design and plan a digital media project in their academic discipline. This course makes it possible for participating doctoral students to build on the theoretical insights gleaned in the first core course and to begin to conceive and develop an IT project in their own discipline.

**LING 83800: Practical and Design Issues in Natural Language Processing Systems**  
***Permission from instructor required.***  
This course is designed to introduce students to practical concepts and ideas in Natural Language Processing (NLP) industry and the challenges of designing systems that perform various intelligent tasks involving human languages. We will explore how computational linguistics theories are put into practice in industry projects and implementations. The course approaches NLP from the practical aspects of system design, usability and practical do’s and don’ts, as well as of tools available for the processing of linguistic information and the underlying
Computational Courses Offered at The Graduate Center – Spring 2016

computational properties of natural languages. In this course students will become familiar with the complexities of modeling human languages for system implementation. We’ll face the challenges to “understand it” automatically using machines while thriving to achieve the same quality level as in human interaction. We will benchmark speech and text analytics systems to dialog management and call center applications which use language models and statistical operations in the era of big data analysis. The course follows the different schools of thought, as well as the historical development that led to our current achievements all in light of “Man – Machine interaction” basic issues. We will analyze real world examples of products and technologies developed in the area and used in real world industry projects.

LING 83600: Advanced Natural Language Processing: Multimodality Design and Applications (Crosslisted with CSc 84010)
***Some programming background required. Permission from instructor required.***
Description: A multimodal user interface for devices requires the integration of several recognition technologies together with sophisticated user interface and distinct tools for input and output of data. Multimodal interaction provides the mobile user with new complex multiple modalities of interfacing with a system such as: speech, gestures and movements, touch, type and more. The course discusses the new world of multimodality User Interface, the technologies and design which are innovation and create a state of the art user interface. We will discuss the commercial challenges and try to offer new approaches to these issues. The objective of the course is to expose the students to state of the art multimodal user interface technologies and to have them face design challenges so they will become familiar with the area of Mobility and Multimodality both from the technological aspect and from the usability aspect. Student will be required to suggest and design the architecture and dialog flow for a multimodal application. The design plane will be done using the tools and best practices acquired in the class.

MALS 75300: Data Visualization Methods (Crosslisted with CSc 87100)
***Open to all GC students.***
This class is designed to teach the students practical skills in visualizing and analyzing cultural and social datasets. The main tool we will use is R, the leading open source platform for data analysis. The students will be also introduced to other popular tools for creating interactive web-based visualizations. We will cover the following practical topics: preparing data for analysis and visualization; summarizing data; basic visualization techniques for 1D, 2D, and multi-variable data; use of visualization for exploratory data analysis; creative data visualization; history of visualization; elements of graphic design for visualization and project web site design; strategies for presenting projects online; how to write effective project descriptions for the web presentation; promoting projects through social media and getting media coverage. We will also examine papers from computational social science and data analysis/visualization projects by designers and artists.
Computational Courses Offered at The Graduate Center – Spring 2016

MALS 75500: Digital Humanities Methods and Practices (Crosslisted with IDS 81640; Digital Praxis Seminar II, Spring 2016)
***Open to all GC students.***
During the Fall 2015 semester, students explored the landscape of the digital humanities, examining a range of ways to approach DH work and proposing potential DH projects. In the spring, we will put that thinking into action. In this praxis-oriented course, we will split into teams and then develop and launch functional versions of projects first imagined in the fall. Students will complete the class having gained hands-on experience in the collaborative planning, production, and dissemination of a digital humanities project, and having picked up a variety of technical, project management, and rhetorical skills along the way. A goal is to produce projects that will have a trajectory and a timeline of their own that extends beyond the Spring 2016 semester. Students will be supported by a range of advisors matched to the needs of the individual projects, and successful completion of the class will require a rigorous commitment to meeting target delivery dates we will establish together at the outset. The class will hold a public launch event at the end of the semester where students will present their proofs-of-concept, and receive feedback from the broader community.

