

**Additional Information Provided
Regarding Revised Proposal
(Round 3)**

UNIVERSITY OF CALIFORNIA, RIVERSIDE

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO



SANTA BARBARA • SANTA CRUZ

PROFESSOR VASSILIS J. TSOTRAS
DEPARTMENT OF COMPUTER SCIENCE
Bourns College of Engineering
University of California
Riverside, CA 92521
Tel: 951-827-2888
Fax: 951-827-4643
Email: tsotras@cs.ucr.edu

To: Professor Dylan Rodriguez, Chair
Academic Senate

and

Professor Harry Tom, Chair
Academic Senate Committee on Planning and Budget

From: Vassilis Tsotras
Professor, Computer Science & Engineering
Co-chair, Adhoc Committee for the Design of the Data Science Major

Re: Response to Senate Review of Bachelor of Science in Data Science

Date: November 5, 2019

A handwritten signature in dark ink, appearing to read "V. Tsotras", is positioned to the right of the "From:" line.

Dear Professors Rodriguez and Tom,

Thank you for sharing the comments from the Planning and Budget Committee (P&B) as well as the Executive Council about the Bachelor of Science in Data Science (BSDS). First, I would like to inform the P&B committee that we were not aware of their comments on the initial version of the Data Science Major proposal and hence we could not address them in the revised version. We would also like to thank both committees for their comments which we will address below.

The new BSDS will be funded as any other major on campus. We apologize if the previously submitted budget justification implied otherwise. Specifically, student tuition will be distributed to colleges as with any other program. This means that the tuition portion due to BSDS major course enrollment (and associated lab fees or the like) will be distributed to colleges proportionally to the enrollments. This will cover teaching costs like any other student enrollment would cover teaching costs (including instructors, TAs, labs, etc.) Similarly, the tuition portion distributed according to majors will also be distributed to colleges proportionally to the BSDS majors in that college. This will cover advising and administrative overhead just as any other major in that college would. In this way, the per-course and the per-major costs of majors in BSDS would be covered just as the same costs would be covered for other majors on campus.

The initial target of 50 majors is not a long-term cap. It is a planning estimate and represents the desire to *initially* keep enrollments around this level to allow the program to grow gracefully to its natural size. As mentioned above, any enrollment growth in BSDS would be supported just as enrollment growth in any other BS major would be supported.

We would separately note that if any new courses are required for the major, they would be added in a particular department, and therefore their teaching costs would be covered by the enrollment in students in that course through the campus budget model, just as any other new course in the department.

Changes to the major, including adding additional "applications" courses would go through the BSDS program committee, as standard for any program. We would also like to note that the current proposal contains an initial list of application course sequences and new such sequences will be added as the program proceeds. In the initial list we included application courses that: (i) are related to data, (ii) have prerequisites that are satisfied by the coursework in the BSDS major and (iii) are given often enough so that the students in the BSDS major are not delayed. We expect and will welcome new such courses from other disciplines as this will make the DSBS major even stronger.

Revised Proposal (Round 2)

Proposal for the new undergraduate major in Data Science Revised in August 2019

1. *Name of the academic program and the department or unit that will administer the program.*

Name: Data Science Undergraduate Major

Administration: The Data Science major will be administered jointly by the Department of Computer Science and Engineering (CSE) and the Department of Statistics.

2. *A thorough justification, including the motivation for the creation of the program in terms of student interest and professional or academic importance.*

Data has become ubiquitous in everyday life, impacting every profession, from entry-level office workers to CEOs, from team coaches to general managers, from accountants to CFOs. Businesses now have data available to them at a scale that is historically unprecedented; harnessing this data for insight on what customers want provides them with a competitive advantage. Traditional companies (Ford, Walmart, General Electric, etc.) today pride themselves as being transformed to big-data businesses. The field of Data Science has emerged to address the proliferation of data and the need to manage and understand it.

Data Science lies at the intersection of Computer Science and Statistics, and its solutions already serve a variety of application domains in science, engineering and business. While there is a plethora of M.S. offerings in Data Science, relatively few undergraduate Data Science programs are currently offered. A recent search for Data Science undergraduate degrees (<http://datascience.community/colleges>) indicates that such offerings are typically provided as (i) minors to existing B.S. programs (Computer Science, Math or Statistics), (ii) certificate programs, or (iii) programs with very specific focuses (Business Data Analytics or Computational Analytics being the most common ones). Within the UC system, currently there are three existing undergraduate Data Science degree programs offered at UC Irvine (2015), UC Berkeley (2018) and UC San Diego (2018). The proposed UCR Data Science undergraduate program has thus the potential to fill an important gap. Creating a new Data Science undergraduate degree will be instrumental in educating future data scientists by building knowledge bottom-up, covering both essential knowledge from Computer Science (in managing data) and Statistics (in analyzing data), and integrating this knowledge with applications to other domains and to real-life problems.

Data permeates all aspects of science, engineering, and other academic disciplines. Yet, a comprehensive program studying how data can be collected, transformed, analyzed, and used to solve problems across academic disciplines does not currently exist at UCR at the undergraduate level. Statistics and Computer Science degree programs discuss some aspects of this, but our proposed program provides an interdisciplinary view of how to employ data and the role of data across other academic disciplines. The rise of data science as a term reflects a new academic discipline, studying data itself. This program will educate students in these analyses, thought processes, and ways of viewing the world through a data lens.

Through its interdisciplinary nature, a Data Science undergraduate program offers a great opportunity to serve as a pathway for professional careers in various areas. The proposed program is different from existing programs in that its students will complete course sequences in other departments (e.g., economics, business, sociology, earth sciences, biology,

bioinformatics, and astronomy) where they will learn how Data Science principles are applied in these domains.

Fueled by the explosion of data, Data Science jobs have proliferated and the demand for data scientists is extremely high; moreover, this demand is expected to be strong for years to come. A recent McKinsey report forecasted a need for hundreds of thousands of data scientists in the next decade. Another study by IBM found more than 2.3 million data science and analytics job listings in 2015, and both job openings and job demand are projected to grow significantly. Three-fifths of the data science and analytics jobs are in the finance and insurance, professional services, and information technology sectors, but the manufacturing, health care, and retail sectors also are hiring significant numbers of data scientists. According to Glassdoor, a recruiting site, Data Scientist is the best job in the US (for the last three years in a row) with around 110K median base salary. We thus expect that the new program will be in high demand among students and will serve well the UCR community.

Furthermore, many application areas related to Data Science, such as astronomy, biology, and economics, historically have gender-balanced enrollment. Blending data science with these core applications will help bring these female populations into the broad domain of computing, thus enriching its diversity.

3. *Relationship of the new program to existing programs.*

The design of the proposed Data Science major was greatly inspired by two recent reports about creating undergraduate Data Science programs from (i) the National Academies of Sciences, Engineering and Medicine and (ii) the Park City Mathematics Institute, a NSF report endorsed by the Board of Directors of the American Statistical Association. Special effort was made to match the suggested guidelines from these reports. We were also influenced by two existing Data Science programs at UCI and University of San Francisco.

The proposed program is different in its requirements from the traditional Computer Science as well as the Statistics undergraduate degrees. For example, the Data Science program does not require upper division courses like CS150, CS152, CS153, CS161, CS120A, CS120B which are all core requirements for the CSE undergraduate major.

The Data Science program is also different from the Statistics undergraduate major. For example, the Data Science program does not require upper division courses like STAT157, STAT160A, STAT160B, and STAT160C, which are all core requirements for the Statistics undergraduate Bachelor of Science major. In addition to the core requirements, the technical and application electives of the Data Science program are also different from the traditional Computer Science and Statistics program, with a strong emphasis on data science applications in specific domain disciplines.

4. *The proposed curriculum. Great care should be given in this area, correct rubrics should be listed for courses, all cross listings should be listed, unit total considerations should be taken into account and totals should be verified by program staff, faculty, and appropriate Executive Committee personnel. A copy of the proposed program change should be provided for inclusion in the Catalog.*

Please see attached program description in Appendix A; the proposed catalog entry is in Appendix B.

5. *A list of faculty who will be involved in the program, including those teaching, advising, and administering.*

Below is the current list of faculty involved in the program (new faculty will be added as the program evolves):

Professors:

Xinping Cui, Statistics
Evangelos Hristidis, CSE
Daniel Jeske, Statistics
Eamonn Keogh, CSE
Stefano Lonardi, CSE
Christian Shelton, CSE
Vassilis Tsotras, CSE

Associate Professors:

James Flegal, Statistics
Jun Li, Statistics
Shujie Ma, Statistics
Weixin Yao, Statistics
Shuheng Zhou, Statistics

Assistant Professors:

Ahmed Eldawy, CSE
Esra Kurum, Statistics
Wenxiu Ma, Statistics
Amr Magdy, CSE
Vagelis Papalexakis, CSE

LPSOEs:

Mariam Salloum* (CSE)
Paea LePendur (CSE)
TBD* (Statistics)

Administration:

Daniel Jeske, Statistics
Vassilis Tsotras, CSE

* The CSE and Statistics Departments were recently awarded a \$400K gift from the Technology Pathways Initiative for creating the new Data Science undergraduate program. A significant part of this gift will support the first year of two LPSOE/LSOE positions. CSE has already hired Dr. Salloum in one of these positions.

6. *For interdisciplinary programs, the degree of participation and the role of each department must be explicitly described. The chairs of all participating departments must provide written approval for the creation of the program and indicate their commitment to provide necessary resources including faculty release.*

The program will be administered through a joint steering committee. The steering committee will consist of three faculty from the Department of Computer Science and Engineering and three faculty from the Department of Statistics. The program director and the program co-director will be from different departments. The Director and co-Director are the coordinators of the program within their respective departments and colleges, and the Director has the added responsibility of coordinating and/or resolving campus-level issues. Normal terms for the Director and co-Director are 3 years and at the end of the 3-year term the co-Director is expected to accede to the Director position. If the Director (or co-Director) is unable to complete their 3-year term, a faculty from the same department will be chosen to assume the duties until the end of that 3-year term.

Directors and co-Directors will be filled through a nomination process that starts with the joint steering committee nominating names for the two positions to the Deans of the two colleges. The Deans will review the names and when they approve they will forward the names as nominees to the Provost/EVC for final approval. The same process will be used if and when the need arises to name an acting co-Director.

Circumstances may intervene that call for consecutive terms of a Director or a co-Director, or consecutive Directors or co-Directors from within the same department. These situations will be recognized, agreed upon, and handled by the joint steering committee, the two Deans, and the EVC/Provost on a case-by-case basis.

Director and co-Director stipend costs will be set by mutual agreement of the Deans of the two colleges, and they will split these costs 50:50.

Proposed changes to the program will need to be approved by the majority of the steering committee (including Director and co-Director). In the case of a tied vote, the Director makes the final decision. The proposed program change will then be reviewed by each college executive committee and then the committee on education policy. If these committees consider the change to be noncontroversial, the proposed change is placed on the Consent Calendar for a meeting of the Division of the Academic Senate.

Each department will be responsible for offering at least once per year any of the program's core courses taught by that department. The two departments will also cooperate in providing materials needed for any appropriate accreditation process (e.g., ABET or WASC.)

Each year the Director and co-Director will issue an annual report to the two Deans that describes the state of the program. Based on the report, the two Deans can agree to initiate procedures for having one college assume full control of the program, creating two parallel programs, or mutually agreeing to retire the program upon the graduation of the last enrolled student.

As with other intercollege majors (e.g., Neuroscience Undergraduate Major), when students declare the new major, they choose from which college they wish to have their B.Sc. degree awarded. Students whose degrees are awarded by the Bourns College of Engineering are advised in and have their records maintained by the Department of Computer Science and Engineering; students whose degrees are awarded by the College of Natural and Agricultural Sciences are advised in and have their records maintained by the CNAS Academic Advising Center. Students must fulfill the breadth requirements of the college they choose.

7. *Projected enrollment in the program.*

The projected enrollment at the start of the program is 25-30 students; we expect a class of 50 students at steady state.

8. *Name of degree, if applicable, and the anticipated number of degrees to be granted when the program reaches steady state.*

BS in Data Science, 40-50 degrees awarded per year

9. *Potential impact of the new program on existing programs. If the proposed program includes required courses from a department other than the administering department, the proposal must include a statement from the department indicating that it has been consulted and that it will provide access to the required courses.*

The new major uses six existing lower division courses offered by the Department of Mathematics (namely: MATH009A, MATH009B, MATH009C, MATH010A, MATH011/CS011, MATH031). All these courses are currently being used by CSE and/or Statistics for satisfying college/major requirements in the BS degrees offered currently by the CSE and Statistics departments. Given that these courses are also been used by many other departments to satisfy college requirements, we do not expect a major impact in their offerings by the new program.

The design of the new major led to the creation of four new upper division courses (namely: CS105, CS167, STAT156A, STAT156B). These courses have being approved by their respective departments and serve as electives in the BS programs offered by the CSE and Statistics departments. All other upper division courses required for the new major are currently being taught in the CSE and Statistics Departments.

Moreover, the new major requires that students take four electives (organized in two-course sequences) in departments outside CSE and Statistics. These sequences are created from courses that are currently taught in their home departments (e.g., ECE, Business, Economics; more such sequences will be added in the future). As the major's students will have various options for their electives, the expected impact on those courses would be minimal. The chairs of the involved departments have been consulted and have agreed to provide access to the required course sequences (see attached letters).

See attached letters of support from Math (Chair) and chairs of the application sequences Depts in Appendix C.

