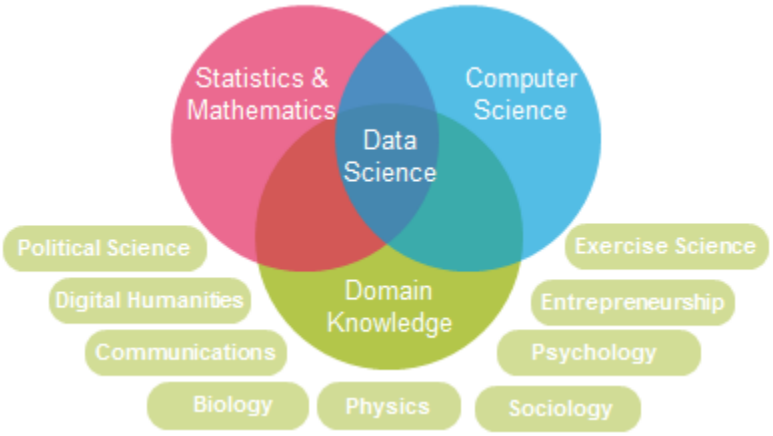
New Degree Program Proposal
Bachelor of Science in Data Science

Table of Contents

1. Broad Description and Purpose of Program

The increasing importance of big data and high demand for data science skills motivates the Department of Mathematics and Computer Science to propose a new program of Bachelor of Science in Data Science, as well as a new minor in Data Science. The goal of the program is to provide students with the technical skills and practical insight to Data Science in a variety of application domains.

The Data Science program utilizes an interdisciplinary approach with a focus on Statistics, Computer Science, and Mathematics applied to a specific domain such as Political Science, Psychology, or Communications. Given its nature, the degree will be organized by the Department of Mathematics and Computer Science along with an interdisciplinary advisory team.



Interdisciplinary Data Science

Justification for Data Science Program

Data science is an emerging academic discipline, a response to an increasing demand for people who are able to understand and analyze an enormous amount of data. Data science is a field at the intersection of statistics, computer science, mathematics, and real world application domains. While data science is not tied exclusively to the field of big data, it capitalizes on its richness and potential to solve challenging problems. Data science provides powerful approaches for transforming large and complex data into information, knowledge, and intelligent decisions.

From IBM - *A data scientist represents an evolution from the business or data analyst role. The formal training is similar, with a strong foundation typically in computer science and applications, modeling, statistics, analytics and math. What sets the data scientist apart is strong application acumen, coupled with the ability to communication findings … in a way that can influence how an organization approaches a business challenge.*

A study by McKinsey Global Institute projects that by 2018, the United States will need to increase the number of graduates with skills handling large amounts of data by as much as 60 percent [1]. The United States will need an additional 140,000 to 190,000 data scientists with "deep analytical skills." The prime illustration of that is Cleveland-based KeyCorp, which has been experiencing an acute need for big-data skills. Amy Brady, the company's chief information officer, said "we need all levels of education - undergraduate, graduate and executive. Even high schools can play a role. I'd encourage all students in any field to have some kind of data-analytic coursework by the time they graduate" [2]. Brady

emphasized that "we're excited about IBM's focus on undergrad programs. We don't need all Ph.Ds. There's also a need for entry-level people that you can grow in your own organization — the undergraduate part is critical to the sustainability of this model" [2].

There is an increasing need for data-science professionals in a wide spectrum of disciplines. Healthcare is one example where the field of big data has a significant impact on transforming clinical practice and medical science through data analytics. McKinsey Global Institute estimates that U.S. healthcare could save more than $300 billion by using big data creatively and efficiently [1]. Even more important, it could significantly improve the quality of health care. The recently established partnership between John Carroll University and Cleveland Clinic's Automated Chemistry Lab highlights a strong demand in data science as a way to address critical clinical problems. The focus of this partnership is on engaging undergraduate students in data analytics. Through this partnership, undergraduate students gained experience in applications of analytics and visualization tools in the biomedical domain. The collaboration with Cleveland Clinic illustrates both the need for and meaningful engagement of undergraduate students in real data science applications.

In 2012, the White House launched a "Big Data Research and Development Initiative." One of the major objectives of the initiative is "to expand the workforce needed to develop and use Big Data technologies" [3]. Specifically, National Science Foundation provides support for training undergraduates to use graphical and visualization techniques for complex data.

In the October 2012 issue of Harvard Business Review, an article titled "Data Scientist: The Sexiest Job of the 21st Century" stated that "there are no university programs offering degrees in data science" [4]. Today, while hundreds of universities worldwide offer graduate programs in data science or data analytics, few offer undergraduate education in this field. Yet, it is critical that learning opportunities in data science are integrated into undergraduate curricula. The proposed B.S. in Data Science at John Carroll University directly addresses the national shortage in data science. John Carroll University will play an important role in the design of a new undergraduate program in data science and make Northeast Ohio one of the key regions in the country to produce entry-level data science skills.

The Department of Mathematics and Computer Science is already moving in this direction. In 2015-2016, three courses were piloted that expose undergraduate students to data science - Data Visualization, Biomedical Exploratory Data Analysis, and Big Data Analytics in Hadoop (Advanced Database Systems). The courses were offered in collaboration with corporate partners, Cleveland Clinic, top hospital in the United States, and SyngGlyphX, the company that offers the cutting edge big data visualization platform.

The idea of the new B.S. in Data Science has received strong support from a data science industry roundtable hosted by BioEnterprise this year. BioEnterprise is an organization that promotes the growth of biomedical, biotechnology and healthcare IT industries in Northeast Ohio. At the roundtable discussion, biomedical industry leaders presented their workforce needs in data science skills. Together, industry leaders and educators discussed the health IT talent pipeline for the biomedical industry, the importance of data science skills, and the importance of gaining real-world experience. John Carroll University's new program is designed to reflect the needs of industries in Northeast Ohio.

Strengthening the Academic Mission of John Carroll University

Can big data make the world a fairer place, or does it simply reinforce the advantages of the privileged few? Big data may bring greater safety, convenience and economic opportunity, yet it may also be used to produce inequality, discrimination and violation of privacy. John Carroll University's integrative curriculum, combined with the interdisciplinary data science curriculum, will encourage students to explore the complex relationships between data, technology and society. The B.S in Data Science addresses several objectives outlined in the university's strategic plan, including Goal 1 Objective 2 (distinctive programs), Objective 4 (experiential learning) and Objective 5 (student thriving).

Data Science for Social Good

Academic and corporate institutions increasingly use data science to solve critical problems in a variety of domains such as transportation, education, health care, community development, energy, and public safety. University of Chicago, University of Washington, and Georgia Tech have established summer fellowships under a program entitled Data Science for Social Good (DSSG) [5] [6] [7]. The fellowships bring data and domain scientists together to work on projects that impact public policy for social benefit. Bayes Impact, a nonprofit backed by the Bill & Melinda Gates Foundation and Y Combinator, uses data science to improve health and social services in order to tackle issues such as unemployment and preventative health care [8]. Online data science competitions in the form of "social good hackathons" are increasingly popular [9].

Strengthening Diversity and Inclusion in STEMM

Data science is an emerging field of STEMM (Science, Technology, Engineering, Mathematics, and Medicine). While the number of jobs and degrees granted in STEMM has consistently risen since 2000, the gender and racial gap has widened, particularly within the field of computing. According to a 2015 Taulbee Survey, fewer than 4% of current computer science degree recipients are Black or African American, and while 37% of U.S. computer science degrees went to women in 1983, today the number is 18% [10]. In contrast, over 40% of current statistics degree recipients are women [11]. Interest in statistics begins in high school, where half of students taking the Advanced Placement Exam in statistics are female. While the number of students taking the computer science AP exam increased 150% between 2010 and 2015, signaling a rapid rise of interest in the field, computer science lags behind other disciplines in terms of participation rates among women and traditionally underrepresented groups [12].

2015 Advanced Placement Exam	**All AP Exams**	Computer Science	Biology	Eng Lang & Composition	Statistics	US History
Female	**55%**	21.9%	60.3%	62.2%	52.2%	53.7%
American Indian	**0%**	0.4%	0.5%	0.6%	0.4%	0.6%
Asian	**15%**	29.2%	18.1%	11.5%	12.3%	17.8%
Black	**7%**	3.8%	7.0%	8.8%	6.0%	7.9%
Mexican American	**8%**	3.6%	6.6%	8.7%	5.9%	8.3%
Other Hispanic	**8%**	5.0%	6.6%	8.1%	5.8%	7.8%
Puerto Rican	**1%**	0.6%	0.9%	1.0%	0.8%	1.0%
White	**54%**	51.7%	54.1%	55.1%	58.4%	56.2%

An interdisciplinary data science degree program, combined with an integrative core, provides John Carroll University an opportunity to broaden participation in a high-growth field of STEMM while also developing data scientists educated in the Jesuit Catholic character.

Undergraduate Data Science Programs at Other Institutions

The rising demand from industry for "statisticians that know computing" and "computer scientists that know statistics" has resulted in the recent development of at least 36 programs within the United States that offer an undergraduate degree in Data Science, Data Analytics or Business Analytics. Although the specific degree title may vary, approximately half of the programs contain a sufficient combination of computational, statistical and mathematical coursework to be classified as data science [13].

Ohio State University offers a Bachelor of Science in Data Analytics jointly managed by the Department of Statistics and the Department of Computer Science and Engineering. The intercurricular program is based on partnerships between the College of Engineering, the College of Medicine, and the Fisher College of Business.

> *"The student response has been overwhelming: More than 80 students had selected data analytics as their major plan by the end of the first year of the program."*
> –Dr. Christopher Hans and Dr. Srinivasan Parthasarathy, Ohio State University [14].

Miami University offers an analytics co-major jointly supported by the Department of Statistics in the College of Arts and Sciences and the Department of Information Systems and Analytics in the Farmer School of Business.

> *"Student reaction has been very positive. We have gone from zero to 65 co-majors in less than two years and have already had 11 students graduate with the co-major. In addition, our analytics co-majors are highly recruited, with nearly all having multiple job offers by the fall of their senior year."* – Dr. John Bailer and Dr. Allison Jones-Farmers, Miami University [14].

Northern Kentucky University began to offer a Bachelor of Science in Data Science in 2013.

> *"Response has been enthusiastic, both by students, with an enrollment of 23 by the end of the second year since launch, and by area employers, who are eager to hire students from this rare undergraduate program in data science."*
> – Dr. Mark Lancaster and Dr. James McGuffee, Northern Kentucky University [14].

Auburn University started its undergraduate degree in Business Analytics in 2015. The program requires several courses involving computation. Over 50 students enrolled in the new major by the second year.

> *"Currently we have about 50 students in the major. … When the program initially started, there were only a few students. Quantitative programs in Business Schools are not initially popular. However, as the word spread on the demand in industry for graduates with such skills, students are now wanting to major in the program."*
> -Dr. Amit Mitra, Department of Systems and Technology, Auburn University [15].

Case Western Reserve University launched a minor in Applied Data Science in the Fall of 2014 and is expected to launch a major in 2016 [16].

The University of San Francisco offers a Bachelor of Science in Data Science. The degree program is jointly managed through the Department of Mathematics and the Department of Computer Science [17].

University of Michigan offers a Data Science undergraduate degree that is a joint program between the EECS Department in the College of Engineering and the Department of Statistics [18].

Denison University is launching an interdisciplinary data analytics major in 2016. The major includes a core of mathematics and computer science courses, several data analytics courses, along with a set of courses in a variety of application domains [19].

2. Program Curriculum – Major in Data Science

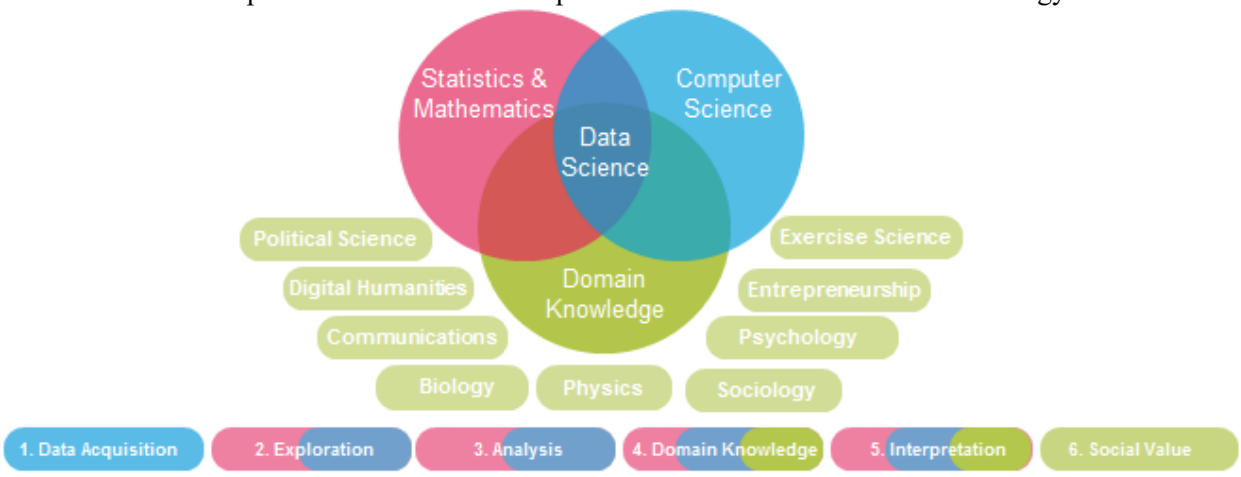
The American Statistical Association's *Curriculum Guidelines for Undergraduate Programs in Statistical Sciences* emphasizes the following key points [20]:

- **Increased importance of data science.** Working with data requires extensive computing skills. To be prepared for statistics and data science careers, students need facility with professional statistical analysis software, the ability to access and manipulate data in various ways, and the ability to perform algorithmic problem solving. In addition to more traditional mathematical and statistical skills, students should be fluent in higher-level programming languages and facile with database systems.
- **Real applications.** Data should be a major component of statistics courses. Programs should emphasize concepts and approaches for working with complex data and provide experiences in designing studies and analyzing non-textbook data.
- **More diverse models and approaches.** Students require exposure to and practice with a variety of predictive and explanatory models in addition to methods for model-building and assessment. They must be able to understand issues of design, confounding, and bias. They need to know how to apply their knowledge of theoretical foundations to the sound analysis of data.
- **Ability to communicate.** Students need to be able to communicate complex statistical methods in basic terms to managers and other audiences and to visualize results in an accessible manner. They must have a clear understanding of ethical standards. Programs should provide multiple opportunities to practice and refine these statistical practice skills.

