QUANTITATIVE ANALYSIS CENTER

The Quantitative Analysis Center (QAC) coordinates support for quantitative analysis across the curriculum and provides an institutional framework for collaboration across departments and disciplines in the area of data analysis. Through its programs, it facilitates the integration of quantitative teaching and research activities and provides experiential learning opportunities in statistical computing across academic fields and disciplines. The Center contributes to the development of digital and computational studies initiatives, sponsors data analysis labs, and oversees the Data Analysis Minor (catalog.wesleyan.edu/departments/qac/ugrd-data-mn) and the Applied Data Science Certificate (catalog.wesleyan.edu/certificates/applied-data-science) programs.

FACULTY

Robert Ira Kabacoff
BA, University of Connecticut; PHD, University of Missouri, St. Louis
Professor of the Practice in Quantitative Analysis

Emmanuel I. Kaparakis
Director of Centers for Advanced Computing; Director, Center of Quantitative Analysis; Adjunct Assistant Professor of Quantitative Analysis

Valerie L. Nazzaro
BA, University of Connecticut; MA, University of Connecticut; MS, University of Connecticut; PHD, University of Connecticut
Assistant Professor of the Practice in Quantitative Analysis

Pavel V Oleinikov
Associate Director, Quantitative Analysis Center; Adjunct Assistant Professor of Quantitative Analysis

AFFILIATED FACULTY

Jennifer S. Rose
BA, University of Rhode Island; MA, University of Rhode Island; PHD, University of Rhode Island
Professor of the Practice in the Center for Pedagogical Innovation; Director, Institutional Review Board; Director, Center for Pedagogical Innovation; Professor of the Practice, Quantitative Analysis Center

VISITING FACULTY

William S. Boyd Jr
BA, Rhodes College; MA, The University of Memphis; PHD, University Tennessee Memph
Visiting Scholar in the Quantitative Analysis Center

Jielu Yao
BS, Fudan University; MSC, National University of Singapore
Research Fellow in the Quantitative Analysis Center

ADVISING BOARD

Francis Starr, Professor of Physics, Chair
Erika Franklin Fowler, Associate Professor of Government
Daniel Krizanc, Professor of Computer Science
Manolis Kaparakis, Director of Centers for Advanced Computing, ex officio
David Baird, Vice President for Information Technology and CIO
Marc Eisner, Professor of Government and Dean of the Social Science
Diane Klare, Head of Research Services, Olin Memorial Library

• Data Analysis Minor (catalog.wesleyan.edu/departments/qac/ugrd-data-mn)

QAC150 Working with SQL and Databases
Many of you have heard of studies that analyzed Twitter messages and predicted some phenomena—spread of flu in New York, consumer confidence index, and so on. Behind the success of these studies are the systems for data storage and retrieval. A regular user can access only the latest nine days of tweets. Any study that aspires to analyze longer periods has to deal with the issues of storing the observations and retrieving them later for analysis. The goal of this course is to show you how to do that—how to connect to various types of databases and how to retrieve and update your data. We will start with relational databases; learn SQL, the language used to query and update the data; and explore the latest developments in the database field—Hadoop and MapReduce.
Offering: Host
Grading: OPT
Credits: 0.25
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: None

QAC151 Working with Excel and VBA
Many of us know Excel for its spreadsheets: a quick and easy way to store some information, share it, and maybe make some charts. The goal of this course is to show you the more advanced features of Excel. We will write code in Visual Basic for Applications (VBA), learn how to import data from external databases and Web-based resources, create custom menus to interact with a user, and examine how Excel can be used in business decision making.
Offering: Host
Grading: OPT
Credits: 0.25
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: None

QAC155 Working with Python
This course introduces students to programming, data management, and analysis with Python. Through a series of hands-on lab exercises, students learn to work with a variety of data using a high-level programming language and associated libraries to effectively manage and analyze their data. The emphasis is on data exploration and visualization and includes work with unstructured data generated by social media interactions. While there are no prerequisites, a basic familiarity with computing tools, an understanding of descriptive statistics, and a willingness to make mistakes and learn from them is expected.
Offering: Host
Grading: OPT
Credits: 0.25
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: None

QAC156 Working with R
This course introduces students to programming, data management, and analysis with R. Through a series of hands-on lab exercises, students learn to work with a variety of data formats and use R’s programming language and associated packages to effectively manage and analyze their data, with an emphasis on
data exploration and visualization. While there are no prerequisites, a basic familiarity with computing tools, an understanding of descriptive statistics, and a willingness to make mistakes and learn from them is expected.