10. *A full listing of resources required for start-up and for operations. In cases where no additional resources will be needed, this must be explicitly stated. This listing may include: personnel (faculty FTE or temporary positions, Teaching Assistants or Readers, administrative staff, technical support); support services including computer facilities and library resources; space requirements. A plan indicating how the resources will be obtained would also be helpful to the committee in reviewing the proposal. A letter of support from the College Dean and/or Executive Vice Chancellor-Provost indicating endorsement as well as a promise of support for the proposal also would be extremely helpful.*

- Faculty FTE: the program will use existing faculty from the two departments as well as two new LSOE positions (one per department). The LSOE for CSE has already been hired (Mariam Salloum) while the LSOE position for Statistics has been approved. These two new LSOEs are partially supported by a recent gift (400K) from the Technology Pathways Initiative for creating the new Data Science undergraduate program.
- Teaching Assistants: at steady state we will have 50 students per year; this corresponds to about 1.5 additional lab/discussion sections (assuming a section is about 30-35 students). Students will take 4 courses per quarter, resulting in about 6 lab/discussion sections. Assuming a TA handles 2-3 sections, we estimate that the new program will need 2-3 TA positions per quarter. Such cost however will be easily covered by the tuition fees of the new students.
- Administrative Staff: the program will be administered by existing staff in the two home departments. BCOE students enrolled in the Data Science program through the CSE Department will be advised by BCOE's Office of Undergraduate Student Academic Affairs (OSAA). OSAA currently has eight full time advisors that will initially accommodate the Data Science students. The BCOE college is actively working on hiring three additional advisors that we be available to reduce the load as program grows. CNAS students enrolled in the Data Science program through the Department of Statistics will be advised by CNAS's Undergraduate Academic Advising Center (UAAC). The UAAC MPS (Math and Physical Sciences) advising unit currently has 1 full time advisor that will initially accommodate Data Science students. CNAS is actively working on hiring one additional MPS advisor that will be available to reduce the load as program grows.
- Computer facilities and library resources: no new facilities required
- Space requirements: no new space requirements.

11. *Both internal and external letters of support should be provided with the proposal. Internal letters of support are often from UCR department chairs and faculty of related programs. The external letters should be from other UC campuses or other peer institutions. Letters from off-campus help to establish the quality of the program and its fit within the context of related programs at other universities. Upon consultation with the CEP the demand for external letters may be waived.*

See Appendix D for letters of support from:

Dr. Walid Najjar, Chair, Department of Computer Science and Engineering

Dr. Xinpeng Cui, Department of Statistics.

Dr. Chris Lynch, Dean (BCOE)

Dr. Kathryn Uhrich, Dean (CNAS)

12. Approvals from program faculty, College faculty (if the new proposal affects a college regulation), and the appropriate Executive Committee should be obtained before forwarding the new program to the attention of the Senate Analyst for CEP.

Approved by Department of Computer Science and Engineering on 5/20/2019

Approved by Department of Statistics on 6/1/2018

Approved by BCOE Executive Committee on 8/26/2019

Approved by CNAS Executive Committee on 6/12/2019

Appendix A:

Draft of the new Data Science Undergraduate Program

Draft of the new Data Science Undergraduate Program

Program Design Committee members:

Daniel Jeske (Stat), Wenxiu Ma (Stat), Vagelis Papalexakis (CSE), Christian Shelton (CSE), Vassilis Tsotras (CSE), Shuheng Zhou (Stat)

Example Four Year Schedule

Year 1	Fall	Winter	Spring
	MATH 9A	MATH 9B	MATH 9C
	CS 10	CS 12	CS 14
	ENGL 1A	ENGL 1B	
	H/SS 1	H/SS 2	Phy_Sci
Year 2	Fall	Winter	Spring
	MATH 31	MATH 10A	CS 111
	CS 100	MATH 11 / CS 11	CS 105
	STAT 100A or equivalent	STAT 100B or equivalent	
	Bio_Sci	Sci 1	Sci 2
Year 3	Fall	Winter	Spring
	STAT 156A	STAT 156B	STAT 167 / CS 171
	CS 166 / CS 167	CS 141	
	STAT 147	ENGR 170/PBPL 170	[TECH ELECTIVE]
	H/SS 3	H/SS 4	H/SS 5
Year 4	Fall	Winter	Spring
	STAT 170A	STAT 170B	STAT 183 / CS 179
	ENGL 1C/ENGR 180		
	[APP 1]	[APP 2]	
	[TECH ELECTIVE]	[TECH ELECTIVE]	[TECH ELECTIVE]

Comments:

Four new courses, CS105, CS167, STAT156A, STAT156B

Course Descriptions:

CS 010 Introduction to Computer Science for Science, Mathematics, and Engineering I (4)

Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): a college mathematics course (may be taken concurrently) or credit for MATH 009A from the Advanced Placement Examination or the Mathematics Advisory Examination. Covers problem solving through structured programming of algorithms on computers using the C++ object-oriented language. Includes variables, expressions, input/output (I/O), branches, loops, functions, parameters, arrays, strings, file I/O, and classes. Also covers software design, testing, and debugging. Credit is not awarded for CS 010 if it has already been awarded for CS 010V or CS 030.

CS 012 Introduction to Computer Science for Science, Mathematics, and Engineering II (4)

Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 010 or CS 010V with a grade of “C” or better; familiarity with C or C++ language. Covers structured and object-oriented programming in C++. Emphasizes good programming principles and development of substantial programs. Topics include recursion, pointers, linked lists, abstract data types, and libraries. Also covers software engineering principles. Credit is awarded for only one of CS 012 or CS 012V or CS 013.

CS 014 Introduction to Data Structures and Algorithms (4)

Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 012 or CS 012V with a grade of “C” or better or CS 013 with a grade of “C” or better; proficiency in C++. Topics include basic data structures such as arrays, lists, stacks, and queues. Covers dictionaries (including binary search trees and hashing) and priority queues (heaps). Offers an introductory analysis of algorithms, sorting algorithms, and object-oriented programming including abstract data types, inheritance, and polymorphism. Explores solving complex problems through structured software development.

CS 100 Software Construction (4)

Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 014 with a grade of “C-” or better. Emphasizes development of software systems. Topics include design and implementation strategies; selection and mastery of programming languages, environment tools, and development processes. Develops skill in programming, testing, debugging, performance evaluation, component integration, maintenance, and documentation. Covers professional and ethical responsibilities and the need to stay current with technology.

CS 105 Data Analysis Tools (4) -- new course --

Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 14. Introduction to data analysis tools including data statistics, simple data storage types, data acquisition from the web and public APIs, data cleaning, crowdsourcing for

data collection and cleaning, supervised and unsupervised learning techniques, and data visualization. The laboratory will also include hands-on exercises on the aforementioned topics in Python and MATLAB.

CS 111 Discrete Structures (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 010 or CS 010V; CS 011/MATH 011; MATH 009C (or MATH 09HC). A study of discrete mathematical structures emphasizing applications to computer science. Topics include asymptotic notation, generating functions, recurrence equations, elements of graph theory, trees, algebraic structures, and number theory.

CS 141 Intermediate Data Structures and Algorithms (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 014 with a grade of “C-” or better; CS 111; MATH 009C or MATH 09HC; proficiency in C++. Explores basic algorithm analysis using asymptotic notations, summation and recurrence relations, and algorithms and data structures for discrete structures including trees, strings, and graphs. Also covers general algorithm design techniques including “divide-and-conquer,” the greedy method, and dynamic programming. Integrates knowledge of data structures, algorithms, and programming.

CS 166 Database Management Systems (4) Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 100, CS 111. Covers architecture of database management systems; relational, network, and hierarchical models; distributed database concepts; query languages; implementation issues; and privacy and security of the database.

CS 167 Introduction to Big Data Management (4) -- new course-- Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 100, CS111. Introduces the design of big-data systems and their application in data management and processing. Describes the common functionality in big-data processing such as distributed storage, resource management, query processing, fault-tolerance, and APIs. Covers the popular big-data technologies such as MapReduce, key-value stores, and semi-structured data management.

CS 171 Introduction to Machine Learning and Data Mining (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 100, CS 111. Introduces formalisms and methods in data mining and machine learning. Topics include data representation, supervised learning, and classification. Covers regression and clustering. Also covers rule learning, function approximation, and margin-based methods.

CS 179 (E-Z) Project in Computer Science (4) For hours and prerequisites, see segment descriptions. Under the direction of a faculty member, student teams propose, design, build, test, and document software and/or hardware devices or systems. Emphasizes professional and ethical responsibilities and the need to stay current on technology and its global impact on economics, society, and the environment.

MATH 009A First-Year Calculus (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): MATH 005 with a grade of “C-” or better or MATH 006B with a grade of “C-” or better or equivalent. Introduction to the differential calculus of functions of one variable. Credit is awarded for only one of MATH 008B, MATH 009A, or MATH 09HA.

MATH 009B First-Year Calculus (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): MATH 008B with a grade of “C-” or better or MATH 009A with a grade of “C-” or better or MATH 09HA with a grade of “C-” or better. Introduction to the integral calculus of functions of one variable. Credit is awarded for only one of MATH 009B or MATH 09HB.

MATH 009C First-Year Calculus (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): MATH 009B with a grade of “C-” or better or MATH 09HB with a grade of “C-” or better. Further topics from integral calculus, improper integrals, infinite series, Taylor’s series, and Taylor’s theorem. Credit is awarded for only one of MATH 009C or MATH 09HC.

MATH 010A Calculus of Several Variables (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): MATH 009B with a grade of “C-” or better or MATH 09HB with a “C-” or better or equivalent. Topics include Euclidean geometry, matrices and linear functions, determinants, partial derivatives, directional derivatives, Jacobians, gradients, chain rule, and Taylor’s theorem for several variables.

MATH 011 Introduction to Discrete Structures (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): MATH 009A (or MATH 09HA); CS 010 or CS 010V or MATH 009B (or MATH 09HB). Introduction to basic concepts of discrete mathematics emphasizing applications to computer science. Topics include propositional and predicate calculi, elementary set theory, functions, relations, proof techniques, elements of number theory, enumeration, and discrete probability. **Cross-listed with CS 011.**

MATH 031 Applied Linear Algebra (5) Lecture, 3 hours; discussion, 2 hours. Prerequisite(s): MATH 009A (or MATH 09HA) with a grade “C-” or better and CS 010 or CS 010V or MATH 009B (or MATH 09HB) with a grade of “C-” or better. A study of matrices and systems of linear equations, determinants, Gaussian elimination, vector spaces, linear independence and linear transformation, orthogonality, eigenvalues, and eigenvectors. Also examines selected topics and applications.

STAT 100A Introduction to Statistics (5) Lecture, 3 hours; discussion, 1 hour; laboratory, 3 hours. Prerequisite(s): MATH 005 or MATH 006B or MATH 009A or MATH 09HA or equivalent. A general introduction to descriptive and inferential statistics. Topics include histograms; descriptive statistics; probability; normal, binomial, and Poisson distributions; sampling distributions; hypothesis testing; and confidence intervals. Credit is awarded for only one of STAT 048 or STAT 100A.

STAT 100B Introduction to Statistics (5) Lecture, 3 hours; discussion, 1 hour; laboratory, 3 hours. Prerequisite(s): STAT 100A “An introduction to statistics” with a grade of C- or better. Topics include linear regression, correlation, analysis of variance, and simple experimental designs.

STAT 147 Introduction to Statistical Computing (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 100A or equivalent. Introduction to computer-assisted data analysis and statistical inference using both the R and SAS packages. Topics include input, output, and editing of data; graphical procedures; descriptive statistics; cross-tabulation; inferential statistical techniques including estimation and testing; and analysis of variance.

STAT 156A Mathematical Statistics with Applications for Data Science I (4) -- new course-- Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): MATH 009C or consent of instructor. Introduction to frequentist probability concepts, random variables, and their distributions. Discussion of key theorems and inequalities in probability theory. Introduction to frequentist methods of point and interval estimation.

STAT 156B Mathematical Statistics with Applications for Data Science II (4) -- new course-- Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 156A or consent of instructor. Introduction to Bayesian probability concepts, illustrative application of Frequentist theory to linear regression, logistic regression and ANOVA, analysis of contingency tables, applications of sequential statistics, methods for observational studies and for missing data analyses.

STAT 167 Introduction to Data Science (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 100B or STAT 155, with a grade of C- or better, or equivalents; STAT 147, with a grade of C- or better. Introduction to data science using the R programming language. Topics include big data management, visualization and analytical skills, unsupervised and supervised statistical learning methods, and real-world data science application examples.

STAT 170A Regression Analysis (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 157, STAT 160C, or equivalents. Topics include simple and multiple linear regression, scatter-plots, and point and interval estimation. Addresses prediction, testing, calibration, interpretation, and practical applications of multiple regression. Explores simple, partial, and multiple correlation; variable selection methods; diagnostic procedures; and regression for longitudinal data.

STAT 170B Design of Experiments (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 170A with a grade of “C-” or better. Topics include principles of design; completely randomized designs; and one-way analysis of variance. Covers complete block designs and two-way analysis of variance; multiple comparisons; and complete factorial experiments. Explores fixed, random, and mixed models; split-plot designs; nested designs; analysis of covariance; sample size determination; and power analysis.

STAT 183 Statistical Consulting (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 170B; STAT 171, may be taken concurrently; Restricted to class level standing of senior. Introduces the statistical consulting process. Promotes consulting skills including developing effective communication skills, applying statistical methodology to client projects, and learning how to manage time and resources in a consulting environment. Satisfactory (S) or No Credit (NC) grading is not available.

ENGR 170 Technology, Policy, and Ethics (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): upper-division standing. Provides contemporary perspectives on interplays between technology, public policy, and ethics. Covers social, legal, and ethical issues such as liability, as well as environmental, patent, and copyright law. Cross-listed with PBPL 170.

TECHNICAL ELECTIVES

MATH 120 Optimization (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): MATH 010A with a grade of "C-" or better; MATH 031 with a grade of "C-" or better. Introduction to classical optimization including unconstrained and constrained problems in several variables, Addresses Jacobian and Lagrangian methods and the Kuhn-Tucker conditions. Covers the basic concepts of linear programming including the simplex method and duality with applications to other subjects.