The guidelines serve as a foundation for John Carroll University's Data Science curriculum and program learning goals.

Program Learning Goals (PLG)

1. **Data Acquisition**: collect, store, preserve, manage and share data in a distributed environment through practical, hands-on experience with programming languages and big data tools;
2. **Problem Exploration**: develop problem solving skills through experiences that foster computational and data-analytic thinking;
3. **Analysis**: develop an in-depth understanding of the key technologies in data science: data mining, machine learning, visualization techniques, predictive modeling, and statistics;
4. **Domain knowledge**: experience discipline-specific data use cases in order to solve real-world problems of high complexity;
5. **Interpretation**: learn methods for effective data communication and visualization, and demonstrate their use in data representation;
6. **Social Value**: explore social and ethical implications of the use of data and technology.



Program Requirements

Course	Course Description	Credit	Status	Prerequisites	Timetable	Notes
[1]DATA100	Introduction to Data Science	3	**New**		Spring 2017. Every semester	QA
MT122	Elementary Statistics or equivalent	3				QA
CS128/ CS128L	Introduction to Software Application Development and LAB	4				
CS150	Database Systems	3				
DATA200	Intermediate Data Science	3	**New**	DATA100, CS128, MT122 or equivalent Co/Pre: CS150	Spring 2018. Annual	
MT223	Intermediate Statistics	3		MT122 or equivalent		
MT288	Network Analysis	3	**New**		Fall 2017. Annual	
EN300	Technical Writing	3				
DATA470	Data Science Capstone Project	3	**New**	Senior standing, DATA200, MT223, MT288	Fall 2018. Annual	
	Domain Knowledge **-** 4 courses	12+				
	Electives **-** 2 courses	6				
	Total Program Requirements	46+				

Data Science Core (28 credits):

The data science core is designed to develop computational, statistical and mathematical thinking, as well as knowledge of the foundations of data science and communication skills. The capstone project allows students to demonstrate the attainment of an advanced level of the program learning goals.

Computational Thinking:
- CS128 –Intro to Software App Development
- CS150 – Database System

Statistical and Mathematical Thinking:
- MT122 – Elementary Statistics
- MT223 – Intermediate Statistics
- MT288 – Network Analysis

Foundations of Data Science
- DATA100 – Introduction to Data Science
- DATA200 – Intermediate Data Science

Communication Skills
- EN300 – Technical Writing
- Data visualization in DATA100, DATA200

Applied Data Science
- DATA470 – Data Science Capstone

[1] A new course code DATA is proposed to indicate a data science course. CS150, CS307, CS322, CS340, CS350, MT122, MT223, MT228, MT229, MT422, and MT424 will be cross-listed with the DATA course code. MT223 will be renumbered as DATA260.

Domain Knowledge – Areas of Specialization (12+ credits):

The majority of data science programs include courses in statistics, computing, and business. Recent proposals for new data science programs have begun to integrate courses in other disciplines as well. Ohio State University requires their data analytics majors to take 4 elective courses from one area of specialization, choosing courses from biomedical informatics, business analytics, computational analytics, or social science analytics. Data analytics majors at Denison University choose 3 electives from anthropology & sociology, biology, economics, physics, political science, or psychology.

The B.S. in Data Science at John Carroll University is designed to expose students to broad issues and applications beyond the data science core. The areas of specialization provide discipline-specific experiences that enhance student understanding of data acquisition and management, data analysis, linguistics and natural language processing, data modeling, data visualization and communication, as well as ethical and social considerations of data. Students select 4 courses from one area of specialization.

4 courses selected from one area of specialization. Prerequisites are included in the course count.		
Communications • CO201 Communications Research • CO225 Journalism • CO325 Investigative Reporting • CO346 Campaign Issues • CO315 Integrated Marketing Communications • CO360 IMC Research • CO455 Health & Environ. Writing	**Digital Humanities** • EN2xx Literature • EN311 Old English • EN312 Late Medieval Literature • EN488 History of the English Lang • EN498 Independent study	**Entrepreneurship** • ER201 Creativity, Innovation and Idea Development • ER301 Intro to Entrepreneurship • ER304 Social Entrepreneurship • ER305 Accounting & Finance for Entrepreneurs • ER306 Entrepreneurial Marketing and Sales • ER480 Entrepreneurship Field Experience
Exercise Science • EPA205/205L Human Anatomy • EPA206/206L Human Physiology • EPA230 Nutrition for Athletics and Physical Activity • EPA407 Exercise Physiology • EPA409 Kinesiology • EPA432 Motor Learning • EPA440 Independent Study	**Health Disparities&Social Justice** • BL155/157 Principles of Biology • BL156/158 Principles of Biology II • BL240 Epidemiology • BL260 Poverty & Disease • BL399 Special Problems Biology • PO160 Health Care and Social Justice in Latin America, and PO2xx Health Care Access in Latin America	**Physics** • PH135,135L Physics I • MT135 Calculus and Analytic Geometry I • PH136, 136L Physics II • MT136 Calculus and Analytic Geometry II
Political Science • PO200 Introduction to Methods • PO203 GIS I • PO300/L Research Methods & Lab • PO319 U.S. Elections • PO324 Crisis Mapping, New Media and Politics • PO337 Comparative Health Politics • PO399/498 Independent Study	**Psychology** • PS100 Introduction to Psychological Science • PS301/301L Experimental Design & Analysis • PS401/401L Research Methods • PS435 Tests & Measurements • PS499 Individual Research	**Sociology & Criminology** • SC101 Intro to Sociology • SC350 Sociological Research Methods I • SC351 Sociological Research Methods II • SC493 Independent Study • SC497 Undergraduate Research

Electives (6 credits):

The electives deepen the understanding of skills and methods used in data science. Students will choose 2 courses from the following list:

- CS/DATA307 BioInformatics
- CS/DATA322 Big Data Analytics
- CS/DATA340 Data Visualization
- CS/DATA350 Advanced Database
- MT/DATA421 Mathematical Statistics
- MT/DATA422 Applied Statistics
- MT/DATA424 Applied Regression

Sample 4 Year Plan

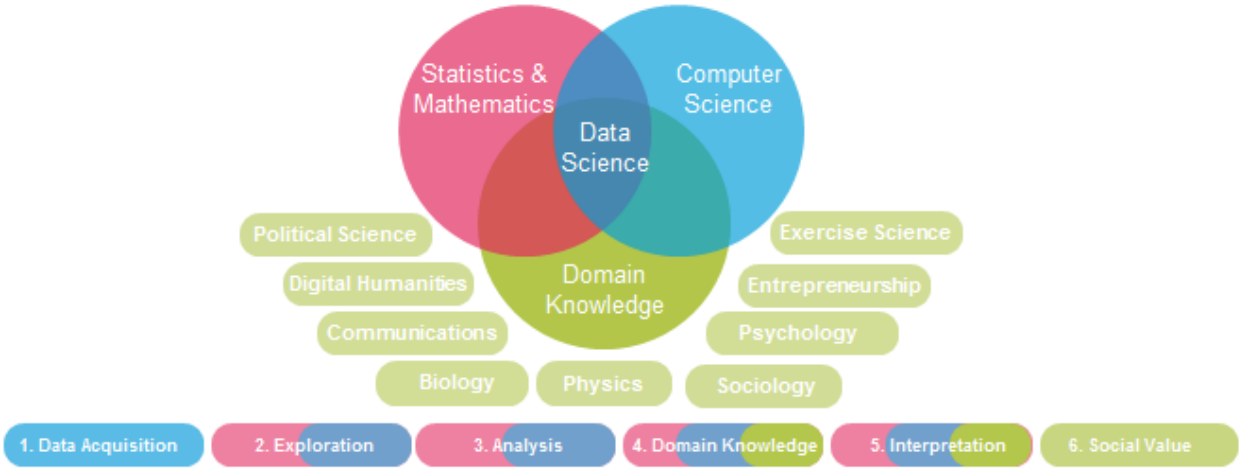
The major is designed for undergraduate students to complete in four years; however, the prerequisite structure makes it is possible to complete in three years as well.

Sample 4 Year Plan	
Fall	**Spring**
DATA100	MT122, CS128
CS150, MT288	DATA200, MT223
EN300, Elective or Domain	2 Electives or Domain
DATA470, Elective or Domain	2 Electives or Domain

First and Second Year
The first year provides the background knowledge and fundamental skills required to develop expertise in data science. Students take classes in programming and database design, along with topics in statistics and mathematics such as exploratory data analysis, probability fundamentals and network analysis. Students encounter programming in Processing (Java-based), Python, and R. In the second year, courses integrate the skills gained in the first year, enabling students to build a portfolio of data science skills including data mining, machine learning, and statistical inference.

Third and Fourth Year
Students focus on applying the skills gained in a variety of problem domains, including individual and group projects that provide practical experience of data science. Students will enhance their data science skills by acquiring domain knowledge in one of several disciplines through courses that include data analysis, modeling, or application. Communication and team working skills are developed in a variety of domains allowing them to operate effectively as data scientists. Students are encouraged to complete an internship in data science (at least one summer between sophomore and junior year). The degree culminates through a capstone project experience that applies coursework to open-ended data science problems.

Program Curriculum Map



Map of Program Learning Goals (PLGs)						
Course	PLG 1	PLG 2	PLG 3	PLG 4	PLG 5	PLG 6
DATA100	Basic	Basic	Basic	Basic	Basic	Basic
CS/DATA128	Basic	Basic				
CS/DATA150	Basic	Basic	Basic	Basic	Basic	
MT/DATA122	Basic	Basic	Basic	Basic	Basic	Basic
DATA200	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate	Basic
MT223/DATA260	Basic	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate
MT288	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate
EN300		Intermediate			Advanced	Intermediate
DOMAIN KNOWLEDGE	Basic	Intermediate	Basic	Advanced	Intermediate	Advanced
ELECTIVES	Advanced	Advanced	Advanced		Advanced	
DATA470	Advanced	Advanced	Advanced	Advanced	Advanced	Advanced

Relationship between Proposed Program and Existing Programs

The Department of Mathematics and Computer Science at John Carroll University presently supports four undergraduate degree programs: (1) B.S. in Mathematics, (2) B.A. in Teaching Mathematics, (3) B.S. in Computer Science, and (4) B.S. in Healthcare Information Technology.

The B.S. in Data Science provides an interdisciplinary curriculum with distinct learning goals and outcomes. In terms of required courses, the B.S. in Data Science, the B.S. in Computer Science, and the B.S. in Healthcare Information Technology share 2 courses:

- CS128,CS128L – Introduction to Software Application Development and Lab
- CS150 – Database Systems

The Data Science major requires a minimum of 46 credits of coursework, which is similar to the Computer Science major. The Computer Science curriculum is highly focused on building computing skills, with all but one course having a CS designation. The Data Science curriculum in contrast represents a blend of computing, statistics, and applied experiences from other disciplines. Students will encounter an early example of this integrative nature in DATA200, which builds upon the knowledge and skills acquired in its three prerequisites and one co-requisite. As a result of the interdisciplinary nature, along with the four course requirement from an area of specialization, the data science major requires fewer upper-level computing courses than the computer science major. This is consistent with the curriculum of other interdisciplinary data science programs at liberal arts institutions, such as the data analytics major at Denison University.

New Courses to be developed

The curriculum integrates existing computer science, statistics, and mathematics courses, requiring the creation of just four new courses:

- DATA100 – Introduction to Data Science
- DATA200 – Intermediate Data Science
- MT288 – Network Analysis
- DATA470 – Data Science Capstone Project

Detailed syllabi for the new courses are available in the appendix.

DATA100 INTRODUCTION TO DATA SCIENCE (3 credits)
Data science capitalizes on big data and focuses on data analytics that turn information into actionable knowledge. This course will introduce students to the key ideas, practices, and challenges of modern data analysis. Students will get an overview of the data, questions, and tools that data scientists deal with in their practice. This course will introduce students to practical approaches to essential exploratory techniques, interactive data discovery, and predictive analytics including basic techniques for collecting, cleaning, and sharing data. Hands-on activities will enable students to learn the practical toolkit of a data scientist.

DATA200 INTERMEDIATE DATA SCIENCE (3 credits)
Prerequisite: DATA100, CS128, MT/DATA122 or equivalent. Pre or Co-requisite: CS/DATA150.
This course will provide a strong foundation in the field of data science and data analytics with a focus on computational approaches and experiential learning. Students will learn about processes and practice of data science that are developed to analyze diverse sources of data including data modeling, machine learning, and natural language processing. This course will present the fundamentals of inference in a practical approach. Students will build a portfolio of Big Data skills.

DATA470 DATA SCIENCE PROJECT (3 credits)
Prerequisite: senior standing in the Data Science major, DATA200, DATA260, MT288. Simulation of the environment of the professional data scientist working in a team on a large data project for a real client. Students will encounter a wide variety of issues that naturally occur in a project of scale, using their skills, ingenuity, and research abilities to address all issues and deliver a usable data product.

MT288 NETWORK ANALYSIS (3 credits)
Introduction to the mathematical and computational theory of network analysis. Fundamentals of graph theory and matrix algebra. Measures of networks including concepts of centrality and structural balance. Study of large-scale structure of networks. Computational representation of and algorithms on networks. Independent student project.

Approval Process for New Courses

New courses will go through a standard process of review and approval from the Department of Mathematics and Computer Science, in conjunction with consultation among faculty from the domain knowledge areas of specialization.

Faculty Expertise and Availability

One of the primary goals in the development of this program was for it to be attainable with limited resources. With the hiring in 2015 of Dr. Elena Manilich and Dr. Billie Marget, the Department of Mathematics and Computer Science has faculty with superior knowledge and expertise to develop and support the Data Science degree program. Additional faculty members who have experience in teaching topics in data science include Drs. Brendan Foreman, Marc Kirschenbaum, Daniel Palmer, and Linda Seiter. This past year the Department of Mathematics and Computer Science voted to terminate the Bachelor of Science in Computer Information Systems (CIS) degree program. As a result of hiring new faculty and the cessation of the CIS degree, it will be possible to support the new program with existing faculty resources.

3. Program Curriculum - Minor in Data Science

In addition to the B.S. in Data Science, the Department of Mathematics and Computer Science proposes a new minor in Data Science. Students who successfully complete the minor will learn basic principles and build skills in the science of how we use data to solve critical problems. The Data Science minor will enhance the credentials for students in a wide variety of disciplines. A recent entry-level job market analysis shows liberal arts graduates with data-analysis skills have access to 137,000 more jobs and an average increase in salary of $12,700 over those without such skills. Computer programming skills provide liberal arts graduates access to an additional 53,000 jobs and $18,000 average salary increase [21]. The Data Science minor provides both data-analytic and programming skills. It is hoped that John Carroll University students will encounter DATA100 or MT122 while fulfilling the quantitative analysis core requirement, and subsequently choose to continue with other courses in the minor.