Offering: Host
Grading: OPT
Credits: 0.25
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: None

QAC157 Working with SAS
This course introduces students to programming, data management, and analysis with SAS. Through a series of hands-on lab exercises, students learn to work with a variety of data formats and use SAS's programming capabilities to effectively manage and analyze their data, with an emphasis on data exploration and visualization. While there are no prerequisites, a basic familiarity with computing tools, an understanding of descriptive statistics, and a willingness to make mistakes and learn from them is expected.

Offering: Host
Grading: OPT
Credits: 0.25
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: None

QAC158 Working with Stata
This course introduces students to programming, data management, and analysis with Stata. Through a series of hands-on lab exercises, students learn to work with a variety of data formats and use Stata's programming capabilities to effectively manage and analyze their data, with an emphasis on data exploration and visualization. While there are no prerequisites, a basic familiarity with computing tools, an understanding of descriptive statistics, and a willingness to make mistakes and learn from them is expected.

Offering: Host
Grading: OPT
Credits: 0.25
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: None

QAC171 Business Modeling with Excel
Broadly speaking, this course is about evidence-based decision-making. It is intended for students with a background in Excel and VBA who want to learn how to use spreadsheets to develop business and financial models and communicate the results. Model building is different from the usual Excel skills as it focuses more on the analytical/mathematical aspects than on data wrangling and requires some basic familiarity with economic/financial models and data analysis tools. We will start with problems where the inputs are deterministic and we must decide on optimal allocation of resources. We will then advance to problems with stochastic inputs and explore solutions either through simulation or through optimization of goal functions. Through this work we will develop the appropriate programming skills (e.g., VBA) and learn to effectively use Excel to implement our models and display the results of our analyses.

Offering: Host
Grading: OPT
Credits: 0.25
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: None

QAC211 Digging the Digital Era: A Data Science Primer
The course introduces students to the practice of what has come to be known as data science. Using a multidisciplinary approach and data from a variety of sources that cover any aspect of everyday life—from credit card transactions to social media interactions and Web searches—data scientists try to analyze and predict events and behavior. The first part of the course defines the area and introduces basic concepts, tools, and emerging applications. We will describe how big data analysis affects both business practices and public policy and discuss applications in different areas/disciplines. We also discuss the ethical, legal, and privacy dimensions of big data analysis. In part two of the course, we work on data acquisition and management and introduce appropriate programming and data management tools. In part three, we concentrate on basic analytical and visualization techniques as we explore and understand the emerging patterns. Using a learning-by-doing approach in a computing laboratory, students will learn how to write computer programs in R—programming in R is a significant part of the course work—to access, organize, and analyze data through a series of small projects designed to illustrate the application of the techniques we develop for a variety of data sets and situations. Students will also engage in a semester-long project where they will access and use data from social media (Twitter) to address their own research questions.

Offering: Host
Grading: A-F
Credits: 1.00
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: None

QAC221 Modeling and Data Analysis: From Molecules to Markets
The development of models to describe physical or social phenomena has a long history in several disciplines, including physics, chemistry, economics, and sociology. With the emergence of ubiquitous computing resources, model building is becoming increasingly important across all disciplines. This course will examine how to apply modeling and computational thinking skills to a range of problems. Using examples drawn from physics, biology, economics, and social networks, we will discuss how to create models for complex systems that are both descriptive and predictive. The course will include significant computational work. No previous programming experience is required, but a willingness to learn simple programming methods is essential.

Offering: Crosslisting
Grading: A-F
Credits: 1.00
Gen Ed Area: NSM-PHYS
Identical With: PHYS221, CIS231
Prereq: None

QAC231 Introduction to (Geo)Spatial Data Analysis and Visualization
Geographic information systems (GIS) provide researchers, policy makers, and citizens with a powerful analytical framework for spatial pattern recognition, decision making, and data exploration. This course is designed to introduce social science and humanities students to spatial thinking through the collection, management, analysis, and visualization of geospatial data using both desktop and cloud-based platforms. Classes will consist of short lectures, hands-on training using different spatial analysis and geodesign technologies (e.g., ESRI ArcGIS, Google Fusion Tables, MapBox), group projects, critiques, and class discussions. Weekly readings and assignments will build skills and reinforce
QAC239 Proseminar: Machine Learning Methods for Text, Audio and Video Analysis