MATH 135A Numerical Analysis (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 010 or CS 010V or equivalent with a grade of "C-" or better; MATH 031 with a grade of "C-" or better (may be taken concurrently). A study of numerical methods for determining solutions to nonlinear equations and simultaneous linear equations. Topics also include interpolation, techniques of error analysis, and computer applications.

CS 170 Introduction to Artificial Intelligence (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 100 with a grade of "C-" or better, CS 111. An introduction to the field of artificial intelligence. Focuses on discrete-valued problems. Covers heuristic search, problem representation, and classical planning. Also covers constraint satisfaction and logical inference.

CS 172 Introduction to Information Retrieval (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 100; CS 111; EE 114 or STAT 155. Introduces information retrieval (IR) principles and techniques for indexing and searching document collections. Topics include Web search, text processing, ranking algorithms, search in social networks, and search evaluation. Also studies scalability issues in search engines. Satisfactory (S) or No Credit (NC) grading is not available.

CS 180 Introduction to Software Engineering (4) Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 100. A study of software engineering techniques for the development,

maintenance, and evolution of large software systems. Topics include requirements and specification; system design and implementation; debugging, testing, and quality assurance; reengineering; project management; software process; tools; and environments.

CS 181 Principles of Programming Languages (4) Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 061, CS 100, CS 111, CS 150. Covers the principles of programming language design. Includes the study and comparison of several programming languages, their features, and their implementations.

STAT 104 Decision Analysis and Management Science (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 008 or equivalent; STAT 048 or STAT 100A or equivalent; upper-division standing. A survey of deterministic and probabilistic models for decision making. Topics include linear programming and extensions, networks, dynamic programming, decision trees, queuing models, and simulation. Explores the application of these models in decision making. Emphasizes use of the computer. Cross-listed with BUS 104.

STAT 127 Introduction to Quality Improvements (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 048 or STAT 100A or consent of instructor. Explores Deming's 14 points for management, graphical methods, fishbone diagram, Pareto analysis, control charts for attributes and variables, cusum and moving average charts, process-capability, economic design, acceptance sampling, Taguchi method, parameter design, tolerance design, reliability, hazard rate, censoring, and accelerated life testing. Cross-listed with BUS 127.

STAT 130 Sampling Surveys (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 100B, or equivalent. Covers simple random sampling, addresses stratified sampling, cluster sampling, and ratio and regression estimates. Explores random response, capture-recapture, and jack-knife techniques.

STAT 140 Nonparametric Techniques (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 100B or equivalent. Covers randomization tests, rank tests, methods of association, and distribution-free tests.

STAT 146 Statistical Forecasting Techniques (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 100B or equivalent. Topics include exponential smoothing, simple and multiple regression analysis, time series, trend analysis, and seasonal analysis

STAT 157 Statistical Computer Packages (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 100B with a grade of C- or better, or equivalents; STAT 147 with a grade of C- or better; or consent of instructor. A study of major statistical packages including SAS with the emphasizing advanced SAS programming. Topics include advanced graphical procedures, linear models (regression and analysis of variance), multivariate techniques, and SAS macros.

STAT 171 General Statistical Models (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 170B. Generalized linear models and least squares. Analysis of covariance, nonlinear regression, nonlinear least squares. Regression methods for discrete data: loglinear models, logistic regression, discriminant analysis. Regression methods for life data. Cox survival model, Kaplan-Meier estimation, Mantel-Haenszel test.

APPLICATION COURSE SEQUENCES

Economics: {ECON 108, ECON 136}

Business: {BUS 104, BUS 123}, {BUS 124, BUS 125}, {BUS 103, BUS 115}.

Earth Sciences: {GEO 111, GEO 161}, {GEO 115, GEO 147}

Electrical and Computer Engineering: {EE142, EE146}

Biology/Bioinformatics: {BIO 5A, BIO 20}

BIOL 005A Introduction to Cell and Molecular Biology (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): BIOL 05LA (may be taken concurrently) or BIOL 020 (may be taken concurrently) with grades of “C-” or better; CHEM 001A and CHEM 01LA with grades of “C-” or better or CHEM 01HA and CHEM 1HLA with grades of “C-” or better; consent of instructor is required for students repeating the course. An intensive course designed to prepare for upper-division courses in cell and molecular biology. Covers biochemical, structural, metabolic, and genetic aspects of cells. Required for Biology majors; recommended for science majors desiring an introduction to biology.

BIOL 020 Dynamic Genome (2) F Laboratory, 6 hours. Prerequisite(s): CHEM 001A or CHEM 01HA, MATH 008B or MATH 009A (MATH 009A may be taken concurrently); freshman standing. Introduces computational and experimental approaches in investigating the genomes of plants and animals. Explores scientific discovery using the tools of bioinformatics and genomics. Includes participation in research projects being conducted on campus. Credit is not awarded for BIOL 020 if it has already been awarded for BIOL 05LA.

BUS 104 Decision Analysis and Management Science (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 008 or equivalent; STAT 048 or STAT 100A or equivalent; upper-division standing. A survey of deterministic and probabilistic models for decision making. Topics include linear programming and extensions, networks, dynamic programming, decision trees, queuing models, and simulation. Explores the application of these models in decision making. Emphasizes use of the computer. Cross-listed with STAT 104.

BUS 123 Spreadsheet Modeling for Decision-Making (4) Lecture, 3 hours; written work, 3 hours. Prerequisite(s): BUS 104/STAT 104 or consent of instructor. Introduces the fundamental techniques of using data to make informed management decisions in the presence of uncertainty of advanced Microsoft Excel functionality. Uses spreadsheet modeling for decision analysis and optimization applications. Enhances and reinforces analytical skills and the ability to intelligently use information for making decisions under uncertainty.

BUS 124 Business Analytics (4) Lecture, 3 hours; term paper, 1 hour; written work, 2 hours. Prerequisite(s): STAT 048 or consent of instructor. Provides fundamental concepts and tools needed to understand the emerging role of business analytics in organizations. Applies basic business analytics tools in a spreadsheet environment. Introduces market-leading techniques that help identify and manage key data from business processes. Provides the essential tools required for data mining and business process re-engineering.

BUS 125 Simulation for Business (4) Lecture, 3 hours; extra reading, 1.5 hours; outside projects, 1.5 hours. Prerequisite(s): BUS 104/STAT 104, STAT 048, or equivalents. Introduces simulation as a tool for analyzing complex systems. Analyzes and discusses the theory and practice of modeling through simulation. Topics include modeling uncertainty and collecting input data, Monte Carlo simulation techniques, model verification and validation, and sensitivity analysis. Examines applications in finance, marketing, operations, and supply chain management.

BUS 103 Marketing and Distribution Management (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): upper-division standing or consent of instructor. An introduction to the role of marketing in society with emphasis on concepts, marketing methods, and institutions.

BUS 115 Marketing Research (4) Lecture, 3 hours; outside research, 3 hours. Prerequisite(s): BUS 103. Covers types and sources of marketing information, the marketing research process, and techniques of data collection and analysis, including consumer and customer surveys and test marketing. Examines both quantitative and qualitative research with analysis of the values and limitations of data. Emphasis is placed on evaluation and interpretation of results.

ECON 108 Introductory Econometrics (5) Lecture, 3 hours; discussion, 1 hour; laboratory, 1 hour; written work, 1 hour; outside research, 1 hour. Prerequisite(s): ECON 107 or consent of instructor. A continuation of ECON 107. Covers, at an introductory level, the basic concepts related to logit and probit models, simultaneous equations models, dynamic time series models, unit roots and auto-regressive conditional heteroskedasticity (ARCH), and forecasting.

ECON 136 Empirical Financial Economics (4) Lecture, 3 hours; individual study, 3 hours. Prerequisite(s): ECON 107 or consent of instructor. Discusses various empirical aspects of financial economics and financial risk management. Addresses both theoretical and applied issues in finance, risk management, and econometrics. Also discusses quantitative analysis, simulation methods, and case studies.

EE 142 Pattern Recognition and Analysis of Sensor Data (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): EE 114 or STAT 155 or consent of the instructor. Introduction to pattern recognition for multi-dimensional, multi-modal sensor data such as images, videos, and smart grids. Classification and decision functions, feature extraction, regression, and neural networks. Clustering and dimensionality reduction for unsupervised learning. Dynamic models and tracking. Applications of pattern recognition in computer vision, robotics, smart grids, etc.

EE 146 Computer Vision (4) Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): senior standing in Computer Science or Electrical Engineering, or consent of instructor. Imaging formation, early vision processing, boundary detection, region growing, two-dimensional and three-dimensional object representation and recognition techniques. Experiments for each topic are carried out.

GEO 111 Numerical Skills in Geoscience (4) Lecture, 3 hours; laboratory, 3 hours; term paper, 1 hour;. Prerequisite(s): MATH 009C. An introduction to computer programming and numerical modelling. The class will teach the basic principles of how computer programs are written and numerical models constructed, and provide data processing and visualization skills. The class will hence foster an ability to apply numerical techniques to problems in Earth and Environmental Sciences.

GEO 161 Cenozoic Climate Change (4) Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): one of the following courses with a grade of "C-" or better; GEO 001 or GEO 002 or GEO 009 or GEO 011. Examines physical, chemical, and biological evidence of climatic and environmental change throughout the Cenozoic Era (last 65 million years) to provide a framework for understanding natural environmental change and for predicting future change. Introduces students to computer-based numerical methods of data analysis for interpreting past records of environmental change.

GEO 115 Geologic Maps and Landforms (5) Lecture, 2 hours; laboratory, 6 hours; field, 30 hours per quarter. Prerequisite(s): GEO 001 (may be taken concurrently); MATH 004 or MATH 005, or MATH 008A. Examines characteristic patterns of bedrock outcrops, surficial deposits, the related landforms, and their representation on maps. Covers unconformities, folds, faults, intrusions, alluvial fans, river terraces, and landforms indicative of glaciers, volcanoes, landslides, and earthquakes. Applies map information to resource and hazard evaluation.

GEO 147 Active Tectonics and Remote Sensing (4) Lecture, 2 hours; discussion, 1 hour; laboratory, 3 hours. Prerequisite(s): GEO 001, GEO 115; or consent of instructor. A computer-based course that introduces active tectonics and the earthquake cycle and how they are studied using remote sensing data. Explores examples of actively deforming areas from around the world using computer visualization software and freely available data sources (satellite imagery, digital topography, GPS and earthquake data).

Appendix B:

Proposed catalog entry for the Data Science Undergraduate Program

Data Science Undergraduate Major (Catalog entry)

Major

Data science studies the collection, management, and analysis of data to extract knowledge. It is a multidisciplinary program with core components from Computer Science and Statistics, and required application study in a variety of empirical disciplines. Courses span the discipline from theory to practice and prepare students for careers or graduate studies in data-intensive fields.

The B.S. in Data Science major is an intercollege major offered by the Bourns College of Engineering and the College of Natural and Agricultural Sciences. A B.S. degree in Data Science is offered by each college. When students declare the major, they choose from which college they wish to have their degree awarded. Students whose degrees are awarded by the Bourns College of Engineering are advised in and have their records maintained by the BCOE Office of Student Academic Affairs; students whose degrees are awarded by the College of Natural and Agricultural Sciences are advised in and have their records maintained by the CNAS Undergraduate Academic Advising Center. Breadth requirements vary by college; and students must fulfill the breadth requirements of the college they choose.

All undergraduates in the Bourns College of Engineering must see an advisor at least annually. Visit student.engr.ucr.edu for details.

University Requirements

See Undergraduate Students section.

College Requirements

College breadth requirements vary depending on which college is chosen to award the degree. For details on breadth requirements, see the Colleges and Programs section of this catalog. Students are encouraged to consult their advisor regarding requirements.

Major Requirements

1. Lower-division requirements (37 units):
 - a) CS 010; CS 012; CS 014
 - b) MATH 009A; MATH 009B; MATH 009C; MATH 010A; MATH 031
 - c) MATH 011/CS 011
2. Upper-division requirements (60 units):
 - a) CS 105; CS 141
 - b) STAT 147; STAT 156A; STAT 156B; STAT 170A; STAT 170B
 - c) ENGR 170 or PBPL 170
 - d) CS 166 or CS 167
 - e) STAT 167 or CS 171

- f) STAT 183 or CS 179 (E-Z)
 - g) Four courses (at least 16 units) from the following list, none of which can also be used to satisfy other major requirements:
CS 166; CS 167; CS 170; CS 172; CS 180; CS 181; MATH 120; MATH 135A;
STAT 104; STAT 127; STAT 130; STAT 140; STAT 146; STAT 157; STAT 171.
3. Major Breadth requirement (8 units)
- One two-course sequence, chosen from the course sequences listed below:
- i. BIO 5A and BIO 20
 - ii. BUS 103 and BUS 115
 - iii. BUS 104 and BUS 123
 - iv. BUS 124 and BUS 125
 - v. ECON 108 and ECON 136
 - vi. EE 142 and EE 146
 - vii. GEO 111 and GEO 161
 - viii. GEO 115 and GEO 147

Note An introductory Statistics class, such as STAT 100A and STAT 100B, is strongly recommended.

Appendix C:


Letters of support/collaboration from other Departments (Mathematics, Earth Sciences, Economics, EEOB, Electrical and Computer Engineering)

DATE: October 18, 2018

TO: Dr. Vassilis Tsotras, Department Chair
Department of Computer Science

FROM: Dr. Yat Sun Poon, Department Chair
Department of Mathematics

RE: B.S. in Data Science proposal



Dear Vassilis,

I am glad to hear that a BS degree in data science is being brought to UCR through a joint effort between the Statistics and Computer Science departments. I believe many students will find it appealing. I cannot see any potential problems with your plans.

