UCONN
UNIVERSITY OF CONNECTICUT

Office of the Provost
Anne D'Alleva, Ph.D.
Provost and Executive Vice President
for Academic Affairs

December 7, 2022

TO: Members of the Board of Trustees

FROM: Anne D'Alleva, Ph.D.
Provost and Executive Vice President for Academic Affairs

RE: Bachelor of Arts in Applied Data Analysis

RECOMMENDATION:

That the Board of Trustees approve a new undergraduate major in Bachelor of Arts in Applied Data Analysis in the College of Liberal Arts and Sciences.

BACKGROUND:

Data science and the analysis of quantitative data are rapidly growing fields that are revolutionizing science and society. The new Bachelor of Arts in Applied Data Analysis will provide undergraduate students with amajor that prepares them for a range of careers in an area of rapid job growth. Data science-related jobs are anticipated to see significant growth in employment over the next decade or two. The U.S. Bureau of Labor Statistics anticipates an increase of about 40,000 data scientist positions between now and 2031, up about 36% (2021-2031), much faster than other occupations. This growth is faster than any other occupation requiring a college degree except for nurse practitioners. The median salary for data scientists is currently $101,000. This data science major will provide skills that will make students competitive for other related high-growth occupations, like statisticians and web developers.

Prospective students for this program will be STEM and non-STEM interested students who are considering careers where the understanding and application of data is central to functional tasks. The Bachelor of Arts degree is organized as an interdepartmental major. The degree has an advisory board which makes decisions about the curriculum, including approval and re-authorization (from time to time) of skill, domain, and elective courses. This new Bachelor of Arts in Applied Data Analysis will provide students with a degree option that will broaden their field of interest and career opportunities. Given the application process to be accepted and course requirements, we expect a cohort of 40-50 students per semester. Based on this distribution, we would estimate around 300 students in the program once it has been established. The New England Regional Tuition Break program may also bring an additional 50-100 students, since most of the participating states do not have this type of undergraduate degree program.

352 MANSFIELD ROAD, UNIT 1086
STORRS, CT 06269-1086
PHONE: 860.486.4037
EMAIL: anne.dalleva@uconn.edu
WEB: provost.uconn.edu