In this course, students will learn machine learning techniques to analyze text, audio, and video data. The course consists of three parts: text analysis, audio analysis, and video analysis. Each part will first introduce how these non-traditional data can be converted into mathematical objects suitable for computer processing and, particularly, for the application of machine learning techniques. Then students will learn a selection of supervised and unsupervised learning algorithms that are effective for text, audio, image/video analysis. Finally, students will explore major applications of these techniques such as sentiment analysis, speech emotion recognition, face recognition, pedestrian detection, keyframe extraction.

Offering: Host
Grading: A-F
Credits: 1.00
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: None

QAC241 Introduction to Network Analysis

This is an interdisciplinary hands-on course examining the application of network analysis in various fields. It will introduce students to the formalism of networks, software for network analysis, and applications from a range of disciplines (history, sociology, public health, business, political science). We will review the main concepts in network analysis and learn how to use the software (e.g., network analysis and GIS libraries in R) and will work through practice problems involving data from several sources (Twitter, Facebook, airlines, medical innovation, historical data). Upon completion of the course, students will be able to conduct independent research in their fields using network analysis tools.

Offering: Host
Grading: A-F
Credits: 1.00
Gen Ed Area: NSM-QAC
Identical With: CIS239
Prereq: COMP112 OR QAC155 OR QAC156

QAC250 An Introduction to Data Journalism

This course is designed to familiarize students with the basic principles and tools of data journalism and to provide a wider understanding of the role of basic data analysis in society. To that end, the course will focus on developing a solid familiarity with basic data analysis and visualization software. It will also focus on developing the tools of journalism: retrieving public data, interviewing people and databases, and the basic principles of journalistic writing. By the end of the course, students will be able to analyze data, identify stories within the data, and create a news story complete with data visualizations of publishable quality—a skill transferable to many fields and disciplines. Both online and traditional print platforms will be covered.

Offering: Host
Grading: A-F
Credits: 1.00
Gen Ed Area: SBS-QAC
Identical With: CSPL250, WRCT250

QAC251 Data Visualization: An Introduction

This course will introduce students to the principles and tools necessary to present quantitative information in a visual way. While tables and graphs are widely used in our daily lives, it takes skill to deconstruct what story is being told. It also takes a perceptive eye to know when information is being misrepresented with particular graphics. The main goals of the course are for students to learn how to present information efficiently and accurately so that we enhance our understanding of complex quantitative information and to become proficient with data visualization tools. Beginning with basic graphing tools, we will work our way up to constructing map visualizations and interactive graphs. This course will require a substantial amount of computation in R. No prior programming experience is necessary, but learning does require willingness and time.

Offering: Host
Grading: A-F
Credits: 1.00
Gen Ed Area: NSM-QAC, SBS-QAC
Identical With: CIS251
Prereq: None

QAC260 Special Topics in Computer Science

This course is designed for nonmajors who wish to pursue some topic in computer science beyond introduction to programming. Topics will vary according to the instructor.

Offering: Crosslisting
Grading: A-F
Credits: 1.00
Gen Ed Area: NSM-MATH
Identical With: COMP260
Prereq: COMP112

QAC261 Project-Based Programming for Research

This project-based course will introduce students to programming in the context of research design, data visualization, and analysis of big data, focusing on the essential concepts and tools needed to carry out research and problem solving and to keep abreast of new technologies. We will survey these topics by combining scientific problems and modern programming approaches, and students will learn the fundamentals of programming required for structuring and conducting research.

Offering: Crosslisting
Grading: A-F
Credits: 1.00
Gen Ed Area: NSM-PSYC
Identical With: PSYC381
Prereq: None

QAC282 Economics of Big Data

"Big data" is a popular buzzword that describes techniques using very large datasets, often from nontraditional sources. Many technology firms essentially base their businesses on big data; Google, Facebook, and Amazon are all examples. Increasingly, there are opportunities and pressures to employ these techniques in other areas of the economy and society such as government, health care, and education. This course examines big-data analysis techniques and how they relate to conventional economic statistics; the effect of big data on the economy, society, and privacy; and practical methods of big-data analysis using the R statistics package.