----- Forwarded Message -----

Subject: Earth Sciences courses for the new Data Science undergraduate program

Date: Fri, 14 Dec 2018 15:52:33 -0800

From: David Oglesby <doglesby@ucr.edu>

To: Vassilis Tsotras <tsotras@cs.ucr.edu>

Dear Vassilis:

I am in support of the Data Science undergraduate program between the Earth Sciences and Statistics Departments. As we discussed, I am in favor of adding the following Earth Sciences two-course sequences as electives available to the undergraduate students in the Data Science program:

GEO 111 and GEO 161
GEO 115 and GEO 147

Even though GEO 115 requires MATH 004 or MATH 005, or MATH 008A, the Data Science students would already have appropriate MATH core courses (e.g. MATH 9A-9C) so they would be allowed to take GEO115.

Sincerely,
David

David D. Oglesby
Professor of Geophysics
Chair, Department of Earth Sciences
University of California, Riverside
Riverside, CA 92521-0423
PH: (951) 827-2036
FAX: (951) 827-4324
david.oglesby@ucr.edu

From: **Steven M Helfand** <steven.helfand@ucr.edu>
Date: Mon, Nov 26, 2018 at 5:32 PM
Subject: Data Science Majors
To: Daniel R Jeske <daniel.jeske@ucr.edu>
Cc: Gloria Gonzalez-Rivera <ggonzale@ucr.edu>, Anthony L Gonzalez
<anthony.gonzalez@ucr.edu>, Bradley Troy Scalf <bradley.scalf@ucr.edu>

Dean Dan,

I write to affirm that the Economics Department will accept STAT 170A as a suitable prerequisite for Econ 108 and Econ 136. This will permit the Data Science majors to take these classes without taking Econ 107 first. Academically, we see no difficulty with this substitution.

Understanding that the data science program is just beginning and that data science students will likely not be taking Econ 108 or 136 until their third year in the program, we would like to move forward with this agreement for a trial period of five years, starting in 2019-20. We wish to do this as a five year experiment because we are concerned about the size of the classes, and Econ 108 in particular. Currently we plan to offer Econ 108 once per year, with a maximum size of 35. This is the size of the lab, and all that we can handle with one TA who teaches the sections and lab. If it were to grow beyond 35 students, we would need to hire a second TA, and we don't currently have funding to do this. At its current size, we expect that there would be 5-10 seats each year for non-econ majors.

Sincerely,

Steven

Steven Helfand
Chair, Department of Economics
University of California
Riverside, CA 92521-0427
Office: (951) 827-1470
steven.helfand@ucr.edu
<http://economics.ucr.edu/people/faculty/helfand/>

November 27, 2018
Professor Helen Regan
Chair
Department of Evolution, Ecology, and
Organismal Biology

Prof. Vassilis Tsotras
Director, Data Science Center at UCR

Dear Tsotras:

The Department of Evolution, Ecology, and Organismal Biology supports initiation of the Data Science undergraduate program at UCR. We are in favor of adding two Biology courses to be taken as electives by the undergraduate students in the Data Science program. The following two Biology courses are most relevant to the Data Science undergraduate program:

BIOL 005A – Introduction to Cell and Molecular Biology
BIOL 020 – Dynamic Genome

Both courses are offered every quarter, often more than once per quarter, and students from CSE and Statistics should be able to take them provided the prerequisites are satisfied.

Sincerely,



Helen Regan,
Professor and Chair
Department of Evolution, Ecology, and Organismal Biology at UCR



Bourns College of Engineering
Department of Electrical and Computer Engineering
343 Winston Chung Hall
Riverside, CA 92521
(951) 827-2484

April 22, 2018

Prof. Vassilis Tsotras
Director, Data Science Center at UCR

Dear Vassilis:

I strongly support initiation of the Data Science undergraduate program at UCR. I am also very much in favor of adding two ECE-based courses to be taken as electives by the undergraduate students in the Data Science program. The following two ECE courses are most relevant to the Data Science undergraduate program:

EE142 - Pattern Recognition and Analysis of Sensor Data
EE146 - Computer Vision

Both courses also allow EE prerequisites to be waived by consent of instructor, and students from CSE and Statistics should be able to take them.

Sincerely,

A handwritten signature in blue ink, appearing to read "Ilya Dumer", with a long horizontal flourish extending to the right.

Ilya Dumer,
Professor and Chair
Department of Electrical and Computer Engineering at UCR

Appendix D:

Letters of support from the chairs of CSE and Statistics and the BCOE and CNAS Deans.

December 28, 2018

To Whom It May Concern:

This letter is in strong support for the proposed undergraduate program in Data Science at UCR, to be jointly offered by the Departments of Computer Science & Engineering and Statistics.

Data Science has grown out of the need to integrate computational and statistical approaches to processing and interpreting data. Tools originating from data science are now becoming indispensable in today's science, technology, and business, fueling the demand for data scientists. Recognizing this need, our department has taken the initiative to develop research and educational programs in Data Science at UCR. In collaboration with other departments on campus, an online MS program in Data Science is already being offered. Recently the Data Science Center has been established that includes multiple newly hired faculty members, and has been given designated space in the new MRB building. Creating an undergraduate program in Data Science is the next step in this endeavor. This program will address critical and documented shortage of college graduates trained in Data Science, in industry, government, and academia.

The CSE Department enthusiastically supports the creation of the Data Science program and is fully committed to providing necessary resources within its capabilities for the instruction and advising of its students. As Data Science is at the intersection of Statistics and Computer Science, we look forward to collaborate with the Statistics department to establish the new undergraduate Data Science program at UCR.



Walid A. Najjar

Professor and Chair
Department of Computer Science and Engineering
Bourns College of Engineering
University of California Riverside

November 14, 2018

Dear Committee Members:

On behalf of the UCR Department of Statistics, I am writing in strong support for the proposed inter-departmental undergraduate program in Data Science at UCR.

Data Science is experiencing a rapid and unplanned growth. The website <http://datascience.community/colleges> currently lists 530 programs in Data Science, analytics and related fields at over 200 universities around the world. The vast majority of these are master's degrees and certificates programs offered both traditionally and online. There has been rapid growth of undergraduate programs at both research institutions and liberal arts colleges. It is expected this number to increase significantly in the near future.

Recognizing this need, our department has taken steps to develop research and educational programs in Data Science at UCR including the new undergraduate course "Introduction to Data Science" that has been taught for the past two years and the new graduate courses "Statistical Computing" and "Statistical Data Mining Methods" that have been taught for both Statistics graduate program and MS online program in Data Science at UCR.

Data Science program at the undergraduate level provides a synergistic approach to real world problem solving, one that leverages the content in Statistics, Computer Science and Mathematics but using case-based focus and hands-on approach. We have successfully boosted undergraduate enrollment into our Statistics program over past few years. The number of students had increased from a dozen to 136. We believe we will be able to attract more students with curriculum uniquely tailored for Data Science program, i.e., "students learn Data Science by doing Data Science". We believe our Data Science program will serve students well whether they join the marketplace or continue on to more advanced study.

Recognizing that Data Science is at the intersection of Statistics and Computer Science, the statistics department embraces the opportunity to partner with the computer science and engineering department to establish an undergraduate Data Science program at UCR.

Sincerely,



Xinping Cui, Ph.D.
Professor and Chair
Department of Statistics
University of California, Riverside
Tel: (951) 827-2563 | Fax: (951) 827-3286
E-mail: Xinping.Cui@ucr.edu

27 November, 2018

To whom it may concern:

I am writing this letter in enthusiastic support for the enclosed proposal to establish an undergraduate degree program in Data Science. This program will be jointly administered between BCOE and CNAS, with degrees granted by each college. I have had detailed conversations with Professor Tsotras and the program committee and fully support the academic program and administrative structure. I commit to working with them and leadership from CNAS to insure the program's success.

This program will help address the critical and documented shortage of college graduates educated in Data Science and the critical interpretation and analysis of large datasets. We expect students attracted to this program to come from a variety of backgrounds and other interests, increasing the diversity among Engineering students, and those in computational fields in particular. This program is partially supported by a grant from the Center for Advancing Women in Technology (CAWIT), which funds two L(P)SOE positions (one in BCOE and one in CNAS). I have recently joined the CAWIT advisory board and believe the data science program will bring more women into computing careers.

The Bourns College of Engineering looks forward to partnering with the College of Agricultural and Natural Sciences to offer this degree to UCR students. It is an important part of keeping our curriculum current and educating our students in the era of big data.

Sincerely,



Prof. Christopher S. Lynch
Dean, Bourns College of Engineering
University of California, Riverside

December 10, 2018

To Whom It May Concern:

RE: Data Science Undergraduate Major

I am writing to convey my support for the proposal to establish an undergraduate degree program in Data Science. The Department of Statistics in the College of Natural and Agricultural Sciences ("CNAS") and the Department of Computer Science and Engineering in the Bourns College of Engineering ("BCOE") will jointly administer this program, which will allow students to obtain a BS degree through an integrated plan of study. The program will provide an efficient and cost-effective academic path for students interested in careers within a variety of areas relating to data science.

Data Science as a discipline is found at the intersection of computer science and statistics and permeates other disciplines as well. For this program, the partnership between CNAS and BCOE is well demonstrated by the list of faculty involved: a healthy mix of faculty from both colleges. Sharing the costs for the director and co-director stipends between the two colleges is another illustration of integration of the two colleges in the program. Costs for director and co-director stipends will be equally split (50% - 50%) between CNAS and BCOE. Exact amounts will be determined at the time the program is approved.

We look forward to partnering with the Bourns College of Engineering to offer this degree program to our students.

Sincerely,



Kathryn Uhrich, Dean

College of Natural and Agricultural Sciences

From: Dylan Rodriguez <dylanr@ucr.edu>
Sent: Wednesday, October 30, 2019 1:11 PM
To: Vassilis Tsotras <tsotras@cs.ucr.edu>
Cc: Cherysa P Cortez <cherysa.cortez@ucr.edu>; Daniel R Jeske <daniel.jeske@ucr.edu>
Subject: Re: Data Science Major Proposal

Dear Vassilis:

I'm writing to provide the attached consultative feedback from CEP and P&B on the proposal for the new major in Data Science. As you will see, P&B has some substantive feedback to which you may wish to respond with a revised proposal. At its regular meeting of October 28 (this past Monday), the Senate's Executive Council engaged in some significant discussion regarding the need for and importance of such a program. While Council agreed on the academic, pedagogical, and institutional merits of the proposed major, it also agreed with P&B's consultative concerns about the absence of a clearly delineated budget (and budgetary accountabilities) in the submitted proposal. Some Council members also expressed concern about the narrow scope of the proposal in relation to the social sciences and whether courses from other disciplines may be included and/or impacted within the present scope of the proposal.

I trust that this consultation will help you consider your next steps in the proposal's development. Please feel free to contact me with any questions.

peace
dylan

Dylan Rodríguez
President-Elect (2020-2021), American Studies Association
Chair of the Academic Senate (2016-2020), UC Riverside Division
Professor, Dept. of Media and Cultural Studies
University of California, Riverside
Riverside, CA 92521

"If one were forced for the sake of clarity to define [fascism] in a word simple enough for all to understand, that word would be 'reform.'"

-George Jackson

organizations i support:

<http://www.socallib.org/>

www.criticalresistance.org

<http://scholarsforsocialjustice.com/>

my books:

[Forced Passages: Imprisoned Radical Intellectuals and the U.S. Prison Regime](#) (Univ. of Minnesota Press, 2006)


[Suspended Apocalypse: White Supremacy, Genocide, and the Filipino Condition](#) (Univ. of Minnesota Press, 2009)

[Critical Ethnic Studies: A Reader](#) (co-editor) (Duke Univ. Press, 2016)



October 7, 2019

To: Dylan Rodríguez, Chair
Riverside Division

From: Stefano Vidussi, Chair 
Committee on Educational Policy

Re: Second Review of Proposed B.S. in Data Science

The Committee on Educational Policy (CEP) reviewed and voted to support the revised proposal for a new B.S. in Data Science at their October 4, 2019 meeting.



PLANNING & BUDGET

October 22, 2019

To: Dylan Rodriguez, Chair
Riverside Division

From: Harry Tom, Chair
Committee on Planning and Budget

A handwritten signature in black ink, appearing to read "Harry WK Tom".

Re: [Campus Review] Proposal: New Undergraduate Major: Data Science Undergraduate Major – Revised

The Committee on Planning & Budget (P&B) reviewed the revised proposal for a new undergraduate major in Data Science and requests more clarification of the budget, including which units will cover the costs, both initially and in the long run, and which units will receive income generated by any increase in or shift of majors. The proposal counts on the full allotment per student apportion to deans as income for this major but does not account for all the expenses associated with teaching the estimated 148 students expected in the steady state. The budget should include as expenses the proportion of the time of existing faculty and other personnel dedicated to dealing with those students (i.e., if half of the students in a given class are from this major, half of the part of the salary of that faculty member associated with that class needs to be covered by the major). The work of administration members, as well as the use of services and facilities (classrooms, library, etc.) should be accounted the same way. At present, the budget provided is not realistic. It is also unclear that the issues raised in P&B's previous memo dated April 12, 2019 were addressed.

Original Proposal (Round 1)

Proposal for the new undergraduate major in Data Science

1. *Name of the academic program and the department or unit that will administer the program.*

Name: Data Science Undergraduate Major

Administration: The Data Science major will be administered jointly by the Department of Computer Science and Engineering (CSE) and the Department of Statistics.

2. *A thorough justification, including the motivation for the creation of the program in terms of student interest and professional or academic importance.*

Data has become ubiquitous in everyday life, impacting every profession, from entry-level office workers to CEOs, from team coaches to general managers, from accountants to CFOs. Businesses now have data available to them at a scale that is historically unprecedented; harnessing this data for insight on what customers want provides them with a competitive advantage. Traditional companies (Ford, Walmart, General Electric, etc.) today pride themselves as being transformed to big-data businesses. The field of Data Science has emerged to address the proliferation of data and the need to manage and understand it.