UCONN

PROPOSAL FOR

BACHELOR OF ARTS IN APPLIED DATA ANALYSIS
30.7001

COLLEGE OF LIBERAL ARTS AND SCIENCES
UNIVERSITY OF CONNECTICUT

Introduction and Rationale

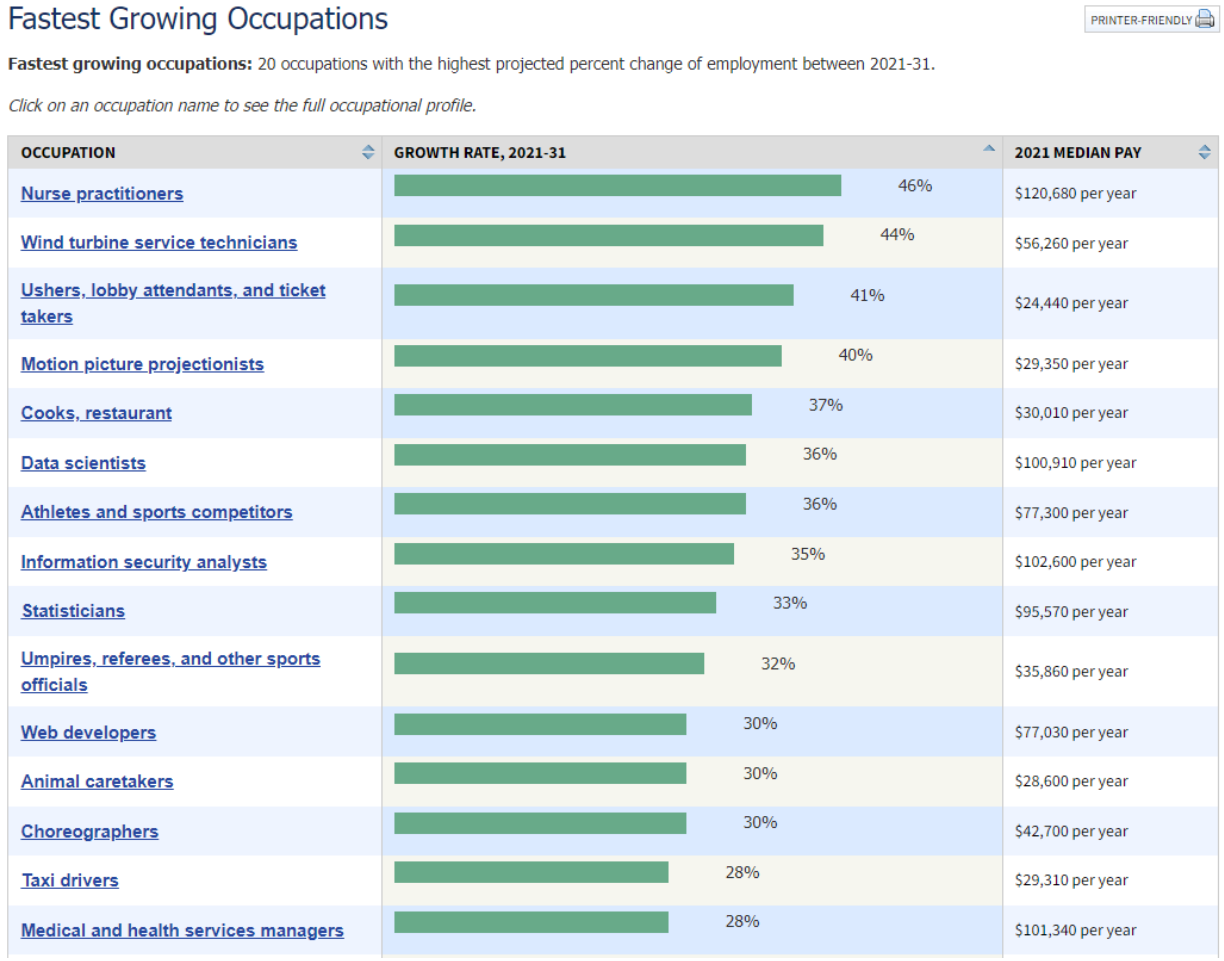
Data science and the analysis of quantitative data are rapidly growing fields that are revolutionizing science and society. Work is becoming increasingly more data-driven, and this affects the jobs that are available and the skills that are required. As data and data analysis tools become more widely available, more aspects of the economy, society, and daily life will become dependent on them. While today the term "data scientist" typically describes a knowledgeable worker who is principally occupied with analyzing complex and massive data resources, data science spans a much broader array of activities. These involve applying data science principles for data collection, storage, integration, analysis, inference, communication, and ethics. In future decades, undergraduates interested in many specialties will benefit from a fundamental awareness of and competence in data science.

The changing workplace requires more and more people with a basic understanding of data science and a substantial cadre of talented graduates with highly developed data science skills, acquired through substantial coursework and practice. Graduates of these types of programs can expect to find work in almost all occupational realms and will serve in a number of roles, including operating and designing the analytical systems, preparing data, coordinating analysis, visualizing output, and supporting data-driven decision making. Journalists, administrators in the public and private sector, artists, lawyers, teachers, and others will also increasingly need to understand and use data. Hence there is a great need to prepare students for the data-enriched world of the rest of this century.

Data science-related jobs are anticipated to see significant growth in employment over the next decade or two. The BLS anticipates an increase of about 40,000 "data scientist" positions between now and 2031, up about 36% (2021-2031) for the occupation of Data scientist", much faster than other occupations, and faster than any other occupation requiring a college degree except for nurse practitioner (Figure 1). The median salary data scientists are currently well above average: $101,000.[1] Data science training provides skills that would make students competitive for other related high growth occupations, like statisticians

[1] https://www.bls.gov/ooh/math/data-scientists.htm

and web developer. About 120 colleges and universities currently offer Data Science BA degrees, including many of UConn's peer and aspirant institutions: such as Boston University, Iowa State, Northeastern, Penn State, Purdue, Rutgers, SUNY-Albany, UC-Davis, UC-Irvine, University of Georgia, University of Iowa, UMass-Dartmouth, and URI.



Fastest Growing Occupations

PRINTER-FRIENDLY

Fastest growing occupations: 20 occupations with the highest projected percent change of employment between 2021-31.

Click on an occupation name to see the full occupational profile.