Minor in Data Science - Program Requirements			
Course	**Course Description**	**Credit**	**Notes**
	Program Core		
DATA100	Introduction to Data Science	3	QA
MT/DATA122	Elementary Statistics or equivalent	3	QA
CS128/128L	Intro to Software App Devl and LAB	4	
CS/DATA150	Database Systems	3	
DATA200	Intermediate Data Science	3	
2 Electives	DATA260, 307, 322, 340, 350	6	
	Total Program Requirements	22	

Program Learning Goals

The data science minor builds a basic portfolio of data science skills. The goals are similar to the major, although the level of mastery is not as advanced. DATA200 is a project-based course that integrates the computational and data-analytic skills acquired in the 100-level prerequisites, thus providing a mini-capstone experience. After completing DATA200, students will have a foundation of data science skills, achieving the following program goals at a basic/intermediate level:

1. **Data Acquisition**: collect, store, preserve, manage and share data in a distributed environment through practical, hands-on experience with programming languages and big data tools;
2. **Problem Exploration**: develop problem solving skills through experiences that foster computational and data-analytic thinking;
3. **Analysis**: develop a basic understanding of the key technologies in data science: data mining, machine learning, visualization techniques, predictive modeling, and statistics;
4. **Domain knowledge**: experience discipline-specific data use cases in order to solve real-world problems of high complexity;
5. **Interpretation**: learn methods for effective data communication and visualization, and demonstrate their use in data representation;
6. **Social Value**: explore social and ethical implications of the use of data and technology.

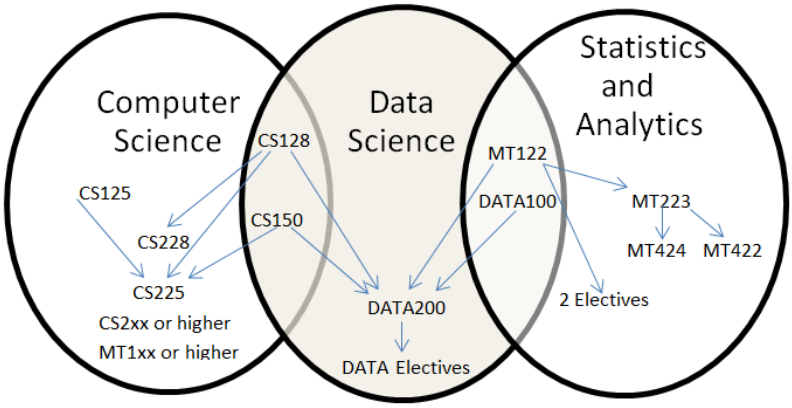
Program Curriculum Map

Map of Program Learning Goals (PLGs)						
Course	PLG 1	PLG 2	PLG 3	PLG 4	PLG 5	PLG 6
DATA100	Basic	Basic	Basic	Basic	Basic	Basic
CS/DATA128	Basic	Basic				
CS/DATA150	Basic	Basic	Basic	Basic	Basic	
MT/DATA122		Basic	Basic	Basic	Basic	Basic
DATA200	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate	Basic
2 Electives	Intermediate	Intermediate	Intermediate		Intermediate	

Relationship between Proposed Minor and Existing Minors

The Department of Mathematics and Computer Science presently offers three separate minors: (1) Mathematics, (2) Statistics, and (3) Computer Science. The name of the Statistics minor is evolving to "Statistics and Analytics."

The curriculum for the Mathematics minor is distinct from Data Science. The figure below shows the program requirements and prerequisite structure among the Computer Science, Data Science, and Statistics and Analytics minors.



Required courses for Computer Science, Data Science, and Statistics and Analytics minors

The Data Science minor builds computational and database skills in CS128 and CS150, both of which are required in the Computer Science minor. Data Science students acquire statistical analysis skills in MT122 (or restricted list of equivalent courses), prerequisite to the required course MT223 in the Statistics and Analytics minor. The Statistics and Analytics minor is presently evolving to include DATA100 in order to ensure students obtain basic statistical programming and data visualization techniques.

4. Organization and Administration of Program

The Data Science major will be housed within the Mathematics and Computer Science department and administered as such. In that context a separate director will not be necessary. An oversight committee consisting of faculty members approved by the department chair will perform duties for the major such as (1) coordinate with the chair of the department and other data-centric programs within the university, (2) approve the development and inclusion of new courses, (3) interface with industry partners, and (4) administer internships for the program.

Recommended Line of Reporting and Structure of Governance

The Data Science oversight committee will consist of at least one Computer Science faculty member, at least one Statistics faculty member, and at least one Mathematics faculty member. An invitation to participate on the oversight committee will also be extended to a faculty member in each domain knowledge area of specialization. Chairs from the relevant departments will recommend appointees. One of the oversight committee members will be elected to chair the committee and report to the Mathematics and Computer Science department chair.

5. Implementation Timetable

- DATA100 – Introduction to Data Science. Spring 2017
- MT288 – Network Analysis. Fall 2018
- DATA200 – Intermediate Data Science. Spring 2018
- DATA470 – Data Science Capstone. Fall 2019

6. Assessment

Achievement of the program learning goals will be assessed with a blend of capstone experience and portfolio-based program assessment. Each portfolio will have four specific items:

- Baseline Student Experience and Skills Assessment.
  In the gateway courses (DATA100, MT122, CS150), students will fill out a survey/skill assessment tool to establish a baseline of their statistical, computational, and data modeling ability and experience, along with their ability for logical reasoning and abstract thinking.
- Data Science Primitives Assessment.
  At the beginning and end of DATA200, all students will complete a statistical analysis and programming skills assessment.
- Capstone Student Assessment.
  At the end of DATA470 (or afterward) all students will have a technical exit interview with a member of the faculty. At this interview, the materials in the portfolio and the student's experiences in the program will be discussed. A report will be placed in the portfolio.

Additional items to be collected in the portfolio for all majors:
DATA100 final project/exam
DATA200 final project/exam
DATA470 capstone project

Program Evaluation plan – indicators of program success

This program will be evaluated based on the number of students entering John Carroll with the intent to major in Data Science, the size and diversity of the graduating class, course evaluations, exit interviews at graduation, and alumni career tracking. Faculty will also meet and consult with industry partners to annually evaluate the program, in particular the quality and preparedness of the student interns and alumni. In the long-term, an advisory board of data science professionals in the region will be established to provide insight and input to the program.

7. Budget

Faculty Resources

The Data Science major will be housed within the Mathematics and Computer Science department. The hiring in 2015 of new faculty with expertise in Data Science, along with the cessation of the CIS degree, makes it possible to support the new program with existing faculty resources.

Enrollment Estimation

Interest in data science among existing John Carroll University students is strong. Enrollment in the spring 2016 semester of *CS480–Data Visualization* exceeded the computer classroom capacity – 29 students. The fall 2016 semester of *CS322–Big Data* had 25 students enrolled.

It is difficult to estimate the potential enrollment in the data science major. Brian Williams, Vice President for Enrollment, explained the traditional method of predicting enrollment involves looking at responses of high school juniors and seniors who have taken the ACT and have expressed interest in a particular field. Given that Data Science is a new field, it is not yet available as an option on the ACT. An additional method for prediction compares current program enrollment in other universities in John Carroll University's comparator set. However at this time, there is no standard Federal classification code (CIP) for Big Data and Analytics degrees or certificate programs. There is no definitive way to see program enrollments or number of graduates generated at other universities in data science programs at the present time.

In the external scan of programs on specific university web sites, programs may be affiliated with the departments of computer science, statistics, engineering, or business. Program names may reflect different specializations (information management, business analytics, statistics and operational research, etc.) and concentration areas (marketing, insurance, financial services). Of the data science/analytics programs identified at other universities, student response has been strong (Ohio State University – 80 students in year 2, Auburn University – 50 students in year 2, Miami University – 65 students in year 2, Northern Kentucky University – 23 students in year 2).

There are still relatively few programs at the undergraduate level, and they vary between "data analytics" and "data science". Some undergraduate programs are being offered online in conjunction with computer science or business. Most programs are at the Master's level (at larger research universities). Again, there is no defined way to get enrollment or high school trend explicitly because it's so new. In consultation with Brian Williams, and in seeing the job demand projections in this field, we believe that this is a new trend that will not go away. Being earlier to adopt and adapt, and be known in this area rather than waiting, is the enrollment imperative that we face.

In lieu of any high school or university trend data, data related to job demand is useful in support of this program. Consider the example of the McKinsey report referenced in the introduction of this proposal which indicates the need for an additional 140,000 to 190,000 data scientists in the United States by 2018 [1]. Another recent report by Chmura Economics & Analytics further indicates a large gap for data science talent in the Potomac Region, with bachelor's students meeting only 44% of local business demand for computer scientists, 7% of demand in data modeling and administration, 14% demand for management information systems, 71-75% demand in mathematics and statistics [22].

Faculty Salary for Summer Course Development

In an effort to inform high school students of Data Science as a potential career, the Department of Mathematics and Computer Science will collaborate with the Greater Cleveland Council of Teachers of Mathematics (GCCTM) and the Ohio Council of Teachers of Mathematics (OCTM) to develop an online data science course designed for high school instructors who teach AP statistics or AP computer science, as well as those who teach the non-AP courses. The course will provide graduate non-degree seeking credit. The budget includes funds to support a faculty member to develop and teach the online course. The anticipated first offering of the course is summer 2017, with a subsequent offering in summer 2018. The course will be marketed to members of the GCCTM and OCTM.

The online summer course material will subsequently be evolved to create a series of online data science workshops intended primarily for high school teachers, although potentially open to anyone wishing to learn data science. Each workshop will require a registration fee to access the online material. It is expected that such an online workshop series will open up the audience of high school teachers to a greater geographical area - from northeast Ohio to the whole country, thus expanding the potential exposure to a large number of high school students.

Beyond the online course and workshops, the data science major will be marketed to high school teachers and their students through frequent contact with the members of the GCCTM and OCTM. The department of mathematics faculty will collaborate with Jerry Moreno, Assistant Professor Emeritus of Statistics, who presently serves as president of GCCTM and is an active member of the OCTM.

Travel – Data Science Education Conferences

The budget includes travel funds to support faculty attendance at data science education conferences.

Software Requirements – Cloud Storage

The software packages used in data science courses are open source and free, including:

- Hadoop Distributed Computing Platform
- Spark, Pig, Hive, Apache Mahout
- MySQL

Cloud storage is expected to be approximately $100 per semester. Students will purchase cloud storage directly. Courses that require cloud storage will indicate the cost in the course details displayed on Banner, thus allowing students to be aware of the cost prior to registration. The majority of required courses listed in the Data Science curriculum do not require cloud storage. The budget includes an estimated annual cost of $500/instructor to allow instructors to test a variety of cloud storage options. The estimated cost covers one instructor in year 1 and 2, two instructors in years 3 and 4.

Budget Summary		Yr 1 1718	Changes	Yr 2 1819	Changes	Yr 3 1920	Changes	Yr 4 2021
Net tuition revenue	$	**6,130**	8,070	$ **14,200**	11,800	$ **26,000**	14,180	$ **40,180**
Costs		**10,800**	-	**10,800**	(7,800)	**3,000**	-	**3,000**
Contribution Margin		**(4,670)**	8,070	**3,400**	19,600	**23,000**	14,180	**37,180**
Math & CS Support		**10,000**	5,000	**15,000**	-	**15,000**	-	**15,000**
Net Contribution Margin	$	**5,330**	13,070	$ **18,400**	19,600	$ **38,000**	14,180	$ **52,180**
Net Tuition Revenue - Details								
Cohort year 1		4	-	4	-	4	-	4
Attrition		-	-	(1)	-	(1)	-	(1)
Total across campus		4	(1)	3	-	3	-	3
Cohort year 2		-	-	6	-	6	-	6
Attrition		-	-	-	(1)	(1)	-	(1)
		-	6	6	(1)	5	-	5
Cohort year 3		-	-	-	8	8	-	8
Attrition		-	-	-	-	-	(1)	(1)
		-	-	-	8	8	(1)	7
Cohort year 4		-	-	-	-	-	10	10
Attrition		-	-	-	-	-	(1)	(1)
		-	-	-	-	-	9	9
Total number of students		4	5	9	7	16	8	24
Courses and credit hours to be taken:								
DATA100 - Spring 2017		3		3		3		3
DATA200 - Spring 2018		-		3		3		3
MT288 - Fall 2018		-		-		3		3
DATA470 - Fall 2019		-						3
Cr Hr for each student		3		6		9		12
Credit Hours Taught		**12**	15	**27**	21	**48**	24	**72**
Net tuition rate		1,277	38	1,315	39	1,354	41	1,395
Discount Rate		(766)	(23)	(789)	(23)	(812)	(25)	(837)
Net tuition per Cr Hr		511	15	526	16	542	16	558
			-		-		-	
Net tuition revenue	$	**6,130**	8,070	$ **14,200**	11,800	$ **26,000**	14,180	$ **40,180**

Costs Details -							
Payroll Fringe Rate -							
Faculty fringe rate	*0.375*		*0.375*		*0.375*		*0.37*
Staff PT fringe rate	*0.08*		*0.08*		*0.08*		*0.08*
Dir Payroll	-	-		-		-	
Fringes	-	-	-	-	-	-	-
Faculty Payroll	6,000	-	6,000	-	-	-	-
Fringes	2,300	-	2,300	(2,300)	-	-	-
Staff PT Payroll	-	-	-	-	-	-	-
Fringes	-	-	-	-	-	-	-
Other	-	-	-	-	-	-	-
Payroll and Fringes	8,300	-	8,300	(2,300)	-	-	-
Operating	2,000	-	2,000	-	2,000	-	2,000
Student Help	-	-	-	-	-	-	-
Capital	-	-	-	-	-	-	-
Software	500	-	500	500	1,000	-	1,000
Rounding		-		-		-	
Operating	2,500	-	2,500	500	3,000	-	3,000
Total Costs	**10,800**	**-**	**10,800**	**(1,800)**	**3,000**	**-**	**3,000**
Math & CS Support Details -	Support committed by the Chair of Mathematics and Computer Science Department.						