Data Science lies at the intersection of Computer Science and Statistics, and its solutions already serve a variety of application domains in science, engineering and business. While there is a plethora of M.S. offerings in Data Science, relatively few undergraduate Data Science programs are currently offered. A recent search for Data Science undergraduate degrees (<http://datascience.community/colleges>) indicates that such offerings are typically provided as (i) minors to existing B.S. programs (Computer Science, Math or Statistics), (ii) certificate programs, or (iii) programs with very specific focuses (Business Data Analytics or Computational Analytics being the most common ones). Within the UC system, currently there are three existing undergraduate Data Science degree programs offered at UC Irvine (2015), UC Berkeley (2018) and UC San Diego (2018). The proposed UCR Data Science undergraduate program has thus the potential to fill an important gap. Creating a new Data Science undergraduate degree will be instrumental in educating future data scientists by building knowledge bottom-up, covering both essential knowledge from Computer Science (in managing data) and Statistics (in analyzing data), and integrating this knowledge with applications to other domains and to real-life problems.

Data permeates all aspects of science, engineering, and other academic disciplines. Yet, a comprehensive program studying how data can be collected, transformed, analyzed, and used to solve problems across academic disciplines does not currently exist at UCR at the undergraduate level. Statistics and Computer Science degree programs discuss some aspects of this, but our proposed program provides an interdisciplinary view of how to employ data and the role of data across other academic disciplines. The rise of data science as a term reflects a new academic discipline, studying data itself. This program will educate students in these analyses, thought processes, and ways of viewing the world through a data lens.

Through its interdisciplinary nature, a Data Science undergraduate program offers a great opportunity to serve as a pathway for professional careers in various areas. The proposed program is different from existing programs in that its students will complete course sequences in other departments (e.g., economics, business, sociology, earth sciences, biology,

bioinformatics, and astronomy) where they will learn how Data Science principles are applied in these domains.

Fueled by the explosion of data, Data Science jobs have proliferated and the demand for data scientists is extremely high; moreover, this demand is expected to be strong for years to come. A recent McKinsey report forecasted a need for hundreds of thousands of data scientists in the next decade. Another study by IBM found more than 2.3 million data science and analytics job listings in 2015, and both job openings and job demand are projected to grow significantly. Three-fifths of the data science and analytics jobs are in the finance and insurance, professional services, and information technology sectors, but the manufacturing, health care, and retail sectors also are hiring significant numbers of data scientists. According to Glassdoor, a recruiting site, Data Scientist is the best job in the US (for the last three years in a row) with around 110K median base salary. We thus expect that the new program will be in high demand among students and will serve well the UCR community.

Furthermore, many application areas related to Data Science, such as astronomy, biology, and economics, historically have gender-balanced enrollment. Blending data science with these core applications will help bring these female populations into the broad domain of computing, thus enriching its diversity.

3. *Relationship of the new program to existing programs.*

The design of the proposed Data Science major was greatly inspired by two recent reports about creating undergraduate Data Science programs from (i) the National Academies of Sciences, Engineering and Medicine and (ii) the Park City Mathematics Institute, a NSF report endorsed by the Board of Directors of the American Statistical Association. Special effort was made to match the suggested guidelines from these reports. We were also influenced by two existing Data Science programs at UCI and University of San Francisco.

The proposed program is different in its requirements from the traditional Computer Science as well as the Statistics undergraduate degrees. For example, the Data Science program does not require upper division courses like CS150, CS152, CS153, CS161, CS120A, CS120B which are all core requirements for the CSE undergraduate major.

The Data Science program is also different from the Statistics undergraduate major. For example, the Data Science program does not require upper division courses like STAT157, STAT160A, STAT160B, and STAT160C, which are all core requirements for the Statistics undergraduate Bachelor of Science major. In addition to the core requirements, the technical and application electives of the Data Science program are also different from the traditional Computer Science and Statistics program, with a strong emphasis on data science applications in specific domain disciplines.

4. *The proposed curriculum. Great care should be given in this area, correct rubrics should be listed for courses, all cross listings should be listed, unit total considerations should be taken into account and totals should be verified by program staff, faculty, and appropriate Executive Committee personnel. A copy of the proposed program change should be provided for inclusion in the Catalog.*

Please see attached program description in Appendix A; the proposed catalog entry is in Appendix B.

5. *A list of faculty who will be involved in the program, including those teaching, advising, and administering.*

Below is the current list of faculty involved in the program (new faculty will be added as the program evolves):

Professors:

Xinping Cui, Statistics
Evangelos Hristidis, CSE
Daniel Jeske, Statistics
Eamonn Keogh, CSE
Stefano Lonardi, CSE
Christian Shelton, CSE
Vassilis Tsotras, CSE

Associate Professors:

James Flegal, Statistics
Jun Li, Statistics
Shujie Ma, Statistics
Weixin Yao, Statistics
Shuheng Zhou, Statistics

Assistant Professors:

Ahmed Eldawy, CSE
Esra Kurum, Statistics
Wenxiu Ma, Statistics
Amr Magdy, CSE
Vagelis Papalexakis, CSE

LPSOEs:

Mariam Salloum* (CSE)
Paea LePendur (CSE)
TBD* (Statistics)

Administration:

Daniel Jeske, Statistics
Vassilis Tsotras, CSE

* The CSE and Statistics Departments were recently awarded a \$400K gift from the Technology Pathways Initiative for creating the new Data Science undergraduate program. A significant part of this gift will support the first year of two LPSOE/LSOE positions. CSE has already hired Dr. Salloum in one of these positions.

6. *For interdisciplinary programs, the degree of participation and the role of each department must be explicitly described. The chairs of all participating departments must provide written approval for the creation of the program and indicate their commitment to provide necessary resources including faculty release.*

The program will be administered through a joint steering committee. The steering committee will consist of three faculty from the Department of Computer Science and Engineering and three faculty from the Department of Statistics. The program director and the program co-director will be from different departments. The Director and co-Director are the coordinators of the program within their respective departments and colleges, and the Director has the added responsibility of coordinating and/or resolving campus-level issues. Normal terms for the Director and co-Director are 3 years and at the end of the 3-year term the co-Director is expected to accede to the Director position. If the Director (or co-Director) is unable to complete their 3-year term, a faculty from the same department will be chosen to assume the duties until the end of that 3-year term.

Directors and co-Directors will be filled through a nomination process that starts with the joint steering committee nominating names for the two positions to the Deans of the two colleges. The Deans will review the names and when they approve they will forward the names as nominees to the Provost/EVC for final approval. The same process will be used if and when the need arises to name an acting co-Director.

Circumstances may intervene that call for consecutive terms of a Director or a co-Director, or consecutive Directors or co-Directors from within the same department. These situations will be recognized, agreed upon, and handled by the joint steering committee, the two Deans, and the EVC/Provost on a case-by-case basis.

Director and co-Director stipend costs will be set by mutual agreement of the Deans of the two colleges, and they will split these costs 50:50.

Proposed changes to the program will need to be approved by the majority of the steering committee (including Director and co-Director). In the case of a tied vote, the Director makes the final decision. The proposed program change will then be reviewed by each college executive committee and then the committee on education policy. If these committees consider the change to be noncontroversial, the proposed change is placed on the Consent Calendar for a meeting of the Division of the Academic Senate.

Each department will be responsible for offering at least once per year any of the program's core courses taught by that department. The two departments will also cooperate in providing materials needed for any appropriate accreditation process (e.g., ABET or WASC.)

Each year the Director and co-Director will issue an annual report to the two Deans that describes the state of the program. Based on the report, the two Deans can agree to initiate procedures for having one college assume full control of the program, creating two parallel programs, or mutually agreeing to retire the program upon the graduation of the last enrolled student.

As with other intercollege majors (e.g., Neuroscience Undergraduate Major), when students declare the new major, they choose from which college they wish to have their B.Sc. degree awarded. Students whose degrees are awarded by the Bourns College of Engineering are advised in and have their records maintained by the Department of Computer Science and Engineering; students whose degrees are awarded by the College of Natural and Agricultural Sciences are advised in and have their records maintained by the CNAS Academic Advising Center. Students must fulfill the breadth requirements of the college they choose.

7. *Projected enrollment in the program.*

The projected enrollment at the start of the program is 25-30 students; we expect a class of 50 students at steady state.

8. *Name of degree, if applicable, and the anticipated number of degrees to be granted when the program reaches steady state.*

BS in Data Science, 40-50 degrees awarded per year

9. *Potential impact of the new program on existing programs. If the proposed program includes required courses from a department other than the administering department, the proposal must include a statement from the department indicating that it has been consulted and that it will provide access to the required courses.*

The new major uses six existing lower division courses offered by the Department of Mathematics (namely: MATH009A, MATH009B, MATH009C, MATH010A, MATH011/CS011, MATH031). All these courses are currently being used by CSE and/or Statistics for satisfying college/major requirements in the BS degrees offered currently by the CSE and Statistics departments. Given that these courses are also been used by many other departments to satisfy college requirements, we do not expect a major impact in their offerings by the new program.

The design of the new major led to the creation of four new upper division courses (namely: CS105, CS167, STAT156A, STAT156B). These courses have being approved by their respective departments and serve as electives in the BS programs offered by the CSE and Statistics departments. All other upper division courses required for the new major are currently being taught in the CSE and Statistics Departments.

Moreover, the new major requires that students take four electives (organized in two-course sequences) in departments outside CSE and Statistics. These sequences are created from courses that are currently taught in their home departments (e.g., ECE, Business, Economics; more such sequences will be added in the future). As the major's students will have various options for their electives, the expected impact on those courses would be minimal. The chairs of the involved departments have been consulted and have agreed to provide access to the required course sequences (see attached letters).

See attached letters of support from Math (Chair) and chairs of the application sequences Depts in Appendix C.

10. *A full listing of resources required for start-up and for operations. In cases where no additional resources will be needed, this must be explicitly stated. This listing may include: personnel (faculty FTE or temporary positions, Teaching Assistants or Readers, administrative staff, technical support); support services including computer facilities and library resources; space requirements. A plan indicating how the resources will be obtained would also be helpful to the committee in reviewing the proposal. A letter of support from the College Dean and/or Executive Vice Chancellor-Provost indicating endorsement as well as a promise of support for the proposal also would be extremely helpful.*

- Faculty FTE: the program will use existing faculty from the two departments as well as two new LSOE positions (one per department). The LSOE for CSE has already been hired (Mariam Salloum) while the LSOE position for Statistics has been approved. These two new LSOEs are partially supported by a recent gift (400K) from the Technology Pathways Initiative for creating the new Data Science undergraduate program.
- Teaching Assistants: at steady state we will have 50 students per year; this corresponds to about 1.5 additional lab/discussion sections (assuming a section is about 30-35 students). Students will take 4 courses per quarter, resulting in about 6 lab/discussion sections. Assuming a TA handles 2-3 sections, we estimate that the new program will need 2-3 TA positions per quarter. Such cost however will be easily covered by the tuition fees of the new students.
- Administrative Staff: the program will be administered by existing staff in the two home departments. BCOE students enrolled in the Data Science program through the CSE Department will be advised by BCOE's Office of Undergraduate Student Academic Affairs (OSAA). OSAA currently has eight full time advisors that will initially accommodate the Data Science students. The BCOE college is actively working on hiring three additional advisors that we be available to reduce the load as program grows. CNAS students enrolled in the Data Science program through the Department of Statistics will be advised by CNAS's Undergraduate Academic Advising Center (UAAC). The UAAC MPS (Math and Physical Sciences) advising unit currently has 1 full time advisor that will initially accommodate Data Science students. CNAS is actively working on hiring one additional MPS advisor that will be available to reduce the load as program grows.
- Computer facilities and library resources: no new facilities required
- Space requirements: no new space requirements.

11. *Both internal and external letters of support should be provided with the proposal. Internal letters of support are often from UCR department chairs and faculty of related programs. The external letters should be from other UC campuses or other peer institutions. Letters from off-campus help to establish the quality of the program and its fit within the context of related programs at other universities. Upon consultation with the CEP the demand for external letters may be waived.*

See Appendix D for letters of support from:

Dr. Walid Najjar, Chair, Department of Computer Science and Engineering

Dr. Xinpeng Cui, Department of Statistics.

Dr. Chris Lynch, Dean (BCOE)

Dr. Kathryn Uhrich, Dean (CNAS)

12. Approvals from program faculty, College faculty (if the new proposal affects a college regulation), and the appropriate Executive Committee should be obtained before forwarding the new program to the attention of the Senate Analyst for CEP.

Approved by Department of Computer Science and Engineering on 12/11/2018

Approved by Department of Statistics on 6/1/2018

Approved by BCOE Executive Committee on 2/28/2019

Approved by CNAS Executive Committee on 1/15/2019

Appendix A:

Draft of the new Data Science Undergraduate Program

Draft of the new Data Science Undergraduate Program

Program Design Committee members:

Daniel Jeske (Stat), Wenxiu Ma (Stat), Vagelis Papalexakis (CSE), Christian Shelton (CSE), Vassilis Tsotras (CSE), Shuheng Zhou (Stat)

Example Four Year Schedule

Year 1	Fall	Winter	Spring
	MATH 9A	MATH 9B	MATH 9C
	CS 10	CS 12	CS 14
	ENGL 1A	ENGL 1B	
	H/SS 1	H/SS 2	Phy_Sci
Year 2	Fall	Winter	Spring
	MATH 31	MATH 10A	CS 111
	CS 100	MATH 11 / CS 11	CS 105
	STAT 100A or equivalent	STAT 100B or equivalent	
	Bio_Sci	Sci 1	Sci 2
Year 3	Fall	Winter	Spring
	STAT 156A	STAT 156B	STAT 167 / CS 171
	CS 166 / CS 167	CS 141	
	STAT 147	ENGR 170/PBPL 170	[TECH ELECTIVE]
	H/SS 3	H/SS 4	H/SS 5
Year 4	Fall	Winter	Spring
	STAT 170A	STAT 170B	STAT 183 / CS 179
	ENGL 1C/ENGR 180		
	[APP 1]	[APP 2]	
	[TECH ELECTIVE]	[TECH ELECTIVE]	[TECH ELECTIVE]

Comments:

Four new courses, CS105, CS167, STAT156A, STAT156B

Course Descriptions:

CS 010 Introduction to Computer Science for Science, Mathematics, and Engineering I (4)

Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): a college mathematics course (may be taken concurrently) or credit for MATH 009A from the Advanced Placement Examination or the Mathematics Advisory Examination. Covers problem solving through structured programming of algorithms on computers using the C++ object-oriented language. Includes variables, expressions, input/output (I/O), branches, loops, functions, parameters, arrays, strings, file I/O, and classes. Also covers software design, testing, and debugging. Credit is not awarded for CS 010 if it has already been awarded for CS 010V or CS 030.