OCCUPATION	GROWTH RATE, 2021-31	2021 MEDIAN PAY
Nurse practitioners	46%	$120,680 per year
Wind turbine service technicians	44%	$56,260 per year
Ushers, lobby attendants, and ticket takers	41%	$24,440 per year
Motion picture projectionists	40%	$29,350 per year
Cooks, restaurant	37%	$30,010 per year
Data scientists	36%	$100,910 per year
Athletes and sports competitors	36%	$77,300 per year
Information security analysts	35%	$102,600 per year
Statisticians	33%	$95,570 per year
Umpires, referees, and other sports officials	32%	$35,860 per year
Web developers	30%	$77,030 per year
Animal caretakers	30%	$28,600 per year
Choreographers	30%	$42,700 per year
Taxi drivers	28%	$29,310 per year
Medical and health services managers	28%	$101,340 per year

Figure 1: Fastest growing Occupations 2021-2021 and median incomes according to US Bureau of Labor Statistics. https://www.bls.gov/ooh/fastest-growing.htm

Enrollment Projections

Prospective students for this program will be STEM and non-STEM interested students who are considering careers where the understanding and application of data is central to functional tasks. Given the emerging nature of data science in institutions with similar profiles to UConn, we can only provide enrollment projections based on educated estimates and understanding of our existing institutional context. Many of our students in the Humanities and Social Sciences have been eager to approach their various disciplines in a quantitative fashion, but our academic plans in these areas have not been particularly conducive to this approach. This new Bachelor of Arts in Applied Data Analysis will provide students with a degree option that will broaden their field of interest and career opportunities. Given the application process to be accepted and course requirements, we expect a cohort of 40-50 students per semester. Based on this distribution, we would estimate around 300 students in the program once it has been established. The New England Regional Tuition Break program may also bring an additional 50-100 students, since most of the participating states do not have this type of undergraduate degree program**.**

Required Resources

Specific courses and requirements are detailed below. The program that we have devised contains courses that are all taught in CLAS. The BA degree draws on courses from multiple departments in the college. The BA degree is organized as an interdepartmental major. The degree has an advisory board which makes decisions about the curriculum, including approval and re-authorization (from time to time) of skill, domain, and elective courses. Depending on the growth in the size of the major and the fact that basic and advanced data science skills are an increasingly important part of the training of the faculty in many existing disciplines–e.g., economics, political science, sociology, biology– we anticipate that the major can grow as faculty in various departments in the college that become more data-science oriented. CLAS has committed to several lines in affiliated departments this year for faculty who will likely contribute to this major.

Justification

In March 2020, Dean Juli Wade called a meeting to discuss the creation of a new Undergraduate Data Science Program within CLAS, and asked interested CLAS Department Heads to appoint members to a CLAS Undergraduate Data Science Committee for this purpose. This committee, consisting of members of the Departments of Statistics, Political Science, Mathematics, Economics, Geography, Geosciences, Public Policy, Ecology and Evolutionary Biology, and Molecular Cell Biology, has since been meeting regularly to develop this major, as well as the BS in Statistical Data Science. In addition, the Departments of Philosophy, Cognitive Sciences, Sociology, and Marine Science were involved in aspects of the curriculum, making it a thoroughly interdisciplinary major. The curriculum for the degree was approved by the CLAS Courses and Curriculum Committee on 10/18/2022.

Analytic training includes courses where students will learn to:

-- formulate good questions and determine the types of data appropriate to answer those questions,
-- collect, retrieve, manipulate, store, analyze, and report on information *in an ethical manner,*
-- conduct work that is reproducible, and
-- make appropriate inferences from data analysis.

Students completing a BA in Applied Data Analysis must attain competence in four core areas of *data science* as suggested in the 2018 National Academy of Science report *Data Science for Undergraduates: Opportunities and Options:*

A. Computer Programming, data generation, and analysis: Almost all data generation and analysis require the manipulation of large amounts of digitized information. Because most tasks associated with data analytics involve amounts of data that cannot be (re)processed by hand or involve processing data to be used by different analytical software and hardware and for different practical applications, degree recipients must have an elementary understanding of computer systems and languages, data structures and control, and algorithmic development and utilization. To be able to address specific, complex problems with students attaining the BA degree will learn to collect and manage the appropriate information such that it can be utilized effectively by individuals and organizations.

B. Data analysis: This skill includes a core set of features consistent with probability and statistics in quantitative data analysis: e.g., sampling, randomness, experimental/observational research designs, parametric/non-parametric estimation and inference up through at least multiple linear regression.