Offering: Crosslisting
Grading: OPT
Credits: 1.00
Gen Ed Area: SBS-ECON
Identical With: ECON282
Prereq: ECON101 OR ECON110
QAC301 Statistics Education Practicum
This course will serve students who are pursuing their undergraduate degree in a variety of disciplines but who want to expand their skills in statistics and applied data analysis in preparation for a future career. It will also serve students who are currently pursuing independent, quantitative research at the undergraduate or graduate level. The course will center on personal interaction in support of introductory statistics students. Active peer mentoring and supporting experiences will be based on the theory that good teachers (and learners) of statistics need to be developed, as opposed to being trained. In line with this theory, this hands-on course will provide an extensive opportunity to build specific knowledge regarding teaching and learning in the area of data-driven statistical inquiry. Students enrolled in this course will (1) attend statistics-mentoring development sessions (one hour per week); (2) provide one-on-one support for introductory statistics students during workshop-oriented class sessions (three hours per week); (3) lead small group-mentored meetings for five to six statistics students (one hour per week); and (4) monitor and critique progress on applied data assignments (one hour per week). In addition to these hands-on experiences, students will pursue a project aimed at furthering the field of statistics education. Projects may take the form of course evaluation, content/conceptual curriculum development, or translation of educational statistical software materials.

Similar to QAC380 (Introduction to Statistical Consulting), this course is aimed at providing students with an opportunity to enhance their statistical skills beyond the introductory level.
Offering: Host
Grading: OPT
Credits: 1.00
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: None

QAC302 Political Science by the Numbers
This course covers the basics of probability theory and statistics. The main purpose of this course is to promote the understanding of statistical concepts and how these concepts can be used to make inferences about the political world. Topics include probability distributions, correlation analysis, linear regression, generalized linear models, maximum likelihood, logistic regression, causal inference, experiments, and non-parametric modeling. Lectures will mainly cover theory, while readings will connect the concepts described during lecture to problems in political science. Whenever possible, the instructor will draw upon research in political science to illustrate the why and how of a given concept or technique. Demonstrations will allow students to "play around" with abstract statistical concepts. Most lectures will have an interactive component involving class participation. Problems sets will cover some of the more technical aspects of what we discuss in class along with applications using real data.
Offering: Crosslisting
Grading: A-F
Credits: 1.00
Gen Ed Area: SBS-GOV
Identical With: GOVT367
Prereq: None

QAC305 Exploratory Data Analysis and Pattern Discovery
The course introduces the theory and practice of exploring, describing, summarizing and detecting patterns of interest in complex datasets. Various approaches including aggregation, clustering, data visualization, and latent variable modeling will be employed. This course will give students an opportunity to develop computational skills (primarily in SAS) and to learn how to discover and interpret relationships in unstructured observational data. The applications and examples for this course will be broad and relevant to many fields of study.
Offering: Host
Grading: A-F
Credits: 1.00

QAC307 Experimental Design and Causal Inference
The course provides the foundations and statistical thinking to design, collect, and analyze experimental data and introduces appropriate techniques for observational data when causal inference is the objective of the analysis. Throughout the course, we introduce and compare various experimental designs. We will discuss sample size and power calculations as well as the advantages and disadvantages of each of these designs. With observational data, we will explore difference-in-difference models, propensity score matching techniques, regression discontinuity designs. This course gives students the opportunity to develop further their computational skills as we learn how to describe, interpret, control, and draw inferences from experimental and observational data.
Offering: Host
Grading: A-F
Credits: 1.00
Gen Ed Area: NSM-QAC, SBS-QAC
Identical With: CIS307
Prereq: QAC201 OR PSYC200 OR MATH132 OR ECON300

QAC311 Longitudinal Data Analysis
Work across different fields, from medicine and public health to social sciences and education, often involves the collection and analysis of longitudinal data--combination of cross-sectional and time series (repeated measures for the unit of observation) data. This rich data structure provides opportunities to explore questions that could not be addressed with simpler data sets, but at the same time requires special considerations because we are analyzing observations that are not independent. The course introduces students to appropriate graphical exploration of the data and the specification and estimation of fixed and random-effects models. It also develops the basic framework for difference-in-differences models and explores their applications.
Offering: Host
Grading: A-F
Credits: 0.50
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: [QAC201 or SOC257 or GOVT201 or PSYC280 or NS&B280] OR [QAC380 or PSYC395] OR ECON300 OR [GOVT367 or QAC302] OR PSYC200