PHYS 85200: Computational Methods in Physics
***Open to all GC students. Students should have basic knowledge in statistical mechanics, thermodynamics, or in physical chemistry. Programming skills are helpful but not required.***
This course is intended to give a broad introduction to tools of computational research in physics. After learning some basic skills (representation of numbers, differentiation, integration, sampling) fully developed case studies emphasizing a few standard computational paradigms (molecular dynamics, Monte-Carlo, quantum spectra) will be used to provide a detailed prospective on modeling, from setting up the model, to designing the simulation and data analysis.

PSC 89101: Quantitative Analysis I
***Open to all GC students.***
The aim of this course is to introduce graduate students to statistical analysis in political science. I want students to think of themselves as future contributors of empirical work, as well as critical consumers. In that spirit, there will be an emphasis on "learning-by-doing." Each student should locate a data set of interest to them by the third week of the semester that will be used to carry out statistical exercises. To help students see the linkages between the material we cover and the work in the discipline, I am also listing some articles from our journals in addition to the assigned books. It is my hope that students will be more tolerant of the technical material if they can see the payoff in terms of a better understanding of political science rather than statistics. For some of the topics, I have also suggested additional readings that may increase your understanding of the technical material. With technical material, I have found that it helps to read two or three different presentations of the same topic to understand it more
clearly. These readings should be done actively with paper and pencil in hand. By the end of this course students will have a working understanding of regression analysis.

**SOC 81100: Social Demography and Geographies of the Disadvantaged** *(Crosslisted with DCP 80300)*

***Prerequisite: Introductory statistics including multiple linear regression.***

In this course we will examine the role of “place” as social geographies which relates to containers of populations. In particular, we are interested in the social geographies of disadvantage. We will explore theoretical treatments and popular sources of data in the analysis of disadvantaged populations. We will also be introduced to ways that public policies, institutional practices and spatial perceptions become institutionalized and influence local contexts to maintain disadvantage. Students in the course will work with data from the US Census Bureau, Centers for Disease Control, and other administrative population level data sources. In addition, students will be introduced to a series of open source software packages commonly used in the application of methods associated with the examination of disadvantaged populations/individuals in localized contexts. Methodological applications include Multilevel modeling (could be listed as HLM), Geographically Weighted Regression (GWR), Spatial Regression, and an introduction to Spatio-Temporal Analyses.

**SOC 81900: Topics in Multivariate Methods**

***Students should have background in statistical analysis at least up to the level of a graduate course covering regression.***

Social science methods have made a lot of progress over the last 25 years; simple regression is no longer state-of-the-art. This course is an introduction to these more recent methods, emphasizing when they are used and how to use them, and minimizing the underlying math. It is taught in a computer lab, and is suitable for anyone who has statistics up to regression and takes a practical approach. The course begins with a review of multiple regression and its limitations, which highlights the rationale for the new techniques. It then introduces several methods whose purpose is to focus on estimating the causal influence of one particular variable of interest on some outcome. These methods: regression discontinuity analyses, propensity score matching, difference-in-difference models and others are particularly useful for evaluating the effects of policy changes or of social/clinical interventions. Another set of methods we will cover has a different goal: uncovering unexpected relations in data. These are the central tools of ‘data mining’ and they identify interactions, non-linear relationships and heterogeneity in datasets. By the end of the course, students should know which techniques to use in what contexts and feel confident that they know how to run each program and interpret its output. Each student will be graded based a term paper that presents an analysis of quantitative data of their choosing using one or more of these techniques. We hope this will serve as the core of a publishable paper. This course uses Stata and JMP pro software.
UED 74100: Quantitative Research Methods in Urban Education

***Permission from instructor required***

This course will instruct students in file management and many quantitative methods used for the analysis of survey data. Each student will undertake an individual project and will work on every aspect of the research endeavor from identifying a topic for investigation to writing and presenting a final project.