C. Data visualization: Visualization refers generally to the effective presentation and communication of data in a manner that can stand alone as a communication tool or that complements the narrative text. As part of data visualization training, students will learn modern visualization standards and how to use computer visualization tools. They will also learn to effectively communicate to different audiences and avoid engaging in the misrepresentation of data.

D. Ethics of data collection and use: The ethical challenges of collecting and using data to inform decision-making are enormous. This is particularly the case when much of the data used concerns observations about behaviors or characteristics collected without the full knowledge of those being observed. The very power of data science makes it important that all parts of the data science curriculum educate students about the ethical use of data science tools.

In addition to these skill requirements, students must learn about a specific substantive domain area topic of social or scientific relevance.

After completing the skill and domain area training, students will conduct a final research project which applies all the core data science skills to a practical problem in the substantive domain training area.

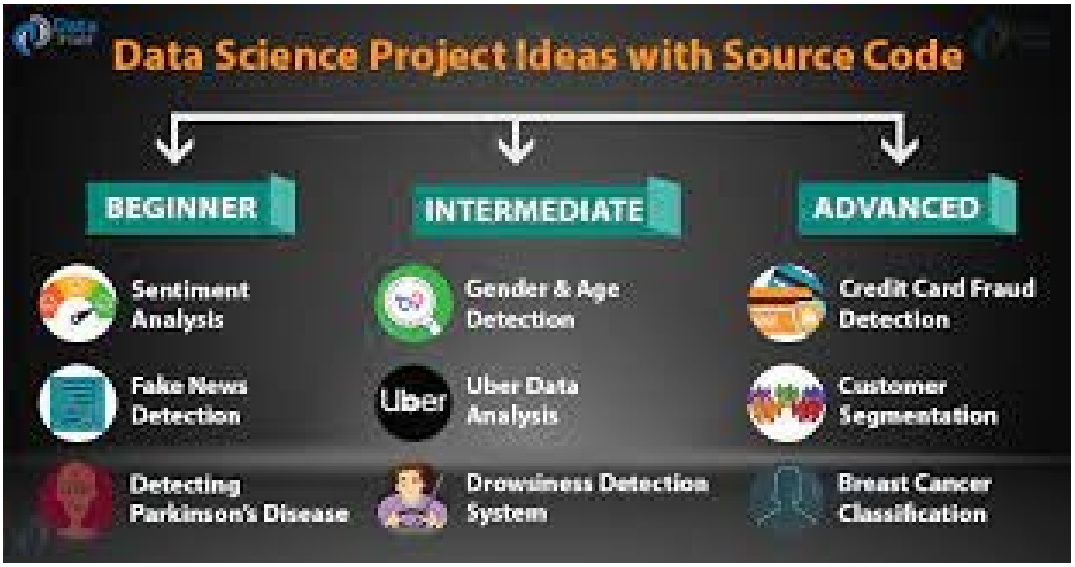


Figure 2: Examples of Data Science final projects (https://data-flair.training/blogs/data-science-project-ideas/)

Bachelor of Arts Degree Curriculum

The BA requires 36 credits, with one course in four core areas, a nine-credit domain sequence, STAT 3255 (Introduction to Data Science), and a Capstone course of at least 3 credits. Students meet the "writing in the major" requirement in a domain-specific W course, or in a Capstone W course.

The four core area requirements are:

1. *Programming and data management:* 1 course (3 credits) STAT 2255 or COGS 2500
2. *Basic Data Analysis:* 1 course (3 credits): STAT 3215Q
3. *Data Ethics* 1 course (3 credits): PHIL 3202
4. *Data Visualization* 1 course (at least 3 credits): STAT 3675Q or GEOG 3510

To meet the domain area requirement, students must select one of the following domains areas, and complete at least 9 credits. One of these domain courses must be a W course.

American Political Institutions: POLS 3600, POLS 3601, POLS 3604, POLS 3606; W course: POLS 3603WQ; Capstone: DSDA 4815

American Political Representation POLS 2607, POLS 3612, POLS 3617, POLS 3618, POLS 3625; W course: POLS 3608W; Capstone: DSDA 4815

Earth Data Science: GSCI 2800, GSCI 3020, GSCI 3710, GSCI 4230, GSCI 4810; W course: GSCI 2050W; Capstone: GSCI 4150

Public Management and Policy: PP 3032, PP 3033, PP 3098, PP 4031, PP 4034; W course: PP 3020W; Capstone: DSDA 4815

Survey Research Methods: PP 2100, PP 3030, PP 3098; W course: PP 3020W; Capstone: DSDA 4815

Population Dynamics: SOCI 2110(W), SOCI 2651(W), SOCI 2660(W), SOCI 2820(W); SOCI 2901(W); SOCI 3971(W); W course: W version of any of the above domain courses; Capstone: DSDA 4815

(Domain areas may be added by petitioning the advisory board.)