8. **Letters**

- Dr. Todd Bruce, Director of Academic Assessment
- Dr. Graciela Lacueva, Associate Dean of Science, Mathematics and Health
  Dr. Margaret Farrar, Dean, College of Arts and Sciences
- Dr. Dennis Hareza, University Budget Advisory Committee
- Dr. Brian Williams, Vice President for Enrollment & Institutional Analytics
- Dr. David Wong, Executive Director for Academic Finance

Department Chairs

- Dr. Paul Shick, Department of Mathematics and Computer Science
- Dr. Phyllis “Penny” Braudy Harris, Department of Sociology & Criminology
- Dr. Rebecca Drenovsky, Department of Biology
- Dr. Jeffrey Dyck, Department of Physics
- Dr. Margaret Finucane, Tim Russert Department of Communication & Theatre Arts
- Dr. Kathleen Manning, Program in Exercise Science, Physical Education, and Sports Studies
- Dr. Mindy Peden, Department of Political Science
- Dr. Debby Rosenthal, Department of English
- Dr. Jackie Schmidt, Entrepreneurship Program
- Dr. Sheri Young, Department of Psychology

John Carroll UNIVERSITY

Office of Academic Assessment
1 John Carroll Blvd.
University Heights, Ohio 44118-4581
Phone 216.397.1600
www.jcu.edu

September 12, 2016

Dr. Barbara D'Ambrosia,
Chair, Faculty Council

Dr. Peifang Tan,
Chair, Committee on Academic Policies
John Carroll University

Dear Dr. D'Ambrosia and Dr. Tan:

I am writing to indicate my support for the proposed Bachelor of Science in Data Science. I encourage the members of the Committee on Academic Policies to recommend that the proposed new program be approved by the faculty and made part of the curriculum.

I have been asked to provide information regarding the relationship of new program to the overall curriculum, the viability of its assessment plan, and the availability of resources to support assessment of the program.

The proposed major and minor are congruent with our existing academic structure, being comparable in size and structure to other majors and minors at the institution. It also fits nicely with the new integrative core curriculum, with the potential to offer QA courses, as well as ENW linked pairs. *The learning goals and course of study show a strong alignment with many of our institutional academic learning goals*. This program obviously has strengths in critical analysis, but the use of data science techniques within other content areas provides a new model for integrative knowledge and a venue for exploring social justice issues.

This proposal provides a solid framework for a viable assessment plan. It describes program-level goals that are specific, measurable, easily aligned with the institutional academic learning goals, and focused on student learning. The proposal also indicates direct and indirect measures of student learning that are closely matched to the goals. Data obtained from these measures should enable the faculty and chair to make curricular changes at appropriate intervals to best ensure student learning.

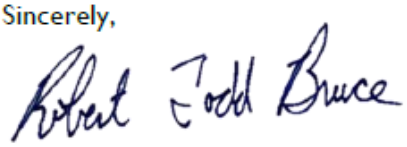
Assessment of student learning is a routine part of faculty work and every academic program at John Carroll University already is involved in assessment. This program will be no different, and its assessment plan shares many commonalities with existing plans within the department. As with all departments and interdisciplinary programs, the department's existing assessment

coordinator will coordinating assessment of this program with my office. *There is nothing in this proposal to suggest that assessment of this program will require additional resources beyond those already provided by the institution in general or my office, in particular.*

If the program is approved, I look forward to the opportunity to work with the math department to further develop their formal assessment plan, curriculum map, and assessment instruments and procedures if the proposal is accepted.

I fully support the approval of this new program. Please contact me at x1600 or rbruce@jcu.edu if there is anything else that I can provide to inform your deliberations.

Sincerely,



Robert Todd Bruce
Director of Academic Assessment

John Carroll UNIVERSITY

1 John Carroll Boulevard
University Heights, Ohio 44118-4581
www.jcu.edu

September 6, 2016

Dr. P. Tian
Chair of CAP

Dear Dr. Tian,

This is an evaluation letter for the Bachelors of Science in Data Science proposal.

The group preparing this proposal has invested considerable time and effort in compiling data, consulting across the university, and presenting a well- researched proposal. In writing this evaluation letter we want to make clear our full support for the development of this new major. Commenting on both the benefits of the program and the concerns about its implementation will, we hope, assist CAP and the MT/CS department.

Favorable points:
1-Fit with the University's Strategic Plan: the proposed program addresses Goal 1 (objectives 2, 4, and 5) and Goal 3 (objective 6).
2- Potential for contributions to the Core: the interdisciplinary nature of this field should provide opportunities for ENW courses.
3-The program addresses the clear need to develop data scientists. The needs are both regional and national.
4-The proposal presents opportunities for collaboration with both high-schools and industry.
5-The department has faculty qualified to teach in this program.
6-There is no need of additional resources to start this program.
7-The new courses will be delivered using engaging pedagogy, and there will be valuable experiences of professional development for the students.
8-The administrative structure proposed for the program will help maintain interdisciplinary connections and assist students pursuing a variety of interests. The proposed external board will be very important in maintaining currency as well as in sustaining opportunities of internships for our students.

9- Dr. Seiter engaged in extensive collaboration in the preparation of this proposal, consulting with the chair of each of the departments whose specializations are listed here, as well as several others not listed but who see future potential in connecting to a Data Science major.

Concerns:

1-Resources:

a) The success of the program relies heavily on the expertise of two faculty members, making it somewhat unstable. We suggest that the department investigates the availability of area professionals willing to offer courses on a part-time basis, and cultivate relations with such professionals. Having such resources will prevent a crisis in case of full-time faculty becoming unavailable, and will allow for a quick fix in case of a surge in demand that necessitates multiple sections of courses.

b) The software costs are unpredictable. The proposal states that currently we have free passes for 22 students and one faculty; it is unclear if those privileges are used only while students are in class. If not, it appears that as the program grows, more students and faculty will need to acquire access. We suggest that for classes requiring cloud computing the department add a lab fee to cover costs.

c) Computer classroom availability. It is not clear from the proposal if the department has evaluated the availability of computer classrooms to offer the data courses. While this may not be an issue initially, it can become a serious issue if the program grows. It was suggested that perhaps students can be required to bring their own laptops; for that solution to work the department should think about contingency plans for students who cannot afford to buy their own.

2- We all know that the fact that there is a great need for scientists and engineers does not automatically translate in interest on the part of students to pursue such fields. The proposal clearly indicates that *current* CS students are very interested in Data Science. One may worry that Data Science might grow at the expense of other CS programs. Enrollment trends in all three programs should be monitored to determine viability.

3-Concerns about the syllabi included:

i. some of the syllabi fail to link course goals to the University Learning Goals
ii. some of the syllabi do not indicate how the learning goals will be assessed
iii. syllabus for the capstone fails to mention Core requirements
iv. the course description for MT228 states: "Introduction to the mathematical and computational theory of network analysis **with applications in globalization studies and other social sciences**"; however, nothing in the course outline indicates time devoted specifically to such applications. We suggest this be made more explicit (or the reference in the description is eliminated).

We look forward to working with CAP in the study of this proposal.

Sincerely,

Dr. Graciela Lacueva
Associate Dean of Science, Mathematics and Health

Dr. Margaret Farrar
Dean, College of Arts and Sciences

Response from the Department of Mathematics and Computer Science to address concerns listed in letter of support from the Dean's Office, College of Arts and Sciences

1-Resources:
a) Page 12 has been updated to indicate the availability of three full-time tenured faculty beyond the two new faculty hires as capable of teaching courses in the new major.
b) Page 18 has been updated to indicate students will directly pay for the cost of cloud storage, which is expected to be less than $100 per semester. The majority of courses listed in the program do not require cloud storage.
c) With the termination of the Computer Information Systems major, there are sufficient computer classrooms available to support the Data Science major.
2-Page 18 describes the budget item for outreach through the development of an online course and workshops for high school teachers as a mechanism for marketing the program to high student students.

3-DATA100, DATA200, MT288 and DATA470 syllabi have been updated to link to the University Learning Goals and indicate methods of assessment. MT288 has been updated with details concerning applications in the social sciences.

John Carroll UNIVERSITY

1 John Carroll Boulevard
University Heights, Ohio 44118-4581
www.jcu.edu

To: Dr. Barbara K. D'Ambrosia, Chair, Faculty Council
From: Dr. Brian Williams, VP for Enrollment & Institutional Analytics
Date: September 16, 2016
Re: Support for a B.S. in Data Science

Please consider this document as a full endorsement for a new program in Data Science from the enrollment division.

Programs focused on data align strongly with marketplace needs and employee demand across nearly every industry. The enrollment division was involved in the review and early development of this program proposal. We feel the program frames well the place in our curriculum where data science can reside. Placing the program at the nexus of Statistics, Computer Science and domain knowledge allows students to have a context to their data skills.

Already being ready to launch specialization in 7 domain areas for the launch of the program is testament to the research and data methods and research focus in our existing programs of study. This program and purpose context will be able to resonate with students JCU currently attracts as well as open new marketing channels. JCU will be able to position a different type of major experience than heavier programming and software engineering programs. Such programs will likely arise out of schools of engineering and be a significantly more technical degree than envisioned here. We can work to attract software engineering towards the domain marketing/messaging while also the data science approach can also be marketed to students with expressed interest in those domain areas. This two-pronged approach provides us opportunities to reach a wider audience and grown enrollment.

Further, we support the minor/major rationale. The ability of a student to tip the balance towards domain as a major with a minor in data science, or gravitate towards the major in data science and gain the domain context (as a minor-like experience) creates very distinct pathways for students.

An enrollment estimation was included in the proposal (p. 18) where we articulated the demand on the employer side and shared the difficulty of identifying high school demand at this time. However, as new majors arise, I submit the additional points (mostly qualitative) to underscore how the trend toward data science is timely and urgent.

- Our regional institutions graduated 1,328 students with bachelor's or higher degrees in computer and information sciences, general (CIP code 11.0101), yet the regional demand for such degrees is estimated to be 3,042 per year. The degree awards only accounted for 44 percent of the business demand. Shortages for other degrees in computer science programs are even more severe.
- "They know that many of the tools they use in their personal lives are driven by data, and they are hearing a lot about the career opportunities for people with knowledge and skills in this area. A number of students have commented that they value the opportunity to combine several interests, rather than

focusing exclusively on either statistics or computer science." -- University of Michigan (upon launch of undergraduate Data Science program)

- "Our motivation for developing the program was to provide students with skills to satisfy the needs of business, industry, and government. Student reaction has been very positive. We have gone from zero to 65 co-majors in less than two years and have already had 11 students graduate with the co-major. In addition, our analytics co-majors are highly recruited, with nearly all having multiple job offers by the fall of their senior year." -- Miami University (Analytics Co-Major)
- "Our primary motivation in developing the program was to construct a coherent curriculum that would prepare students to work in this exciting and growing area. While previously students could only piece together courses from various departments that would address particular aspects of analytics (with no guarantee of consistency of curricular structure), development of the major has provided an integrated approach to data analytics education that provides a natural and cohesive curricular path from start to finish. More than 80 students had selected data analytics as their major plan by the end of the first year of the program." - Ohio State University (BS in Data Analytics)

Data science programs are new enough that no federal CIP Code has been established to define this emerging field of study. However, the momentum clearly shows the demand and possibility for us to claim a niche and get into this space as an early adopter. Often development of new majors presents JCU with an opportunity to add majors and specializations that are well established and exist already at other schools. With regard to data science, we would actually be able to be involved in defining emerging big data and data science approaches and be early to market. Further I believe from the data science proposal and working groups, we have sufficient talent and passion within our existing faculty to add a progressive new major wherein the core need is developing coursework. Offering courses that are at a capacity when we have not promoted the major and minor in any way in the recruitment process is a further indication of the ground swell of interest and demand that is possible.

If I can elaborate any further on our support, please do not hesitate to contact me.

Sincerely,

Brian G. Williams
Vice President for Enrollment

Cc:Dr. Linda Seiter

John Carroll UNIVERSITY

September 16, 2016

Dr. Peifang Tian
Chair, Faculty Council Committee on Academic Policies
Associate Professor, Department of Physics

Re: Proposed Bachelor of Science in Data Science

Dear Dr. Tian:

I have reviewed the budget for the proposed new degree program, Bachelor of Science in Data Science. Based upon the results of my review, I believe that this budget was prepared in accordance with JCU budget guidelines and that this budget is realistic.

The projected revenue was prepared on an incremental basis. As in any start-up operation, whether the projected revenue will materialize depends on many factors, including the merit of the program itself and how the program is marketed and executed. Although the projected incremental revenue is modest, we would need only about two new students in year one to break even, assuming that there will be committed support from the Department of Mathematics and Computer Science (see further discussion below). The potential is there for this program to grow tremendously and supplement the current mathematics and computer science programs. The number of students signing up or attending classes for this degree program will demonstrate the demand.

It should be noted that the proposed budget for incremental operating expenses (payroll, fringes and departmental costs) is modest, ranging from about $6,000 to $13,000 annually. Dr. Paul Shick, Chairperson of the Department of Mathematics and Computer Science, has committed to reallocating resources in his department, ranging from $10,000 to $15,000 annually, in order to subsidize this new degree program. Taking this commitment into consideration, it is expected that a small net contribution margin will be achieved in year one.

Of special note is that the hiring of additional faculty will not be necessary to implement this program. The recent termination of the Computer Information System (CIS) major by the Department of Mathematics and Computer Science has created excess capacity within the department.

In summary, I believe that this degree program will enrich John Carroll University with a minimal investment. I wish to express my enthusiastic support.

Please do not hesitate to contact me if you have any questions.

Sincerely,

John Carroll UNIVERSITY

Department of Mathematics and Computer Science
1 John Carroll Boulevard
University Heights, Ohio 44118-4581
Phone 216.397.4351
Fax 216.397.3033
http://www.jcu.edu/math

September 15th, 2016

To the Faculty Council Committee on Academic Policies:

I am writing to offer my full and enthusiastic support the proposal for a new Major and Minor in Data Science submitted on behalf of the Department of Mathematics and Computer Science by Linda Seiter. The Department is strongly behind this effort, and we view the area of Data Science as an essential part of our future.

The proposal itself is beautifully crafted, carefully laying out the rationale behind it, as well as the details and potential costs of implementation. In particular, the proposal argues convincingly the new program will potentially benefit a very wide variety of academic programs by providing students in these disciplines the opportunity to pose meaningful questions, then "harvest" and analyze data proficiently. This is a truly interdisciplinary opportunity for JCU that fits beautifully with the goals of our new Integrative Core. The program is very likely to attract new students to JCU, given the boom in career opportunities in the field.