CS 012 Introduction to Computer Science for Science, Mathematics, and Engineering II (4)

Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 010 or CS 010V with a grade of “C” or better; familiarity with C or C++ language. Covers structured and object-oriented programming in C++. Emphasizes good programming principles and development of substantial programs. Topics include recursion, pointers, linked lists, abstract data types, and libraries. Also covers software engineering principles. Credit is awarded for only one of CS 012 or CS 012V or CS 013.

CS 014 Introduction to Data Structures and Algorithms (4)

Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 012 or CS 012V with a grade of “C” or better or CS 013 with a grade of “C” or better; proficiency in C++. Topics include basic data structures such as arrays, lists, stacks, and queues. Covers dictionaries (including binary search trees and hashing) and priority queues (heaps). Offers an introductory analysis of algorithms, sorting algorithms, and object-oriented programming including abstract data types, inheritance, and polymorphism. Explores solving complex problems through structured software development.

CS 100 Software Construction (4)

Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 014 with a grade of “C-” or better. Emphasizes development of software systems. Topics include design and implementation strategies; selection and mastery of programming languages, environment tools, and development processes. Develops skill in programming, testing, debugging, performance evaluation, component integration, maintenance, and documentation. Covers professional and ethical responsibilities and the need to stay current with technology.

CS 105 Data Analysis Tools (4) -- new course --

Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 14. Introduction to data analysis tools including data statistics, simple data storage types, data acquisition from the web and public APIs, data cleaning, crowdsourcing for

data collection and cleaning, supervised and unsupervised learning techniques, and data visualization. The laboratory will also include hands-on exercises on the aforementioned topics in Python and MATLAB.

CS 111 Discrete Structures (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 010 or CS 010V; CS 011/MATH 011; MATH 009C (or MATH 09HC). A study of discrete mathematical structures emphasizing applications to computer science. Topics include asymptotic notation, generating functions, recurrence equations, elements of graph theory, trees, algebraic structures, and number theory.

CS 141 Intermediate Data Structures and Algorithms (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 014 with a grade of “C-” or better; CS 111; MATH 009C or MATH 09HC; proficiency in C++. Explores basic algorithm analysis using asymptotic notations, summation and recurrence relations, and algorithms and data structures for discrete structures including trees, strings, and graphs. Also covers general algorithm design techniques including “divide-and-conquer,” the greedy method, and dynamic programming. Integrates knowledge of data structures, algorithms, and programming.

CS 166 Database Management Systems (4) Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 100, CS 111. Covers architecture of database management systems; relational, network, and hierarchical models; distributed database concepts; query languages; implementation issues; and privacy and security of the database.

CS 167 Introduction to Big Data Management (4) -- new course-- Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 100, CS111. Introduces the design of big-data systems and their application in data management and processing. Describes the common functionality in big-data processing such as distributed storage, resource management, query processing, fault-tolerance, and APIs. Covers the popular big-data technologies such as MapReduce, key-value stores, and semi-structured data management.

CS 171 Introduction to Machine Learning and Data Mining (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 100, CS 111. Introduces formalisms and methods in data mining and machine learning. Topics include data representation, supervised learning, and classification. Covers regression and clustering. Also covers rule learning, function approximation, and margin-based methods.

CS 179 (E-Z) Project in Computer Science (4) For hours and prerequisites, see segment descriptions. Under the direction of a faculty member, student teams propose, design, build, test, and document software and/or hardware devices or systems. Emphasizes professional and ethical responsibilities and the need to stay current on technology and its global impact on economics, society, and the environment.

MATH 009A First-Year Calculus (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): MATH 005 with a grade of “C-” or better or MATH 006B with a grade of “C-” or better or equivalent. Introduction to the differential calculus of functions of one variable. Credit is awarded for only one of MATH 008B, MATH 009A, or MATH 09HA.

MATH 009B First-Year Calculus (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): MATH 008B with a grade of “C-” or better or MATH 009A with a grade of “C-” or better or MATH 09HA with a grade of “C-” or better. Introduction to the integral calculus of functions of one variable. Credit is awarded for only one of MATH 009B or MATH 09HB.

MATH 009C First-Year Calculus (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): MATH 009B with a grade of “C-” or better or MATH 09HB with a grade of “C-” or better. Further topics from integral calculus, improper integrals, infinite series, Taylor’s series, and Taylor’s theorem. Credit is awarded for only one of MATH 009C or MATH 09HC.

MATH 010A Calculus of Several Variables (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): MATH 009B with a grade of “C-” or better or MATH 09HB with a “C-” or better or equivalent. Topics include Euclidean geometry, matrices and linear functions, determinants, partial derivatives, directional derivatives, Jacobians, gradients, chain rule, and Taylor’s theorem for several variables.

MATH 011 Introduction to Discrete Structures (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): MATH 009A (or MATH 09HA); CS 010 or CS 010V or MATH 009B (or MATH 09HB). Introduction to basic concepts of discrete mathematics emphasizing applications to computer science. Topics include propositional and predicate calculi, elementary set theory, functions, relations, proof techniques, elements of number theory, enumeration, and discrete probability. **Cross-listed with CS 011.**

MATH 031 Applied Linear Algebra (5) Lecture, 3 hours; discussion, 2 hours. Prerequisite(s): MATH 009A (or MATH 09HA) with a grade “C-” or better and CS 010 or CS 010V or MATH 009B (or MATH 09HB) with a grade of “C-” or better. A study of matrices and systems of linear equations, determinants, Gaussian elimination, vector spaces, linear independence and linear transformation, orthogonality, eigenvalues, and eigenvectors. Also examines selected topics and applications.

STAT 100A Introduction to Statistics (5) Lecture, 3 hours; discussion, 1 hour; laboratory, 3 hours. Prerequisite(s): MATH 005 or MATH 006B or MATH 009A or MATH 09HA or equivalent. A general introduction to descriptive and inferential statistics. Topics include histograms; descriptive statistics; probability; normal, binomial, and Poisson distributions; sampling distributions; hypothesis testing; and confidence intervals. Credit is awarded for only one of STAT 048 or STAT 100A.

STAT 100B Introduction to Statistics (5) Lecture, 3 hours; discussion, 1 hour; laboratory, 3 hours. Prerequisite(s): STAT 100A “An introduction to statistics” with a grade of C- or better. Topics include linear regression, correlation, analysis of variance, and simple experimental designs.

STAT 147 Introduction to Statistical Computing (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 100A or equivalent. Introduction to computer-assisted data analysis and statistical inference using both the R and SAS packages. Topics include input, output, and editing of data; graphical procedures; descriptive statistics; cross-tabulation; inferential statistical techniques including estimation and testing; and analysis of variance.

STAT 156A Mathematical Statistics with Applications for Data Science I (4) -- new course-- Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): MATH 009C or consent of instructor. Introduction to frequentist probability concepts, random variables, and their distributions. Discussion of key theorems and inequalities in probability theory. Introduction to frequentist methods of point and interval estimation.

STAT 156B Mathematical Statistics with Applications for Data Science II (4) -- new course-- Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 156A or consent of instructor. Introduction to Bayesian probability concepts, illustrative application of Frequentist theory to linear regression, logistic regression and ANOVA, analysis of contingency tables, applications of sequential statistics, methods for observational studies and for missing data analyses.

STAT 167 Introduction to Data Science (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 100B or STAT 155, with a grade of C- or better, or equivalents; STAT 147, with a grade of C- or better. Introduction to data science using the R programming language. Topics include big data management, visualization and analytical skills, unsupervised and supervised statistical learning methods, and real-world data science application examples.

STAT 170A Regression Analysis (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 157, STAT 160C, or equivalents. Topics include simple and multiple linear regression, scatter-plots, and point and interval estimation. Addresses prediction, testing, calibration, interpretation, and practical applications of multiple regression. Explores simple, partial, and multiple correlation; variable selection methods; diagnostic procedures; and regression for longitudinal data.

STAT 170B Design of Experiments (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 170A with a grade of “C-” or better. Topics include principles of design; completely randomized designs; and one-way analysis of variance. Covers complete block designs and two-way analysis of variance; multiple comparisons; and complete factorial experiments. Explores fixed, random, and mixed models; split-plot designs; nested designs; analysis of covariance; sample size determination; and power analysis.

STAT 183 Statistical Consulting (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 170B; STAT 171, may be taken concurrently; Restricted to class level standing of senior. Introduces the statistical consulting process. Promotes consulting skills including developing effective communication skills, applying statistical methodology to client projects, and learning how to manage time and resources in a consulting environment. Satisfactory (S) or No Credit (NC) grading is not available.

ENGR 170 Technology, Policy, and Ethics (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): upper-division standing. Provides contemporary perspectives on interplays between technology, public policy, and ethics. Covers social, legal, and ethical issues such as liability, as well as environmental, patent, and copyright law. Cross-listed with PBPL 170.

TECHNICAL ELECTIVES

MATH 120 Optimization (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): MATH 010A with a grade of "C-" or better; MATH 031 with a grade of "C-" or better. Introduction to classical optimization including unconstrained and constrained problems in several variables, Addresses Jacobian and Lagrangian methods and the Kuhn-Tucker conditions. Covers the basic concepts of linear programming including the simplex method and duality with applications to other subjects.

MATH 135A Numerical Analysis (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 010 or CS 010V or equivalent with a grade of "C-" or better; MATH 031 with a grade of "C-" or better (may be taken concurrently). A study of numerical methods for determining solutions to nonlinear equations and simultaneous linear equations. Topics also include interpolation, techniques of error analysis, and computer applications.

CS 170 Introduction to Artificial Intelligence (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 100 with a grade of "C-" or better, CS 111. An introduction to the field of artificial intelligence. Focuses on discrete-valued problems. Covers heuristic search, problem representation, and classical planning. Also covers constraint satisfaction and logical inference.

CS 172 Introduction to Information Retrieval (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 100; CS 111; EE 114 or STAT 155. Introduces information retrieval (IR) principles and techniques for indexing and searching document collections. Topics include Web search, text processing, ranking algorithms, search in social networks, and search evaluation. Also studies scalability issues in search engines. Satisfactory (S) or No Credit (NC) grading is not available.

CS 180 Introduction to Software Engineering (4) Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 100. A study of software engineering techniques for the development,

maintenance, and evolution of large software systems. Topics include requirements and specification; system design and implementation; debugging, testing, and quality assurance; reengineering; project management; software process; tools; and environments.

CS 181 Principles of Programming Languages (4) Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): CS 061, CS 100, CS 111, CS 150. Covers the principles of programming language design. Includes the study and comparison of several programming languages, their features, and their implementations.

STAT 104 Decision Analysis and Management Science (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 008 or equivalent; STAT 048 or STAT 100A or equivalent; upper-division standing. A survey of deterministic and probabilistic models for decision making. Topics include linear programming and extensions, networks, dynamic programming, decision trees, queuing models, and simulation. Explores the application of these models in decision making. Emphasizes use of the computer. Cross-listed with BUS 104.

STAT 127 Introduction to Quality Improvements (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 048 or STAT 100A or consent of instructor. Explores Deming's 14 points for management, graphical methods, fishbone diagram, Pareto analysis, control charts for attributes and variables, cusum and moving average charts, process-capability, economic design, acceptance sampling, Taguchi method, parameter design, tolerance design, reliability, hazard rate, censoring, and accelerated life testing. Cross-listed with BUS 127.

STAT 130 Sampling Surveys (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 100B, or equivalent. Covers simple random sampling, addresses stratified sampling, cluster sampling, and ratio and regression estimates. Explores random response, capture-recapture, and jack-knife techniques.

STAT 140 Nonparametric Techniques (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 100B or equivalent. Covers randomization tests, rank tests, methods of association, and distribution-free tests.

STAT 146 Statistical Forecasting Techniques (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 100B or equivalent. Topics include exponential smoothing, simple and multiple regression analysis, time series, trend analysis, and seasonal analysis

STAT 157 Statistical Computer Packages (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 100B with a grade of C- or better, or equivalents; STAT 147 with a grade of C- or better; or consent of instructor. A study of major statistical packages including SAS with the emphasizing advanced SAS programming. Topics include advanced graphical procedures, linear models (regression and analysis of variance), multivariate techniques, and SAS macros.

STAT 171 General Statistical Models (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): STAT 170B. Generalized linear models and least squares. Analysis of covariance, nonlinear regression, nonlinear least squares. Regression methods for discrete data: loglinear models, logistic regression, discriminant analysis. Regression methods for life data. Cox survival model, Kaplan-Meier estimation, Mantel-Haenszel test.

APPLICATION COURSE SEQUENCES

Economics: {ECON 108, ECON 136}

Business: {BUS 104, BUS 123}, {BUS 124, BUS 125}, {BUS 103, BUS 115}.