To reach 36 credits, additional credits may be taken from approved domain areas or the list of courses below.
GEOG 2500, GEOG 3500Q, STAT 2215Q, STAT 3025Q, STAT3515Q, STAT 3375Q

Explanation for core courses

STAT 2255 (Statistical Programming) or **COGS 2500Q** (Coding for Cognitive Scientists) addresses *programming and data management.* STAT 2255 introduces statistical programming via Python including data types, control flow, object-oriented programming, and graphical user interface-driven applications such as Jupyter notebooks. The emphasis of the course is on algorithmic thinking, efficient implementation of different data structures, control and data abstraction, file processing, and data analysis and visualization. COGS 2500 is an introduction to programming for students with little or no prior programming experience. Its goal is to familiarize students with core concepts and essential skills. Like STAT 2255, COGS 2500 also uses the Python programming language because it is both accessible to beginners and widely used in real-world scientific programming. The concepts and skills are general, however, and will be helpful in mastering other programming languages as well.

STAT3215Q (Applied Linear Regression in Data Science) addresses *basic data analysis* as it covers simple linear regression and correlation analysis, multiple linear regression, analysis of variance, goodness of fit, comparing regression models through partial and sequential F tests, dummy variables, regression assumptions and diagnostics, model selection and penalized regression, prediction and model validation, principles of design of experiments, one-way and two-way analysis of variance. (Additionally, STAT1000Q/STAT1100Q or equivalent is a prerequisite for entry into the major, and covers sampling, randomness, and experimental/observational research designs, among other topics.)

STAT3255 (Introduction to Data Science) addresses *all core areas of data science* by introducing data science for effectively storing, processing, analyzing, and making inferences from data. Topics include project management, data preparation, data visualization, statistical models, machine learning, distributed computing, and ethics. It also provides training in the ability to formulate good questions; assess which kinds of data are appropriate to answer those questions; conduct ethical data collection, manipulation, and analysis that is reproducible; and make appropriate inferences based on the data.

To meet the core area of *data visualization*, students must take at least three credits of **GEOG3510** (Cartographic Techniques) or **STAT3675Q** (Statistical Computing); GEOG3510 covers methods for representing geographic data in tables, graphs, and maps emphasizing proper application, integration, and interpretation of methods in data visualization. STAT3675Q, while arguably also a *programming* course, covers dynamic reports, and both basic and advanced graphics (with ggplot2) in the R programming language.

The core area of *ethics* will be addressed in both **PHIL 3202** (Data Ethics), and also **STAT 3255** (Introduction to Data Science). PHIL 3202 will introduce students to issues of ethics and equity in the contemporary practice of data science. The ability to collect, store, process, and analyze ever greater amounts of data offers great opportunities, as well as potential perils. Topics to be covered will include

systematic approaches to assessing ethical issues; privacy and confidentiality; defining research and the responsibilities associated with conducting ethical research; implicit and structural biases in data collection and analysis. STAT 3255 further covers the American Statistical Association's *Ethical Guidelines for Statistical Practice*, designed to help statistical practitioners make decisions ethically.

The capstone course, **DSDA 4815** (or GSCI 4150 for those completing the Geosciences domain), requires students to combine their domain knowledge with the core areas of data science in a final culminating research project. Capstone projects will include computational analyses of big datasets, including problem-specific programming (e.g., using shell, R, and/or Python), statistical analysis, and data visualization.

Students meet the university "writing in the major" requirement with a W course in their domain area.

Information literacy involves a general understanding of and competency in three integrally related processes:

- Information development and structure – an understanding of how information is created, disseminated and organized;
- Information access – an understanding of information communication processes and a facility with the tools required to tap into these processes;
- Information evaluation and integration – an ability to evaluate, synthesize and incorporate information into written, oral, or media presentations.

In addition to the basic competency achieved in ENGL 1007, ENGL 1010, ENGL 1011, ENGL 2011 or equivalent, students will receive instruction on how to conduct an effective search for information on the web for applicable topics in the required capstone and W courses.