The main purpose of this letter is to clarify how the proposed new program can be implemented, at least from the perspective of the Math & CS Department. First, both the proposed minor and major are multidisciplinary in structure, so roughly 20% of the courses in any student's curriculum will be drawn from other departments, somewhat lessening the burden on our department. Second, the new program is constructed largely from existing courses, so that course development costs (both in dollars and in time) will be relatively low. Of the new courses that will be needed in the program, two also have very natural "outside" audiences and are classes that we would be developing with or without this proposal. The introductory course, DATA 100, is a badly-needed course that is likely to become one of the most popular QA courses on campus. The topics covered in DATA 100 could form a great starting point for building analytical capabilities for students in science, the Boler School, social sciences and even humanities. MT 288, Network Analysis, is likely to become popular with Computer Science students as well as those in our DS programs. Third, the current Math & CS Department faculty can implement the program, at least in the short term, without additional hires. The recent elimination of our Computer Information Systems major provides enough "extra" coverage for us to handle the additional DATA courses, at least until the program grows significantly. We also anticipate some changes in the demand for MT courses due to the gradual elimination of the

old Core, which will help us cover the Math and Stats courses in the program. If the program attracts as many new students to JCU as we hope, we will eventually need to hire another faculty member in Applied Math or Statistics. The MT & CS Department is quite aware of this shift in our needs, and we have been planning on our next hire being in these more applied areas of the mathematical sciences. Fourth, we have the physical facilities to offer all of the coursework for the program in our existing computer classrooms, so no new construction is necessary. Currently, there are enough available timeslots to offer the needed sections of the DATA 100 and 200 courses for the next few years. We also anticipate some lessening in our computer classroom needs due to the phasing out of the old Core (and perhaps some changes in the Boler core). If the program were to grow enormously, we would eventually need to consider adding a fourth computer classroom or using laptops in regular classrooms, but such considerations are a long way off. Finally, the department should be able to cover the "marginal" costs of the program (including computer lab space and limited summer support) for the first few years out of our existing budget.

In summary, the Department of Mathematics and Computer Science strongly supports the proposal for both a Major and a Minor in Data Science, and we have committed the time and resources necessary to make this important new program work.

If you need any other information or have any questions, please feel free to contact me at 216-397-4352 or at shick@jcu.edu.

Sincerely,

Paul L. Shick
Professor of Mathematics
Chair, Dept. of Mathematics and Computer Science
John Carroll University

John Carroll UNIVERSITY

DEPARTMENT OF SOCIOLOGY AND CRIMINOLOGY
20700 NORTH PARK BOULEVARD
UNIVERSITY HEIGHTS, OHIO 44118-4581
PHONE 216.397.4381
WWW.JCU.EDU

September 19, 2016

Linda Seiter, Ph.D.
Professor
Department of Mathematics and Computer Science

Dear Linda:

As chairperson of the Department of Sociology & Criminology, it is a pleasure to support the proposed Data Science Program. The program's learning goals and curriculum are well thought out and clearly articulated. It provides student with the breadth, depth, and skills in data science plus the opportunity to apply that knowledge in a variety of domains such as: Social Science, Health and Biomedical Informatics, Business, and Psychology. Students graduating JCU with this academic background should be in strong demand in the work force in a variety of fields.

As we discussed, the Department of Sociology and Criminology would be interested in participating in the Data Science Program by the integration of some of our sociology courses: SC 101, SC 350, SC 351, SC 493, and SC 497 as an area of specialization in the Data Science Major. These courses are already offered on a regular basis in our curriculum, and we welcome students from other disciplines in our courses. If there is anything else I can do to move this proposal forward, do not hesitate to contact me.

Sincerely,

Phyllis "Penny" Braudy Harris, Ph.D.
Professor and Chair, Department of Sociology & Criminology

John Carroll UNIVERSITY

Department of Biology
1 John Carroll Boulevard
University Heights, Ohio 44118-4581
Phone 216.397.4451
Fax 216.397.4482
www.jcu.edu

10 October 2016

Dear Members of the Committee on Academic Policies:

I am writing in strong support of the proposed major in Data Science. As a university, we are seeking new and creative ways to repackage current course offerings in ways that will attract students, engage faculty, and generate revenue. I am excited by the prospect of the interdisciplinary degree in Data Science. If the proposal moves forward, JCU will be only the second regional institution to offer this (or a similar) degree at the undergraduate level. This program also offers horizontal pathways for students who start in a more traditional STEM program but desire a more interdisciplinary framework for their studies. Specifically, the proposed area of specialization, "Health Disparities and Social Justice", would be of interest to our students who start in the typical pre-health pathway but decide that professional school in medicine or allied health is not the right fit for them. I also see it as a draw for new students to the institution, who are interested in a data-driven program that also speaks to their commitment to social justice. The Biology courses in this area of specialization are regularly offered by the department (BL 155/157, multiple sections every fall; BL 156/158, multiple sections every spring; BL 240, every fall; BL 260, alternate falls; BL 399, independent study). Therefore, their inclusion in the major does not reflect new expenditures for the university. I am happy to offer my support to this new program, and I hope to find additional pathways through which Biology students could move through this proposed major. Please do not hesitate to contact me if you have additional questions (rdrenovsky@jcu.edu; 216.397.4451).

Best regards,

Rebecca E. Drenovsky
Professor and Chair
Biology Department

John Carroll UNIVERSITY

Physics Department
1 John Carroll Blvd.
University Heights, Ohio 44118-4581
Phone 216.397.4301
Fax 216.397.4499
www.jcu.edu/physics

October 12, 2016

Linda Seiter, Ph.D.
Professor
Department of Mathematics and Computer Science

Dear Linda:

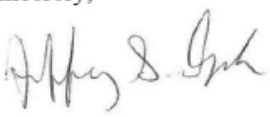
I am happy to provide a letter in support of your new Data Science major. During our meeting recently to discuss the integration of Physics courses as an area of specialization in this major, we agreed that the right set of courses from our area would include:

PH 135 and 135L: Physics 1 with Workshop (lab)
PH 136 and 136L: Physics 2 with Workshop (lab)
together with
MT 135 and MT 136: Calculus I and II (co-/pre-requisites of the above courses).

These calculus-based introductory physics courses cover the foundations of classical physics. In the accompanying labs, we focus on connecting the formal physical theory to actual experiments and data taken by the students. Further, the lab activities are supported by model building in a computing environment. Here students learn the Python programming language (which is integrated into the coursework plans for the Data Science major) and apply numerical calculations to aid in their study of both the experimental and theoretical side of the physics.

The Physics Department very much in favor of this integration with Data Science. It will bring more students into our introductory courses. In addition, first year students that have started as Physics majors will already have this area specialization for the Data Science major. At the end of very year, some of these freshman students find out that the Physics major is not for them; and to have a potential path that takes advantage of all four of these courses will be a real advantage to them. This type of arrangement therefore can strengthen both the Physics and the Data Science program.

Sincerely,



Jeffrey S. Dyck
Chair, Dept. of Physics
John Carroll University

John Carroll UNIVERSITY

University Heights, OH 44118
www.jcu.edu

September 16, 2016

Barbara D'Ambrosia, Ph.D., Chair, Faculty Council
Peifang Tian, Ph.D., Chair, Committee on Academic Policies
John Carroll University
1 John Carroll Boulevard
University Heights, OH 44118

Dear Barbara and Peifang,

I am pleased to offer full support for the proposed major and minor in Data Science. The Tim Russert Department of Communication and Theatre Arts is excited for the opportunity to be an area of specialization for students who might choose the major. We offer multiple courses within the department meet the data-rich expectation, allowing students to apply the knowledge learned in the foundational courses to those in the Russert Department.

I presented this opportunity to the department faculty and we unanimously supported the decision to collaborate with program, assuming its approval. We would be willing to have a member of the department serve as a liaison on the advising committee for the major as well. We see this major as a real opportunity for our students' growth and development as well, whether they choose one course to fulfill their Quantitative Analysis requirement, or the minor to complement their Communication major.

We are pleased to offer our support of this proposal. If I may provide any additional information, please contact me (mfinucane@jcu.edu or 216.397.1608). Thank you for your consideration.

Sincerely,

Margaret Finucane

Margaret Finucane, Ph.D., Chair
Tim Russert Department of Communication & Theatre Arts

CC: Linda Seiter, Ph.D.

John Carroll UNIVERSITY

1 John Carroll Boulevard
University Heights, OH 44118-4581
www.jcu.edu

November 29, 2016

TO: Dr. Linda Seiter, Mathematics and Computer Science

From: Dr. Kathleen Manning, Exercise Science and Sports Studies

RE: Additional Letter of Support for the Data Science Program

---

I am following up on my initial letter of support (10/13/2016) to provide additional information related to Exercise Science & Sports Studies support of this major.

There are 3 tenured and tenure track faculty in Exercise Science & Sports Studies. Each of them supports this major, and supports the inclusion of Exercise Science courses in the major. Dr. Jackie Nagle Zera, is an exercise physiologist, and was your original contact for this major and she provided you with the initial courses that she thought would fit within the content domain of the major. Dr. Greg Farnell, is an exercise physiologist and when we met, expressed support for this major, and the inclusion of our courses within the content domain. Based on my own experiences in development and learning, specifically motor learning, I believe this major and minor is an excellent interdisciplinary opportunity to use our collective resources to integrate the field of data into our major programs in a way not possible as individual departments.

The specific courses that we believe have potential within this major are:

EPA 205/205L:	Human Anatomy (fall semester)
EPA 206/206L:	Human Physiology (spring semester)
EPA 230:	Nutrition for Athletics & Physical Activity (fall, spring)
EPA 407:	Exercise Physiology (fall, spring, beginning fall, 2016)
EPA 409:	Kinesiology (fall, spring, beginning fall, 2016)
EPA 432:	Motor Learning (fall semester)
EPA 440:	Independent Study

As I said in my original letter, I believe this is a major that will attract prospective students because of the integrative nature of the major, and the potential for careers beyond John Carroll. We are happy to support this major. Please let me know if there is anything I can do to assist in this process.

Dear CAP,

It is with great enthusiasm that I am writing in support of the Data Science Program. I have reviewed the proposal and talked with Linda Seiter as well as members of my department, including Colin Swearingen and Elizabeth Stiles who were asked by Dean Farrar to be co-chairs of committee to explore Data Analytics as a program (which I understand may be subsumed by the Data Science Major but I believe both faculty are more than happy about). In addition to being able to offer on a regular basis the courses that the major will draw on for "domain knowledge" we are excited to be developing other courses which could work as well. Moreover, we are looking forward to having Data Science students in our classes and working with Linda and others in a collaborative way.

Right now we have the following courses that we offer on a regular basis that can be taken by Data Science students in fulfillment of their major; PO 102 (Introduction to Comparative Politics), PO 200 (Introduction to Methods), PO 203 (Geographic Information Systems), PO 300/L (Political Science research Methods), PO 324 (Crisis Mapping, New Media, and Politics), PO 332 (African Politics), and PO 337 (Comparative Health Policy). Dr. Colin Swearingen will be teaching two classes in the future that could also be possibilities for this domain: PO160 (Health Care and Social Justice in Latin America) and PO2xx (Health Care Access in Latin America). On behalf of the Department of Political Science I would like to support this innovative proposal and assure the committee that there will be plenty of courses offered on a regular basis to support our inclusion as "domain knowledge."

Sincerely,

Mindy Peden

Chair, Political Science Department

John Carroll UNIVERSITY

October 13, 2016

Dear Members of the Committee on Academic Policies,

I am very pleased that the English Department can contribute to the new proposed major in Data Science. The burgeoning field of Digital Humanities is rich with the possibilities of textual and visual analysis and the representation of data and information. I think it will be a wonderful opportunity for Data Science majors to be able to take courses in the English Department that contribute to their field of knowledge.

Dr. Emily Butler, a specialist in Medieval Studies, offers three courses that could contribute to the major: EN 311: Old English Language and Literature, 312: Late Medieval Literature, and 488: History of the English Language. David Adams, an adjunct in our department, offers 300/409: Technical Writing, which teaches the writing skills necessary for Data Science and that of other science majors. The English Department could also offer an independent study (EN 498) that would allow professors to engage students with data-driven research projects that would benefit both teacher and student.

With our new core curriculum and the drive for more interdisciplinarity, the English Department is happy to support and contribute to the Data Science major.

Thank you.

D. Rosenthal

Debby Rosenthal, PhD
Professor and Chair
English Department
x1721
drosenthal@jcu.edu

John Carroll UNIVERSITY

Department of Psychology
1 John Carroll Boulevard
University Heights, Ohio 44118-4581
Phone 216.397.4355
Fax 216.397.1633
www.jcu.edu

October 13, 2016

Dear Colleagues,

As the chair of the Psychology department, I am writing to you in support of the proposed Data Science major, offered in the Math and Computer Science department. For students selecting Psychology as an area of specialization, the program will require students to complete four courses in Psychology as partial fulfillment of their program requirements. All students will complete PS100 Introduction to Psychology: Perspectives in Social Justice. PS 100 will carry an ISJ designation. The content of this course is the same as PS 101- Introduction to Psychology, with the exception of the additional social justice course requirements. While it is intended for non-majors, the department will permit CS/DS majors to use this course as a perquisite in selecting the three remaining courses.

The Data Science students will then complete three of the following: PS301/301L Experimental Design and Analysis, PS401/401L Advanced Research Methods, PS435 Tests and Measurements, and PS 499 Individual Research Projects in Psychology. Dr. Seiter and I have met to discuss the course descriptions, course expectations, and the prerequisites for each course to ensure that they are viable options for the students. The courses are offered each semester, with the exception of PS 401. PS 401, Advanced Research Methods, is offered in the spring semester, only. We have faculty with the expertise to offer PS 401 each semester if enrollment in the course is significant enough to warrant adding a section.

I have discussed the program with the members of the department. All are in agreement with the plan to offer the above courses as support courses for the Data Science students. We look forward expanding the collaborative work we do with our colleagues in Math and Computer Science Department. Please feel free to contact me with any questions you may have about the above courses, enrollments, or other concerns. I thank you for your time and consideration.

Sincerely,

Sheri Young, Chair
Psychology

John Carroll
UNIVERSITY

November 29, 2016

Dear Colleagues,

I am writing to you in support of the proposed Data Science major, offered in the Math and Computer Science department. We are excited that the entrepreneurship minor has the opportunity to be an area of specialization for students who might choose the major. Dr. Seiter and I have met to discuss entrepreneurship courses ER 201, ER 301, ER 304, ER 305, ER 306 and ER 480 to ensure they are viable options for the students. We offer four of these courses (ER 201, ER 301, ER 304, and ER 480) each semester and two once a year (ER 305 Fall, ER 306 Spring).

I have discussed this opportunity with faculty (full and adjunct) teaching in the minor. All were in favor of the proposal and agreed to ensure that their classes would fulfill expectations. The entrepreneurship program looks forward to working more directly with faculty in the Math and Computer Science Department and support for this proposal. If you need any additional information, please contact me (schmidt@jcu.edu) or 216-397-4242.