QAC312 Hierarchical Linear Models
Research questions cannot always be explored by collecting data with independent observations. Sometimes this is due to limitations or constraints on the data collection method, and other times our questions pertain to data that are measured at both the individual and group levels (e.g., patients from different hospitals or students from different schools that belong to different districts). Hierarchical linear models (HLM), also called multi-level or mixed models, explicitly model such nested data structures and address analytical and estimation issues not accounted within the framework of the classical linear model. Using data sets from different fields of study (e.g., education, medicine, and health) students will learn to formulate multilevel research questions, estimate and critically examine HLM applications.
Offering: Host
Grading: A-F
Credits: 0.50
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: [QAC201 or SOC257 or GOVT201 or PSYC280 or NS&B280] OR [QAC380 or PSYC395] OR ECON300 OR [GOVT367 or QAC302] OR PSYC200

QAC313 Latent Variable Analysis
The course is an introduction to latent variable modeling. Students will learn the fundamental statistical methods for structural equation modeling (SEM), including principal component analysis, confirmatory factor analysis, path analysis, and SEM for both quantitative and binary observed variables. In addition, students will learn the basic components of SEM, such as assumptions,
testing model fit and indices of fit, testing competing models, estimation methods, and issues in model identification. Students will learn to develop structural equation modeling using AMOS, R, and/or Mplus statistical software.

Offering: Host
Grading: A-F
Credits: 0.50
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: [QAC201 or SOC257 or GOVT201 or PSYC280 or NS&B280] OR [QAC380 or PSYC395] OR ECON300 OR [GOVT367 or QAC302] OR PSYC200

QAC314 Survival Analysis
Survival or event history analysis focuses on modeling and analysis of time-to-event data--such as onset of a disease; duration of a strike; failure of a biological, physical, or social system; or recidivism. This course introduces students to survival and hazard functions, the analysis of censored data using parametric and non-parametric estimation methods; compares survival curves for different groups; and discusses competing risk models. The emphasis is on the applications of the different methods with the objective of broadening computational skills in R and/or SAS and to reinforce statistical writing and communication. These skills will be applied to a variety of problems in political science, public health, engineering, and medicine.

Offering: Host
Grading: A-F
Credits: 0.50
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: [QAC201 or SOC257 or GOVT201 or PSYC280 or NS&B280] OR [QAC380 or PSYC395] OR ECON300 OR [GOVT367 or QAC302] OR PSYC200

QAC323 Bayesian Data Analysis: A Primer
This course introduces the applied principles of Bayesian statistical analysis. The Bayesian paradigm is particularly appealing in research where prior research and historical data are available on parameters of interest. This course will teach students appropriate techniques for analyzing data of this nature as well as broaden computational skills in R. The course will lay the foundation for Bayesian data analysis that students can use to further develop skills in decision making.

Offering: Host
Grading: A-F
Credits: 0.50
Gen Ed Area: NSM-QAC, SBS-QAC
Identical With: CIS323
Prereq: MATH132 OR ECON300 OR [GOVT367 or QAC302]

QAC344 Advanced GIS and Spatial Analyses
A geographic information system (GIS) is a powerful database that allows for the collection, manipulation, analysis, and presentation of spatially referenced data. GIS technologies facilitate natural science, social science, and humanities research and any other project that uses location-based data. This course will focus on individual projects conducted within a collaborative learning framework. Each student is responsible for developing and producing a semester-long project focused on advanced spatial data analyses and/or advanced cartographic design using a GIS. Students will enter the course with an individual or small team (2–3 students) project in mind. The project may be a component of a senior thesis, work on a faculty member's research project, a community-based service-learning project, and so on. Course sessions will be a mix of studio time for projects (e.g., work time, critiques), skill development (e.g., lectures, student-led skills-training sessions), and intellectual advancement (e.g., guest speakers, conference attendance). Specific skills-training sessions will be determined by components of each project.

Offering: Crosslisting
Grading: A-F
Credits: 1.00
Gen Ed Area: NSM-EES
Identical With: E&ES380, E&ES590

Prereq: QAC231 OR EES322

QAC356 Advanced R: Building Open-Source Tools for Data Science
An in-depth exploration of the R programming language and development environment; from syntax and data structures, to data analysis and visualization, report writing, project management, and software development. Students will work in teams to create open-source data science applications that can be used to manage and prepare data, perform statistical analysis, and create compelling visualizations and reports. Students will also have an opportunity to contribute to a shared Wesleyan R package that contains datasets and functions that are useful for teaching introductory and advanced data science courses.