Earth Sciences: {GEO 111, GEO 161}, {GEO 115, GEO 147}

Electrical and Computer Engineering: {EE142, EE146}

Biology/Bioinformatics: {BIO 5A, BIO 20}

BIOL 005A Introduction to Cell and Molecular Biology (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): BIOL 05LA (may be taken concurrently) or BIOL 020 (may be taken concurrently) with grades of “C-” or better; CHEM 001A and CHEM 01LA with grades of “C-” or better or CHEM 01HA and CHEM 1HLA with grades of “C-” or better; consent of instructor is required for students repeating the course. An intensive course designed to prepare for upper-division courses in cell and molecular biology. Covers biochemical, structural, metabolic, and genetic aspects of cells. Required for Biology majors; recommended for science majors desiring an introduction to biology.

BIOL 020 Dynamic Genome (2) F Laboratory, 6 hours. Prerequisite(s): CHEM 001A or CHEM 01HA, MATH 008B or MATH 009A (MATH 009A may be taken concurrently); freshman standing. Introduces computational and experimental approaches in investigating the genomes of plants and animals. Explores scientific discovery using the tools of bioinformatics and genomics. Includes participation in research projects being conducted on campus. Credit is not awarded for BIOL 020 if it has already been awarded for BIOL 05LA.

BUS 104 Decision Analysis and Management Science (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 008 or equivalent; STAT 048 or STAT 100A or equivalent; upper-division standing. A survey of deterministic and probabilistic models for decision making. Topics include linear programming and extensions, networks, dynamic programming, decision trees, queuing models, and simulation. Explores the application of these models in decision making. Emphasizes use of the computer. Cross-listed with STAT 104.

BUS 123 Spreadsheet Modeling for Decision-Making (4) Lecture, 3 hours; written work, 3 hours. Prerequisite(s): BUS 104/STAT 104 or consent of instructor. Introduces the fundamental techniques of using data to make informed management decisions in the presence of uncertainty of advanced Microsoft Excel functionality. Uses spreadsheet modeling for decision analysis and optimization applications. Enhances and reinforces analytical skills and the ability to intelligently use information for making decisions under uncertainty.

BUS 124 Business Analytics (4) Lecture, 3 hours; term paper, 1 hour; written work, 2 hours. Prerequisite(s): STAT 048 or consent of instructor. Provides fundamental concepts and tools needed to understand the emerging role of business analytics in organizations. Applies basic business analytics tools in a spreadsheet environment. Introduces market-leading techniques that help identify and manage key data from business processes. Provides the essential tools required for data mining and business process re-engineering.

BUS 125 Simulation for Business (4) Lecture, 3 hours; extra reading, 1.5 hours; outside projects, 1.5 hours. Prerequisite(s): BUS 104/STAT 104, STAT 048, or equivalents. Introduces simulation as a tool for analyzing complex systems. Analyzes and discusses the theory and practice of modeling through simulation. Topics include modeling uncertainty and collecting input data, Monte Carlo simulation techniques, model verification and validation, and sensitivity analysis. Examines applications in finance, marketing, operations, and supply chain management.

BUS 103 Marketing and Distribution Management (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): upper-division standing or consent of instructor. An introduction to the role of marketing in society with emphasis on concepts, marketing methods, and institutions.

BUS 115 Marketing Research (4) Lecture, 3 hours; outside research, 3 hours. Prerequisite(s): BUS 103. Covers types and sources of marketing information, the marketing research process, and techniques of data collection and analysis, including consumer and customer surveys and test marketing. Examines both quantitative and qualitative research with analysis of the values and limitations of data. Emphasis is placed on evaluation and interpretation of results.

ECON 108 Introductory Econometrics (5) Lecture, 3 hours; discussion, 1 hour; laboratory, 1 hour; written work, 1 hour; outside research, 1 hour. Prerequisite(s): ECON 107 or consent of instructor. A continuation of ECON 107. Covers, at an introductory level, the basic concepts related to logit and probit models, simultaneous equations models, dynamic time series models, unit roots and auto-regressive conditional heteroskedasticity (ARCH), and forecasting.

ECON 136 Empirical Financial Economics (4) Lecture, 3 hours; individual study, 3 hours. Prerequisite(s): ECON 107 or consent of instructor. Discusses various empirical aspects of financial economics and financial risk management. Addresses both theoretical and applied issues in finance, risk management, and econometrics. Also discusses quantitative analysis, simulation methods, and case studies.

EE 142 Pattern Recognition and Analysis of Sensor Data (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): EE 114 or STAT 155 or consent of the instructor. Introduction to pattern recognition for multi-dimensional, multi-modal sensor data such as images, videos, and smart grids. Classification and decision functions, feature extraction, regression, and neural networks. Clustering and dimensionality reduction for unsupervised learning. Dynamic models and tracking. Applications of pattern recognition in computer vision, robotics, smart grids, etc.

EE 146 Computer Vision (4) Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): senior standing in Computer Science or Electrical Engineering, or consent of instructor. Imaging formation, early vision processing, boundary detection, region growing, two-dimensional and three-dimensional object representation and recognition techniques. Experiments for each topic are carried out.

GEO 111 Numerical Skills in Geoscience (4) Lecture, 3 hours; laboratory, 3 hours; term paper, 1 hour;. Prerequisite(s): MATH 009C. An introduction to computer programming and numerical modelling. The class will teach the basic principles of how computer programs are written and numerical models constructed, and provide data processing and visualization skills. The class will hence foster an ability to apply numerical techniques to problems in Earth and Environmental Sciences.

GEO 161 Cenozoic Climate Change (4) Lecture, 3 hours; laboratory, 3 hours. Prerequisite(s): one of the following courses with a grade of "C-" or better; GEO 001 or GEO 002 or GEO 009 or GEO 011. Examines physical, chemical, and biological evidence of climatic and environmental change throughout the Cenozoic Era (last 65 million years) to provide a framework for understanding natural environmental change and for predicting future change. Introduces students to computer-based numerical methods of data analysis for interpreting past records of environmental change.

GEO 115 Geologic Maps and Landforms (5) Lecture, 2 hours; laboratory, 6 hours; field, 30 hours per quarter. Prerequisite(s): GEO 001 (may be taken concurrently); MATH 004 or MATH 005, or MATH 008A. Examines characteristic patterns of bedrock outcrops, surficial deposits, the related landforms, and their representation on maps. Covers unconformities, folds, faults, intrusions, alluvial fans, river terraces, and landforms indicative of glaciers, volcanoes, landslides, and earthquakes. Applies map information to resource and hazard evaluation.

GEO 147 Active Tectonics and Remote Sensing (4) Lecture, 2 hours; discussion, 1 hour; laboratory, 3 hours. Prerequisite(s): GEO 001, GEO 115; or consent of instructor. A computer-based course that introduces active tectonics and the earthquake cycle and how they are studied using remote sensing data. Explores examples of actively deforming areas from around the world using computer visualization software and freely available data sources (satellite imagery, digital topography, GPS and earthquake data).

Appendix B:

Proposed catalog entry for the Data Science Undergraduate Program

Data Science Undergraduate Major (Catalog entry)

Major

Data science studies the collection, management, and analysis of data to extract knowledge. It is a multidisciplinary program with core components from Computer Science and Statistics, and required application study in a variety of empirical disciplines. Courses span the discipline from theory to practice and prepare students for careers or graduate studies in data-intensive fields.

The B.S. in Data Science major is an intercollege major offered by the Bourns College of Engineering and the College of Natural and Agricultural Sciences. A B.S. degree in Data Science is offered by each college. When students declare the major, they choose from which college they wish to have their degree awarded. Students whose degrees are awarded by the Bourns College of Engineering are advised in and have their records maintained by the BCOE Office of Student Academic Affairs; students whose degrees are awarded by the College of Natural and Agricultural Sciences are advised in and have their records maintained by the CNAS Undergraduate Academic Advising Center. Breadth requirements vary by college; and students must fulfill the breadth requirements of the college they choose.

All undergraduates in the Bourns College of Engineering must see an advisor at least annually. Visit student.engr.ucr.edu for details.

University Requirements

See Undergraduate Students section.

College Requirements

College breadth requirements vary depending on which college is chosen to award the degree. For details on breadth requirements, see the Colleges and Programs section of this catalog. Students are encouraged to consult their advisor regarding requirements.

Major Requirements

1. Lower-division requirements (37 units):
 - a) CS 010; CS 012; CS 014
 - b) MATH 009A; MATH 009B; MATH 009C; MATH 010A; MATH 031
 - c) MATH 011/CS 011
2. Upper-division requirements (76 units, minimum):
 - a) CS 100; CS 105; CS 111; CS 141
 - b) STAT 147; STAT 156A; STAT 156B; STAT 170A; STAT 170B
 - c) ENGR 170 or PBPL 170

- d) CS 166 or CS 167
- e) STAT 167 or CS 171
- f) STAT 183 or CS 179 (E-Z)
- g) Four courses (at least 16 units) from the following list, none of which can also be used to satisfy other major requirements:
CS 166; CS 167; CS 170; CS 172; CS 180; CS 181; MATH 120; MATH 135A;
STAT 104; STAT 127; STAT 130; STAT 140; STAT 146; STAT 157; STAT 171.
- h) One two-course sequence, chosen from the course sequences listed below:
 - i. BIO 5A and BIO 20
 - ii. BUS 103 and BUS 115
 - iii. BUS 104 and BUS 123
 - iv. BUS 124 and BUS 125
 - v. ECON 108 and ECON 136
 - vi. EE 142 and EE 146
 - vii. GEO 111 and GEO 161
 - viii. GEO 115 and GEO 147

Note An introductory Statistics class, such as STAT 100A and STAT 100B, is strongly recommended.

Appendix C:


Letters of support/collaboration from other Departments (Mathematics, Earth Sciences, Economics, EEOB, Electrical and Computer Engineering)

DATE: October 18, 2018

TO: Dr. Vassilis Tsotras, Department Chair
Department of Computer Science

FROM: Dr. Yat Sun Poon, Department Chair
Department of Mathematics

RE: B.S. in Data Science proposal



Dear Vassilis,

I am glad to hear that a BS degree in data science is being brought to UCR through a joint effort between the Statistics and Computer Science departments. I believe many students will find it appealing. I cannot see any potential problems with your plans.

----- Forwarded Message -----

Subject: Earth Sciences courses for the new Data Science undergraduate program

Date: Fri, 14 Dec 2018 15:52:33 -0800

From: David Oglesby <doglesby@ucr.edu>

To: Vassilis Tsotras <tsotras@cs.ucr.edu>

Dear Vassilis:

I am in support of the Data Science undergraduate program between the Earth Sciences and Statistics Departments. As we discussed, I am in favor of adding the following Earth Sciences two-course sequences as electives available to the undergraduate students in the Data Science program:

GEO 111 and GEO 161
GEO 115 and GEO 147

Even though GEO 115 requires MATH 004 or MATH 005, or MATH 008A, the Data Science students would already have appropriate MATH core courses (e.g. MATH 9A-9C) so they would be allowed to take GEO115.

Sincerely,
David

David D. Oglesby
Professor of Geophysics
Chair, Department of Earth Sciences
University of California, Riverside
Riverside, CA 92521-0423
PH: (951) 827-2036
FAX: (951) 827-4324
david.oglesby@ucr.edu

From: **Steven M Helfand** <steven.helfand@ucr.edu>
Date: Mon, Nov 26, 2018 at 5:32 PM
Subject: Data Science Majors
To: Daniel R Jeske <daniel.jeske@ucr.edu>
Cc: Gloria Gonzalez-Rivera <ggonzale@ucr.edu>, Anthony L Gonzalez
<anthony.gonzalez@ucr.edu>, Bradley Troy Scalf <bradley.scalf@ucr.edu>

Dean Dan,

I write to affirm that the Economics Department will accept STAT 170A as a suitable prerequisite for Econ 108 and Econ 136. This will permit the Data Science majors to take these classes without taking Econ 107 first. Academically, we see no difficulty with this substitution.

Understanding that the data science program is just beginning and that data science students will likely not be taking Econ 108 or 136 until their third year in the program, we would like to move forward with this agreement for a trial period of five years, starting in 2019-20. We wish to do this as a five year experiment because we are concerned about the size of the classes, and Econ 108 in particular. Currently we plan to offer Econ 108 once per year, with a maximum size of 35. This is the size of the lab, and all that we can handle with one TA who teaches the sections and lab. If it were to grow beyond 35 students, we would need to hire a second TA, and we don't currently have funding to do this. At its current size, we expect that there would be 5-10 seats each year for non-econ majors.

Sincerely,

Steven

Steven Helfand
Chair, Department of Economics
University of California
Riverside, CA 92521-0427
Office: (951) 827-1470
steven.helfand@ucr.edu
<http://economics.ucr.edu/people/faculty/helfand/>

November 27, 2018
Professor Helen Regan
Chair
Department of Evolution, Ecology, and
Organismal Biology

Prof. Vassilis Tsotras
Director, Data Science Center at UCR

Dear Tsotras:

The Department of Evolution, Ecology, and Organismal Biology supports initiation of the Data Science undergraduate program at UCR. We are in favor of adding two Biology courses to be taken as electives by the undergraduate students in the Data Science program. The following two Biology courses are most relevant to the Data Science undergraduate program:

BIOL 005A – Introduction to Cell and Molecular Biology
BIOL 020 – Dynamic Genome

Both courses are offered every quarter, often more than once per quarter, and students from CSE and Statistics should be able to take them provided the prerequisites are satisfied.

Sincerely,



Helen Regan,
Professor and Chair
Department of Evolution, Ecology, and Organismal Biology at UCR



Bourns College of Engineering
Department of Electrical and Computer Engineering
343 Winston Chung Hall
Riverside, CA 92521
(951) 827-2484

April 22, 2018

Prof. Vassilis Tsotras
Director, Data Science Center at UCR

Dear Vassilis:

I strongly support initiation of the Data Science undergraduate program at UCR. I am also very much in favor of adding two ECE-based courses to be taken as electives by the undergraduate students in the Data Science program. The following two ECE courses are most relevant to the Data Science undergraduate program:

EE142 - Pattern Recognition and Analysis of Sensor Data
EE146 - Computer Vision

Both courses also allow EE prerequisites to be waived by consent of instructor, and students from CSE and Statistics should be able to take them.