Sincerely,

Jackie Schmidt

Jackie Schmidt
Director of the Entrepreneurship Minor

9. References

[1] McKinsey Global Institute, "Big Data: The next frontier for innovation, competition and productivity," 2011. [Online]. Available: http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation.

[2] K. Noyes, "Educating the "big data" generation," 27 May 2014. [Online]. Available: http://fortune.com/2014/05/27/educating-the-big-data-generation/.

[3] Office of Science and Technology Policy, Executive Office of the President, "Big Data Initiative," 29 March 2012. [Online]. Available: https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf.

[4] T. H. D. a. D. Patil, "Data Scientist: The Sexiest Job of the 21st Century," *Harvard Business Review,* October 2012.

[5] University of Chicago, "Data Science For Social Good," [Online]. Available: https://dssg.uchicago.edu/.

[6] University of Washington - eScience Institute, "Data Science for Social Good," [Online]. Available: http://escience.washington.edu/get-involved/incubator-programs/data-science-for-social-good/.

[7] Georgia Tech, "Atlanta Data Science for Social Good," [Online]. Available: http://dssg-atl.io/.

[8] Bayes Impact, [Online]. Available: http://www.bayesimpact.org/.

[9] DrivenData, [Online]. Available: https://www.drivendata.org/.

[10] CRA, "2015 Taulbee Survey," *Computing Research News,* May 2016.

[11] B. Schulte, "Women flocking to statistics, the newly hot, high-tech field of data science," The Washington Post, 19 December 2014. [Online]. Available: https://www.washingtonpost.com/local/women-flocking-to-statistics-the-new-hot-high-tech-field-of-data-science/2014/12/19/f3e2e486-62ed-11e4-9fdc-d43b053ecb4d_story.html.

[12] College Board, "AP Program Participation and Performance Data 2015," [Online]. Available: https://research.collegeboard.org/programs/ap/data/participation/ap-2015.

[13] DataScience.Community, "COLLEGE & UNIVERSITY DATA SCIENCE DEGREES," [Online]. Available: http://datascience.community/college.

[14] AMSTAT News, "New Undergraduate Data Science Programs," [Online]. Available: http://magazine.amstat.org/blog/2015/08/01/new-undergraduate-data-science-programs-2.

[15] Amit Mitra, PhD. Department of Systems and Technology, Auburn University, "Electronic Correspondence," 19 July 2016. [Online].

[16] Case Western Reserve University, School of Engineering, "Applied Data Science Undergraduate Degree," [Online]. Available: http://case.edu/datascience/for-students/degree-programs/bachelor-of-science/.

[17] University of San Francisco, "Data Science Undergraduate Degree," [Online]. Available: https://www.usfca.edu/arts-sciences/undergraduate-programs/data-science.

[18] University of Michigan, "Data Science Undergraduate Degree," [Online]. Available: https://www.eecs.umich.edu/eecs/undergraduate/data-science/.

[19] Denison University, [Online]. Available: http://denison.edu/academics/data-analytics.

[20] American Statistical Association, "Curriculum Guidelines for Undergraduate Programs in Statistical Sciences," 2014.

[21] G. Blumenstyk, "Liberal-Arts Majors Have Plenty of Job Prospects, if They Have Some Specific Skills, Too," *Chronicle of Higher Education,* 9 June 2016.

[22] Chmura Economics & Analytics, "Big Data and Analytics in Northern Virginia and the Potomac Region," [Online]. Available: https://gwtoday.gwu.edu/sites/gwtoday.gwu.edu/files/downloads/BigData%20report%202014%20for%20Web.pdf.

Appendix A – Course Descriptions

NEW COURSES

DATA100 INTRODUCTION TO DATA SCIENCE (3 credits)
Data science capitalizes on big data and focuses on data analytics that turn information into actionable knowledge. This course will introduce students to the key ideas, practices, and challenges of modern data analysis. Students will get an overview of the data, questions, and tools that data scientists deal with in their practice. This course will introduce students to practical approaches to essential exploratory techniques, interactive data discovery, and predictive analytics including basic techniques for collecting, cleaning, and sharing data. Hands-on activities will enable the students to learn the practical toolkit of a data scientist.

DATA200 INTERMEDIATE DATA SCIENCE (3 credits)
Prerequisites: CS128, DATA100, MT/DATA122 or equivalent. Pre or Co-requisite: CS/DATA150.
This course will provide a strong foundation in the field of data science and data analytics with a focus on computational approaches and experiential learning. Students will learn about processes and practice of data science that are developed to analyze diverse sources of data including data modeling, machine learning, and natural language processing. This course will present the fundamentals of inference in a practical approach. Students will build a portfolio of a Big Data skills.

DATA470 DATA SCIENCE PROJECT (3 credits)
Prerequisite: senior standing in the Data Science major, DATA200, MT223/DATA260, MT288.
Simulation of the environment of the professional data scientist working in a team on a large data project for a real client. Students will encounter a wide variety of issues that naturally occur in a project of scale, using their skills, ingenuity, and research abilities to address all issues and deliver a usable data product.

MT288 NETWORK ANALYSIS (3 credits)
Introduction to the mathematical and computational theory of network analysis with applications in globalization studies and other social sciences. Fundamentals of graph theory and matrix algebra. Measures of networks including concepts of centrality and structural balance. Study of large-scale structure of networks. Computational representation of and algorithms on networks. Independent student project.

EXISTING COURSES

CS/DATA150 DATABASE SYSTEMS (3 credits)
Introduction to relational database design and implementation. Topics include database systems concepts and architectures, structure query language (SQL) entity relationship (ER) modeling, relational database design, functional dependencies and normalization.

CS128 INTRODUCTION TO SOFTWARE APPLICATION DEVELOPMENT (3 credits)
Corequisite: CS128L. Fundamentals of computing with an emphasis on mobile technology. Utilize a visual programming environment to design, build and test mobile apps. Introduction to application development, inquiry-based simulation, rapid prototyping, incremental problems solving and graphical user interface programming.

CS128L INTRODUCTION TO SOFTWARE APPLICATION DEVELOPMENT LAB (1 credit)
Corequisite: CS128. Programming laboratory intended to provide hands-on experience in applying the programming concepts learned in CS128. Experience in learning the process of program development, with emphasis on techniques for testing and debugging. CS128 and CS128L must be taken together in the same semester.

MT/DATA122 ELEMENTARY STATISTICS I (3 credits)
Describing data by graphs and measures, sampling distributions, confidence intervals and tests of hypotheses for one and two means and proportions, Chi-square tests, correlation and regression. Methods are illustrated in the context of quantitative research, with applications in disciplines such as sports, psychology, and social and natural sciences. Use of appropriate statistical software

MT223/DATA260 INTERMEDIATE STATISTICS (3 credits)
Prerequisite: MT/DATA 122. Power analysis, factorial and repeated measures analysis of variance, nonparametric procedures, contingency tables, introduction to multiple regression. Use of appropriate statistical software.

MT/DATA228 STATISTICS FOR THE BIOLOGICAL SCIENCES (3 credits) Exploratory data analysis, probability fundamentals, sampling distributions and the Central Limit Theorem, estimation and tests of hypotheses through one-factor analysis of variance, simple linear regression, and contingency tables using appropriate statistical software. Course content in biology context.

MT/DATA229 PROBABILITY AND STATISTICS (3 credits) Prerequisite: MT 135. Probability, discrete and continuous distributions, sampling distributions and the Central Limit Theorem, introduction to data analysis, estimation and hypothesis testing, simple linear regression and correlation; exact, normal-theory, and simulation-based inference; use of appropriate statistical software. Methods are illustrated in the context of quantitative research, with applications in disciplines including sports, psychology, and social and natural sciences.

EN300 TECHNICAL WRITING (3 credits)
Introduction to effective workplace writing practices; emphasis on technical and digital writing, audience and organizational needs, information design, ethics, usability testing, and team writing. This course is a "W" course and a culminating course for students in the Professional Writing Track within the English Department. As such it serves as a pre-professional course for those students in that track and an intensive value-added course for students from other majors. The structure of the course is best revealed through its Learning Outcomes and Assignment Menu. The course certainly qualifies as an extended exercise in active learning.

CS/DATA307 BIOINFORMATICS (3 credits)
Prerequisite: CS228 or DATA200. The Application of computational methods and principles to solve data-intensive and pattern-discovery problems in biology, especially molecular and systems biology, without prior knowledge of college-level biology. Topics may gene sequence assembly, sequence alignment, phylogenetic tree inference, gene expression, and protein interaction networks.

CS/DATA322 BIG DATA ANALYTICS (3 credits)
Prerequisite: CS/DATA150 and either CS228 or DATA200. What is "Big Data"? Data mining algorithms, machine learning algorithms. Emphasis on real analyses that are being performed every day by businesses, governments, and online social networks.

CS/DATA340 DATA VISUALIZATION (3 credits)
Prerequisite: CS228 or DATA200. Introduction to basic data visualization techniques. Discussion of different techniques to view data, and analysis of classic data representations. Students will use advanced tools for generating and exploring, static and dynamic visual representations of very large datasets.

CS/DATA350 ADVANCED DATABASE (3 credits)
Prerequisite: CS/DATA150 and either CS228 or DATA200. Alternative data models and advanced database techniques. Object-oriented data models, Web-DBMS integration technology, data-warehousing and date-mining techniques, database security and optimization, other advanced topics.

MT/DATA421 MATHEMATICAL STATISTICS (3 credits)
Prerequisites: DATA/MT 229, MT233. Moment generating functions, transformations, properties of estimators, foundations of hypothesis tests, one- and two-factor analysis of variance, and nonparametric analyses.

MT/DATA422 APPLIED STATISTICS (3 credits)
Prerequisites: DATA 260/MT 223. Two factor analysis of variance; categorical data analysis, logistic regression, factor analysis, simulation, analysis of large datasets; use of appropriate statistical software.

MT/DATA424 APPLIED REGRESSION ANALYSIS (3 credits)
Prerequisite: DATA 260/MT 223. Multiple linear regression, collinearity, model diagnostics, variable selection, nonlinear models; autocorrelation, time series, and forecasting; use of appropriate statistical software.

Appendix B – New Course Syllabi

DATA100
DATA200
MT288
DATA470

DATA 100 Introduction to Data Science

Description

Data science capitalizes on big data and focuses on data analytics that turn information into actionable knowledge. This course will introduce students to the key ideas, practices, and challenges of modern data analysis. Students will get an overview of the data, questions, and tools that data scientists deal with in their practice. This course will introduce students to practical approaches to essential exploratory techniques, interactive data discovery, visualization, and predictive analytics including basic techniques for collecting, cleaning, and sharing data. Hands-on activities will enable the students to learn the practical toolkit of a data scientist.

Learning Objectives

- Introduce students to fundamental principles of data science.
- Teach basics of data science, stimulate data-analytic thinking, and develop data-analytic skills.
- Teach basic methods for effective data communication and visualization, and demonstrate their use in data representation.
- Provide hands-on experience in applying basic data science methods and algorithms and interpreting their results.

Learning goals:
The goals for this course meet at an introductory level the overall goals for the Data Science Program including:

1. **Data Acquisition**: collect, store, preserve, manage and share data in a distributed environment through practical, hands-on experience with programming languages and big data tools;
2. **Problem Exploration**: develop problem solving skills through experiences that foster computational and data-analytic thinking;
3. **Analysis**: develop a basic understanding of the key technologies in data science: data mining, machine learning, visualization techniques, predictive modeling, and statistics;
4. **Domain knowledge**: experience discipline-specific data use cases in order to solve real-world problems of high complexity;
5. **Interpretation**: learn methods for effective data communication and visualization, and demonstrate their use in data representation;
6. **Social Value**: explore social and ethical implications of the use of data and technology.

Specifically, after taking this class, students should:

- Understand the role of the data science in today's world, in business, medicine, entertainment, social networking, and other domains.
- Be able to identify the kinds of problems for which data science may provide intelligent solutions.
- Understand the components and challenges of the entire data science pipeline.
- Be able to use basics techniques for collecting, cleaning and sharing data.
- Understand basic methods and related tools that are used for effective data analysis and visualization.
- Be able to apply basic data science methods, processes, and algorithms using open-source tools.
- Demonstrate an ability to ask questions, express data creativity, and communicate data-driven discoveries.

This course relates to the John Carroll University Academic Learning Goals:

- Demonstrate an integrative knowledge of human and natural worlds.
- Make connections across disciplines, perspectives, and methods.
- Develop habits of critical analysis and aesthetic appreciation.
- Identify and understand the fundamental element of a problem.
- Demonstrate quantitative literacy.
- Demonstrate the ability to analyze multiple forms of expression (oral, written, digital or visual).
- Develop critical thinking skills.
- Apply creative and innovative thinking.
- Demonstrate problem-solving skills.
- Communicate skillfully in multiple forms of expression.
- Employ technology effectively to access and convey information.
- Apply a framework for examining ethical dilemmas.

Required Texts:

R for Data Science 1st Edition by Wickham and Grolemund, 2016

Software:

Students need access to the Internet and open-source software - R in a cloud-based environment.

Tentative Schedule		
	Topics	
Week 1	Welcome to Data Science	Syllabus, textbooks, motivational videos, journal articles
Week 2	Introduction to the Data Scientist's Toolbox	R and RStudio, markdown, and git
Week 3	Getting Started with R	Install and configure software for statistical programming, generic programming language concepts
Week 4	R Programming	Practical statistical computing: reading data into R, working with R packages
Week 5	R Programming	Writing R functions, organizing and commenting R code.
Week 6	Getting and Cleaning Data	Basic ways to obtain data from the web, from APIs, and from databases.
Week 7	Getting and Cleaning Data	The basics techniques for collecting, cleaning and sharing data.
Week 8	Getting and Cleaning Data	
Week 9	Exploratory Data Analysis in R	Exploratory techniques for summarizing data
Week 10	Data Visualization in R and SynGlyphX	Fundamentals of creating visualizations: choosing tools to visualize data, visualizing patterns, relationships, spotting differences.
Week 11	Data Visualization in R and SynGlyphX	Common techniques used to visualize high-dimensional data.
Week 12	What is Data Mining?	Introduction to Data Mining. Basic examples of Data Mining algorithms, supervised vs. unsupervised learning.
Week 13	Data Mining Algorithms	Examples of clustering and decisions tree algorithms
Week 14	Data Science Project	Students will demonstrate their understanding of basic data science methods by working on a real world data science project. Students engage in the process of solving a data science project.
Week 15	Course Review and Presentations	

Grades

30% Homework / Programming Assignments
10% Quizzes
25% Midterms (2)
10% Project
25% Final Exam

Attendance

You are learning a new set of concepts and new programming languages – you need to be in class to ensure that learning takes place. Excused absences need to be arranged ahead of time, or sufficiently documented (doctor's note, dean's email. etc.). Three unexcused absences result in a lowering of your grade by a full letter. Additional absences will result in a meeting with the instructor and potentially the department chair.