Offering: Host
Grading: A-F
Credits: 1.00
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: QAC201 or MATH132 or ECON300 or QAC302 AND COMP112 AND QAC156 or QAC211 or QAC251

QAC380 Introduction to Statistical Consulting
In this course, students will be exposed to realistic statistical and scientific problems that appear in typical interactions between statisticians and researchers. The goal is for students to apply what they have learned in their basic statistics and data analysis courses to gain greater experience in the areas of research collaboration, data management and analysis, and writing and presenting reports on the results of the analyses. An important objective of the course is to help develop communication skills, both written and verbal, as well as the professional standards and interpersonal skills necessary for effective statistical consulting.

Offering: Host
Grading: A-F
Credits: 1.00
Gen Ed Area: NSM-QAC, SBS-QAC
Identical With: PSYC395
Prereq: MATH132 OR ECON300 OR PSYC200 OR [QAC201 or SOC257 or GOVT201 or PSYC280 or NS&B280]

QAC381 QAC Praxis Service Learning Lab
As a service learning lab, this course provides students with an opportunity to further develop their abilities to analyze data and apply their knowledge and statistical computing skills as they work closely with nonprofit community partners on data analytic projects. Students in the course will identify research questions of interest to the community partner, gather and manage data, conduct statistical analyses, and interpret and summarize results. The service component involves providing statistical consulting to community partners by formulating and completing data analytic projects, the results of which may be used to improve services, identify areas requiring increased services and areas in which services can be made more efficient, as well as improving data collection, data reporting, and organizational functioning.

Offering: Host
Grading: A-F
Credits: 1.25
Gen Ed Area: SBS-QAC
Prereq: MATH132 OR ECON300 OR PSYC200 OR QAC201 OR SOC257 OR GOVT201 OR PSYC280 OR NS&B280

QAC385 Applications of Machine Learning in Data Analysis
The course provides a broad overview of machine learning algorithms and focuses on their application in data mining. Building on a basic background of regression analysis, and following a learning-by-doing approach, students are introduced to data mining tools and techniques that are used to identify patterns and relationships in large and complex data. While the emphasis is on intuition and application rather than theoretical results, through different case studies, students are introduced to the fundamentals of the different methods and
learn how to conceptualize a problem, analyze it using appropriate tools, and communicate their results.

Offering: Host
Grading: A-F
Credits: 1.00
Gen Ed Area: NSM-QAC, SBS-QAC
Prereq: QAC211 OR [PHYS221 or QAC221 or CIS231] OR ECON300 OR [GOVT367 or QAC302] OR MATH231 OR MATH232

QAC386 Quantitative Textual Analysis: Introduction to Text Mining
We encounter computerized processing of text in almost every field of life. Google tries to infer the meaning of our search queries, online review engines try to extract information about what products are popular with the users, and across different fields scholars analyze text for insights into the processes and phenomena they study. This course will introduce you to the skills necessary to mine text for information and knowledge. You will learn how to use R to retrieve text from a variety of sources, how to use regular expressions to identify which pieces of text are useful to your study, and how to use techniques from data mining to analyze the processed text to extract information and for classification and prediction.

Offering: Host
Grading: A-F
Credits: 1.00
Gen Ed Area: SBS-QAC
Prereq: QAC211 OR ECON300 OR [GOVT367 or QAC302]

QAC401 Individual Tutorial, Undergraduate
Topic to be arranged in consultation with the tutor.

Offering: Host
Grading: OPT

QAC402 Individual Tutorial, Undergraduate
Topic to be arranged in consultation with the tutor.

Offering: Host
Grading: OPT

QAC419 Student Forum
Student-run group tutorial, sponsored by a faculty member and approved by the chair of a department or program.

Offering: Host
Grading: Cr/U

QAC420 Student Forum
Student-run group tutorial, sponsored by a faculty member and approved by the chair of a department or program.

Offering: Host
Grading: Cr/U

QAC492 Teaching Apprentice Tutorial, Undergraduate
The teaching apprentice program offers undergraduate students the opportunity to assist in teaching a faculty member's course for academic credit.

Offering: Host
Grading: OPT