Sincerely,

A handwritten signature in blue ink, appearing to read "Ilya Dumer", with a long horizontal flourish extending to the right.

Ilya Dumer,
Professor and Chair
Department of Electrical and Computer Engineering at UCR

Appendix D:

Letters of support from the chairs of CSE and Statistics and the BCOE and CNAS Deans.

December 28, 2018

To Whom It May Concern:

This letter is in strong support for the proposed undergraduate program in Data Science at UCR, to be jointly offered by the Departments of Computer Science & Engineering and Statistics.

Data Science has grown out of the need to integrate computational and statistical approaches to processing and interpreting data. Tools originating from data science are now becoming indispensable in today's science, technology, and business, fueling the demand for data scientists. Recognizing this need, our department has taken the initiative to develop research and educational programs in Data Science at UCR. In collaboration with other departments on campus, an online MS program in Data Science is already being offered. Recently the Data Science Center has been established that includes multiple newly hired faculty members, and has been given designated space in the new MRB building. Creating an undergraduate program in Data Science is the next step in this endeavor. This program will address critical and documented shortage of college graduates trained in Data Science, in industry, government, and academia.

The CSE Department enthusiastically supports the creation of the Data Science program and is fully committed to providing necessary resources within its capabilities for the instruction and advising of its students. As Data Science is at the intersection of Statistics and Computer Science, we look forward to collaborate with the Statistics department to establish the new undergraduate Data Science program at UCR.



Walid A. Najjar

Professor and Chair
Department of Computer Science and Engineering
Bourns College of Engineering
University of California Riverside

November 14, 2018

Dear Committee Members:

On behalf of the UCR Department of Statistics, I am writing in strong support for the proposed inter-departmental undergraduate program in Data Science at UCR.

Data Science is experiencing a rapid and unplanned growth. The website <http://datascience.community/colleges> currently lists 530 programs in Data Science, analytics and related fields at over 200 universities around the world. The vast majority of these are master's degrees and certificates programs offered both traditionally and online. There has been rapid growth of undergraduate programs at both research institutions and liberal arts colleges. It is expected this number to increase significantly in the near future.

Recognizing this need, our department has taken steps to develop research and educational programs in Data Science at UCR including the new undergraduate course "Introduction to Data Science" that has been taught for the past two years and the new graduate courses "Statistical Computing" and "Statistical Data Mining Methods" that have been taught for both Statistics graduate program and MS online program in Data Science at UCR.

Data Science program at the undergraduate level provides a synergistic approach to real world problem solving, one that leverages the content in Statistics, Computer Science and Mathematics but using case-based focus and hands-on approach. We have successfully boosted undergraduate enrollment into our Statistics program over past few years. The number of students had increased from a dozen to 136. We believe we will be able to attract more students with curriculum uniquely tailored for Data Science program, i.e., "students learn Data Science by doing Data Science". We believe our Data Science program will serve students well whether they join the marketplace or continue on to more advanced study.

Recognizing that Data Science is at the intersection of Statistics and Computer Science, the statistics department embraces the opportunity to partner with the computer science and engineering department to establish an undergraduate Data Science program at UCR.

Sincerely,



Xinping Cui, Ph.D.
Professor and Chair
Department of Statistics
University of California, Riverside
Tel: (951) 827-2563 | Fax: (951) 827-3286
E-mail: Xinping.Cui@ucr.edu

27 November, 2018

To whom it may concern:

I am writing this letter in enthusiastic support for the enclosed proposal to establish an undergraduate degree program in Data Science. This program will be jointly administered between BCOE and CNAS, with degrees granted by each college. I have had detailed conversations with Professor Tsotras and the program committee and fully support the academic program and administrative structure. I commit to working with them and leadership from CNAS to insure the program's success.

This program will help address the critical and documented shortage of college graduates educated in Data Science and the critical interpretation and analysis of large datasets. We expect students attracted to this program to come from a variety of backgrounds and other interests, increasing the diversity among Engineering students, and those in computational fields in particular. This program is partially supported by a grant from the Center for Advancing Women in Technology (CAWIT), which funds two L(P)SOE positions (one in BCOE and one in CNAS). I have recently joined the CAWIT advisory board and believe the data science program will bring more women into computing careers.

The Bourns College of Engineering looks forward to partnering with the College of Agricultural and Natural Sciences to offer this degree to UCR students. It is an important part of keeping our curriculum current and educating our students in the era of big data.

Sincerely,



Prof. Christopher S. Lynch
Dean, Bourns College of Engineering
University of California, Riverside

December 10, 2018

To Whom It May Concern:

RE: Data Science Undergraduate Major

I am writing to convey my support for the proposal to establish an undergraduate degree program in Data Science. The Department of Statistics in the College of Natural and Agricultural Sciences ("CNAS") and the Department of Computer Science and Engineering in the Bourns College of Engineering ("BCOE") will jointly administer this program, which will allow students to obtain a BS degree through an integrated plan of study. The program will provide an efficient and cost-effective academic path for students interested in careers within a variety of areas relating to data science.

Data Science as a discipline is found at the intersection of computer science and statistics and permeates other disciplines as well. For this program, the partnership between CNAS and BCOE is well demonstrated by the list of faculty involved: a healthy mix of faculty from both colleges. Sharing the costs for the director and co-director stipends between the two colleges is another illustration of integration of the two colleges in the program. Costs for director and co-director stipends will be equally split (50% - 50%) between CNAS and BCOE. Exact amounts will be determined at the time the program is approved.

We look forward to partnering with the Bourns College of Engineering to offer this degree program to our students.

Sincerely,



Kathryn Uhrich, Dean

College of Natural and Agricultural Sciences

February 28, 2019

To Whom It May Concern:

Re: Proposal for the new undergraduate major in Data Science

The finance and administration team in the BCOE Dean's Office conducted an enrollment and financial analysis to determine the financial implications of establishing an undergraduate program in Data Science at UCR. This program will be jointly offered by the Department of Computer Science & Engineering and the Department of Statistic. The analysis is attached to this letter.

Based on the attached analysis, the BCOE Dean's Office finance and administration team does not have any concerns with the establishment of this degree program. The analysis was shared with the CNAS Dean's Office who did not indicate any concerns with the establishment of this degree program.

Sincerely,

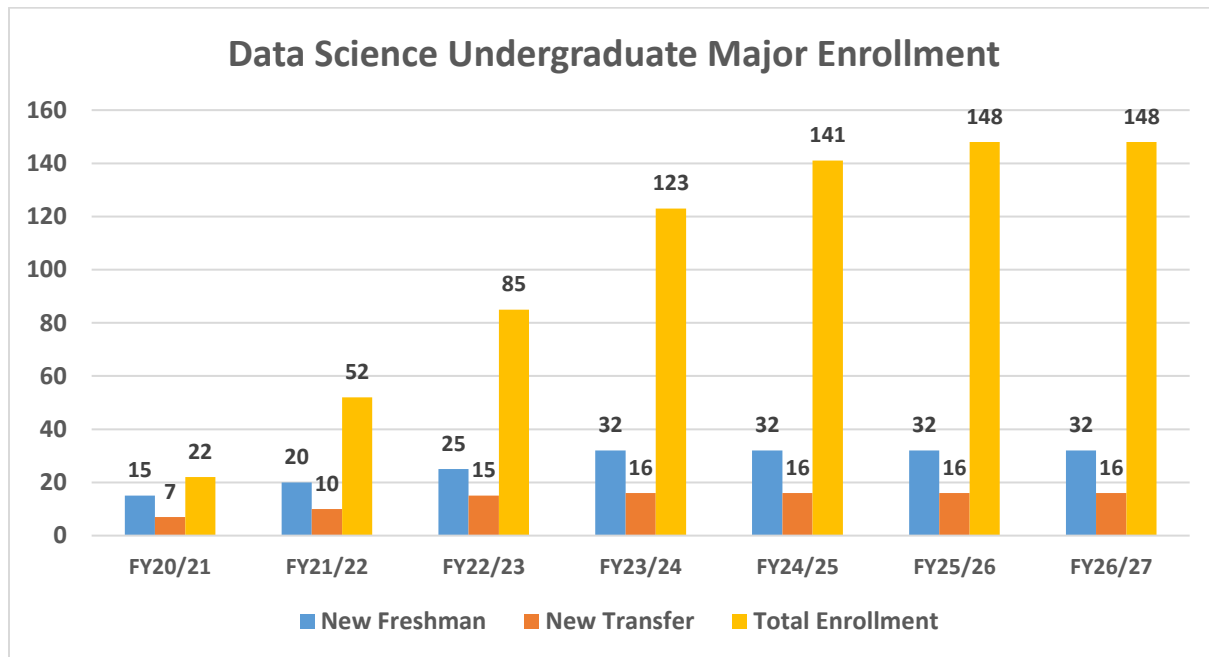


Veronica Ruiz
Assistant Dean, Finance and Administration
Marlan & Rosemary Bourns College of Engineering

Proposal for the new undergraduate major in Data Science

Enrollment, Degree Attainment, and Financial Resources

Projected Enrollment: The projected enrollment at the start of the program is 25-30 students; we expect a class of 50 students at steady state.



Anticipated Number of Degrees: We anticipate the BS in Data Science major will award about 40-50 degrees awarded per year at steady state.

	# of Majors	# of Degrees
FY2020/2021	22	0
FY2021/2022	52	0
FY2022/2023	85	7
FY2023/2024	123	10
FY2024/2025	141	30
FY2025/2026	148	41
FY2026/2027	148	48

Tuition Revenue and Financial Resources: The Data Science major will use existing faculty from BCOE and CNAS. New LPSOE's will teach some courses and the percentage of their time will increase as the program size grows. There will be a Director and Co-Director with stipends paid by BCOE and CNAS. TA appointments and the need for additional student advisor support will increase as the program grows, causing the need to hire additional TAs and advisors. No other new resources will be needed.

Data Science UG Major	FY20/21	FY21/22	FY22/23	FY23/24	FY24/25	FY25/26	FY26/27
No. of DS Majors	22	52	85	123	141	148	148
Headcount Major	1,070	1,072	1,072	1,074	1,074	1,074	1,075
Wrkload/Crdt Hrs (45 units = 1)	3,355	3,357	3,357	3,359	3,359	3,359	3,360
Tuition Revenue Data Science	97,350	230,308	376,465	545,259	625,053	656,084	656,380
Faculty - using existing	0	0	0	0	0	0	0
LPSOE - 1 @ 25% & existing	42,425	70,409	100,414	135,583	169,576	205,486	211,651
Directors - covered by Schools	0	0	0	0	0	0	0
TA appointments	0	21,029	35,009	49,520	51,308	57,997	59,541
Admin staff - add 25% & incrise	0	22,226	23,033	47,506	48,961	50,461	78,010
Library Acquisitions- no new	0	0	0	0	0	0	0
IE/Facilities/Labs - no new	0	0	0	0	0	0	0
Space/infrastructure - no new	0	0	0	0	0	0	0
Total Expenses >	42,425	113,664	158,456	232,609	269,844	313,944	349,201
Net Income for Data Science	54,925	116,644	218,009	312,650	355,209	342,140	307,179

The Bourns College of Engineering Dean's Office prepared this document, using estimates for enrollment and anticipated number of degrees provided by Dr. Tsotras and Dr. Jeske in the Data Science Undergraduate major proposal application. Tuition revenue was calculated using the estimated enrollment and calculations for headcount major and workload credits from the Campus Budget Process Appendix 2 – Tuition Revenue Distribution, with projected increases for future years. Expenses were calculated using data from the proposal related to the need for faculty, LPSOE's, directors, TAs, administrative staff, and other standard items related to the cost of undergraduate instruction. The LPSOE cost is based on the current BCOE LPSOE salary average (\$116,027) and current benefit rate (39.1%) with projected increases salary increases (3%) and estimated benefit increases. The percent of time increases over the years based on the enrollment projection growths. The same calculation method was used for the administrative staff, which is based on a Student Academic Advisor 2 salary at the 25th percentile with estimated merit increases (3%) and benefit costs increases from the current 58.8%. The need for increasing administrative support rises as the enrollment rises. The director stipends are not included in the calculation as those costs will be covered by BCOE and CNAS. While this data was calculated by BCOE staff, it was shared and discussed with the financial team in the CNAS Dean's Office and both parties agree that the initial calculations are appropriate based on the proposal information and current revenue distributions and salary/benefit costs.



April 15, 2019

To: Dylan Rodríguez, Chair
Riverside Division

From: Paul Lyons, Chair
Committee on Educational Policy

A handwritten signature in black ink that reads "Paul Lyons".

Re: Proposed New B.S. in Data Science


The Committee on Educational Policy (CEP) reviewed the proposal for a new B.S. in Data Science at their April 12, 2019 meeting and voted to support the proposal. The Committee does recommend that the proposal consider adding an experimental sciences sequence of courses to the curriculum for the program's first year of course work, which will help students to understand the background of the discipline of Data Science.



PLANNING & BUDGET

April 12, 2019

To: Dylan Rodriguez, Chair
Riverside Division

From: Katherine Kinney, Chair 
Committee on Planning and Budget

Re: Proposal: New Undergraduate Major: Data Science

Planning and Budget (P&B) discussed the proposal for an undergraduate major in Data Science at their April 9, 2019 meeting. P&B would like confirmation that the Deans will plan to work together to provide TA support. The costs for the program generally are well accounted at the modest enrollment target of 50 students. The committee would like to know how this number was set and if 50 represents a cap on enrollment. If it is not a cap, we would also like to know how higher enrollments would be supported.