Academic Honesty

The nature of this class is such that plagiarizing either from classmates or the internet is a potential problem. There is simply no way to learn to program without doing the work. Like learning a foreign language, you must practice and you must encounter real issues and develop real solutions. In later courses and professional practice, it makes sense to build programs on top of the work of others. As Newton said, "If I can see farther it is because I am standing on the shoulders of giants". This approach does not work when learning the basics of a field. If you are struggling with a concept, idea, or technique from class, that tells you it's time to visit the professor during their office hours. If you do choose to use the work of others and present it as your own, or use it in an unattributed way, this constitutes plagiarism and will result in a 0 for the first offense, the second offense will be a grade of "F" in the course.

Notes on Academic Honesty and Working with Classmates:
Learning in groups and studying with a partner or partners has a long tradition and great value. However, it is critically important to fully understand where the line is between helping each other learn and violating course policy. Studying together, working though concepts with other students, sharing insights are all important and useful aspects of learning. These types of activities should all be done BEFORE writing up the solutions to the homework. Consider the following example: Two students are struggling with a concept at the core of the current homework assignment. They have a study session in which they go over the class notes, read the text, and look on the internet for information. After working together for, say two hours, they feel that they now understand the concept and can apply it. At that point they should go separate ways and apply the concept to the assignment independently. All work turned in MUST be produced individually.

Students with Disabilities

In accordance with federal law, if you have a documented disability (learning, psychological, sensory, physical, or medical) you may be eligible to request accommodations from the Office of Services for Students with Disabilities (SSD). To make a request for accommodations, please contact SSD Director Allison West at (216) 397-4967 or visit the SSD office, located in Room 7A, on the garden (lower) level of the Administration Building. Please keep in mind that accommodations are not retroactive so it is best to register with SSD at the beginning of each semester. Only those accommodations approved by SSD will be recognized by your instructors. Please contact SSD if you have further questions.

Sexual Harassment and Bias

John Carroll University is committed to fostering a learning and working environment based open communication, mutual respect, and ethical and moral values consistent with Jesuit and Catholic traditions. The University seeks to provide an environment that is free of bias, discrimination, and harassment, including sexual harassment.

If you have experienced, sexual harassment/assault/misconduct/gender/sex/sexual orientation, and you share this with a faculty member, the faculty member must notify the Title IX Coordinator, Kendra Svilar, J.D., who will discuss options with you. She can be reached by email at ksvilar@jcu.edu or (216) 397-1559. For more information about your options and resources, please go to http://sites.jcu.edu/hr/pages/resourcespolicies/title-ix/.

If you have experienced bias or discrimination based on race, age, sex, sexual orientation*, religion, ethnic or national origin, disability, military or veteran status, genetic information or any factor protected by law, you are encouraged to report this to the Bias Reporting System at* ***http://sites.jcu.edu/bias*** *or to Dr. Terry Mills, Assistant Provost for Diversity and Inclusion, at tmills@jcu.edu, or (216) 397-4455. For more information about the University commitment to diversity and inclusion, please see* ***http://sites.jcu.edu/ diversity****.*

**You can report concerns anonymously through the Bias Reporting System.*

DATA 200 – Intermediate Data Science

Course Objectives

This course will provide a strong foundation in the field of data science and data analytics with a focus on computational approaches and experiential learning. Students will learn about processes and practice of data science that are developed to analyze diverse sources of data including data modeling, machine learning, and natural language processing. This course will present the fundamentals of inference in a practical approach. Students will build a portfolio of Big Data skills.

Data Science Tools

We will use various tools and programming languages to analyze data such as Python, IPython, Mahout, Pig, NumPy, pandas, SciPy, Scikit-learn, the Natural Language Toolkit (NLTK), and Spark MLlib.

Prerequisites

DATA100, CS128, MT122. Co or pre-requisite: CS150.

Data Science Program Learning Goals

The goals for this course meet at an intermediate level the Data Science Program Learning Goals:

1. **Data Acquisition**: collect, store, preserve, manage and share data in a distributed environment through practical, hands-on experience with programming languages and big data tools;
2. **Problem Exploration**: develop problem solving skills through experiences that foster computational and data-analytic thinking;
3. **Analysis**: develop a basic understanding of the key technologies in data science: data mining, machine learning, visualization techniques, predictive modeling, and statistics;
4. **Domain knowledge**: experience discipline-specific data use cases in order to solve real-world problems of high complexity;
5. **Interpretation**: learn methods for effective data communication and visualization, and demonstrate their use in data representation;
6. **Social Value**: explore social and ethical implications of the use of data and technology.

Course learning goals:

- Provide detailed overview of computational approaches that are developed to analyze diverse sources of data.
- Introduce technologies that are used to collect, store, preserve, manage, analyze, and share large quantities of data in a distributed environment.
- Understand the fundamentals of data modeling, data mining, and basic statistical inference and use this information for making informed choices in analyzing data.

This course relates to the John Carroll University goals of:

- Demonstrate an integrative knowledge of human and natural worlds.
- Make connections across disciplines, perspectives, and methods.
- Develop habits of critical analysis and aesthetic appreciation.

- Identify and understand the fundamental element of a problem.
- Demonstrate quantitive literacy.
- Demonstrate the ability to analyze multiple forms of expression (oral, written, digital or visual).
- Develop critical thinking skills.
- Apply creative and innovative thinking.
- Demonstrate problem-solvent skills.
- Communicate skillfully in multiple forms of expression.
- Employ technology effectively to access and convey information.
- Apply a framework for examining ethical dilemmas.

Learning Outcomes

After taking this course, students will:

- Describe the data science life cycle.
- Recognize use cases for data science.
- Describe supervised and unsupervised learning differences.
- Understand machine learning tasks.
- Use Pig (SQL-like) language to transform and prepare data on a data cluster.
- Write a Python script using Scientific Python Ecosystem to analyze big data.
- Use the data structure classes in the pandas library.
- Describe use cases for Natural Language Processing (NLP)
- Understand machine learning techniques used for NLP.
- Be able to use narratives and visualizations to communicate findings.

Schedule

	Topics	Tools / Programming Languages / Packages for Scientific Computing
Week 1	Use Cases of Data Science What is Data Science? Data Science vs. Business Intelligence What is a Data Scientist?	
Week 2	Understanding Distributed File System and Block Storage Understanding MapReduce WordCount in MapReduce	Basic Linux skills (e.g. navigate the file system, copy files, create folders) Hadoop - MapReduce
Week 3	What is Machine Learning? Computational Approach to Modeling The History of Machine Learning Supervised vs. Unsupervised Learning One or more examples Examples of Unsupervised Learning: Clustering, Outlier Detection , Affinity Analysis Examples of Supervised Learning: Classification, Regression, Decision Trees, Recommendation	Mahout , an open-source Apache project whose goal is to build a scalable machine learning library
Week 4	Preprocessing of unstructured data without a schema High- level, SQL-like programming language	Apache Pig Queries, Grunt shell
Week 5	Python Programming (2 - 3 weeks)	Python IDE
Week 6	The Scientific Python Ecosystem Analyzing Data with Python	Exploring fundamental packages of scientific computing: NumPy, SciPy, matplotlib, pandas, IPython, SymPy
Week 7	Running Python on a Data Cluster	

	Topics	Tools / Programming Languages / Packages for Scientific Computing
Week 8	Designing and Executing Python Machine Learning Algorithm on a cluster: One or more of the the following algorithms will be discussed: Support Vector Machines, Naive Bayes for Classification, Nearest Neighbors, Tree Based Methods Challenges of Machine Learning on a data cluster	Scikit-learn, an open-source machine learning library for Python that features various classification, regression and clustering algorithms:
Week 9	Introduction to Natural Language Processing What is NLP? NLP in Big Data Common Tasks in NLP One or more of the following topics will be discussed: Optical Character Recognition Sentence Segmentation Part-of-speech Tagging Named Entity Recognition Topic Modeling NLTK - Natural Language Toolkit Classifying Text The NaiveBayesClassifier Decision Trees	Natural Language Toolkit Spark MLlib, a fast, in-memory computing system that runs on Hadoop.
Week 10	Data Science Project	
Week 11	Data Science Project	
Week 12	Data Science Project	
Week 13	Review	
Week 14	Review	

Grades

30% Homeworks / Programming Assignment
10% Quizzes
25% Midterms (2)
10% Project
25% Final Exam

Attendance

You are learning a new set of concepts and new programming languages – you need to be here to ensure that that learning takes place. Excused absences need to be arranged ahead of time, or sufficiently documented (doctor's note, dean's email. Etc.). Three unexcused absences result in a lowering of your grade by a full letter. Additional absences will result in a meeting with the instructor and potentially the department chair.

Academic Honesty

The nature of this class is such that plagiarizing either from classmates or the internet is a potential problem. There is simply no way to learn to program without doing the work. Like learning a foreign language, you must practice and you must encounter real issues and develop real solutions. In later courses and professional practice, it makes sense to build programs on top of the work of others. As Newton said, "If I can see farther it is because I am standing on the shoulders of giants". This approach does not work when learning the basics of a field. If you are struggling with a concept, idea, or technique from class, that tells you it's time to visit the professor during their office hours. If you do choose to use the work of others and present it as your own, or use it in an unattributed way, this constitutes plagiarism and will result in a 0 for the first offense, the second offense will be a grade of "F" in the course.

Notes on Academic Honesty and Working with Classmates:
Learning in groups and studying with a partner or partners has a long tradition and great value. However, it is critically important to fully understand where the line is between helping each other learn and violating course policy. Studying together, working though concepts with other students, sharing insights are all important and useful aspects of learning. These types of activities should all be done BEFORE writing up the solutions to the homework. Consider the following example: Two students are struggling with a concept at the core of the current homework assignment. They have a study session in which they go over the class notes, read the text, and look on the internet for information. After working together for, say two hours, they feel that they now understand the concept and can apply it. At that point they should go separate ways and apply the concept to the assignment independently. All work turned in MUST be produced individually.

Students with Disabilities

In accordance with federal law, if you have a documented disability (learning, psychological, sensory, physical, or medical) you may be eligible to request accommodations from the Office of Services for Students with Disabilities (SSD). To make a request for accommodations, please contact SSD Director Allison West at (216) 397-4967 or visit the SSD office, located in Room 7A, on the garden (lower) level of the Administration Building. Please keep in mind that accommodations are not retroactive so it is best to register with SSD at the beginning of each semester. Only those accommodations approved by SSD will be recognized by your instructors. Please contact SSD if you have further questions.

Sexual Harassment and Bias

John Carroll University is committed to fostering a learning and working environment based open communication, mutual respect, and ethical and moral values consistent with Jesuit and Catholic traditions. The University seeks to provide an environment that is free of bias, discrimination, and harassment, including sexual harassment.

If you have experienced, sexual harassment/assault/misconduct/gender/sex/sexual orientation, and you share this with a faculty member, the faculty member must notify the Title IX Coordinator, Kendra Svilar, J.D., who will discuss options with you. She can be reached by email at ksvilar@jcu.edu or (216) 397-1559. For more information about your options and resources, please go to http://sites.jcu.edu/hr/pages/resourcespolicies/title-ix/.

If you have experienced bias or discrimination based on race, age, sex, sexual orientation*, religion, ethnic or national origin, disability, military or veteran status, genetic information or any factor protected by law, you are encouraged to report this to the Bias Reporting System at* ***http://sites.jcu.edu/bias*** *or to Dr. Terry Mills, Assistant Provost for Diversity and Inclusion, at tmills@jcu.edu, or (216) 397-4455. For more information about the University commitment to diversity and inclusion, please see* ***http://sites.jcu.edu/ diversity****.*

**You can report concerns anonymously through the Bias Reporting System.*

MT 288
Introduction to Network Analysis

Course Description

Introduction to the mathematical and computational theory of network analysis. Fundamentals of graph theory and matrix algebra. Measures of networks including concepts of centrality and structural balance. Study of large-scale structure of networks. Computational representation of and algorithms on networks. Independent student project.

Required Materials:

M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, 2010.

Other materials may be assigned on an *ad hoc* basis.

Accommodation

In accordance with federal law, if you have a documented disability (Learning, Psychological, Sensory, Physical, or Medical) you may be eligible to request accommodations from the Office of Services for Students with Disabilities (SSD). Please contact the Director, Allison West at (216) 397-4967 or come to the office located in room 7A, in the Garden Level of the Administration Building. Please keep in mind that accommodations are not retroactive so it is best to register at the beginning of each semester. Only accommodations approved by SSD will be recognized in the classroom. Please contact SSD if you have further questions.

Anti-Bias Policy

At John Carroll University, we are committed to fostering a respectful and inclusive campus community. Incidents of bias which are intentional or unintentional actions against someone on the basis of an actual or perceived aspect of their identity, including actions that occur in classrooms, can and should be reported on the Bias Incident Reporting Form, accessible at http://sites.jcu.edu/bias/. Questions about bias can be directed to members of the Bias Response Team: Lauren Bowen, Associate Academic Vice President (bowen@jcu.edu), Bud Stuppy, Director of Human Resources (cstuppy@jcu.edu) or Danielle Carter, Director of the Center for Student Diversity and Inclusion (dcarter@jcu.edu).

Academic Honesty

"Whoever repeats something he heard and names its source brings salvation to the world." – Rabbi Chanina (c. 200 B.C.E.)

Academic honesty, expected of every student, is essential to the process of education and to upholding high ethical standards. Cheating, including plagiarism, inappropriate use of technology, or any other kind of unethical behavior, may subject the student to severe academic penalties, including dismissal.

All work submitted for evaluation in a course, including tests, term papers, and computer programs, must represent only the work of the student unless indicated otherwise. Material taken from the work of others must be acknowledged. Materials submitted to fulfill requirements in one course may not be submitted in another course without prior approval of the instructor(s). Concerns about the propriety of obtaining outside assistance and acknowledging sources should be addressed to the instructor of the course before the work commences and as necessary as the work proceeds.

You will receive a grade of 0 for any work in this class on which you have been academically dishonest, and I will write and submit a report to the Dean of the College of Arts and Science describing the incident. Further details regarding the University's policy on academic dishonesty may be found in the current *Undergraduate Bulletin.*

Attendance Policy

Your participation in this course, both during class sessions and in outside studies, is a requisite of the course. A portion of your final grade will be dependent on the level of your participation during the class sessions. You are expected to attend each class meeting time with your full attention and participation. It is customary for me to call on non-volunteer students to answer questions, give feedback and provide insight into discussions.

Assessment

Your grade in this course depends on your mastery of the course content. There are three content units to the course (described below). You will be assessed for your mastery of the content of each unit. Your score within each unit will be determined as a weighted average of homework, participation and examination scores. The specific weight of each type of assessment within each unit is listed below. Furthermore, there will be a course synthesis score, determined by the score on the Independent Project.

Your final score in the course will be the average of the three unit scores and the course synthesis score. Grades will be assigned by the following percentages. Borderline cases will be decided by class participation and general disposition of student.

Grade	A	A-	B+	B	B-	C+	C	D	F
Lowest %	93	90	87	83	80	77	70	60	0

The following table describes each content unit and the assessment method utilized to determine your score within each unit.

Assessment Category	Description
Unit One The Mathematics of Networks (25% of Final Grade)	The grade for this unit will be calculated as the weighted average consisting of • 30% of the total homework grade percentage • 30% of the total participation grade percentage • 40% of the Exam 1 grade The examination for this unit will occur on *. Late or missed work in this category will result in a score of 0 points.
Unit Two Measures of Networks (25% of Final Grade)	The grade for this unit will be calculated as the weighted average consisting of • 30% of the total homework grade percentage • 30% of the total participation grade percentage • 40% of the Exam 2 grade The examination for this unit will occur on * Late or missed work in this category will result in a score of 0 points.
Unit Three Computational Aspects of Networks (25% of Final Grade)	The grade for this unit will be calculated as the weighted average consisting of • 30% of the total homework grade percentage • 30% of the total participation grade percentage • 40% of the Exam 3 grade The examination for this unit will occur on *

	Late or missed work in this category will result in a score of 0 points.
Unit Four Course Synthesis (25% of Final Grade)	The grade for this unit will be calculated as the grade of the Final Independent Project The examination for this unit will occur on *. Late or missed work in this category will result in a score of 0 points.

Homework & Lab Assignments

There will be periodic homework assignments covering the content material of the course. These assignments will range from short writing assignments to larger synthesis papers. It is expected that you will submit neat, legible assignments. Passages within your submissions, which are overly messy or un-readable within your submissions, will be ignored and graded as a 0 towards your grade. Late or missing assignments – barring extreme circumstances – will be counted as a 0 grade.

Exams

There will be three examinations in this course. Three exams will cover topics from each of the three content units.

Independent Project

The course synthesis portion of your final grade will be based on the Final Independent Project. This project, which you will be working on from the first week of the course to the last, will give you the opportunity to explore how network analysis can be used to understand better and solve problems within an area of your interest. You will be expected to provide a poster for the final poster session of the course as well as a paper detailing your findings. A schedule of the required work for this project will be provided early in the semester.

Course Outline: Below is a rough schedule of the course of study with the relevant sections from the text. You are expected to have read the text materials prior to each session. This is always subject to change.

Each content unit is indicated by color as follows:

Content Unit 1	Content Unit 2	Content Unit 3
The Mathematics of Networks	Measures of Networks	Computational Aspects of Networks

Dates	**Topics**
Week 1	Basic Network Concepts and Applications ■ Basic graph theory ■ Review of common networks
Week 2	Matrix Algebra ■ Review of matrix algebra ■ Adjacency matrix of networks
Week 3	Special Networks ■ Acyclic networks & trees ■ Bipartite networks
Week 4	Characteristics of Networks ■ Degrees ■ Paths ■ Components ■ Graph Laplacian ■ Random Walks
Week 5	Degree-related Centrality ■ Degree centrality ■ Eigenvector centrality ■ Katz centrality ■ PageRank centrality
Week 6	Relational Centrality ■ Hubs/Authorities ■ Closeness centrality ■ Between-ness centrality

Week 7	Social-oriented Measures, pt. 1 ■ Topology of sets of vertices ■ Transitivity & clustering coefficient ■ Reciprocity
Week 8	Social-oriented Measures, pt. 2 ■ Structural balance ■ Similarity ■ Homophily
Week 9	Large-scale Structure of Networks, pt. 1 ■ Components ■ Small-world effect ■ Degree distribution ■ Power laws
Week 10	Large-scale Structure of Networks, pt. 2 ■ Distribution of other measures ■ Cluster coefficients ■ Assortative mixing
Week 11	Representing & Storing Network Data, pt. 1 ■ Introduction to computational complexity ■ Adjacency matrices ■ Adjacency lists
Week 12	Representing & Storing Network Data, pt. 1 ■ Trees ■ Heaps
Week 13	Fundamental Network Algorithms, pt. 1 ■ Degree & degree distributions ■ Clustering coefficients
Week 14	Fundamental Network Algorithms, pt. 2 ■ Shortest paths ■ Maximum flow & minimum cuts

DATA 470 **Data Science Capstone**

Purpose and Content
Prerequisites: senior Data Science major, DATA 200, 260, MT288.

The purpose of this Data Science Capstone is to simulate the experience of a real data science team through the development of a new data product. The capstone project class will allow students to create a usable/public data product that can will demonstrate their skills. Projects will be drawn from real-world problems and will be conducted with industry, government, and academic partners. Data science teams will make widespread use of previously learned tools and methods. Student developers will encounter a wide variety of issues that naturally occur in a project of scale, using their skills, ingenuity, and research abilities to address all issues and deliver a data product.

Class

This class meets twice a week. During one class meeting each week, you will be given time in class to work on your project. I will spend every third class embedded with your project team. (For example – in the second week of class, I will meet with team 1, the third week with team 2, and the fourth week with team 3. We then return to team 1 the following week. My role will be observer, collaborator, and evaluator. Each session I will give your team a progress report and grade. The second meeting each week will be devoted to an aspect of professional data science development. These classes will include lectures, guest speakers, activities/exercises, discussing readings, and meetings with clients. Throughout the semester we will have code walkthroughs, status reports and project demonstrations. There will be occasional readings assigned throughout the semester. Teams will be required to prepare a discussion of the assigned readings. The preparation includes distributing a list of suggested discussion questions prior to the date of the discussion. For the project reviews, both the developers and participants will be evaluated on their contribution to the event.

Learning Goals

The goals for this course meet at an intermediate level the Data Science Learning Goals:

1. **Data Acquisition**: collect, store, preserve, manage and share data in a distributed environment through practical, hands-on experience with programming languages and big data tools;
2. **Problem Exploration**: develop problem solving skills through experiences that foster computational and data-analytic thinking;
3. **Analysis**: develop a basic understanding of the key technologies in data science: data mining, machine learning, visualization techniques, predictive modeling, and statistics;
4. **Domain knowledge**: experience discipline-specific data use cases in order to solve real-world problems of high complexity;
5. **Interpretation**: learn methods for effective data communication and visualization, and demonstrate their use in data representation;
6. **Social Value**: explore social and ethical implications of the use of data and technology.

Course Learning Outcomes: After successfully completing this course, students will:

1. experience in an academic setting the structure of the data science pipeline
2. work with an independent client with a real data science need for an entire semester
3. evaluate, research, and apply tools, and concepts necessary for a project
4. create all documentation for a project – keep Jupiter Data Science notebook
5. learn how to develop, evaluate, and communicate with a team with complementary skill sets and roles
6. learn practical skills that will help you overcome the common challenges of data science projects

Stand-up Meetings

As is done in industry, there will be stand-up meetings in this class. Instead of every day, we will have one every other week. Everyone will stand up, pass around a physical token of some sort, the person with the token will briefly state what they are currently working on/planning to work on, and any obstacles that need to be addressed. Successful participation in the stand-up meeting will contribute towards your class participation grade. Note: In the past these meetings were sometimes the first time that a student thought about what they did the previous week and what they plan to do in the coming week. This made for ineffective meetings. This semester each student and each team will be expected to **plan** what their contribution to the stand-up meeting will be **before** class and coordinate what they will say with their teammates. Each person will speak for 1 minute or less. The entire meeting will take less than 15 minutes of class time.

Jupiter Data Science Journals

Each student will keep a Jupiter Notebook, a web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text. Your Notebook will include data cleaning and transformation, numerical simulation, statistical modeling, machine learning and much more. In addition, every time you work on a project, you will make a note of the date and time, the people you worked with, and what was done. Plan to submit your Jupiter Notebook as a part of a final data product presentation.

Project

The crux of this course is a semester-long, group project that will be written for a "client" who will actually use the data product. The groups will consist of 3 to 5 students working together. Your data science product ***will*** be developed in a professional manner using data science techniques and project management strategies. The nature of the course *requires* that every team schedule *at least* 8-10 hours per week outside of class to work on the projects as a team. Any amount less than that will either not allow you to finish the project or produce a disorganized rush at the end of the course.

Exams

There will be no exams in this class. At the final exam, professional oral presentations of the projects will be made by each group. The presentations will be open to any interested parties and ***interview attire*** is required.

Note: interview attire includes jacket and tie for the males (no T-shirts!, no sneakers!, no jeans!) and a professional outfit for the females. You will NOT be allowed to present in unacceptable attire.

Requirements
Your final grade will be calculated as follows:

30 % Data product
35 % Jupiter Notebook
15 % Class participation
20 % Final Presentation

* The project grade will consist of both individual and group components. The team will receive a grade for the project and each individual will receive a grade for their contribution, participation and success in the project. This information is culled from many sources including the “boss”, the client, and fellow team members.

Schedule

Week	Description	Commitment outside the classroom
1	Meet with your client, understand and introduce your projects, and begin working with the dataset.	8-10 hours
2	Getting and Cleaning the Data. Exploratory data analysis and modeling.	8-10 hours
3	Creative exploration. Submit your milestone report and review submissions from your classmates.	8-10 hours
4	Modeling.	8-10 hours
5	Build and evaluate your prediction model - efficiency and accuracy. Milestone report.	8-10 hours
6	Evaluate the predictive accuracy, computational runtime and model complexity.	8-10 hours
7	Improve the predictive accuracy, reduce computation runtime and model complexity.	8-10 hours
8	First component of your final project, your data product.	8-10 hours
9	First component of your final project, your data product.	8-10 hours
10	Milestone report - Jupiter Notebook.	8-10 hours
11	Review your data product with your classmates.	8-10 hours
12	Second component of your final project, a Jupiter Notebook to accompany your data product.	8-10 hours
13	Second component of your final project.	8-10 hours
13	Second component of your final project.	8-10 hours
14	Final Project Submission. Submit your final project and review the work of your classmates.	8-10 hours

15	Final Presentations	

Attendance
You are part of a team. The other members of your team depend on your presence. We will be doing pair programming, so if someone misses a class, they are directly impacting the other members of their team. As such, attendance is required in this course. Any sickness must have documentation or it will count as an unexcused absence. If, for some **valid** reason you think that you must miss class, you **must** submit a written request to the instructor explaining the circumstances at least 24 hours in advance. If the instructor approves the request you may not attend on the specified day without consequence. Otherwise your class participation portion of the grade will be 20% for the first absence. If will be reduced 50% for the second absence. If you miss three classes your class participation grade will be 0. This effectively means that if everything else in the class is perfect (that has yet to happen) your course grade will be in the C range.

Academic Honesty
The nature of this class is such that plagiarizing is not a real potential, as each project will be developing a custom piece of software. Using existing Open-Source code with the guidelines of its release is a standard industry practice and will be expected. Any other violation of original work through use of unattributed code will be given an F for the course.

Students with Disabilities
In accordance with federal law, if you have a documented disability (learning, psychological, sensory, physical, or medical) you may be eligible to request accommodations from the Office of Services for Students with Disabilities (SSD). To make a request for accommodations, please contact SSD Director Allison West at (216) 397-4967 or visit the SSD office, located in Room 7A, on the garden (lower) level of the Administration Building. Please keep in mind that accommodations are not retroactive so it is best to register with SSD at the beginning of each semester. Only those accommodations approved by SSD will be recognized by your instructors. Please contact SSD if you have further questions.

Sexual Harassment and Bias
John Carroll University is committed to fostering a learning and working environment based open communication, mutual respect, and ethical and moral values consistent with Jesuit and Catholic traditions. The University seeks to provide an environment that is free of bias, discrimination, and harassment, including sexual harassment.
If you have experienced, sexual harassment/assault/misconduct/gender/sex/sexual orientation, and you share this with a faculty member, the faculty member must notify the Title IX Coordinator, Kendra Svilar, J.D., who will discuss options with you. She can be reached by email at ksvilar@jcu.edu *or (216) 397-1559. For more information about your options and resources, please go to* http://sites.jcu.edu/hr/pages/resourcespolicies/title-ix/*.*

If you have experienced bias or discrimination based on race, age, sex, sexual orientation*, religion, ethnic or national origin, disability, military or veteran status, genetic information or any factor protected by law, you are encouraged to report this to the Bias Reporting System at* ***http://sites.jcu.edu/bias*** *or to Dr. Terry Mills, Assistant Provost for Diversity and Inclusion, at* tmills@jcu.edu*, or (216) 397-4455. For more information about the University commitment to diversity and inclusion, please see* ***http://sites.jcu.edu/diversity****.*

**You can report concerns anonymously through the Bias Reporting System.*

Appendix C – Universities with multiple data-centric programs

George Mason University

- Computational and Data Science major https://cos.gmu.edu/cds/bs-in-computational-and-data-sciences-2/
- Computational and Data Science Minor https://cos.gmu.edu/cds/minor-in-computational-and-data-sciences/?_ga=1.154917693.405328958.1469653466
- Business Analytics Minor http://business.gmu.edu/undergraduate/minors/analytics/?_ga=1.127587208.405328958.1469653466
- Data Analysis Minor (6 required stats classes + 3 electives) http://statistics.gmu.edu/pages/data_analysis_minor.html?_ga=1.61093929.405328958.1469653466
- Statistics Minor http://statistics.gmu.edu/pages/statistics_minor.html?_ga=1.94009240.405328958.1469653466

Massachusetts Institute of Technology

- Business analytics major and minor http://mitsloan.mit.edu/undergrad/15-2-business-analytics/
- Statistics and Data Science minor https://stat.mit.edu/news/new-minor-in-statistics-and-data-science/

University of Michigan
Major in statistics, major in data science, major in informatics (stats+social science/ethics track)
http://lsa.umich.edu/stats/undergraduate-students/undergraduate-programs.html

Western Michigan University

- Business Analytics major https://wmich.edu/academics/undergraduate/business-analytics
- Data Science major https://wmich.edu/academics/undergraduate/datascience
- Statistics Major https://wmich.edu/academics/undergraduate/statistics

Drexel University

- Business analytics major, minor http://catalog.drexel.edu/undergraduate/collegeofbusiness/businessanalytics/
- Data Science major, minor http://drexel.edu/cci/programs/undergraduate-programs/bs-datascience/